Effect of Clinical Information on Diagnostic Performance in Breast Sonography

Song-Ee Baek, MD, Min Jung Kim, MD, PhD, Eun-Kyung Kim, MD, PhD, Ji Hyun Youk, MD, PhD, Hye-Jeong Lee, MD, Eun Ju Son, MD, PhD

Objective. The purpose of this study was to assess whether the clinical information (CI) of patients affects the degree of suspicion for malignancy by radiologists performing breast sonography. ***Methods.*** We included 150 breast lesions in 144 patients who underwent breast sonography and sonographically guided core needle biopsy. A pathologic diagnosis was available for all 150 breast lesions: 78 (52%) were malignant, and 72 (48%) were benign. Three radiologists retrospectively reviewed the sonograms of all lesions twice at 8-week intervals first without any CI for the patients (first review) and then with CI such as patient age, palpability, and personal history of risk factors for breast cancer (second review). The reviewers categorized the final assessment according to the American College of Radiology Breast Imaging Reporting and Data System. We compared diagnostic performance such as sensitivity and specificity and the degree of suspicion for malignancy between the image reviews with and without CI. ***Results.*** In the second review, sensitivity was improved in all 3 reviewers (94.0 to 99.2%; $P < .05$), and specificity was decreased (39.8 to 30.8%; $P = .04$). There was a significant increase of suspicion for malignancy with the patients' CI ($P < .05$). ***Conclusions.*** Clinical information about a patient's breast cancer history and clinical presentation with a palpable mass can increase the suspicion for malignancy on sonography and the sensitivity of sonographic interpretation.
Key words: Breast Imaging Reporting and Data System lexicon; breast sonography; clinical information.

Abbreviations
ACR, American College of Radiology; A_z, area under the curve; BI-RADS, Breast Imaging Reporting and Data System; CI, clinical information; NPV, negative predictive value; PPV, positive predictive value; ROC, receiver operating characteristic

Received March 5, 2009, from the Department of Radiology, Research Institute of Radiological Science, Yonsei University College of Medicine, Seoul, Korea. Revision requested March 30, 2009. Revised manuscript accepted for publication May 4, 2009.

Address correspondence to Min Jung Kim, MD, PhD, Department of Radiology, Yonsei University Heath System, 250 Seongsanno, Seodaemun-gu, Seoul 120-752, Korea.
E-mail: mines@yuhs.ac

The American College of Radiology (ACR) developed the Breast Imaging Reporting and Data System (BI-RADS) and thereby established the criteria for the terminology and evaluation range for lesions detected on mammography and sonography.[1,2] The BI-RADS recommended consideration of clinical information (CI) and physical examination findings in interpreting the results of imaging studies.[2] To date, most studies have focused on sonographic features of lesions.[1,3,4] According to our review of the literature, however, no studies have examined whether CI such as a history of breast cancer and symptoms of palpable lesions can affect the assessment of lesions detected on sonography.

Given this background, we evaluated whether there was a significant difference in diagnostic value depending on the method of interpretation of sonographic findings associated with the availability of CI. We also studied whether CI could affect the interpretation of sonographic findings.

© 2009 by the American Institute of Ultrasound in Medicine

Materials and Methods

Our Institutional Review Board approved this study and waived the informed consent requirement because it was retrospective.

Case Selection

Between January 2004 and March 2004, sonographically guided core needle biopsy was performed for 236 lesions in 224 patients at our institution. One radiologist reviewed the images and follow-up results for all 224 patients and excluded 80 patients with 86 histopathologically benign breast lesions that did not undergo subsequent larger tissue sampling such as excision or sufficient imaging follow-up. The remaining 150 histopathologically proven lesions in 144 patients were malignant on histopathologic examination or histopathologically benign and showed no appreciable change in size on sonography for at least 2 years. These 150 lesions from 144 patients made up the study population.

