European Radiology (2019) 29:4468–4476
https://doi.org/10.1007/s00330-018-5934-9

BREAST

Effect of training on ultrasonography (US) BI-RADS features for radiology residents: a multicenter study comparing performances after training

Jung Hyun Yoon[1] · Hye Sun Lee[2] · You Me Kim[3] · Ji Hyun Youk[4] · Sung Hun Kim[5] · Sun Hye Jeong[6] · Ji Young Hwang[7] · Jin Hee Moon[7] · Young Mi Park[8] · Min Jung Kim[1]

Received: 31 July 2018 / Revised: 17 October 2018 / Accepted: 29 November 2018 / Published online: 7 January 2019

Abstract

Objectives To evaluate the effect of training radiology residents on breast ultrasonography (US) according to the Breast Imaging Reporting And Data System (BI-RADS) and the factors that influence the training effect.

Methods This multicenter, prospective study was approved by eight institutional review boards. From September 2013 to July 2014, 248 breast masses in 227 women were included for US image acquisition. Representative B-mode and video images of the breast masses were recorded, among which 54 cases were included in the education set and 66 in the test set. Sixty-one radiology residents scheduled for breast imaging training individually reviewed the test set, immediately before, 1 month after, and 6 months after training. Diagnostic performances and US descriptors of the residents were evaluated and compared against those of expert radiologists.

Results Agreements between residents and experienced radiologists showed improvement after training, while agreements between post-training and post-6-month training descriptors did not show significant differences (all $p > 0.05$, respectively). Sensitivity, negative predictive value (NPV), and AUC were significantly improved for residents post-training and post-6-month training (all $p < 0.05$), while approximating the performances of expert radiologists except for AUC (0.836, 0.840, and 0.908, respectively, $p < 0.05$). Low levels of pre-training AUC, total number of breast US examinations, and the number of sessions per week that residents were involved in were factors influencing the improvement of AUC.

Conclusion Training using education material dedicated for breast US imaging effectively improved the diagnostic performances of radiology residents and agreements with experienced radiologists on US BI-RADS features.

Key Points

- *Agreements on lesion descriptors between residents and experienced radiologists showed improvement after training, regardless of test point.*
- *Sensitivity, NPV, and AUC were significantly improved for residents in post-training and post-6-month training (all $p < 0.05$).*
- *Low levels of pre-training AUC, total number of breast US examinations, and the number of sessions per week that residents were involved in were factors influencing the improvement of AUC.*

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00330-018-5934-9) contains supplementary material, which is available to authorized users.

✉ Min Jung Kim
mines@yuhs.ac

1 Department of Radiology, Research Institute of Radiological Science, Yonsei University, College of Medicine, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, South Korea

2 Biostatistics Collaboration Unit, Medical Research Center, Yonsei University, College of Medicine, Seoul, South Korea

3 Department of Radiology, Dankook University Hospital, Dankook University, College of Medicine, Cheonan, South Korea

4 Department of Radiology, Gangnam Severance Hospital, Yonsei University, College of Medicine, Seoul, South Korea

5 Department of Radiology, Seoul St. Mary's Hospital, Seoul, South Korea

6 Department of Radiology, Soonchunhyang University, Bucheon Hospital, Bucheon, South Korea

7 Department of Radiology, Kangnam Sacred Hospital, Hallym University, College of Medicine, Seoul, South Korea

8 Department of Radiology, Inje University Busan Paik Hospital, Inje University College of Medicine, Gimhae, South Korea

Keywords Breast · Ultrasound · Education · Residency · Training

Abbreviations

ACR	American College of Radiology
AUC	Area under the receiver operating characteristic curve
BI-RADS	Breast Imaging Reporting And Data System
NPV	Negative predictive value
PPV	Positive predictive value
US	Ultrasonography

Introduction

Breast ultrasonography (US) is currently used as an adjunctive imaging modality to mammography, a primary imaging modality for women < 40 years, and for imaging guidance during percutaneous biopsy procedures [1]. In the past, US was used to differentiate solid masses from benign cysts among mammography-detected abnormalities, but with the application of high-resolution transducers, it is currently used to detect breast masses more comprehensively. Nowadays, the clinical utilization of breast US is not just limited to abnormalities identified on other imaging modalities or physical examinations, but has expanded to breast cancer screening in selected populations [2, 3].

