Diagnostic performance of screening breast ultrasonography in women with mammographically dense breast: Application of the downgrade criteria

Soo-Yeon Kim

Department of Medicine

The Graduate School, Yonsei University

Diagnostic performance of screening breast ultrasonography in women with mammographically dense breast: Application of the downgrade criteria

Directed by Professor Eun-Kyung Kim

The Doctoral Dissertation
submitted to the Department of Medicine,
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree
of Doctor of Philosophy

Soo-Yeon Kim

June 2016

This certifies that the Doctoral Dissertation of Soo-Yeon Kim is approved.

Thesis Supervisor : Eun-Kyung Kim

Thesis Committee Member#1 : Eun-Cheol Park

Thesis Committee Member#2 : Woo Kyung Moon

Thesis Committee Member#3: Inkyung Jung

Thesis Committee Member#4: Joon Jeong

The Graduate School
Yonsei University

June 2016

ACKNOWLEDGEMENTS

I acknowledge my deep gratitude to Professor Eun-Kyung Kim, who is my thesis supervisor, for supporting my efforts with total commitment and facilitating every step of the process. My appreciation for her guidance and encouragement is tremendous. I am also indebted to Professor Eun-Cheol Park, Woo Kyung Moon, Inkyung Jung, and Joon Jeong for their help for pertinent advice to assure the superior quality of this paper. Additionally, I also thank Mi Kyung Song and Bo Kyung Ma for their help in statistical analysis.

<TABLE OF CONTENTS>

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Diagnostic performance of screening breast ultrasonography in women with mammographically dense breast:
Application of the downgrade criteria

Soo-Yeon Kim

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Eun-Kyung Kim)

Purpose: To investigate whether the application of the downgrade criteria on supplemental screening US for women with negative mammography but dense breasts can reduce Breast Imaging Reporting and Data System (BI-RADS) category 3-4a without a loss of cancer detection.

Methods: This retrospective study included 4730 women with negative mammography but dense breast who underwent radiologist-performed, hand-held screening US from March 2009 to February 2013. To reduce BI-RADS category 3-4a on screening US, since March 2010, complicated cysts 5mm or smaller and circumscribed oval-shaped solid masses 5mm or smaller were downgraded to BI-RADS category 2. The distribution of BI-RADS category, cancer detection yield and diagnostic performances of screening US were analyzed and compared before (year 2009: from March 2009 to February 2010) and after (year 2010-2012: from March 2010 to February 2013) the application of the downgrade criteria, with the reference standard defined as a combination of pathology and 12-month follow-up. Performances between less-experienced (19 fellows with 1-2 years of experience) and experienced (3 staffs with 12-18 years of experience) radiologists were compared.

Results: The application of downgrade criteria reduced BI-RADS category 3-4a in both less-experienced and experienced radiologists. Screening US detected 12 cancers (three ductal carcinoma in situs and nine invasive cancers) during four years. There were no differences in cancer detection yields before and after the application of downgrade criteria. Interval cancers were not detected. Specificity, accuracy and the area under the receiver operating characteristic curve significantly increased after the application of downgrade criteria.

Conclusion: The application of downgrade criteria reduced BI-RADS category 3-4a without a loss of cancer detection.

Key words : Screening mammography, Dense breast, Screening breast ultrasound, Breast Imaging Reporting and Data System

Diagnostic performance of screening breast ultrasonography in women with mammographically dense breast: Application of the downgrade criteria

Soo-Yeon Kim

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Eun-Kyung Kim)

I. INTRODUCTION

Mammography is the only screening modality that has been proven to reduce mortality caused by breast cancer.[1,2] However, the sensitivity of mammography decreases in dense breasts, and invasive cancer is often mammographically subtle or occult when the breast tissue is dense.[3] These limitations of screening mammography have led to the introduction of breast density inform legislation in the United States, and the increasing demand for a supplemental screening tool adjuvant to screening mammography.[4]

Breast ultrasound (US) is an attractive screening tool with no ionizing radiation, of wide availability, and known for its tolerability by patients. Supplemental screening US in women with dense breasts and/or the elevated risk of breast cancer has been shown to benefit additional cancer detection by an average of 4.2 cancers per 1000 women screened.[5-16] Most of the breast cancers identified by supplemental US are smaller less than 1cm in size, invasive and node-negative.[5-11,13,14,16] Asian women tend to have small and dense breasts than Caucasian women, and the peak age of breast cancer diagnosis is between 45 and 55 years old, about 10-20 years younger than that in Caucasian women.[17-19] In these situations, mammography has been shown to be less effective, thus the supplemental screening US in Asian women may have more potential to detect additional cancers.[17-19] However, the addition of screening US after negative mammography has resulted in a high rate of positive test results (Breast Imaging Reporting and Data System (BI-RADS) final assessment category 3-5).[5-13,20-22] The supplemental screening US prompted a median 5.6% of women screened to undergo biopsy owing to their lesions being classified as BI-RADS category 4-5.[5-13,20] Also, the screening US resulted in nearly 20% of women screened to undergo short-term follow-up owing to a BI-RADS category 3 assessment.[11,21,22] The average positive predictive value (PPV) for biopsies performed was only 9.5%, leading to unnecessary harm, discomfort, and emotional stress associated with biopsy.[5,8,11,23-29]

Since 2001, we have collected all data for breast US exams performed at our hospital, and we have performed a periodic audit for screening breast US from 2009.[14] In our 2009 results, supplemental screening US was shown to detect additional cancers (3.0 cancers per 1000 exams) in women with dense breasts who were at average risk for breast cancer, but there were concerns regarding the high rates of positive test results (BI-RADS category 3-5) compared to mammography alone, and a low PPV of less than 4%.[14] Since March 2010, to reduce the BI-RADS category 3-4a on screening US, the downgrade criteria which classify complicated cysts 5mm or smaller and circumscribed oval-shaped solid masses 5mm or smaller as BI-RADS category 2 were established based on our experience and other published studies,[14,30-32] and were applied in our daily practice. Complicated cysts and circumscribed oval shaped masses 5mm or smaller are often seen on screening US, and according to BI-RADS, they are assigned as category 3, probable benign finding and then short-term follow-ups are required.[32-34] Some complicated cysts may be assigned as category 4a, low suspicious finding, then be aspirated or biopsied, because they sometimes may be seen as a hypoechoic lesion with an indistinct or microlobulated margin and mild posterior shadowing, a mimic of malignancy.[32-34] Less-experienced radiologists might have more possibility to assign them as category 4a. However, breast cancer presenting as a complicated cyst or circumscribed oval shaped mass 5mm or smaller was very rare less than 0.5%.[14,30-32] Thus, it was expected that category 3-4a rate could be reduced without a significant loss of cancer detection, if they are downgraded to category 2.

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

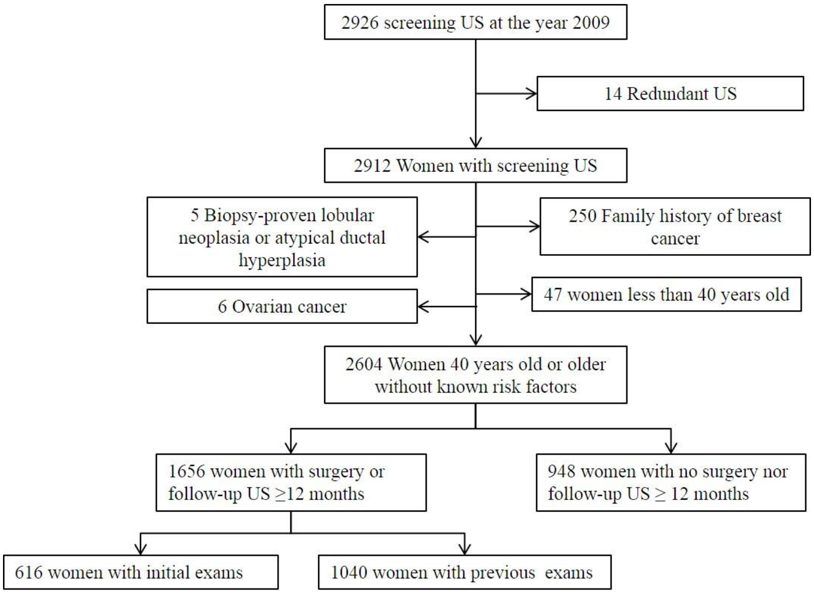
II. MATERIALS AND METHODS

Our Institutional Review Board approved this retrospective study, and patient approval nor informed consent was required for the review of medical records.