The sizes of the 150 lesions ranged from 4 to 40 mm (mean, 13.5 mm). Of them, 78 (52.0%) cases were malignant, and 72 (48.0%) were benign. Patient ages ranged from 23 to 73 years (mean, 45.3 years). The number of patients with a personal history of risk factors for breast cancer was 19 (13%). Risk factors for breast cancer included a history of breast cancer diagnosis, a family history of breast cancer, breast cancer 1 gene positivity, nulliparity, and a history of lobular carcinoma in situ. The number of patients with a palpable mass was 58 (39%).

Sonographically guided core needle biopsy was performed with a 5- to 12-MHz linear probe, an automated gun (Pro-Mag 2.2; Manan Medical Products, Inc, Northbrook, IL), and a 14-gauge Tru-Cut needle with a 22-mm throw (SACN biopsy needle; Medical Device Technologies, Inc, Gainesville, FL) under guidance from an HDI 5000 sonography unit (Philips Healthcare, Bothell, WA) by 1 of 4 radiologists who had between 1 and 7 years of clinical experience in breast imaging studies.

Image Review

To interpret the images, 1 radiologist who selected the study population converted the transverse and longitudinal sonograms of each lesion to the 300-dpi tagged image file format and then arbitrarily arranged them in PowerPoint XP (Microsoft Corporation, Redmond, WA), thus preparing 2 files. In these 2 PowerPoint files with embedded tagged image file format images, the order and number of each lesion were identical, and only 1 file contained CI (patient age, palpability, and history). Three radiologists reviewed each image file twice at 8-week intervals. The file without CI was reviewed first. The radiologists were asked to select the category of each lesion as 3, 4a, 4b, 4c, or 5 based on the BI-RADS. Reviewers were blinded to histopathologic findings and the ratio of malignant to benign lesions included in this study while they evaluated the files. They were also unaware that the second review session would include the same cases with CI provided because the information for the second session might have affected the strictness of the diagnostic criteria. Eight weeks later, files containing the age, medical history, and symptoms of the patients were evaluated. Records that were made in the first session were not permitted to be reviewed.

Three radiologists who reviewed those 2 files were all board certificated. They had an average of 5.3 years (2, 4, and 10 years) of experience in performing breast sonography and interpreting mammography and breast sonography. The numbers of sonographic examinations performed and interpreted by the radiologists in their own practices varied from 200 to 350 examinations per month.

Data Analysis

We assessed whether the diagnostic sensitivity, specificity, and area under the curve (A_z) of receiver operating characteristic (ROC) analysis changed between the first and second reviews for each reviewer and all reviewers using histopathologic examination as the reference standard. For this purpose, the category of final assessment was dichotomized as negative (category 3) or positive (categories 4a, 4b, 4c, and 5). Parametric estimates of A_z were calculated by the probability of malignancy based on a 5-point scale (categories 3, 4a, 4b, 4c, and 5) and compared for reader performance in the 2 sessions by Dorfman-Berbaum-Metz multiple readers and multiple cases analysis.[5,6] The statistical significance of the results was reported at 95% confidence intervals for the mean differences

in A_z values for reader performance in the 2 sessions. The mean differences were regarded as statistically significant at the 5% level when the corresponding confidence interval did not encompass 0. Additionally, intergroup differences in the final assessment of lesions were analyzed by the Wilcoxon signed rank test. This was also determined on the basis of the criteria that $P < .05$ was considered statistically significant.

All of the cases were divided into 3 subgroups according to CI: a subgroup with a history of risk factors for breast cancer, a subgroup with palpability, and a subgroup without a history of risk factors for breast cancer or palpability. In each subgroup, the patterns of changes in diagnostic sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were evaluated for the first and second sessions. We also assessed whether the increased degree of suspicion for malignancy would be different according to CI such as palpability and a history of breast cancer by a χ^2 test.

Statistical analysis, except for ROC analysis, was performed with SPSS version 10 for Windows software (SPSS Inc, Chicago, IL).

