

Interobserver agreement on the interpretation of automated whole breast ultrasonography

ULTRA SONO GRAPHY

Eun Jeong Kim, Sung Hun Kim, Bong Joo Kang, Yun Ju Kim

Department of Radiology, Seoul St. Mary's Hospital, The Catholic University of Korea College of Medicine, Seoul, Korea

ORIGINAL ARTICLE

<http://dx.doi.org/10.14366/usg.14015>
pISSN: 2288-5919 • eISSN: 2288-5943
Ultrasonography 2014;33:252-258

Purpose: The purpose of this study was to prospectively evaluate the interobserver agreement on lesion characterization and the final assessment of automated whole breast ultrasonography (ABUS) images.

Methods: Between March and August 2012, 172 women underwent bilateral ABUS before biopsy guided by handheld ultrasonography (HHUS) and mammography. A total of 206 breast lesions were confirmed histopathologically by biopsy. Three-dimensional volume data from ABUS scans were analyzed by two radiologists without the knowledge of HHUS results or patient clinical information. The two readers described the type, shape, orientation, margin, echogenicity, posterior acoustic features, and categorization of the final assessment of detected breast lesions. Kappa statistics were used to analyze the described characteristics of the breast lesions detected by both of the two readers.

Results: Of the 206 histopathologically confirmed lesions, reader 1 detected 166 lesions and reader 2 detected 150 lesions. A total of 145 lesions were detected by both readers using ABUS images. There was substantial agreement on shape ($\kappa=0.707$), and moderate agreement on type, margin, mass orientation, echogenicity, and posterior acoustic features ($\kappa=0.592, 0.438, 0.472, 0.524, \text{ and } 0.541$, respectively). Breast Imaging Reporting and Data System final assessment values yielded a kappa value of 0.3971 when category subdivisions 4A, 4B, and 4C were included. With respect to the C2, C3, C4, and C5 categories, the interobserver agreement was moderate ($\kappa=0.505$).

Conclusion: ABUS is a promising diagnostic tool with a good interobserver agreement, comparable to that of HHUS.

Keywords: Breast; Ultrasonography; Observer variation

Received: March 7, 2014
Revised: April 20, 2014
Accepted: April 21, 2014

Correspondence to:

Sung Hun Kim, MD, Department of Radiology, Seoul St. Mary's Hospital, The Catholic University of Korea College of Medicine, 222 Banpo-daero, Seocho-gu, Seoul 137-701, Korea
Tel. +82-2-2258-1452
Fax. +82-2-599-6771
E-mail: rad-ksh@catholic.ac.kr

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2014 Korean Society of Ultrasound in Medicine (KSUM)

Introduction

Breast ultrasonography is a well-established diagnostic tool with mammography for evaluating breast abnormalities [1]. However, the modality has several drawbacks. Because of operator dependency, the skill and knowledge of the operator affect the diagnostic accuracy [2–5]. Additionally, poor standardization and reproducibility of breast ultrasonography reduce the diagnostic yield [6,7].



How to cite this article:

Kim EJ, Kim SH, Kang BJ, Kim YJ. Interobserver agreement on the interpretation of automated whole breast ultrasonography. Ultrasonography. 2014 Oct;33(4):252-258.

To improve upon these problems, a new approach, the automated breast ultrasonography, has been developed [8,9]. Automated whole breast ultrasonography (ABUS) scans the entire breast in a standardized manner and sends all resulting images to a workstation [10]. Radiologists are then able to interpret breast lesions efficiently and within a short time frame [11].

ABUS is expected to minimize operator dependency [10,11], but another important point to be considered is consistent characterization and interpretation of lesions. Consistent reporting with reproducible descriptions of lesion location, size, and features is required. The purpose of this prospective study was to evaluate the interobserver agreement on lesion characterization and final assessment when reviewing ABUS images.

