Ultrasound in Med. & Biol., Vol. 40, No. 5, pp. 854–863, 2014

0301-5629/$ - see front matter

http://dx.doi.org/10.1016/j.ultrasmedbio.2013.10.005

● *Original Contributions*

DIAGNOSTIC PERFORMANCE OF ULTRASOUND AND ULTRASOUND ELASTOGRAPHY WITH RESPECT TO PHYSICIAN EXPERIENCE

Su Yeon Ko,*† Eun-Kyung Kim,* Ji Min Sung,‡ Hee Jung Moon,* and Jin Young Kwak*

*Department of Radiology, Severance Hospital, Research Institute of Radiologic Science, Yonsei University College of Medicine, Seoul, Republic of Korea; †Department of Radiology, Jeju National University Hospital, Jeju National School of Medicine, Jeju City, Republic of Korea; and ‡Department of Research Affairs, Yonsei University College of Medicine, Seoul, Republic of Korea

(*Received* 13 *May* 2013; *revised* 2 *October* 2013; *in final form* 7 *October* 2013)

Abstract—The aim of this study was to compare the diagnostic performance of gray-scale ultrasound (US), elastography and a combination of gray-scale ultrasound and elastography (US-E) in differentiating benign and malignant thyroid nodules with respect to the level of physician experience. Three hundred fifty-eight patients with 367 thyroid nodules who underwent both gray-scale US and elastography, from November 2011 to January 2012, were included in this study. The diagnostic performance of US performed by experienced and less experienced physicians was compared. Comparisons of the diagnostic performance of US, elastography and US-E were evaluated for each group separately. Of 367 nodules, 121 were malignant and 246 were benign. When we compared the diagnostic performance of the experienced and less experienced physician groups, specificity was statistically higher in the experienced physician group for both US alone ($p = 0.001$) and US-E ($p = 0.048$). However, the experienced and less experienced physician groups did not differ significantly on other measures of diagnostic performance, regardless of modality. For the experienced physicians, the specificity and positive predictive value US were 88.0% and 76.8%, respectively; both of them were significantly higher than the corresponding values for US-E. For the less experienced physicians, specificity was significantly higher on elastography (93.8%) than on US (71.4%) ($p < 0.001$). However, diagnostic performance did not differ significantly between US and US-E for the less experienced physicians. Experienced physicians had superior specificity compared with less experienced physicians. The diagnostic performance of elastography and US-E was inferior compared with that of US alone, irrespective of the level of experience of the physician. (E-mail: docjin@yuhs.ac)

Key Words: **Elastography, Gray-scale ultrasound, Level of experience, Thyroid nodule.**

INTRODUCTION

With continuing improvements in diagnostic ultrasound (US) technique, the rate of detection of thyroid nodules is increasing and is as high as 67% on US (Tan and Gharib 1997). Conventional gray-scale US is the first diagnostic method for thyroid nodules. Known suspicious US features include solid appearance, hypoechogenicity or marked hypoechogenicity, microcalcifications, taller than wide shape, microlobulated or irregular margins and abnormal Doppler flow (Kim et al. 2002; Kwak et al. 2011). Each US feature may not on its own be sufficient to predict malignancy because of variable diagnostic performance and issues with reproducibility (Choi et al. 2010; Kim et al. 2010; Moon et al. 2012b).

To overcome the inherent limitations of conventional US, elastography can be a potential adjunctive diagnostic technique for thyroid nodules. It has been reported that firm or hard consistency observed through elastography can be associated with an increased risk of malignancy (Hong et al. 2009; Lyshchik et al. 2005; Rago et al. 2007). Although elastography has been reported to evaluate the hardness or stiffness of a tissue objectively, results on the reproducibility of elastography remain controversial (Lyshchik et al. 2005; Moon et al. 2012d; Rago et al. 2007; Russ et al. 2013; Sebag et al. 2010; Shweel and Mansour 2013; Trimboli et al. 2012; Unluturk et al. 2012). The diagnostic performance of both US and elastography improves with the experience of the physician (Kim et al. 2010, 2012; Moon et al. 2012b). Therefore,

Address correspondence to: Jin Young Kwak, Department of Radiology, Research Institute of Radiologic Science, Yonsei University College of Medicine, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-752, Republic of Korea. E-mail: docjin@yuhs.ac

experience can be a critical factor affecting the accuracy of the aforementioned modalities in diagnosing thyroid nodules. However, there have been no investigations into the diagnostic performance of US and elastography in the differentiation of benign and malignant thyroid nodules with respect to the experience of the performer. Therefore, we compared the diagnostic performance of US, elastography and the combination of US and elastography (US-E) in differentiating benign from malignant thyroid nodules according to the level of experience of physicians.

