Ultrasound in Med. & Biol., Vol. 42, No. 9, pp. 2083–2088, 2016

0301-5629/$ - see front matter

http://dx.doi.org/10.1016/j.ultrasmedbio.2016.05.006

● *Original Contribution*

COMPARISON OF INTER-OBSERVER VARIABILITY AND DIAGNOSTIC PERFORMANCE OF THE FIFTH EDITION OF BI-RADS FOR BREAST ULTRASOUND OF STATIC VERSUS VIDEO IMAGES

JI HYUN YOUK,* INKYUNG JUNG,[†] JUNG HYUN YOON,[‡] SUNG HUN KIM,[§] YOU ME KIM,[¶] EUN HYE LEE,[||] SUN HYE JEONG,[||] and MIN JUNG KIM[‡]

*Department of Radiology, Gangnam Severance Hospital, Research Institute of Radiological Science, College of Medicine, Yonsei University, Seoul, South Korea; [†]Department of Biostatistics and Medical Informatics, Yonsei University College of Medicine, Seoul, South Korea; [‡]Department of Radiology, Severance Hospital, Research Institute of Radiological Science, College of Medicine, Yonsei University, Seoul, South Korea; [§]Department of Radiology, Seoul St. Mary's Hospital, College of Medicine, Catholic University of Korea, Seoul, South Korea; [¶]Department of Radiology, Dankook University Hospital, College of Medicine, Dankook University, Yongin, South Korea; and [||]Department of Radiology, Soonchunhyang University, Bucheon Hospital, Bucheon, South Korea

(*Received* 10 *January* 2016; *revised* 2 *May* 2016; *in final form* 4 *May* 2016)

Abstract—Our aim was to compare the inter-observer variability and diagnostic performance of the Breast Imaging Reporting and Data System (BI-RADS) lexicon for breast ultrasound of static and video images. Ninety-nine breast masses visible on ultrasound examination from 95 women 19–81 y of age at five institutions were enrolled in this study. They were scheduled to undergo biopsy or surgery or had been stable for at least 2 y of ultrasound follow-up after benign biopsy results or typically benign findings. For each mass, representative long- and short-axis static ultrasound images were acquired; real-time long- and short-axis B-mode video images through the mass area were separately saved as cine clips. Each image was reviewed independently by five radiologists who were asked to classify ultrasound features according to the fifth edition of the BI-RADS lexicon. Inter-observer variability was assessed using kappa (κ) statistics. Diagnostic performance on static and video images was compared using the area under the receiver operating characteristic curve. No significant difference was found in κ values between static and video images for all descriptors, although κ values of video images were higher than those of static images for shape, orientation, margin and calcifications. After receiver operating characteristic curve analysis, the video images (0.83, range: 0.77–0.87) had higher areas under the curve than the static images (0.80, range: 0.75–0.83; $p = 0.08$). Inter-observer variability and diagnostic performance of video images was similar to that of static images on breast ultrasonography according to the new edition of BI-RADS. (E-mail: mines@yuhs.ac) **© 2016 World Federation for Ultrasound in Medicine & Biology.**

Key Words: **Ultrasonography, Breast neoplasms, Breast diseases, Observer variation, BI-RADS.**

INTRODUCTION

Breast ultrasonography is currently considered to be integral in the detection and diagnosis of benign and malignant breast masses as a first-line examination or an adjunct to mammography (Abdullah et al. 2009). Although the use of hand-held transducers remains standard practice throughout the world, breast ultrasonography has inherent drawbacks of operator dependence and lack of reproducibility. To minimize variability in characterization and final assessment of breast masses identified on ultrasonography, the Breast Imaging Reporting and Data System (BI-RADS) was developed by the American College of Radiology (2003). Previous studies have assessed the reliability of the BI-RADS lexicon or final assessment for breast ultrasonography and validated its value in predicting malignancy (Hamy et al. 2012; Kim et al. 2008; Lazarus et al. 2006). Recently, the new edition of BI-RADS ultrasound was published with the addition of new sections and changes in terminology (Mendelson et al. 2013). For example, lesion boundary is no longer a major category, and for the terms used to describe echo pattern, "complex" was changed to "complex cystic and solid," and

Address correspondence to: Min Jung Kim, Department of Radiology, College of Medicine, Yonsei University, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 120-752, South Korea. E-mail: mines@yuhs.ac

"heterogeneous," a mixture of echogenic patterns within a solid mass, has been added. Regarding the descriptors used for calcification, the distinction between "macrocalcification" and "microcalcification" was deleted and "intra-ductal calcifications" was added.

