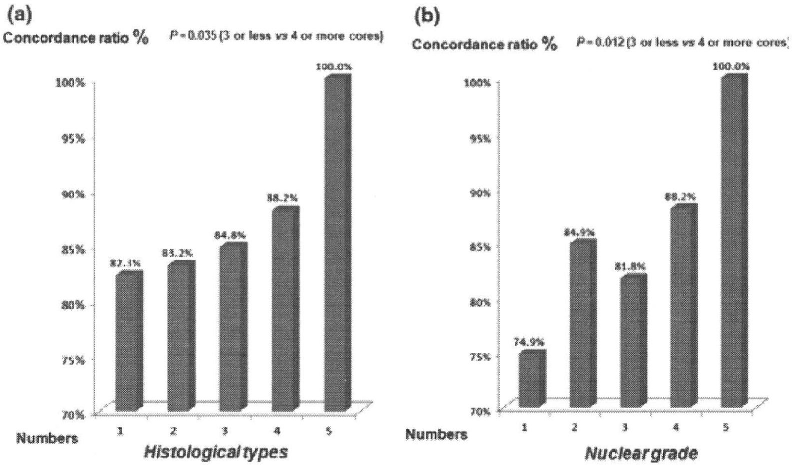


Fig. 1. Analysis of concordance rate of (a) histological type and (b) nuclear grade according to the number of core needle biopsies (CNB).

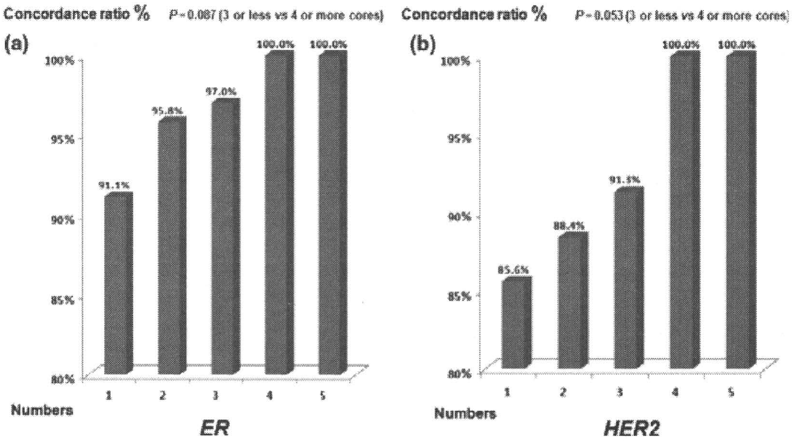


Fig. 2. Analysis of concordance rate of (a) estrogen receptor (ER) and (b) human epidermal growth factor receptor 2 (HER2) according to the number of core needle biopsies (CNB).

Discussion

Core needle biopsy (CNB) has been performed using a variety of devices to evaluate the nature of breast lesions. Core needle biopsy (CNB) has become the gold standard because of its lower inconclusive rate and the histological information it can provide.(19) The histological type of the lesions obtained by CNB was reported to be correlated closely with that of the excision specimen in 87 of 105 (83%) of cases.(5) In our present study, we demonstrated that the concordance rate of histological types between CNB and surgical specimens was 84.4% (298/353), which is very consistent with the results of report above. When the lesions are malignant, the presence or absence of invasion can be documented and the grade and type of tumor present can be assessed by histopathological evaluation of CNB.(19) It is also possible to examine ER, PgR, and HER2 status using immunohistochemistry in CNB specimens.(15) Therefore, information obtained from CNB greatly helps clinicians to determine a treatment plan for individual patients with regard to conservative management or primary chemotherapy.(19) However, some complications of CNB have been also reported in previous studies,(20–23) including postbiopsy pain,(20) hematoma,(21) infection(22) and seeding of carcinoma cells,(23) although such complications are considered rare if done in appropriate institutions by qualified staff. These indicate above all that CNB is a reliable and safe measure to diagnose breast disorders in a preoperative setting.

In our present study, 52 cases diagnosed as DCIS by CNB were subsequently diagnosed as IDC by surgical specimens; however, among these cases, 63.5% (33/52) were T1mic and T1a, and the tumor size was smaller than 2 cm. As for nuclear grades including nuclear atypia and mitotic counts, the concordance rates were almost 80%. In addition, more than 90% of discordant cases were within one grade discrepancies. In Table 2, we also demonstrate the tendency of nuclear atypia, mitotic counts, and nuclear grade of CNB to be lower compared to those of surgical specimens. This phenomenon is considered to be caused by the differences in methodologies employed between CNB and surgical specimens. We examined the higher atypical and mitotic area in surgical specimens, whereas we examined the narrow and limited area in CNB specimens. Therefore, the nuclear factors of surgical specimens tended to be worse than those of CNB. As for ER and PgR expression, results of previous studies demonstrated that the absolute concordance of ER and PgR between CNB and surgical specimens were 61.7–99% and 61.5–97.1%, respectively.(1,2,5–7) We demonstrated that the concordance rate of ER and PgR was 94.1–96.0% and 86.1–89.5%, respectively. The concordance of PgR was lower than that of ER, due to the fact that PgR immunoreactivity was weaker and more heterogenous than ER. We examined ER expression by Allred score and proportions of immunoreactive tumor cells were scored as follows: cut-off ≤1%(15) and 10%≤,(14,16) but there were no statistically significant differences. In addition, the concordance rate of HER2 status was 64–96%.(1,2,5–7) Recently, there have been increasing reports evaluating the use of HER2-targeted agents in neoadjuvant therapy for both primary operable and primary inoperable HER-positive breast cancer.(24,25) It is therefore important to achieve a more definitive diagnosis of HER2 status in pre-operative CNB. We also demonstrated that there was discordance in judgments of ER, PgR, and HER2 between CNB specimens and surgically resected specimens in some cases. However, many of these discordant cases were detected more frequently in equivocal or borderline categories (Tables 3,4). In Table 4, it can be seen that four cases of HER2 score 2+ for CNB were changed to 1+ for surgical specimens, and four cases of CNB 3+ were changed to surgical 2+. We detected the strongest HER2 expression area in these tumors by CNB. Therefore, HER2 scores from surgical specimens were lower than those from CNB in these cases. We demonstrated that the disagreements were due to technical problems and intratumoral heterogeneity. If there were discrepancies between two evaluators, the evaluation of the third experienced pathologist was selected. In addition, histopathological staining was performed by a single and experienced technician. Two authors independently evaluated CNB samples and surgical specimens twice on different occasions. Therefore, these findings suggest that interobserver difference, different immunohistochemical technique, and different pre-analytical conditions were not the causes of discordance in the judgment of these factors. It is important for diagnostic accuracy to be established more definitively, and if possible, an increment of the number of CNB specimens may be considered more important because of intratumoral heterogeneity.

Several previous studies have tried to determine the optimum number of specimens to be obtained for ultra sound (US)-guided CNB to accurately diagnose histological subtypes.(26–29) One study demonstrated that among 73 lesions, cells indicating the diagnosis were present in the first specimen in 51 (70%), in the second specimen in 67 (92%), in the third specimen in 70 (96%), and in the fourth specimen in all 73 (100%) of cases.(27) This result suggested that a minimum of four specimens should be obtained with 14-gauge US-guided breast biopsy.(27) However, a study by Melotti *et al.*(29) examined the comparison of the quantity and quality of tissue harvested from breast biopsy when using 14-, 16-, and 18-gauge long-throw needles. The results in that study clearly demonstrated that when comparing 14-, 16-, and 18-gauge needles, accuracy rose with needles of increasing size.(29) These results also suggested that diagnostic accuracy of CNB increased with the increase of harvested specimens.(27,29)

To the best of our knowledge, this is the first study to evaluate the correlation between the number of core biopsies obtained and an accuracy of histological types, nuclear grades, hormone receptors, and HER2 status. Statistically significant differences were detected between patients who received four or more cores and those who had received three or less cores in the discrepancy of the following factors between CNB and subsequent surgical specimens: histological types, nuclear grade, and PgR. In addition, a similar tendency was also detected in ER and HER2 as above, and the cases that received four cores reached to 100% concordance in diagnosis between CNB and surgical specimens. Therefore, the optimal numbers of CNB may be considered four cores, which represent sufficient volume for histopathologic diagnosis. Core biopsy can provide reliable information on histological types, invasion, nuclear grade, hormone receptors, and HER2 status of patients.

Acknowledgments

We appreciate Yayoi Takahashi, MT, for her excellent technical assistance. This work was partly supported by grants from the Ministry of Health, Labour and Welfare of Japan for Research on Intractable Diseases, Risk Analysis Research on Food and Pharmaceuticals, and Development of Multidisciplinary Treatment Algorithm with Biomarkers and Modeling of the Decision-making Process with Artificial Intelligence for Primary Breast Cancer. This work was also partly supported by a Grant-in-Aid for Scientific Research (18390109) from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and the Yasuda Medical Foundation.

References

1 Sutela A, Vanninen R, Sudah M *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 Arnedos M, Nerurkar A, Osin P *et al.* Discodance between core needle biopsy (CNB) and excision biopsy (EB) for estrogen receptor (ER), progesterone receptor (PgR) and HER2 status in early breast cancer (EBC). *Ann Oncol* 2009; **20**: 1948–52.

3 Pettine S, Place R, Babu S *et al.* Stereotactic breast biopsy is accurate, minimally invasive, and cost effective. *Am J Surg* 1996; **171**: 474–6.

4 Britton PD. Fine needle aspiration or core biopsy. *The Breast* 1999; **8**: 1–4.

5 Usami S, Moriya T, Amari M *et al.* Reliability of prognostic factors in breast carcinoma determined by core needle biopsy. *Jpn J Clin Oncol* 2007; **37**: 250–5.

6 Cavaliere A, Sidoni A, Scheibel M *et al.* Biopathologic profile of breast cancer core biopsy: is it always a valid method? *Cancer Lett* 2005; **14**: 22–7.

7 Park SY, Kim KS, Lee TG *et al.* The accuracy of preoperative core biopsy in determining histologic grade, hormone receptors, and human epidermal growth factor receptor 2 status in invasive breast cancer. *Am J Surg* 2009; **197**: 266–9.

8 Dronkers DJ. Stereotaxic core biopsy of breast lesions. *Radiology* 1992; **183**: 631–4.

9 Brenner RJ, Fajardo L, Fisher PR *et al.* Percutaneous core biopsy of the breast: effect of operator experience and number of samples on diagnostic accuracy. *Am J Roentgenol* 1996; **166**: 341–6.

10 Tavassoli FA, Devilee P. *World health organization classification of tumors. Tumor of the breast and females gentia organs.* Lyon: IARC Press, 2003.

11 Rosen PP. *Rosen's Breast Pathology*, 3rd edn. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2009.

12 Tsuda H, Akiyama F, Kurosumi M *et al.* Establishment of histological criteria for high-risk node-negative breast carcinoma for a multi-institutional randomized clinical trial of adjuvant therapy. Japan National Surgical Adjuvant Study of Breast Cancer (NSAS-BC) pathology section. *Jpn J Clin Oncol* 1998; **28**: 486–91.