METHODS

This retrospective study was approved by the institutional review board and required neither patient approval nor informed consent for review of patients' images and medical records. However, written informed consent was obtained from all patients for US-guided fine-needle aspiration (US-FNA) before each procedure as part of our hospital's daily practice.

Patient population

From November 2011 to January 2012, 459 consecutive thyroid nodules were imaged with gray-scale US and elastography and subjected to US-FNA was performed. Of these, 92 thyroid nodules with indeterminate (n = 36) or inadequate (n = 53) results at cytologic evaluation or suspicious for papillary thyroid carcinoma (n = 3) were excluded because they had not undergone surgery or repeat US-FNA. Inclusion criteria were as follows: (i) thyroid nodules on which thyroid surgery had been performed (n = 114); (ii) benign or malignant results at cytologic evaluation (n = 250); (iii) benign or malignant results at US-FNA or thyroid surgery after nondiagnostic cytologic results (n = 3). Finally, 367 thyroid nodules in 358 patients (80 men and 278 women, mean age: 49 y, age range: 19–83 y) were enrolled in this study.

Real-time gray-scale US

Real-time gray-scale US was performed using a 6- to 14-MHz linear array transducer (EUB-7500, Hitachi Medical, Tokyo, Japan) by one of seven board-certificated radiologists who specialized in thyroid imaging and was assigned arbitrarily according to the hospital's daily schedule. Of 367 thyroid nodules, 198 were examined by three experienced physicians (>7 y) and 169 nodules were examined by four less experienced physicians (<1 y).

Interpretations of US features (internal components, echogenicity, margins, calcifications and shape) of all thyroid nodules were prospectively recorded for clinical use together with a final assessment by the radiologist who performed the US examination. The internal component was classified as completely solid, cystic portion >50% or cystic portion ≤50%. Echogenicity was classified as hyperechogenicity, isoechogenicity, hypoechogenicity (with respect to normal thyroid parenchyma) or marked hypoechogenicity (defined as decreased echogenicity compared with the strap muscle). Margins were classified as well defined or not well defined. Calcifications were classified as microcalcifications (≤1 mm in diameter; tiny, punctuate, hyperechoic foci, either with or without acoustic shadows), macrocalcifications or no calcifications. Shape was defined as taller than wide (ratio of anteroposterior diameter to transverse diameter ≥ 1) or wider than tall. Suspicious features of thyroid nodules on US included marked hypoechogenicity, not well-defined margins, microcalcifications and taller than wide shape (Kim et al. 2002; Moon et al. 2010). All radiologists classified the thyroid nodules into two categories: positive for malignancy or negative for malignancy. Thyroid nodules that one of the suspicious features were classified as positive for malignancy. Thyroid nodules that had none of the suspicious features were classified as negative for malignancy.

Strain elastography

Elastography was performed after the conventional US examination by the same radiologist who performed the US. Before performing elastography, each of the seven radiologists had 2 mo of experience with the machine used in the study and participated in weekly thyroid imaging conferences regarding elastography images. During training, each radiologist performed more than 205 US-FNA procedures on thyroid nodules. Also, each of the three more experienced radiologists had previously performed elastography on more than 300 thyroid nodules, and each of the four less experienced radiologists had previously performed elastography on more than 100 thyroid nodules.

Elastography measurements were made using a free-hand technique, and all images were obtained in longitudinal planes. The probe was placed on the neck with light pressure exerted by light repetitive compression on the skin above the index malignant lesion; the level of pressure was maintained constant throughout the examination. For elastography acquisition, a square region of interest was set with the thyroid nodule positioned at the center of the box by moving the transducer. Elastography was displayed over the gray-scale image on a color scale from red (no strain, hardest components) to blue (greatest elastic strain, softest components) (Moon et al. 2012d).

Elastographic images were classified according to Rago et al. (2007) on a scale of 1 to 5, where 1 = even strain in the whole nodule; 2 = strain in a large part of the nodule; 3 = strain at the periphery of the lesion,