J Med Ultrasonics (2006) 33:239-244
DOI 10.1007/s10396-006-0110-2

ORIGINAL ARTICLE

Eriko TOHNO · Kiyoshi SAWAI
Kazuhiro SHIMAMOTO · Ei UENO · Tokiko ENDOU
Hiroko TSUNODA-SHIMIZU · Hideaki SHIRAI
Etsuo TAKADA

Establishment of seminars to improve the diagnostic accuracy and effectiveness of breast ultrasound

Received: August 16, 2005 / Accepted: April 10, 2006

Abstract
Purpose. To improve the ability of technicians and physicians to find and diagnose breast lesions in breast ultrasound screening.
Methods. Seminars were organized for technicians and physicians engaged in breast ultrasound screening, and tests were carried out to evaluate the usefulness of the seminars. Each seminar lasted 2 days and comprised lectures and group activities. Pretests and post-tests conducted before and after each seminar, respectively, consisted of 100 questions: 50 about animated images, and 50 about static images. The tests required the participant to find lesions in animated images and estimate the probability of malignancy from static images.
Results. In the animated image tests, sensitivity was greater after the seminar, although specificity did not change significantly. In the static image tests, sensitivity increased, and a significant increase was also observed in the receiver operating curve analysis for degree of certainty in diagnosing cancer.
Conclusion. The seminars improved the participants' ability to find and diagnose breast lesions during ultrasound screening.

Keywords breast · education · screening · ultrasonography

E. Tohno (✉) · E. Ueno
Graduate School of Comprehensive Human Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8575, Japan
Tel. +81-29-853-3205; Fax +81-29-853-7985
e-mail: etohno@sakura.cc.tsukuba.ac.jp

K. Sawai
Department of Endocrine and Breast Surgery, Kyoto Prefectural University of Medicine, Kyoto, Japan

K. Shimamoto
School of Health Sciences, Nagoya University, Nagoya, Japan

T. Endou
National Hospital Organization Nagoya Medical Center, Nagoya, Japan

H. Tsunoda-Shimizu
St. Luke's International Hospital, Tokyo, Japan

H. Shirai
Sapporo Kotoni Breast Clinic, Sapporo, Japan

E. Takada
Dokkyo Medical University School of Medicine, Tochigi, Japan

Introduction

Breast ultrasound is widely used in Japan to diagnose breast cancer, and is also expected to be adopted in routine health examinations in the future. Before breast ultrasound can be widely used in screening, however, education and qualification systems to improve the accuracy of screening must be in place. We designed seminars to educate technicians and physicians engaged in breast ultrasound screening, and evaluated their diagnostic ability after attending the seminars. The effectiveness of the seminar was evaluated based on tests prepared for this purpose.

Materials and methods

Subjects

The subjects of this study were 86 technicians and 30 physicians who participated in three seminars conducted by the Educational Committee of the Japan Association of Breast and Thyroid Sonology from October 2003 through May 2004. Table 1 shows the number of participants and their occupations.

Seminar program

Table 2 shows the basic program of the 2-day seminar. The lectures covered basic aspects of breast diseases and ultrasound. The first three group activity themes were titled not according to pathologic changes but according to the appearance of lesions on ultrasound images, using

classifications such as accentuating type, intermediate/attenuating-type, and non-mass image forming diseases. This is a more practical approach because the examiner does not know the pathology of the lesion at the time of the examination. During these group activities the participants discussed ways to evaluate cases represented in static and animated images. During the hands-on activity using handmade konjac jelly phantoms, participants were taught how to hold the probe, scan, and complete a results form. In actual screening, ability to detect abnormalities in real-time images of the entire breast scan is essential. In the group activity, in order to teach the skills required for detecting lesions, we asked each participant to find lesions in animated images from cases edited and recorded on compact discs. This method was the same as that used in the tests carried out before and after the seminars. Further, another group activity involved presenting ultrasound images of typical lesions of various types, which were explained using animated or static images also contained on the compact disc.