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 Arihiro K, Umemura S, Kurosumi M *et al.* Comparison of evaluations for hormone receptors in breast carcinoma using two manual and three automated immunohistochemical assays. *Am J Clin Pathol* 2007; **127**: 356–65.

15 Goldhirsch A, Ingle JN, Gelber RD *et al.* Threshholds for therapies: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2009. *Ann Oncol* 2009; **20**: 1319–29.

16 Kurosumi M. Immunohistochemical assessment of hormone receptor status using a new scoring system (J-score) in breast cancer. *Breast Cancer* 2007; **14**: 189–93.

17 Chivukula M, Bhargava R, Brufsky A *et al.* Clinical importance of HER2 immunohistologic heterogenous expression in core-needle biopsies vs resection specimens for equivocal (immunohistochemical score 2+) cases. *Mod Pathol* 2008; **21**: 363–8.

18 Fleiss JL. *Statistical Methods for Rates and Proportions*. New York: John Wiley & Sons, 1981; 212–36.

19 Shannon J, Douglas-Jones AG, Dallimore NS. Conversion to core biopsy in preoperative diagnosis of breast lesions. Is it justified by the results? *J Clin Pathol* 2001; **54**: 762–5.

20 Frank SL, Frank JL, March D *et al.* Does therapeutic touch ease the discomfort or distress of patients undergoing stereotactic core breast biopsy? A randomized clinical trial. *Pain Med* 2007; **8**: 419–24.

21 Somerville P, Seifert P, Destounis SV *et al.* Anticoagulation and bleeding risk after core needle biopsy. *AJR Am J Roentgenol* 2008; **191**: 1194–7.

22 Witt A, Yavuz D, Walchetseder C *et al.* Preoperative core needle biopsy as an independent risk factor for wound infection after breast surgery. *Obstet Gynecol* 2003; **101**: 745–50.

23 Liberman L, Vuolo M, Dershaw DD *et al.* Epithelial displacement after stereotactic 11-gauge directional vacuum-assisted breast biopsy. *AJR Am J Roentgenol* 1999; **172**: 677–81.

24 Buzdar AU, Ibrahim NK, Francis D *et al.* Significantly higher pathologic complete remission rate after neoadjuvant therapy with trastuzumab, paclitaxel, and epirubicin chemotherapy: results of a randomized trial in

doi: 10.1111/j.1349-7006.2010.01630.x

human epidermal growth factor receptor 2–positive operable breast cancer. *J Clin Oncol* 2005; **23**: 3676–85.

25 Baselga J, Semiglazov V, Manikhas GM *et al.* Efficacy of neoadjuvant trastuzumab in patients with inflammatory breast cancer: data from the noah (Neoadjuvant Herceptin) phase iii trial [abstract 2030]. *Eur J Cancer* 2007; **5**: 193.

26 Dagrada GP, Mezzelani A, Alasio L *et al.* HER-2/neu assessment in primary chemotherapy treated breast carcinoma: no evidence of gene profile changing. *Breast Cancer Res Treat* 2003; **80**: 207–14.

27 Fishman JE, Milikowski C, Ramsinghani R *et al.* US-guided core-needle biopsy of the breast: how many specimens are necessary? *Radiology* 2003; **226**: 779–82.

28 Kaplan SS, Racenstein MJ, Wong WS. US-guided core biopsy of the breast with a coaxial system. *Radiology* 1995; **194**: 573–5.

29 Melotti MK, Berg WA. Core needle breast biopsy in patients undergoing anticoagulation therapy: preliminary results. *AJR Am J Roentgenol* 2001; **174**: 245–9.

JJCO Japanese Journal of Clinical Oncology

Jpn J Clin Oncol 2010;40(10)905–912
doi:10.1093/jjco/hyq070
Advance Access Publication 21 May 2010

Original Articles

The Correlation Between Ultrasonographic Findings and Pathologic Features in Breast Disorders

Kentaro Tamaki[1,2,3,*], **Hironobu Sasano**[2], **Takanori Ishida**[1], **Kazuyuki Ishida**[2], **Minoru Miyashita**[1], **Motohiro Takeda**[1], **Masakazu Amari**[1], **Narumi Harada-Shoji**[1], **Masaaki Kawai**[1], **Toshiyuki Hayase**[4], **Nobumitsu Tamaki**[3] and **Noriaki Ohuchi**[1]

[1]Department of Surgical Oncology, Tohoku University Graduate School of Medicine, [2]Department of Pathology, Tohoku University Hospital, Sendai, Miyagi, [3]Department of Breast Surgery, Nahanishi Clinic, Naha, Okinawa and [4]Institute of Fluid Science, Tohoku University, Sendai, Miyagi, Japan

*For reprints and all correspondence: Kentaro Tamaki, Department of Surgical Oncology, Tohoku University Graduate School of Medicine, 1-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi 980-8574, Japan.
E-mail nahanisikenta@yahoo.co.jp

Received January 9, 2010; accepted April 7, 2010

Objective: Breast ultrasonography has gained widespread acceptance as a diagnostic tool for the evaluation of human breast disorders. It is important to evaluate the correlation of ultrasonography findings with the corresponding histopathological features.
Method: We retrospectively reviewed the 154 cases of breast disorders. We evaluated the correlation the ultrasonography findings and carcinoma cells extension with their corresponding histopathological findings. In addition, we also studied the information on estimation of histological types and cancer extension used by the other modalities such as computed tomography and magnetic resonance imaging.
Results: The concordance rate for margins between ultrasonography findings and histopathological features was 91.6% ($P < 0.001$) and that for boundary zone was 87.0% ($P < 0.001$). Histopathological correlation of internal and posterior echoes demonstrated that internal low echo masses were composed of fibroblastic cells with marked collagenization in the stroma, or the cases in which carcinoma cells proliferated in a monotonous, solid and/or expanding manners. Attenuation of posterior echo was detected in the cases associated with hyperplasia of collagenized fibroblastic stroma. An increased cellularity in the mass with prominent large tumor nests and little fibrous stroma demonstrated the accentuation or no alterations of the posterior echo. The concordance rate of borders was 84.4% ($P < 0.001$). The correlation between estimated histological type by ultrasonography diagnosis and actual histological types was 87.0%. An overall detection rate of carcinoma extension by ultrasonography was 86.4%. In addition, an overall detection rate of carcinoma extension by ultrasonography, magnetic resonance imaging and computed tomography was 93.8%.
Conclusion: These results demonstrated correlation between histopathological and ultrasonographic findings of the breast lesions is cardinal for quality control or improving the quality of ultrasonography.

Key words: breast ultrasound – histopathologic findings – carcinoma extension

INTRODUCTION

Breast cancer has become one of the leading causes of death among women. Early clinical detection of breast cancer through screening has led to the detection of the tumor at a relatively earlier clinical stage, which definitely reduced its mortality. The mammographic appearance of breast carcinoma has been well known to vary greatly (1). On the other hand, breast ultrasonography (US) has gained widespread acceptance as a diagnostic tool for the evaluation of breast disorders (2). It is true that some breast diseases that are obscured by dense breast tissue at mammography can be detected with US. US has been in general proposed to serve better in the detection of breast cancer if the patient is young or the masses are small (3,4). Results of many previously published studies have demonstrated the diagnostic benefits in differentiating benign from malignant breast disease in the evaluation using US (5). It was well known that carcinomas are classically described as irregular solid masses with a heterogenous texture and reduced sound transmission in the US, resulting in what is called 'shadowing' behind the lesion (5,6). In addition, a vertical orientation of the lesion is described more often in breast carcinoma in US evaluation (5,6). It is also true that not all carcinomas fulfill these criteria and some do only partially (5). In general, an accurate correlation of US findings with their corresponding histopathologic features is considered most important in US evaluation in this setting.

Breast-conserving surgery is being widely applied in the treatment of early breast cancer. In order to perform conserving surgery, it is very important to detect carcinoma extension and determine the excision area as accurately as possible in the preoperative setting for the benefits of the patients (7,8). Complete removal of a breast tumor with tumor-negative surgical margins is considered most important for avoiding local recurrence in breast-conserving surgery. With high-resolution equipment available, US can detect smaller non-palpable cancers not necessarily detected on high-quality mammography. Excellent visualization of extended intraductal component has been reported using US in some institutions (9,10). However, few have demonstrated the limitation of the US to detect small lesions. Therefore, we attempted to evaluate carcinoma infiltration based on US findings, through revealing histopathologic features of the carcinoma cells infiltration which cannot be detected by US. In addition, in order to overcome these possible limitations, magnetic resonance imaging (MRI) and computed tomography (CT) are being increasingly utilized for the preoperative evaluation of carcinoma extension (11–13). Therefore, we also evaluated the information regarding the detection of cancer infiltration by US in conjunction with MRI and CT.

It is very important to evaluate the correlation of US findings with the corresponding histopathological features. The purpose of this study is therefore, to evaluate the correlation of the US findings including shape, boundary zone, internal and posterior echo, anterior and posterior borders, estimated histological types and carcinoma infiltration with their corresponding histopathological findings of the same lesions. In particular, for internal and posterior echoes, attenuation has been considered to be provided by a highly cellular fibroblastic proliferation (2,14). However, none has ever reported that internal and posterior echo were indeed based on the ratio of intratumoral carcinoma cells and fibroblastic stroma, and histological stromal characteristics of the same lesions. We therefore indicated that anterior and posterior echoes were indeed caused by the ratio of intratumoral carcinoma cells and fibroblastic stroma, and histological stromal characteristics. In addition, some histological types demonstrated low concordance rates between estimated or the histological types estimated by ultrasonographic findings and actual histological types. Therefore, we also discussed this particular discordance between estimated US findings and histologic types, in detail.

Table 1. Histological types of examined cases

Histological types (all)	154
Invasive ductal carcinoma (IDC)	132
Ductal carcinoma in situ (DCIS)	7
Invasive lobular carcinoma (ILC)	10
Mucinous carcinoma	5

PATIENTS AND METHODS

PATIENTS

We retrospectively reviewed the US findings and the histopathologic features of 154 breast lesions for which surgery was performed in Tohoku University Hospital from 1 January 2006 to 31 December 2007 and in which the patients were initially detected by US. The cases treated with neo-adjuvant chemotherapy were excluded from this study of correlating preoperative US findings with histopathlogical analyses. We received informed consents from the patients and the protocol for this study was approved by the Ethics Committee at Tohoku University School of Medicine. The median age of the patients was 57 years (range, 27–85). Of the remaining consecutive 154 patients, 132 were diagnosed histopathologically as invasive ductal carcinomas (IDC), 7 with ductal carcinoma in situ (DCIS), 10 with invasive lobular carcinomas (ILC) and 5 with mucinous carcinomas (Table 1).

US AND HISTOPATHOLOGIC ANALYSES

The US were assessed by one of experienced eight breast surgeons of Tohoku University Hospital. They got the consensus meeting of US for a week to standardize the US exam. In addition, two surgical oncologists independently evaluated the US findings in a retrospective manner, without the knowledge of subsequent histopathological diagnosis. These two investigators were also blinded to the clinical outcome of the patients. The US examination was carried out using the following mechanical scanners: Aloka SSD 3500 (Aloka Co., Tokyo, Japan) with a 10-MHz transducer.