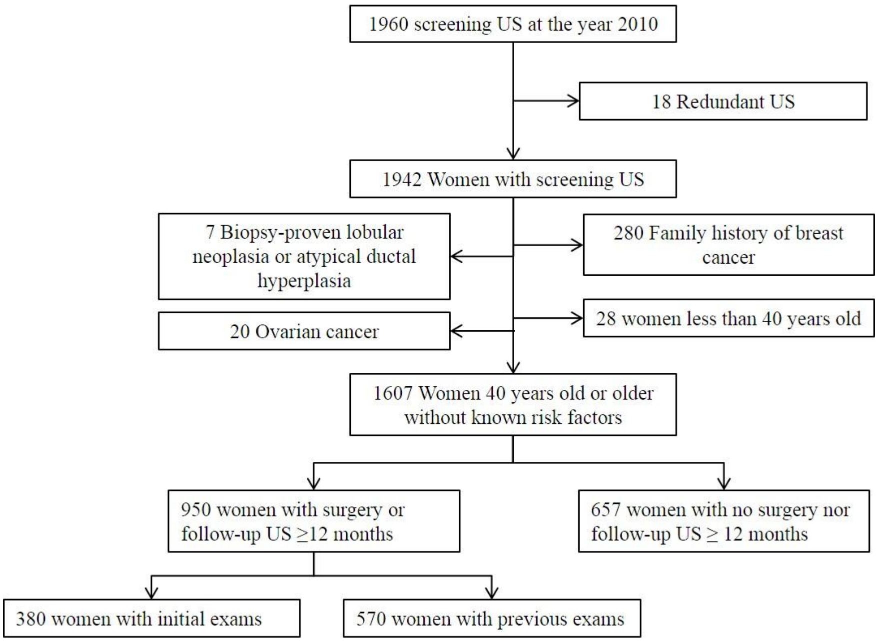
1. Study population

From March 2009 to February 2013, 42956 bilateral mammograms were consecutively performed at our institution. Of these, mammograms with the following inclusion criteria were found: (a) screening mammography, (b) dense breast defined as BI-RADS density grade 3 (heterogeneously dense) or 4 (extremely dense),[35] (c) negative findings defined as BI-RADS final assessment category 1 (negative) or 2 (benign),[35] and (d) supplemental US exams performed within 3 months after mammography. Mammography examinations that did not meet the inclusion criteria were excluded for the following reasons (n=34582): (a) diagnostic or post-operative follow-up of breast cancer (n=15294), (b) fatty breast (n=3809), (c) BI-RADS category 0 and 3-6 (n=3044), and (d) no supplemental US exams within 3 months (n=5464). As a result, 8374 mammograms and supplemental screening US exams (2026 exams from March 2009 to February 2010 (from now on, year 2009), 1960 exams from March 2010 to February 2011 (year 2010), 2362 exams from March 2011 to February 2012 (year 2011), and 2026 exams from March 2012 to February 2013 (year 2012), Figure 1a-d, respectively) were considered for inclusion. At each year, the exclusion criteria were as follows: (a) redundant US exams which were performed twice or more on the same woman during a one year period due to early visits despite a BI-RADS 1 or 2 assessment on a prior examination, (b) women less than 40 years old, (c) women with known risk factors for breast cancer (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; BRCA mutation was not considered because BRCA mutation analysis had not been routinely performed at our institution), and (d) women without surgery nor follow up US exams for at least 12 months. Finally, 4730 women 40 years or older and at average risk of breast cancer [mean age± standard deviation (years), 51.7±7.3; range (years), 40-78] were included from the year 2009-2012. There were 1656 women [51.8±7.3, 40-78] in the year 2009, 950 women [51.2±7.2, 40-75] in the year 2010, 1132 women [51.4±7.0, 40-78] in the year 2011, and 992 women [52.4±7.5, 40-78] in the year 2012. For each year, women were divided into two groups; women who had undergone initial US exams at our institution (prevalence screening), or who had undergone previous US exams at our institution (incidence screening). The numbers of women with initial exams were 616, 380, 385, and 175 women at the year 2009, 2010, 2011, and 2012, respectively.

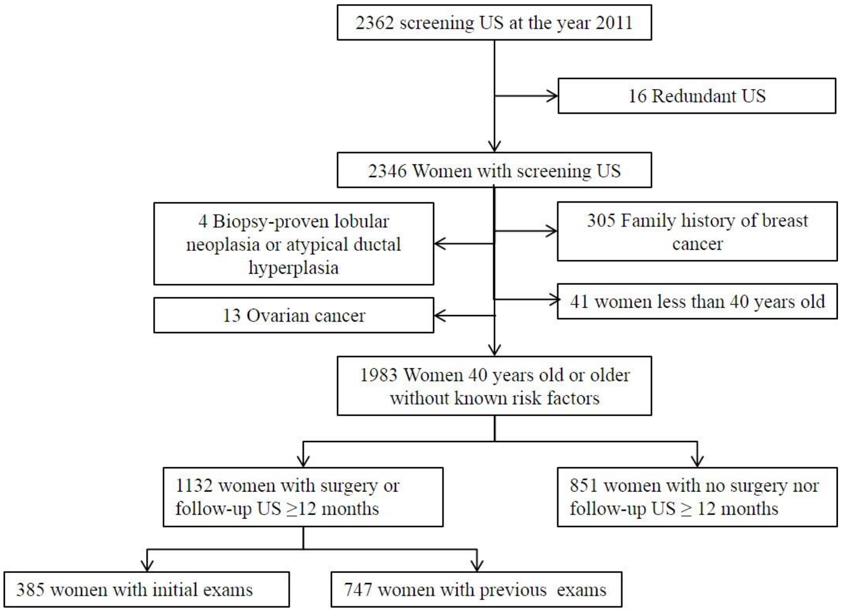
(a)


2926 screening US at the year 2009
14 Redundant US
2912 Women with screening US
5 Biopsy-proven lobular neoplasia or atypical ductal hyperplasia
250 Family history of breast cancer
6 Ovarian cancer
47 women less than 40 years old
2604 Women 40 years old or older without known risk factors
1656 women with surgery or follow-up US ≥12 months
948 women with no surgery nor follow-up US ≥ 12 months
616 women with initial exams
1040 women with previous exams

(b)


1960 screening US at the year 2010
18 Redundant US
1942 Women with screening US
7 Biopsy-proven lobular neoplasia or atypical ductal hyperplasia
280 Family history of breast cancer
20 Ovarian cancer
28 women less than 40 years old
1607 Women 40 years old or older without known risk factors
950 women with surgery or follow-up US ≥12 months
657 women with no surgery nor follow-up US ≥ 12 months
380 women with initial exams
570 women with previous exams

(c)



(d)

