Medicine®

OPEN

Application of the downgrade criteria to supplemental screening ultrasound for women with negative mammography but dense breasts

Soo-Yeon Kim, MD, PhD, Min Jung Kim, MD, PhD, Hee Jung Moon, MD, PhD, Jung Hyun Yoon, MD, PhD, Eun-Kyung Kim, MD, PhD[*]

Abstract

We investigated whether the application of the downgrade criteria to supplemental screening ultrasound (US) for women with negative mammography but dense breasts can reduce the rate of Breast Imaging Reporting and Data System (BI-RADS) categories 3 to 4a without a loss of cancer detection.

This retrospective study was approved by the Institutional Review Board, and the need to obtain informed consent was waived. A total of 3171 consecutive women (978 women, 1173 women, and 1020 women in the first, second, and third year, respectively) with negative mammography but dense breast who underwent radiologist-performed, hand-held supplemental screening US from March 2010 to February 2013 were included. Downgrade criteria for BI-RADS category 2 were complicated cysts ≤5 mm observed as circumscribed, homogeneous, and hypoechoic lesions and circumscribed oval-shaped solid masses ≤5 mm. Changes in the distribution of BI-RADS category, biopsy rate, and cancer detection yield over 3 years were analyzed. Performances of less-experienced (12 fellows with <2 years of experience) and experienced (3 staffs with >12 years of experience) radiologists were compared. Outcomes of initial examinations (prevalence screening) and noninitial examinations (incidence screening) were compared.

Application of the downgrade criteria reduced BI-RADS categories 3 to 4a in both less-experienced (from 39.4% to 16.0%, $P < 0.001$) and experienced radiologists (from 22.6% to 11.1%, $P < 0.001$) over 3 years. Biopsy rates also significantly decreased from 6.5% to 2.4% ($P < 0.001$). Cancer detection yield of supplemental screening US was 2.8 per 1000 examinations (9 of 3171: 2 ductal carcinoma in situ and 7 invasive cancers). There were no differences in cancer detection yield per each year ($P = 0.539$). There was no interval cancer. In noninitial examinations, BI-RADS categories 3 to 4a rates, biopsy rates, and cancer detection rates were lower compared to initial examinations.

Application of the downgrade criteria reduced BI-RADS categories 3 to 4a without a loss of cancer detection. We suggest that our downgrade criteria can be used to reduce the false positive rate in the supplemental screening US. Further large-scale, multicenter, prospective studies are needed to validate the effectiveness of the downgrade criteria.

Abbreviations: BI-RADS = Breast Imaging Reporting and Data System, BRCA = breast cancer susceptibility gene, DCIS = ductal carcinoma in situ, HR = hormone receptor, HER2 = human epidermal growth factor receptor 2, IHC = immunohistochemistry, PPV = positive predictive value, US = ultrasound.

Keywords: breast imaging reporting and data system, dense breast, downgrade criteria, mammography, supplemental screening ultrasound

1. Introduction

Mammography is the only screening modality that has been proven to reduce mortality caused by breast cancer.[1] However, the sensitivity of mammography decreases in dense breasts, and invasive cancer is often mammographically subtle or occult when the breast is dense.[2] These limitations have led to the introduction of breast density notification legislation in the United States and the increasing demand for a supplemental screening tool.[3] Breast ultrasound (US) is an attractive screening tool which is widely available, well tolerated by patients, and without ionizing radiation. Supplemental screening US in women with dense breasts and/or the elevated risk of breast cancer has been shown to detect additional cancers by an average of 4.2 cancers per 1000 women screened.[4–11] Most of the cancers identified by supplemental US are less than 1 cm in size, invasive, and node-negative.[4–6,8,9,11,12] However, screening US has resulted in a high rate of positive test results (Breast Imaging Reporting and Data System [BI-RADS] final assessment categories 3–5).[4–8,13–15] The average positive predictive value (PPV) for biopsies was only 9.5% with increased false positive biopsies which may be associated with discomfort, emotional stress, and medical cost.[16] In addition, operator dependence has long been a concern for hand-held breast US when even performed by experienced

physicians, but few studies have evaluated the interoperator variability of physician-performed hand-held US.[17,18]

In our 2009 results, radiologist-performed, hand-held supplemental screening US was shown to detect additional cancers (1.8 cancers per 1000 examinations [3 of 1656]) in women with negative but dense breasts, but there were concerns regarding the high rates of category 3 lesions (30.4%, 504 of 1656) compared to mammography alone, and a low PPV of less than 3%.[9] Since March 2010, to reduce the false positive rate, the downgrade criteria which classify complicated cysts 5 mm or smaller which are observed as circumscribed, homogeneous, and hypoechoic lesions 5 mm or smaller and circumscribed oval-shaped solid masses 5 mm or smaller as category 2 were established and were applied in our daily practice. These cysts and masses are often seen on screening US, and according to BI-RADS, they are assigned to category 3, probable benign finding, and then undergo short-term follow-up examinations.[19] Some complicated cysts may be assigned to category 4a, low-suspicious finding. These cysts are then aspirated or biopsied, because they can sometimes look like hypoechoic solid lesions with indistinct or microlobulated margins and mild posterior shadowing, a mimic of malignancy.[20] There may be a higher possibility of less-experienced radiologists assigning them to category 4a. However, a breast cancer presenting as a complicated cyst or circumscribed oval-shaped solid mass 5 mm or smaller is very rare, with reported malignancy rates being less than 0.5%.[9,20,21] Thus, we expected the categories 3 to 4a rate to be reduced without a significant loss of cancer detection, if the cysts or masses are downgraded to category 2.