Materials and Methods

Patients

The institutional review board approved this study. Patients who agreed to participate in the study gave written informed consent. Between March and August of 2012, study participation was offered to patients who underwent mammography and handheld ultrasonography (HHUS) and were scheduled for biopsy. If a category 4 or 5 lesion was found on mammography or HHUS, biopsy was done. Further, the biopsy of C2 and C3 lesions was performed when the patient or the clinician wanted pathological confirmation. In total, 172 women (age range, 20 to 80 years; mean, 48 years) were included. ABUS was performed before biopsy for all patients included in the study. Thirty-two patients and 1 patient underwent biopsy for 2 lesions and 3 lesions, respectively. Of the 206 lesions included, 191 were visible and accessible using HHUS, and ultrasonography-guided 14 G core needle biopsies were performed. The other 15 lesions were confirmed by mammography-guided, 11 G vacuum-assisted biopsy.

Ultrasonography

Mammography images were obtained in the standard craniocaudal and mediolateral oblique views using a Mammomat 3000 unit (Siemens Medical Solutions, Solna, Sweden) and a Lorad M3 mammography unit (Hologic Inc., Boston, MA, USA).

Bilateral HHUS examinations of the breast were randomly performed by any of the four radiologists, each with 8-, 6-, 1-, and 1-year experience in breast imaging, using a 7–15-MHz linear transducer (iU22 Ultrasonography System, Philips Healthcare, Bothell, WA, USA) or a 6–14-MHz linear transducer (EUB-8500 scanner, Hitachi Medical, Tokyo, Japan).

ABUS images were obtained using the ACUSON S2000 Automated Breast Volume Scanner (ABVS; Siemens Medical Solutions, Mountain

View, CA, USA), operated by a trained radiographer. The ABVS acquired 15.4 cm × 16.8 cm × maximum 6 cm volume data sets of the breast in one sweep by using a 5–14-MHz wide-aperture linear probe. The bilateral breast was initially scanned with an anterior-posterior view that included the nipple and most parts of the breast, with the patient in the supine position. Lateral and medial views, which mainly included the outer and the inner breast, were then scanned with the patient in an oblique position. After acquisition, the three-dimensional (3D) volume data were automatically sent from the ACUSON S2000 ABVS to the workstation and reviewed in multiple orientations (transverse, coronal, and sagittal plane) by using a multi-planar reconstruction display. The scan thickness was displayed at intervals of 1 mm without overlap.

Image Analysis

Two breast radiologists, with either 8 or 6 years of experience with breast imaging and 60 cases of experience with ABUS, analyzed all the ABUS data independently after reviewing the mammography results. They knew that patients were scheduled for biopsy after the ABUS examination. They were blinded to the HHUS image findings and to other clinical information. The two readers evaluated various features of the breast lesions and described the type ("not special" or special cases defined as any of the following: cyst, clustered cyst, intraductal lesion, postoperative scar, or calcifications), shape (oval, round, or irregular), orientation (parallel or non-parallel), margin (circumscribed, indistinct, microlobulated, angular, or spiculated), echogenicity (anechoic, hypoechoic, isoechoic, hyperechoic, or complex), posterior acoustic features (none, enhancement, shadowing, or combined), and categorization of final assessment (category 2, category 3, category 4A, category 4B, category 4C, or category 5) according to Breast Imaging Reporting and Data System (BI-RADS) [12].

Statistical Analysis

Statistical analysis was performed using SAS ver. 9.2 (SAS Institute Inc., Cary, NC, USA). The ABUS detection rate of 206 breast lesions was calculated for each radiologist, and the detection rates of benign and malignant lesions of ABUS and HHUS were compared. When lesions were detected by both of the radiologists, the interobserver agreement was assessed using kappa statistics. We used the following definition to interpret the kappa coefficients: a kappa (κ) value of equal to or less than 0.20 indicated a slight agreement; values from 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, almost perfect agreement [13]. For variables that included ordering of the values, a weighted κ -value was obtained.

Results

Of the 206 lesions, 46 were malignant (35 invasive ductal carcinomas, 1 invasive lobular carcinoma, 1 mixed invasive ductal and mucinous carcinoma, 1 mucinous carcinoma, and 8 ductal carcinomas *in situ*), and 160 were benign (5 atypical ductal hyperplasia, 1 radial scar, and 154 other benign lesions such as fibrocystic change or fibroadenoma). Nine ultrasound-guided vacuum-assisted breast biopsies and 12 surgical excisions were done for 2 tubular adenomas, 4 intraductal papillomas, 7 fibroadenomas, 6 atypical ductal hyperplasias, 1 fibrocystic change, and 1 complex sclerosing lesion. The result of 1 lesion was upgraded from atypical ductal hyperplasia to ductal carcinoma *in situ*.