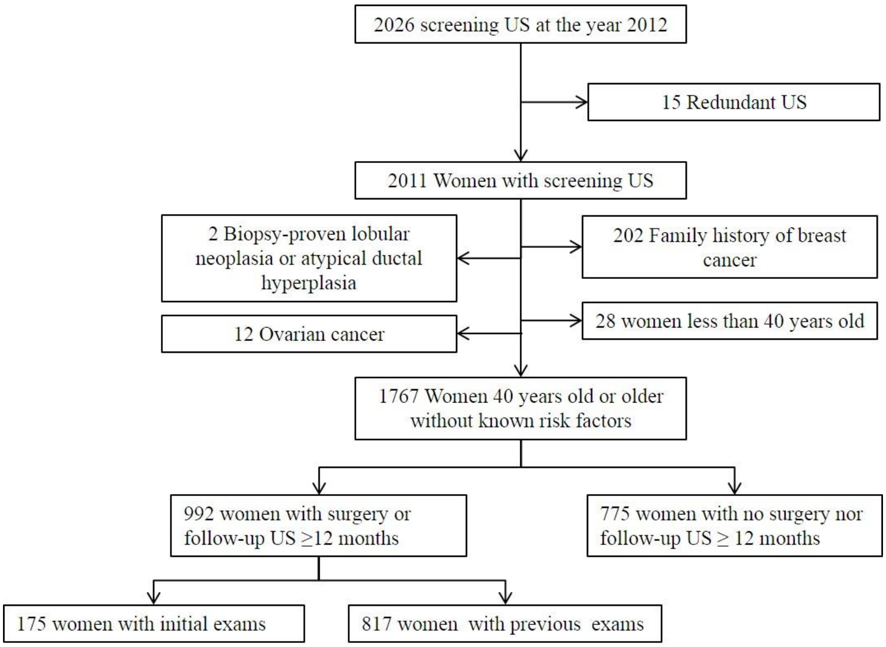


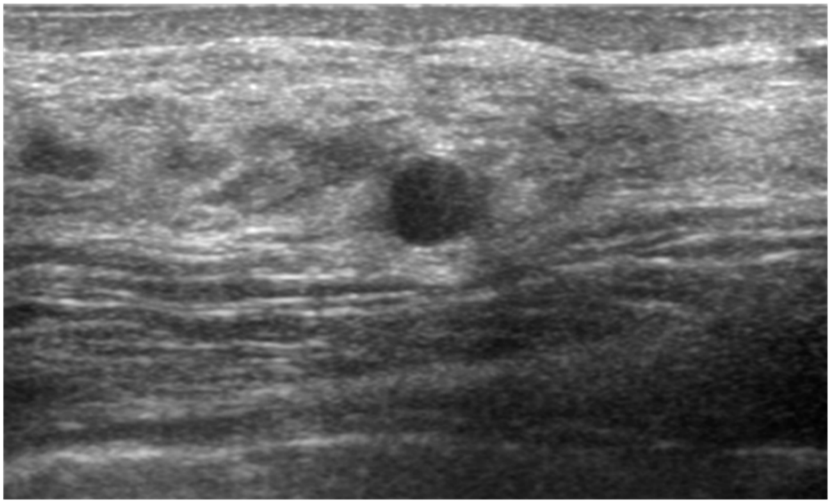
Fig. 1. Flow chart of the study population at each year (a-d, year 2009-2012, respectively)

2. Screening mammography and US

Digital mammography was performed with a full-filled digital mammography system (Lorad/Hologic Selenia, Lorad/Hologic, Danbury, CT, USA and SENOGRAPHE 2000D, GE Medical Systems, Milwaukee, WI, USA). Standard mediolateral oblique and craniocaudal views were routinely obtained. All mammography were interpreted by one of 22 dedicated breast radiologists (19 fellows with 1 to 2 years of experience and 3 staffs with 12 to 18 years of experience). The final assessment was prospectively determined according to BI-RADS.[35]

Radiologist-performed, hand-held bilateral whole-breast US was performed by one of the aforementioned breast radiologists with a 12 – 5MHz linear array transducer (HDI 5000 or iU22, Phillips-Advanced Technology Laboratories, Bothell, WA, USA; Logic 9, GE Medical Systems, Milwaukee, WI, USA). 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 exams and the clinical information of patients. The final assessment was prospectively determined by the same radiologists who performed US according to BI-RADS.[35] Since March 2010, in order to reduce the false positive rate, we have trained our radiologists to classify the following findings as category 2 – 1) a complicated cyst 5mm or smaller which demonstrated as a circumscribed, homogeneous and hypoechoic lesion (Fig. 2a) and 2) a circumscribed oval-shaped solid mass 5mm or smaller without any suspicious US features (Fig. 2b). The two criteria for downgrading were selected in consensus after an in-depth discussion between staff radiologists based on the experience of our institution and other publications.[14,30-32]

(a)



(b)

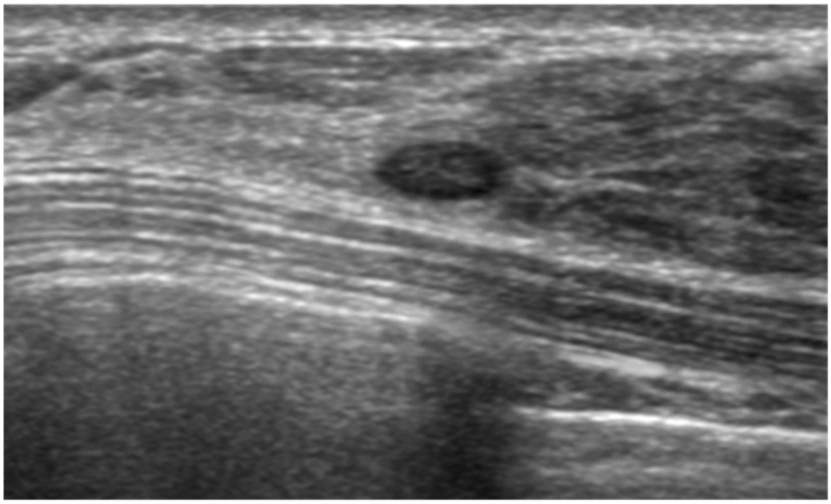


Fig. 2. An example of downgrade criteria: a complicated cyst 5mm or smaller (a) and a circumscribed oval-shaped solid mass 5mm or smaller (b).

3. Follow-up

Short-interval follow-up US exams were recommended for category 3 lesions at 6 months, 12 months, and 24 months after category 3 assessments. If the lesion demonstrated stability during 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, TSK Laboratory, Japan) was performed for category 4 or 5 lesions immediately after US. 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, atypical papilloma and phyllodes tumor.[9]

4. Data and statistical analysis

The pathologic results of biopsy and surgery performed within 1 year of screening US were reviewed. In patients with surgery, surgical results showing DCIS or invasive cancer were considered as disease positive, and those showing benign results were considered as disease negative. In patients without surgery, the absence of a diagnosis of cancer on follow-up US ≥12 months after the US were considered as disease negative. BI-RADS category 3 or higher on screening US were considered as test positive, and category 1 and 2 were considered as test negative.[21] Interval cancers were defined as those diagnosed because of clinical abnormalities occurring in an interval less than 1 year after the last screening US.

Because the aforementioned downgrade criteria were established and applied since March 2010, all data were compared between year 2009 and year 2010-2012 to evaluate the differences before (year 2009) and after (year 2010-2012) the application of the downgrade criteria using Chi-square test. The changing trends in data over four years from 2009 to 2012 were evaluated using

Mantel-Haenszel Chi-square test. The distribution of BI-RADS final assessment category, total cancer yield per 1000 exams, invasive cancer yield per 1000 exams, positive predictive value (PPV), biopsy rate, and diagnostic performances (sensitivity, specificity, negative predictive value, and accuracy) were calculated. PPV_1 was defined as the malignancy rate among positive screening US tests (cancer/ lesions with a BI-RADS 3 or higher).[35] PPV_2 was defined as the malignancy rate among positive tests with biopsy recommendations (cancer/BI-RADS 4 or 5 lesions).[35] PPV_3 was defined as the malignancy rate among positive tests with biopsy recommendations for which the same lesion underwent biopsy (cancer/ lesions that underwent biopsy among BI-RADS 4 or 5 lesions).[35] The area under the receiver operating characteristic curves (AUC) of screening US of year 2009 and year 2010-2012 were calculated and compared.[36] Differences in BI-RADS 3-4a and 3-5 rates, total and invasive cancer yields, PPVs, and biopsy rates were compared between the two groups with initial or non-initial exams using Chi-square test. BI-RADS 3-4a rates and total cancer yield were compared between less-experienced and experienced using Chi-square or Fisher's exact test. 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, USA).

III. RESULTS

1. Distributions of BI-RADS category of screening US over 4 years

Mean age of the included women was 51.2±7.3 years with a range of 40 to 78 years. Mean ages were not different between women with screening US in year 2009 and in year 2010-2012 (51.8±7.3 years vs. 51.7±7.3 years, $p = 0.569$). Initial exams were more present in year 2009 compared to in year 2010-2012 (37.2% (616 of 1656) vs. 30.6% (940 of 3074), $p<0.001$).

The distributions of BI-RADS final assessment category on screening US are summarized in Table 1. In total and non-initial exams, BI-RADS category 3-4a and 3-5 showed the decreasing trend over 4 years from year 2009 to 2012, and they decreased in year 2010-2012 compared to in year 2009. In initial exams, BI-RADS category 3-4a and 3-5 did not decrease in year 2010-2012 compared to in year 2009, but they showed the decreasing trend over 3 years from year 2010 to 2012. BI-RADS category 3-4a and 3-5 were more present in initial exams compared to non-initial exams (All, p<0.001).

