Breast Cancer Res Treat (2014) 144:391–396
DOI 10.1007/s10549-014-2863-4

EPIDEMIOLOGY

The value of substratified combined imaging assessment with mammography and ultrasonography for Chinese women with palpable breast masses

Jiguang Li · Peng Xing · Liang Feng · Huiting Dong · Feng Jin · Yunfei Wu · Xinyu Zheng

Received: 14 July 2013 / Accepted: 27 January 2014 / Published online: 12 February 2014

Abstract A 5-point breast imaging classification modified from the seven-category Breast Imaging-Reporting and Data System has been applied for mammographic and ultrasonographic examinations in patients with palpable breast masses. The aim of this study was to confirm the value of combined imaging assessment. We included 5,296 cases (3,002 benign and 2,294 cancer) from January 2004 to December 2011. Ultrasonography showed a significantly ($P < 0.01$) higher sensitivity and specificity and lower false-negative rate and false-negative predictive value (false-NPV) than mammography. The sensitivity of combined imaging was significantly ($P < 0.01$) increased and the false-negative rate and false-NPV were significantly ($P < 0.01$) reduced compared to mammography or ultrasonography alone. However, the specificity was significantly ($P < 0.01$) declined for combined imaging versus mammography or ultrasonography alone. Compared with combined imaging assessment, a significant ($P < 0.01$) improvement was noted with substratified scoring, with increased specificity and false-negative rate and decreased sensitivity. In conclusion, the substratified combined imaging score has the potential to provide additional value in the workup of palpable breast lesions.

Keywords Breast neoplasm · Mammography · Ultrasonography · BI-RADS

J. Li (✉) · P. Xing · L. Feng · H. Dong · F. Jin · Y. Wu · X. Zheng
Breast Division, The First Affiliated Hospital of China Medical University, China Medical University Cancer Institute, Shenyang 110001, People's Republic of China
e-mail: jgli@mail.cmu.edu.cn

Abbreviations

MG	Mammography
US	Ultrasonography
False-NPV	False-negative predictive value (1 − NPV)
BI-RADS	Breast Imaging-Reporting and Data System

Introduction

Mammography is currently the principal imaging modality for detection and evaluation of breast diseases. Ultrasonography is used as an adjunct or supplement to mammography for clinically symptomatic patients or for screening populations with dense breast tissue [1, 2]. The Breast Imaging-Reporting and Data System (BI-RADS) was originally developed for mammographic interpretation and later adapted for ultrasonography and other imaging techniques. The first edition of BI-RADS, which was published in 1993, solely addressed the standardization of mammography reporting. The BI-RADS–Ultrasound lexicon was added to the fourth edition of BI-RADS in 2003 [3].

Before 2003, the physicians in our hospital usually conducted a stratified assessment of ultrasonographic features to predict the likelihood of cancer for the patients with palpable breast mass. The assessment system consisted of five categories and may be considered to be a modification of BI-RADS. The first mammography machine was installed in 2003. BI-RADS was used for mammography assessment. In our hospital, different physicians interpreted mammography and ultrasonography examinations. Then, the surgeons would receive two imaging reports (the two scores may be same or different). Most of our symptomatic clinic patients refuse to undergo repeated fine needle aspirations, usually requesting whether

open biopsy or surgical removal of mass is necessary or not. The surgeons give management suggestions to the patients according to the combined assessment score of mammography and ultrasonography and the substratified combined score. We usually called this clinical approach as pre-biopsy/pre-operative combined imaging assessment or dual imaging assessment. To confirm the value of combined imaging assessment, in this study we analyzed the data of our patients with palpable breast masses and undergoing biopsy or surgery.

Patients and methods

Patient data were retrieved from the database of the Breast Division of The First Hospital of China Medical University and China Medical University Cancer Institute. From January 2004 to December 2011, a total of 9,515 female patients underwent breast biopsy or surgical treatment in our hospital. Of these, 5,296 patients with palpable breast masses underwent both mammography and ultrasonography before treatment. This subset of patients, comprising 3,002 benign cases and 2,294 cancer cases, was included for analysis. Patients with nipple discharge without palpable mass, Paget's disease confined to the nipple, breast lymphoma, or sarcoma were excluded. If the patients received neoadjuvant chemotherapy, pre-chemotherapy images were chosen for analysis. If the patients underwent more than one imaging examinations before biopsy or surgical treatment, the latest was analyzed. In cases with multiple lesions in one breast, images of the largest lesion were analyzed. Bilateral breast lesions were recorded as two independent cases.