Among the 206 pathologically confirmed lesions, 194 lesions were noted on HHUS. Among them, reader 1 detected 164 lesions and reader 2 detected 149 lesions on ABUS. Two lesions were noted only on ABUS. In all, the two readers detected 166 and 150 lesions, respectively. The 25 lesions missed on ABUS were all benign.

A total of 145 lesions were detected by both of the two readers using ABUS. Of the 145 lesions detected in common, 144 were seen on HHUS images and 1 lesion was not noted on HHUS; the pathological examination revealed 38 malignant lesions and 107 benign lesions (Fig. 1).

The interobserver agreement was assessed for the 145 lesions detected by both of the readers. There was substantial agreement on shape ($\kappa=0.707$) and moderate agreement on type, margin, mass orientation, echogenicity, and posterior acoustic features ($\kappa=0.592$, 0.438, 0.472, 0.524, and 0.541, respectively). Final BI-RADS assessments yielded a kappa value of 0.397 and weighted kappa

value of 0.661 when subdivisions 4A, 4B, and 4C were included. With respect to categories 1, 2, 3, 4, and 5, the κ -value was 0.505 and the weighted kappa value 0.611 (Table 1, Figs. 2-4).

Table 1. Interobserver agreement on ultrasonography features of 145 lesions

Lexicon	Subgroup	Lesion no. of reader 1	Lesion no. of reader 2	κ -value
Type	Not special	130	128	0.592
	Special	15	17	
	Cyst	8	13	
	Clustered cyst	3	2	
	Intraductal lesion	4	2	
	Postoperative scar	0	0	
	Calcification	0	0	
Shape	Oval	93	96	0.707
	Round	4	5	
	Irregular	48	44	
Orientation	Parallel	128	113	0.472
	Non-parallel	17	32	
Margin	Circumscribed	70	92	0.438
	Indistinct	31	12	
	Microlobulated	14	8	
	Angular	12	17	
	Spiculated	18	16	
Echogenicity	Anechoic	13	14	0.524
	Hypoechoic	118	127	
	Isoechoic	7	3	
	Hyperechoic	0	0	
	Complex	7	1	
Posterior acoustic features	None	71	84	0.541
	Enhancement	57	43	
	Shadowing	13	15	
	Combined	4	3	
BI-RADS final assessment	Category 2	11	21	0.397 ^{a)} – 0.505 ^{b)}
	Category 3	57	61	
	Category 4	60	46	
	Category 4A	32	27	
	Category 4B	13	7	
	Category 4C	15	12	
	Category 5	17	17	

BI-RADS, Breast Imaging Reporting and Data System.

^{a)}Interobserver agreement obtained with the categorization of BI-RADS category 2, 3, 4A, 4B, 4C, and 5. ^{b)}Interobserver agreement obtained with the categorization of BI-RADS category 2, 3, 4, and 5.

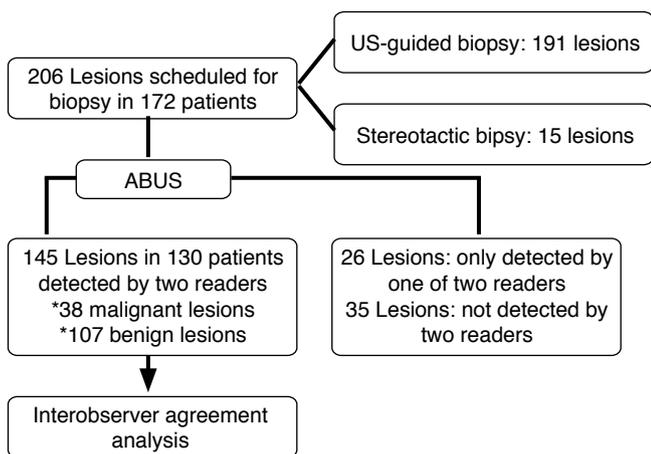


Fig. 1. Flow chart summarizes the study sample in terms of lesions and their number. ABUS, automated whole breast ultrasonography; US, ultrasonography.

