Susan G. Orel, MD
Nicole Kay, BA
Carol Reynolds, MD
Daniel C. Sullivan, MD

Author contributions:
Guarantor of integrity of entire study, S.G.O.; study concepts, S.G.O.; study design, D.C.S., S.G.O.; definition of intellectual content, S.G.O.; literature research, S.G.O.; statistical analysis, S.G.O.; manuscript preparation and editing, D.C.S., S.G.O.; manuscript review, S.G.O., N.K., C.R., D.C.S.

BI-RADS Categorization As a Predictor of Malignancy[1]

PURPOSE: To determine the positive predictive value (PPV) of the American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) categories 0, 2, 3, 4, and 5 by using BI-RADS terminology and by auditing data on needle localizations.

MATERIALS AND METHODS: Between April 1991 and December 1996, 1,400 mammographically guided needle localizations were performed in 1,109 patients. Information entered into the mammographic database included where the initial mammography was performed (inside vs outside the institution), BI-RADS category, mammographic finding, and histopathologic findings. A recorded recommendation was available for 1,312 localizations in 1,097 patients, who composed the study population.

RESULTS: The 1,312 localizations yielded 449 (34%) cancers (139 [31%] were ductal carcinoma in situ [DCIS]; 310 [69%] were invasive cancers) and 863 (66%) benign lesions. There were 15 (1%) category 0 lesions; the PPV was 13% (two of 15 lesions). There were 50 (4%) category 2 lesions; the PPV was 0% (0 of 40 lesions). There were 141 (11%) category 3 lesions; the PPV was 2% (three of 141 lesions). The three cancers in this group were all non-comedotype DCIS. There were 936 (71%) category 4 lesions; the PPV was 30% (279 of 936 lesions). There were 170 (13%) category 5 lesions; the PPV was 97% (165 of 170 lesions).

CONCLUSION: Placing mammographic lesions into BI-RADS categories is useful for predicting the presence of malignancy. Perhaps, most important, a lesion placed into BI-RADS category 3 is highly predictive of benignity, and short-term interval follow-up as an alternative to biopsy would decrease the number of biopsies performed in benign lesions.

While screening mammography is highly sensitive for the detection of clinically occult breast cancer, most mammographically detected suspicious lesions for which biopsy is recommended will prove to be benign. In the United States, the positive predictive value (PPV) for biopsy performed because of mammographic findings, or the number of cancers detected divided by the total number of biopsies performed, is reported to be in the range of 15%–40% (1–4). The cost associated with the resultant large number of biopsies of benign lesions has been cited as one obstacle to the widespread acceptance of mammographic screening (4–7). In a study (4) of a low-cost screening project sponsored by the American Cancer Society in Orange County, Calif, it was shown that the cost of screening mammograms accounted for less than one-third of the total costs associated with breast cancer detection and diagnosis, while the cost associated with surgical consultation and biopsy of benign lesions represented the major induced cost of screening (4).

Needle localization is performed for a wide variety of lesions of varying levels of suspicion, only a minority of which are highly suspicious for malignancy (1,2). Also included are lesions believed to be benign or probably benign for which routine screening or short-term interval follow-up is recommended but for which the patient, her physician, or both request biopsy; low-suspicion lesions (ie, probable fibroadenoma) for which tissue diagnosis is suggested; and indeterminate lesions for which biopsy is recommended. The percentage of lesions that will prove to be malignant (the PPV) will vary with the level of suspicion (1,2).

The American College of Radiology has developed the Breast Imaging Reporting and Data System (BI-RADS), which is intended to standardize the terminology in the mammographic report, the assessment of the findings, and the recommendation of the action to be

taken (8). On the basis of the level of suspicion, mammographically detected lesions can be placed into one of four BI-RADS assessment categories: category 2 for benign lesions, category 3 for probably benign lesions, category 4 for suspicious lesions, and category 5 for highly suspicious lesions. With the use of the BI-RADS lexicon, lesions in which needle localization is performed can now be placed into one of the assessment categories, and the PPV can be determined for each category. The purpose of this study was to calculate the PPV for each BI-RADS assessment category to determine if the BI-RADS assessment category is a good predictor of malignancy.