Table 1. Distribution of BI-RADS final assessment category of screening US

	Year 2009	Year 2010-2012	Year 2010	Year 2011	Year 2012	*P* value[1]	*P* value[2]
			Total exams				
Category 1	566 (34.2)	891 (29.0)	262(27.6)	327(28.9)	302(30.4)	**<.001**	**0.021**
Category 2	498 (30.1)	1397(45.5)	369(38.8)	480(42.4)	548 (55.2)	**<.001**	**<.001**
Category 3	504 (30.4)	663 (21.6)	266 (28.0)	274 (24.2)	123 (12.4)	**<.001**	**<.001**
Category 4	88 (5.3)	122 (4.0)	53 (5.6)	50 (4.4)	19 (1.9)	**0.032**	**<.001**
Cat 4a	80 (4.8)	114 (3.7)	48 (5.1)	48 (4.2)	18 (1.8)	0.063	**<.001**
Cat 4b	6 (0.4)	4 (0.1)	3 (0.3)	1 (0.1)	0 (0.0)	0.108	**0.027**
Cat 4c	2 (0.1)	4 (0.1)	2 (0.2)	1 (0.1)	1 (0.1)	>.999	0.864
Category 5	0 (0.0)	1 (0.0)	0 (0.0)	1 (0.1)	0 (0.0)	>.999	0.799
Cat 3-4a	584 (35.3)	777 (25.3)	314 (33.1)	322 (28.4)	141 (14.2)	**<.001**	**<.001**
Cat 3-5	592 (35.7)	786 (25.6)	319 (33.6)	325 (28.7)	142 (14.3)	**<.001**	**<.001**
Total	1656	3074	950	1132	992		
			Initial exams				
Category 1	242 (39.3)	307 (32.7)	104 (27.4)	130 (33.8)	73 (41.7)	**0.008**	0.685
Category 2	87 (14.1)	196 (20.9)	75 (19.7)	73 (19.0)	48 (27.4)	**<.001**	**<.001**
Category 3	236 (38.3)	356 (37.9)	163 (42.9)	145 (37.7)	48 (27.4)	0.861	**0.044**
Category 4	51 (8.3)	80 (8.5)	38 (10.0)	36 (9.4)	6 (3.4)	0.872	0.255
Cat 4a	45 (7.3)	74 (7.8)	34 (9.0)	34 (8.8)	6 (3.4)	0.681	0.459
Cat 4b	4 (0.6)	4 (0.4)	3 (0.8)	1 (0.3)	0 (0.0)	0.720	0.244
Cat 4c	2 (0.3)	2 (0.2)	1 (0.3)	1 (0.3)	0 (0.0)	0.651	0.647
Category 5	0 (0.0)	0 (0.1)	0 (0.0)	1 (0.3)	0 (0.0)	>.999	0.756
Cat 3-4a	281 (45.6)	430 (45.7)	197 (51.8)	179 (46.5)	54 (30.9)	0.969	**<.001**[a]
Cat 3-5	287 (46.6)	437 (46.5)	201 (52.9)	182 (47.3)	54 (30.9)	0.969	**<.001**[a]
Total	616	940	380	385	175		
			Non-initial exams				
Category 1	324 (31.2)	584 (27.4)	158 (27.7)	197 (26.4)	229 (28.0)	**0.027**	0.082

Category 2	411 (39.5)	1201 (56.3)	294 (51.6)	407 (54.5)	500 (61.2)	**<.001**	**<.001**
Category 3	268 (25.8)	307 (14.4)	103 (18.1)	129 (17.3)	75 (9.2)	**<.001**	**<.001**
Category 4	37 (3.6)	42 (2.0)	15 (2.6)	14 (1.9)	13 (1.6)	**0.007**	**0.004**
Cat 4a	35 (3.4)	40 (1.9)	14 (2.5)	14 (1.9)	12 (1.5)	**0.009**	**0.005**
Cat 4b	2 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.107	0.174
Cat 4c	0 (0.0)	2 (0.1)	1 (0.2)	0 (0.0)	1 (0.1)	>.999	0.560
Category 5	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	NA	NA
Cat 3-4a	303 (29.2)	347 (16.3)	117 (20.7)	143 (19.1)	87 (10.7)	**<.001**	**<.001**
Cat 3-5	305 (29.3)	349 (16.4)	118 (20.7)	143 (19.1)	88 (10.8)	**<.001**	**<.001**
Total	1040	2134	570	747	817		

Values given are numbers (percentages).
Cat: category
P value[1] for difference between year 2009 versus year 2010-2012
P value[2] for the trend over four years
P value[2] <.001[a] means the trend for category 3-4a and 3-5 of initial exams over three years from 2010 to 2012

2. Cancer detection yield of screening US

Screening US detected twelve additional cancers over 4 years: three were DCIS and nine were invasive cancers. Total cancer yield per 1000 exams was 2.5 (Table 2, 12 of 4730; 95% confidence interval (CI), 1.3-5.4) and invasive cancer yield per 1000 exams was 1.9 (9 of 4730; 95% CI, 0.9-4.5). There were no differences in total and invasive cancer yields between the year 2009 and the year 2010-2012, and per each four year. Total cancer yield and invasive cancer yield of initial exams were slightly higher with those of non-initial exams without statistical significance (Total cancer yield, 2.6 (4 of 1556) for initial exam vs. 2.5 (8 of 3174) for non-initial exam, p=1.000; Invasive cancer yield, 2.6 (4 of 1556) for initial exam vs. 1.6 (5 of 3174) for non-initial exam, p=0.702). Characteristics of the US-detected cancers are demonstrated in Table 3 and Figs. 3 and 4. The median size of the cancers was 9mm, ranging from 5 to 15mm. All had low or intermediate histologic grades. Molecular subtypes were either luminal A (75.0%, 9 of 12) or triple negative (25.0%, 3 of 12). None had lymph node metastasis. There were no interval cancers detected within 12 months after the last screening US.

Table 2. Total and invasive cancer yields of screening US

	Year 2009	Year 2010-2012	Year 2010	Year 2011	Year 2012	*P* value[1]	*P* value[2]
			Total exams				
Total cancer yield						0.559	0.621
‰ (95% CI)	1.8 (0.4-5.3)	2.9 (1.3-5.6)	2.1 (0.3-7.6)	4.4 (1.4-0.3)	2.0 (0.2-7.3)		
No./total	3/1656	9/3074	2/950	5/1132	2/992		
Invasive cancer yield						0.509	0.570
‰ (95% CI)	1.2 (0.1-4.4)	2.3 (0.9-4.7)	2.1 (0.3-7.6)	2.7 (0.5-7.7)	2.0 (0.2-7.3)		
No./total	2/1656	7/3074	2/950	3/1132	2/992		
			Initial exams				
Total cancer yield						0.157	0.236
‰ (95% CI)	0.0 (0.0)	4.3 (1.2-10.9)	2.6 (0.1-4.6)	7.8 (1.6-2.6)	0.0 (0.0-0.9)		
No./total	0/616	4/940	1/380	3/385	0/175		
Invasive cancer yield						0.157	0.236
‰ (95% CI)	0.0 (0.0)	4.3 (1.2-10.9)	2.6 (0.1-14.6)	7.8 (1.6-22.6)	0.0 (0.0-20.9)		
No./total	0/616	4/940	1/380	3/385	0/175		
			Non-initial exams				
Total cancer yield						0.722	>.999
‰ (95% CI)	2.9 (0.6-8.4)	2.3 (0.8-5.5)	1.8 (0.0-9.7)	2.7 (0.3-9.6)	2.4 (0.3-8.8)		
No./total	3/1040	5/2134	1/570	2/747	2/817		
Invasive cancer yield						0.665	>.999
‰ (95% CI)	1.9 (0.2-6.9)	1.4 (0.3-4.1)	1.8 (0.0-9.7)	0.0 (0.0-4.9)	2.4 (0.3-8.8)		
No./total	2/1040	3/2134	1/570	0/747	2/817		