Results

First Versus Second Review

Table 1 shows the changes in diagnostic sensitivity, specificity, and A_z values for the first and second reviews. Compared with the first review, the second review showed increased sensitivity ranging from 1.3% to 9% (mean, 5.2%) for each reviewer. Two reviewers showed sensitivity of 100%. Specificity was decreased from 2.8% to 18.0% for all reviewers (Figure 1). Overall sensitivity was increased from 94% to 99.2%, and overall specificity was decreased from 39.8% to 30.8%. These differences reached statistical significance ($P < .05$). For the A_z values of the ROC analysis, all reviewers showed an increase in A_z from 0.0241 to 0.0621 (mean difference, 0.0385; $P = .051$).

Compared with the first review, the probability of malignancy of lesions in the second review showed that the suspicion for malignancy was raised to a significant degree ($P < .05$; Figure 2). As shown in Table 2, category 3 was decreased by 7.1% (100 to 68) in the second review. On the other hand, categories 4 and 5 were increased by 5.1 and 2%, respectively. Among 100 lesions that were initially classified as category 3 without CI, 32 cases were classified as higher than category 4 after the provision of CI. Among them, 13 cases were diagnosed as malignant (Figure 3). Furthermore, categories 4a and 4c were increased by 2.2% and 4.0%, respectively, after the availability of CI. However, category 4b showed a 1.1% decrease.

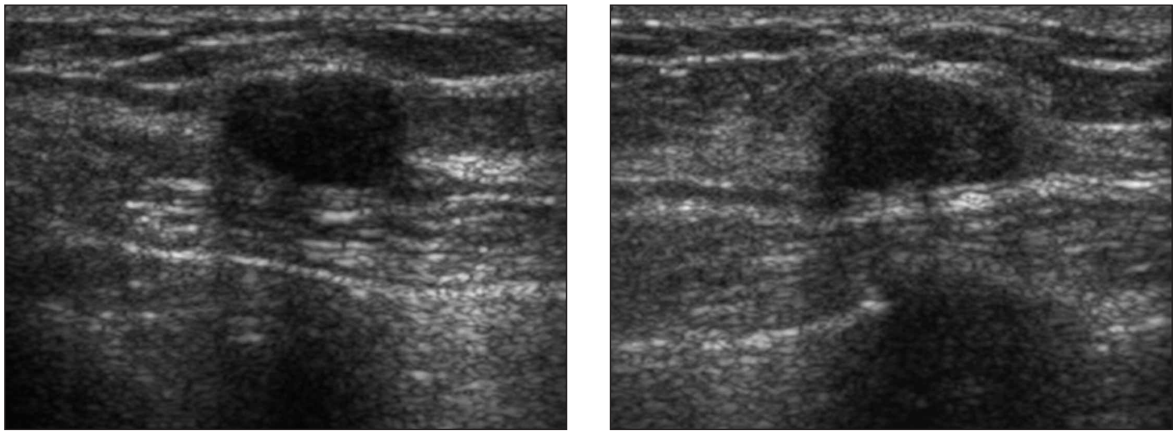
Subgroups According to CI

After CI was provided to the reviewers, the sensitivity and NPV were increased and the specificity and PPV were decreased in all subgroups ($P > .05$; Table 3). Of 57 cases in which a history of risk factors for breast cancer was present, suspicion of malignancy increased in 21 in the second review (Table 4 and Figures 4 and 5). Among 13 cases in which the initial category was 3, categorization was increased in 10 (76.9%) as category 4a in 9 and category 4c in 1 in the second review (Figure 4). Of these cases, 8 were shown to be malignant. Only in 2 of 57 cases was the category decreased (4c to 4b and 4a to 3). Those 2 cases were confirmed by histopathologic examination as malignant and benign, respectively.

Table 1. Diagnostic Performance With and Without CI

	Sensitivity (n = 78)		Specificity (n = 72)		A_z		
Reviewer	Without CI	With CI	Without CI	With CI	Without CI	With CI	95% Confidence Interval of Mean Difference
1	92.3 (72)	97.4 (76)	48.6 (35)	30.6 (22)	0.8827	0.9068	–0.06631 to 0.01813
2	98.7 (77)	100.0 (78)	36.1 (26)	33.3 (24)	0.9225	0.9519	–0.09113 to 0.03245
3	91.0 (71)	100.0 (78)	34.7 (25)	30.6 (22)	0.8729	0.9350	–0.12741 to 0.00325
Overall	94.0	99.2[a]	39.8	30.8[a]	0.8927	0.9312[b]	–0.07720 to 0.00019

Sensitivity and specificity values are percentages with raw data in parentheses. Clinical information included patient age, breast pain, lesion palpability, history of breast cancer, familial history of breast cancer, breast cancer 1 gene positivity, and history of lobular carcinoma in situ.
[a] $P < .05$.
[b] $P = .051$.