One major limitation of breast US is that lesion detection and image interpretation rely considerably on the radiologist. The American College of Radiology Breast Imaging Reporting And Data System (ACR BI-RADS) lexicon [4] has helped standardize image description and interpretation reports, but even still, interobserver agreement ranges from poor to moderate among radiologists [5–9]. This variation in lesion description and assessment critically affects patient management, as false-positive results lead to additional imaging or invasive biopsies that increase medical costs, not to mention patient anxiety. Several reports have proven that training on the US BI-RADS lexicon improves lesion characterization and interpretation of US examinations [10–12], but little has been reported on how training affects the performances of radiology residents, especially on the long-term outcomes in performance after training. As the requirement for breast US increases, the importance of appropriate training for radiology residents on breast US has also increased, but consensus has not been reached regarding the adequate training curriculums or requirements; the ACR Society specifies the training duration of 3 months in the breast radiology department, but the number of examinations required during training is not specified, while the European Federation of Societies for Ultrasound in Medicine and Biology (EFSUMB) guideline specifies the number of breast examinations according to the levels of trainees that should be undergone for adequate training [13, 14]. Also, the training effect of the present radiology curriculum has yet to be investigated.

Based on this, the purpose of this study was to evaluate the effect of training radiology residents on breast US and to reveal factors that may influence this training effect in the radiology residency curriculum.

Materials and methods

This multicenter, prospective study was conducted at eight institutions of the Korean Society of Breast Imaging & Korean Society for Breast Screening (KSBI & KSFBS) and was approved by each institutional review board. Informed consent for video recording during US examinations was obtained from all patients. Informed consent was also obtained from all radiology residents who volunteered to participate in this study.

Patients

From September 2013 to July 2014, women with visible breast masses on US examinations were candidates for this study if any of the following criteria was met: women who (1) were scheduled for US-guided biopsy, (2) were scheduled for surgery or vacuum-assisted excision, (3) had been followed for more than 2 years after benign biopsy results, or (4) had masses showing typically benign features on US which had not been biopsied were candidates for this study. Two hundred and forty-eight breast masses among 227 women (mean age 47.3 years, range 19–81 years) from five institutions were finally included in this study. Mean size of the 248 masses was 15.9 mm (range 4–60 mm). Of the 248 masses, 71 (24.6%) were malignant, and 177 (71.4%) were benign.

Breast US examinations

An in-house US machine was used for breast US examinations (*i*U22, PhilipsMedical Systems; GE LOGIQ E9, GE Medical System; SuperSonic Imagine; EUB-8500, Hitachi Medical) that was equipped with a high-frequency linear-array transducer. Seven board-certified radiologists with 3 to 13 years of experience in breast imaging were involved in image acquisition. After localizing the target breast mass during US examination, representative transverse and longitudinal B-mode scans of the breast mass were obtained. The maximum diameter of the mass was measured and recorded for data analysis. Video images were recorded at both transverse and longitudinal axes of the mass, with scans starting at the normal breast parenchyma at one end of the mass and moving in one direction to include the entire mass until reaching the normal breast parenchyma located at the other end of the mass at which the video scan would end. The B-mode scans and video images were collected separately in a hard drive and underwent processing to eliminate any patient-related information within the images prior to review.

Image selection and interpretation

US and video images of the 248 breast masses were reviewed by 2 radiologists dedicated to breast imaging (M.J.K. and J.H.Y.), who classified the images into three groups: (1) the education set, cases that had typical US features or that were representative of special cases described in the ACR BI-RADS [4]; (2) the test set, cases that were considered suitable for testing image interpretation; and (3) the inadequate file set, B-mode scans with artifacts or video images that were recorded at speeds too slow or fast for interpretation or those that did not include scanning of the entire breast mass. The test set contained cases containing representative BI-RADS descriptors that the residents were being tested on (Table 1). Associated features in BI-RADS [4] were not considered when selecting images because these associated features mostly require real-time examinations of the whole breast and areas located beyond the scanning range of the mass. Of the 248 breast masses, 102 were selected for the education set, 66 were selected for the test set, and 80 were included in the inadequate file set. For each breast mass, B-mode representative images and video images were displayed as PowerPoint files (Microsoft).

Table 1 Proportion of the US descriptors and final assessment used for image analysis among the 66 cases of the test set

Descriptor	Contents	*N* (%)
Shape	Oval	37 (56.1)
	Round	11 (16.6)
	Irregular	18 (27.3)
Orientation	Parallel	39 (59.1)
	Non-parallel	27 (40.9)
Margin	Circumscribed	25 (37.9)
	Indistinct	9 (13.6)
	Angular	6 (9.1)
	Microlobulated	13 (19.7)
	Spiculated	13 (19.7)
Echo pattern	Anechoic	1 (1.5)
	Hyperechoic	2 (3.0)
	Complex cystic and solid	5 (7.6)
	Hypoechoic	42 (63.7)
	Isoechoic	7 (10.6)
	Heterogeneous	9 (13.6)
Posterior features	Absent	42 (63.6)
	Enhancement	14 (21.2)
	Shadowing	6 (9.1)
	Combined	4 (6.1)
Calcifications	No calcifications	53 (80.3)
	Calcifications in mass	13 (19.7)
	Calcifications out of mass	0 (0.0)
	Intraductal calcifications	0 (0.0)