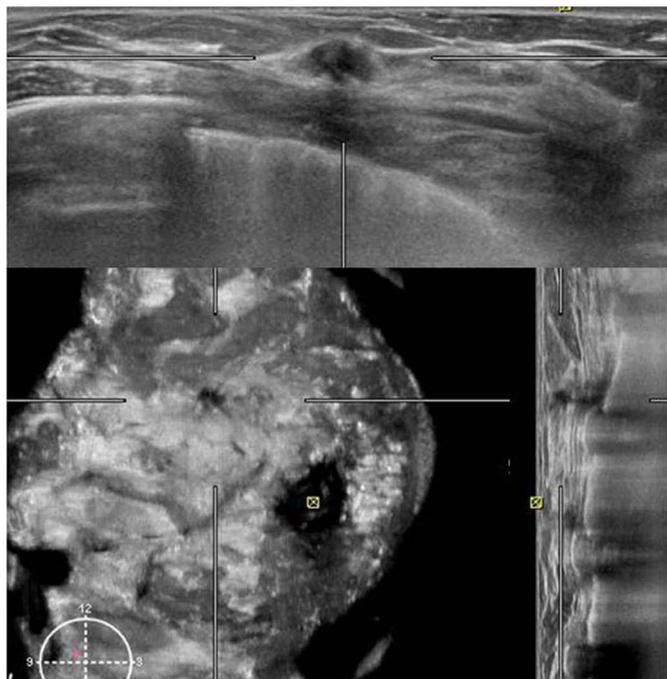


Fig. 2. Automated whole breast ultrasonography (ABUS) images of a breast lesion in a 51-year-old woman. (Upper, transverse plane; lower left, coronal reconstruction; lower right, sagittal plane; yellow square mark, position of the nipple). Both radiologists described this breast lesion as not-special, hypoechoic, and irregular in shape, with an abrupt boundary, parallel orientation, and no posterior acoustic features. One radiologist described the lesion boundary as spiculated, while the other described the lesion boundary as indistinct. One radiologist categorized the lesion as category 4C, and the other as category 4A. The mass was pathologically confirmed as an invasive ductal carcinoma.

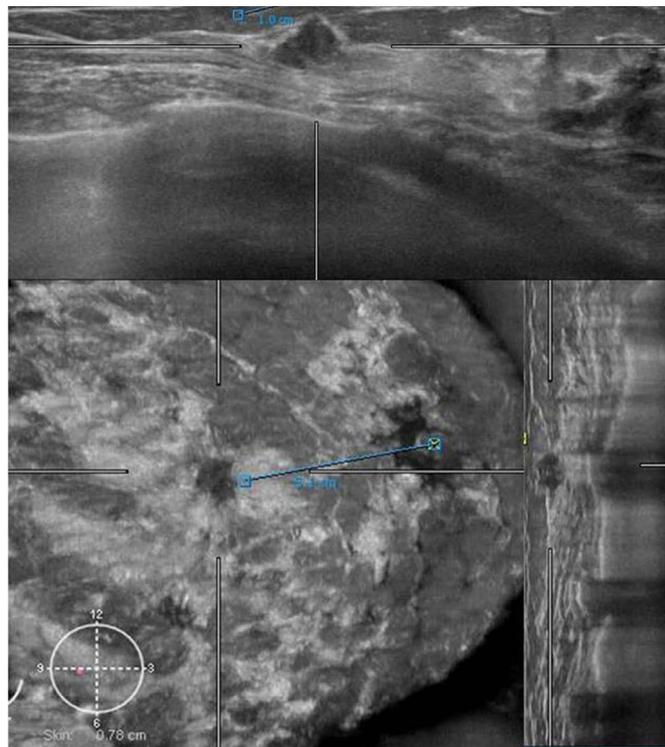


Fig. 3. Automated whole breast ultrasonography (ABUS) images of a breast lesion in a 43-year-old woman. (Upper, transverse plane; lower left, coronal reconstruction; lower right, sagittal plane; yellow square mark, position of the nipple). Both radiologists described this breast lesion as not-special, hypoechoic, and irregular in shape, with an abrupt boundary and a parallel orientation. One radiologist described the lesion as having a spiculated margin and posterior enhancement. The other described the lesion as having an angular margin and no posterior acoustic features. One radiologist categorized this lesion as category 4C, and the other as category 4A. The mass was pathologically confirmed as an invasive ductal carcinoma.