Table 1. Number of participants and their occupations

Seminar	No. participants	No. technicians	No. physicians
1st	37	37	0
2nd	38	8	30
3rd	41	41	0
Total	116	86	30

Preparation of the test questions

Animated images

Records of 50 cases were obtained from collaborating institutions. The images of some cases showed lesions and others did not. The images were continuously recorded on videotape. The speed of probe handling varied from recorder to recorder, and images recorded at inadequate speed were excluded from the analysis with the consent of the collaborators. We copied the videotapes onto digital videotape and sampled about 30 s from each case. We then deleted all personal and hospital-related information, transferred the remaining data to a server in AVI format, and converted the files to MPEG1 movie files. These files were played from compact discs for the participants on individual computer terminals, or over a local area network. The 50 questions on the animated images included 25 questions related to 25 abnormal cases whose ultrasound images should have correctly been classified as category 3 or higher. The categories are described in Table 3 and are based on criteria set forth by the Subcommittee for Breast Ultrasound Screening of the Japan Association of Breast and Thyroid Sonology.[1]

Static images

Fifty static images from 50 cases were obtained from collaborating institutions. Personal and hospital-related

Table 2. Basic program of the breast ultrasound seminar

Pretest (100 min)
Lecture
- Basic knowledge regarding breast diseases (30 min)
- Pathology of breast diseases (40 min)
- Tissue ultrasonic characteristics of breast diseases (40 min)
- Breast ultrasound examination method (30 min)
  - Breast ultrasound terminology (mass-image forming diseases) (40 min)
  - Breast ultrasound terminology (non-mass image forming diseases) (30 min)
  - Screening criteria and completing the results form (20 min)

Group activities (50 min per activity)
- Themes
  - Mass-image forming diseases: accentuating-type diseases
  - Mass-image forming diseases: intermediate-type and attenuating-type diseases)
  - Non-mass image forming diseases
  - Hands-on activity
  - Skills for finding lesions
  - Ultrasound images of various lesions

Post-test (100 min)
Answers to the test questions given (30 min)

Table 3. Criteria for determining categories during ultrasonography

Category	Remarks
1. Normal	Including such normal variations as fat islands.
2. Typically benign	Not requiring recall after the screening. Includes mastopathy, cyst, typical fibroadenoma not exceeding 20 mm in longest diameter, and calcified fibroadenoma.
3. Probably benign	Requiring further examination after screening. Including papilloma within the dilated duct or cyst, and other lesions in which the nature of the lesion is not typically benign.
4. Suspicious abnormality	Biopsy should be considered.
5. Typically malignant	

information was deleted, and the images were converted to JPEG format. Questions concerning the 50 static images included 24 on ultrasound images showing breast cancers and four on images showing probably benign lesions; all 28 of these lesions should have been classified as category 3 or higher. The categories ranged from 1 to 5, and were verified by three physician collaborators.

Pretests and post-tests

Pretests and post-tests were used to estimate the effectiveness of the program. The total time allotted for evaluating both the animated and the static images was 100 min. We explained the criteria for deciding categories, shown in Table 3, in about 10 min before the pretest. Two sample questions for both the animated and static image tests were also presented at this time to show participants how to use the computer and answer the questions. The pretest and the post-test showed the same cases, but presented in a different order. We spent about 30 min explaining the correct answers after the post-test.

Animated image test

Participants were required to answer 50 questions on the animated images by deciding if a lesion of category 3 or higher was present or absent. The images were displayed on Windows Media Player, a standard Microsoft Windows application. Participants were thus able to replay and pause the images at will, but were unable to change the playing speed. The participants viewed the images on the 15-inch monitors of their desktop computers.

Static image test

Participants were asked to play 50 static images and to provide information on (1) the percentage probability of cancer, (2) the category of the lesion, and (3) the name of the most probable disease. For item (1), participants were further instructed to specify their confidence level for malignancy using a continuous distribution scale.