Associated features such as architectural distortion, ductal changes, skin changes, edema, or vascularity/elastographic features were not included in analysis

Six dedicated breast radiologists from the five institutions held regular meetings to review the cases selected for the education and test sets. During the review sessions, PowerPoint slides containing both B-mode scans and video clips for each case were displayed on a computer monitor for image analysis. Each case was analyzed according to the lesion descriptors and final assessment categories (2, 3, 4a, 4b, 4c, and 5) of the fifth edition of ACR BI-RADS [4], and data were first recorded individually by each of the six radiologists and then in consensus after discussion. During image review, all radiologists were blinded to clinical, histopathologic, and mammographic findings to minimize biases that could affect US interpretation.

According to the consensus data obtained from the image interpretation meetings, the 102 cases from the education set were organized into educational material consisting of a lexicon review including typical cases or US features described in the ACR BI-RADS [4], and case reviews with video files and feature analyses illustrated according to the consensus results reached by the six experienced radiologists. Figure 1 shows the format in which case review slides were organized. The 66 cases of the test set were organized similarly, but without the consensus results (Supplementary 1).

The 66 cases included in the test set consisted of 21 (31.8%) cancers and 45 (68.2%) benign masses. Of the 21 malignant masses, 19 were invasive carcinomas (mean size, 13.9 mm; range, 7–22 mm) and 2 were ductal carcinoma in situ. The 45 benign cases consisted of 1 atypical ductal hyperplasia confirmed on surgical pathology, 1 adenosis, 15 fibroadenomas, 11 fibrocystic changes, 5 galactoceles, 1 hamartoma, 4 intraductal papillomas, 1 intramammary lymph node (benign), 2 stromal fibroses, and 4 simple cysts.

Image review by the radiology residents

Our Radiology Society recommends minimal requirement of 300 cases for mammography interpretation, 150 cases for breast US examinations, and 15 cases for US-guided biopsy under supervision during the 4 years of residency. From August 2014 to October 2016, 61 radiology residents from