A B

Figure 1. Images from a 42-year-old female patient with a well-defined hypoechoic ovoid lesion on sonography. All 3 reviewers classified it as category 3. After CI revealed that the patient had lesion palpability, the 3 reviewers evaluated this lesion as category 3, 4a, and 4a, respectively. The lesion was proven to be a fibroadenoma at biopsy. **A**, Transverse plane. **B**, Longitudinal plane.

In cases with palpability, the suspicion of malignancy was increased after CI was provided in 47 cases (Table 4). However, it was decreased in 22 cases. Of 35 cases with an initial category of 3, 16 had a category higher than 3 in the second review (45.7% [16 of 35]; Figure 4). Of these cases, 4 (25% of 16) were shown to be malignant.

Particularly, a history of risk factors for breast cancer tended to increase suspicion for malignancy (76.9%) more than palpability (45.7%), but there was no statistical significance ($P > .05$). There were 6 cases in which a history of risk factors for breast cancer as well as palpability were present. Among them, 1 lesion that was evaluated as category 3 in the first review was eventually classified as 4c by 1 of 3 radiologists in the second review. The histopathologic result of this case was malignancy. The initial probabilities of malignancy for the remaining 5 cases were category 4a in 2, category 4c in 1, and category 5 in 2. In the second review, however, 1 of 2 category 4a cases was converted to category 4c, and the category 4c case was converted to category 5. All 6 cases with a history of risk factors for breast cancer as well as palpability were malignant.

Figure 2. Images from a 61-year-old female patient with an irregularly shaped, nonparallel, hypoechoic lesion with microlobulations on sonography. All 3 reviewers classified this lesion as category 4a. The patient had a history of contralateral ductal carcinoma in situ. The reviewers saw the partially spiculated margin (arrows) and upgraded the category to 4a, 4b, and 4c after CI was provided. The lesion was confirmed as ductal carcinoma in situ at surgery. **A**, Transverse plane. **B**, Longitudinal plane.

A B

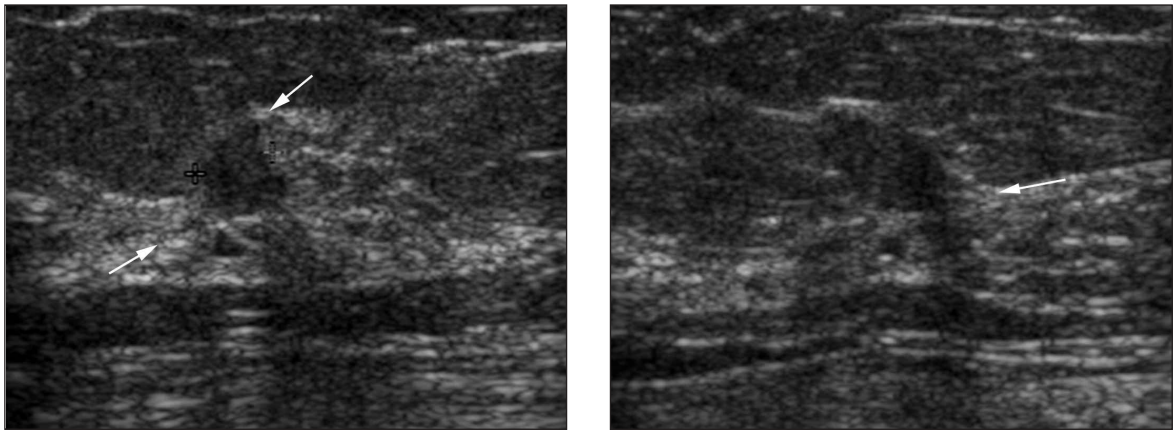


Table 2. Degree of Suspicion for Malignancy With and Without CI

Category	Without CI	With CI
3	100 (22.2)	68 (15.1)
4	276 (61.4)	299 (66.5)
4a	182 (40.5)	192 (42.7)
4b	26 (5.8)	21 (4.7)
4c	68 (15.1)	86 (19.1)
5	74 (16.4)	83 (18.4)
Total	450 (100.0)	450 (100.0)

Values are raw data with percentages in parentheses.
CIinical information was as in Table 1.
$P < .05$.