Surgical specimens had been fixed in 10% formaldehyde solution and cut into serial 5-mm thick slices. Histopathological slides in each tumor were reviewed by two pathologists independently without knowledge of the breast US findings. They used Olympus (Tokyo, Japan) BX50 and 20X objectives for the analysis.

Of two or more hardcopy transverse and sagittal plane images of breast lesions, only the largest lesion was analyzed in this study. In the patients with multiple breast lesions, only the largest lesion was evaluated. US findings were subsequently analyzed according to the American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) sonographic classification (2) and the Japan Association of Breast and Thyroid Sonology (JABTS) breast sonographic classification (14). The presence of a mass, margin, boundary zone, internal echoes, posterior echoes and associated findings were each recorded. Histopathological evaluations were based on the Japanese Breast Cancer Society (2008) (15), World Health Organization (WHO) histological classification of tumors of the breast (1) and Rosen's Breast Pathology (16). (i) Margin was tentatively classified into circumscribed or not and also histopathologically classified into these two categories above. (ii) For boundary zone, we analyzed the presence or absence of halo in US. Ultrasonographic 'Halo' corresponded to the histopathologic features in which carcinoma cells invade into fat tissue admixed with adipocytes and elastic fiber (14). We termed the histopathologic feature 'histopathologic halo' and evaluated the existence of the 'histopathologic halo' (Fig. 1). (iii) Internal echo was tentatively classified into low and equal/heterogenous, and posterior echo was tentatively classified into accentuating, no change and attenuating (14). Histopathologic features corresponding to internal and posterior echoic findings were defined by the ratio of carcinoma cells to stroma and the following characteristics related to stromal architecture; collagenization or poor collagenization. We analyzed the intratumoral stroma in five representative fields per case (×200) (Fig. 2). (iv) We analyzed relevant findings about interruption of the anterior and posterior borders of the mammary gland. Interruption of the borders demonstrated extension in adipose tissue, whereas non-interruption demonstrated extension in gland (Fig. 3). (v) We examined the concordance between the estimated and actual histological types. We estimated histological types as followings; IDC, DCIS, ILC and mucinous carcinoma by US without knowledge of histopathological diagnoses.

US, CT and MRI were performed prior to breast-conserving surgery. A 16-row detector CT system (Somatom Sensation Cardiac; Siemens Medical Solution, Erlangen, Germany) was used with CT skin marker, consisting of a paper seal and seven 75-mm non-lead lines with an open window between each line, over the location of the target (13). The breast MRI was obtained using a 1.5 tesla MRI clinical scanner (Magnetom Vision, Siemens, Erlangen, Geramany) (17). The histopathologic diagnosis and the carcinoma extension in all slices were determined by the two pathologists. The surgical margin was defined as positive margin when there were malignant cells at the surgical margin and within 5 mm of the surgical margin. The accurate ratio between the cancer extension detected by the US and the histopathologic cancer extension was evaluated. We also studied the information on detecting cancer extension used by the other modalities such as CT and MRI. In addition, the histopathological characteristics of the cases which could not to be detected by the US were also evaluated. If there were discrepancies of carcinoma extension and estimated histological types among these modalities, we returned to examine the discrepant lesions by US again. When the US findings of the lesions represented desmoplastic change or stromal reaction, we accepted the diagnoses by MRI and CT. On the other hand, when the US findings represented normal variations, we accepted the US diagnoses.

Statistically analysis, such as the one-factor ANOVA and simple regression analysis, were performed using StatMate III for Windows ver. 3.18 (ATMS, Tokyo, Japan). The results were considered significant at $P < 0.05$.

RESULTS

EVALUATION OF THE MARGINS OF THE LESION

Twenty-six out of the 154 were circumscribed masses. Of the 26 circumscribed tumors detected by US, 18 cases

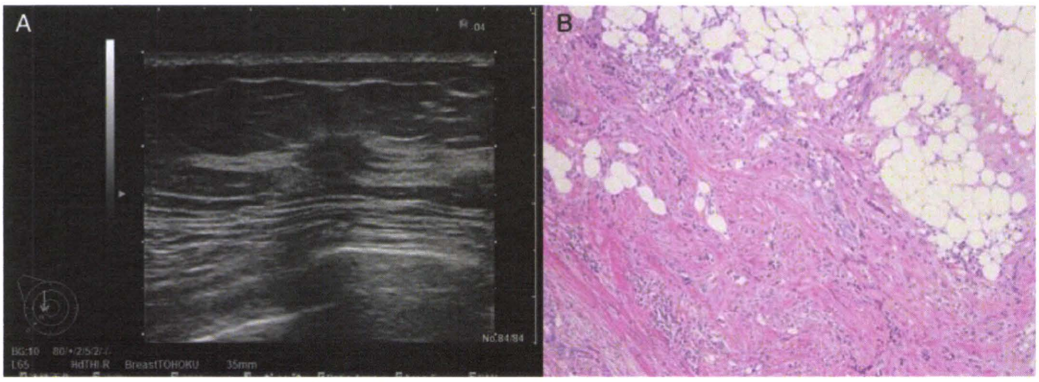


Figure 1. Representative illustrations of 'halo'. (A) 'Halo' of the US finding. (B) The histopathologic feature representing infiltration of carcinoma cells into the surrounding tissues, such as fat tissue and elastic fiber.

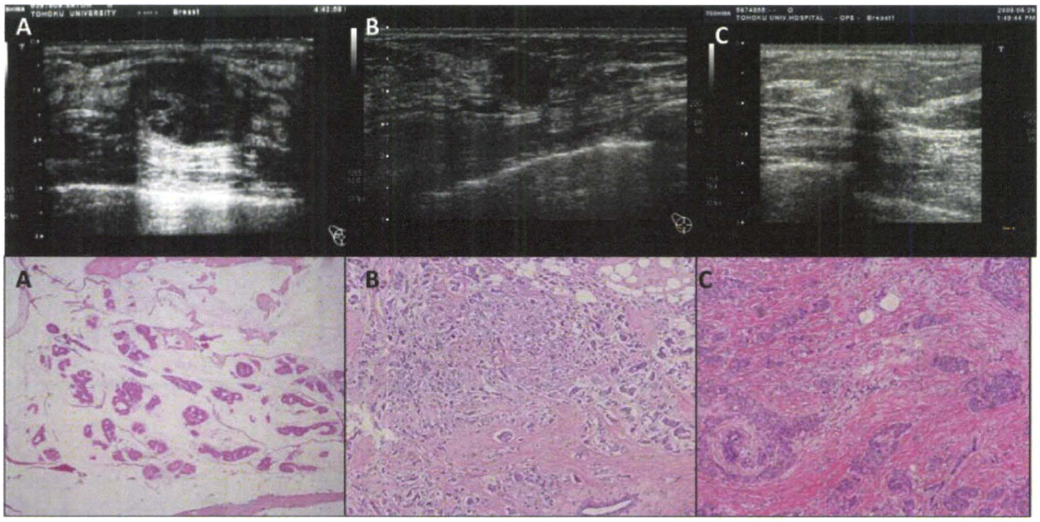


Figure 2. Representative illustrations of internal echoe and posterior echoe. (A) Shows internal echo is heterogenous and posterior echo is accentuating. The histologic type is mucinous carcinoma in which intratumoral structure is heterogenous and constructed by mucin. (B) Shows internal echo is low or hetrogenous and posterior echo is no change. The intratumoral histopathologic feature is heterogenous and poor collagenized stroma. Whereas (C) shows internal echo is low and posterior echo is attenuation. The intratumoral histopathologic feature is heterogenous but the stroma is marked collagenized.

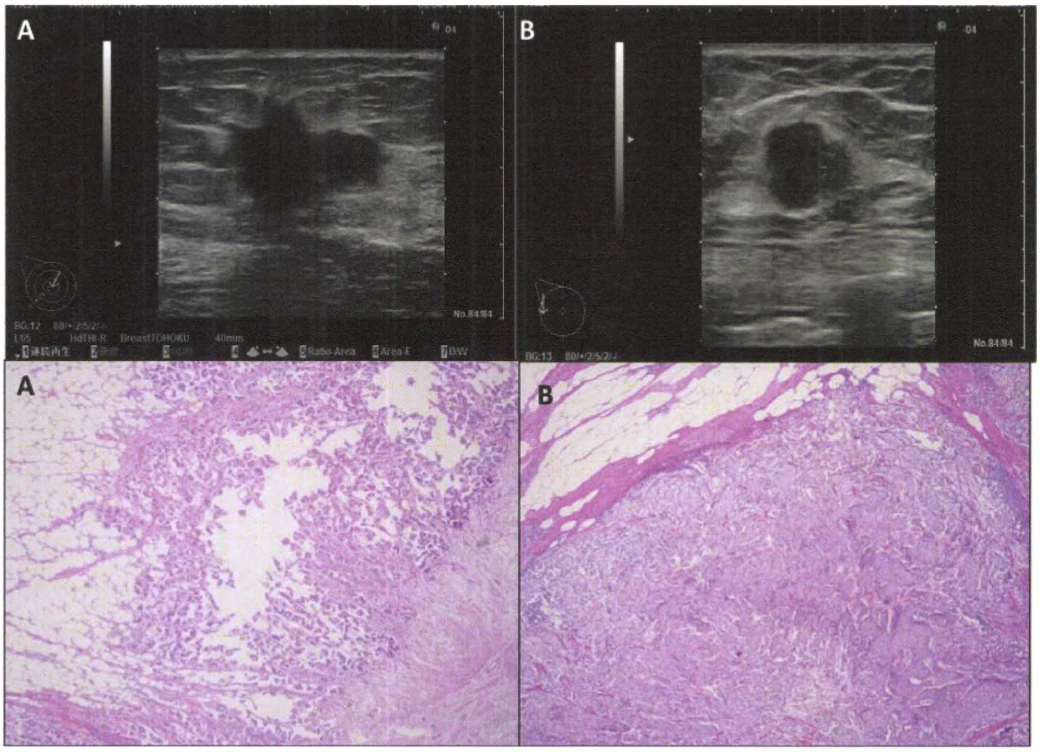


Figure 3. Representative illustrations of interruption and not interruption of the mammary borders. (A) Shows interruption of the anterior border. Histopathologically, carcinoma cells extend to fat. (B) Shows not interruption of the borders. Histopathologically, carcinoma cells extend in the mammary gland.

(69.2%) were also histopathologically circumscribed. Not circumscribed masses were 128 tumors in our present study. One hundred and sixteen out of these 128 tumors (90.6%) were also histopathologically 'not circumscribed'. The rate of concordance between US and histopathological findings was 87.0% ($P < 0.001$).

BOUNDARY ZONE (HALO)

Eighty-nine out of these154 tumors were recognized with halo using US. Seventy-eight out of these 89 tumors with halo were defined as 'histopathologic halo' (Table 2). The rate of concordance was 87.6%. Sixty-five out of the 154 tumors turned out to be the masses without halo. Fifty-six out of the 65 non-halo tumors (86.2%) were also non-histopathologic halo tumors. The rate of concordance for boundary zone between US and histopathologic findings was 87.0% ($P < 0.001$).