MATERIALS AND METHODS

Needle localizations performed at our institution from April 1991 through December 1996 were retrospectively reviewed. Needle core biopsies performed with ultrasonographic (US) or stereotactic guidance, begun in September 1994, were not included in this study. The cases were divided into those in which the mammogram prompting localization was obtained at our institution (inside case) and those in which the mammogram was obtained at an outside institution (outside case). For inside cases, the mammographer's recommendation was recorded. For outside cases, localizations were included in this study only if the mammogram had been reviewed by one of our mammographers (including S.G.O. and D.C.S.) and a recommendation had been recorded prior to localization. The outside recommendation was not routinely recorded in our review of the outside images.

During the time of this study, the mammograms were reviewed by one of eight dedicated breast imaging radiologists (including S.G.O. and D.C.S.). A total of 1,400 localizations were performed in 1,109 patients. A recorded recommendation was available for 1,312 localizations in 1,097 patients, who made up the study population.

The mammographic findings were recorded. Because the mammograms preceding the needle localizations were interpreted prior to routine use of the BI-RADS lexicon and assessment categories (not in routine use until 1997), a BI-RADS assessment category was retrospectively assigned for each case on the basis of the report for inside cases or the inside mammographer's recorded recommendation at review of the outside images. The BI-RADS category was assigned as follows: BI-RADS category 0 for additional imaging evaluation recommended; BI-RADS category 2 for lesions classified as benign; BI-RADS category 3 for lesions classified as probably benign and for which 6-month follow-up was recommended; BI-RADS category 4 for lesions classified as suspicious or indeterminate; and BI-RADS category 5 for lesions classified as highly suggestive or diagnostic of carcinoma.

In each case, the mammographic finding, the BI-RADS assessment category, and the histopathologic result were reviewed. The PPV was defined as the number of malignant results divided by the total number of biopsies. The PPV was calculated for each BI-RADS assessment category, and the PPVs for each BI-RADS assessment category were compared by using a comparison of two proportions (9).

For cases classified as either BI-RADS category 0 or BI-RADS category 2, available surgical charts were reviewed to determine why needle localization was performed despite either the report of a benign finding (BI-RADS category 2) or the recommendation for additional imaging evaluation or comparison of the present mammogram with prior mammograms (BI-RADS category 0). The surgical charts were reviewed in 30 of 65 patients with a BI-RADS category 0 or 2 lesion.

TABLE 4
Frequency of Carcinoma according to BI-RADS Category for Different Mammographic Lesions

Mammographic Finding	BI-RADS Category				
	0	2	3	4	5
Mass (n = 596)	1/9 (11)	0/32 (0)	0/76 (0)	101/379 (27)	97/100 (97)
Microcalcifications (n = 582)	1/6 (17)	0/13 (0)	3/53 (6)	148/463 (32)	47/47 (100)
Mass with calcifications (n = 36)	. . .	0/3 (0)	0/3 (0)	10/23 (43)	7/7 (100)
Asymmetric density (n = 32)	. . .	0/1 (0)	0/3 (0)	8/26 (31)	2/2 (100)
Developing density (n = 31)	. . .	. . .	0/4 (0)	4/23 (17)	3/4 (75)
Architectural distortion (n = 29)	. . .	. . .	0/2 (0)	5/17 (29)	9/10 (90)
Architectural distortion with calcifications (n = 3)	. . .	. . .	. . .	3/3 (100)	. . .
Dilated duct (n = 3)	. . .	0/1 (0)	. . .	0/2 (0)	. . .