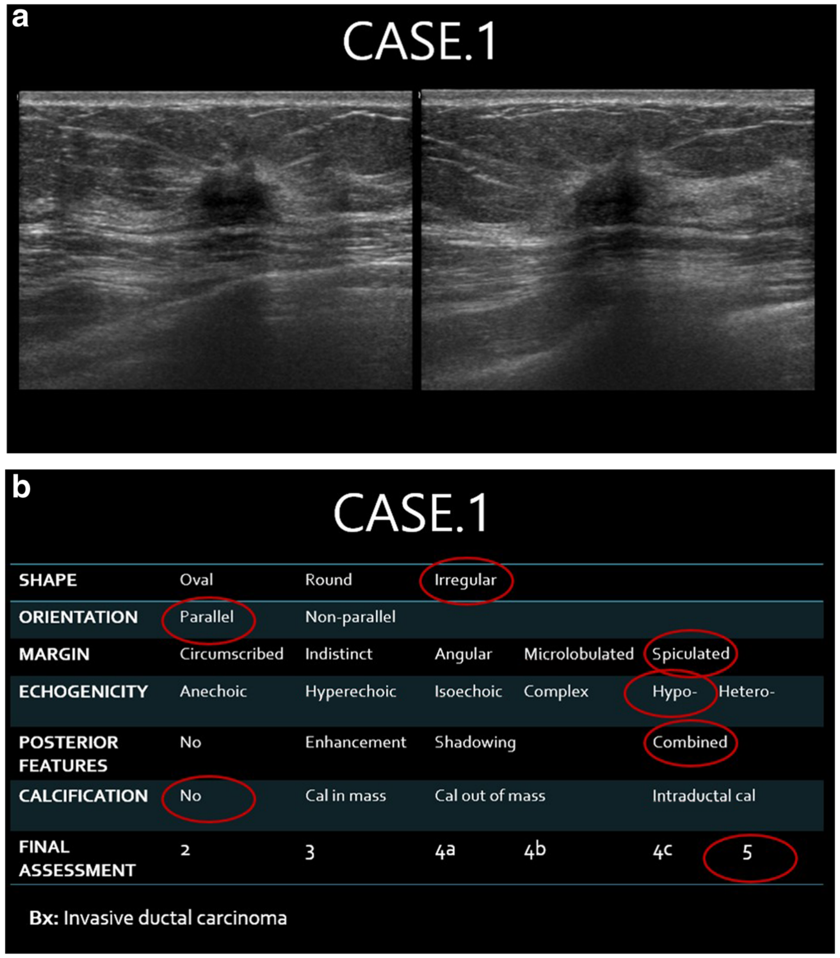

Fig. 1 Example of the education material used for training. Each case consists of three slides, with the first slide containing representative B-mode transverse and longitudinal images (**a**) and the second slide displaying two axis-view video images side by side (not shown in this figure). Expert consensuses on lesion description and pathologic diagnosis are displayed in the third slide (**b**)

8 institutions who were scheduled for training in the breast imaging section were enrolled in this study. Before training (pre-training), the residents were asked to review the 66 cases of the test set and fill out an Excel data sheet (Microsoft) with the US BI-RADS descriptor and imaging category that was considered the most appropriate for each lesion [4]. After filling out the data sheet for pre-training data, the educational material made with the education set was distributed and residents were asked to review the material freely during their training month in breast imaging. After training, the same test set was reviewed and the data sheet was filled out by the residents twice, first immediately after the scheduled training period ended (post-training) and then 6 months after the training ended (post-6-month training). Information regarding the training environment of each resident was obtained with a questionnaire, with the questionnaire requesting information regarding the year of residency, training experience in breast imaging, the number of breast examinations performed during training, and the number of staff radiologists dedicated to breast imaging at each institution.

Data and statistical analysis

For data analysis, BI-RADS categories 2 and 3 were considered negative, and BI-RADS categories 4a to 5 were considered positive. Diagnostic performances of the residents and experienced breast radiologists were calculated, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy. Area under the receiver operating characteristic (ROC) curves (AUCs) was calculated. Diagnostic performances were compared between the residents' pre-training, post-training, and post-6-month analytic test set data and the expert radiologists' consensus. Forest plots were used to display the tasks performed using GLIMMIX procedures to test the equality of group means.

To evaluate factors that affected the diagnostic performances of the residents during training, univariate and multivariate analyses were performed using the improvement of AUC as a dependent variable. Spearman's correlation analysis was performed to evaluate the correlation between the total number of breast US examinations performed and the number of training sessions for breast US per week. Kappa analysis was used to evaluate the agreement between residents and experienced radiologists in selecting US descriptors and final assessment categories: kappa values ranging from 0.00 to 0.20 indicated slight agreement, from 0.20 to 0.40 indicated fair agreement, from 0.40 to 0.60 indicated moderate agreement, from 0.60 to 0.80 indicated substantial agreement, and from 0.80 to 1.00 indicated almost perfect agreement. Differences in kappa values were calculated by the bootstrapping method for comparison.

Statistical analysis was performed using SAS software (version 9.2, SAS Inc.) and R package (version 3.2.5, http://www.R-project.org). All p values < 0.05 were considered to have statistical significance.

Results

Table 2 summarizes the demographics and training environments of the 61 residents included in this study. Mean time interval from the pre-training analysis to the post-training analysis and from the pre-training analysis to the post-6-month training analysis was 30.8 days ± 13.1 (range, 10–86 days) and 176.3 days ± 44.9 (range, 90–306 days), respectively. Thirty-two (50.8%) residents had 3 or more staff radiologists in the breast imaging section of their institution. Mean number of breast US examinations performed by the residents during their training month was 33.9 cases (standard deviation, 26.6). Fifty (82.0%) residents had more than 4 sessions of breast US during their training month, and 35 (56.4%) performed more than 5 cases of breast US examinations per session under the supervision of staff radiologists.

Table 2 Demographics and training environments of the 61 residents enrolled in this study

Characteristics*	*N* (%)
Years of residency	
1–2 years	35 (57.4)
3–4 years	26 (42.6)
First time in breast imaging training?	
Yes	29 (47.5)
No	32 (52.5)
No. of staff radiologists in the breast imaging section?	
1 or 2	30 (49.2)
3	8 (13.1)
4	4 (6.6)
5	19 (31.1)
How many training session of breast US per week?	
1–3 sessions per week	11 (18.0)
4–6 sessions per week	33 (54.1)
7–10 sessions per week	17 (27.9)
How many cases of breast US performed by the resident during each session?	
0–5 cases per session	26 (42.6)
6–10 cases per session	26 (42.6)
More than 10 cases per session	9 (14.8)

*Answered by the residents filling out a questionnaire

Agreement between experienced radiologists and residents on US descriptors

Agreements on US descriptors between the residents and the experienced radiologists are summarized in Supplementary data 2. Agreements between the residents and the experienced radiologists were significantly improved both post-training and post-6-month training for orientation, echogenicity, margin, and calcifications (all $p < 0.05$, respectively). κ values did not show significant differences between the post-training agreement and the post-6-month agreement for US descriptors (all $p > 0.05$, respectively).