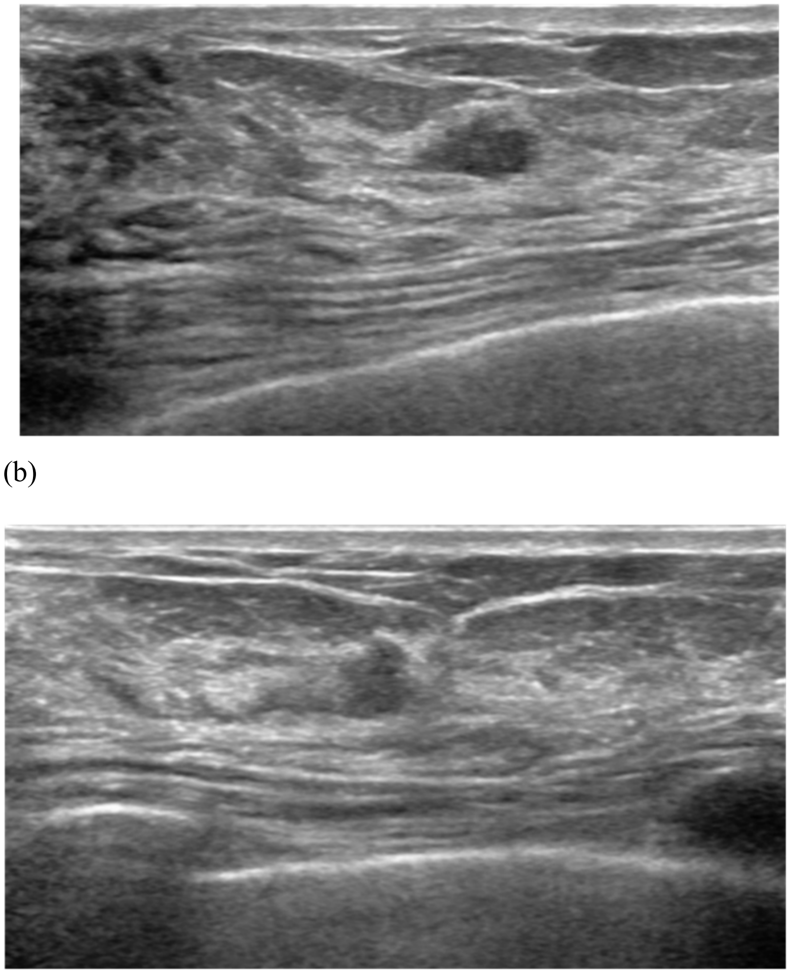
CI, confidence interval
P value[1] for difference between year 2009 versus year 2010-2012
P value[2] for difference of each year

Table 3. Clinicopathologic and imaging characteristics of screening ultrasound (US)-detected cancers

No.	Year	Age	US category	Size on US [12]	Size on surgical pathology [12]	Comparison with the previous US	Histologic subtype	Histologic grade	Molecular Subtype
1	2009	53	3	11	9	New	Invasive lobular carcinoma	Low	Luminal A
2	2009	46	4a	4	13	New	Ductal carcinoma in situ	Low	Luminal A
3	2009	69	4a	11	9	New	Invasive ductal carcinoma	Low	Triple negative
4[a]	2010	56	4c	10	15	New	Invasive ductal carcinoma	Intermediate	Luminal A
5	2010	47	4c	9	12	Initial	Mixed invasive ductal and lobular carcinoma	Low	Luminal A
6[b]	2011	57	4a	6	7	New	Ductal carcinoma in situ	Low	Luminal A
7	2011	52	4a	9	7	Initial	Invasive ductal carcinoma	Intermediate	Triple negative
8	2011	50	4c	6	5	Initial	Invasive lobular carcinoma	Low	Luminal A
9	2011	47	4a	14	14	New	Ductal carcinoma in situ	Low	Luminal A
10	2011	42	5	15	15	Initial	Invasive ductal carcinoma	Low	Luminal A
11	2012	58	4a	12	5	Increased	Invasive apocrine carcinoma	Intermediate	Triple negative
12	2012	44	4a	5	8	New	Mucinous carcinoma	Low	Luminal A

Initial: There was no previous US exams.
4[a]: Figure 4 is this case. 6[b]: Figure 3 is this case.
All screening US-detected cancers were node-negative.

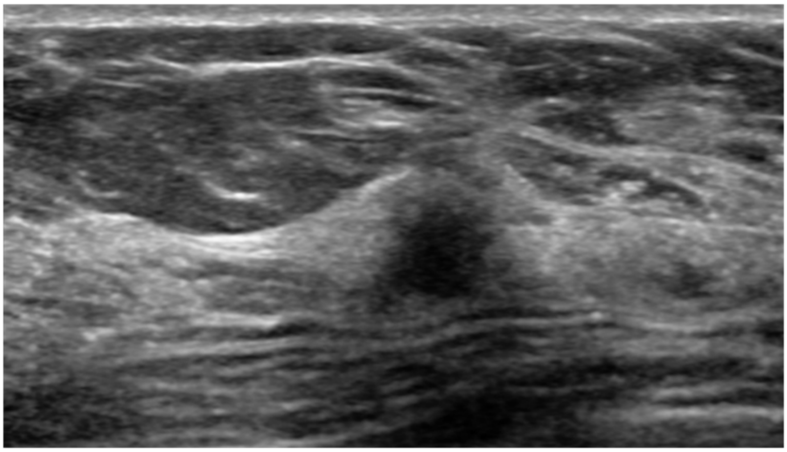
(a)



(b)

Fig. 3. 57-year-old woman with supplemental screening US-detected breast cancer (Ductal carcinoma in situ, low histologic grade and luminal A subtype). Transverse (a) and longitudinal (b) gray-scale US images show a 6mm-sized irregular hypoechoic solid mass with microlobulated margins. It was assigned BI-RADS category 4a.

(a)



(b)

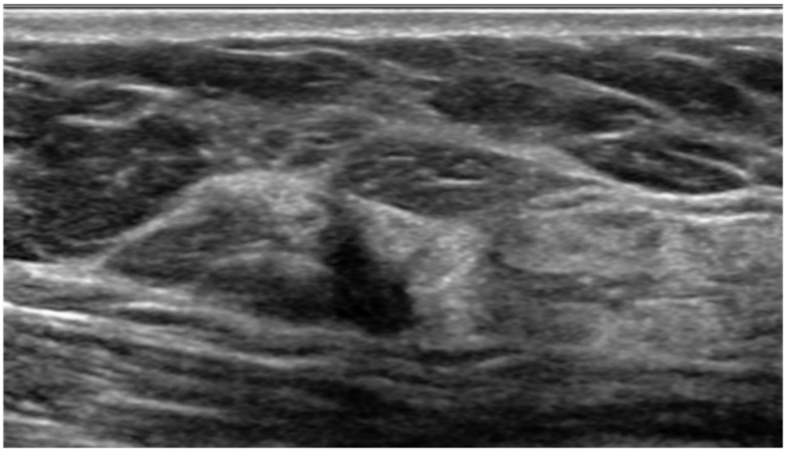


Fig. 4. 56-year-old woman with supplemental screening US-detected breast cancer (invasive ductal carcinoma, intermediate histologic grade and luminal A subtype).

Transverse (a) and longitudinal (b) gray-scale US images show a 10mm-sized irregular hypoecohic solid mass with speculated margins and posterior shadowing. It was assigned BI-RADS category 4c.

3. PPV and biopsy rate

On the whole, PPVs increased in year 2010-2012 compared to in year 2009 although statistical significance was not achieved (Table 4). In total exams, PPV_2 and PPV_3 showed the increasing trend over 4 years with statistical significance (Table 4). Overall PPVs of initial exams were smaller than those of non-initial exams without statistical significance (PPV_1, 0.6% (4 of 724) vs. 1.2% (8 of 654), p=0.301; PPV_2, 3.0 % (4 of 132) vs. 8.9% (7 of 79), p= 0.135; PPV_3, 3.1% (4 of 131) vs. 9.0% (7 of 78), p=0.098).

338 lesions (127 (10.9%) of 1167 BI-RADS 3 lesions, and all BI-RADS 4 (n=210) and 5 (n=1) lesions) underwent US-guided 14-gauge core needle biopsy. Biopsy results consisted of 12 cancers (3 DCIS and 9 invasive cancers), 10 high-risk lesions (5 atypical ductal hyperplasia, 2 atypical papilloma, 2 phyllodes tumor, and 1 radial scar), and 316 benign lesions. Surgery was performed for 12 cancers and 10 high-risk lesions, and 10 high-risk lesions were finally confirmed to be benign by surgery. In total and non-initial exams, the biopsy rates of year 2010-2012 were significantly decreased compared to those of year 2009, and they significantly decreased over 4 years from 2009 to 2012 (Table 4). In initial exams, the biopsy rate of year 2010-2012 did not decreased compared to year 2009, but they showed the decreasing trend over 3 years from 2010 to 2012 (Table 4). The biopsy rate of initial exams was significantly higher than that of non-initial exams (12.0% (187 of 1556) vs. 4.8% (151 of 3174), p<0.001).