In our hospital, the first mammography machine, GE Senography 2000D with Full-field digital mammography (Horten, Norway), was installed in 2003. A second mammography machine, Siemens Mammomat inspiration with Full-field digital mammography (Erlangen, Germany), was installed in 2010. Standard craniocaudal and mediolateral oblique images were obtained with further magnification or additional views taken as required. In the early period of this study, the five categories of BI-RADS were used, and the seven categories (including 1, 2, 3, 4a, 4b, 4c, and 5) were used after their introduction. Category 0 was not reported, which was usually assessed as Category 1 or 3. All ultrasonography examinations included real-time bilateral whole-breast and power Doppler blood flow scans, using Toshiba Xario (Tochigi, Japan), Aplio XV (Tokyo, Japan), Aplio80 (Tokyo, Japan), Aloka 10 (Tokyo, Japan), or Hitachi Preirus (Tokyo, Japan) with 7–14 mHz probes. Sonoelastography was not conducted in routine examinations. Each ultrasonography examination was assigned a score from 1 to 5.

Table 1 Comparison of our institute's 5-point breast imaging classification with UK (RCR) 5-point category and BI-RADS

Our institute's 5-point category		UK (RCR) 5-point category [5, 6]	
Score and definition	Equivalent to BI-RADS	Score and definition	Equivalent to BI-RADS
1. Normal without abnormal findings	1	Normal	1, 2
2. Benign abnormal findings	2	Benign	3
3. Indeterminate or uncertain	3, 4a	Indeterminate/ probably Benign findings	4a, 4b
4. Suspicious of malignancy	4b, 4c	Finding suspicious of malignancy	4c
5. Highly suspicious of malignancy	5	Finding highly suspicious of malignancy	5

Our 5-point scoring classification followed similar definitions to the BI-RADS, the European Society of Mastology Breast Radiology Diagnostic classification (EUSOMA) [4], and the Australian National Breast Cancer Center (NBCC) classification [5]. The major difference was in the definition of Category 3, changing from "probably benign" in BI-RADS, "indeterminate" in EUSOMA, "indeterminate/equivocal" in NBCC to "indeterminate or uncertain" in our classification system. Another 5-point scoring classification for mammography and ultrasonography has been suggested by the Royal College of Radiologists (RCR) Breast Group of UK [5, 6]. Comparisons of our and RCR 5-point scoring classifications to the BI-RADS 7-point system are shown in Table 1.

In practice, mammographic Categories 3 and 4a were considered as Category 3 and mammographic Categories 4b and 4c as Category 4. Therefore, both mammography and ultrasonography assessments had the same five categories. The higher of the two category numbers for mammography and ultrasonography was assigned as the combined imaging score. For instance, in the case of a breast lesion with Category 3 by ultrasonography but Category 4b or 4c by mammography, the combined imaging assessment would be Category 4. Category 3, 4, and 5 for combined imaging assessments may be further substratified according to different mammography and ultrasonography scores. But usually substratified combined Category 3 was applied: Category 3a (mammography Category 3 and ultrasonography Category 1 or 2), Category 3b (ultrasonography Category 3 and mammography Category 1 or 2), and Category 3c (mammography Category 3 and ultrasonography Category 3).

Table 2 The cancer incidence for each group of categories by modalities in patients with palpable masses (2,294/5,296 = 43.3 % with malignancy)

Assessment categories	No. of malignant	No. of benign	Total patients	Malignancy (%)
Mammography alone[a]				
1. Normal without abnormal findings	53	699	752	7.0
2. Benign abnormal findings	35	479	514	6.8
3. Indeterminate or uncertain	400	1,427	1,827	21.9
4. Suspicious of malignancy	793	375	1,168	67.9
5. Highly suspicious of malignancy	1,013	22	1,035	97.9
Total	2,294	3,002	5,296	
US alone				
1. Normal without abnormal findings	5	45	50	10.0
2. Benign abnormal findings	43	1,446	1,489	2.9
3. Indeterminate or uncertain	158	1,092	1,250	12.6
4. Suspicious of malignancy	887	393	1,280	69.3
5. Highly suspicious of malignancy	1,201	26	1,227	97.9
Total	2,294	3,002	5,296	
Combined assessment with MG and US[b]				
1. Normal without abnormal findings	0	13	13	0
2. Benign abnormal findings	13	732	745	1.7
3. Indeterminate or uncertain	95	1,547	1,642	5.8
4. Suspicious of malignancy	681	665	1,346	50.6
5. Highly suspicious of malignancy	1,505	45	1,550	97.1
Total	2,294	3,002	5,296	
Substratified combined assessment				
Category 1	0	13	13	0
Category 2	13	732	745	1.7
Category 3				
3a. MG 3 + US 1 or 2	19	633	652	2.9
3b. US 3 + MG 1 or 2	12	311	323	3.7
3c. MG 3 + US 3	64	603	667	9.6
Category 4				
4a. MG 4 + US 1, 2, or 3	65	281	346	18.8
4b. US 4 + MG 1, 2, or 3	250	299	549	45.5
4c. MG 4 + US 4	366	85	451	81.2
Category 5				
5a. MG 5 + US 1 to 4	304	19	323	94.1
5b. US 5 + MG 1 to 4	492	23	515	95.5
5c. MG 5 + US 5	709	3	712	99.6
Total	2,294	3,002	5,296	