Therefore, the purpose of this study was to investigate whether the application of the downgrade criteria to supplemental screening US for women with negative mammography but dense breasts can reduce BI-RADS categories 3 to 4a without a loss of cancer detection.

2. Methods

2.1. Study population

The study was approved by the Institutional Review Board, and informed consent requirement was waived for this retrospective study. From March 2010 to February 2013, 31,373 consecutive women underwent mammograms in our institution. Using the database of the radiology department, we searched for women who met the following criteria. The inclusion criteria were as follows: women who underwent screening mammography, women who had dense breast defined as BI-RADS density grade 3 (heterogeneously dense) or 4 (extremely dense) at mammography,[19] women who had negative findings defined as BI-RADS final assessment category 1 or 2 at mammography,[19] and women who had radiologist-performed, hand-held supplemental US examinations performed within 3 months after mammography. The exclusion criteria were as follows: women who had redundant US examinations which were performed twice or more during a 1-year period due to early visits despite a BI-RADS category 1 or 2 assessment on a prior examination, women with known risk factors for breast cancer other than dense breast (biopsy-proven lobular neoplasia or atypical ductal hyperplasia, history of ovarian cancer, family history of breast cancer; there was no case with chest irradiation history and breast cancer susceptibility gene (BRCA) mutation was not considered because BRCA mutation analysis had not been routinely performed at our institution), and women who did not undergo surgery nor follow-

up US examinations for at least 12 months. Finally, a total of 3171 women with negative mammography but dense breast who underwent radiologist-performed, hand-held supplemental screening US were included (mean age±standard deviation [years], 51.2±7.7; range [years], 24–78): 978 women (50.7±7.6; 27–75) in the first year (from March 2010 to February 2011), 1173 women (50.9±7.6; 24–78) in the second year (from March 2011 to February 2012), and 1020 women (52.0±7.9; 31–78) in the third year (from March 2012 to February 2013). For each year, the women were divided into 2 groups; women who underwent initial US examinations at our institution (prevalence screening) or who underwent previous US examinations at our institution (incidence screening) to test the following hypothesis. We hypothesized that the cancer detection rate, positive test rate, and biopsy rate would be lower while PPV would be higher in noninitial examinations compared to initial examinations for the following reasons: in noninitial examinations, women with cancer and/or positive tests at previous US examination had already been excluded from the study population, and the degree of attention or concentration of the radiologists performing US may be lower for the noninitial examinations compared to initial examinations.

2.2. Screening mammography and US

Digital mammography was performed with a full-filled digital mammography system (Lorad/Hologic Selenia, Lorad/Hologic, Danbury, CT; SENOGRAPHE 2000D, GE Medical Systems, Milwaukee, WI). Standard mediolateral oblique and craniocaudal views were routinely obtained. All mammograms were interpreted by 1 of 15 breast radiologists (12 fellows with 1–2 years of experience [less-experienced group] and 3 faculties with 12–18 years of experience [experienced group]). The final assessment was prospectively determined according to BI-RADS.[19] Radiologist-performed, hand-held bilateral whole-breast US was performed by one of the aforementioned breast radiologists with a 12- to 5-MHz linear array transducer (HDI 5000 or iU22, Phillips-Advanced Technology Laboratories, Bothell, WA; Logic 9, GE Medical Systems, Milwaukee, WI). At the radiologist's discretion, color or power Doppler imaging and harmonic imaging were performed. Elastography was not performed. The radiologists were not blinded to mammography and were able to review prior examinations and the clinical information of patients. The final assessment was prospectively determined by the same radiologists who performed US according to BI-RADS.[19] Since March 2010 (the starting year of this study), in order to reduce the false positive rate, we have trained our radiologists to classify the following findings as category 2: a complicated cyst 5 mm or smaller which were observed as a circumscribed, homogeneous, and hypoechoic lesion (Fig. 1A) and a circumscribed oval-shaped solid mass 5 mm or smaller without any suspicious US features (Fig. 1B). The 2 criteria for downgrading were selected in consensus after an in-depth discussion between staff radiologists based on experience and other publications.[9,20,21] During the study period, staff radiologists continued to emphasize the downgrade criteria to fellow radiologists at the weekly conference.

