JJCO Japanese Journal of Clinical Oncology

Jpn J Clin Oncol 2012;42(3)168–174
doi:10.1093/jjco/hyr197
Advance Access Publication 4 January 2012

Breast Ultrasonographic and Histopathological Characteristics Without Any Mammographic Abnormalities

Kentaro Tamaki[1,2,3,*], Takanori Ishida[2], Minoru Miyashita[2], Masakazu Amari[2], Noriaki Ohuchi[2], Yoshihiko Kamada[1], Kano Uehara[1], Nobumitsu Tamaki[1] and Hironobu Sasano[3]

[1]Department of Breast Surgery, Nahanishi Clinic, Okinawa, [2]Department of Surgical Oncology, Tohoku University Graduate School of Medicine and [3]Department of Pathology, Tohoku University Hospital, Miyagi, Japan

*For reprints and all correspondence: Kentaro Tamaki, 1-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi 980-8574, Japan. E-mail: nahanisikenta@yahoo.co.jp

Received October 13, 2011; accepted December 13, 2011

Objective: We evaluated ultrasonographic findings and the corresponding histopathological characteristics of breast cancer patients with Breast Imaging Reporting and Data System (BI-RADS) category 1 mammogram.
Methods: We retrospectively reviewed the ultrasonographic findings and the corresponding histopathological features of 45 breast cancer patients with BI-RADS category 1 mammogram and 537 controls with mammographic abnormalities. We evaluated the ultrasonographic findings including mass shape, periphery, internal and posterior echo pattern, interruption of mammary borders and the distribution of low-echoic lesions, and the corresponding histopathological characteristics including histological classification, hormone receptor and human epidermal growth factor receptor 2 status of invasive ductal carcinoma and ductal carcinoma *in situ*, histological grade, mitotic counts and lymphovascular invasion in individual cases of BI-RADS category 1 mammograms and compared with those of the control group.
Results: The ultrasonographic characteristics of the BI-RADS category 1 group were characterized by a higher ratio of round shape ($P < 0.001$), non-spiculated periphery ($P = 0.021$), non-interruption of mammary borders ($P < 0.001$) and non-attenuation ($P = 0.011$) compared with the control group. A total of 52.6% of low-echoic lesions were associated with spotted distribution in the BI-RADS 1 group, whereas 25.8% of low-echoic lesions were associated with spotted distribution in the control group ($P = 0.012$). As for histopathological characteristics, there was a statistically higher ratio of triple-negative subtype ($P = 0.021$), and this particular tendency was detected in histological grade 3 in the BI-RADS category 1 group ($P = 0.094$).
Conclusion: We evaluated ultrasonographic findings and the corresponding histopathological characteristics for BI-RADS category 1 mammograms and noted significant differences among these findings in this study. Evaluation of these ultrasonographic and histopathological characteristics may provide a more accurate ultrasonographic screening system for breast cancer in Japanese women.

Key words: breast US – BI-RADS category 1 mammogram – histopathological characteristics

INTRODUCTION

The incidence of breast cancer has increased worldwide, which is partly considered to be due to mass screening programs resulting in the discovery of clinically occult or early breast lesions (1). Early clinical detection of breast cancer through screening has therefore led to the detection of the tumor at a relatively earlier clinical stage. The effectiveness of screening mammography on reduction in mortality by breast cancer has been well established in both Western countries and Japan (2). Mammography has thus become the gold standard for

detecting breast disorders. Therefore, it has become very important to increase the rate of mammographic screening among the general public toward reducing the breast cancer mortality. However, it is also true that 7.2% of the malignant cases were associated with no mammographic abnormalities (3). In addition, the malignant ratio of 20, 30 and 40 years without mammographic abnormalities was statistically higher than the ratio of the other age groups (3). Ultrasonography (US) has been in general proposed to prove much more effective in the detection of breast cancer if the patient is young, has dense breast or their detected masses are small (4–6). Therefore, it has become very important to improve the quality of US diagnoses.

The effectiveness of ultrasound screening for women aged 40 years has been evaluated in detecting and reducing mortality of the breast cancer in Japan in order to complement this particular pitfall of mammography (7). This study named J-START (The Japan Strategic Anti-cancer Randomized Trial) evaluates the effectiveness of screening mammography with US breast cancer screening compared with mammography alone in 40 years, with a design to study 50 000 women with mammography and US and 50 000 controls with mammography only (7). The participants are scheduled to take a second-round screening with the same modality 2 years onwards (7). The primary endpoints are sensitivity and specificity, and the secondary endpoint as the rate of advanced breast cancer (7). Whether or not breast US screening is adopted in the future large-scale screening therefore largely depends on the results of this research. Considerable efforts will be required to successfully carry out this massive undertaking done in Japanese population.

Strict or rigorous conformity to high quality of interpretation of US finding among those involved in this screening is therefore mandatory for the very success of an US diagnosis in such a large scale. We previously examined the correlation between US findings and the corresponding histopathological features in breast disorders in our previous study (6). There have been relatively few reported studies on assessing US performance and its resolution without any mammographic abnormalities (8). Therefore, in this study, we evaluated US findings and the corresponding histopathological characteristics for breast cancer patients with Breast Imaging Reporting and Data System (BI-RADS) (9) category 1 mammogram.

PATIENTS AND METHODS

Patients

We retrospectively reviewed the US findings and their corresponding histopathological features of 45 breast cancer patients with BI-RADS category 1 mammogram and 537 controls with mammographic abnormalities. The patients underwent needle biopsies or surgical resection at the Tohoku University Hospital from January 2006 to December 2010. We received informed consents from all the patients and the protocol for this study was approved by the Ethics Committee at Tohoku University Graduate School of Medicine.

Imaging Devices and Breast Tissue Specimens

The US examinations were assessed by one of the experienced eight breast specialists in Tohoku University Hospital. The consensus meeting of US was held for 1 whole week in order to standardize the US examination among these eight doctors. In addition, two of them independently evaluated the US findings in a retrospective manner, without the knowledge of clinical and histopathological information of individual patients. All US evaluations were carried out using Aloka SSD 3500 and ProSound α7 (Aloka Co., Tokyo, Japan) with a 10 MHz transducer.

We stained the corresponding tissue slides of the cases using hematoxylin–eosin (H&E) and immunohistochemistry for estrogen receptor (ER) and human epidermal growth factor receptor 2 (HER2). Surgical specimens had been fixed in 10% formaldehyde solution, cut into serial 5 mm-thick slices, embedded in paraffin, cut into 4 μm-thick sections and placed on the glue-coated glass slides. We employed the avidin–streptavidin immunoperoxidase method using the clone 6F11 antibody (Ventana, Tucson, AZ, USA) in automated immunostainer (Benchmark System; Ventana). A standardized immunohistochemistry kit (HercepTest for Immunoenzymatic Staining; Dako, Copenhagen, Denmark) was used for HER2 staining. Histopathological slides were reviewed by two pathologists independently without the knowledge of clinical information. Olympus (Tokyo, Japan) BX50 and 20× objectives were used for the analyses.

Imaging and Histopathological Analyses

Two or more hardcopy transverse and sagittal plane images of breast lesions were analyzed in this study. We recorded tumor shape, periphery, internal and lateral echo pattern, interruption of mammary borders and the distribution of low-echoic lesions, according to the BI-RADS sonographic classification (9) and the Japan Association of Breast and Thyroid Sonology (JABTS) breast sonographic classification (10). Tumor shape was tentatively classified into round, oval, lobular and irregular (9,10). Periphery was tentatively classified into circumscribed, obscured, indistinct and spiculated (9,10). Internal echo was classified into low and heterogeneity or high (9,10). Lateral echo was also classified into accentuation, no change and attenuation (9,10). Interruption of mammary borders was classified into interruption, indeterminate and no (9,10). Distribution of low-echoic lesions was classified into spotted and segmental (9,10) (Fig. 1).

Two of the experienced pathologists independently evaluated surgical pathology specimens, respectively. Histopathological evaluations were based on World Health Organization (WHO) histological classification of tumor of the breast (11) and Rosen's breast pathology (12). ER was determined by nuclear staining graded from 0 to 8 using the Allred score, and ER positivity was Grade 3 or more (13). With regard to HER2 evaluation, membranous staining was graded as the following: score 0–1+, 2+ and 3+ (14).

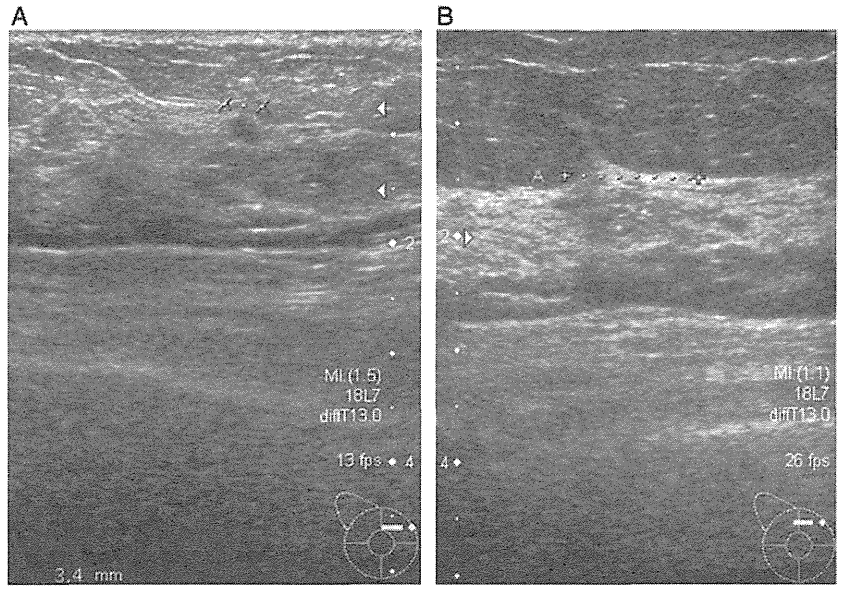


Figure 1. Representative illustrations of the distribution of low-echoic lesions. (A) Spotted and (B) segmental.