Table 2. US findings, the number of cases and the ratio of accuracy between US findings and histopathological features, of margin, boundary zone and associated findings

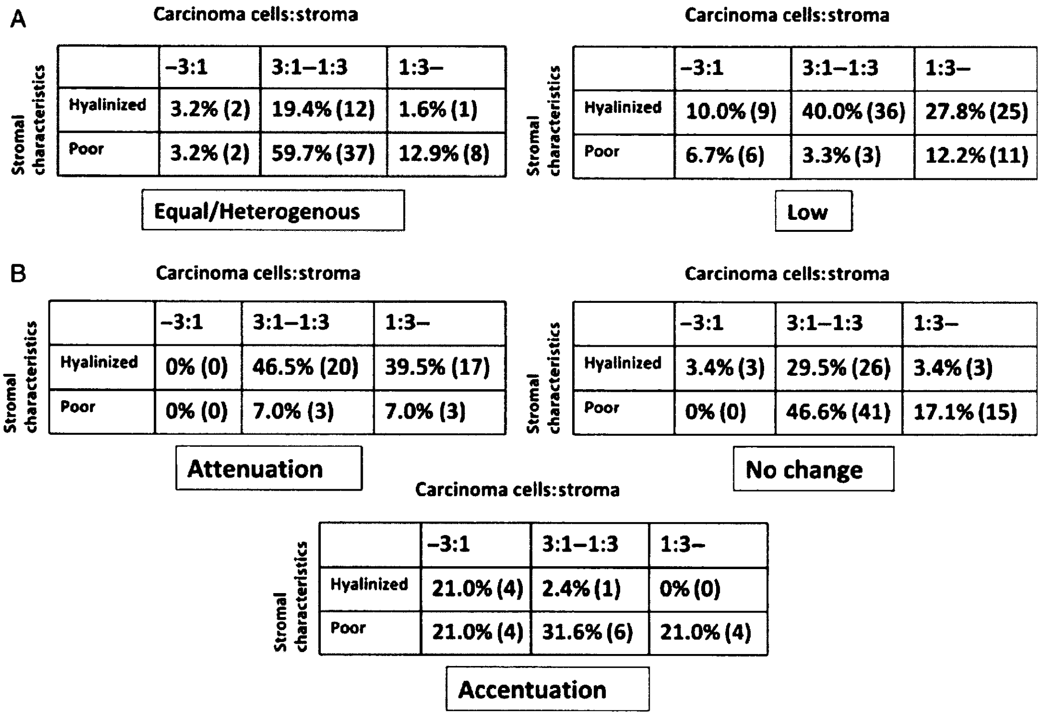
US findings	No. of cases	Rate of concordance (%)
Margin		
Circumscribed	26	69.2
Not circumscribed	128	90.6
Boundary zone (halo)		
halo (+)	89	87.6
halo (−)	65	86.2
Associated findings (Interruption of the mammary borders)		
Interruption	112	83.0
Non-interruption	42	88.1

INTERNAL AND POSTERIOR ECHOES

About 59.7% of the tumors in which internal echoes were equal/heterogeneous histopathologically were associated with poor collagenized stroma and heterogenous intratumoral structure (the ratio of carcinoma cells to stroma was 3:1–1:3) (Fig. 4A). However, tumors associated with low echo levels demonstrated marked collagenized stroma and the higher fibroblastic stromal ratio. As for posterior echo, accentuating tumors histopathologically demonstrated carcinoma cells proliferated in pushing, encapsuled and monotonous fashions, and were also demonstrated in all mucinous carcinoma examined. About 76.1% of the tumors classified as 'no changes' also demonstrated the patterns of marked intratumoral heterogeneity. In addition, ulrtasonographically attenuating cases (43 out of the 153 tumors, 28.1%) were associated with marked collagenized stroma and higher fibroblastic stromal ratio (Fig. 4B).

HISTOPATHOLOGICAL CORRELATIONS WITH OTHER ULTRASONOGRAPHIC FINDINGS (INTERRUPTION OF THE ANTERIOR OR POSTERIOR BORDERS OF THE MAMMARY GLAND)

Interruption of anterior and posterior borders tumors were detected in 112 out of the 153 tumors. Ninety-three out of the 112 tumors (83.0%) were also histopathologically interpreted as extension into adipose tissue. Non-interruption tumors were seen in 42 cases. Thirty-seven out of the



A

Carcinoma cells:stroma

Stromal characteristics	-3:1	3:1–1:3	1:3–
Hyalinized	3.2% (2)	19.4% (12)	1.6% (1)
Poor	3.2% (2)	59.7% (37)	12.9% (8)

Equal/Heterogenous

Carcinoma cells:stroma

Stromal characteristics	-3:1	3:1–1:3	1:3–
Hyalinized	10.0% (9)	40.0% (36)	27.8% (25)
Poor	6.7% (6)	3.3% (3)	12.2% (11)

Low

B

Carcinoma cells:stroma

Stromal characteristics	-3:1	3:1–1:3	1:3–
Hyalinized	0% (0)	46.5% (20)	39.5% (17)
Poor	0% (0)	7.0% (3)	7.0% (3)

Attenuation

Carcinoma cells:stroma

Stromal characteristics	-3:1	3:1–1:3	1:3–
Hyalinized	3.4% (3)	29.5% (26)	3.4% (3)
Poor	0% (0)	46.6% (41)	17.1% (15)

No change

Carcinoma cells:stroma

Stromal characteristics	-3:1	3:1–1:3	1:3–
Hyalinized	21.0% (4)	2.4% (1)	0% (0)
Poor	21.0% (4)	31.6% (6)	21.0% (4)

Accentuation

Figure 4. Analysis of internal echoes and posterior echoes according to the ratio of carcinoma cells to stroma and the stromal characteristics. (A) Is the results of internal echo and (B) is the results of posterior echoes.

42 tumors (88.1%) were histopathologically infiltration in mammary gland or non-invasive carcinomas. The rate of concordance of these borders was 84.4% ($P < 0.001$).

THE CORRELATION BETWEEN FINAL US AND HISTOPATHOLOGICAL DIAGNOSES

The ratio of the correlation between estimated histological types by US diagnosis and histopathological types was 91.6% (141 out of the 154 tumors). The concordance rates between US findings and the following histologic types; IDC, DCIS, ILC and mucinous carcinoma were 98.5% (130 out of the 132 tumors), 14.3% (1 out of the 7 tumors), 60.0% (6 out of the 10 tumors) and 80.0% (4 out of the 5 tumors), respectively (Table 3). US was limited in its ability to detect the lesions with <1 mm in diameter. The concordance rate of combined modalities was 96.1% (148 of 154 tumors) with US, CT with CT skin marker and MRI; IDC, DCIS, ILC and mucinous carcinoma were 99.2% (131 out of the 132 tumors), 57.1% (4 out of the 7 tumors), 80.0% (8 out of the 10 tumors) and 100% (5 out of 5 tumors), respectively.

Table 3. The concordance rates between US diagnosis and the histological types

Histological types	Concordance	Not concordance	Rate of concordance (%)
IDC	130	2	98.5
DCIS	1	6	14.3
ILC	6	4	60.0
Mucinous	4	1	80.0

CORRELATION BETWEEN ULTRASONOGRAPHIC AND HISTOPATHOLOGICAL CARCINOMA EXTENSION

The overall detection rate of carcinoma extension by US was 86.4% (133 out of the 154 tumors). Seventeen out of the 21 tumors (81%) in which US could not detect carcinoma extension demonstrated low-grade intraductal components (Fig. 5A and B), and these lesions were also <1 mm in diameter. Three out of these 21 cases turned out to be invasive ILC in which carcinoma cells invaded with forming single cell pattern (Fig. 5C) and small LCIS extension (Fig. 5D). One was the infiltrated lesion of IDC. This tumor was invaded with forming single cell pattern with poor stromal reaction, similar to ILC. The detection rate of combined modalities was 93.8% (144 of 154 tumors) with US, CT with CT skin marker and MRI.

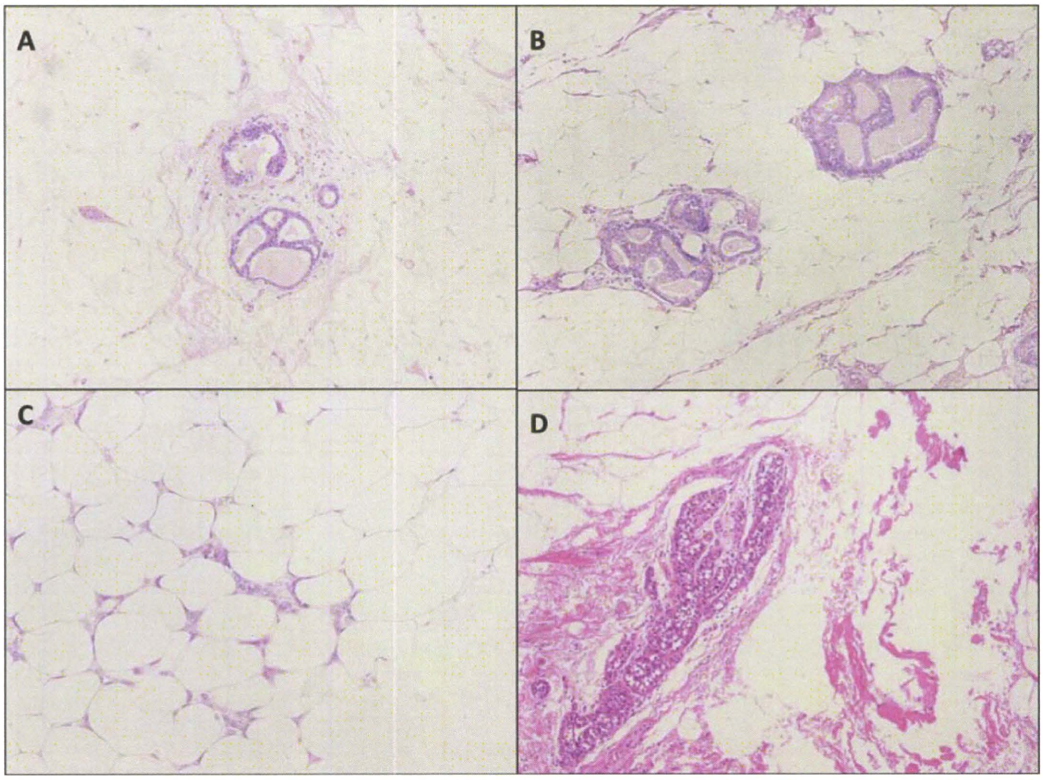


Figure 5. Representative illustrations of the carcinoma extensions which could not be detected by ultrasonography (US). (A) and (B) are low-grade intraductal components. (C) Is invasive lobular carcinoma extension and (D) is LCIS lesion.