Determining question suitability

The same test was given to 10 of the seminar lecturers at the time of the pretest for the third seminar to confirm the suitability of the images used in the questions. Among the 25 animated images in which lesions of greater than category 3 were not present, eight were omitted because five or more lecturers answered that they were present. Thus, in total, the animated image test included 25 questions concerning images of lesions of greater than category 3 and 17 questions concerning images without such lesions (as described above). Similarly, in the static image test, 12 questions on images considered to be category 1 or 2 were omitted because five or more lecturers rated the associated images as being category 3 or higher. Two benign cases considered to be category 3 were also omitted because five or more lecturers rated them as category 4 or 5. Benign lesions that appeared questionable or typically malignant in ultrasound images were considered unsuitable for the test. More than half the lecturers were in agreement in all other cases. Thus, the static image test included 26 lesions that should have been categorized as greater than category 3, and 10 lesions that should have been categorized as category 1 or 2.

Statistical analysis

Receiver operating characteristics (ROC) analysis was used to compare the diagnostic performance of the participants before and after the seminar,[2,3] and the area under the binormal ROC curve (Az) was used as an index of performance.[4]

Results

Because the analysis was limited to data from participants who answered all questions in both tests, the number of subjects was 36. Those unable to answer the questions within the allotted time were excluded to avoid bias resulting from improvement due to learning how to use their computers during the pretest; however, those who could not answer or who overlooked any questions, and those unable to finish answering the questions because of mistakes in using the computers were also excluded.

Results obtained from the animated and static image pretests and post-tests appear in Table 4. McNemar's test was used to analyze significance. In the animated image test, sensitivity in finding lesions of greater than category 3 increased significantly between the pretest and the post-test, from 79.7% to 88.0% ($P < 0.01$), but specificity did not decrease significantly. Accuracy also increased significantly, from 76.6% to 82.5% ($P < 0.01$). In the static image test, ROC analysis of the probability of cancer found that the Az was significantly greater in the post-test than that in the pretest, indicating a significant improvement in diagnostic performance after the seminar (Fig. 1). Category sensitivity was considered to be true positive when participants selected the category determined by the collaborators. Category sensitivity also increased significantly between the pretest and the post-test, from 67.0% to 71.7% ($P < 0.05$).

Discussion

Breast ultrasound is widely used for the clinical diagnosis of breast diseases. Many reports claim that breast ultrasound examination can detect breast cancer that cannot be detected by palpation.[5–7] Although the majority opinion in

Table 4. Sensitivity, specificity, accuracy, category sensitivity, and Az of each participant, and average score for each test