Discussion

ABUS has been proposed as a promising tool for overcoming poor standardization and reproducibility of HHUS results. In addition to favorable standardization and reproducibility, the application of a computer-aided detection (CAD) system may be useful in improving the screening efficiency [14]. Indeed, several studies have shown equivalent performance between ABUS and HHUS [1,15]. The results of this study are meaningful in that we included a large population with breast lesions that were confirmed pathologically. In our study, ABUS found all of the malignant lesions noted on HHUS.

Consistent reporting with reproducible characterization of breast lesions is also critical for the clinical application of ABUS. We found substantial agreement on the description of shape ($\kappa=0.707$) and moderate agreement on type, margin, mass orientation,

echogenicity, and posterior acoustic features ($\kappa=0.592, 0.438, 0.472, 0.524,$ and 0.541 , respectively). The BI-RADS final assessment yielded fair agreement ($\kappa=0.397$) and a higher weighted kappa value (0.661) when weighing the close misses more heavily.

The interagreement of descriptors using ABUS was comparable to the interobserver agreement achieved using HHUS. With respect to lesion shape, margin, echogenicity, and posterior acoustic features, ABUS showed higher levels of interobserver agreement than those previously reported using HHUS ($\kappa=0.42-0.64, 0.32-0.36, 0.36-0.58,$ and $0.47-0.53$, respectively) (Table 2) [2,4,5]. For the final assessment, Park et al. [4] and Abdullah et al. [2] reported similar interobserver agreements for the final assessment (0.30 with the subcategorization of category 4 and 0.49 without the subcategorization of category 4, respectively). Lee et al. [5] reported a higher interobserver agreement than that of our study

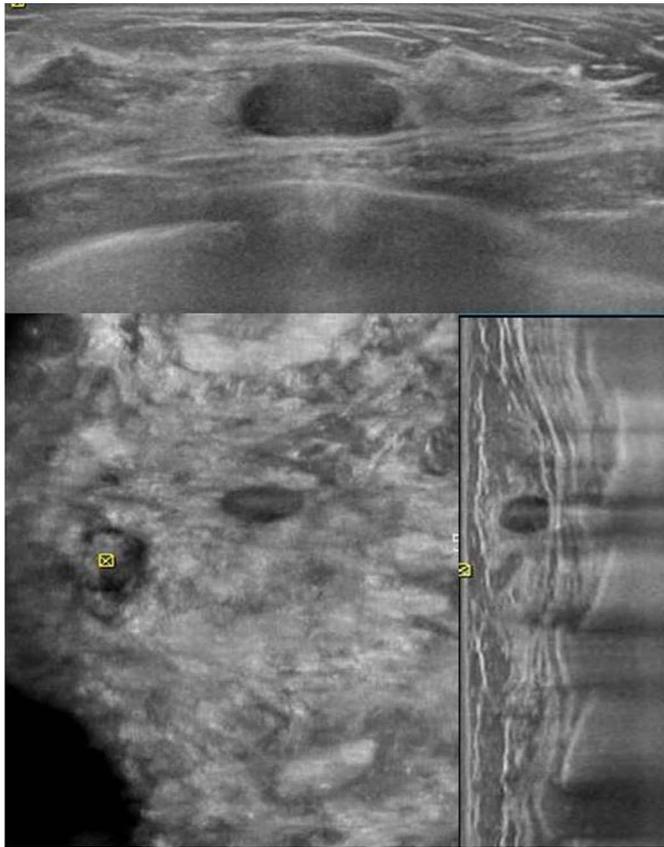


Fig. 4. Automated whole breast ultrasonography (ABUS) images of a breast lesion in a 34-year-old woman. (Upper, transverse plane; lower left, coronal reconstruction; lower right, sagittal plane; yellow square mark, position of the nipple). Both radiologists described this breast lesion as not-special, hypoechoic, and oval in shape, with a circumscribed margin, an abrupt boundary, and a parallel orientation. The two radiologists only disagreed in their descriptions of the posterior acoustic features. Both radiologists categorized this lesion as category 3. The mass was pathologically confirmed as a fibroadenoma.