Diagnostic performances of the 61 residents according to the test point

Table 3 summarizes the diagnostic performances of the experienced radiologists and the 61 residents according to the test point. Post-training sensitivity, NPV, and AUC were significantly improved compared to pre-training performances, while specificity, PPV, and accuracy were significantly decreased (all $p < 0.05$). Post-6-month training performances showed a significant improvement in sensitivity, NPV, and AUC when compared to pre-training performances and did not show significant differences when compared to post-training performances (all $p > 0.05$). No significant differences were seen between the performances of the experienced radiologists and the residents post-training and post-6-month training (Supplementary data 3). Although the residents' post-training and post-6-month training AUC improved, the AUC of the experienced radiologists (0.908) was still significantly higher (all $p < 0.001$, Fig. 2).

Factors affecting improvements in the diagnostic performances of the residents during training

Of the 61 residents included, 21 (34.4%) showed a significant improvement in AUC, and the remaining 40 (65.6%) did not show changes after training (Table 4). Residents whose AUC improved included a significantly higher number of residents with pre-training AUC below the mean AUC of the total 61 residents (57.5% vs. 23.8%, $p = 0.025$), residents who performed a higher total number of breast US examinations each month (38.88 ± 29.00 vs. 24.27 ± 18.50, $p = 0.020$), and residents who performed more training sessions of breast US per week (5.20 ± 1.57 vs. 4.19 ± 1.33, $p = 0.015$). Multivariable analysis showed that a pre-training AUC below the mean

Table 3 Diagnostic performances of the 61 residents included in this study according to the test point

%	Residents Pre-training	p*	Post-training	p*	$p^{\dagger}$	Post-6 months	p*	$p^{\ddagger}$	Experienced radiologists
Sensitivity	91.0 (88.6–93.4)	< 0.001	96.3 (95.0–97.6)	0.520	< 0.001	96.3 (94.7–97.8)	0.583	> 0.999	97.6 (94.9–100.0)
Specificity	48.2 (43.7–52.4)	0.767	40.7 (37.3–44.0)	0.339	< 0.001	39.4 (35.7–43.1)	0.283	0.535	45.9 (37.4–54.5)
PPV	46.4 (44.5–48.2)	0.899	43.8 (42.4–45.2)	0.346	0.010	43.4 (41.8–45.0)	0.309	0.630	46.0 (42.1–49.9)
NPV	93.1 (91.6–94.7)	0.003	96.4 (95.3–97.5)	0.445	< 0.001	96.5 (95.3–97.8)	0.530	0.874	97.8 (95.3–100.0)
Accuracy	61.7 (59.2–64.2)	0.871	58.3 (56.2–60.5)	0.250	0.009	57.5 (55.1–59.8)	0.205	0.508	62.4 (56.7–68.0)
AUC	0.793 (0.773–0.813)	< 0.001	0.836 (0.823–0.848)	< 0.001	< 0.001	0.840 (0.827–0.853)	< 0.001	0.555	0.908 (0.891–0.926)

95% confidence intervals are in parentheses

PPV positive predictive value, *NPV* negative predictive value, *AUC* area under the receiver operating characteristic (ROC) curve

*p values compared to the experienced radiologists

† Compared to pre-training data

‡ Compared to post-training data

value was a significant factor for AUC improvement, with the odds ratio being 4.329 (95% CI 1.325–14.143, $p = 0.015$). As the total number of breast US examinations performed per month and the number of training sessions of breast US per week showed high correlation ($r = 0.71$, $p < 0.001$), an additional multivariable analysis was performed separately to include only one of the two factors described above. On the additional multivariable analysis, a pre-training AUC below the mean value, the total number of breast US examinations per month, and the number of training sessions of breast US per week were statistically significant factors influencing the improvement of AUC (all $p < 0.05$).