DISCUSSION

Previous studies demonstrated that the following ultrasonographic features such as oval or round shape, parallel orientation, circumscribed margins, abrupt interface, enhancement or absence of posterior acoustic features, absence of surrounding tissue alterations represented a benign breast lesion, whereas, irregular shape, non-parallel orientation, echogenic halo, posterior acoustic shadowing and abnormalities of the surrounding tissue regardless of echo pattern were considered to be consistent with a malignant lesion (5,18). It is also true that not all carcinomas fulfill these criteria and some do only partially (5). We therefore examined the details of preoperative US findings and compared these findings to their corresponding histopathological features of the surgical specimens.

It has been well known that circumscribed masses are usually detected in the cases in which carcinoma cells proliferated in both solid and expanded fashion. On the other hand, not circumscribed masses are detected in the cases in which carcinoma cells are arranged in cords, clusters and/or trabeculae, and/or are associated with mixed intraductal component and invasive areas. Halo is generally defined as one of back scattering in US (14). Reflection and scattering are detected and are considered to be caused by the structure which is heterogenous and smaller than sound wave, such as cellular tissue (14). Such opposite sound wave of incident element is generally defined as back scattering (19). Results of our present study demonstrated that halo was indeed characterized by the following histopathologic features, carcinoma cells infiltrated into fat tissue, mixed fat tissue, carcinoma cells and fibroblastic stroma.

Results of previous studies demonstrated that the degree of internal hypoechogenicity determines its sensitivity in predicting malignancy of the lesion (6,20,21). In addition, posterior shadowing has been also reported as one of the important US features suggestive of malignant nature of the lesions (6,20,22). It has been known that shadowing is provided by a highly cellular fibroblastic proliferation (2,14). This is the first study which demonstrated that anterior and posterior echoes were caused by the ratio of intratumoral carcinoma cells and fibroblastic stroma, and histological stromal characteristics. Results of our present study demonstrated that internal echoes and posterior echoes were defined by histopathological intratumoral construction. The acoustic characteristic impedance of the medium becomes larger, reflection subsequently becomes bigger and low internal echo and attenuation of posterior echo finally occur (14). In addition, low internal echo are also detected in the cases histopathologically associated with poor fibroblastic stroma, and carcinoma cells proliferating monotonously, solidly and confertusly. Internal and posterior acoustic shadowing is considered the direct result of US beam attenuation by the desmoplastic reaction to breast cancer (23). Results of our present study of histopathological correlation therefore demonstrated that internal low echoic masses represented the high ratio of fibroblastic stroma and the stroma in these lesions turned out to have marked degrees of interstitial collagenization or the tumors in which carcinoma cells proliferated monotonously, solidly and pushingly. In addition, attenuation of posterior echo was detected in the tumors histopathologically associated with hyperplasia of collagenized fibroblastic stroma. However, it is also true that increased cellularity in the mass with prominent large tumor nests and very little fibrous stroma demonstrated the accentuation or no change of the posterior echo. In addition, poor collagenized fibrous stroma exhibits high acoustic impedance. Histopathological intratumoral construction is important for internal and posterior echoes.

Results of our present study demonstrated that the concordance rate between the US findings and the histopathologic features was 91.6%. According to the comparison between the US findings in details and histopathologic features, we could reasonably postulate histological types from the US findings. Our results demonstrated that DCIS and ILC were lower concordance rates between estimated and actual histological types than the other types. As for DCIS, in some cases, intraductal components were gathered and formed nodule histopathologically. Therefore, the US findings of some DCIS cases were similar to some invasive carcinomas. On the other hand, some IDC tumors such as T1mic which have microinvasion 0.1 cm or less in greatest dimension were similar to DCIS because we proved that US is limited to detect the lesions with <1 mm in diameter. On the other hand, ILC is classified in the following types, classical pattern, solid pattern, alveolar pattern, pleomorphic lobular carcinoma and mixed type carcinoma (1). Therefore, US findings of ILC showed various features and US is limited in its ability to diagnose as ILC correctly. Previous study demonstrated that MR imaging has been shown to be particularly useful in the evaluation of ILC and DCIS (24). Our study also demonstrated the concordance rate of combined modalities was 96.8% (149 of 154 cases) with US, CT with CT skin marker and MRI. Especially, the concordance rate of ILC and DCIS were up to 57.1% (4 out of 7 cases) and 80.0% (8 out of the 10 cases), respectively. It is therefore very important to use a variety of imaging modalities for examining histological types of breast lesions.

Breast conservation therapy has become the treatment standard for the great majority of breast carcinoma (13,25). Several investigations reported the association of higher tumor recurrence rates with positive or close margins than with negative margins following breast-conserving therapy (25,26). It therefore becomes very important to precisely evaluate carcinoma extension preoperatively and determine the excision areas for performing breast-conserving surgery as accurate as possible (13). Results of our study clearly demonstrated the detection rate of carcinoma extension was 86.4% by US. Many tumors in which US could not appreciate carcinoma extension corresponded to DCIS following histopathological evaluations, especially DCIS associated with low-grade malignancy. In such low-grade malignant

DCIS, intraductal components are very small, similar to non-pathologic breast ducts or lobes, the differentiation between low-grade malignant DCIS and non-pathologic breast tissue is generally defined only in cell or nuclear pleomorphism. In addition, such DCIS were rarely detected in stromal desmoplastic reaction. As for ILC, ILC can be insidious and difficult to detect on routine physical examination and/or imaging including US. Histopathologically, ILC is generally characterized by a proliferation of small cells, which lack cohesion and appear individually dispersed through a fibrous connective tissue or arranged in single file linear cords (1). There is often little host reaction or disturbance of the background architectures (1). In addition, this is the first study demonstrated that US is limited to detect the lesions with less than 1 mm in diameter. The results of this study demonstrated that the combined modalities such as US, CT and MRI increase the accuracy of detection from 86.4 to 93.8%. Therefore, it has become very important that a variety of imaging modalities has been used for examining tumor extension and multifocality in breast cancer patients.

CONCLUSION

We clarified the histopathological features of the breast lesions of the tumors in which carcinoma extension could not be preoperatively detected using US. The results of this study demonstrated correlation between histopathological and ultrasonographic findings of the breast lesions is cardinal for quality control or improving the quality of US.

Conflict of interest statement

None declared.

References

1. Tavassoli FA, Devilee P. World Health Organization Classification of Tumors. Tumor of the Breast and Females Genital Organs. Lyon: IARC Press.
2. *American College of Radiology Breast Imaging Reporting and Data System (BI-RADS), Ultrasound*, 4th edn. Reston, VA: ACR. http://www.acr.org/SecondaryMainMenuCategories/quality_safety/BIRADSAtlas/BIRADSAtlasexcerptedtext/BIRADSUltrasoundFirstEdition/ACRBIRADSUSLexiconClassificationFormDoc1.aspx (2 May 2010, date last accessed).
3. Tsunoda HS, Tohno E, Ueno E. Examination of effectiveness of breast cancer detection by modalities and age groups. *J Jpn Assoc Breast Cancer Screen* 1998;7:281–5.
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 to 39 years. *Breast Cancer* 2007;14:229–33.
5. Zonderland HM, Hermans J, Coerkamp EG. Ultrasound variables and their prognostic value in a population of 1103 patients with 272 breast cancers. *Eur Radiol* 2000;10:562–1568.
6. Chao TC, Lo YF, Chen SC, Chen MF. Prospective sonographic study of 3039 breast tumors. *J Ultrasound Med* 1999;18:363–70.
7. Ishida T, Furuta A, Moriya T, Ohuchi N. Pathological assessment of intraductal spread of carcinoma in relation to surgical margin state in breast-conserving surgery. *Jpn J Clin Oncol* 2003;33:61–166.
8. Ohuchi N. Breast-conserving surgery for invasive cancer: a principle based on segmental anatomy. *Tohoku J Exp Med* 1999;108:03–118.
9. Tsunoda-Shimizu H, Ueno E, Tohno E. Ultrasonographic evaluation for breast conservative therapy. *Jpn J Breast Cancer* 1996;11:649–55.
10. Kamio T, Kameoka S, Hamano K, Kimura T. Indication for breast conservative surgery by ultrasonography. *Jpn J Breast Cancer* 1996;11:656–64.
11. Amano G, Yajima M, Moroboshi Y, Kuriya Y, Ohuchi N. MRI accurately depicts underlying DCIS in a patient with Paget's disease of the breast without palpable mass and mammography findings. *Jpn J Clin Oncol* 2005;35:49–153.
12. Hata T, Takahashi H, Watanabe K, Takahashi M, Taguchi K, Itoh T, et al. Magnetic resonance imaging for preoperative evaluation of breast cancer: comparative study with mammography and ultrasonography. *J Am Coll Surg* 2004;198:90–197.
13. Harada-Shoji N, Yamada T, Ishida T, Amari M, Suzuki A, Moriya T, et al. Usefulness of lesion image mapping with multidetector-row helical computed tomography using a dedicated skin marker in breast-conserving surgery. *Eur Radiol* 2009;19:868–74.
14. Japan Association of Breast and Thyroid Sonology. *Guideline for Breast Ultrasound-Management and Diagnosis*, 2nd edn. Tokyo: Japanese 2008.
15. Sakamoto G, Inaji H, Akiyama F, Haga S, Hiraoka M, Inai K, et al. General rules for clinical and pathological recording of breast cancer 2005. *Breast Cancer* 2005;12(Suppl.): S1–27.
16. Rosen PP. *Rosen's Breast Pathology*, 3rd edn. Philadelphia, PA: Lippincott Williams & Wilkins, 2009.
17. Amano G, Ohuchi N, Ishibashi T, Ishida T, Amari M, et al. Correlation of three-dimensional magnetic resonance imaging with precise histopathological map concerning carcinoma extension in the breast. *Breast Cancer Res Treat* 2000;60:3–55.
18. Costantini M, Belli P, Ierardi C, Franceschini G, La Torre G, Bonomo L. Solid breast mass characterization: use of the sonographic BI-RADS classification. *Radiol med* 2007;112:877–94.
19. Koushi N. Basis of Ultrasound Diagnostic System and Technology. Tokyo: Japanese 2009.
20. Leucht WJ, Rade DR, Humbert KD. Diagnostic value of different interpretative criteria in real-time sonography of the breast. *Ultrasound Med Biol* 1988;14:59–73.
21. Jackson VP. Sonography of malignant breast disease. *Semin Ultrasound CT MR* 1989;10:119–31.
22. Harper AP, Kelly-Fry E, Noe JS, Bies JR, Jackson VP. Ultrasound in the evaluation of solid breast masses. *Radiology* 1983;146:731.
23. Kobayashi T. Diagnostic ultrasound in breast cancer: analysis of retro-tumorous echo patterns correlated with sonic attenuation by cancerous connective tissue. *JCU J Clin Ultrasound* 1979;7:471–9.
24. Lee CH. Problem solving MR imaging of the breast. *Radiol Clin N Am* 2004;42:919–34.
25. Singletary SE. Surgical margins in patients with early-stage breast cancer treated with breast conservation therapy. *Am J Surg* 2002;184:383–93.
26. Lindley R, Bulman A, Parsons P, Phillips R, Henry K, Ellis H. Histologic features predictive of an increased risk of early local recurrence of treatment of breast cancer by local tumour excision and radical radiotherapy. *Surgery* 1989;105:13–20.

