Ultrasound in Med. & Biol., Vol. 42, No. 11, pp. 2622–2629, 2016

0301-5629/$ - see front matter

http://dx.doi.org/10.1016/j.ultrasmedbio.2016.06.017

● *Original Contribution*

INTER- AND INTRA-OBSERVER AGREEMENT IN ULTRASOUND BI-RADS CLASSIFICATION AND REAL-TIME ELASTOGRAPHY TSUKUBA SCORE ASSESSMENT OF BREAST LESIONS

Fabienne Schwab,* Katharina Redling,* Matthias Siebert,* Andy Schötzau,† Cora-Ann Schoenenberger,‡ and Rosanna Zanetti-Dällenbach*

*Department of Gynecology and Obstetrics, University Hospital of Basel, Basel, Switzerland; †Statistics, Department of Gynecology and Obstetrics, University Hospital of Basel, Basel, Switzerland; and ‡Department of Chemistry, University of Basel, Basel, Switzerland

(*Received* 11 *January* 2016; *revised* 2 *May* 2016; *in final form* 8 *June* 2016)

Abstract—Our aim was to prospectively evaluate inter- and intra-observer agreement between Breast Imaging Reporting and Data System (BI-RADS) classifications and Tsukuba elasticity scores (TSs) of breast lesions. The study included 164 breast lesions (63 malignant, 101 benign). The BI-RADS classification and TS of each breast lesion was assessed by the examiner and twice by three reviewers at an interval of 2 months. Weighted κ values for inter-observer agreement ranged from moderate to substantial for BI-RADS classification (κ = 0.585–0.738) and was substantial for TS (κ = 0.608–0.779). Intra-observer agreement was almost perfect for ultrasound (US) BI-RADS (κ = 0.847–0.872) and TS (κ = 0.879–0.914). Overall, individual reviewers are highly self-consistent (almost perfect intra-observer agreement) with respect to BI-RADS classification and TS, whereas inter-observer agreement was moderate to substantial. Comprehensive training is essential for achieving high agreement and minimizing the impact of subjectivity. Our results indicate that breast US and real-time elastography can achieve high diagnostic performance. (E-mail: rosanna.zanetti@usb.ch)

Key Words: **Breast Imaging Reporting and Data System, BI-RADS, Tsukuba score, Breast ultrasound, Inter-observer agreement, Intra-observer agreement, Breast elastography.**

INTRODUCTION

In breast ultrasonography (US), lesions are classified on the basis of their sonographic features into categories 2–6 according to the Breast Imaging Reporting and Data System (BI-RADS) (American College of Radiology [ACR] 2013). Solid breast lesions are characterized according to shape (oval, round, irregular); orientation (parallel, not parallel); margins (circumscribed, not circumscribed); echo pattern (anechoic, hyper-echoic, complex cystic and solid, hypo-echoic, iso-echoic and heterogeneous); and posterior acoustic features. Benign masses are round or oval and wider than tall, with smooth, defined margins. In contrast, malignant masses tend to be irregular, with ill-defined to spiculated margins, and are taller than wide. In addition, BI-RADS classifies solid breast lesions according to their risk of malignancy. Cysts, solid lesions with unchanged or diminishing size and benign features in follow-up, as well as solid lesions with known histology, are classified as BI-RADS 2 or benign lesions. A solid breast lesion without any suspicious features is considered BI-RADS 3 or probably benign, whereas BI-RADS 4 indicates a suspicious finding and BI-RADS 5 is highly suggestive of malignancy. A known biopsy-proven malignancy is classified as BI-RADS 6. A normal sonographic breast examination without any lesions is considered BI-RADS 1. An incomplete examination or the need for additional imaging evaluation is described as BI-RADS 0.