Discussion

Our study results show that the diagnostic performances of radiology trainees improve significantly after training under

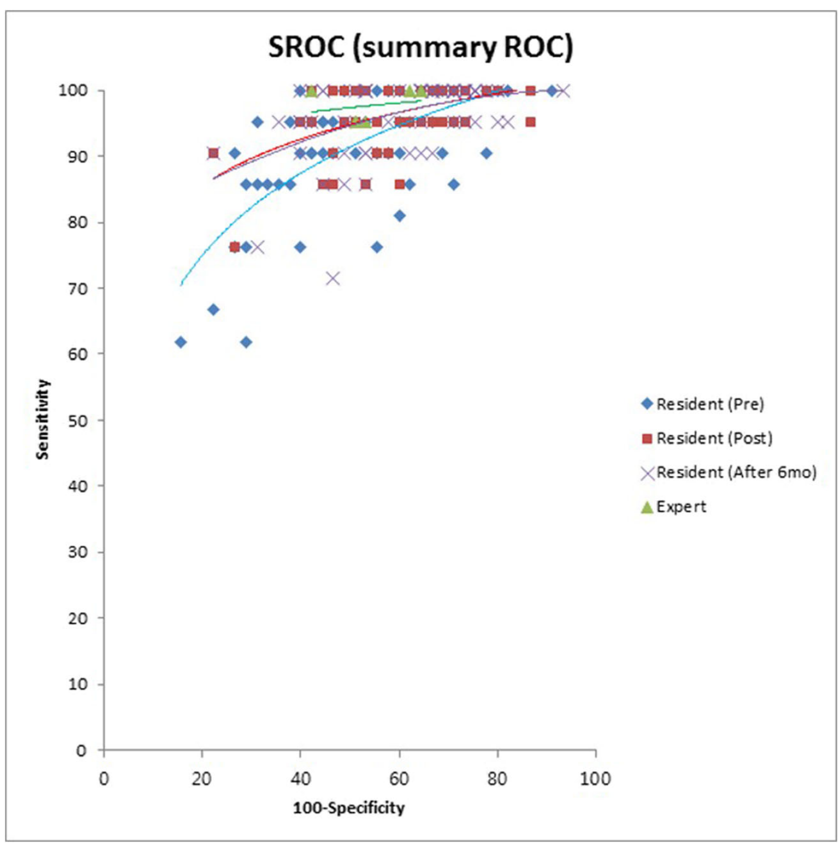

Fig. 2 Summary receiver operating characteristic (summary ROC, SROC) curve for the experienced radiologists and the 61 residents according to each test point. With training, AUCs for post-training (red curve) and post-6-month training (purple curve) are both at significantly higher levels compared to those for pre-training (blue curve), but they do not reach the AUC of the experienced radiologists (green curve)

Table 4 Factors affecting the diagnostic performances of residents during training

	Demographics			p	Univariate analysis		Multivariate analysis		Multivariate analysis†		Multivariate analysis‡	
	Total	No change ($n=21$)	AUC improvement ($n=40$)		OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p
Years in residency				0.568								
1–2 year(s)	35 (57.4%)	11 (52.4%)	24 (60.0%)		–							
3–4 years	26 (42.6%)	10 (47.6%)	16 (40.0%)		0.733 (0.253–2.127)	0.568						
Experience in breast imaging training				0.596								
No	29 (47.5%)	9 (42.9%)	20 (50.0%)		–							
Yes	32 (52.5%)	12 (57.1%)	20 (50.0%)		0.750 (0.259–2.172)	0.569						
Pre-training AUC				0.012								
Above mean	33 (54.1%)	16 (76.2%)	17 (42.5%)		–		–		–	–		
Below mean*	28 (45.9%)	5 (23.8%)	23 (57.5%)		4.329 (1.325–14.143)	0.015	6.068 (1.643–22.419)	0.007	5.836 (1.640–20.776)	0.007	5.630 (1.540–20.581)	0.009
No. of staff radiologists in breast imaging section				0.520								
	3.03 ± 1.527	2.86 ± 1.49	3.13 ± 1.56		1.125 (0.791–1.601)	0.513						
Total no. of breast US examinations performed by the resident per month				0.020								
	33.85 ± 26.65	24.27 ± 18.50	38.88 ± 29.00		1.029 (0.999–1.060)	0.060	1.458 (0.801–2.655)	0.352	1.037 (1.003–1.073)	0.034		
No. of training sessions of breast US per week				0.015								
	4.85 ± 1.56	4.19 ± 1.33	5.20 ± 1.57		1.618 (1.076–2.434)	0.021	1.020 (0.979–1.062)	0.217			1.791 (1.129–2.842)	0.013
No. of breast US performed by the residents per session				0.105								
	6.50 ± 3.42	5.52 ± 2.63	7.01 ± 3.69		1.162 (0.965–1.401)	0.114						

OR odds ratio, *95% CI* 95% confidence interval

*Mean AUC of 0.808 for pre-training test set analysis of the 61 radiology residents

† Multivariable analysis performed including total no. of breast US examinations performed by the resident per month