Note.—Data are the number of cancers detected divided by the number of localizations performed. The numbers in parentheses are the PPVs.
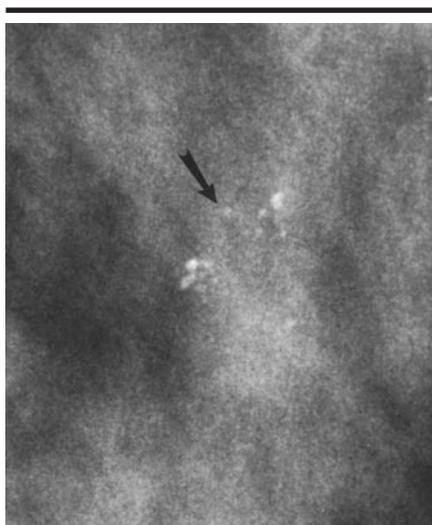
Figure 1. Calcifications had been identified on an outside a screening mammogram in a 40-year-old woman, and follow-up diagnostic mammography was performed. Spot magnification mammogram obtained in the craniocaudal projection reveals a cluster of calcifications (arrow). The interpretation was punctate, questionably new calcifications, and a 6-month follow-up examination was recommended. Needle localization was performed, and solid-type DCIS measuring 0.8 cm was found. Repeat excision revealed residual DCIS measuring 1.2 cm, with negative margins of resection.

For all cases in which a lesion was classified as BI-RADS category 3 but the biopsy result was malignancy, the mammograms were retrospectively reviewed, and the histopathologic findings were reviewed with the breast pathologist (C.R.).

RESULTS

Of the total 1,312 localizations, 793 (60%) were inside cases and 519 (40%) were outside cases. The mammographic findings are summarized in Table 1. The biopsy results are summarized in Table 2. The histopathologic result was malignancy in 449 (34%) cases and benignity in 863 (66%) cases. Of the 449 malignant cases, 139 (31%) were ductal carcinoma in situ (DCIS), 16 (4%) were invasive lobular carcinoma, and 294 (65%) were invasive ductal carcinoma.

For lesions placed into BI-RADS category 0 (work-up incomplete), 15 (1%) biopsies were performed. The PPV for category 0 lesions was 13% (two of 15 biopsies). For BI-RADS category 2 (benign) lesions, 50 (4%) biopsies were performed. The PPV for category 2 lesions was 0% (0 of 50 biopsies). For BI-RADS category 3 (probably benign) lesions, 141 (11%) biopsies were performed. The PPV for category 3 lesions was 2% (three of 141 biopsies). For BI-RADS category 4 (indeterminate) lesions, 936 (71%) biopsies were performed. The PPV for category 4 lesions was 30% (279 of 936 biopsies). For BI-RADS category 5 (highly suggestive of malignancy) lesions, 170 biopsies (13%) were performed. The PPV for category 5 lesions was 97% (165 of 170 biopsies). The PPV for BI-RADS categories 4 and 5 combined was 40% (444 of 1,106 biopsies) compared with the PPV for categories 0–3 combined of 2% (five of 206 biopsies; $P < .001$). The results were not significantly different ($P = .9$) between inside and outside cases (Table 3).

The frequencies of carcinoma according to BI-RADS category for different mammographic lesions are summarized in Table 4. The most frequent BI-RADS category 2 lesion in which needle localization was performed was a mass with or without associated calcifications, which describes 35 of 50 masses. For the 35 masses, the most frequent descriptions were stable for 1–3 years (n = 13), recurrent simple cyst (n = 10), and stable or minimally enlarging benign-appearing intramammary lymph node (n = 6). For outside cases (n = 19), the biopsy was performed on the basis of the outside recommendation. For inside cases (n = 16), the biopsy was performed at the request of the patient, her physician, or both; biopsy in eight cases was performed for recurrent cysts.