[a] Original BI-RADS categories have been modified to our 5-point breast imaging assessment classification

[b] All assessments were made with a combination of mammography (MG) and ultrasonography (US). The category was determined by the higher score from MG or US

Overall, we evaluated four types of imaging classification: (1) mammography alone, (2) ultrasonography alone, (3) combined imaging with mammography and ultrasonography, and (4) substratified combined imaging. The sensitivity and specificity for each type were calculated. False-negative rate (cancer cases with negative imaging assessment/total cancer cases) and false-negative predictive value (false-NPV; cancer cases/all cases with negative imaging assessment) were also estimated. For mammography alone, ultrasonography alone, and combined imaging assessment, Categories 1 and 2 were regarded as negative for cancer, and 3, 4, and 5 as positive for cancer. For substratified combined imaging, Categories 1, 2, 3a, and 3b were regarded as negative for cancer, and

Table 3 Cases with false-negative results

	2	3a	3b	Total
Bilateral breast cancer	1	1	3	5
Papilloma with malignant change	4			4
Intracystic tumor		3		3
Dense breasts	8	15	9	32
Total	13	19	12	44

Table 4 Comparison of results by different assessment methods

Assessment method	Sensitivity (%)	Specificity (%)	False-negative rate (%)	False-NPV (%)
US alone[a]	97.9	49.7	2.1	3.1
Mammography alone[b]	96.2	39.2	3.8	7.0
Conventional combined MG and US[c]	99.4	24.8	0.6	1.7
Substratified combined MG and US[d]	98.1	56.3	1.9	2.5

[a] Comparison between ultrasonography (US) alone and mammography (MG) alone, $P < 0.01$ for all comparisons

[b] Comparison between MG alone and conventional combination of MG and US, $P < 0.01$ for all comparisons

[c] Comparison between US alone and conventional combination of MG and US, $P < 0.01$ for all comparisons

[d] Comparison between combination of MG and US and substratified combined imaging assessment, $P < 0.01$ for all comparisons

3c, 4, and 5 as positive for cancer. Differences in the sensitivity, specificity, false-negative rate, and false-NPV were measured using Chi square test. A P value less than 0.05 was considered significant.

Results

The incidences of cancer detection using different types of imaging classification are shown in Table 2. Of the 5,296 patients, 13 with combined Category 1 underwent biopsy or surgery. Indications included mild nipple discharge, prophylactic breast excision, and removal of palpated "mass" or glandular tissues according to the patient's request. 745 patients with combined Category 2 underwent biopsy or surgery, because of both patients and surgeons preferred to remove a palpable breast mass if the patient's age was over 35, or the mass was larger than 2 cm. Among them, 13 cases (1.7 %) had a cancer diagnosis. For the patients with combined Category 3a, 3b, and 3c, the incidence of cancer detection was 2.9, 3.7, and 9.6 %, respectively.

Substratified combined categories 2, 3a, and 3b represented false-negative results in patients diagnosed with cancer. These cases showed dense breasts, bilateral synchronous breast cancer, papilloma with malignant change, and intracystic cancer (Table 3).