Breast Cancer
DOI 10.1007/s12282-010-0221-x

ORIGINAL ARTICLE

Educational program and testing using images for the standardization of breast cancer screening by ultrasonography

**Eriko Tohno · Hidehito Takahashi ·
Takayuki Tamada · Yasuhisa Fujimoto ·
Hidemitsu Yasuda · Noriaki Ohuchi**

Received: 24 March 2010 / Accepted: 16 August 2010

Abstract
Background To improve the quality of breast cancer screening by ultrasonography, both effective training and evaluation of the performance of the examiners of breast ultrasound are essential.
Methods The Educational Committee of the Japan Association of Breast and Thyroid Sonology, an NPO, has established 2-day training programs on breast ultrasonography with tests at the end of the programs. The tests are performed using images to evaluate the ability of observers to detect and evaluate lesions on ultrasound. Ability to detect lesions was examined by using videos, and ability to evaluate lesions was examined by using still images. The results of tests taken by 422 physicians and 415 technologists were analyzed.
Results In a comparison between physicians and technologists, the video specificity, the still image sensitivity, and the percentage of category agreement did not show any significant differences. The video sensitivity, the still image specificity, and the percentage of disease name agreement were significantly higher in technologists. Observers who had experienced <100 cases showed significantly poorer results in all subjects except for the video specificity in physicians and the still image specificity in technologists.
Conclusions Ultrasound technologists perform as well as physicians in recognizing and interpreting cancers on breast ultrasound, which supports their role in performing the initial screening examination.

Keywords Breast · Screening · Ultrasonography · Education

Introduction

Cancer screening by biennial mammography is recommended for women aged 40 years or above in Japan. Ultrasonography can detect breast cancers that escape mammography, and the supplemental contribution is particularly high in high-density breasts of young women. In Japan, a large-scale comparative study of whether the concomitant use of ultrasonography is effective for breast cancer screening in women in their 40s is in progress [1].

Ultrasonography is already used for breast cancer screening at many private health care organizations. Quality control is important for breast cancer screening by ultrasonography. While the equipment is an important

E. Tohno (✉)
Graduate School of Comprehensive Sciences,
University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, Japan
e-mail: etohno@sakura.cc.tsukuba.ac.jp

H. Takahashi
Epidemiology and Biostatistics, School of Medicine,
University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, Japan

T. Tamada
School of Medicine, Cluster of Medical Sciences,
University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, Japan

Y. Fujimoto
Tachibana Hospital, 4-3-18 Tachibana, Amagasaki,
Hyogo, Japan

H. Yasuda
Fourth Department of Surgery, International Medical Center
of Japan, 1-21-1 Toyama, Shinjuku, Tokyo, Japan

N. Ohuchi
Division of Surgical Oncology, Graduate School of Medicine,
Tohoku University, 1-1 Seiryouchou Aoba-ku, Sendai,
Miyagi, Japan

component of quality control, the ability of the examiner is even more important, because ultrasonographic detection and diagnosis of the lesions are performed in real-time. To improve the skill of the examiner for the future nationwide introduction of breast cancer screening by ultrasonography, the Educational Committee of the Japan Association of Breast and Thyroid Sonology (JABTS), an NPO, has organized 2-day training programs on breast ultrasonography for physicians and technologists with tests using images to evaluate the ability of the participants at the end of the programs. In this report, the test results were analyzed.

Subjects and methods

Between April 2008 and March 2009, the Educational Committee of the JABTS sponsored training programs on breast ultrasonography 18 times (9 times for physicians, 9 times for technologists). The target population was physicians and technologists engaged, or expected to be engaged, in breast cancer screening by ultrasound and those at hospitals accepting secondary examinations of breast cancer screening. Each training program was performed over 2 days, with 49 physicians or 48 technologists participating at most. The subjects of this study were 422 physicians and 415 technologists. The specialties of the physicians and the numbers practicing each speciality were as follows: breast surgeons (196), surgeons of other or unspecified fields (118), gynecologists (37), radiologists (33), physicians working at screening institutions (21), and physicians of internal medicine (15); data were missing for 2. The technologists included medical technologists and radiographers; both are allowed to perform ultrasound in Japan. The experience levels were divided into four groups according to the numbers of ultrasound examinations performed during the past 5 years: <100, 100–499, 500–999, and ≥1,000. The distribution of the physicians in the different experience levels was 66, 126, 74, and 156, respectively, and that of technologists was 105, 110, 53, and 147, respectively.

Table 1 shows a training program.

Tests using images were performed at the end of the training program using laptop-type personal computers. The observers answered using computers. The contents of the tests were as follows.

Fifty questions using videos (each 15–25 s long)

> Based on the video images that mimicked ultrasonography and that did not pause at the lesion, participants were asked whether a lesion that requires secondary examination (category 3 or more advanced) was present or absent. The questions consisted of 25 videos including lesions of category 3 or greater and 25 videos that included no lesions or included lesions of category 2.

Table 1 Training program for physicians and technologists

- Physicians
  - Lectures for all participants
    - Basic knowledge of breast diseases (20 min)[a]
    - Pathology of breast diseases (40 min)
    - Ultrasonographic tissue characteristics of various breast diseases (40 min)
    - Method of breast ultrasonography (30 min)
    - Terminology of breast ultrasonography (mass image-forming lesions) (30 min)
    - Terminology of breast ultrasonography (non-mass image-forming lesions) (30 min)
    - Criteria for secondary examinations and method to describe findings on screening (30 min)
  - Group training (50 min, each)
    - Mass image-forming diseases, pt. 1
    - Mass image-forming diseases, pt. 2
    - Non-mass image-forming diseases
    - Hands on, pt. 1 (scanning method, description of findings)
    - Hands on, pt. 2 (interventional technique)[a]
    - Skill of finding lesions
    - Ultrasound images of various diseases
  - Tests
    - Tests using images (50 questions using videos, 50 questions using still images; 100 min)
- Technologists
  - Lectures for all participants
    - Basic knowledge of breast diseases (30 min)[a]
    - Pathology of breast diseases (40 min)
    - Ultrasonographic tissue characteristics of various breast diseases (40 min)
    - Method of breast ultrasonography (30 min)
    - Terminology of breast ultrasonography (mass image-forming lesions) (30 min)
    - Terminology of breast ultrasonography (non-mass image-forming lesions) (30 min)
    - Criteria for secondary examinations and method to describe findings on screening (30 min)
  - Group training (50 min)
    - Mass image-forming diseases, pt. 1
    - Mass image-forming diseases, pt. 2
    - Non-mass image-forming diseases
    - Hands on, pt. 1 (scanning method, description of findings)
    - Skill of finding lesions
    - Ultrasound images of various diseases
  - Tests
    - Written test (25 questions in 40 min)[a]
    - Tests using images (50 questions using videos, 50 questions using still images; 100 min)

[a] Differs between physicians and technologists

Fifty questions using still images (each consisting of 1 or 2 orthogonal images)

Based on the still images, the category and disease considered most likely were selected. The questions included 18 sets of malignancies, 7 sets of benign lesions that need further examination, 17 sets of benign lesions that do not need recall (category 2), and 8 sets of normal breasts (including normal variations).

The videos were played using Windows Media Player. The observers were taught how to use the software during group training sessions. The answers (recommended categories and disease names) to questions presented using still images were determined based on a conference involving two doctors with more than 20 years of experience with breast ultrasound and one technologist with more than 10 years of experience. Table 2 defines the categories we use in Japan in comparison with BIRADS categories. Figure 1 and Table 3 show the criteria for category judgment. Table 4 shows the choices of disease names presented in the tests using still images. There were multiple possibly correct answers regarding both the category and disease name in still images, and the answer was regarded as correct if one of them was selected. The duration of tests using videos and those using still images combined was 100 min.

The following six items were evaluated for comparison:

1. Percentage of correct judgments of diseases on video images (video sensitivity)
2. Percentage of correct judgments of non-disease conditions on video images (video specificity)

Table 2 Breast cancer categories and their definitions

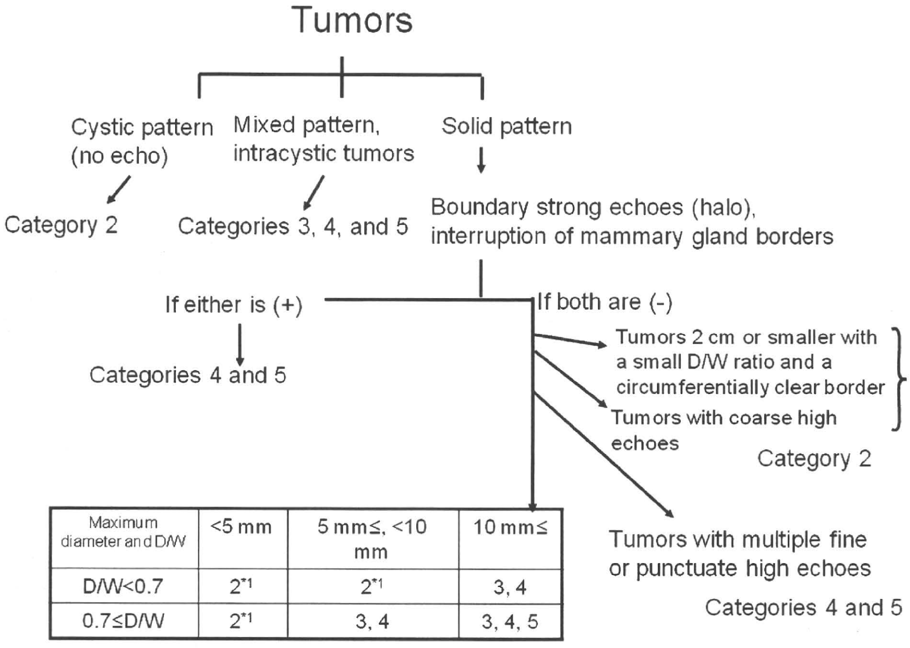
Category 1	No abnormality. Normal variations are included.
Category 2	Benign lesions not requiring secondary examinations on screening.
Category 3 (BIRADS category 3 and 4a)	Lesions that are probably benign but may also be malignant. They require further examinations including cytological or histological examinations or follow-up at hospitals.
Category 4 (BIRADS category 4b and 4c)	Lesions suspected to be malignant.
Category 5	Highly suggestive of malignancy—appropriate action should be taken.

The category classification in Japan is slightly different from that of BIRADS. Screening institutions usually do not follow-up the patient in Japan. If lesions that need either follow-up or further examinations are detected, the patients are referred to hospitals for secondary examinations

3. Percentage of correct judgments of diseases in still images (still image sensitivity)
4. Percentage of correct judgments of non-disease conditions in still images (still image specificity)
5. Percentage of category agreement for still images
6. Percentage of disease agreement for still images

Here, the video sensitivity and specificity are defined, respectively, as answering that a disease is "present" when viewing a video containing a category 3 or more advanced

Fig. 1 Mass image-forming lesions



Maximum diameter and D/W	<5 mm	5 mm≤, <10 mm	10 mm≤
D/W<0.7	2*1	2*1	3, 4
0.7≤D/W	2*1	3, 4	3, 4, 5

*1: Tumors with irregular morphology may be classified as Category 3 or above.