Scoring of 2+ was added fluorescence *in situ* hybridization (FISH) that was used to calculate the gene copy ratio of HER2-to-CEP17 (the PathVysion HER2 DNA Probe Kit; Abbott, Chicago, IL, USA). Positive is defined as either HER2:CEP17 signal ratio (FISH score) >2.2 (14). Histological grades and mitotic counts were assessed according to the criteria of Elston and Ellis (15). Van Nuys classifications were also assessed for ductal carcinoma *in situ* and invasive ductal carcinoma (IDC) with predominant intraductal components cases (16,17). We also identified the presence or absence of lymphovascular invasion according to the Rosen's Breast Pathology (12).

At first, we examined the differences of the patients' characteristics between these two groups including the distribution of age, menopausal status, past history of the benign proliferative disease, background of detection, clinical stage, breast density of mammography according to the BI-RADS lexicon (9) and surgical strategy as the breast-conserving ratio. We evaluated the US findings including mass shape, periphery, internal and posterior echo pattern, interruption of mammary borders and the distribution of low-echoic lesions and compared them with histopathological characteristics including histological classification, hormone receptor and HER2 status of IDC, tumor size confirmed by histopathology, histological grade, mitotic counts and lymphovascular invasion of BI-RADS category 1 mammograms. We then compared these findings with those of control group patients.

Statistical Analyses

Statistical analyses were performed using StatMate III for Windows ver. 3.18 (ATMS, Tokyo, Japan). The results were considered significant at $P < 0.05$.

RESULTS

The Details of Both BI-RADS 1 and Control Groups

Table 1 summarizes the difference in the patients' characteristics including the distribution of age, menopausal status, past history of the benign proliferative disease, background of detection, clinical stage, breast density of mammography and surgical strategy. The median ages of the study group and the control group were 48 years (range, 32–84) and 56 years (range, 26–88), respectively ($P = 0.047$). There was a statistically significant higher ratio of Stages 0 and I, heterogeneously and extremely dense, and conserving surgery in the BI-RADS 1 group ($P < 0.001$, <0.001 and 0.002, respectively). However, there was a statistically significant lower ratio of menopause and self-palpation in the BI-RADS 1 group ($P < 0.001$, respectively; Table 1).

Table 1. The details of patients

	BI-RADS 1	Control	*P* value	Odds ratio
Age	48 (32–84)	56 (26–88)	0.047	—
Menopausal ratio	37.8%	63.4%	<0.001	0.31
Benign proliferative disease	2.2%	9.5%	NS	2.34
Cause of detection (self-palpation ratio)	24.4%	59.4%	<0.001	0.22
Stage (Stages 0 and I)	93.3%	66.4%	<0.001	7.08
Heterogeneously and extremely dense ratio	91.1%	39.1%	<0.001	15.97
Surgical strategy (conserving ratio)	95.6%	74.6%	0.002	7.82

The Ratios of Mass Cases and the Tumor Size

Twenty-six out of the 45 were US mass cases in the BI-RADS 1 group and 370 out of the 490 were US mass cases in the control group. There was a statistically significant difference between the BI-RADS 1 and control groups ($P = 0.003$). The US tumor size of BI-RADS 1 and control groups was 12.1 mm (range, 3.2–24.9 mm) and 18.5 mm (range, 6.5–150 mm) with statistically significant differences ($P < 0.001$).

Evaluation of the US Characteristics

Figure 2 summarizes the results of the numbers and ratios of mass shape (Fig. 2A), periphery (Fig. 2B), internal echo pattern (Fig. 2C), lateral echo pattern (Fig. 2D) and interruption of mammary borders (Fig. 2E) of the BI-RADS 1 and control groups. There were statistically higher ratios of round mass shape ($P < 0.001$), no change of lateral echo pattern ($P = 0.028$) and no or indeterminate interruption of mammary borders ($P < 0.001$) in the BI-RADS 1 group. There were statistically lower ratios of spiculated periphery ($P = 0.021$), attenuation of lateral echo pattern ($P = 0.011$) and interruption of mammary borders ($P < 0.001$) in the BI-RADS 1 group. Figure 3 summarizes the results of the numbers and ratios of distribution of low-echoic lesions. There were statistically higher ratios of spotted distribution and lower cases of segmental distribution in the BI-RADS 1 group than in the control group ($P = 0.012$).

Evaluation of the Corresponding Histopathological Characteristics

Figure 4 summarizes the results of the numbers and ratios of results classified by histological subtypes (Fig. 4A), hormone receptor and HER2 expression of IDC (Fig. 4B), tumor size of the invasive lesion as confirmed by the histopathological examination (Fig. 4C), histological grade (Fig. 4D), mitotic counts (Fig. 4E) and lymphovascular invasion (Fig. 4F). There was statistically higher ratios of triple-negative subtype, smaller tumor size and lower case of lymphovascular invasion in the BI-RADS 1 group ($P = 0.021$, $P < 0.001$ and $P = 0.012$, respectively) compared with the control group. In addition, a higher ratio of histological grade 3 was detected in the BI-RADS 1 group but this difference did not reach the statistical significance ($P = 0.094$).

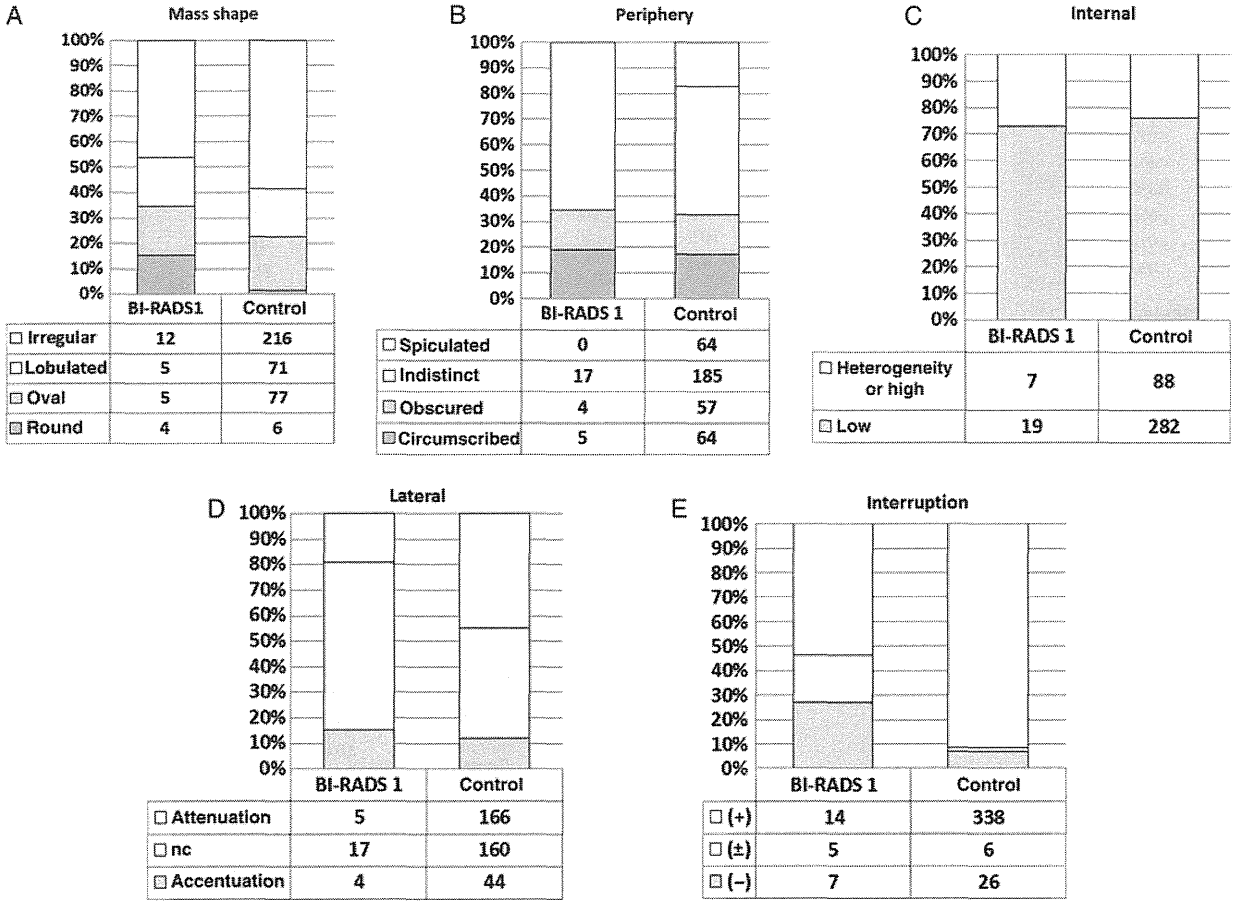


Figure 2. The US characteristics of BI-RADS category 1 and control groups. (A) Mass shape, (B) periphery, (C) internal echo pattern, (D) lateral echo pattern and (E) interruption of mammary borders.

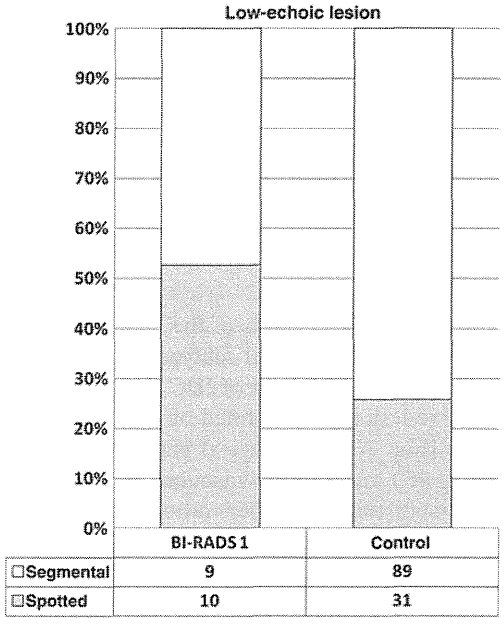


Figure 3. The distribution of low-echoic lesions of BI-RADS category 1 and control groups.

DISCUSSION

Mammography has been considered a gold standard for breast cancer screening system. However, US screening combined with mammography may have the potential to become one of the useful screening systems to decrease breast cancer mortality according to the results of the J-START trial (7). Therefore, strict or rigorous conformation to high quality of interpreting the US findings is required or mandatory for the future success of an US diagnosis especially at the level of mass screening. Our present study is the first study to focus upon incremental detection of breast cancer by US in asymptomatic women with mammography-negative breasts, and focused on the US findings and the corresponding histopathological characteristics of the cases with BI-RADS category 1 mammograms.