‡ Multivariable analysis performed including No. of training sessions of breast US per week

the present curriculum, with improvements being especially more prominent in trainees with low levels of pre-training diagnostic performances, and we also found that with training, radiology residents were able to reach similar levels of image description and interpretation to expert radiologists. Starting with the Connecticut Public Act 09-41 [15], the breast density notification legislation has become widely enacted, which in turn has brought about a substantial increase in demand for supplementary screening tools such as breast US or MRI for women with mammographically dense breasts [16]. Adding supplementary US to mammography detects 1.2 to 7.2 additional cancers per 1000 women [2, 3], which has also led to an increase in recall rates and benign biopsies. To minimize these adverse effects of implementing screening US, well-educated radiologists are needed for the standardized interpretation of breast US examinations. As with the positive effect of training on breast US reported for experienced radiologists and residents [10, 12], standardized training programs on breast US for training residents may prove to be beneficial.

While most studies evaluating training have reported immediate results after training, the long-term effects of training have not been well documented, with one study reporting no significant decline in 2–3-month post-training results for mammographic descriptors [17]. Post-6-month training tests were added to this study to see the long-term effects of training, and our results showed that the post-6-month training performances were similar to the immediate post-training performances, and these performances even came close to the performances of expert radiologists, along with an improved agreement for US descriptors. The current training curriculum for breast US in Korea proves to be effective for residents because it first improves the sensitivity of US interpretation and second is based on the sustained training effect for a considerable post-training time.

The dedicated education material used during training contained breast US images with expert consensus descriptions, and pathologic diagnoses for the case images were displayed right after the US images for a feedback effect (Fig. 1), since a recent ACRIN 6666 investigator study showed that feedback improves performances in US interpretation [12]. This organization used for immediate feedback, and the usage of video images for both education material and test files may have contributed to the training effect, factors that should be considered in building education material for residents in the future. Higher diagnostic performances have been reported for expert radiologists using video clips which provide a more similar environment to everyday practice compared to using B-mode images in US image analysis [18], supporting the usefulness of video clips in educational material for breast US training.

As for environmental factors influencing the training effect, low levels of pre-training AUC, the number of training sessions on breast US per week, and the total number of breast US examinations performed during the training month were factors that significantly influenced improvements in AUC. Another study has also reported the education effect to be maximized in residents with low-level diagnostic performances before training [12], which fits the goal of training programs. While the number of training sessions per week or total number of examinations performed during the training month had an effect on improving AUC, the number of breast US examinations performed by the resident per session did not show a significant effect, supporting the assumption that constant exposure to breast US examinations is important to improving performances. This is a finding that should be considered when organizing training curriculums for trainees in breast imaging including residents or fellows. In addition, the number of US sessions per week and the total number of breast US examinations are also factors used for education maintenance for level 2 trainees in the EFSUMB guidelines, our results coming close to the number of examinations they recommend, 40 cases per month (at least 10 cases per week). The EFSUMB guidelines do not specify the number of sessions required per week, and our results show that the training effect for breast US may be enhanced by considering the number of sessions per week that the trainee is involved.

There are several limitations to this study. First, the images collected for the education or test sets were obtained from different US machines of different institutions, which may have affected US interpretation. During clinical practice, we use more than one type of machine both for real-time examinations and for image interpretation; therefore, we consider this to represent everyday practice. In addition, although we used video clips to approximate real-time examinations, our study still has limitations on reflecting the effect of education on improving real-time performances. Second, the training conditions for the 61 residents were not the same, varying under the circumstances of each institution. Factors other than the environmental data we obtained from the questionnaires may have affected our results. Also, due to the small population of trainees, building a competency-based education model with quality assurance standards considered is difficult based on our results, in which further profound investigation regarding training conditions is needed. Third, the same test set was used for pre-, post-, and post-6-month data acquisition, and in some cases, image analysis may have been made by memory [19]. However, pathologic diagnoses for the cases included in the test set or the consensus data of the experienced radiologists were not revealed to any of the residents. Therefore, using the same test set would have had little effect on our results. In addition, the test interval for long-term follow-up was set at 6 months, based on realistic approach since obtaining data may be more difficult from the residents if the test interval was set at more than 1 year after initial testing, which may have shown different results.

In conclusion, radiology residents showed a significant improvement in diagnostic performances including sensitivity, NPV, and AUC and in agreements for US descriptors compared to experienced radiologists immediately after training and 6 months after training on breast US. Involving residents in more sessions of breast US examinations may enhance the effect of training, which should be considered when organizing future curriculums for radiology residents.

Funding This study has received funding by the Korean Society of Breast Imaging & Korean Society for Breast Screening (KSBI & KSFBS-2013-No.001).