Table 3 Non-mass image-forming lesions

1. Unidirectional or segmental duct ectasia with internal echoes
   (a) If the internal echoes sharply protrude → category 3
   (b) If the internal echoes are broadly based → category 4 or 5
2. Localized or segmental hypoechoic (or isoechoic) areas with unclear borders → category 3, 4
   If punctate high-echo spots suggestive of calcification are observed in the lesion, lesions of higher malignancy must be suspected → category 4 or 5
3. Clustered small cysts with localized or segmental distribution → category 3
4. Architectural distortion
   (a) If its presence is questionable → category 3
   (b) If its presence is definite → category 3, 4
   (c) If its location coincides with the site of prior surgery → category 2

Table 4 Choices of breast diseases

Normal or benign	Malignant
Normal (including normal variation)	Ductal carcinoma in situ
Fibroadenoma (including involuted fibroadenoma)	Scirrhous carcinoma[a]
Phyllodes tumor	Papillotubular carcinoma[a]
Intraductal papilloma	Solid-tubular carcinoma[a]
Cyst (including complicated cyst)	Invasive lobular carcinoma
Adenosis	Mucinous carcinoma
Sclerosing adenosis	Medullary carcinoma
Fibrosis	Squamous cell carcinoma
Duct ectasia	Inflammatory carcinoma
Fat necrosis (including oil cyst)	Malignant lymphoma
Panniculitis	
Mastitis	
Lipoma	
Hamartoma	
Atheroma of the skin	
Silicone granuloma	
Other benign mass	

[a] Subgroups of infiltrating ductal carcinoma no special type (NOS) defined in the "General Rules for Clinical and Pathological Recording of Breast Cancer" by the Japanese Breast Cancer Society [7]

lesion and answering that a disease is "absent" when viewing a video not containing a category 3 or more advanced lesion. The still image sensitivity and specificity were defined, respectively, as the percentage of category 3 or more advanced lesions scored as 3 or higher and the percentage of category 2 or less advanced lesions scored as 2 or lower. Category and disease agreement for still images were defined, respectively, as the percentages of answers that agreed with the recommended categories in all cases and the percentage of answers that agreed with the recommended disease names in all cases. Figure 2 shows an example of still image questions and answers and how the answers were evaluated.

Regarding each of these six items, the significance of the following differences was examined:

1. Difference between physicians and technologists
2. Differences among physicians in various fields of specialty such as breast surgery, other surgery, gynecology, radiology, working at screening institutions, and internal medicine; and between technologists working at a screening facility and those working at a hospital
3. Differences according to the number of patients the observers had examined by breast ultrasonography during the past 5 years (self-reported) (<100, 100–499, 500–999, and ≥1,000)

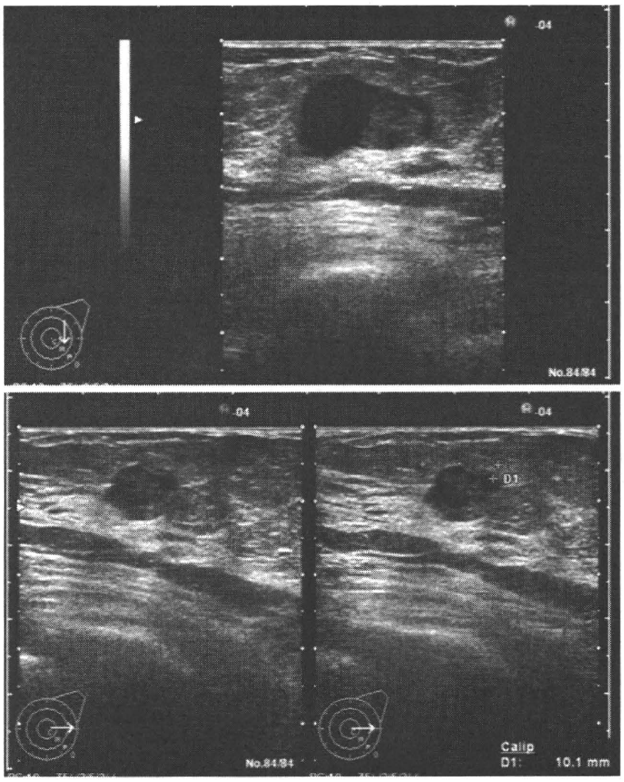


Fig. 2 Example of still image tests. The recommended answers for this image set are category 3 or 4. Although this lesion was actually an intraductal papilloma, the differential diagnoses include intraductal papilloma or DCIS (non-invasive intracystic papillary carcinoma is classified as DCIS in Japan). The answers were evaluated as follows. Category 2, intraductal papilloma: sensitivity, false negative; category agreement, incorrect; disease agreement, correct. Category 3, intraductal papilloma: sensitivity, true positive; category agreement, correct; disease agreement, correct. Category 4, DCIS: sensitivity, true positive; category agreement, correct; disease agreement, correct. Category 5, DCIS: sensitivity, true positive; category agreement, incorrect; disease agreement, correct

Table 5 Comparisons between physicians and technologists

	Physicians	Technologists	p value
Number	422	415	
Video sensitivity (%) (no of loss)	84.0	85.9 (1)	0.0370
Video specificity (%) (no of loss)	79.4	80.3 (1)	0.3469
Still image sensitivity (%)	95.9	95.8	0.7502
Still image specificity (%)	85.1	86.6	0.0260
Category agreement (%)	84.8	85.1	0.6157
Disease agreement (%)	78.4	81.1	<0.0001

4. Differences according to age (20s, 30s, 40s, and 50s or above) of the observers

The differences between two groups such as physicians and technologists were examined by the *t* test, and those among three or more groups were examined by one-way ANOVA and Tukey's multiple comparison procedure. All statistical analyses were performed using SAS 9.13, with a significance level of 5%.

Results

Comparisons between physicians and technologists

The video sensitivity was 84.0% in physicians and 85.9% in technologists, being significantly higher in technologists ($p = 0.037$) (Table 5). The still image specificity was 85.1% in physicians and 86.6% in technologists, being significantly higher in technologists ($p = 0.026$). The percentage of disease name agreement was 78.4% in physicians and 81.1% in technologists, being significantly higher in technologists ($p < 0.0001$).

Comparisons according to physician fields of specialty

Among physicians, breast surgeons and radiologists showed better performance in many measures than doctors from other fields, especially gynecologists (Table 6). No significant difference was noted in the still image sensitivity.

No significant difference was observed between technologists working at a screening facility and those working at a hospital.

Comparisons according to the number of patients the subjects examined by breast ultrasonography during the past 5 years (self-reported)

The video sensitivity improved with increases in the number of patients for both physicians and technologists (Table 7). In physicians, video sensitivity was 76.1% in those who had examined <100 patients and 85.5% in those who had examined 100 or more, with a significant difference ($p < 0.0001$). In technologists, it was 79.8% in those who had examined <100 patients and 87.9% in those who had examined 100 or more, again showing a significant

Table 6 Differences among physicians in various fields of specialty and between technicians working at a screening facility and those working at a hospital

	Physicians ($N = 420$)[a]								Technologists ($N = 415$)		
	Breast surgery 1	Other surgery 2	Gynecology 3	Radiology 4	Screening 5	Internist 6	p value (global)	Multiple comparison ($p < 0.05$)	Screening facility	Hospital	p value
Number	196	118	37	33	21	15			130	285	
Video sensitivity (%)	86.9	81.1	74.9	87.9	84.2	82.7	<0.0001	1-2, 1-3, 3-4	85.6	86.0	0.7499
Video specificity (%)	81.1	77.6	73.2	84.5	77.3	79.5	0.0027	1-3, 3-4	80.1	80.4	0.8394
Still image sensitivity (%)	96.6	95.3	95.0	96.1	95.4	94.1	0.1831		95.1	96.1	0.0690
Still image specificity (%)	87.8	83.2	77.6	86.4	84.2	83.5	<0.0001	1-2, 1-3, 2-3, 3-4	86.5	86.7	0.8803
Category agreement (%)	86.9	83.2	79.3	86.2	83.0	84.7	<0.0001	1-2, 1-3, 3-4	84.6	85.3	0.4144
Disease agreement (%)	80.9	76.3	72.6	81.7	74.4	75.7	<0.0001	1-2, 1-3, 1-5, 3-4	80.9	81.2	0.8257

[a] Speciality was missing in two doctors

Table 7 Differences according to the number of patients the subjects had examined by breast ultrasonography during the past 5 years

	1 <100 patients	2 100–499 patients	3 500–999 patients	4 ≥1,000 patients	*p* value (global)	Multiple comparison ($p < 0.05$)
Physicians ($N = 422$)						
Number	66	126	74	156		
Video sensitivity (%)	76.1	82.4	84.3	88.5	0.0019	1-2, 1-3, 1-4, 2-4
Video specificity (%)	76.2	80.1	78.3	80.7	0.1162	
Still image sensitivity (%)	93.3	95.8	96.9	96.7	<0.0001	1-2, 1-3, 1-4
Still image specificity (%)	81.7	85.0	82.4	88.0	<0.0001	1-4, 2-4, 3-4
Category agreement (%)	81.0	84.2	84.0	87.3	<0.0001	1-2, 1-4, 2-4, 3-4
Disease agreement (%)	73.0	77.7	77.6	81.7	<0.0001	1-2, 1-3, 1-4, 2-4, 3-4
Technologists ($N = 415$)						
Number	105	110	53	147		
Video sensitivity (%)	79.8	87.2	87.2	88.7	<0.0001	1-2, 1-3, 1-4
Video specificity (%)	73.1	81.1	82.6	84.1	<0.0001	1-2, 1-3, 1-4
Still image sensitivity (%)	95.0	95.3	96.5	96.5	0.0773	
Still image specificity (%)	84.1	86.1	87.6	88.4	0.0054	1-4
Category agreement (%)	82.4	84.5	86.5	87.0	<0.0001	1-3, 1-4
Disease agreement (%)	76.6	80.6	83.0	84.0	<0.0001	1-2, 1-3, 1-4, 2-4

difference ($p < 0.0001$). The video specificity was 73.1% in technologists who had examined <100 patients and 82.8% in those who had examined 100 or more, with a significant difference ($p < 0.0001$).

The still image sensitivity was 93.3% in physicians who had examined <100 patients and 96.4% for those who had examined 100 or more, with a significant difference ($p < 0.0001$). The still image specificity was 81.7% in physicians who had examined <100 patients, being significantly lower than the 85.7% in those who had examined 100 or more ($p = 0.0037$) and 88.0% for those who had examined 1,000 or more patients ($p < 0.0001$). The value in those who had examined 1,000 or more patients was significantly higher than the 85.0% in those who had examined 100–499 ($p = 0.0389$) and 82.4% for those who had examined 500–999 patients ($p = 0.0002$). In technologists, the still image specificity was significantly higher at 88.4% in those who had examined 1,000 or more patients than the value in those who had examined <100 ($p = 0.0033$).