US detected cancers are in general smaller than those identified with mammography. Results of our present study demonstrated that the BI-RADS category 1 group was associated with a statistically higher ratio of low-echoic lesions than the control group. In addition, 52.6% of low-echoic lesions demonstrated spotted distribution in the BI-RADS 1 group, whereas 25.8% of low-echoic lesions spotted

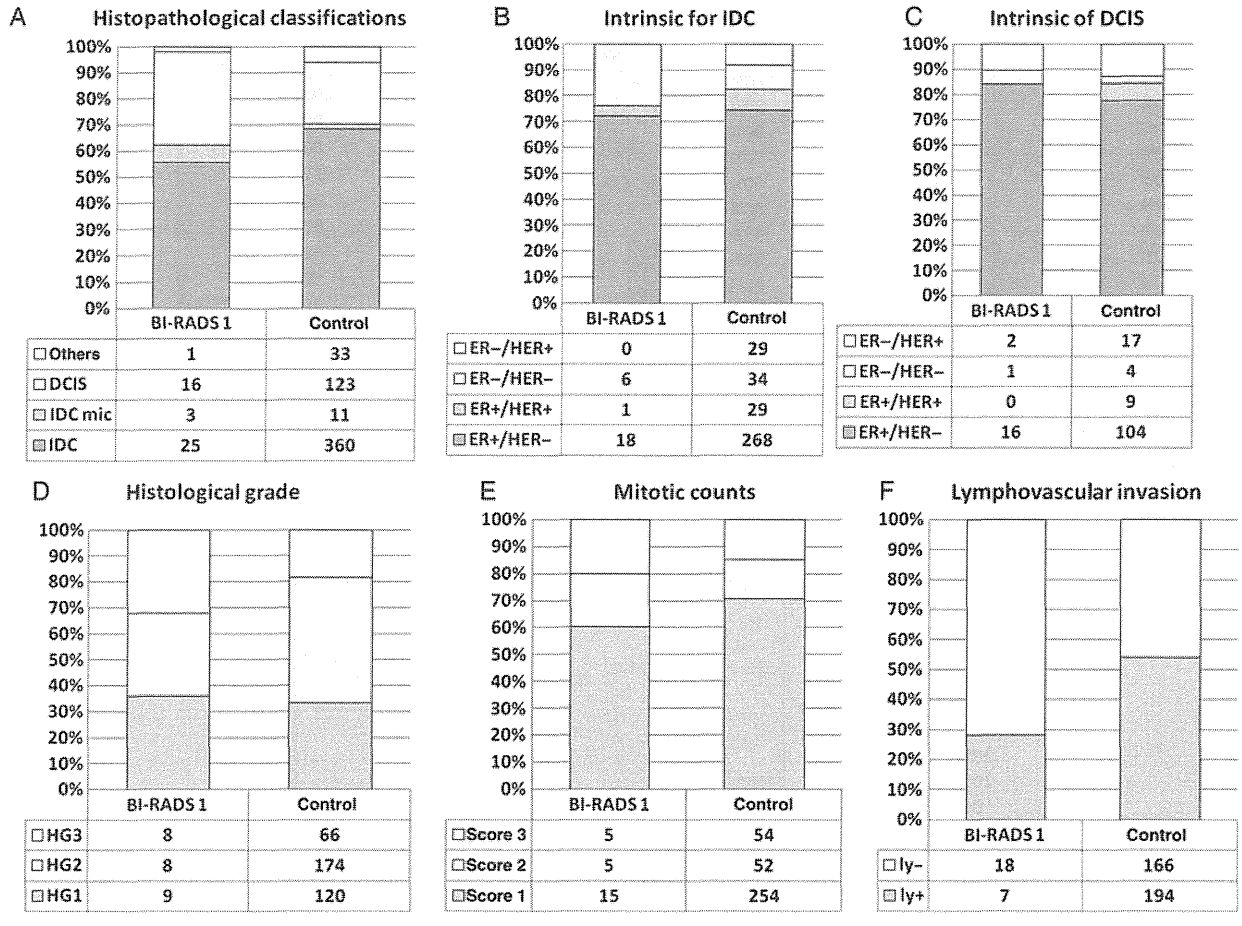


Figure 4. The histopathological characteristics of BI-RADS category 1 and control groups. (A) Histological classification, (B) hormone receptor and human epidermal growth factor receptor 2 expression of invasive ductal carcinoma, (C) tumor size of the invasive lesion, (D) histological grade, (E) mitotic counts and (F) lymphovascular invasion.

distribution in the control group. A low-echoic lesion with spotted distribution is therefore considered one of the predicting factors of malignancy in the BI-RADS category 1 group. In addition, the tumor size of the BI-RADS 1 group was smaller, and the detected masses were characterized by a higher ratio of round shape, non-spiculated periphery, non-interruption of mammary borders and non-attenuation in the BI-RADS category 1 group. These results could be mainly affected by mammographic breast density. In addition, results of our present study also demonstrated that there was a statistically higher ratio of heterogeneously and extremely dense breast in the BI-RADS 1 group and the tumors with well-collagenized stromal reaction were also detected as architectural distortion or spiculation in dense breast mammogram. Therefore, mammographic breast density was reasonably postulated to influence characteristics of breast cancers with BI-RADS category 1. Results of previous studies demonstrated that the most breast cancer cases of BI-RADS category 1 were relatively hypoechoic within a background of hyperechoic fibroglandular tissue, which may make the lesions more conspicuous and detectable (18). However, it is also true that previous studies have not evaluated the US findings of BI-RADS category 1 cases and this is the first study demonstrating the US findings such as mass shape and periphery of BI-RADS category 1 cases. In addition, this is the first reported study to demonstrate histopathological characteristics of BI-RADS category 1 cases. The statistically higher ratio of triple-negative subtype was detected in BI-RADS category 1 cases, and histological grade 3 tended to be also higher in the BI-RADS category 1 group. Results above did indicate that the BI-RADS category 1 group was histologically characterized by a higher malignant level than those with mammographic abnormalities, but it awaits further investigations for clarification.

Previous study also demonstrated that earlier detection of breast cancer resulted in a decrement in mortality, which parallels the reduction in size distribution of cancers depicted and closely parallels the reduction in rates of node-positive breast cancer (19). Screening US also appears to detect many breast cancer cases at a smaller size and earlier stage compared with mammographic screening. In addition, in women with mammography dense breast, US was reported to be able to detect a substantially larger number of cancers with a supplemental cancer detection of 0.3–0.5% by US alone (18). Therefore, it is important to detect the US findings with the localized low-echoic lesion. In addition, among the BI-RADS category 1 group, particular attention should be paid to the US findings such as solidary differentiated masses such as oval or round shape and non-spiculated periphery because the corresponding histopathological features of the cases associated with these US findings above include a much higher ratio of triple-negative subtype and/or histological grade 3. Therefore, early detection of such solidary masses with triple-negative subtype and/or high histological grade by US may possibly contribute to the eventual reduction in breast cancer mortality.

We evaluated US findings and the corresponding histopathological characteristics for BI-RADS category 1 mammograms and noted significant differences among these findings in this study. Evaluation of these US and histopathological characteristics may provide a more accurate US screening system for Japanese women.

Acknowledgement

We thank Yayoi Takahashi, MT, for her excellent technical assistance for immunohistochemical staining.

Funding

This work was supported in part by a Grant-in Aid from 'Kurokawa Cancer Research Foundation'.

Conflict of interest statement

None declared.

References

1. Sutela A, Vanninen R, Sudah M, Berg M, Kiviniemi V, Rummukainen J, et al. Surgical specimen can be replaced by core samples in assessment of ER, PR and HER-2 for invasive breast cancer. *Acta Oncol* 2008;47:38–46.
2. Kawai M, Kuriyama S, Suzuki A, Nishino Y, Ishida T, Ohnuki K, et al. Effect of screening mammography on breast cancer survival in comparison to other detection methods: a retrospective cohort study. *Cancer Sci* 2009;100:1479–84.
3. Tamaki K, Sasano H, Miyashita M, Ishida T, Amari M, Ohuchi N, et al. A new mammographic classification: as a potential predictor of breast disorders for Asian women. 12th International St Gallen Breast Cancer Conference, St Gallen, Switzerland, 2011, SG-BCC2011-1204.
4. Osako T, Takahashi K, Iwase T, Iijima K, Miyagi Y, Nishimura S, et al. Diagnostic ultrasonography and mammography for invasive and noninvasive breast cancer in women aged 30–39 years. *Breast Cancer* 2007;14:229–33.
5. Crystal P, Strano SD, Shcharynski S, Koretz MJ. Using sonography to screen women with mammographically dense breasts. *Am J Roentgenol* 2003;181:177–82.
6. Tamaki K, Sasano H, Ishida T, Ishida K, Miyashita M, Takeda M, et al. The correlation between ultrasonographic findings and pathologic features in breast disorders. *Jpn J Clin Oncol* 2010;40:905–12.
7. Ohuchi N, Ishida T, Kawai M, Narikawa Y, Yamamoto S, Sobue T. Randomized controlled trial on effectiveness of ultrasonography screening for breast cancer in women aged 40–49 (J-START): research design. *Jpn J Clin Oncol* 2011;41:275–7.
8. Zanello PA, Robim AFC, Goncalves de Oliveira TM, Elias Junior J, Andrade JM, Monteiro CR, et al. Breast ultrasound diagnostic performance and outcomes for mass lesions using Breast Imaging Reporting and Data System category 0 mammogram. *Clinics* 2011;66:443–8.
9. American College of Radiology (ACR). *Breast Imaging Reporting and Data System (BI-RADS™)*. 4th edn. Reston, VA: American College of Radiology 2003.
10. Japan Association of Breast and Thyroid Sonography. *Guideline for Breast Ultrasound-Management and Diagnosis*. 2nd edn. Tokyo: Japanese 2008.
11. Tavassoli FA, Devilee P. *World Health Organization Classification of Tumors. Tumor of the Breast and Females Genital Organs*. Lyon: IARC Press 2003.
12. Rosen PP. *Rosen's Breast Pathology*. 3rd edn. Philadelphia, PA: Lippincott Williams & Wilkins 2009.