Table 2. ABUS and HHUS studies evaluating the interobserver agreement

Variable	ABUS (κ -value)			HHUS (κ -value)		
	Zhang et al. [11]	Shin et al. [21]	Kim et al. [20]	Abdullah et al. [2]	Park et al. [4]	Lee et al. [5]
Lesion number	234	145	26	267	314	150
Reader number	2	2	2	5	4	4
Type	–	0.75 (mass); 0.63 (special case)	–	–	–	–
Shape	0.79	0.71	0.45	0.64	0.42	0.49
Orientation	0.74	0.72	0.50	0.70	0.61	0.56
Margin	0.76	0.61	0.25	0.36	0.32	0.33
Echogenicity	0.69	0.45	0.65	0.58	0.36	0.37
Posterior acoustic features	0.68	0.42	0.45	0.47	0.53	0.49
BI-RADS final assessment	0.70	0.63	0.57	0.30	0.49	0.53 ^{a)} –0.62 ^{b)}

ABUS, automated whole breast ultrasonography; HHUS, handheld ultrasonography; BI-RADS, Breast Imaging Reporting and Data System.

^{a)}Interobserver agreement obtained with the categorization of BI-RADS category 3, 4A, 4B, 4C, and 5. ^{b)}Interobserver agreement obtained with the categorization of BI-RADS category 3, 4, and 5.

(0.53 with the subcategorization of category 4 and 0.62 without the subcategorization of category 4). However, a direct comparison is not available because category 2 is not included in this study. Some possible reasons for the higher rate of agreement for some descriptors when using ABUS can be suggested. Readers can reproduce whole breast scans in multiple orientations by using the 3D volume data transmitted to the workstation. Scans of the coronal

plane can be performed using ABUS and provide an advantage for evaluating breast lesions. A review of coronal images may be helpful for distinguishing between real lesions and non-homogeneous areas [16]. Several studies have suggested that the retraction phenomenon with an irregular margin on the coronal plane is a characteristic of breast cancer [8,17,18], and lobular carcinoma presents an architectural distortion on the coronal plane images

that is not apparent when using mammography or conventional 2D ultrasonography [18,19]. Additionally, a multislice observation of ABUS images leads to a more consistent interpretation of the lesion margin than a static HHUS image [11].

A few studies have evaluated the interobserver agreement for breast lesions when using ABUS (Table 2) [11,20,21]. Compared to these studies, the agreement level for most ultrasonographic characteristics in this study was lower, except for shape and posterior acoustic features [11,21]. This result can be explained by the heterogeneity of lesion types included in this study. We analyzed breast lesions that included special lesion types, such as cysts, clustered cysts, intraductal lesions, postoperative scarring, and calcifications, while the other studies evaluated mass lesions.

Our study had several limitations. We only studied the agreement between two different examiners because ABUS is a new modality at our institution. For the same reason, both of the readers involved were not familiar with ABUS imaging. A further evaluation of the interobserver agreement among multiple radiologists is needed. Additionally, selection bias may exist in this study because only biopsy-confirmed lesions were included. The duration of follow-up for benign lesions confirmed via a 14-core needle biopsy was less than two years; therefore, the possibility of false negative results remains.

ABUS is a promising diagnostic tool with a good interobserver agreement comparable to HHUS. Once a good interobserver agreement is established, ABUS can be used clinically and may overcome the weaknesses of HHUS, which include poor standardization and reproducibility.

ORCID: Eun Jeong Kim: <http://orcid.org/0000-0003-2053-8855>; Sung Hun Kim: <http://orcid.org/0000-0003-4478-9720>; Bong Joo Kang: <http://orcid.org/0000-0002-5991-6035>; Yun Ju Kim: <http://orcid.org/0000-0001-7658-8726>

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

The author wishes to acknowledge the financial support of the Catholic Medical Center Research Foundation, made in the program year of 2012.