The percentage of category agreement was highest in physicians who had examined 1,000 or more patients at 87.3% and was significantly higher than the 81.0% in those who had examined <100 ($p < 0.0001$) and 84.2% in those who had examined 100–499 patients ($p = 0.004$). In technologists, also, the percentage of category agreement was 82.4% in those who had examined <100 patients, being significantly lower than the 85.1% in those who had examined 100 or more patients ($p < 0.0001$), 86.5% in those who had examined 500–999 patients ($p = 0.0084$), and 87.0% in those who had examined 1,000 or more patients ($p < 0.0001$).

The percentage of disease name agreement was 73.0% in physicians who had examined <100 patients, being significantly lower than the 79.4% in those who had examined 100 or more ($p < 0.0001$). It was also lower than the 77.7% in those who had examined 100–499 patients ($p = 0.0102$), 77.6% in those who had examined 500–999 patients ($p = 0.0297$), and 81.7% in those who had examined 1,000 or more patients ($p = 0.0001$). The results were also better in those who had examined 1,000 or more patients than in those who had examined 100–499 patients ($p = 0.0029$).

In technologists, the percentage of disease name agreement was 76.6% in those who had examined <100 patients, being significantly lower than the 81.1% in those who had examined 100 or more ($p < 0.0001$). It was also lower than the 80.6% in those who had examined 100–499 patients ($p = 0.0079$), 83.0% in those who had examined 500–999 patients ($p = 0.003$), and 84.0% in those who had examined 1,000 or more patients ($p < 0.0001$). It was also higher in those who had examined 1,000 or more patients than in those who had examined 100–499 ($p = 0.0201$).

Comparisons according to age

In physicians, the video specificity was 81.8% in those aged 50 years and above, being significantly lower than in those aged <50 years ($p < 0.0001$) (Tables 8, 9). It was also significantly lower than the 82.4% in those in their 30s ($p < 0.0001$) and 81.3% in those in their 40s ($p < 0.0001$). In technologists, video sensitivity and video specificity were not significantly different for those aged 50 years and

Table 8 Differences according to physician or technologist age

	1 20s	2 30s	3 40s	4 50s or above	p value (global)	Multiple comparison ($p < 0.05$)
Physicians ($N = 419$)[a]						
Number	30	118	149	122		
Video sensitivity (%)	81.2	85.3	86.4	80.3	<0.0001	2-4, 3-4
Video specificity (%)	81.9	82.4	81.3	73.2	<0.0001	1-4, 2-4, 3-4
Still image sensitivity (%)	94.4	96.9	96.7	94.4	<0.0001	2-4, 3-4
Still image specificity (%)	84.4	85.0	87.7	82.2	<0.0001	3-4
Category agreement (%)	84.1	85.2	87.0	82.0	<0.0001	2-4, 3-4
Disease agreement (%)	78.0	79.2	80.9	74.6	<0.0001	2-4, 3-4
Technologists ($N = 415$)						
Number	104	151	129	31		
Video sensitivity (%)	83.3 (1)	87.0	87.6	81.9	0.0120	1-3
Video specificity (%)	79.8 (1)	81.3	80.9	74.8	0.1852	
Still image sensitivity (%)	95.8	96.2	95.8	93.5	0.1149	
Still image specificity (%)	87.0	87.7	86.0	83.2	0.1120	
Category agreement (%)	85.5	86.2	84.7	80.4	0.0023	1-4, 2-4, 3-4
Disease agreement (%)	80.8	82.4	80.7	77.5	0.0627	

[a] Age information was missing in three doctors

Table 9 Comparison of differences between physicians and technologists according to age

	Physicians ($N = 419$)[a]			Technologists ($N = 415$)		
	Under 50	50s or above	p value (global)	Under 50	50s or above	p value (global)
Number	297	122		384	31	
Video sensitivity (%)	85.5	80.3	0.0034	86.2	81.9	0.1607
Video specificity (%)	81.8	73.2	<0.0001	80.8	74.8	0.1221
Still image sensitivity (%)	96.5	94.4	0.0020	96.0	93.5	0.1153
Still image specificity (%)	86.3	82.2	0.0003	86.9	83.2	0.1425
Category agreement (%)	86.0	82.0	<0.0001	85.5	80.4	0.0250
Disease agreement (%)	79.9	74.6	<0.0001	81.4	77.5	0.0656

[a] Age information was missing in three doctors

above compared with that in those aged <50 years. This is considered to have been partly due to the small number of subjects.

In physicians, the still image sensitivity was 94.4% in those aged 50 years or above, being significantly lower than the 96.5% in those aged <50 years ($p = 0.0020$). It was also significantly lower than the 96.9% in those in their 30s ($p = 0.0024$) and 96.7% in those in their 40s ($p = 0.0037$). The still image specificity was highest at 87.7% in physicians in their 40s, being significantly higher than the 82.2% in those aged 50 years or above ($p < 0.0001$). In technologists, no significant difference was noted in either still image sensitivity or still image specificity according to age.

The percentage of category agreement using still images was 82.0% in physicians aged 50 years or above, being significantly lower than the 84.8% in those aged <50 years ($p = 0.0001$). It was also significantly lower than the 85.2% in those in their 30s ($p = 0.0061$) and 87.0% in those in their 40s ($p < 0.0001$). In technologists, the percentage of category agreement was 80.4% in those aged 50 years or above, also being significantly lower than the 85.5% in those aged <50 years ($p < 0.0023$). It was also lower than the 85.5% in those in their 20s ($p = 0.0078$), 86.2% in those in their 30s ($p = 0.0012$), and 84.7% in those in their 40s ($p = 0.0339$).

The percentage of disease name agreement using still images was 74.6% in physicians aged 50 years or above, being significantly lower than the 78.4% in those aged <50 years ($p < 0.0001$). It was also lower than the 79.2% in those in their 30s ($p = 0.002$) and 80.9% in those in their 40s ($p < 0.0001$). In technologists, no significant difference was noted in the percentage of disease name agreement according to age.

Discussion

In Japan, both the morbidity and mortality rates of breast cancer are increasing, and breast cancer became the most frequent type of cancer among Japanese women, overtaking stomach cancer in 1993 [2]. The incidence of breast cancer in Japanese women reaches a peak in the late 40s, unlike in Western countries. In Japan, screening for breast cancer has long been performed by palpation, but mammography was introduced after the results of randomized comparative studies in Western countries were reported. Today, breast cancer screening primarily by mammography has become widely available for women aged 40 years or above. However, the sensitivity of mammography screening for breast cancer in women in their 40s is 71.4%, which is lower than the figures in those in their 50s and 60s [3]. This is probably because many women in their 40s present high-density to extremely dense breasts on mammography. To increase the detection rate of breast cancer by screening, the use of MRI for high-risk groups is recommended in the United States [4]. In Japan, however, the concomitant use of ultrasonography, which can be performed at most hospitals, may be more convenient, and a large-scale comparative study is presently being conducted involving women in their 40s [1].

Since breast ultrasonographic diagnosis is performed in real-time, the ability of the examiner to detect lesions, evaluate them, and create appropriate records greatly affects the sensitivity and specificity of breast cancer screening. In Japan, it is stipulated that examination results must be evaluated by physicians but that breast ultrasonography can actually be performed by either a physician or a technologist. In the United States, the results of screening ultrasonography performed by physicians have been reported [5], but screening by technologists is considered to be advantageous from the viewpoints of the number of available examiners and cost. In this study, the video sensitivity was evaluated to assess the subjects' ability to detect breast cancer. Since it was significantly higher in technologists, screening by technologists is not considered to lead to the overlooking of lesions.

A low specificity is a problem with breast ultrasonography, although many lesions other than breast cancer are detected in this way. We previously performed tests using images before and after a 2-day training program and reported that the program was effective, because the sensitivity of screening using video and still images improved significantly after the program despite a slight but non-significant decrease in specificity [6]. In the program, the participants were trained to discriminate changes in the breast that should be eliminated from secondary examinations (clearly benign lesions or very small lesions that are likely to be benign, e.g., normal variations such as a lactating breast and the interposition of fat, surgically enlarged breasts, cysts, and typical fibroadenoma). The results of this study were obtained from participants of a 2-day training program, who had acquired the same knowledge about the assessment of lesions. The results of this study indicate that technologists have an ability to read both still images and videos comparable to physicians and are sufficiently capable of conducting primary screening.

Among physicians of various specialties, gynecologists did poorly in the tests. This may be partly because they are slightly older than other physicians. The average ages of gynecologists and other specialties were 50 and 44 years, respectively. There is also the concept in breast cancer screening that secondary examinations should not be indicated for changes detected by breast ultrasonography unlikely to be cancer, which differs from the guidelines for uterine cancer screening. Gynecologists may not have been accustomed to the stricter elimination policy. There was no difference between technologists working at a screening facility and those working at a hospital. This probably suggests that technologists at a screening facility are also learning to categorize and identify diseases as well as acquiring skills for detecting lesions, but may also have been due to the difficulty in discriminating clearly between screening and hospital technologists, because some hospital technologists are working in screening departments.

In terms of the number of patients the subjects had examined, those who had experienced <100 cases did poorly in the tests. This suggests that learning in a training program is not sufficient and that clinical experience is important to improve the screening ability. This also shows that if an examiner has experience with fewer than 100 patients, the screening ultrasound should be done with the supervision of experienced examiners.

The results of the subjects aged 50 years or above were poor. Fatigue caused by having to watch the computer screen for a long time and the degree of familiarity with computer operation may have been related to this outcome, and considerations such as extending the testing time and allowing the taking of breaks may be necessary for this age group.

Conclusion

We analyzed the results of tests using images after a breast ultrasonography training program. The test results were comparable between physicians and technologists and were poorer for participants with less clinical experience.

Acknowledgments The work was partially funded by Japan Strategic Anti-cancer Randomized Trial (J-START) (a project in the Third-term Comprehensive Strategy for Cancer Control).

References

1. Ohuchi N, Suzuki A, Sakurai Y, Kawai M, Marikawa Y, Narimatsu H, et al. Current status and problems of breast cancer screening. JMAJ. 2009;52:45–9.
2. Cancer Statistics in Japan 2007. http://ganjoho.ncc.go.jp/public/statistics/backnumber/2007_jp.html.
3. Saslow D, Boetes C, Burke W, Harms S, Leach MO, Lehman CD, et al. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. CA Cancer J Clin. 2007;57:75–89.
4. Suzuki A, Kuriyama S, Kawai M, Amari M, Takeda M, Ishida T, et al. Age-specific interval breast cancers in Japan: estimation of the proper sensitivity of screening using a population-based cancer registry. Cancer Sci. 2008;99:2264–7.
5. Berg WA, Blume JD, Cormack JB, Mendelson EB, Lehrer D, Böhm-Vélez M, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. JAMA. 2008;14:2151–63.
6. Tohno E, Sawai K, Shimamoto K, Ueno E, Endou T, Tsunoda-Shimizu H, et al. Establishment of seminars to improve the diagnostic accuracy and effectiveness of breast ultrasound. J Med Ultrasonics. 2006;33:239–44.
7. Histological classification. Breast Cancer. 2005;12 Suppl:12–14.