In many clinical situations, real-time imaging is required for accurate interpretation, which ultrasonography can provide. The strength of ultrasonography is that it can be performed and evaluated in real time. Nevertheless, only static representative ultrasound images selected by investigators are typically evaluated, which makes the merits of real-time ultrasonography invalid (Sung 2014). With most ultrasound machines, video images are acquired by recording the breast scan as real-time video clips, which could be expected to provide more information in real time on a breast lesion than do static images. To date, there has been only one study that compared video and static images of breast lesions with respect to sonographic assessment, but the assessment was based on the fourth edition of BI-RADS (Foldi et al. 2011).

We performed the present study to compare the inter-observer variability and diagnostic performance of the fifth edition of BI-RADS on breast ultrasound static and video images.

METHODS

This prospective study was conducted at five different institutions and approved by their institutional review boards (Gangnam Severance Hospital, Severance Hospital, St. Mary's Hospital, Dankook University Hospital and Soonchunhyang University Bucheon Hospital). Informed consent was obtained from all patients.

Patients

Eligible patients were women with a breast mass visible on breast ultrasound examination performed at five different institutions from November 2013 to April 2014, who were scheduled to undergo ultrasound-guided core needle biopsy or surgery or had been stable for at least 2 y of ultrasound follow-up after a benign biopsy result or typically benign findings, such as simple cysts, intramammary lymph nodes and postsurgical fluid collection (Mendelson et al. 2013). In total, 99 breast masses from 95 women 19–81 y of age (mean: 46.8 ± 12.4 y) were enrolled in this study.

Ultrasound examination

Each mass was examined with the institution's house system (iU22, Philips Medical Systems, Bothell, WA, USA; GE LOGIQ E9, GE Medical Systems, Milwaukee, WI, USA; SuperSonic Imagine, Aix-en-Provence, France; EUB-8500, Hitachi Medical, Tokyo, Japan) equipped with high-frequency linear array transducers. During scanning, spatial compounding was used in all units, and harmonic imaging was used in two units (SuperSonic Imagine and EUB-8500). Seven radiologists who specialized in breast imaging with 3–13 y of experience performed the ultrasound examinations. After study entry, representative long- and short-axis B-mode images of the mass were acquired, and the mass diameter (the longest dimension of the mass on ultrasound) was recorded. To obtain video images of the mass, real-time long- and short-axis B-mode images were separately saved on a hard drive as cine clips of at least 5 s per image. Scanning of each video image started from the area of normal breast parenchyma surrounding the mass, through the entire mass and to the area of normal breast parenchyma surrounding the other end of the mass in one-directional movement of the transducer. Doppler images or elastograms were not obtained for review.

Image evaluation

For each breast mass, two Microsoft Power Point 2010 data sets of static images and cine clips consisting of two representative long- and short-axis images were analyzed with a 3-wk interval between reading sessions. Because all reviewers contributed to image acquisition and to minimize reader bias, images of each data set were made anonymous, presented in randoom order and reviewed 3 mo after data collection. Each ultrasound image was reviewed independently by five radiologists who were dedicated to breast imaging and had 6–13 y of experience in breast ultrasound and BI-RADS. During the review, radiologists were blinded to the clinical, mammographic and histopathologic findings and asked to describe ultrasound features according to the BI-RADS lexicon terminology (shape, orientation, margin, echo pattern, posterior features and calcifications) and to assign BI-RADS final assessment categories (2, 3, 4a, 4b, 4c and 5) to the masses to indicate the probability of malignancy (Mendelson et al. 2013).

Statistical analysis

After review of the histopathology of core biopsy or surgery, at least 2 y of follow-up results and a typically benign finding on ultrasound, the final diagnoses were categorized as benign or malignant.

The Fleiss κ-value was calculated with 95% confidence intervals (95% CIs) to assess the proportion of inter-observer agreement for ultrasound features and final assessment categories according to BI-RADS among the five reviewers, as well as the dichotomized BI-RADS final assessment categories (positive assessments [categories 4a, 4b, 4c and 5] and negative assessments [categories 2 and 3]). We applied the degrees of agreement for different ranges of κ-values suggested by