Thirteen of the 50 BI-RADS category 2 lesions in which biopsy was performed represented microcalcifications. Nine of 13 were outside cases. Biopsy was performed in these cases on the basis of outside recommendation, despite review of the outside images that showed either benign-appearing calcifications (milk of calcium, fat necrosis, probable fibroadenoma; n = 5) or stable (for at least 12 months) calcifications (n = 4). For the four inside cases, biopsy was performed due to either a cancer or a suspicious lesion elsewhere in the breast (n = 2), a prior history of ipsilateral cancer (n = 1), or a strong family history of cancer (n = 1).

Of the 15 BI-RADS category 0 lesions in which biopsy was performed, 11 were outside cases. Biopsy was performed on the basis of the outside recommendation, despite the inside review that resulted in the request for additional mammography, US, or both, or comparison of present mammogram with prior mammograms. In the four inside cases, biopsy was performed despite the recommendation for comparison with outside (prior or older) mammograms in three cases and further evaluation with US in one case.

No masses classified as BI-RADS category 3 proved to be malignant. The mammographic finding in each of the three BI-RADS category 3 lesions that proved to be malignant was calcifications. In each case, diagnostic mammograms, including magnification views, were obtained. The calcifications were described as questionably new but punctate in one case (Fig 1), new and coarse in one case (Fig 2), and coarsening (following reduction mammoplasty) in one case (Fig 3). The histopathologic finding was solid-type DCIS in each case. The lesions measured 5, 18, and 20 mm. Biopsy was performed in two of the cases in 1993 and in one of the cases in 1994.

DISCUSSION

The American College of Radiology BI-RADS was developed to provide a standardized

reporting system for mammography (8,10,11). Prior to the implementation of the BI-RADS lexicon, there was no uniformity in mammography reporting. The absence of uniformity often resulted in ambiguous reports, which left the referring health care provider in a quandary as to what management strategy was required. In addition, the absence of uniform reports made the performance of a medical audit including outcome analyses very difficult, if not impossible. One of the elements of the audit is to compare the mammographic interpretation with the outcome of biopsy or follow-up. Ambiguous or equivocal reports made it difficult to determine which mammograms were interpreted as positive and which were interpreted as negative. In addition, the degree of concern (ie, probably benign vs indeterminate vs highly suspicious) often could not be determined on the basis of the report.

The main impetus for the implementation of the BI-RADS lexicon was to eliminate the confusion surrounding mammography reports such that the findings and the recommendations would be made clear. One of the main components of the BI-RADS report is the overall impression and assessment. By assigning a final assessment category on the basis of the mammographic findings, referring health care providers can now receive reports that include both a succinct interpretation of the mammographic findings and a clear management recommendation. The assignment of final assessment categories also serves to facilitate the performance of the medical audit of mammography practices. Mammographic reports with a BI-RADS category 1 (negative), category 2 (benign finding), and category 3 (probably benign finding) are classified as negative. Mammographic reports with a BI-RADS category 4 (suspicious) and category 5 (highly suggestive of malignancy) are classified as positive. For those cases where additional evaluation is needed (BI-RADS category 0), the assessment is classified as incomplete (10).

Historically, studies in which results of mammographic needle localization are reviewed have lacked standardized mammographic terminology for either lesion description or degree of suspicion (ie, minimally suspicious vs slightly suspicious vs moderately suspicious), which makes it difficult to draw clear conclusions as to what lesions and what levels of suspicion resulted in the recommendation for biopsy (1,2). With the implementation of the BI-RADS lexicon, needle localization procedures can now be divided on the basis of the assigned BI-RADS category, and the PPV can be correlated for each category.

In our series, the overall PPV for all needle localizations was 34% (449 of 1,312 localizations). However, when the procedures are divided into BI-RADS categories, it becomes apparent that the PPV varies with the assigned category. Fifteen percent (191 of 1,312 localizations) of the needle localizations were performed when the mammographic report indicated benignity (BI-RADS category 2, 4% of 1,312 localizations) or probable benignity (BI-RADS category 3, 11% of 1,312 localizations), and the PPVs were 0% (0 of 50 localizations) for BI-RADS category 2 and 2% (three of 141 localizations) for BI-RADS category 3.