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Min Jung Kim.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry One of the authors has significant statistical expertise, Hye Sun Lee.

Informed consent Written informed consent was obtained from all subjects (patients) in this study.

Ethical approval Institutional Review Board approval was obtained.

Study subjects or cohorts overlap Some study subjects or cohorts have been previously reported in Ultrasound Med Biol 2016;42(9):2083–2088.

The performances of the experienced radiologists in reviewing static and real-time video clips obtained from the patients included in this study have been published. While our study focuses on comparing the performances and agreements between residents and experienced radiologists on breast US examinations, the prior study focuses on comparing the diagnostic performances between static and video clip images.

Methodology
- prospective
- cross-sectional study/observational
- multicenter study

References

1. Hooley RJ, Scoutt LM, Philpotts LE (2013) Breast ultrasonography: state of the art. Radiology 268:642–659
2. Berg WA, Blume JD, Cormack JB et al (2008) Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. JAMA 299:2151–2163
3. Berg WA, Zhang Z, Lehrer D et al (2012) 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 307:1394–1404
4. American College of Radiology (2013) Breast imaging reporting and data system, 5th edn. American College of Radiology, Reston, VA
5. Lee HJ, Kim EK, Kim MJ et al (2008) Observer variability of Breast Imaging Reporting and Data System (BI-RADS) for breast ultrasound. Eur J Radiol 65:293–298
6. Lee YJ, Choi SY, Kim KS, Yang PS (2016) Variability in observer performance between faculty members and residents using Breast Imaging Reporting and Data System (BI-RADS)-ultrasound, fifth edition (2013). Iran J Radiol 13:e28281
7. Park CS, Lee JH, Yim HW et al (2007) Observer agreement using the ACR Breast Imaging Reporting and Data System (BI-RADS)-ultrasound, First Edition (2003). Korean J Radiol 8:397–402
8. Schwab F, Redling K, Siebert M, Schotzau A, Schoenenberger CA, Zanetti-Dällenbach R (2016) Inter- and intra-observer agreement in ultrasound BI-RADS classification and real-time elastography Tsukuba score assessment of breast lesions. Ultrasound Med Biol 42:2622–2629
9. Berg WA, Campassi C, Langenberg P, Sexton MJ (2000) Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment. AJR Am J Roentgenol 174:1769–1777
10. Ortiz-Perez T, Trevino EJ, Sepulveda KA, Hilsenbeck SG, Wang T, Sedgwick EL (2013) Does formal instruction about the BI-RADS ultrasound lexicon result in improved appropriate use of the lexicon? AJR Am J Roentgenol 201:456–461
11. Scheel JR, Peacock S, Orem J et al (2016) Improving breast ultrasound interpretation in Uganda using a condensed Breast Imaging Reporting and Data System. Acad Radiol 23:1271–1277
12. Berg WA, Blume JD, Cormack JB, Mendelson EB (2012) Training the ACRIN 6666 investigators and effects of feedback on breast ultrasound interpretive performance and agreement in BI-RADS ultrasound feature analysis. AJR Am J Roentgenol 199:224–235
13. Monticciolo DL, Rebner M, Appleton CM et al (2013) The ACR/Society of Breast Imaging Resident and Fellowship Training Curriculum for Breast Imaging, updated. J Am Coll Radiol 10:207–210 e204
14. Education and Practical Standards Committee, European Federation of Societies for Ultrasound in Medicine and Biology (2006) Minimum training recommendations for the practice of medical ultrasound. Ultraschall Med 27:79–105
15. Hooley RJ, Greenberg KL, Stackhouse RM, Geisel JL, Butler RS, Philpotts LE (2012) Screening US in patients with mammographically dense breasts: initial experience with Connecticut Public Act 09-41. Radiology 265:59–69
16. Berg WA, Mendelson EB (2014) Technologistperformed handheld screening breast US imaging: how is it performed and what are the outcomes to date? Radiology 272(1):12–27
17. Berg WA, D'Orsi CJ, Jackson VP et al (2002) Does training in the Breast Imaging Reporting and Data System (BI-RADS) improve biopsy recommendations or feature analysis agreement with experienced breast imagers at mammography? Radiology 224:871–880
18. Youk JH, Jung I, Yoon JH et al (2016) Comparison of inter-observer variability and diagnostic performance of the fifth edition of BI-RADS for breast ultrasound of static versus video images. Ultrasound Med Biol 42:2083–2088
19. Cook DA, Beckman TJ (2010) Reflections on experimental research in medical education. Adv Health Sci Educ Theory Pract 15:455–464