Of 225 cases in which there was neither a personal history of risk factors for breast cancer nor a palpable mass, 8 classified as category 4a were reevaluated as category 3 in the second review. All of these cases were shown to be benign on histopathologic examination.

Discussion

The ACR recommends that clinical data and physical examination findings should be considered when the BI-RADS is used for the interpretation of sonographic findings.[2] To date, however, no studies to our knowledge have elucidated whether CI of patients can affect the interpretation of sonographic findings.

In our series, CI, including patient age, the presence of a palpable mass, and a history of risk factors for breast cancer, raised sensitivity and lowered specificity in the interpretation of sonographic findings ($P < .05$) and tended to raise the A_z value in the ROC analysis ($P = .051$).

For both mammographic and sonographic findings, BI-RADS category 3 applies to lesions that are not palpable. There is still controversy as to whether this criterion can also be applied to palpable lesions.[2,7,8] According to Graf et al,[8] a 6-month follow-up observation rather than a histopathologic examination would be appropriate for solid lesions corresponding to category 3, in which there were no findings suggestive of malignancy on sonography or mammography and no calcification present despite the presence of a palpable mass. In our study, despite the recognition of the presence of a palpable mass, 3 reviewers adhered to category 3 in 19 cases and raised the suspicion of malignancy in 16 (Figures 4 and 5). If physicians have information about the presence of palpability during examination of a lesion, they could be more conscious of a suspicious finding. Although palpability is not a suspicious finding in itself, a record of palpability could make the radiologist more alert, possibly resulting in leading the observer to search for subtle but suspicious findings for the lesion that were overlooked. Of these 16 cases that were reclassified as having a higher probability of malignancy in the second review, 4 cases were indeed malignant.

Figure 3. Images from a 32-year-old female patient with a hypoechoic ovoid breast lesion on sonography. At first, the reviewers categorized this lesion as category 4a, 3, and 3. After CI revealed that the patient had a history of contralateral breast cancer, all of the reviewers upgraded the category of this lesion to 4a because of the presence of microlobulations (arrows) in some areas of the lesion. This lesion was proven to be invasive ductal carcinoma. **A**, Transverse plane. **B**, Longitudinal plane.