In contrast, 84% ($n = 1,106$) of the 1,312 needle localizations were performed when the mammographic finding was positive (BI-RADS category 4, 71% of all localizations; BI-RADS category 5, 13% of all localizations), and the PPVs were 30% (279 of 936 localizations) for category 4 and 97% (165 of 170 localizations) for category 5.

When comparing the outcomes of localizations performed for lesions classified as BI-RADS category 3 with those of localizations performed for lesions classified as BI-RADS category 4 or 5, there was a statistically significant difference, with a PPV of 2% (three of 141 localizations) for BI-RADS category 3 lesions compared with a PPV of 40% for BI-RADS category 4 or 5 lesions (444 of 1,106 localizations; $P < .001$).

The various BI-RADS assessment categories appear to be useful predictors of the malignancy of a lesion. Placing a lesion into BI-RADS category 0 indicates that additional work-up is required prior to rendering a final management recommendation. While these localizations represented only 1% (15 of 1,312 localizations) of all localizations, 87% ($n = 13$) of the 15 category 0 lesions proved to be benign. Review of the cases placed into BI-RADS category 0 demonstrated that, for inside cases, comparison of the present mammogram with prior (older) outside mammograms was the most frequent recommendation, although prior images were not obtained. For the outside cases, while the inside review resulted in recommendation of additional mammographic or US work-up, the decision to perform biopsy was based on the original outside report. As we did not record whether the outside mammogram leading to a biopsy recommendation represented a screening or diagnostic mammogram, we have no re-
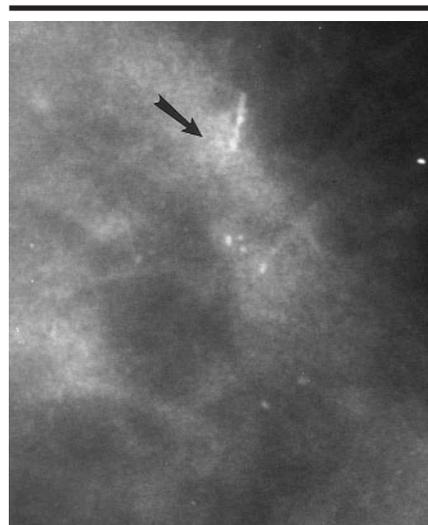
Figure 2. New calcifications had been identified on a screening mammogram in a 68-year-old woman, and follow-up diagnostic mammography was performed. Spot magnification image reveals a cluster of calcifications (arrow). The interpretation was new but coarse calcifications. There were other coarse benign calcifications elsewhere in the breast. A 6-month follow-up examination was recommended. Needle localization was performed, and solid-type DCIS measuring 0.5 cm was found, with negative margins of resection.

cord as to whether additional imaging was performed prior to the decision to recommend biopsy. While the number of BI-RADS category 0 lesions in which biopsy was performed is small, the results do suggest that by obtaining prior mammograms or performing additional imaging work-up, biopsy might have been avoided.

Placement of a lesion into BI-RADS category 2 was 100% predictive of a benign lesion. Similar to the case of BI-RADS category 0 lesions, the most frequent explanation for the biopsy of BI-RADS category 2 lesions identified on outside images was an outside recommendation for biopsy, despite the fact that inside review of these images demonstrated findings that were believed to be benign. There are several potential explanations for the discrepant interpretations of these mammograms, including interobserver variability in lesion description, variable thresholds for recommending biopsy, and variable levels of experience in interpreting mammograms.