The sensitivity (97.9 vs. 96.2 %, $P < 0.01$) and specificity (49.7 vs. 39.2 %, $P < 0.01$) of ultrasonography were significantly higher than that of mammography, while the false-negative rate (2.1 vs. 3.8 %, $P < 0.01$) and the false-NPV (3.1 vs. 7.0 %, $P < 0.01$) were significantly reduced. Moreover, the sensitivity of combined imaging was significantly greater than that of mammography (99.4 vs. 96.2 %, $P < 0.01$) or ultrasonography alone (99.4 vs. 97.9 %, $P < 0.01$). The false-negative rate and the false-NPV of combined imaging were significantly lower than those of mammography ($P < 0.01$) or ultrasonography alone ($P < 0.01$). In contrast, the specificity of the combined imaging was significantly lowered compared with mammography (24.8 vs. 39.2 %, $P < 0.01$) or ultrasonography alone (24.8 vs. 49.7 %, $P < 0.01$). Compared with the combined imaging assessment, significant improvement was noted with the substratified classification, with a significant increase in the specificity (56.3 vs. 24.8 %, $P < 0.01$) and decrease in the sensitivity (98.1 vs. 99.4 %, $P < 0.01$). The false-negative rate of the substratified classification was significantly increased from 0.6 to 1.9 % ($P < 0.01$). The false-NPV was also increased, but no statistical significance was observed (2.5 vs. 1.7 %, $P = 0.26$) (Table 4).

Discussion

The BI-RADS initially consisted of categories 1–5. Since patients with Category 4 lesions have a widely varying risk of cancer, ranging from 2 to 95 %, Category 4 has been subdivided into 4a, 4b, and 4c in the latest (fourth) edition of BI-RADS [7]. It has been reported that the subcategories 4a, 4b, and 4c are beneficial in stratifying lesions by likelihood of malignancy [8]. In our hospital, the seven categories of BI-RADS were applied for mammography and the five categories with a modification of BI-RADS for ultrasonography. For the combined imaging modality, the five categories with modified definitions were preferred. Category 4a of mammography was considered as Category 3 in the five-category classification, thus narrowing the range of cancer likelihood of Category 4 and the biopsy suggestion would be more definite. Complicated management would be focused in the subsets of Category 3.

To the best of our knowledge, the most published studies of mammography and ultrasonography are not double-blind examinations. As ultrasonography is usually performed as an adjunct to mammography, ultrasonography assessment

may be affected by prior mammography reporting [2, 9]. Chan et al. [4] reported that the sensitivity of ultrasonography was 91 %, which was higher than that of mammography (78 %). Ultrasonography and mammography were performed and evaluated independently in a large multicenter trial comparing mammography and ultrasonography as screening tools [10]. But this study involved a population at a high risk for breast cancer. In our hospital, mammographers and sonographers give interpretations of imaging findings independently, which allows double-blind comparisons of mammography, ultrasonography, and combined modalities. The diagnostic accuracy of ultrasonography seems to be higher than that of mammography. Such improvement in diagnosis may be partially ascribed to the introduction of Power Doppler blood flow examination in our patients [11, 12].

Our data revealed that the cancer incidence for Categories 1–5 using mammography was 7.0, 6.8, 21.9, 67.9, and 97.9 %, respectively, and using the ultrasonography, the incidence was 10.0, 2.9, 12.6, 69.3, and 97.9 %, respectively. It is quite difficult to give management suggestions to patients with Category 1, 2, or 3 according to either mammography or ultrasonography reporting. Because of the false-negative results of mammography or ultrasonography alone [4, 13], it could not exclude the possibility of malignancy in patients with negative findings on one imaging examination.

Although the seven-category BI-RADS is applied primarily in mammography assessment and our five-category classification in ultrasonography assessment, the correlations between the two assessment scores have been well-defined. It is ready to generate the combined imaging score and the final substratified combined imaging score from the two imaging reports. We found that combined imaging assessment with mammography and ultrasonography increased the sensitivity to 99.4 % and decreased the false-negative rate to 0.6 % and the false-NPV to 1.7 %. But in this protocol, all the patients with Category 3 lesions on the combined imaging were suggested to biopsy, and the specificity was only 24.8 %. Therefore, it is necessary and valuable to substratify combined Category 3 into 3a, 3b, and 3c. Patients with combined Categories 1 and 2 would have no indication for biopsy, and patients with combined Category 3a and 3b would be assigned to short-term follow-up. When Categories 3c, 4, and 5 of the substratified combined imaging assessment were considered as positive for cancer, the specificity significantly increased from 24.8 to 56.3 %, while the sensitivity altered from 99.4 to 98.1 %.

Although Categories 2, 3a, and 3b were regarded as negative for cancer, we prefer to remove a palpable mass in the selected patients with these negative imaging, and found 44 cases diagnosed as cancer. The false-negative rate was as high as 1.9 % and the false-NPV was 2.5 %. These false-negative results were related to mammographic dense breasts, bilateral breast cancer, papilloma with malignant change, and intracystic cancer. Besides imaging score, imaging features and clinical assessment, including mammographic dense breasts, imaging suspected as a papilloma or intracystic tumor, age, mass size, physician's palpation, history of breast cancer, should be considered when physicians give suggestions to the patients with negative imaging. This management may avoid delayed breast cancer diagnosis in the patients with Categories 2, 3a, and 3b.