Participant no.	Animated image test						Static image test			
	Sensitivity		Specificity		Accuracy		Category sensitivity		Az	
	Pretest	Post-test	Pretest	Post-test	Pretest	Post-test	Pretest	Post-test	Pretest	Post-test
1	88.0	88.0	52.9	100.0	73.8	92.9	70.3	62.2	0.890	0.938
2	80.0	84.0	52.9	41.2	69.1	66.7	61.1	51.4	0.928	0.962
3	50.0	52.0	76.5	52.9	61.0	52.4	43.2	62.2	0.759	0.745
4	92.0	96.0	41.2	70.6	71.4	85.7	64.9	75.7	0.667	0.692
5	56.0	72.0	23.5	58.8	42.9	66.7	74.8	64.9	0.933	0.908
6	91.3	80.0	29.4	52.9	65.0	69.1	59.5	64.9	0.750	0.899
7	92.0	92.0	70.6	100.0	83.3	95.2	67.6	70.3	0.858	0.792
8	76.0	92.0	100.0	70.6	85.7	83.3	63.9	67.6	0.815	0.911
9	72.0	88.0	82.4	88.2	76.2	88.1	78.4	73.0	0.961	0.959
10	76.0	84.0	100.0	88.2	85.7	85.7	67.6	73.0	0.935	0.923
11	60.0	84.0	52.9	94.1	57.1	88.1	75.7	77.8	0.908	0.957
12	84.0	80.0	94.1	94.1	88.1	85.7	64.9	62.2	0.907	1.000
13	84.0	92.0	100.0	64.7	90.5	81.0	64.9	75.7	0.914	0.899
14	66.7	80.0	88.2	76.5	76.3	78.6	62.2	67.6	0.809	0.948
15	68.0	80.0	29.4	52.9	52.4	69.1	56.8	56.8	0.830	0.877
16	72.0	80.0	47.1	64.7	61.9	73.8	56.8	62.2	0.811	0.894
17	76.0	100.0	88.2	70.6	81.0	87.8	62.2	70.3	0.836	0.898
18	96.0	96.0	76.5	76.5	88.1	88.1	72.2	75.7	0.926	0.939
19	80.0	88.0	76.5	88.2	78.6	88.1	75.0	78.4	0.933	0.982
20	84.0	96.0	64.7	64.7	76.2	83.3	78.4	78.4	0.832	1.000
21	72.0	88.0	76.5	70.6	73.8	81.0	68.6	62.9	0.868	0.942
22	84.0	92.0	100.0	82.4	90.5	88.1	61.1	75.7	0.928	0.892
23	68.0	79.2	88.2	100.0	76.2	87.8	75.7	67.6	0.882	0.969
24	56.0	72.0	88.4	82.4	66.7	76.2	48.7	64.9	0.969	0.933
25	80.0	96.0	88.2	76.5	83.3	88.1	63.9	81.1	0.728	0.924
26	82.6	96.0	81.3	70.6	82.1	85.7	75.0	82.4	0.944	0.989
27	96.0	80.0	47.1	47.1	76.2	66.7	59.5	78.4	0.826	0.910
28	88.0	88.0	64.7	76.5	78.6	83.3	67.6	81.1	0.918	0.942
29	92.0	100.0	76.5	82.4	85.7	92.9	73.0	75.7	0.948	0.954
30	76.0	92.0	58.8	64.7	69.1	81.0	73.0	81.1	0.890	0.962
31	96.0	96.0	100.0	76.5	97.6	88.1	75.7	83.8	0.805	0.908
32	69.6	96.0	75.0	52.9	71.8	78.6	48.7	78.4	0.898	0.926
33	100.0	100.0	87.5	88.2	94.4	95.2	81.1	78.4	0.883	0.922
34	88.0	92.0	76.5	88.2	83.3	90.5	62.2	73.0	0.873	0.962
35	92.0	96.0	52.9	88.2	76.2	92.9	78.4	73.0	0.931	1.000
36	84.0	100.0	94.1	64.7	88.1	85.7	81.1	75.7	0.844	0.921
Average	79.7	88.0*	72.3	74.5	76.6	82.5*	67.0	71.7**	0.870	0.922

Significant difference between pretest and post-test: * $P < 0.01$, ** $P < 0.05$

the United States and Europe[8] is that breast ultrasound is not appropriate for routine screening, the American College of Radiology Imaging Network 6666 has initiated trial ultrasound screening for women at risk. Researchers in Japan have also indicated that breast ultrasound examination is effective in routine screening;[9,10] ultrasound examination may be particularly suitable for Japanese women, whose breasts are typically small and rich in mammary glands.