13. Allred DC, Harvey JM, Berardo M, Clark GM. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol* 1998;11:155–68.
14. Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, et al. American society of clinical oncology/college of American pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J Clin Oncol* 2007;25:118–45.
15. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histopathological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 1991;19:403–10.
16. Silverstein MJ. Prognostic classification of breast ductal carcinoma in situ. *Lancet* 1995;345:1154–7.
17. Tamaki K, Moriya T, Sato Y, Ishida T, Maruo Y, Yoshinaga K, et al. Vasohibin-1 in human breast carcinoma: a potential negative feedback regulator of angiogenesis. *Cancer Sci* 2009;100:88–94.
18. Youk JH, Kim EK. Supplementary screening sonography in mammographically dense breast: pros and cons. *Korean J Radiol* 2010;11:589–93.
19. Berg WA. Beyond standard mammographic screening: mammography at age extremes, ultrasound, and MR imaging. *Radiol Clin North Am* 2007;45:895–906.

Radiol Phys Technol (2013) 6:70–77
DOI 10.1007/s12194-012-0171-4

Usefulness of presentation of similar images in the diagnosis of breast masses on mammograms: comparison of observer performances in Japan and the USA

Chisako Muramatsu · Robert A. Schmidt · Junji Shiraishi · Tokiko Endo · Hiroshi Fujita · Kunio Doi

Received: 21 May 2012 / Revised: 6 July 2012 / Accepted: 22 July 2012 / Published online: 8 August 2012

Abstract Computer-aided diagnosis has potential in improving radiologists' diagnosis, and presentation of similar images as a reference may provide additional useful information for distinction between benign and malignant lesions. In this study, we evaluated the usefulness of presentation of reference images in observer performance studies and compared the results obtained by groups of observers practicing in the United States and Japan. The results showed that the presentation of the reference images was generally effective for both groups, as the areas under the receiver operating characteristic curves improved from 0.915 to 0.924 for the group in the US and from 0.913 to 0.925 for the group in Japan, although the differences were marginally ($p = 0.047$) and not ($p = 0.13$) statistically significant, respectively. There was a slight difference between the two groups in the way that the observers reacted to some benign cases, which might be due to differences in the population of screenees and in the socio-clinical environment. In the future, it may be worthwhile to investigate the development of a customized system for physicians in different socio-clinical environments.

Keywords Similar images · Computer-aided diagnosis · Breast masses · Mammograms · Image retrieval · Observer study

C. Muramatsu (✉) · H. Fujita
Department of Intelligent Image Information,
Graduate School of Medicine, Gifu University,
1-1 Yanagido, Gifu, Japan
e-mail: chisa@fjt.info.gifu-u.ac.jp

R. A. Schmidt · K. Doi
Department of Radiology, The University of Chicago,
5841 S Maryland Ave, Chicago, IL 60637, USA

J. Shiraishi
Department of Medical Information Systems,
School of Health Sciences, Kumamoto University,
2-31-1 Kurokami, Kumamoto, Japan

T. Endo
Department of Radiology, Nagoya Medical Center,
4-1-1 Nakaku Sannomaru, Nagoya, Aichi, Japan

K. Doi
Gunma Prefectural Collage of Health Sciences,
328-1 Kamiokimachi, Maebashi, Gunma, Japan

1 Introduction

Breast cancer is the most frequently diagnosed cancer and one of the leading causes of cancer deaths in women in Japan, the United States (US), and European countries [1–3]. Mammography is considered the most effective screening method for early detection of breast cancer for women at normal risk [4–6]. For improving the diagnostic accuracy and efficiency, computer-aided detection (CADe) was introduced [7–10], and its potential usefulness was indicated in an observer performance study [11] and in prospective studies [12–16]. Once a suspicious lesion is found, radiologists may determine whether it should be biopsied or followed up. However, diagnosis on mammograms can be difficult and requires proper training and reading experience. For assisting radiologists' reading, investigators have suggested computer-aided diagnosis (CADx), in which a computer provides the likelihood of malignancy of an unknown lesion and have reported the potential usefulness of CADx in distinguishing between benign and malignant lesions on mammograms [17–19]. In these studies, radiologists' performance in terms of the area (AUC) under the receiver operating characteristic (ROC)

curve was improved with use of CADx; however, the studies indicated that the AUCs by many observers with CADx were lower than the AUC by the computer alone. One of the reasons might be that the result of computer analysis was summarized only in one numeral, i.e., the likelihood of malignancy, and the evidence was not clear to radiologists.

In recent years, mammography practice has been shifting from analogue to digital images. With implementation of Picture Archiving and Communication System (PACS), it became much easier to store and retrieve images from the previous examinations, and an effective use of stored data is expected. Radiologists, on the other hand, are trained and gain experience by reading many images in their clinical practice, in textbooks, and in training courses. Therefore, presentation of images that are similar to an unknown image can be an intuitive guide to reinforce the numerical likelihood of malignancy [20]. Different methods for automated selection of similar images have been investigated for diagnosis of chest radiographs [21, 22], thoracic computed tomographs [23, 24], and mammograms [25–31]. In some studies, reference images were selected on the basis of the predicted diagnosis [21, 22, 27, 28], whereas in other studies, images were selected by the similarity of the feature values [24, 26, 28]. For selecting similar images from the point of view of diagnosis, we have been investigating a method for quantifying the subjective ratings by radiologists [29, 30], as well as, a similarity index that takes into account the subjective similarity rated by radiologists [23, 31–33]. In our method, the similarity measure, called a psychophysical measure, was determined using an artificial neural network (ANN) which would be trained to learn the relationship between the subjective similarity ratings by radiologists and the computer-extracted image features.

Some of the above studies have indicated the potential usefulness of providing reference images together with other information such as the predicted diagnosis [22–24, 27, 34]. In these studies, it was not clear whether the presentation of reference images itself or the both images and other information together was helpful. Therefore, in order to evaluate the usefulness of providing similar images and to investigate the effect on radiologists in detail, we conducted the observer performance study to evaluate the radiologists' abilities in distinguishing between benign and malignant masses without and with similar images [35]. The result of this study was that, although the presentation of similar images provided beneficial effects, the average AUC was almost unchanged. One of the important findings in this study was that a reference image database must be carefully created so that it does not include "confusing" cases. When textbook-type cases are shown, radiologists will feel comfortable and can confidently and properly react to the given information. However, when atypical cases are presented, radiologists may become anxious, especially if they are cases of cancer. Another finding was that when a new case in question may be a benign-looking malignant case or a malignant-looking benign case, similar images would not be helpful. For these cases, radiologists' initial assessment is likely to be incorrect, and the presentation of similar images would only reinforce their incorrect decisions. This effect is, in fact, one that any type of CADx could have in common. When radiologists' initial judgment for a malignant-looking benign lesion was malignant, and the computer-estimated likelihood was also malignant, radiologists would become confident of their incorrect decision.

In this paper, we report the result from our second observer study after manual refinement of the database to exclude confusing cases. A group of radiologists practicing in the US and another group of radiologists and breast surgeons in Japan participated in the observer study, and the results from the two groups were compared.

2 Materials and methods

2.1 Case selection

Regions of interest (ROIs) including breast masses were used in this study. They were obtained from the Digital Database for Screening Mammography (DDSM) [36], which was made available by the University of South Florida. We initially collected 1568 ROIs, including 728 malignant and 840 benign masses [31]. ROIs containing microcalcifications which may influence the diagnosis of masses were excluded from the database. In the previous observer study [35], potentially confusing cases (benign-looking malignant and malignant-looking benign masses) were excluded from the reference database, which consisted of 365 malignant and 442 benign masses, by use of the computer-estimated likelihood of malignancy. However, because of the imperfect accuracy of the estimated likelihood, the database was suboptimal. In this study, for creating a better reference database, all of the cases were rated by a co-author (C.M.) for their difficulty in distinguishing between benign and malignant from 1 to 4, with 1 being difficult and 4 being easy. The images with unclear masses or with low quality that may not be helpful when presented as a reference were rated 0.

For the observer study, 100 cases, including 50 malignant and 50 benign masses, were selected from those rated 2, 3, and 4 to serve as study cases (unknown cases). They were selected by stratified randomization according to the size distributions of the database with 2.5 mm size bins, excluding those less than 5 mm and those larger than

25 mm. After removing all of the ROIs obtained from the same patients that were selected as the unknown cases, the cases rated 3 and 4 comprised the reference database, which included 429 malignant and 480 benign mass ROIs.

For each unknown image, 8 images each from the malignant and benign groups were selected as "similar" reference images, although 4 images each were presented in the monitor, and the next 4 images were provided only if an observer requested it. In the selection of similar reference images, the size criterion of no more than 50 % difference in the effective diameters was applied first. From the remaining cases, reference images were selected on the basis of the psychophysical similarity measures, which were determined by the ANN trained with 300 sample pairs in the previous study [32]. In the training of the ANN, image features characterizing the shape, contrast, and margin were used as input, and the subjective similarity data by radiologists based on the overall impression for diagnosis were used as the teacher. For avoiding having the same image presented more than 5 times as the first 4 images in 100 cases, the top 10 images with the highest similarity measures were preselected, and 4 of them were used. Note that they were called the reference images because, in some cases, there may be no "similar" images with very high similarity measures, especially the benign reference images for the malignant unknown cases and the malignant reference images for the benign unknown cases.

2.2 Observer performance studies

Observer studies for evaluating the usefulness of presenting reference images in the distinction between benign and malignant masses were conducted at the University of Chicago, Chicago, USA, and at Nagoya Medical Center, Nagoya, Japan. During the studies, the images were shown on a monochrome liquid crystal display monitor (ME511L/P4, 21.3 in., 2048 × 2560 pixels, 410 cd/m^2 luminance; Totoku Electric Co., Ltd.). The readings were conducted in the sequential reading mode, in which an observer was asked to provide his/her confidence level of a lesion being malignant on a continuous rating scale from 0.00 to 1.00, corresponding to "definitely benign" and "definitely malignant," respectively. Immediately after the first rating, four "similar" benign images and four "similar" malignant images were presented on the right and left sides of the unknown case, and the observer was asked to reconsider his/her confidence level. If an observer requested it, next four benign and four malignant images would be shown.