2.3. Follow-up

Short-interval follow-up US examinations were recommended for category 3 lesions at 6, 12, and 24 months after category 3 assessments. If the lesion demonstrated stability during
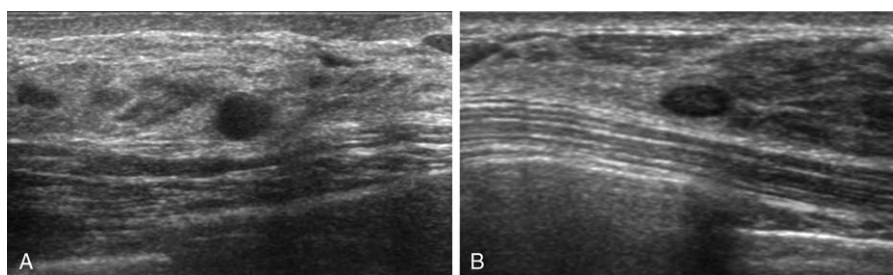
Figure 1. A representative example of downgrade criteria: a complicated cyst 5 mm or smaller (A) and a circumscribed oval-shaped solid mass 5 mm or smaller (B).

24 months, the final assessment was downgraded to category 2. US-guided core needle biopsy using a 14-gauge automated core biopsy needle (TSK Stericut biopsy needle, standard type with co-axial, T SK Laboratory, Soja, Tochigi, Japan) was performed for category 4 or 5 lesions immediately after US or for lesions with increased size or newly developed suspicious lesions during follow-up. Surgical excision was performed for biopsy results of malignancy including ductal carcinoma in situ (DCIS) or invasive cancer, and atypical or high-risk lesions including atypical ductal hyperplasia, atypical lobular hyperplasia, papilloma, and phyllodes tumor.[4] Immunohistochemistry (IHC) analysis for receptors of breast cancers was performed.[22,23] Hormone receptor (HR) positivity was defined as estrogen receptor and/or progesterone receptor positivity ≥1% nuclear staining. Human epidermal growth factor receptor 2 (HER2) positivity was defined as having IHC HER2 score of 3+ or gene amplification by fluorescence in situ hybridization in tumors with IHC HER2 score of 2+. Tumor subtypes were categorized based on the receptor status as HR+ and HER2−, HR+ and HER2+, HR− and HER2+, and HR− and HER2− (triple negative).[24]

2.4. Data and statistical analysis

The pathologic results of biopsy and surgery performed within 1 year of the screening US were reviewed. In patients with surgery, surgical results showing malignancy were considered as disease positive, and those showing benign results were considered as disease negative. In patients without surgery, the absence of a cancer diagnosis on follow-up US ≥ 12 months after the initial US was considered as disease negative. BI-RADS category 3 or higher were considered as test positive and categories 1 and 2 were considered as test negative.[14] Interval cancers were defined as those diagnosed because of clinical abnormalities occurring in an interval less than 1 year after the last screening US.

PPV_1 was defined as the malignancy rate among positive tests (cancer/lesions with a BI-RADS 3 or higher).[19] PPV_2 was defined as the malignancy rate among positive tests with biopsy recommendations (cancer/BI-RADS 4 or 5 lesions).[19] PPV_3 was defined as the malignancy rate of lesions with biopsy among BI-RADS 4 or 5 lesions (cancers/lesions with biopsy among BI-RADS 4 or 5 lesions).[19] The rates of BI-RADS category, PPVs, and biopsy rates for trend over 3 years were evaluated using Mantel–Haenszel chi-square test. The differences in total and invasive cancer yields per year were evaluated using chi-square or Fisher exact test. Differences in BI-RADS 3 to 4a rates, total, and invasive cancer yields, PPVs, and biopsy rates were compared between the 2 groups with initial or noninitial examinations using chi-square test. The rates of BI-RADS 3 to 4a and total cancer yield were compared between less-experienced and experienced radiologists using Mantel–-Haenszel chi-square test for trend over 3 years and chi-square or Fisher exact test for difference per year. P values of less than 0.05 were considered statistically significant. All statistical analyses were performed with SAS statistical software (version 9.2, SAS institute, Cary, NC).

3. Results

With the application of the downgrade criteria, the rate of BI-RADS categories 1 to 2 increased, and categories 3 to 4a decreased over 3 years in total examinations, initial examinations, and noninitial examinations (Table 1, all P < 0.001). In total examinations, BI-RADS categories 3 to 4a decreased from 33.3% to 14.6%. BI-RADS categories 3 to 4a were more present

Table 1					
Distribution of BI-RADS final assessment category in supplemental screening ultrasound.					
	Overall	**First year**	**Second year**	**Third year**	**P**
Total examination					
Categories 1–2	73.8 (2340/3171)	66.1 (646/978)	70.2 (824/1173)	85.3 (870/1020)	<0.001
Categories 3–4a	25.9 (820/3171)	33.3 (326/978)	29.4 (345/1173)	14.6 (149/1020)	<0.001
Initial examination					
Categories 1–2	52.8 (527/998)	46.8 (185/395)	51.2 (212/414)	68.8 (130/189)	<0.001
Categories 3–4a	46.3 (462/998)	51.9 (205/395)	47.8 (198/414)	31.2 (59/189)	<0.001
Noninitial examination					
Categories 1–2	83.4 (1813/2173)	79.1 (461/583)	80.6 (612/759)	89.0 (740/831)	<0.001
Categories 3–4a	16.5 (358/2173)	20.8 (121/583)	19.4 (147/759)	10.8 (90/831)	<0.001

Overall = from first to third year. Total examination = initial plus noninitial examination. Values given are percentages (number of women with corresponding category/total women). P value for the trend over 3 years using Mantel–Haenszel chi-square test.