In Japan, breast cancer screening by visual observation and palpation has been carried out among the general public as a component of general health examinations since 1965.[11] However, a study conducted by the Hisamichi Group, funded by the Japan Ministry of Health and Welfare, found that visual examination and palpation alone is not an effective method of breast cancer screening,[12] and therefore mammography was introduced, which is widely used throughout the United States and Europe. The Japanese Ministry of Health and Welfare thus brought mammography into wide use in Japan. The main targets of this program have been women in their fifties, in whom the effectiveness of mammography has been confirmed in the United States and Europe. Because peak breast cancer morbidity in Japan occurs in women in their late forties,[13] however, proposal no. 0427001, which was issued on April 27, 2004, by the Section of Health and Welfare for the Aged in the Ministry of Health, Labour, and Welfare now states that mammography should be used to examine women in their forties. This proposal thus further provides a rationale for continuous investigation and study of the effectiveness of ultrasound screening.

One reason breast cancer screening by ultrasound has not been adopted for population-based check-ups is the lack of a method to assure the quality of breast cancer ultrasound screening. For mammography screening, there are standardized and widely recognized qualification systems for radiotechnologists and diagnosing physicians interpreting mammograms are generally accepted; 2-day seminars

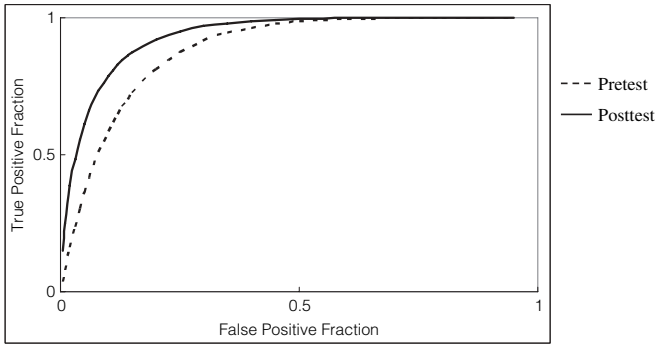

Fig. 1. ROC curve for the static image test. Az increased significantly in the post-test (*solid line*) relative to the pretest (*dashed line*)

for such workers are offered at various institutions in Japan. However, education and qualification systems for technicians and physicians engaged in breast ultrasound screening are not yet in place. Because ultrasound screening entails real-time scanning, there are marked differences in the ratios of cancer detected by workers with differing levels of skill. Some studies have indicated that the cancer discovery ratio is higher for physicians than for technicians.[14] Nevertheless, cost considerations encourage the education of medical technologists and radiographers as ultrasound operators, who, when well trained and adequately experienced, can obtain a level of ability equal to that of qualified physicians.

The purpose of the present study was to establish a seminar system that would improve the quality of breast ultrasound screening and establish a system of diagnostic criteria for detecting breast cancer. The lecture titled "Screening criteria and rules for completing a results form" represents an attempt to standardize breast ultrasound screening. In the group activities described above, animated images displayed on the monitors of computers operated by individual seminar participants were used to teach the skills necessary for finding lesions and the characteristics of ultrasound images of various lesions.

We previously reported on the effectiveness of an ultrasound seminar conducted in Ibaraki Prefecture.[15] The tests given before and after the seminar included fewer questions, which rendered them insufficient for judging and evaluating worker ability. We thus increased the number of questions in the present study and used methods more appropriate for testing skills in ultrasound screening.

In actual screening, any lesions that are present must be found during real-time scanning; any lesion identified can be examined more thoroughly later. Therefore, using the animated images, we focused on evaluating participants' ability to judge the presence or absence of lesions, requiring recall. Static images were used to evaluate the participants' ability to analyze the details of the lesions. Because participants' skills in both areas had improved when the seminar ended, we consider that this program met this aims.

We used only one static image of each lesion to determine participants' ability to evaluate lesions. For the static images, operator subjectivity is unavoidable when the lesions are represented as images. Further, the nature of a lesion is not easily discerned from a single image. These factors led us to use only those images on which more than half the lecturers concurred. Better results would be obtained when the evaluation process also considered animated images that focus on or around the lesion.

