Ultrasound in Med. & Biol., Vol. ■, No. ■, pp. 1–7, 2016

0301-5629/$ - see front matter

http://dx.doi.org/10.1016/j.ultrasmedbio.2016.04.020

● *Original Contribution*

ASSESSMENT OF DIAGNOSTIC ACCURACY AND EFFICIENCY OF CATEGORIES 4 AND 5 OF THE SECOND EDITION OF THE BI-RADS ULTRASOUND LEXICON IN DIAGNOSING BREAST LESIONS

Xuebin Zou,* Jianwei Wang,* Xiaowen Lan,† Qingguang Lin,* Feng Han,* Longzhong Liu,* and Anhua Li*

*Department of Ultrasound, Cancer Center of Sun Yat-Sen University, Guangzhou, Guangdong Province, China; and †Department of Radiation Oncology, Cancer Center of Sun Yat-Sen University, Guangzhou, Guangdong Province, China

(*Received* 22 *January* 2016; *revised* 14 *April* 2016; *in final form* 27 *April* 2016)

Abstract—The purpose of this study was to evaluate the diagnostic accuracy and efficiency of categories 4 and 5 of the second edition of the Breast Imaging Reporting and Data System (BI-RADS) ultrasound (US) lexicon in diagnosing breast lesions. In our retrospective study, 579 lesions in 544 patients were assessed by US as the preliminary diagnosis and classified in subcategories 4a–4c and category 5 based on the second edition of the BI-RADS US lexicon with some obvious changes, such as the redefined margin, new calcification type, associated features and some special cases. Inter-observer agreement was determined. Ultrasound results were compared with the pathologic results for confirmation. Positive predictive values (PPVs) of subcategories 4a–4c were compared with theoretical values using the χ^2 test; the binomial test was used for category 5 lesions. Of the 579 lesions, 212 were confirmed as benign (36.61%), and the remaining 367 lesions were confirmed as borderline/malignant (63.39%). Inter-observer agreement was moderate for subcategories 4a–4c ($\kappa = 0.52$), moderate for subcategories 4a–4c and category 5 ($\kappa = 0.56$) and substantial for categories 4 and 5 ($\kappa = 0.67$). The PPVs for subcategories 4a–4c were 23.74%, 70.67% and 81.25%, respectively. In addition, the total PPV for category 4 was 46.92% (183/390), and the total PPV for category 5 was 97.35% (184/189). Statistical results revealed that the PPVs of subcategories 4a and 4b differed significantly from the theoretical values ($p < 0.05$); the PPVs of subcategory 4c and category 5 were significantly correlated with the theoretical PPVs ($p > 0.05$). In conclusion, subcategories 4a and 4b have lower diagnostic efficiency than subcategory 4c and category 5. Inter-observer agreement for subcategories 4a–4c remains to be improved. The most common features of subcategories 4a–4c differ, but overlap. It is recommended that inexperienced doctors in primary hospitals not classify lesions into subcategories in clinical practice. (E-mail: liah1314@163.com) © **2016 World Federation for Ultrasound in Medicine & Biology.**

Key Words: **Breast lesions, Ultrasound, Breast Imaging Reporting and Data System, BI-RADS categories 4 and 5, pathology.**

INTRODUCTION

The imaging evaluation of breast abnormalities commonly includes conventional mammography, ultrasonography (US) and magnetic resonance imaging (Sickles et al. 2014). Because the use of mammography in breast examination is limited, especially by high-density mammary glands, which form a barrier to mammography, ultrasound, which plays a significant role in modern medicine, has today become a common tool for breast examination in most hospitals (Corsetti et al. 2006; Hille et al. 2004; Hong et al. 2005; Lazarus et al. 2006).

The American College of Radiology published the first edition of its Breast Imaging Reporting and Data System (BI-RADS) US lexicon in 2003 to standardize diagnostic characterization of breast lesions with US (American College of Radiology 2003; Hong et al. 2005; Lazarus et al. 2006; Mendelson et al. 2001). The first edition of the BI-RADS US lexicon included shape, orientation, margins, boundary, echo pattern and posterior features, as well as surrounding tissue alterations (American College of Radiology 2003). After 10 y of practical clinical use, some features were found to have high specificity and sensitivity, whereas other features were not found suitable for diagnosis at all. Therefore,

Address correspondence to: Anhua Li, Ultrasonography Department, Cancer Center of Sun Yat-Sen University, Guangzhou, 510060, Guangdong Province, China. E-mail: liah1314@163.com

in 2013, the second edition of the BI-RADS US lexicon of the fifth edition of the BI-RADS atlas was published (American College of Radiology 2013).