A first potential explanation, interobserver variability in both lesion description and recommendation, may be overcome as the BI-RADS lexicon, with its standardized terminology and assessment categories, gains wider acceptance
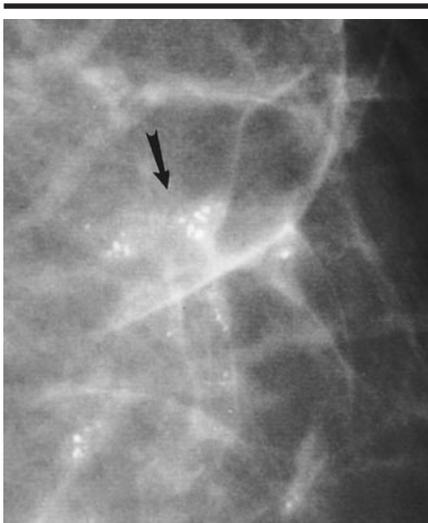
Figure 3. Outside 90° lateral spot magnification mammogram demonstrates calcifications (arrow) in a 47-year-old woman who had undergone left mastectomy and right reduction mammoplasty and who had outside images, including this image, at presentation. Review of several mammograms demonstrated coarsening of the calcifications, and the interpretation was probable fat necrosis following reduction mammoplasty. Needle localization revealed solid-type DCIS measuring 0.8 cm, with negative margins of resection.

and as lesion atlases that illustrate the various BI-RADS lesion descriptors become available.

A second potential explanation for the performance of biopsy in BI-RADS category 2 lesions was the absence of a clear recommendation in the mammographic report, both for inside cases and the reviews of outside cases. While the mammographic finding was described as benign or stable for cases retrospectively classified as BI-RADS category 2, a stated recommendation for routine follow-up may have avoided biopsy. Again, the increasing utilization of the BI-RADS lexicon with assessment categories and associated recommendations has the potential to reduce the ambiguity surrounding mammographic reports so that patients and their physicians will have a clear understanding as to where further imaging is needed, whether prior mammograms are required prior to a final recommendation, or whether the mammographic findings are benign and routine follow-up is recommended.

A third potential explanation for the performance of biopsy in BI-RADS category 2 lesions, especially for outside cases, was the presence of two conflicting reports, one recommending biopsy and one not, which led to confusion on the part of the patient and her physician as to which recommendation was correct. The results of an audit of needle localizations, such as the one performed here, can be shared with patients and their health care providers so that when either inside or outside mammograms are reviewed, the likelihood of malignancy for a lesion assigned to a given BI-RADS category can be quantified.

Perhaps one of the most important observations to be made is that for lesions in which needle localization is performed, placing a lesion into BI-RADS category 3 was highly predictive of a benign lesion. Studies reviewing mammographic follow-up of probably benign lesions have reported a very low incidence of malignancy in this group of lesions in which surveillance mammography is performed, from 0.5% to 4% (5,6,12,13). In the largest of these series, Sickles (12) reported a PPV of 0.5% for lesions classified as probably benign. In this series (12), 17 cancers were found during mammographic follow-up, and all were stage 0 or stage 1. Our results are in keeping with the results of these studies. The PPV of BI-RADS category 3 lesions in which localization was performed was 2% (three of 141 localizations), and all three lesions were solid-type DCIS. It must be noted that we are reporting only those BI-RADS category 3 lesions in which needle localization ultimately was performed. This does not enable us to determine the frequency of carcinoma in lesions classified as BI-RADS category 3 in which surveillance mammography was performed.

There are several limitations of this study. First, the results of our study reflect an audit of an academic mammography practice staffed by full-time dedicated breast imaging radiologists. All outside images in this study were reviewed by one of our radiologists (including S.G.O. and D.C.S.) prior to localization. The similar outcomes for inside and for outside cases are reflective of the review of all cases by the same radiologists. The outside radiologists' recommendations were not routinely recorded; therefore, we were unable to determine the PPV for biopsies initiated because of outside mammograms on the basis of the original interpretation. Results will vary from practice to practice. Each mammography center must perform a medical audit to determine the correlation between the BI-RADS categories and the outcome of needle localizations in its practice.

Second, this study was a retrospective analysis. The BI-RADS categories were assigned on the basis of the mammographic report. In our mammography center, prior to the advent of BI-RADS, terminology was used that paralleled the BI-RADS assessment categories, so that BI-RADS assessment categories could easily be assigned in each case.