The main limitation of this study is patient selection bias. As we selected the patients at a high risk of cancer for biopsy, the percentage likelihood of cancer in the patients without undergoing biopsy may be lower than that in the patients with undergoing biopsy. In addition, there is lack of follow-up of the patients who did not undergo biopsy, especially those with combined Categories 1, 2, and 3a and 3b.

In summary, ultrasonography has a better diagnostic value for cancer detection in patients with palpable breast masses than mammography. However, neither mammography nor ultrasonography alone would be satisfactory because of the high false-negative results and the low specificity. It is thus recommended that mammography and ultrasonography should be used simultaneously for all patients with palpable breast masses. Combined imaging assessment increases the sensitivity but decreases the specificity of cancer detection. The substratified combined imaging score is valuable due to increased specificity and avoiding unnecessary biopsies.

References

1. McCavert M, O'Donnell ME, Aroori S, Badger SA, Sharif MA, Crothers JG, Spence RA (2009) Ultrasound is a useful adjunct to mammography in the assessment of breast tumours in all patients. Int J Clin Pract 63:1589–1594. doi:10.1111/j.1742-1241
2. Weigert J, Steenbergen S (2012) The Connecticut experiment: the role of ultrasound in the screening of women with dense breasts. Breast J 18:517–522. doi:10.1111/tbj.12003
3. Burnside ES, Sickles EA, Bassett LW, Rubin DL, Lee CH, Ikeda DM, Mendelson EB, Wilcox PA, Butler PF, D'Orsi CJ (2009) The ACR BI-RADS experience: learning from history. J Am Coll Radiol 6:851–860. doi:10.1016/j.jacr.2009.07.023
4. Chan SW, Cheung PS, Chan S, Lau SS, Wong TT, Ma M, Wong A, Law YC (2008) Benefit of ultrasonography in the detection of clinically and mammographically occult breast cancer. World J Surg 32:2593–2598. doi:10.1007/s00268-007-9273-2
5. Maxwell AJ, Ridley NT, Rubin G, Wallis MG, Gilbert FJ, Michell MJ, Royal College of Radiologists Breast Group (2009) The Royal College of Radiologists Breast Group breast imaging classification. Clin Radiol 64:624–627. doi:10.1016/j.crad.2009.01.010
6. Taylor K, Britton P, O'Keeffe S, Wallis MG (2011) Quantification of the UK 5-point breast imaging classification and mapping to BI-RADS to facilitate comparison with international literature. Br J Radiol 84:1005–1010. doi:10.1259/bjr/48490964

7. American College of Radiology (2003) Breast imaging reporting and data system, breast imaging atlas, 4th edn. American College of Radiology, Reston, VA
8. Lazarus E, Mainiero MB, Schepps B, Koelliker SL, Livingston LS (2006) BI-RADS lexicon for US and mammography: inter observer variability and positive predictive value. Radiology 239:385–391. doi:10.1148/radiol.2392042127
9. Kim EK, Ko KH, Oh KK, Kwak JY, You JK, Kim MJ, Park BW (2008) Clinical application of the BI-RADS final assessment to breast ultrasonography in conjunction with mammography. AJR 190:1209–1215. doi:10.2214/AJR.07.3259
10. Berg WA, Blume JD, Cormack JB, Mendelson EB, Lehrer D, Böhm-Vélez M, Pisano ED, Jong RA, Evans WP, Morton MJ, Mahoney MC, Larsen LH, Barr RG, Farria DM, Marques HS, Boparai K, ACRIN 6666 Investigators (2008) Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. JAMA 299:2151–2163. doi:10.1001/jama.299.18.2151
11. Kwak JY, Kim EK, Kim MJ, Choi SH, Son E, Oh KK (2008) Power Doppler ultrasonography: evaluation of solid breast lesions and correlation with lymph node metastasis. Clin Imaging 32:167–171. doi:10.1016/j.clinimag.2007.12.004
12. Tozaki M, Fukuma E (2011) Does power Doppler ultrasonography improve the BI-RADS category assessment and diagnostic accuracy of solid breast lesions? Acta Radiol 52:706–710. doi:10.1258/ar.2011.110039
13. Murphy IG, Dillon MF, Doherty AO, McDermott EW, Kelly G, O'Higgins N, Hill AD (2007) Analysis of patients with false negative mammography and symptomatic breast carcinoma. J Surg Oncol 96:457–463. doi:10.1002/jso.20801