Table 2

Total and invasive cancer yields of supplemental screening ultrasound.

	Overall	First year	Second year	Third year	P
Total examination					
Total cancer yield					
‰ (95% CI)	2.8 (1.3–5.4)	2.0 (0.2–7.4)	4.3 (1.4–9.9)	2.0 (0.2–7.1)	0.539
No./total	9/3171	2/978	5/1173	2/1020	
Invasive cancer yield					
‰ (95% CI)	2.2 (0.9–4.5)	2.0 (0.2–7.1)	2.6 (0.5–7.5)	2.0 (0.2–7.1)	>0.999
No./total	7/3171	2/978	3/1173	2/1020	
Initial examination					
Total cancer yield					
‰ (95% CI)	4.0 (1.1–10.2)	2.5 (1.0–14.0)	7.3 (1.5–21.0)	0.0 (0.0–19.3)	0.535
No./total	4/998	1/395	3/414	0/189	
Invasive cancer yield					
‰ (95% CI)	4.0 (1.1–10.2)	2.5 (1.0–14.0)	7.3 (1.5–21.0)	0.0 (0.0–19.3)	0.535
No./total	4/998	1/395	3/414	0/189	
Noninitial examination					
Total cancer yield					
‰ (95% CI)	2.3 (0.1–0.5)	1.7 (0.0–9.5)	2.6 (0.3–9.5)	2.4 (0.3–8.7)	>0.999
No./total	5/2173	1/583	2/759	2/831	
Invasive cancer yield					
‰ (95% CI)	1.4 (0.3–4.0)	1.7 (0.0–9.5)	0.0 (0.0–4.8)	2.4 (0.3–8.7)	0.491
No./total	3/2173	1/583	0/759	2/831	

Overall = from first to third year. Total examination = initial plus noninitial examination. P value for difference per each year using chi-square or Fisher exact test. No./total = number of women with cancers/total women, CI = confidence interval.

in initial examinations compared to noninitial examinations (46.3% [462 of 998] vs 16.5% [358 of 2173], $P < 0.001$).

3.1. Supplemental cancer detection yield

Screening US detected 9 additional cancers: 2 were DCIS and 7 were invasive cancers. Total cancer yield per 1000 examinations was 2.8 (Table 2, 9 of 3171; 95% confidence interval [CI], 1.3–5.4) and invasive cancer yield per 1000 examinations was 2.2 (7 of 3171; 95% CI, 0.9–4.5). There were no differences in total and invasive cancer yields per year. Total cancer yield of initial examinations was higher than that of noninitial examinations without statistical significance (4.0 [4 of 998] vs 2.3 [5 of 2173], $P = 0.475$). Characteristics of the 9 US-detected cancers are demonstrated in Table 3 and Figs. 2 and 3. The median size of the 9 cancers was 8 mm, ranging from 5 to 15 mm. All had low or intermediate histologic grades. IHC subtypes were either HR positive/HER2 negative (77.8%, 7 of 9) or triple negative (22.2%, 2 of 9). None had lymph node or distant metastasis. There were no interval cancers.

3.2. PPV and biopsy rate

PPVs of total examinations showed an increasing trend over 3 years, although statistical significance was not achieved (Table 4. $PPV_1 = 0.6\%$, 1.4%, and 1.3% in first, second, and third year, respectively, $P = 0.494$; PPV_2 and $PPV_3 = 3.6\%$, 9.1%, and 9.5% in first, second, and third year, respectively, $P = 0.338$). As all BI-RADS 4 or 5 lesions underwent biopsy, PPV_2 and PPV_3 were the same. Overall, PPVs of initial examinations were smaller than those of noninitial examinations without statistical significance (PPV_1, 0.9% vs 1.4%, $P = 0.512$; PPV_2 and PPV_3, 4.5% vs 11.9%, $P = 0.145$).

Overall, 147 lesions underwent US-guided 14-gauge core needle biopsy. A total of 135 lesions (all BI-RADS 4 [n = 132] and 5 lesions [n = 1], and 2 lesions out of 700 BI-RADS 3 lesions due to the request of the patient) immediately underwent biopsy. The

remaining 12 lesions underwent biopsy before a full year of follow-up because of increased lesion size (n = 5) or a newly developed low-suspicious lesion (n = 7) on follow-up US for BI-RADS category 3. Overall, biopsy results consisted of 9 cancers (2 DCIS and 7 invasive cancers), 6 high-risk lesions (4 atypical ductal hyperplasia, 1 atypical papilloma, and 1 phyllodes tumor), and 132 benign lesions. Surgery was performed for 9 cancers and 6 high-risk lesions, and 6 high-risk lesions were finally confirmed to be benign by surgery. Biopsy rates significantly decreased over 3 years for total, initial, and noninitial examinations (Table 4). The overall biopsy rate of initial examinations was significantly larger than that of noninitial examinations (9.6% vs 2.3%, $P < 0.001$).