Table 4. Positive predictive values (PPV) and Biopsy rates of screening US

	Year 2009	Year 2010-2012	Year 2010	Year 2011	Year 2012	*P* value[1]	*P* value[2]
			Total exams				
PPV_1							
% (95% CI)	0.5 (0.0-1.1)	1.2 (0.4-1.9)	0.6 (0.1-2.3)	1.5 (0.5-3.6)	1.4 (0.2-5.0)	0.207	0.124
No./total	3/592	9/786	2/319	5/325	2/142		

PPV$_2$						0.126	**0.044**
%	2.3	7.3	3.8	9.8	10.5		
(95% CI)	(0.3-8.0)	(3.4-13.4)	(0.5-13.0)	(3.3-21.4)	(1.3-33.1)		
No./total	3/592	9/786	2/319	5/325	2/141		
PPV$_3$						0.130	**0.045**
%	2.3	7.3	3.8	9.8	10.5		
(95% CI)	(0.3-8.2)	(3.4-13.4)	(0.5-13.0)	(3.3-21.4)	(1.3-33.1)		
No./total	2/86	9/123	2/53	5/51	2/19		
Biopsy rate						**<.001**	**<.001**
%	9.7	5.8	8.4	6.4	2.5		
(95% CI)	(8.3-11.3)	(5.0-6.6)	(6.7-10.2)	(4.9-7.8)	(1.5-3.5)		
No./total	161/1656	177/3074	80/950	72/1132	25/992		
			Initial exams				
PPV$_1$						0.156	0.200
%	0.0	0.9	0.5	1.7	0.0		
(95% CI)	(0.1-1.3)	(0.3-2.3)	(0.0-2.7)	(0.3-4.7)	(0.0-6.6)		
No./total	0/287	4/437	1/201	3/182	0/54		
PPV$_2$,						0.158	0.105
%	0.0	4.9	2.6	8.1	0.0		
(95% CI)	(0.0-7.0)	(1.4-12.2)	(0.1-13.8)	(1.7-21.9)	(0.0-45.9)		
No./total	0/51	4/81	1/38	3/37	0/6		
PPV$_3$						0.297	0.106
%	0.0	4.9	2.6	8.1	0.0		
(95% CI)	(0.0-7.1)	(1.4-12.2)	(0.1-13.8)	(1.7-21.9)	(0.0-45.9)		
No./total	0/50	4/81	1/38	3/37	0/6		
Biopsy rate						0.336	**<.001**[a]
%	11.0	12.7	15.0	13.8	5.1		
(95% CI)	(8.6-13.5)	(10.5-14-8)	(11.4-18.6)	(10.3-17.2)	(1.9-8.4)		
No./total	68/616	119/940	57/380	53/385	9/175		
			Non-initial exams				
PPV$_1$						0.730	0.428
%	1.0	1.4	0.8	1.4	2.3		
(95% CI)	(0.2-2.9)	(0.5-3.3)	(0.0-4.6)	(0.2-5.0)	(0.3-8.0)		
No./total	3/305	5/349	1/118	2/143	2/88		
PPV$_2$						0.438	0.232
%	5.4	11.9	6.7	14.3	15.4		
(95% CI)	(0.7-18.2)	(4.0-25.6)	(0.2-32.0)	(1.8-42.8)	(1.9-45.5)		
No./total	2/37	5/42	1/15	2/14	2/13		
PPV$_3$						0.442	0.233
%	5.6	11.9	6.7	14.3	15.4		
(95% CI)	(0.7-18.7)	(4.0-25.6)	(0.2-32.0)	(1.8-42.8)	(1.9-45.5)		
No./total	2/36	5/42	1/15	2/14	2/13		
Biopsy rate						**<.001**	**<.001**

% (95% CI)	8.9 (7.2-10.7)	2.7 (2.0-3.4)	4.0 (2.4-5.7)	2.5 (1.4-3.7)	2.0 (1.0-2.9)	
No./total	93/1040	58/2134	23/570	19/747	16/817	

CI, confidence interval

P value[1] for difference between year 2009 versus year 2010-2012

P value[2] for the trend over four years

P value[2] <.001[a] means the trend for the biopsy rates of initial exams over three years from 2010 to 2012

4. Diagnostic performances of screening US

In total and non-initial exams, specificity and accuracy significantly increased in year 2010-2012 compared to in year 2009 (Table 5). However, in initial exams, specificity and accuracy were similar between the two periods. Sensitivity and NPV were 100.0% because all detected cancers were assigned as BI-RADS category 3, 4 or 5, and there were no interval cancers among cases assigned as category 1 or 2 during follow-up at least 12 months. AUC of screening US in year 2010-2012 was significantly higher than that in year 2009 (Fig. 5, p<0.001).

Table 5. Diagnostic performances of screening US before and after the application of the downgrade criteria

	Year 2009	**Year 2010-2012**	***P* value**
	Total exams		
Sensitivity			NA
% (95% CI)	100.0 (NA)	100.0 (NA)	
No./total	3/3	9/9	
Specificity			<.001
% (95% CI)	64.4 (62.1-66.7)	74.7 (73.1-76.2)	
No./total	1064/1653	2288/3065	
Negative predictive value			
% (95% CI)	100.0 (NA)	100.0 (NA)	NA
No./total	1064/1064	2288/2288	
Accuracy			<.001

% (95% CI)	64.4 (62.1-66.7)	74.7 (73.2-76.3)	
No./total	1067/1656	2297/3074	
Initial exams			
Sensitivity			NA
% (95% CI)	- (-)	100.0 (NA)	
No./total	0/0	4/4	
Specificity			0.898
% (95% CI)	53.4 (49.5-57.4)	53.7 (50.6-56.9)	
No./total	329/616	503/936	
Negative predictive value			NA
% (95% CI)	100.0 (NA)	100.0 (NA)	
No./total	329/329	503/503	
Accuracy			0.838
% (95% CI)	53.4 (49.5-57.4)	53.9 (50.8-57.1)	
No./total	329/616	507/940	
Non-initial exams			
Sensitivity			NA
% (95% CI)	100.0 (NA)	100.0 (NA)	
No./total	3/3	5/5	
Specificity			<.001
% (95% CI)	70.9 (68.1-73.6)	83.8 (82.3-85.4)	
No./total	735/1037	1785/2129	
Negative predictive value			NA
% (95% CI)	100.0 (NA)	100.0 (NA)	
No./total	735/735	1785/1785	
Accuracy			<.001
% (95% CI)	71.0 (68.2-73.7)	83.9 (82.3-85.4)	
No./total	738/1040	1790/2134	

CI, confidence interval

P value for difference between year 2009 versus year 2010-2012

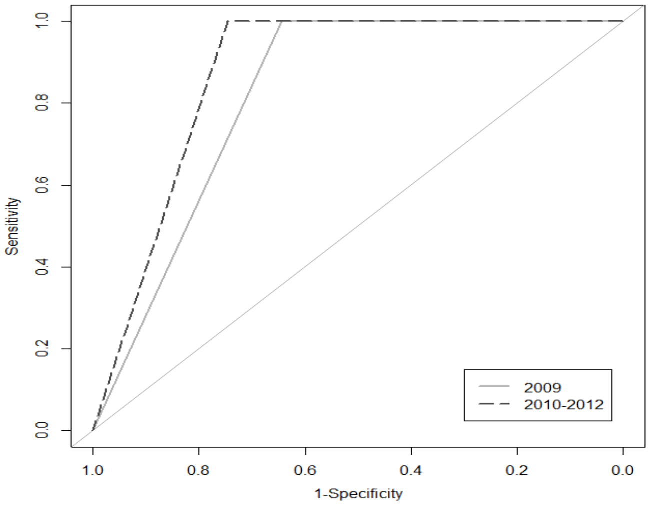


Figure 5. Comparison of the area under the receiver operating characteristic curve before and after the application of the downgrade criteria.
The area under the receiver operating characteristic curve (AUC) was significantly higher after the application of the downgrade criteria (year 2010-2012) compared to before the application of the downgrade criteria (year 2009) (AUC, 0.822 of year 2009 vs. 0.873 of year 2010-2012, p<0.001).

5. Comparison of performances of screening US between less-experienced and experienced

On the whole, category 3-4a rates in less-experienced radiologists were higher than those in experienced radiologists (Table 6). In total and non-initial exams, category 3-4a rates decreased in year 2010-2012 in both less-experienced and experienced radiologists. However, in initial exams, category 3-4a rates did not decrease in both groups. There were no differences in total cancer yields between the two groups. In both groups, category 3-4a rates of initial exams were significantly higher than those of non-initial exams (All, p<0.001).