A number of factors, however, influence BI-RADS classification. For example, considerable overlap in the US appearance between malignant and benign breast lesions introduces observer variability. More significantly, US is an operator-dependent examination because the

Address correspondence to: Rosanna Zanetti-Dällenbach, Department of Gynecology and Obstetrics, University Hospital of Basel, Spitalstrasse 21, 4031 Basel, Switzerland. E-mail: rosanna.zanetti@usb.ch

detection, description and interpretation of breast lesions are based solely on the examiner (Berg et al. 2006). Operator dependence has been investigated by determining weighted κ values of US BI-RADS final assessment between different examiners (see Table 1 and references therein). An early study by Skaane et al. (1997) reported moderate agreement between examiners in their interpretation of breast US (κ = 0.48). Subsequently, Baker et al. (1999) confirmed a moderate (κ = 0.51) inter-observer agreement. These authors also explored intra-observer agreement, which they found to be substantial (κ = 0.66).

In addition to the morphologic features seen in ultrasound, the elastic properties of soft tissues of the breast can be used as a parameter for diagnostic procedures. For example, real-time elastography (RTE) explores the difference in stiffness in benign and malignant breast lesions compared with the surrounding normal parenchyma at the macroscopic scale (Itoh et al. 2006; Sadigh et al. 2012). More recently, indentation-type atomic force microscopy has emerged as a novel diagnostic tool that allows mapping of the stiffness of unadulterated breast biopsies at nanoscale resolution (Plodinec et al. 2012).

In RTE, strain data obtained from tissue displacement produced by external compression with the probe is used to form a strain image (Itoh et al. 2006). Harder areas exhibit less tissue displacement, whereas softer areas exhibit more displacement. Strain distribution, which is inversely related to tissue stiffness, is visualized as a color-coded map that is superimposed on the B-mode image of conventional US. To standardize the interpretation of strain images, Itoh et al. (2006) developed the Tsukuba elasticity score (TS). Like the BI-RADS lexicon for breast US, TS defines parameters to standardize the interpretation of RTE. According to the Tsukuba elasticity score illustrated in Figure 1, TS1 and TS2 represent benign breast lesions, TS3 probably benign, and TS4 and TS5 malignant.

In clinics, RTE is not used as the sole examination procedure, but as a non-invasive adjunct in combination with breast US to further classify breast lesions (Wojcinski et al. 2010). However, the corresponding strain image is influenced by the compression technique of the individual examiner (Ciurea et al. 2011). Similar to US, not only data acquisition, but also interpretation of strain images, is operator dependent. Thus, TS is vulnerable to inter-observer variability. For example, Yoon at al. (2011) reported moderate (κ = 0.46) inter-observer agreement for TSs of static images. However, in real-time examination, inter-observer agreement was decreased to fair (κ = 0.28).

Overall, inter- and intra-observer variations in visual classification in breast US and RTE raise concern regarding the diagnostic performance of these procedures. Whereas a limited number of studies exploring inter- and intra-observer agreement in B-mode US exist, corresponding studies in RTE are rare. Here, we describe a prospective single-center study investigating inter- and intra-observer agreement of US BI-RADS classification and TSs in RTE.

METHODS

The single-center prospective study was approved by the institutional review board and conducted according to good clinical practice guidelines. Women with a solid breast lesion who were at least 18 y of age and scheduled for ultrasound-guided invasive breast biopsy at the outpatient breast clinic of the Women's Hospital

Table 1. Studies investigating inter- and intra-observer agreement

	κ Value			
	BI-RADS final classification		BI-RADS final classification	
Reference	Inter-observer	Intra-observer	Inter-observer	Intra-observer
Berg et al. 2006	0.52			
Lee et al. 2008	0.53	0.72–0.79		
Abdullah et al. 2009	0.30			
Calas et al. 2010		0.37–0.69		
Berg et al. 2012	0.53–0.59			
Elverici et al. 2013	0.35	0.64–0.83		
Thomas et al. 2006a			0.73	
Thomas et al. 2006b			0.86	
Yoon et al. 2011	0.37		0.28*/0.46	
Cho et al. 2011			0.587	
Cho et al. 2012			0.481	
Park et al. 2015	0.478		0.591	
Schafer et al. 2013	0.634	0.784	0.561	0.720
This study	0.585–0.738	0.847–0.872	0.608–0.779	0.879–0.914

* Real time.