3.3. Comparison of the performance between less-experienced and experienced radiologists

On the whole, the categories 3 to 4a rates were higher in less-experienced compared to experienced radiologists (Table 5). The categories 3 to 4a rates showed a decreasing trend over 3 years in both groups. The gaps between the 2 groups for categories 3 to 4a rates decreased over 3 years. There were no differences in total cancer yields between the 2 groups. In both groups, categories 3 to 4a rates of initial examinations were significantly higher than those of noninitial examinations (all $P < 0.001$).

4. Discussion

The application of the downgrade criteria to supplemental screening US reduced BI-RADS categories 3 to 4a in both less-experienced and experienced radiologists over 3 years. Despite the reduction of categories 3 to 4a, PPVs did not significantly increase, as the number of US-detected cancers was small. Our total cancer yield of 2.8 cancers per 1000 examinations was within the reported ranges of 0.3 to 6.8 cancers per 1000 examinations (median: 4.2),[4–11] and interval cancers were not detected. Compared to our 2009 results which were found before

Table 3

Clinicopathologic and imaging characteristics of supplemental screening ultrasound-detected cancers.

| No.	Year	Age	MG density	US BI-RADS category	Size (mm) on US	Size (mm) on surgical pathology	Comparison with the previous US	Histologic subtype	Histologic grade	Subtype
1[a]	1st	56	4	4c	10	15	New	Invasive ductal carcinoma	Intermediate	HR+/HER2−
2	1st	47	3	4c	9	12	Initial	Mixed invasive ductal and lobular carcinoma	Low	HR+/HER2−
3[b]	2nd	57	3	4a	6	7	New	Ductal carcinoma in situ	Low	HR+/HER2−
4	2nd	52	3	4a	9	7	Initial	Invasive ductal carcinoma	Intermediate	HR−/HER2−
5	2nd	50	3	4c	6	5	Initial	Invasive lobular carcinoma	Low	HR+/HER2−
6	2nd	47	3	4a	14	14	New	Ductal carcinoma in situ	Low	HR+/HER2−
7	2nd	42	4	5	15	15	Initial	Invasive ductal carcinoma	Low	HR+/HER2−
8	3rd	58	3	4a	12	5	Increased	Invasive apocrine carcinoma	Intermediate	HR−/HER2−
9	3rd	44	3	4a	5	8	New	Mucinous carcinoma	Low	HR+/HER2−

Initial: there were no previous US examinations. New: the lesion was newly developed compared with previous US examinations. 1[a]: Fig. 3 is this case. 3[b]: Fig. 2 is this case. All were node-negative.

BI-RADS = Breast Imaging Reporting and Data System, HR = hormone receptor, HER2 = human epidermal growth factor receptor 2, MG = mammography, US = ultrasound.

the downgrade criteria was applied, the total cancer detection yield of the present study was higher without statistical significance (1.8 cancers/1000 examinations in 2009[9] vs 2.8 cancers/1000 examinations in 2010–2012, $P = 0.559$). These results mean that the downgrade criteria effectively reduced BI-RADS categories 3 to 4a without a loss of cancer detection—that is, the false positive rates decreased.

We considered a complicated cyst 5 mm or smaller which was observed as a circumscribed, homogeneous, and hypoechoic lesion as BI-RADS category 2. Complicated cysts are cysts with internal debris.[25] When the debris is mobile or a fluid-debris level is seen, they are regarded as benign.[25] But, when the debris is homogeneous and hypoechoic, it is often difficult to distinguish a complicated cyst from a solid mass, so they have been generally classified as probable benign, BI-RADS category 3, or unnecessary aspiration or core needle biopsy has been occasionally performed.[20,25,26] However, breast cancers presenting as complicated cysts on US are very rare at percentages of less than 0.5%, with a range of 0% to 0.4%.[6,20,26] A circumscribed oval-shaped mass 5 mm or smaller without any suspicious US features was also considered as category 2. In previous studies, there were no malignancies among masses less than 5 mm detected on screening US.[9,21] A Connecticut study by Hooley et al and the ACRIN 6666 trial by Berg et al showed the similar results compared to our study. The study by Hooley et al[6] showed that the category 3 rate can decrease nearly 50% by retrospectively classifying nonsimple cysts in the presence of multiple cysts, and solitary, oval, circumscribed complicated cysts 5 mm or smaller as benign lesions without a loss of sensitivity. The ACRIN 6666 trial had no malignancies among multiple bilateral, similar-appearing circumscribed masses, and had a very low malignancy rate (0.4%) for solitary circumscribed masses (including complicated cyst, clustered microcyst, and oval, round, or gently lobulated circumscribed solid mass).[26] When we performed screening US with our downgrade criteria, categories 3 to 4a rates were significantly reduced, and there were no interval cancers. Therefore, complicated cysts and circumscribed oval-shaped solid masses 5 mm or smaller can be downgraded to category 2, and a follow-up after 1 year is suitable for these lesions.