Breast ultrasound is highly accurate in differentiating between benign and malignant lesions. Each lesion is classified in an independent category that indicates the clinical strategy recommended, with the requirement by the second edition of the BI-RADS US lexicon that the same pattern as in the mammogram be used (Chala et al. 2007; Lazarus et al. 2006). There are seven categories, from 0 to 6: category 0 = need for additional imaging; category 1 = no finding or negative finding; category 2 = benign lesion with no suspicious findings in the breast; category 3 = <2% likelihood of malignancy; category 4 = absence of classic features of malignancy, but sufficient suspicion to recommend a biopsy, with a likelihood of malignancy from 2% to 95%; category 5 = >95% likelihood of malignancy with classic features; category 6 = pathologic confirmation of specific malignancy before US. Because of its wide-ranging likelihood of malignancy, category 4 has been subdivided into subcategories 4a, 4b and 4c, which represent 2%–10%, 10%–50% and 50%–95% likelihoods of malignancy, respectively (American College of Radiology 2013).

To our knowledge, there have been studies comparing the accuracy and efficiency of categories 4 and 5 of the first edition of the BI-RADS US lexicon. The conclusions were inconsistent. Some authors noted that results differed from the theoretical values, and others concluded that categories 4 and 5 of the first edition were not sufficiently accurate and efficient (Gweon et al. 2012; Hamy et al. 2012; Heinig et al. 2008; Kim et al. 2008; Lazarus et al. 2006; Lee et al. 2008; Raza et al. 2008). Inter-observer agreement for the first edition was only fair (Abdullah et al. 2009). Some of the inconsistencies in classification by different medical institutions directly influenced and delayed clinical management (Raza et al. 2010). On the other hand, the subcategory system is not fully popularized to use in clinic at that time so that abundant of studies stayed at the stage of evaluating categories only (Raza et al. 2008).

The purpose of our study was to evaluate in depth the diagnostic accuracy and efficiency of categories 4 and 5 of the second edition of the BI-RADS US lexicon in diagnosing lesions.

METHODS

Patient population

In this study, we reviewed a total of 579 lesions in 544 patients, which were assessed with US and classified into subcategories 4a–4c and category 5 based on different likelihoods of malignancy, from May 2015 to August 2015. Lesions never confirmed by pathology were excluded from the study even if reported and assigned to subcategories 4a–4c or category 5. Lesions with sonographic images were recorded by LOGIQ S8 (GE Medical Systems, Milwaukee, WI, USA), iU-22 (Philips Medical Systems, Bothell, WA, USA) and DC-8 (Mindray Medical International Limited, Shenzhen, Guangdong, China) before biopsy or surgery. This retrospective study was approved by the institutional review board of the Sun Yat-Sen University Cancer Center. Written consent was waived; oral consent was obtained.

Imaging and interpretation

All doctors in our department, with work experience ranging from 3 to 20 y, received systematic training in use of the second edition of the BI-RADS US lexicon and were qualified to diagnose and categorize lesions They identified and described the US features of lesions in detail and classified each lesion into a category or subcategory according to the following rules.

For each lesion, US examinations in B-mode and Doppler were recorded in JPEG format. Observations were carried out in at least two planes of the lesion. And measurements were performed in paralleled and vertical planes to obtain three diameter lines to depict the size and shape of the lesion. For each patient, the most diagnostic symptoms of lesions were described in the reports with the new standards of the second edition, and the category or subcategory of each lesion was based on position, size, orientation, shape, margins, internal echo pattern, posterior features, calcification, associated features, vascularity and so on.

Three doctors with different levels of experience (3, 5–10 and 10–15 y, respectively) worked with 100 lesions randomly selected from the population and categorized them. Their results were used to calculate the κ value, which assesses inter-reader agreement. The guidelines of Landis and Koch (1977) were followed in interpreting κ values: 0.00–0.20 = slight agreement; 0.21–0.40 = fair agreement; 0.41–0.60 = moderate agreement; 0.61–0.80 = substantial agreement; 0.80–1.00 = almost perfect agreement.

Pathologic analysis

Lesions were biopsied using by core biopsy, surgical excision or both, according to the clinical routine. US-guided core biopsy was performed with an automated biopsy gun equipped with a 16- or 18-gauge needle, which was used by experienced interventional US doctors and nurses to acquire tissue samples. An experienced pathologist performed the histologic examination. The pathologic results of surgery and the malignant pathologic results of biopsy were defined as confirmed results. Any benign pathologic results of biopsy were