Table 6. Comparison of the rate of category 3-4a and total cancer yield between less-experienced and experienced

	Year 2009			Year 2010-2012				
	Less-experienced	**Experienced**	***P* value¹**	**Less-experienced**	**Experienced**	***P* value²**	***P* value³**	***P* value⁴**
Total exams								
Category 3-4a%	41.9 (419/999)	25.1 (165/657)	**<.001**	28.7 (623/2171)	17.1 (154/903)	**<.001**	**<.001**	**<.001**
Total cancer yield ‰	1.0 (1/999)	3.0 (2/657)	0.567	3.2 (7/2171)	2.2 (2/903)	0.916	0.448	>.999
Initial exams								
Category 3-4a%	53.1 (208/392)	32.6 (73/224)	**<.001**	50.3 (353/702)	32.4 (77/238)	**<.001**	0.378	0.957
Total cancer yield ‰	0.0 (0/392)	0.0 (0/224)	NA	4.0 (3/702)	4.0 (1/238)	>.999	0.557	>.999
Non-initial exams								
Category 3-4a%	34.8 (211/607)	21.2 (92/433)	**<.001**	18.4 (270/1469)	11.6 (77/665)	**<.001**	**<.001**	**<.001**
Total cancer yield ‰	1.6 (1/607)	4.6 (2/433)	0.574	2.7 (4/1469)	1.5 (1/665)	>.999	>.999	0.566

P value[1]: comparison between less-experienced and experienced in Year 2009
P value[2]: comparison between less-experienced and experienced in Year 2010-2012
P value[3]: comparison of performances of less-experienced between Year 2009 and Year 2010-2012
P value[4]: comparison of performances of experienced between Year 2009 and Year 2010-2012

IV. DISCUSSION

To reduce false positive screening US, complicated cysts and circumscribed oval-shaped solid masses 5mm or smaller were downgraded to BI-RADS category 2 in our practice. This application of downgrade criteria significantly reduced category 3-4a rates in both less-experienced and experienced radiologists. Specificity, accuracy and AUC also increased. Our total cancer yield of 2.5 cancers per 1000 exams was within the reported ranges of 0.3 to 6.8 cancers per 1000 exams (median: 4.2),[9-16] and interval cancers were not detected. There were no differences in total and invasive cancer yields before and after the application of downgrade criteria. Interval cancers were not detected. These results showed that screening US had been effectively performed without a loss of the cancer detection rate.

A complicated cyst 5mm or smaller seen as a circumscribed, homogeneous and hypoechoic lesion was downgraded to BI-RADS category 2. Complicated cysts are cysts with internal debris.[33] When the debris is mobile or a fluid-debris level is seen, they are regarded as benign.[33] But, when the debris is homogeneous and hypoecohic, 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.[32-34] However, breast cancers presenting as a complicated cyst on US was very rare less than 0.5%, with a range of 0-0.4%.[11,32,34] A circumscribed oval-shaped mass 5mm or smaller without any suspicious US features was also considered as category 2. In the previous studies on screening US, there were no malignancies among masses less than 5mm detected on screening US.[14,31] Research from Connecticut 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 5mm or smaller as benign lesions without a loss of sensitivity.[11] 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).[34] When we performed screening US with our downgrade criteria, category 3-4a significantly reduced, and there were no interval cancers. Therefore, complicated cyst and circumscribed oval-shaped solid masses 5mm or smaller can be safely downgraded to BI-RADS 2, and a follow-up after one year is suitable for these lesions.

One of the major problems of screening US is a high rate of BI-RADS category 3.[11,22,34] Our overall category 3 rate during four years was 24.7% (1167 of 4730), which was slightly higher than that of ACRIN 6666 (19.5%)[22] and Connecticut study (20.0%).[11] However, the category 3 rate continuously decreased from 30.4% to 12.4% over four years. Thus, the application of downgrade criteria was effective in reducing category 3. Although the biopsy rate also significantly decreased over 4 years, a total of 338 lesions underwent biopsies, and among these, 316 (93.5%) had benign biopsy results. This high rate of false-positive biopsies, over 90%, is consistent with that reported in previous studies.[11,21,29]

Few studies have evaluated the inter-operator variability in the physician-performed hand-held US.[37,38] Bosch et al. found the high inter-examination agreement in both detection and classification across three radiologists with different experience level; one resident and two senior investigators with each 3 and 5 years of experience.[38] Among 11 breast imaging radiologists with at least 3 years of experience (ranges, 3-26 years) who were trained and qualified for ACRIN 6666, there were moderate agreement for BI-RADS final assessment category.[37] In the study, the comparison analysis according to the experience level of the radiologists was not performed.[37] In our study, less-experienced radiologists assigned more numbers of BI-RADS category 3-4a, but total cancer yields were not different between less-experienced and experienced radiologists.

In total and non-initial exams, category 3-4a decreased after the application of the downgrade criteria, but in initial exams, category 3-4a did not decrease: these contrasting results between non-initial and initial exams are considered because the radiologists might have more attention or more passive attitude when they performed initial exams.

Our total cancer yield of 2.5 cancers per 1000 exams was similar to the results of the studies performed at Connecticut (1.8-3.2 cancers per 1000 exams), which may reflect the performance of screening US performed on the general population who are at average risk of breast cancer (e.i. women without known risk factors of breast cancer other than dense breast).[11,12,16] 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 1.0cm in size (range, 6.5-19mm),[9-11,13,14,16] consistent with our results. In non-initial exams (incidence screening), cancer detection rate, positive test rate (BI-RADS 3-5), and biopsy rate were lower, and PPVs were higher compared to initial exams (prevalence screening) with statistical significance for positive test rate and biopsy rate. These results were consistent with ACRIN 6666 results.[21]

There were several limitations in our study. First, this study was retrospectively conducted in a single institution, third-referral center by the breast radiologists. Generalization of the results may be limited for other study populations, and for exams 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. Second, inter-observer variability among different radiologists may affect the results. More systematic training programs using videos, still images, and tests may lead to better results of reduction in the rate of category 3-4a. Third, our downgrade criteria was based on gray-scale US images. Elastography was not used in this study. There have been efforts to reduce false positives using the addition of elastography to gray-scale US.[39,40] Future large-scale, multicenter prospective study to consider elastographic images in the development of downgrade criteria is needed.

V. CONCLUSION

In conclusion, the application of downgrade criteria reduced BI-RADS category 3-4a without a loss of cancer detection.

REFERENCES

1. Tabar L, Yen M-F, Vitak B, Chen H-HT, Smith RA, Duffy SW. Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. The Lancet 2003;361:1405-10.
2. Moss SM, Cuckle H, Evans A, Johns L, Waller M, Bobrow L. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. The Lancet 2006;368:2053-60.
3. Mandelson MT, Oestreicher N, Porter PL, White D, Finder CA, Taplin SH, et al. Breast density as a predictor of mammographic detection: comparison of interval-and screen-detected cancers. Journal of the National Cancer Institute 2000;92:1081-7.
4. Dehkordy SF, Carlos RC. Dense breast legislation in the United States: state of the States. Journal of the American College of Radiology 2013;10:899-902.
5. Crystal P, Strano SD, Shcharynski S, Koretz MJ. Using sonography to screen women with mammographically dense breasts. American Journal of Roentgenology 2003;181:177-82.
6. De Felice C, Savelli S, Angeletti M, Ballesio L, Manganaro L, Meggiorini M, et al. Diagnostic utility of combined ultrasonography and mammography in the evaluation of women with mammographically dense breasts. Journal of ultrasound 2007;10:143-51.
7. Brancato B, Bonardi R, Catarzi S, Iacconi C, Risso G, Taschini R, et al. Negligible advantages and excess costs of routine addition of breast ultrasonography to mammography in dense breasts. Tumori 2007;93:562.
8. Corsetti V, Houssami N, Ferrari A, Ghirardi M, Bellarosa S, Angelini O, et al. Breast screening with ultrasound in women with mammography-negative dense breasts: evidence on incremental cancer detection and false positives, and associated cost. European journal of cancer 2008;44:539-44.
9. Berg WA, Blume JD, Cormack JB, Mendelson EB, Lehrer D, Böhm-Vélez M, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. Jama 2008;299:2151-63.
10. Leong LC, Gogna A, Pant R, Ng FC, Sim LS. Supplementary Breast Ultrasound Screening in Asian Women with Negative But Dense Mammograms—A Pilot Study. Annals of the Academy of Medicine-Singapore 2012;41:432.
11. Hooley RJ, Greenberg KL, Stackhouse RM, Geisel JL, Butler RS, Philpotts LE. Screening US in patients with mammographically dense