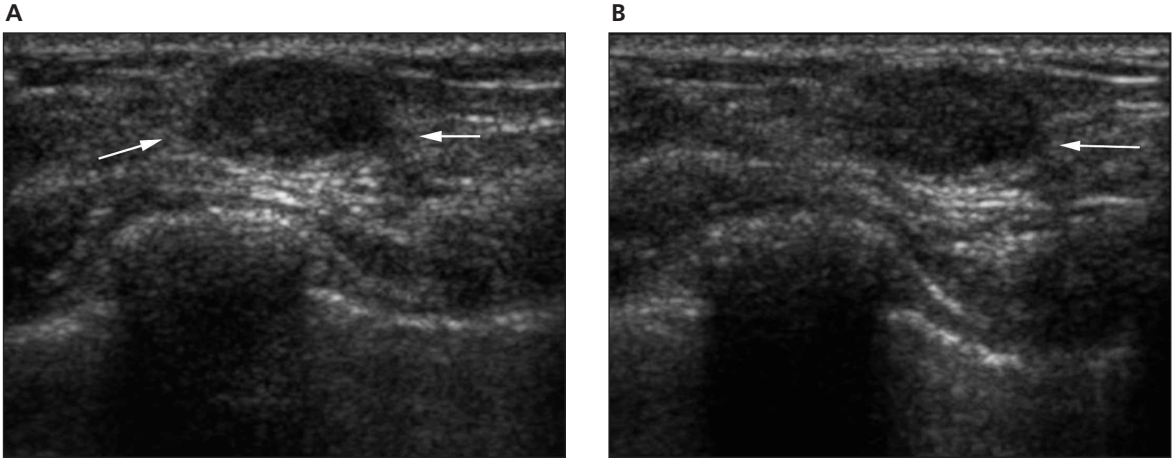


Table 3. Diagnostic Performance According to CI

Parameter	With Age and History of Risk Factors for Breast Cancer (n = 57)[a]		With Age and Palpability (n = 174)[a]		With Age but Without Other CI (n = 225)	
	Without CI	With CI	Without CI	With CI	Without CI	With CI
Sensitivity	72.7 (24/33)	97.0 (32/33)	97.4 (111/114)	100.0 (114/114)	96.8 (90/93)	98.9 (92/93)
Specificity	16.7 (4/24)	12.5 (3/24)	53.3 (32/60)	35.0 (21/60)	37.9 (50/132)	31.8 (42/132)
PPV	54.5 (24/44)	60.4 (32/53)	79.6 (111/139)	74.5 (114/153)	52.3 (90/172)	50.5 (92/182)
NPV	30.8 (4/13)	75 (3/4)	91.4 (32/35)	100.0 (21/21)	94.3 (50/53)	97.7 (42/43)

Values are percentages with raw data in parentheses. Clinical information was as in Table 1.
[a]Six cases with a history of breast cancer and palpability were included in these groups.

Furthermore, if the patient had a history of breast cancer or a history of risk factors for breast cancer, images could be interpreted with more caution because of the increased risk of breast cancer development compared with a patient without such a history. Of 10 cases in which the category was increased from 3 to higher than 4a (Figure 4), 8 proved to be malignant on histopathologic examination. This suggests that the interpretation should be made by increasing sensitivity when a patient has risk factors for or a past history of breast cancer rather than when a patient has no such history. This would be consistent with the evaluation of synchronous breast nodules detected in patients with breast cancer.[9] Kim et al[9] reported that uniform application of the current sonographic BI-RADS classification can lead to underestimation of the risk of malignancy in a patient with breast cancer. Regarding the percentage of cases in which the final assessment was initially classified as category 3 but raised after the provision of CI, the percentage related to palpability (45.7%, 16 of 35 category 3 lesions in Figure 4) was lower than that related to a history of breast cancer (76.9%, 10 of 13 category 3 lesions in Figure 4; $P > .05$). This might suggest that an emphasis could be placed on a history of breast cancer rather than the presence of a palpable mass in the interpretation of imaging findings. However, further study with a larger series should be done to prove this.

Category 4 is used to classify lesions as 4a, 4b, and 4c based on the degree of malignancy observed on sonography.[2] However, category 4b is a gray zone between 4a and 4c in the implication of possible malignancy, and the management of category 4b lesions is indeterminate. Radiologists would put a lesion in category 4b when they are not sure about their interpretations. Therefore, some investigators have suggested that dividing category 4 into 2 subcategories (ie, 4a and 4b) may be much simpler and more useful.[10,11] Our study showed that the number of cases in which category 4b was assigned decreased after CI was given. This could indicate that the CI of the patients was helpful to reviewers for a more reliable interpretation of the results. Area under the curve values in the ROC analysis tended to improve in the second review, although the difference was not statistically sig-

Table 4. Change in Category Classification According to CI

Parameter	With Age and History of Breast Cancer (n = 57)[a]		With Age and Palpability (n = 174)[a]		With Age but Without Other CI (n = 225)	
	Without CI	With CI	Without CI	With CI	Without CI	With CI
3	13 (22.8)	4 (7.0)	35 (20.1)	21 (12.1)	53 (23.6)	43 (19.1)
4	40 (70.2)	46 (80.7)	96 (55.2)	102 (58.6)	143 (63.6)	154 (68.5)
4a	33 (57.9)	33 (57.9)	48 (27.6)	57 (32.8)	103 (45.8)	103 (45.8)
4b	1 (1.8)	4 (7.0)	9 (5.2)	10 (5.7)	16 (7.1)	7 (3.1)
4c	6 (10.5)	9 (15.8)	39 (22.4)	35 (20.1)	24 (10.7)	44 (19.6)
5	4 (7.0)	7 (12.3)	43 (24.7)	51 (29.3)	29 (12.8)	28 (12.4)