The instructions to the observers were: (1) The purpose of this study is to investigate whether providing the similar known images can assist radiologists in the distinction between benign and malignant lesions on mammograms. (2) 100 unknown cases are included in this study. The training session including four cases is provided at the beginning of the study. (3) You are asked to provide your confidence level regarding the malignancy (or benignity) of a lesion with a bar displayed on the monitor by use of a mouse first without similar images, and then after observing the similar images. (4) For each unknown case, four most similar images each from benign and malignant lesions in the database are provided. If desired, you may observe additional four similar images by clicking a "show similar images 5–8" button. (5) There is no time limit.

Ten observers, including two attending breast radiologists, three breast imaging fellows, and five radiology residents, participated in the observer study in the US. The two attendings had 13 and 3 years of experience in reading mammograms, whereas the three fellows were in the first or second year of a breast imaging fellowship, and the residents were third- or fourth-year senior residents who had been trained in the breast-imaging section in their rotation. Eleven observers, including 10 radiologists and breast surgeons who were certified for breast image reading and one in training to be certified, participated in Japan. The ten certified physicians had a mean of 13 years of experience in reading mammograms. The results were evaluated by use of multi-reader multi-case (MRMC) ROC analysis (the University of Chicago, IL, USA) [37].

3 Results

The overall results indicated that AUCs without and with the presentation of the reference images were both high, probably because difficult cases were not included in this study. The AUCs without and with the reference images for the individual observers are listed in Table 1. The mean AUCs were slightly improved by providing the reference images, from 0.915 to 0.924 for the group in the US and from 0.913 to 0.925 for the group in Japan; however, the differences were not statistically significant for the Japan group and marginally significant for the US group. If a change in the confidence level of more than 0.1 in the direction of the correct diagnosis is considered a beneficial effect, on average, there were larger numbers of cases that the presentation of the reference images affected beneficially than those affected detrimentally. Figure 1 shows the numbers of beneficially and detrimentally affected cases for each observer. The average numbers of beneficially and detrimentally changed cases were 13 and 8, respectively, for the US group and 16 and 7, respectively, for the Japan group with p values of 0.04 and 0.01 by paired t test.

It may be noticed in the results that for the US observers, there were more beneficial effects to the malignant cases than to the benign cases, and the majority of the detrimental cases were benign cases. On the other hand, for the

Table 1 Areas under the receiver operating characteristic curves without and with the presentation of the reference images for the individual observers

Observers	US group		Japan group	
	Without	With	Without	With
A	0.951	0.962	0.939	0.942
B	0.972	0.978	0.893	0.924
C	0.940	0.938	0.936	0.947
D	0.947	0.942	0.969	0.941
E	0.942	0.943	0.880	0.880
F	0.906	0.931	0.888	0.919
G	0.874	0.877	0.879	0.888
H	0.887	0.913	0.905	0.958
I	0.871	0.874	0.918	0.940
J	0.863	0.885	0.901	0.885
K			0.936	0.952
Average (p value)	0.915	0.924 ($p = 0.047$)	0.913	0.925 ($p = 0.13$)

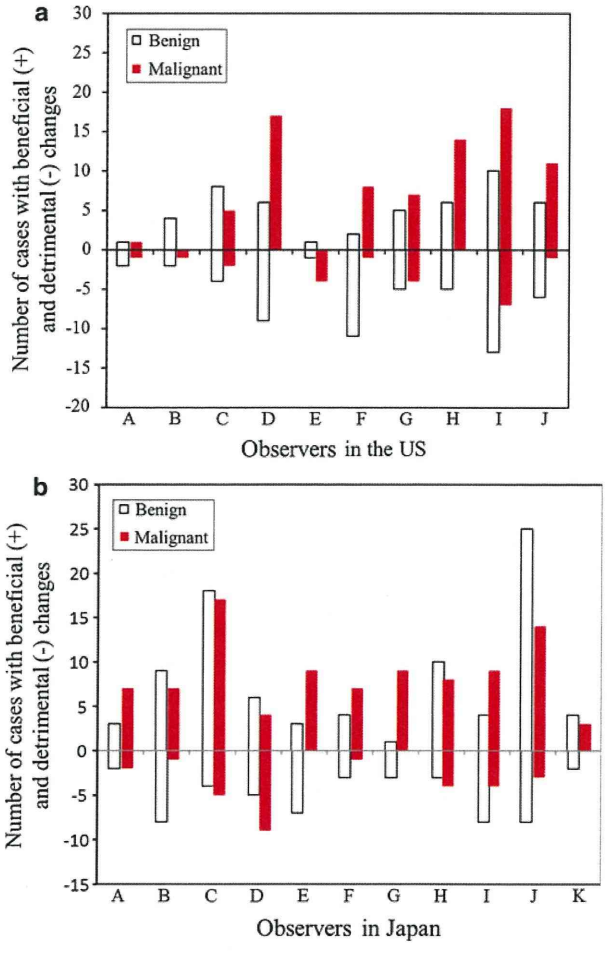


Fig. 1 Numbers of cases that the presentation of the reference images affected beneficially (positive) and detrimentally (negative) for the individual observers. **a** Observers in the US, and **b** observers in Japan

Japanese observers, the presentation of the reference images was helpful for both the malignant and the benign cases. Figure 2 shows the relationships between the average initial confidence levels and their changes, where positive changes correspond to the changes toward a correct diagnosis, for the two groups of observers. It is apparent in the figures that the presentation of the reference images had beneficial effects for many of the malignant cases. For the benign cases, however, it caused the US observers to increase their confidence levels toward malignant (indicated by an arrow in Fig 2a). The results indicate that some observers, regardless of their initial judgments as benign or uncertain, became worried after the reference images were presented. On the other hand, the average initial confidence levels for the benign cases by the Japanese observers were, on average, low for many cases, and the observers became confident of their judgment after the presentation of the reference images (indicated by an arrow in Fig 2b).

Figure 3 shows the relationship between the changes in the confidence levels before and after the presentation of the reference images by the two groups. The points in the right upper quadrant correspond to the cases in which the presentation of the reference images, on average, resulted in the beneficial changes for both groups of observers. An example of such cases is shown in Fig. 4. In this case, the unknown case was malignant. The initial judgments by the observers in both groups were mostly uncertain, and their confidence levels increased after the reference images were presented, with six of them increasing more than 0.1. On the other hand, there are some benign cases in the upper left quadrant in Fig. 3 for which the image presentation resulted in beneficial changes for the Japanese observers, but caused the detrimental changes for the US observers. Figure 5 shows one of such cases. In this case, the observers' initial judgments ranged from somewhat uncertain to likely benign, and the presentation of the reference images caused some US observers to increase their ratings, whereas most Japanese observers remained unchanged.

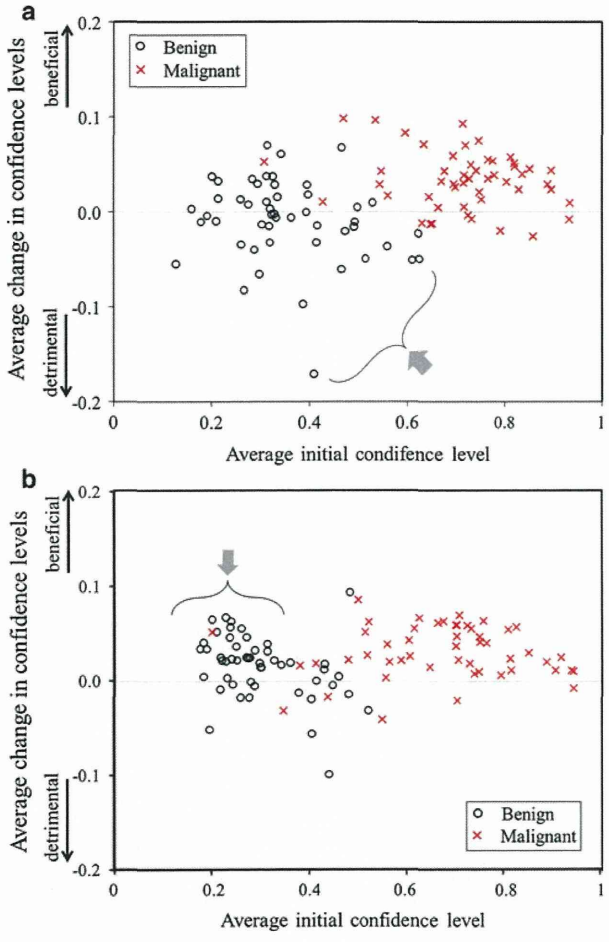


Fig. 2 Relationships between the average initial confidence levels and the changes in confidence levels toward (+) correct and (−) incorrect diagnosis by the **a** observers in the US and **b** observers in Japan. An *arrow* in **a** indicates the benign cases in which presentation of similar images caused detrimental effects, whereas an *arrow* in **b** indicates the benign cases in which the presentation caused beneficial effects

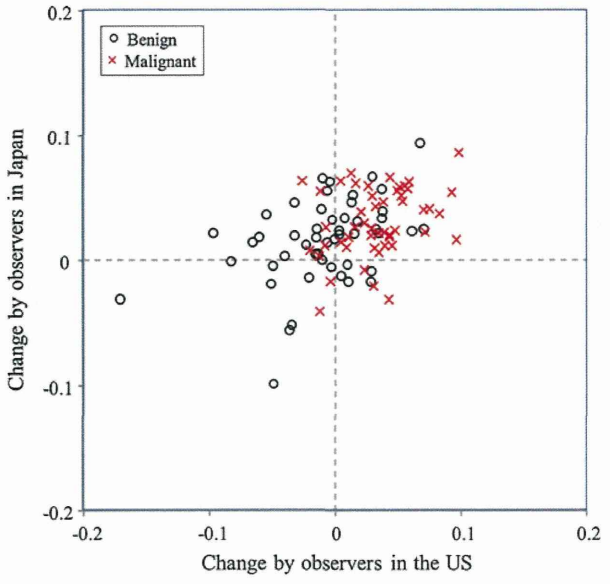
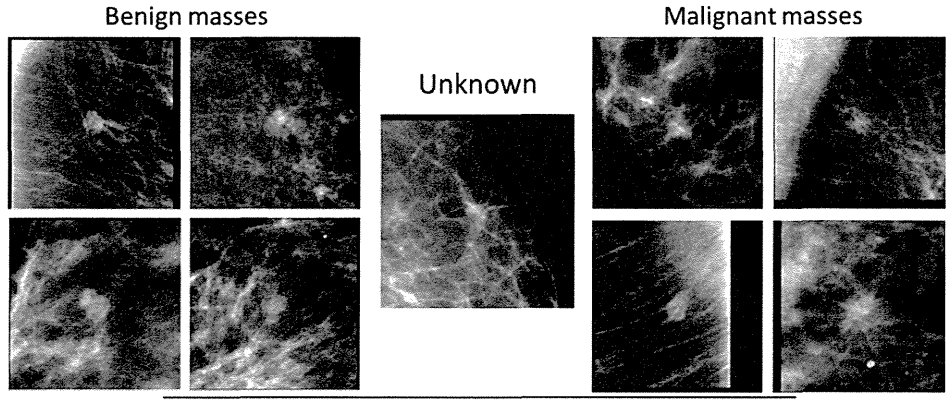