References

- Bassett LW. Imaging of breast masses. *Radiol Clin North Am* 2000;38:669-691.
- Abdullah N, Mesurrolle B, El-Khoury M, Kao E. Breast imaging reporting and data system lexicon for US: interobserver agreement for assessment of breast masses. *Radiology* 2009;252:665-672.
- Lazarus E, Mainiero MB, Schepps B, Koelliker SL, Livingston LS. BI-RADS lexicon for US and mammography: interobserver variability and positive predictive value. *Radiology* 2006;239:385-391.
- Park CS, Lee JH, Yim HW, Kang BJ, Kim HS, Jung JI, et al. Observer agreement using the ACR Breast Imaging Reporting and Data System (BI-RADS)-ultrasound, First Edition (2003). *Korean J Radiol* 2007;8:397-402.
- Lee HJ, Kim EK, Kim MJ, Youk JH, Lee JY, Kang DR, et al. Observer variability of Breast Imaging Reporting and Data System (BI-RADS) for breast ultrasound. *Eur J Radiol* 2008;65:293-298.
- Skaane P, Engedal K, Skjennald A. Interobserver variation in the interpretation of breast imaging: comparison of mammography, ultrasonography, and both combined in the interpretation of palpable noncalcified breast masses. *Acta Radiol* 1997;38:497-502.
- Baker JA, Kornguth PJ, Soo MS, Walsh R, Mengoni P. Sonography of solid breast lesions: observer variability of lesion description and assessment. *AJR Am J Roentgenol* 1999;172:1621-1625.
- Kotsianos-Hermle D, Hiltawsky KM, Wirth S, Fischer T, Friese K, Reiser M. Analysis of 107 breast lesions with automated 3D ultrasound and comparison with mammography and manual ultrasound. *Eur J Radiol* 2009;71:109-115.
- Richter K, Heywang-Kobrunner SH, Winzer KJ, Schmitt KJ, Prihoda H, Froberg HD, et al. Detection of malignant and benign breast lesions with an automated US system: results in 120 cases. *Radiology* 1997;205:823-830.
- Lin X, Wang J, Han F, Fu J, Li A. Analysis of eighty-one cases with breast lesions using automated breast volume scanner and comparison with handheld ultrasound. *Eur J Radiol* 2012;81:873-878.
- Zhang J, Lai XJ, Zhu QL, Wang HY, Jiang YX, Liu H, et al. Interobserver agreement for sonograms of breast lesions obtained by an automated breast volume scanner. *Eur J Radiol* 2012;81:2179-2183.
- American College of Radiology. Breast Imaging Reporting and Data System Atlas (BI-RADS Atlas). Reston, VA: American College of Radiology, 2003.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174.
- Kim JH, Cha JH, Kim N, Chang Y, Ko MS, Choi YW, et al. Computer-aided detection system for masses in automated whole breast ultrasonography: development and evaluation of the effectiveness. *Ultrasonography* 2014;33:105-115.
- Wenkel E, Heckmann M, Heinrich M, Schwab SA, Uder M, Schulz-Wendtland R, et al. Automated breast ultrasound: lesion detection and BI-RADS classification--a pilot study. *Rofo* 2008;180:804-808.
- Chang JM, Moon WK, Cho N, Park JS, Kim SJ. Breast cancers initially detected by hand-held ultrasound: detection performance

- of radiologists using automated breast ultrasound data. *Acta Radiol* 2011;52:8-14.
17. Watermann DO, Foldi M, Hanjalic-Beck A, Hasenburg A, Lughausen A, Prompeler H, et al. Three-dimensional ultrasound for the assessment of breast lesions. *Ultrasound Obstet Gynecol* 2005;25:592-598.
 18. Cho KR, Seo BK, Lee JY, Pisano ED, Je BK, Lee JY, et al. A comparative study of 2D and 3D ultrasonography for evaluation of solid breast masses. *Eur J Radiol* 2005;54:365-370.
 19. Chapellier C, Balu-Maestro C, Bleuse A, Ettore F, Bruneton JN. Ultrasonography of invasive lobular carcinoma of the breast: sonographic patterns and diagnostic value: report of 102 cases. *Clin Imaging* 2000;24:333-336.
 20. Kim SH, Kang BJ, Choi BG, Choi JJ, Lee JH, Song BJ, et al. Radiologists' performance for detecting lesions and the interobserver variability of automated whole breast ultrasound. *Korean J Radiol* 2013;14:154-163.
 21. Shin HJ, Kim HH, Cha JH, Park JH, Lee KE, Kim JH. Automated ultrasound of the breast for diagnosis: interobserver agreement on lesion detection and characterization. *AJR Am J Roentgenol* 2011;197:747-754.