One of the major problems of supplemental screening US is a high rate of BI-RADS category 3.[6,15,26] In this study, the overall category 3 rate during 3 years was 22.1%, significantly lower compared to 30.4% (504 of 1656) of our 2009 results ($P < 0.001$). The category 3 rate continuously decreased from 28.3% to 12.6% over 3 years. Thus, the application of downgrade criteria was effective in reducing the category 3 rate. The category 3 rate was 19.5% in the ACRIN 6666,[15] and 20% in the Connecticut study,[6] similar to the 22.1% of our study. Although the biopsy rate significantly decreased over 3 years, 89.9% of the lesions (132 of 147) had false positive biopsy results, which was consistent with previously reported false positive biopsy rates.[6,14,16]

Few studies have evaluated the interoperator variability of physician-performed hand-held US.[17,18] Bosch et al[18] found high interexamination agreement in both detection and classification across 3 radiologists with different experience levels; 1 resident and 2 senior investigators with each 3 and 5 years of experience. Among 11 breast imaging radiologists with at least 3 years of experience (range, 3–26 years) who were trained and qualified for ACRIN 6666, there was moderate agreement for the BI-RADS final assessment category.[17] In that study, a comparison analysis according to the experience level of the
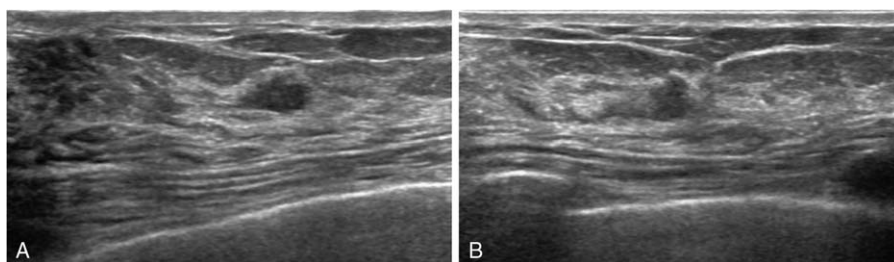
Figure 2. A 57-year-old woman with supplemental screening ultrasound (US)-detected breast cancer (ductal carcinoma in situ, low histologic grade, hormone receptor positive/HER2 negative). Transverse (A) and longitudinal (B) gray-scale US images show a 6-mm-sized irregular hypoechoic solid mass with microlobulated margins. It was assigned BI-RADS category 4a.
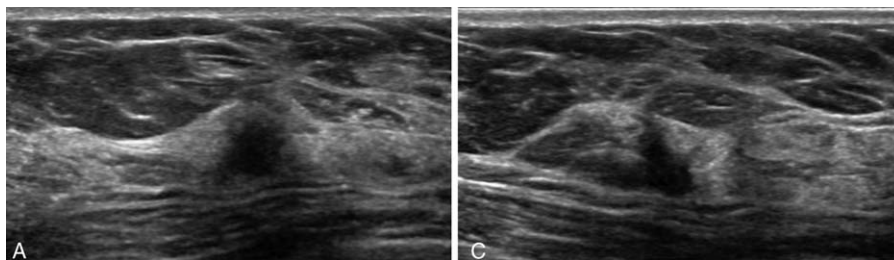
Figure 3. A 56-year-old woman with supplemental screening ultrasound (US)-detected breast cancer (invasive ductal carcinoma, intermediate histologic grade, hormone receptor positive/HER2 negative). Transverse (A) and longitudinal (B) gray-scale US images show a 10-mm-sized irregular hypoechoic solid mass with spiculated margin and posterior shadowing. It was assigned BI-RADS category 4c.

Table 4

PPV and biopsy rates of supplemental screening ultrasound.

	Overall	First year	Second year	Third year	P
Total examination					
PPV_1					
% (95% CI)	1.1 (0.5–2.1)	0.6 (0.1–2.2)	1.4 (0.5–3.3)	1.3 (0.2–4.7)	0.494
No./total	9/831	2/332	5/349	2/150	
PPV_2, PPV_3					
% (95% CI)	6.9 (3.2–12.6)	3.6 (0.4–12.5)	9.1 (3.0–20.0)	9.5 (1.2–30.4)	0.338
No./total	9/131	2/55	5/55	2/21	
Biopsy rate					
% (95% CI)	4.6 (4.1–6.8)	6.5 (4.9–8.3)	5.0 (3.8–6.8)	2.4 (1.4–3.5)	<0.001
No./total	147/3171	64/978	59/1173	24/1020	
Initial examination					
PPV_1					
% (95% CI)	0.9 (0.2–2.2)	0.5 (0.0–2.6)	1.5 (0.3–4.3)	0.0 (0.0–6.1)	>0.999
No./total	4/471	1/210	3/202	0/59	
PPV_2, PPV_3					
% (95% CI)	4.5 (1.2–11.1)	2.5 (0.1–13.2)	7.3 (1.5–19.9)	0.0 (0.0–37.0)	>0.999
No./total	4/89	1/40	3/41	0/8	
Biopsy rate					
% (95% CI)	9.6 (7.8–12.0)	11.1 (8.6–15.8)	10.4 (7.8–14.7)	4.8 (2.5–9.5)	0.040
No./total	96/998	44/395	43/414	9/189	
Noninitial examination					
PPV_1					
% (95% CI)	1.4 (0.5–3.2)	0.8 (0.2–4.5)	1.4 (0.2–4.8)	2.2 (0.3–7.7)	0.563
No./total	5/360	1/122	2/147	2/91	
PPV_2, PPV_3					
% (95% CI)	11.9 (4.0–25.6)	6.7 (0.5–32.0)	14.3 (1.8–42.8)	15.4 (1.9–45.5)	0.574
No./total	5/42	1/15	2/14	2/13	
Biopsy rate					
% (95% CI)	2.3 (1.7–3.2)	3.4 (1.9–4.7)	2.1 (1.1–3.2)	1.8 (1.0–3.1)	0.033
No./total	51/2173	23/583	16/759	15/831	

Overall = from first to third year. Total examination = initial plus noninitial examination. P value for trends over 3 years using Mantel–Haenszel chi-square test. CI = confidence interval, PPV = positive predictive value.