Fig. 3 Relationship between the changes in confidence levels before and after the presentation of the reference images between the two groups of observers

4 Discussion

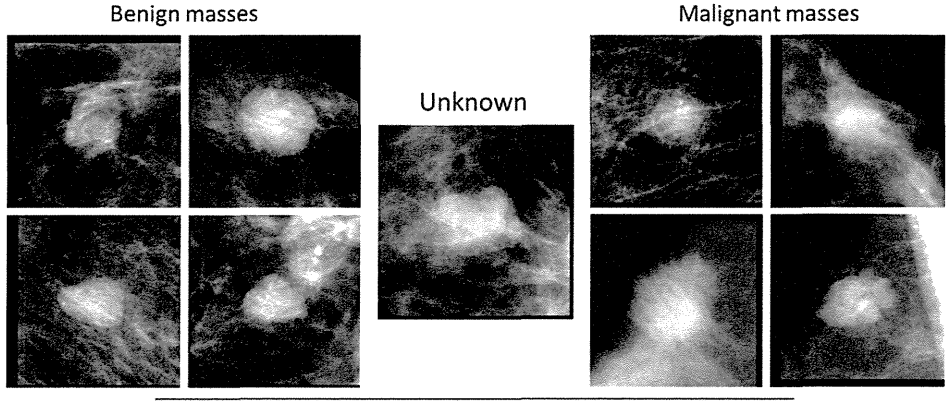
The results of the observer studies showed a somewhat notable difference between the practitioners in the US and those in Japan in the sense of their reaction when the reference images were presented. The differences between the two groups seemed more prominent in the benign than in the malignant cases. One difference we observed during the reading sessions and also obtained in the observers' feedback was that the practitioners in the US primarily and dominantly consider margin characteristics in distinguishing between benign and malignant masses, whereas the practitioners in Japan consider the density of the masses, which relates to their elasticity, in addition to the margin characteristics. This may be due to the fact that Japanese women tend to have dense breasts, and physicians have a difficult time assessing margins more often than those in the US. It is also related to the fact that the breast cancer incidence rate increases with age in the US, whereas it peaks around the late 40 s in Japan. Therefore, the observers in Japan often complained about the use of ROIs without the availability of whole mammographic views during the observer study. When they read mammograms, the relative mass density in comparison with the normal breast tissue density of the patient is one of the important factors that they consider. However, with the lack of a whole view, it was difficult to see the grandular tissue density of the whole breast and the symmetry against the opposite breast.

Another perspective could be related to the number of law suits on missed cancers in the US. Although nobody wants to miss a cancer, physicians in the US may be particularly sensitive to missing one. This is manifested in the reported higher recall rates in the US than those in other countries. According to the study by the Physicians Insurers Association of America, breast cancer is the subject of the most frequent malpractice lawsuits filed, in which 41 % of all claims resulted in compensation averaging about $438,000 [38]. Dick et al. [39] have reported that, in their surveys, about a half of US radiologists responded that they had had a malpractice claim filed against them. In Japan, based on the statistics by the Supreme Court (http://www.courts.go.jp/saikosai/iinkai/izikankei/index.html), the number of medical lawsuits in each year is about 1000, in which internal medicine, surgery, orthopedics, and gynecology are the top four



	Average change	Number of observers who changed
Observers in the US	0.10 (0.47 → 0.57)	3
Observers in Japan	0.09 (0.50 → 0.59)	3

Fig. 4 A malignant case in which the presentation of the reference images resulted in beneficial changes for the both groups



	Average change	Number of observers who changed
Observers in the US	-0.10 (0.39 → 0.48)	-5
Observers in Japan	0.02 (0.33 → 0.31)	1

Fig. 5 A benign case in which the presentation of the reference images resulted differently for the observers in the US and those in Japan

frequently filed, accounting for more than 60 % of cases; no number was provided for radiology. Although it is difficult to compare these statistics, it can be conjectured that US radiologists tend to practice more defensive medicine. The difference is also seen in the diagnostic assessment of probably benign lesions. In the US, cases assessed as BI-RADS 3, "probably benign finding", should have less than 2 % risk of malignancy, whereas a breast imaging guideline in Japan was created on the basis of the BI-RADS, and cases assessed as category 3 "benign but malignancy can't be ruled out" may have about a 2–10 % chance of cancer. These facts may explain the tendency of the US observers to give slightly higher ratings than those in Japan.

There were some differences in the years of experience between the two groups. It is difficult to determine whether the different reaction to the benign cases could be due to the years of experience, because US attending radiologists had a tendency to make only small changes in confidence levels. Note that the years of experience is one index; US attending radiologists and fellows only practice in breast imaging section in their routine work, whereas Japanese radiologists, although experts in breast image reading, may also read images of other organs, and surgeons may spend limited time in image reading. In addition, Japanese observers work at several different clinical facilities, and their practice may be somewhat different. The population of the test cases which were obtained in the US and primarily included Caucasians and African Americans may have affected the performance. Although the average years of experience and their background were different between two groups, the mean AUCs without and with similar images were comparable.

One of the limitations in this study was that we excluded difficult cases from the test dataset, which resulted in high AUCs both without and with similar images. Because of the limitation of time, the population of test cases selected for the observer performance studies were generally different from the clinical population. In this study, we excluded difficult cases because it was believed that CAD likely has no impact or detrimental effect on such cases. When a benign lesion looks very similar to typical malignant cases, a computer likely selects similar malignant lesions and outputs a high likelihood of malignancy. Even if a computer provided a low likelihood of malignancy, it is unlikely that radiologists would change their initial decision. Although we believe that such atypical cases are relatively rare, we did not include them in the present study because the number of the study cases was limited. As a result, the impact of the overall beneficial effect observed in this study could be much smaller in an actual clinical population. On the other hand, the high AUCs without similar images might have decreased the chances of gain.

5 Conclusion

The results of the observer studies indicate a potential utility of presenting reference images in the distinction between benign and malignant masses on mammograms by physicians. The overall effects in terms of the mean AUC were comparable for the observers in both counties. However, there was a slight difference in the reactions by the observers for some benign cases. This difference could be due to the differences in the patient population and the diagnostic environment in the two countries. In this study, the similarity measures used for the selection of reference images were based on the subjective similarity ratings determined by breast radiologists who practice in the US. Although subjective similarities noted by different groups of observers were expected to be comparable for most of the cases, there could be some differences in the impression due to the diagnostic environment. For improving the utility of computer-aided diagnosis systems, it may be worthwhile to investigate the development of a customized CAD system with an effective image selection scheme for physicians in different socio-clinical environments.

Acknowledgments This study was partly supported by the US Department of Defense Breast Cancer Research Program and Grants-in-Aid for Scientific Research in Japan. The authors are grateful to the physicians who participated in the observer studies.

References

1. Matsuda T, Marugame T, Kamo KI, Katanoda K, Ajiki W, Sobue T, The Japan Cancer Surveillance Research Group. Cancer incidence and incidence rates in Japan in 2006: based on data from 15 population-based cancer registries in the Monitoring of Cancer Incidence in Japan (MCIJ) Project. Jpn J Clin Oncol. 2011;42:139–47.
2. Americal Cancer Society. Cancer facts & figures 2012. Atlanta: American Cancer Society; 2012.
3. Ferlay J, Autier P, Boniol M, Heanue M, Colombet M, Boyle P. Estimates of the cancer incidence and mortality in Europe in 2006. Ann Oncol. 2007;18:581–92.
4. Tabar L, Fagerberg G, Duffy SW, Day NE, Gad A, Grontoft O. Update of the Swedish two-county program of mammographic screening for breast cancer. Radiol Clin North Am. 1992;30: 187–210.
5. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Selection, follow-up, and analysis in the health insurance plan study: a randomized trial with breast cancer screening. J Natl Cancer Inst Monogr. 1985;67:65–74.
6. Humphrey LL, Helfand M, Chan BKS, Woolf SH. Breast cancer screening: a summary of the evidence for the U.S. preventive services task force. Ann Intern Med. 2002;137:E-347–67.
7. Doi K, Giger ML, MacMahon H. Computer-aided diagnosis: development of automated schemes for quantitative analysis of radiographic images. Smin Ultrasound CT MRI. 1992;13:140–52.
8. Vyborny CJ. Can computers help radiologists read mammograms? Radiology. 1994;191:315–7.
9. Giger ML, Huo Z, Kupinski MA, Vyborny CJ. Computer-aided diagnosis in mammography. In: Fitzpatrick JM, Sonka M, editors. The handbook of medical imaging, medical imaging processing and analysis, vol 2. SPIE;2000. p. 915–1004.
10. Doi K. Diagnostic imaging over the last 50 year: research and development in medical imaging science and technology. Phys Med Biol. 2006;51:R5–27.
11. Chan HP, Doi K, Vyborny CJ, Schmidt RA, Metz CE, Lam KL, Ogura T, Wu Y, MacMahon H. Improvement in radiologists' detection of clustered microcalcifications on mammograms. Invest Radiol. 1990;25:1102–10.
12. Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. Radiology. 2001;220:781–6.
13. Birdwell RL, Bandodkar P, Ikeda DM. Computer-aided detection with screening mammography in a university hospital setting. Radiology. 2005;236:451–7.
14. Cupples TE, Cunningham JE, Reynolds JC. Impact of computer-aided detection in a regional screening mammography program. Am J Roentgenol. 2005;185:944–50.
15. Morton MJ, Whaley DH, Brandt KR, Amrami KK. Screening mammograms; interpretation with computer-aided detection—prospective evaluation. Radiology. 2006;239:375–83.
16. Dean JC, Ilvento CC. Improved cancer detection using computer-aided detection with diagnostic and screening mammography: prospective study of 104 cancers. Am J Roentgenol. 2006;187: 20–8.
17. Chan HP, Sahiner B, Roubidoux MA, Wilson TE, Adler DD, Paramagul C, Newman JS, Sanjay-Gopal S. Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: an ROC study. Radiology. 1999;212: 817–27.
18. Huo Z, Giger ML, Vyborny CJ, Metz CE. Breast cancer: effectiveness of computer-aided diagnosis—observer study with independent database of mammograms. Radiology. 2002;224: 560–8.
19. Jiang Y, Nishikawa RM, Schmidt RA, Metz CE, Giger ML, Doi K. Improving breast cancer diagnosis with computer-aided diagnosis. Acad Radiol. 1999;6:22–33.
20. Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput Med Imaging Graph. 2007;31:198–211.