We have yet to determine if this test could be used to qualify breast ultrasound operators. One problem that we encountered was the inability of some participants to answer all the questions within the allotted time. We feel, however, that the number of questions is adequate when the test is used to judge skill level in breast ultrasound screening. In terms of the efficiency of screening time, the time provided seems adequate. Another problem resulted from confusion arising from requiring participants to use personal computers, a factor that clearly troubled some of them. Also, results obtained by the participants in the post-test might have been better than would otherwise have been expected because they had learned to use the computer during the group activities. Unfortunately, some data were accidentally lost during operation of the computers. In these cases, improvement in test results was not associated with enhanced ability to carry out breast ultrasound screening. So we included data from only those participants who had answered all the questions in both tests. Problems associated with the data input process indicate a need to improve the software. Answers, for example, should be saved automatically, and the input method should be easy to use. In any case, however, participants must be sufficiently trained before they take the examination.

Conclusion

A seminar program was planned in order to improve the accuracy of breast ultrasound screening, and the seminars were carried out. Comparison of participants' pretest and post-test results indicate that the diagnostic ability of participants is improved by participation the seminars and suggests that such seminars can effectively improve the ability of participants to detect breast cancer.

Acknowledgments The present study was conducted by the Study Group Regarding Improvement of the Accuracy and Efficiency of Breast Cancer Screening, which is supported by the cancer research fund for 2004 provided by the Japanese Ministry of Health, Labour, and Welfare. We acknowledge the cooperation of the many physicians who served as lecturers for the seminars in this study.

References

1. Japan Association of Breast and Thyroid Sonology (ed). Guidelines for breast ultrasound: management and diagnosis. Tokyo, Nankodo; 2004 (in Japanese).
2. Dorfoman DD, Alf E Jr. Maximum likelihood estimation of parameters of signal detection theory: a direct solution. Psychometrika 1968;33:117–24.
3. Metz CE. ROC methodology in radiographic imaging. Invest Radiol 1986;21:720–33.

4. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983;148:839–43.
5. Buchberger W, Niehoff A, Obrist P, et al. Clinically and mammographically occult breast lesions: detection and classification with high-resolution sonography. Semin Ultrasound CT MR 2000;21: 325–36.
6. Leconte I, Feger C, Galant C, et al. Mammography and subsequent whole-breast sonography of nonpalpable breast cancers: the importance of radiologic breast density. Am J Roentgenol 2003;180:1675–9.
7. Berg WA. Supplemental screening sonography in dense breast. Radiol Clin North Am 2004;42:845–51.
8. Elmore JG, Armstrong K, Lehman CD, et al. Screening for breast cancer. JAMA 2005;9:1245–56.
9. Yamasaki M, Masu S, Koga S, et al. Detection rate of breast cancer by mammography and ultrasonography screening in women aged between 40 and 49. J Jpn Assoc Breast Cancer Screen 2002;11: 265–9 (in Japanese).
10. Namba K, Watanabe R, Furusawa H, et al. Significance of breast cancer screening by mammography and ultrasound for women aged 49 and under. J Jpn Assoc Breast Cancer Screen 2002;11: 172–8 (in Japanese).
11. Izuo M. Memorial Address of the Tenth Annual Meeting: the history of breast cancer screening in Japan. J Jpn Assoc Breast Cancer Screen 2001;10:3–16 (in Japanese).
12. Hisamichi S (ed). Evaluation of effectiveness of cancer screening. Sendai, Japan Public Health Association; 2001.
13. Reports of the Research Group for Population-Based Cancer Registration in Japan. Tokyo, National Cancer Center; 2003.
14. Morita S, Mitta T, Kobayashi T, et al. Breast cancer screening using ultrasonography in Hirakata city. J Jpn Assoc Cancer Detect Diagn 2004;11:123–6 (in Japanese).
15. Tohno E, Ueno U, Tsunoda H, et al. Evaluation of the effectiveness of a teaching course for doctors and sonographers engaged in ultrasound breast screening. J Jpn Assoc Breast Cancer Screen 2003;12:108–13.