Values are raw data with percentages in parentheses. Clinical information was as in Table 1.
[a]Six cases with a history of breast cancer and palpability were included in these groups.

nificant. This may further assist clinicians in planning treatment and follow-up observations because improved diagnostic performance could enable clearer communication between the radiologist and the clinician.

The ACR BI-RADS recommends that the final diagnosis be made on the basis of the most suspicious findings, considering both mammography and sonography and a comparison of previous images and clinical data.[2] Although mammographic findings are important in breast evaluation, in this study, they were not included in the image files, which could have affected the reviewers' interpretations and could have been a limitation. It is, indeed, very unusual in practice to interpret sonograms without CI or mammograms. Sonography of the breast alone without accompanying mammography should be reserved for younger patients in whom there is no suspicion of malignancy. Nevertheless, the effect of CI and its importance in the interpretation of images in this study could make a beginner or trainee in breast imaging alert to obtaining patients' CI. We suggest that computer-aided diagnosis on sonography should be done with consideration of CI and mammograms, although further study of the impact of CI on mammographic interpretation and the combined interpretation of mammography and sonography should be done. Additionally, only 2 types of static images, transverse and longitudinal, rather than full-volume image sets or real-time images, were used in the final assessment of the lesions in this study, which might have compromised the results. In the actual clinical setting, when sonography is performed, more imaging data can be obtained. Finally, CI about detection of palpable masses was given to radiologists as "present" or "absent" in this study. In the clinical setting, such criteria as the mobility of a palpable mass and the tactile sense of the mass can affect the interpretation of imaging results. Moreover, additional CI, such as whether the palpable mass was new or increasing and the duration of palpability, can be also helpful. This information was not differentially considered in this study.

This retrospective study included cases with core biopsy data from early 2004. Two of 3 radiologists who participated in the image reviews were working at the hospital where the cases were collected at the time that they were biopsied. Therefore, we were worried that the radiologists might have recognized the cases and CI. Following an appropriate study protocol, image reviews should be done without CI. Thus, we gathered cases from more than 5 years ago in an

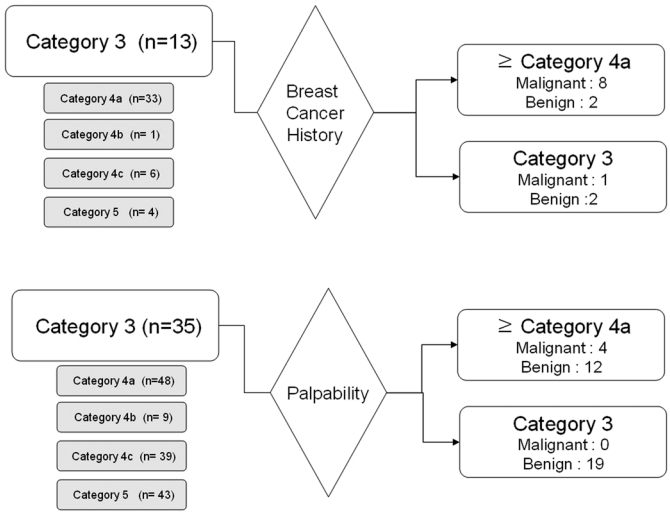
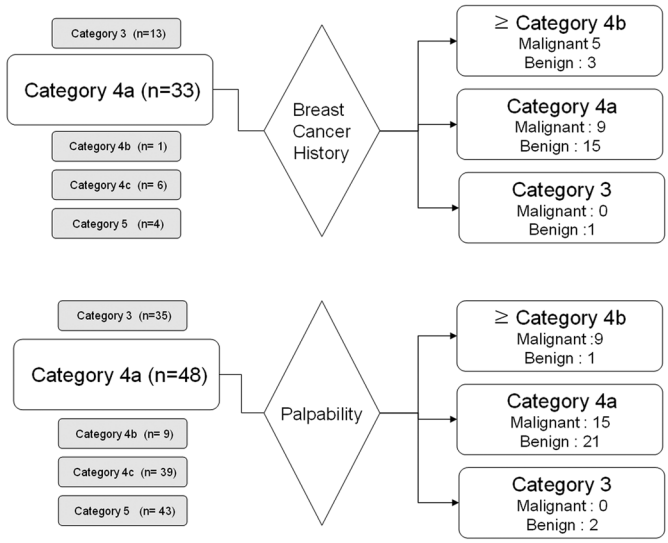