21. Swett HA, Fisher PR, Cohn AI, Miller PL, Mutalik PG. Expert system-controlled image display. Radiology. 1989;172:487–93.
22. Aisen AM, Broderick LS, Winer-Muram H, Brodley CE, Kak AC, Pavlopoulou C, Dy J, Shyu CR, Marchiori A. Automated storage and retrieval of thin-section CT images to assist diagnosis: system description and preliminary assessment. Radiology. 2003;228:265–70.
23. Li Q, Li F, Shiraishi J, Katsuragawa S, Sone S, Doi K. Investigation of new psychophysical measures for evaluation of similar images on thoracic CT for distinction between benign and malignant nodules. Med Phys. 2003;30:2584–93.
24. Kawata Y, Niki N, Ohmatsu H, Moriyama N. Example-based assisting approach for pulmonary nodule classification in three-dimensional thoracic computed tomography images. Acad Radiol. 2003;10:1402–15.
25. Swett HA, Mutalik PG, Neklesa VP, Horvath L, Lee C, Richter J, Tocino I, Fisher P. Voice-activated retrieval of mammography reference images. J Digit Imaging. 1998;11:65–73.
26. Qi H, Snyder WE. Content-based image retrieval in picture archiving and communications systems. J Digit Imaging. 1999; 12:81–3.
27. Sklansky J, Tao EY, Bazargan M, Ornes CJ, Murchison RC, Teklehaimanot S. Computer-aided, case-based diagnosis of mammographic regions of interest containing microcalcifications. Acad Radiol. 2000;7:395–405.
28. Giger ML, Huo Z, Vyborny CJ, Lan L, Bonta I, Horsch K, Nishikawa RM, Rosenbourgh I. Intelligent CAD workstation for breast imaging using similarity to known lesions and multiple visual prompt aids. Proc SPIE. 2002;4684:768–73.
29. Muramatsu C, Li Q, Schmidt RA, Suzuki K, Shiraishi J, Newstead GM, Doi K. Experimental determination of subjective similarity for pairs of clustered microcalcifications on mammograms: observer study results. Med Phys. 2006;33:3460–8.
30. Muramatsu C, Li Q, Schmidt RA, Shiraishi J, Suzuki K, Newstead GM, Doi K. Determination of subjective similarity for pairs of masses and pairs of clustered microcalcifications on mammograms: comparison of similarity ranking scores and absolute similarity ratings. Med Phys. 2007;34:2890–5.
31. Muramatsu C, Li Q, Suzuki K, Schmidt RA, Shiraishi J, Newstead G, Doi K. Investigation of psychophysical measure for evaluation of similar images for mammographic masses: preliminary results. Med Phys. 2005;32:2295–304.
32. Muramatsu C, Li Q, Schmidt RA, Shiraishi J, Doi K. Determination of similarity measures for pairs of mass lesions on mammograms by use of BI-RADS lesion descriptors and image features. Acad Radiol. 2009;16:443–9.
33. Muramatsu C, Li Q, Schmidt RA, Shiraishi J, Doi K. Investigation of psychophysical similarity measures for selection of similar images in the diagnosis of clustered microcalcifications on mammograms. Med Phys. 2008;35:5695–702.
34. Horsch K, Giger ML, Vyborny CJ, Lan L, Mendelson EB, Hendrick ER. Classification of breast lesions with multimodality computer-aided diagnosis: observer study results on an independent clinical data set. Radiology. 2006;240:357–68.
35. Muramatsu C, Schmidt RA, Shiraishi J, Li Q, Doi K. Presentation of similar images as a reference for distinction between benign and malignant masses on mammograms: analysis of initial observer study. J Digit Imaging. 2010;23:592–602.
36. Heath M, Bowyer K, Kopans D, Moore R, Kedelmeyer P. Current states of the digital database for screening mammography. Digital mammography. Dordrecht: Kluwer; 1998.
37. Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis: generalization to the population of readers and patients with the jackknife method. Invest Radiol. 1992;27:723–31.
38. Physician Insurers Association of America. PIAA 2002 breast cancer study. Rockville: Physician Insurers Association of America; 2002.
39. Dick JF III, Gallagher TH, Brenner RJ, Yi JP, Reisch LM, Abraham L, Miglioretti DL, Carney PA, Cutter GR, Elmore JG. Predictors of radiologists' perceived risk of malpractice lawsuits in breast imaging. Am J Roentgenol. 2009;192:327–33.

原　著 ● Original Article

乳がん超音波検診の精度管理に関するアンケート結果

つくば総合健診センター[1]，日本乳癌検診学会超音波検診精度管理委員会委員長[1]，
NPO法人日本乳腺甲状腺超音波診断会議(JABTS)乳房超音波講習会実行委員長[1]
立花病院[2]，JABTS教育委員会委員[2]，静岡県立静岡がんセンター乳腺外科[3]，JABTS教育委員会委員長[3]
国立国際医療研究センター外科，前JABTS理事長[4]，昭和大学医学部乳腺外科[5]，JABTS理事長[5]
川崎医科大学乳腺甲状腺外科[6]，日本乳癌検診学会超音波検診精度管理委員会副委員長[6]

東野英利子[1]　藤本　泰久[2]　田中久美子[3]　安田　秀光[4]
中村　清吾[5]　園尾　博司[6]

要旨：2010年度のJABTS主催乳房超音波講習会受講者を対象に，乳がん超音波検診の精度管理に関するアンケートを行い，精度管理基準作成の参考とすることを目的とした。技師対象の講習会で結果が得られた169の検診施設を解析対象とした。検診受診者数は1,000から5,000名の施設と100から500名の施設が多かった。超音波実施者は5名以下の施設が多いが，実施者の多い施設では臨床検査技師が検査に従事していることが多い。1名の受診者を検査するのに要する時間は平均でスキャン時間8.7分，入室から退室までで12.7分であった。結果判定における過去画像との比較は84％で行われており，画像の電子化は79％で行われていた。精密検査結果の把握は3割程度しか行われていなかった。

以上の結果から，乳がん超音波検診実施者・判定者の資格の制度化，検査方法，結果判定方法の標準化が必要と考えられる。また精密検査結果の把握を義務付けることも必要と考えられる。

索引用語：乳癌，検診，超音波，精度管理

緒　言

検診においては精度管理が重要である。超音波による乳がん検診は日本で最も多く行われているが，任意型検診で行われていることが多く，精度管理指針はまだ公には作成されていない。日本乳癌検診学会超音波検診精度管理委員会では今後精度管理指針を作成する予定であるが，それに先立ち，現在すでに行われている検診において，精度管理がどのように行われているかを知ることは重要である。そこで，NPO法人日本乳腺甲状腺超音波診断会議(JABTS)教育委員会の協力を得て，JABTS教育委員会が主催する講習会において受講者に精度管理に関するアンケートを行った。

別冊請求先：〒305-0005　つくば市天久保1丁目2番地
公益財団法人筑波メディカルセンター
つくば総合健診センター　東野英利子
e-mail address: tohno@tmch.or.jp

1. 対象および方法

アンケートの内容に関しては，日本乳癌検診学会超音波検診精度管理委員会で検討した。その内容を**表1**に示す。2010年度に開催されたJABTS教育委員会主催の乳房超音波講習会は9回で，技師対象が6回，医師対象が3回であった。アンケート用紙はすべての講習会において，あらかじめ受講者に送付し，記載してもらい，講習会時に回収した。アンケートの結果を確認したところ，医師からの回答には未記入部分や辻褄の合わない部分が多く，今回は技師講習会で回収したアンケートに関して解析を行った。6回の講習会の受講者数と開催地を**表2**に示す。総受講者数は287名であったが，期間中に同一施設から複数の受講者がある場合には受講時期が遅い方，受講番号の遅い方を対象とした。未回収および同一施設からの回答が54あり，また，検診施設ではないという回答が64あったため，解析の対象となったのは169のアンケート結果である。

アンケートのうち，4，5に関してはなかなか情報

表1. アンケート

乳がん超音波検診精度管理に関するアンケート

検診施設名＿＿＿＿＿＿＿＿＿＿＿＿　記入者のお名前＿＿＿＿＿＿＿＿＿＿＿＿

1．貴施設では年間に何人の乳がん超音波検診を行っていますか？
約＿＿＿＿＿＿名
□ 検診施設ではない(この場合は以下の質問にはお答えいただかなくて大丈夫です。)

2．貴施設での乳がん超音波検査検診の対象は以下のどれでしょうか。○を付けてください。
(　　)主に市町村検診
(　　)主に人間ドック，職域検診
(　　)上記の両方
(　　)その他　(具体的に ＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿)

3．貴施設では乳がん超音波検診従事者は何人いますか？

① 検診超音波検査を行っている技師・医師の数　総数 ＿＿＿＿＿名

うち臨床検査技師＿＿＿＿名，診療放射線技師＿＿＿＿名，医師＿＿＿＿名

② 超音波検診の結果判定を行っている医師の数 ＿＿＿＿＿名

4．超音波検診の超音波検査を実際に行っている方の資格についてお伺いします。
技師の場合
日本超音波学会認定超音波検査士の資格
体表　　人
検診　　人
両方　　人
JABTS主催あるいは共催の超音波講習会修了者
人
医師の場合
日本超音波医学会の専門医
総合　　人
乳腺甲状腺　人
日本乳癌学会認定
乳腺認定医　人
乳腺専門医　人
JABTS主催あるいは共催の超音波講習会修了者
人