breasts: initial experience with Connecticut Public Act 09-41. Radiology 2012;265:59-69.
12. Parris T, Wakefield D, Frimmer H. Real world performance of screening breast ultrasound following enactment of Connecticut Bill 458. The breast journal 2013;19:64-70.
13. Girardi V, Tonegutti M, Ciatto S, Bonetti F. Breast ultrasound in 22,131 asymptomatic women with negative mammography. The Breast 2013;22:806-9.
14. Moon H, Jung I, Park S, Kim M, Youk J, Kim E. Comparison of Cancer Yields and Diagnostic Performance of Screening Mammography vs. Supplemental Screening Ultrasound in 4394 Women with Average Risk for Breast Cancer. Ultraschall in der Medizin-European Journal of Ultrasound 2015;36:255-63.
15. 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 Radiologica 2015;56:652-8.
16. Weigert J, Steenbergen S. The Connecticut experiment: the role of ultrasound in the screening of women with dense breasts. The breast journal 2012;18:517-22.
17. Li J, Zhang B-N, Fan J-H, Pang Y, Zhang P, Wang S-L, et al. A nation-wide multicenter 10-year (1999-2008) retrospective clinical epidemiological study of female breast cancer in China. BMC cancer 2011;11:364.
18. Shen S, Zhou Y, Xu Y, Zhang B, Duan X, Huang R, et al. A multi-centre randomised trial comparing ultrasound vs mammography for screening breast cancer in high-risk Chinese women. British journal of cancer 2015;112:998-1004.
19. Ishida T, Suzuki A, Kawai M, Narikawa Y, Saito H, Yamamoto S, et al. A Randomized Controlled Trial to Verify the Efficacy of the Use of Ultrasonography in Breast Cancer Screening Aged 40–49 (J-START): 76 196 Women Registered. Japanese journal of clinical oncology 2014:hyt199.
20. Scheel JR, Lee JM, Sprague BL, Lee CI, Lehman CD. Screening ultrasound as an adjunct to mammography in women with mammographically dense breasts. American journal of obstetrics and gynecology 2015;212:9-17.
21. Berg WA, Zhang Z, Lehrer D, Jong RA, Pisano ED, Barr RG, 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.
22. Barr RG, Zhang Z, Cormack JB, Mendelson EB, Berg WA. 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.
23. Gordon PB, Goldenberg SL. Malignant breast masses detected only by ultrasound. A retrospective review. Cancer 1995;76:626-30.
24. Buchberger W, Niehoff A, Obrist P, DeKoekkoek-Doll P, Dünser M. Clinically and mammographically occult breast lesions: detection and classification with high-resolution sonography. Seminars in Ultrasound, CT and MRI: Elsevier; 2000. p.325-36.
25. Kaplan SS. Clinical Utility of Bilateral Whole-Breast US in the Evaluation of Women with Dense Breast Tissue 1. Radiology 2001;221:641-9.
26. Kolb TM, Lichy J, Newhouse JH. Comparison of the Performance of Screening Mammography, Physical Examination, and Breast US and Evaluation of Factors that Influence Them: An Analysis of 27,825 Patient Evaluations 1. Radiology 2002;225:165-75.
27. Leconte I, Feger C, Galant C, Berlière M, Berg BV, D'Hoore W, et al. Mammography and subsequent whole-breast sonography of nonpalpable breast cancers: the importance of radiologic breast density. American Journal of Roentgenology 2003;180:1675-9.
28. Kelly KM, Dean J, Comulada WS, Lee S-J. Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. European radiology 2010;20:734-42.
29. Merry GM, Mendelson EB. Update on Screening Breast Ultrasonography. Radiologic Clinics of North America 2014;52:527-37.
30. Kim MJ, Kim E-K, Park SY, Jung HK, Oh KK, Seok JY. Galactoceles mimicking suspicious solid masses on sonography. Journal of ultrasound in medicine 2006;25:145-51.
31. Kim SJ, Chang JM, Cho N, Chung S-Y, Han W, Moon WK. Outcome of breast lesions detected at screening ultrasonography. European journal of radiology 2012;81:3229-33.
32. Daly CP, Bailey JE, Klein KA, Helvie MA. Complicated breast cysts on sonography: is aspiration necessary to exclude malignancy? Academic radiology 2008;15:610-7.
33. Berg WA, Sechtin AG, Marques H, Zhang Z. Cystic breast masses and the ACRIN 6666 experience. Radiologic clinics of North America 2010;48:931-87.
34. Berg WA, Zhang Z, Cormack JB, Mendelson EB. Multiple bilateral circumscribed masses at screening breast US: consider annual follow-up. Radiology 2013;268:673-83.
35. American College of Radiology. Breast imaging reporting and data system (BI-RADS). 4th ed. Reston, VA: American College of Radiology; 2003.
36. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a

nonparametric approach. Biometrics 1988:837-45.
37. Berg WA, Blume JD, Cormack JB, Mendelson EB. Operator Dependence of Physician-performed Whole-Breast US: Lesion Detection and Characterization 1. Radiology 2006;241:355-65.
38. Bosch A, Kessels A, Beets G, Vranken K, Borstlap A, Von Meyenfeldt M, et al. Interexamination variation of whole breast ultrasound. The British journal of radiology 2014;DOI: http://dx.doi.org/10.1259/bjr/17252624
39. Berg WA, Cosgrove DO, Doré CJ, Schäfer FK, Svensson WE, Hooley RJ, et al. Shear-wave elastography improves the specificity of breast US: the BE1 multinational study of 939 masses. Radiology 2012;262:435-49.
40. Lee SH, Chang JM, Kim WH, Bae MS, Seo M, Koo HR, et al. Added Value of Shear-Wave Elastography for Evaluation of Breast Masses Detected with Screening US Imaging. Radiology 2014;273:61-9.

ABSTRACT(IN KOREAN)

검진 유방촬영술에서 치밀 유방을 가진 여성에서 병용된 유방 초음파의 진단 성적: 등급 하향 기준의 적용

<지도교수 김은경>

연세대학교 대학원 의학과

김 수 연

목적: 검진 유방촬영술에서 이상 소견이 없으나 치밀 유방을 가진 여성들에게 병용된 유방초음파에서 등급 하향 기준을 적용하는 것이 유방암의 발견율을 감소시키지 않으면서 유방영상 판독과 자료 체계 (Breast Imaging Reporting and Data System : BI-RADS) 범주 3-4a 를 감소시키는지 알아보고자 하였다.

방법: 이 연구는 후향적 연구로 검진 유방촬영술에서 이상소견은 없었으나 치밀 유방이 있었고 2009년 3월부터 2013년 2월까지 검진 유방 초음파를 병용하여 시행한 4730명의 여성을 포함하였다. 유방초음파는 영상의학과 의사에 의해서 수동으로 시행되었다. 검진 유방 초음파에서 BI-RADS 범주 3-4a 를 감소시키기 위해서 2010년 3월부터 5mm 미만의 복합 낭종과 경계가 좋은 난원형의 고형 종괴를 BI-RADS 범주 2 로 하향하는 기준을 적용하였다. 병리 결과와 적어도 12개월 이상의 추적 초음파 검사 결과를 참조 기준으로 하여 4년간의 검진 유방초음파의 BI-RADS 범주의 분포, 유방암의 발견율, 진단

성적이 분석되었고, 하향 기준 적용 전 (2009년도: 2009년 3월부터 2010년 2월까지)과 후(2010-2012년도: 2010년 3월부터 2013년 2월까지)의 성적이 비교되었다. 경험이 부족한 영상의학과 의사군 (1년에서 2년의 경험을 가진 19명의 강사)과 경험이 많은 영상의학과 의사군 (12년에서 18년의 경험을 가진 3명의 교수) 간의 성적이 비교되었다.

결과: 5mm 미만의 복합 낭종과 경계가 좋은 난원형의 고형 종괴를 BI-RADS 범주 2 로 하향하는 기준을 적용한 결과, 경험이 부족한 의사군과 경험이 많은 의사군 모두에서 BI-RADS 범주 3-4a 가 유의하게 감소하였다. 검진 초음파는 4년간 12개의 유방암을 발견하였고, 그 중에 3개는 상피내암이었고 9개는 침윤성 유방암이었다. 등급 하향 기준의 적용 전과 후에 유방암 발견율에는 차이가 없었다. 간격암은 발견되지 않았다. 등급 하향 기준을 적용한 이후에 특이도와 정확도, 진단능은 유의하게 증가하였다.

결론: 등급 하향 기준의 적용은 유방암의 발견율을 감소시키지 않으면서 BI-RADS 범주 3-4a 를 유의하게 감소시켰다.

핵심되는 말 : 검진 유방촬영술, 치밀 유방, 검진 유방 초음파, 유방영상 판독과 자료 체계 (Breast Imaging Reporting and Data System)