A third, and related limitation, was our inability to retrospectively describe the morphology of the lesions such as mass margin, mass density, and type and distribution of the calcifications. Prospective studies are advantageous, as the PPV of specific mammographic features and final assessment categories can be determined. In one such recent analysis, Liberman et al (14) found that the standardized terminology of the BI-RADS lexicon does allow quantification of the likelihood of malignancy for various lesions. In that study, the features with the highest PPV were spiculated margins, irregular shape, linear morphology of microcalcifications, and segmental or linear distribution of microcalcifications. The PPVs for lesions classified as BI-RADS categories 4 and 5 were 34% and 81%, respectively, results similar to those reported in our retrospective study. However, it should be stressed that our results, along with those of Liberman et al, reflect practices staffed by dedicated mammographers with academic appointments. It is very important for radiologists who interpret mammograms to perform analyses similar to ours to determine the PPV for various lesions in each BI-RADS assessment category in their practice, because these values will most likely vary among different radiologists and different patient populations.

The overlap in the mammographic appearances of benign and malignant lesions remains one of the limitations of mammographic screening and diagnosis. When a lesion is identified at mammography, a woman may be told by her health care provider that the only way to determine if the lesion is truly benign is by excising it. However, mammographically detected lesions can now be placed into one of the BI-RADS assessment categories, each with its own PPV. There are lesions that can be classified as benign (BI-RADS category 2) that require no further evaluation. There are lesions that can be classified as probably benign (BI-RADS category 3) for which mammographic surveillance is appropriate and for which biopsy can be avoided. We must further educate our patients and our referring physicians about the BI-RADS assessment categories and the correlation between the various categories and outcome so that tissue diagnosis is reserved for those lesions that are indeterminate

(BI-RADS category 4) or highly suggestive of malignancy (BI-RADS category 5).

References

1. Hall FM, Storella JM, Silverstone DZ, Wyshak G. Nonpalpable breast lesions: recommendation for biopsy based on suspicion of carcinoma at mammography. Radiology 1988; 167:353–358.
2. Ciatto S, Cataliotti L, Distante V. Nonpalpable lesions detected with mammography: review of 512 consecutive cases. Radiology 1987; 165:99–102.
3. Linver MN, Osuch JR, Brenner RJ, Smith RA. The mammography audit: a primer for the Mammography Quality Standards Act (MQSA). AJR 1995; 165:19–25.
4. Cyrlak D. Induced costs of low-cost screening mammography. Radiology 1988; 168:661–663.
5. Helvie MA, Pennes DR, Rebner M, Adler DD. Mammographic follow-up of low suspicion lesions: compliance rate and diagnostic yield. Radiology 1991; 178:155–158.
6. Brenner RJ, Sickles EA. Acceptability of periodic follow-up as an alternative to biopsy for mammographically detected lesions interpreted as probably benign. Radiology 1989; 171:645–646.
7. Elmore JG, Barton MB, Moceri VM, et al. Ten year risk of false positive screening mammograms and clinical breast examinations. N Engl J Med 1998; 338:1089–1096.
8. American College of Radiology. Breast imaging reporting and data system (BI-RADS). 2nd ed. Reston, Va: American College of Radiology, 1995.
9. Colton T. Statistics in medicine. Boston, Mass: Little Brown, 1974; 163.
10. Bassett LW. Standardized reporting for mammography: BI-RADS. Breast J 1997; 3:207–210.
11. D'Orsi CJ. American College of Radiology mammography lexicon: an initial attempt to standardize terminology. AJR 1996; 166:779–780.
12. Sickles EA. Periodic mammographic follow-up of probably benign lesions: results of 3,184 consecutive cases. Radiology 1991; 179:463–468.
13. Varas X, Leborgne F, Leborgne JH. Nonpalpable, probably benign lesions: role of follow-up mammography. Radiology 1992; 184:409–414.
14. 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.