Table 5

Comparison of the rate of categories 3 to 4a and total cancer yield between less-experienced and experienced radiologists.

	Overall			First year			Second year			Third year				
	Less-experienced	Experienced	P value*	Less-experienced	Experienced	P value*	Less-experienced	Experienced	P value*	Less-experienced	Experienced	P value*	P value†	P value‡
Total examinations														
Categories 3–4a %	29.4 (661/2249)	17.2 (159/922)	<0.001	39.4 (246/624)	22.6 (80/354)	<0.001	33.4 (298/893)	16.8 (47/280)	<0.001	16.0 (117/732)	11.1 (32/288)	0.059	<0.001	<0.001
Total cancer yield ‰	3.1 (7/2249)	2.1 (2/922)	0.932	1.6 (1/624)	2.8 (1/354)	0.932	4.5 (4/893)	3.6 (1/280)	1.000	2.7 (2/732)	0.0 (0/288)	1.000	0.597	0.622
Initial examinations														
Categories 3–4a %	51.0 (382/749)	32.1 (80/249)	<0.001	59.4 (158/266)	36.4 (47/129)	<0.001	51.7 (178/344)	28.6 (20/70)	<0.001	33.1 (46/139)	26.0 (13/50)	0.453	<0.001	<0.001
Total cancer yield ‰	4.0 (3/749)	4.0 (1/249)	1.000	3.8 (1/266)	0.0 (0/129)	1.000	5.8 (2/344)	14.3 (1/70)	0.427	0.0 (0/139)	0.0 (0/50)	–	0.655	0.277
Noninitial examinations														
Categories 3–4a %	18.6 (279/1500)	11.7 (79/673)	<0.001	24.6 (88/358)	14.7 (33/225)	0.006	21.9 (120/549)	12.9 (27/210)	0.007	12.0 (71/593)	8.0 (19/238)	0.121	<0.001	<0.001
Total cancer yield ‰	2.7 (4/1500)	1.5 (1/673)	1.000	0.0 (0/358)	4.4 (1/225)	0.386	3.6 (2/549)	0.0 (0/210)	1.000	3.4 (2/593)	0.0 (0/238)	1.000	0.531	0.369

Overall = from first to third year. Total examination = initial plus noninitial examination. Lesion number/total number is in parentheses.

* P value for comparison between less-experienced and experienced radiologists using chi-square or Fisher exact test.

† P value for trend of categories 3 to 4a over 3 years using Mantel–Haenszel chi-square test, or difference in total cancer yield per each year in less-experienced radiologists using chi-square or Fisher exact test.

‡ P value for trend of categories 3 to 4a over 3 years using Mantel–Haenszel chi-square test, or difference in total cancer yield per each year in experienced radiologists using chi-square or Fisher exact test.

radiologists was not performed.[17] In our study, categories 3 to 4a rates were significantly higher in less-experienced than experienced radiologists, although there were no differences in total cancer yields. In fact, the more important observation was that the gap in the categories 3 to 4a rate between the 2 groups decreased over 3 years, which supported the effectiveness of the downgrade criteria not only for generally decreasing categories 3 to 4a but also for decreasing the performance gaps between less-experienced and experienced radiologists.

Our total cancer yield of 2.5 cancers per 1000 examinations was similar to the results of the Connecticut studies (1.8–3.2 cancers per 1000 examinations), which may reflect the performance of screening US performed on the general population who are at average risk of breast cancer (i.e., women without known risk factors of breast cancer other than dense breast).[6,7,11] The characteristics of screening US-detected cancers have been reported to be invasive (median, 91%, range, 50%–100%), node-negative (median, 87.5%, range, 78%–100%), and less than 1.0 cm in size (range, 6.5–19 mm),[4–6,8,9,11] consistent with our results.

In noninitial examinations (incidence screening), the cancer detection rate, positive test rate, and biopsy rate were lower, and the PPVs were higher compared to initial examinations (prevalence screening) with statistical significance for the positive test rate and biopsy rate.[14] These results were consistent with our initial hypothesis (refer to the study population in Section 2.1 for a more detailed explanation) and the ACRIN 6666 trial results of combined mammography plus US screening which showed that cancer detection rates, biopsy rates, short-term follow-up rates decreased, and PPVs increased in the incidence screening.