5．超音波検診の判定者(通常は医師)についてお伺いします。
日本超音波医学会の専門医
総合　　人
乳腺甲状腺　人
日本乳癌学会認定
乳腺認定医　人
乳腺専門医　人
JABTS主催あるいは共催の超音波講習会修了者
人

6．貴施設は1名を検査するのにどのくらいの時間がかかっていますか？
① 受診者が入室してから退室するまでの時間(総時間を検査数で割っても結構です。)
約＿＿＿＿分

表1. アンケート(つづき)

② スキャン開始からスキャン終了までの時間(記入者の方の時間でも結構です。)
約　　分

7．貴施設では超音波検査結果の判定をする際に過去の超音波検査画像と比較していますか？(要精査にするかどうか，迷う場合のみでも比較しているとみなします。)
(はい・いいえ)　どちらかをお選びください。

8．貴施設の超音波診断装置の名称(会社名と装置名)と探触子の周波数について
① 装置1　(会社名)
(装置名)　　周波数(または帯域)　　MHz

② 装置2　(会社名)
(装置名)　　周波数(または帯域)　　MHz

③ 装置3　(会社名)
(装置名)　　周波数(または帯域)　　MHz

④ 装置4　(会社名)
(装置名)　　周波数(または帯域)　　MHz

⑤ 装置5　(会社名)
(装置名)　　周波数(または帯域)　　MHz

9．貴施設では画像の電子保存(施設内サーバあるいはハードディスクへの保存)を行っていますか？
(はい・いいえ)　どちらかをお選びください。

10．貴施設では以下の検診結果を調べていますか？

① 要精検率：(1),(2),(3),(4)下記を参照し，どれか一つをお選びください。

② 精検受診率：(1),(2),(3),(4)

③ 最終診断，病理診断：(1),(2),(3),(4)

④ がん発見率：(1),(2),(3),(4)

⑤ 早期がん割合：(1),(2),(3),(4)

(1) 調べており，定期的に知らされている
(2) 多分調べているが，結果は聞かされていない
(3) 調べていない
(4) わからない

11．精密検査結果(最終診断，病理診断)と検診結果(画像，判定)を検討する会を行っていますか？
どれか一つをお選びください。

(1) 定期的に行っている
(2) 自主的に(個人で)行っている
(3) あまり行っていない

ご協力ありがとうございました。

表2. 2010年度に開催されたJABTS教育委員会主催技師対象乳房超音波講習会の開催地(県)と受講者数

開催地	受講者数
茨城県	48
岡山県	48
東京都	48
宮城県	47
愛知県	48
栃木県	48
合　計	287

が得られないようで，未記入が多く，信頼性に欠けると考え，今回の解析の対象外とした。また8の装置に関しても今回は検討を行っていない。

2. 結　果

1)受診者数

各施設における年間の乳がん検診受診者数を**図1**に示す。1,000例以上5,000例未満が最も多く，ついで100例以上500例未満が多かった。

2)検診の種類の分布

主に市町村検診	22施設(13%)
主に人間ドック，職域検診	75施設(44%)
上記の両方	63施設(37%)
その他，あるいは無回答	9施設(5%)

3)超音波検診実施者の人数と人数別の職種

結果を**図2**に示す。超音波検査を実施するのは5名以下の施設が多く，また全体としては臨床検査技師が多い。実施者が少ない施設では医師も検査を行っている。

4)1名を検査するのに要する時間

①受診者が入室してから退室するまでの時間(未回答7)

平均12.7分，最短3分，最長30分

②スキャン開始からスキャン終了までの時間(未回答6)

平均8.7分，最短2分，最長20分

検査時間の分布を**図3**に示す。

5)判定における過去の超音波検査画像との比較

要精査にするかどうか，迷う場合のみを含め，判定において過去画像との比較を行っているのは142施設(84%)，行っていないのは10施設，無回答17施設であった。

6)画像の電子保存

超音波画像を施設内サーバあるいはハードディスクへ電子保存を行っている施設は133施設(79%)，行っていないのは30施設，無回答6施設であった。

7)検診結果の把握

要精検率，精検受診率，最終診断あるいは病理診断，がん発見率，早期癌割合の把握割合を**図4**に示す。最終診断に関しては定期的に調べている施設が約半数であったが，その他に関しては3割以下であった。また精密検査結果(最終診断，病理診断)と検診結果(画像，判定)を検討する会を定期的に行っている施設は55施設(33%)であったが，自主的に行っている施設も58施設あり，6割以上の施設で，何らかの検討を行っていることが分かった。

3. 考　察

日本における乳がん検診の基本はマンモグラフィである[1]が，超音波検査を併用することの有効性に関する評価が現在進行中である[2]。欧米に比して若い世代に乳がんの罹患率の高い日本では，超音波を用いた乳がん検診は任意型検診を中心にすでに広く行われてい

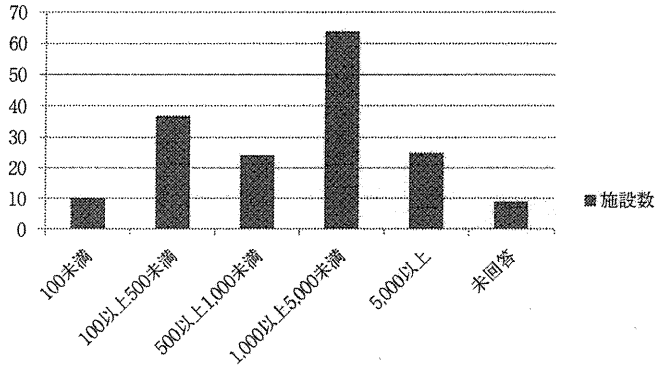
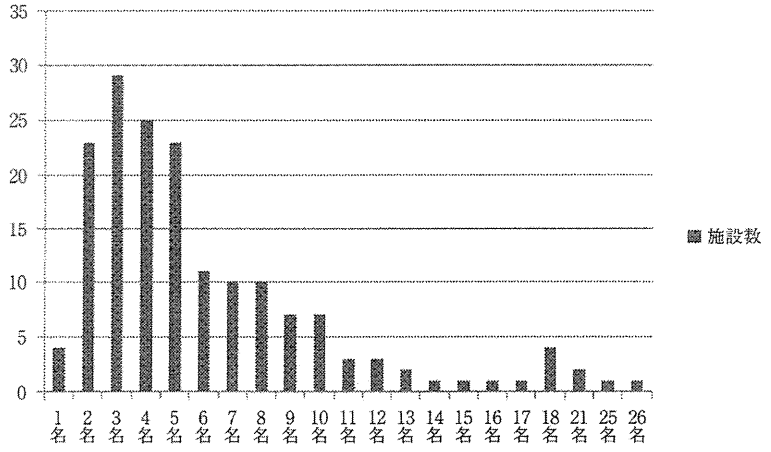
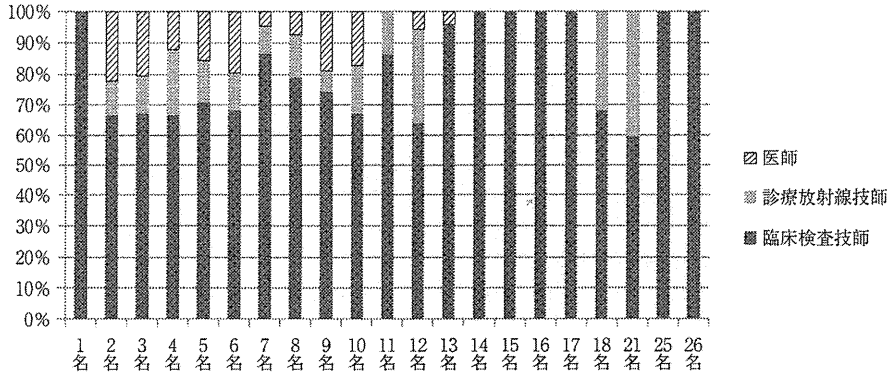


図1. 各施設における年間の乳がん検診受診者数



2－1．超音波検診実施者数とその分布



2－2．実施者数別職種割合

図 2. 超音波検診実施者の人数と人数別の職種

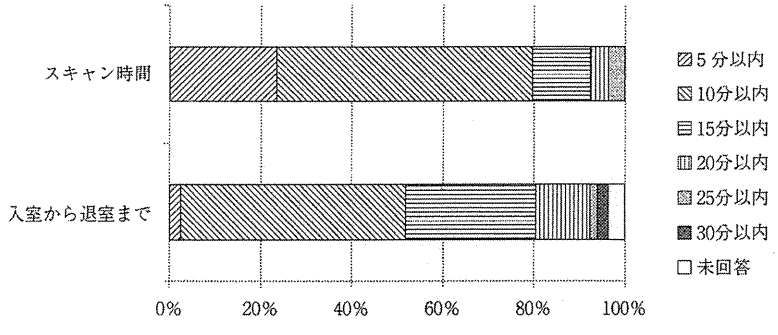


図 3. 検査時間の分布

る。そこで，超音波検診の精度管理を確立することは急務である。それには現在行われている検診の精度管理の状況を明らかにすることが必要である。JABTS教育委員会が主催する２日間の乳房超音波講習会は乳がん超音波検診の精度向上を目的としており[3]，検診施設からの受講者も多く，またすでに検査に従事している者が多い。受講者からのアンケート結果はある程度日本の現状を反映していると考えられる。

日本における乳がん検診の形態には検診専門施設での大規模な検診，病院等に付随した中規模の検診施設，さらに医療機関で行う比較的小規模の検診などがある。中小の施設では医師や診療放射線技師も超音波検査を行っているが，大規模施設では臨床検査技師が検査を行っている施設が多く，診療放射線技師はマンモグラフィの撮影，また医師は結果の評価という役割分担がされていると考えられる。今回は技師講習会のみを対象としたために医師のみで行っている施設はなかったが，医師一人で検査と判定を行っている医療機