Figure 4. Changes in final assessment with consideration of CI in 48 cases that were initially classified as category 3 without CI among 231 cases (number of lesions [77] × number of reviewers [3]) with a history of breast cancer or palpability.

Figure 5. Changes in final assessment with consideration of CI in 71 lesions that were initially classified as category 4a without CI among 231 cases (number of lesions [77] × number of reviewers [3]) with a history of breast cancer or palpability.



attempt to reduce bias caused by recollection of cases by the radiologists. Although the images were taken about 5 years ago, the image quality of the cases included in this study was good and well preserved. The sonography system used for evaluation of the cases was a high-resolution and high-quality unit that is currently used in our practice. Images were stored on a picture archiving and communications system, not film. The picture archiving and communications system was used to prevent damage to images; therefore, we think that the quality of images was not a factor.

To summarize, CI such as a history of risk factors for breast cancer and palpability could affect the diagnostic performance of sonography and would give radiologists a chance to search for subtle but suspicious findings in the interpretation of lesions on sonography. These results suggest that the final assessment based on the BI-RADS should be made together with CI as well as imaging findings.

References

1. Hong AS, Rosen EL, Soo MS, Baker JA. BI-RADS for sonography: positive and negative predictive values of sonographic features. AJR Am J Roentgenol 2005; 184:1260–1265.
2. American College of Radiology. Breast Imaging Reporting and Data System. Reston, VA: American College of Radiology; 2003.
3. Costantini M, Belli P, Lombardi R, Franceschini G, Mulè A, Bonomo L. Characterization of solid breast masses: use of the sonographic Breast Imaging Reporting and Data System lexicon. J Ultrasound Med 2006; 25:649–659.
4. Paulinelli RR, Freitas-Júnior R, Moreira MA, et al. Risk of malignancy in solid breast nodules according to their sonographic features. J Ultrasound Med 2005; 24:635–641.
5. Kurt Rossmann Laboratories for Radiologic Image Research website. Chicago, IL: University of Chicago. http://www-radiology.uchicago.edu/krl/KRL_ROC/software_index.htm. Accessed July 22, 2008.
6. Wagner RF, Metz CE, Campbell G. Assessment of medical imaging systems and computer aids: a tutorial review. Acad Radiol 2007; 14:723–748.
7. Shin JH, Han BK, Ko EY, Choe YH, Nam SJ. Probably benign breast masses diagnosed by sonography: is there a difference in the cancer rate according to palpability? AJR Am J Roentgenol 2009; 192:W187–W191.
8. Graf O, Helbich TH, Fuchsjaeger MH, et al. Follow-up of palpable circumscribed noncalcified solid breast masses at mammography and US: can biopsy be averted? Radiology 2004; 233:850–856.
9. Kim SJ, Ko EY, Shin JH, et al. Application of sonographic BI-RADS to synchronous breast nodules detected in patients with breast cancer. AJR Am J Roentgenol 2008; 191: 653–658.
10. Lee HJ, Kim EK, Kim MJ, et al. Observer variability of Breast Imaging Reporting and Data System (BI-RADS) for breast ultrasound. Eur J Radiol 2008; 65:293–298.
11. Stavros AT. Introduction to breast ultrasound. In: Stavros AT (ed). Breast Ultrasound. 1st ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2004:850–856.