There were several limitations in our study. This study was retrospectively conducted in a single institution, third-referral center by breast radiologists. Generalization of the results may be limited for other study populations, and for examinations performed by technologist or less-experienced physicians. Selection bias might have occurred owing to the exclusion of women without follow-up US for at least 1 year. Due to the retrospective nature of our study, we could not analyze from our collected data whether the downgrade criteria was properly applied per patient-level by each radiologist. More systematic training programs and quality control programs using videos, still images, or tests are needed to monitor the quality of each radiologist's classification abilities with the downgrade criteria. Further large-scale, multicenter, prospective studies are needed to validate the effectiveness of the downgrade criteria.

In conclusion, the application of the downgrade criteria to supplemental screening US for women with negative mammography but dense breasts reduced BI-RADS categories 3 to 4a without a loss of cancer detection.

Acknowledgment

References

[1] Moss SM, Cuckle H, Evans A, et al. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. Lancet 2006;368:2053–60.

[2] Mandelson MT, Oestreicher N, Porter PL, et al. Breast density as a predictor of mammographic detection: comparison of interval-and screen-detected cancers. J Natl Cancer Inst 2000;92:1081–7.

[3] Dehkordy SF, Carlos RC. Dense breast legislation in the United States: state of the States. J Am Coll Radiol 2013;10:899–902.

[4] Berg WA, Blume JD, Cormack JB, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. JAMA 2008;299:2151–63.

[5] Leong LC, Gogna A, Pant R, et al. Supplementary breast ultrasound screening in Asian women with negative but dense mammograms—a pilot study. Ann Acad Med Singapore 2012;41:432.

[6] Hooley RJ, Greenberg KL, Stackhouse RM, et al. Screening US in patients with mammographically dense breasts: initial experience with Connecticut Public Act 09-41. Radiology 2012;265:59–69.

[7] Parris T, Wakefield D, Frimmer H. Real world performance of screening breast ultrasound following enactment of Connecticut Bill 458. Breast J 2013;19:64–70.

[8] Girardi V, Tonegutti M, Ciatto S, et al. Breast ultrasound in 22,131 asymptomatic women with negative mammography. Breast 2013;22:806–9.

[9] Moon H, Jung I, Park S, et al. Comparison of cancer yields and diagnostic performance of screening mammography vs. supplemental screening ultrasound in 4394 women with average risk for breast cancer. Eur J Ultrasound 2015;36:255–63.

[10] Chang JM, Koo HR, Moon WK. Radiologist-performed hand-held ultrasound screening at average risk of breast cancer: results from a single health screening center. Acta Radiol 2015;56:652–8.

[11] Weigert J, Steenbergen S. The Connecticut experiment: the role of ultrasound in the screening of women with dense breasts. Breast J 2012;18:517–22.

[12] Moon HJ, Kim E-K. Characteristics of breast cancer detected by supplementary screening ultrasonography. Ultrasonography 2015;34:153.

[13] Scheel JR, Lee JM, Sprague BL, et al. Screening ultrasound as an adjunct to mammography in women with mammographically dense breasts. Am J Obstet Gynecol 2015;212:9–17.

[14] Berg WA, Zhang Z, Lehrer D, et al. Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. JAMA 2012;307:1394–404.

[15] Barr RG, Zhang Z, Cormack JB, et al. Probably benign lesions at screening breast US in a population with elevated risk: prevalence and rate of malignancy in the ACRIN 6666 trial. Radiology 2013;269: 701–12.

[16] Merry GM, Mendelson EB. Update on screening breast ultrasonography. Radiol Clin N Am 2014;52:527–37.

[17] Berg WA, Blume JD, Cormack JB, et al. Operator dependence of physician-performed whole-Breast US: lesion detection and characterization 1. Radiology 2006;241:355–65.

[18] Bosch A, Kessels A, Beets G, et al. Interexamination variation of whole breast ultrasound. Br J Radiol 2003;76:328–31.

[19] American College of RadiologyBreast Imaging Reporting and Data System (BI-RADS). 4th ed.Reston, VA:American College of Radiology; 2003.

[20] Daly CP, Bailey JE, Klein KA, et al. Complicated breast cysts on sonography: is aspiration necessary to exclude malignancy? Acad Radiol 2008;15:610–7.

[21] Kim SJ, Chang JM, Cho N, et al. Outcome of breast lesions detected at screening ultrasonography. Eur J Radiol 2012;81:3229–33.

[22] Hammond MEH, Hayes DF, Dowsett M, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). Arch Pathol Lab Med 2010;134:e48–72.

[23] Wolff AC, Hammond MEH, Hicks DG, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. Arch Pathol Lab Med 2013;138: 241–56.

[24] Nguyen PL, Taghian AG, Katz MS, et al. Breast cancer subtype approximated by estrogen receptor, progesterone receptor, and HER-2 is associated with local and distant recurrence after breast-conserving therapy. J Clin Oncol 2008;26:2373–8.

[25] Berg WA, Sechtin AG, Marques H, et al. Cystic breast masses and the ACRIN 6666 experience. Radiol Clin N Am 2010;48:931–87.

[26] Berg WA, Zhang Z, Cormack JB, et al. Multiple bilateral circumscribed masses at screening breast US: consider annual follow-up. Radiology 2013;268:673–83.