Medicine®

DIAGNOSTIC ACCURACY STUDY

OPEN

Acoustic Radiation Force Impulse Elastography: A Useful Tool for Differential Diagnosis of Thyroid Nodules and Recommending Fine-Needle Aspiration

A Diagnostic Accuracy Study

Yi-Feng Zhang, MD, Jun-Mei Xu, MD, Hui-Xiong Xu, MD, PhD, Chang Liu, MD, Xiao-Wan Bo, MD, Xiao-Long Li, MD, Le-Hang Guo, MD, Bo-Ji Liu, MD, Lin-Na Liu, MD, and Xiao-Hong Xu, MD

Abstract: To investigate the diagnostic performance of combined use of conventional ultrasound (US) and elastography, including conventional strain elastography such as elasticity imaging (EI) and acoustic radiation force impulse (ARFI) elastography, and to evaluate their usefulness in recommending fine-needle aspiration (FNA).

A total of 556 pathologically proven thyroid nodules were evaluated by US, EI, and ARFI examinations in this study. Three blinded readers scored the likelihood of malignancy for 4 datasets (ie, US alone, US and EI, US and virtual touch tissue imaging [VTI], and US and virtual touch tissue quantification [VTQ]). The diagnostic performances of 4 datasets in differentiating malignant from benign thyroid nodules were evaluated. The decision-making changes for FNA recommendation in the indeterminate nodules or the probably benign nodules on conventional US were evaluated after review of elastography.

The diagnostic performance in terms of area under the ROC curve did not show any change after adding EI, VTI, or VTQ for analysis; and no differences were found among different readers; however, the specificity and positive predictive value (PPV) improved significantly after adding VTI or VTQ for analysis in the senior reader. For the indeterminate nodules on US that were pathologically benign, VTQ made correct decision-making changes from FNA biopsy to follow-up in a mean of 82.6% nodules, which was significantly higher than those achieved by EI (46.8%) and VTI (54.4%) (both $P < 0.05$). With regard to the probably benign nodules on US that were pathologically malignant, EI made the highest correct decision-making change from follow-up to FNA biopsy in a mean of 62.6% nodules (compared with 41.5% on VTQ, $P < 0.05$).

The results indicated that ARFI increases the specificity and PPV in diagnosing thyroid nodules. US combined VTQ might be helpful in reducing unnecessary FNA for indeterminate nodules on US whereas US combined EI is useful to detect the false negative nodules that are probably benign on conventional US.

(*Medicine* 94(42):e1834)

Abbreviations: ARFI = acoustic radiation force impulse, BI-RADS = Breast Imaging Reporting and Data System, EI = elasticity imaging, FNA = fine-needle aspiration, NPV = negative predictive value, OR = odds ratio, PPV = positive predictive value, ROI = region of interest, SWV = shear wave velocity, US = ultrasound, VTI = virtual touch tissue imaging, VTQ = virtual touch tissue quantification.

Editor: Yanjun Gong.
Received: February 11, 2015; revised: September 21, 2015; accepted: September 23, 2015.
From the Department of Medical Ultrasound, Shanghai Tenth People's Hospital, Ultrasound Research and Education Institute, Tongji University School of Medicine, Shanghai, China (Y-FZ, J-MX, H-XX, CL, X-WB, X-LL, L-HG, B-JL, L-NL); Thyroid Institute, Tongji University School of Medicine, Shanghai, China (Y-FZ, J-MX, H-XX, CL, X-WB, X-LL, L-HG, B-JL, L-NL); and Department of Ultrasound, Guangdong Medical College Affiliated Hospital, Zhanjiang, China (H-XX, X-HX).
Correspondence: Hui-Xiong Xu, Department of Medical Ultrasound, Shanghai Tenth People's Hospital, Ultrasound Research and Education Institute, Tongji University School of Medicine, No. 301, Yanchangzhong Road, Shanghai 200072, China (e-mail: xuhuixiong@126.com).
This work was supported in part by Grant SHDC12014229 from Shanghai Hospital Development Center, Grants 14441900900 and 15411969000 from Science and Technology Commission of Shanghai Municipality, Grant 20144Y0148 from Shanghai Municipal Commission of Health and Family Planning, and Grants 81401417 and 81472579 from the National Natural Science Foundation of China.
The authors have no conflicts of interest to disclose.
Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.
This is an open access article distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0, where it is permissible to download, share and reproduce the work in any medium, provided it is properly cited. The work cannot be changed in any way or used commercially.
ISSN: 0025-7974
DOI: 10.1097/MD.0000000000001834

INTRODUCTION

Thyroid cancer has been documented as the most common endocrine malignancy with a rapid worldwide rise in incidence in the past few decades.[1–3] The main goal in the assessment of patients with thyroid nodules is to distinguish thyroid cancers from benign nodules. High-frequency ultrasound (US) is a highly accurate and sensitive method for detecting thyroid nodules; however, conventional US becomes highly predictive of malignancy only when multiple US patterns are simultaneously present in the same thyroid nodule.[4] When the US features are atypical, an assisted diagnostic tool is needed.

Conventional strain elastography has been confirmed to be a useful tool for differentiating benign thyroid nodules from malignant ones, which provides an estimation of tissue stiffness by measuring the degree of strain under the application of an external force. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) of combined use of US and conventional strain elastography for diagnosing thyroid malignancy were 41% to 98%, 34% to 94.8%, 22.3% to 86.3%, and 83% to 100%, respectively.[5–10] The variety in diagnostic performance for conventional strain elastography is largely attributed to its limitations such as unstable image quality because of different external forces applied, unreliable interobserver agreement, and lack of quantitative information.[11]

Compared with conventional strain elastography using quasi-static methods such as manual compression or cardiovascular/respiratory pulsation inducing strain, a new method called acoustic radiation force impulse (ARFI) has gained increasing attention in recent years, in which the strain and shear wave are excited by ARFI with a fixed transmit frequency. ARFI elastography includes measurement of strain/displacement (ie, Virtual Touch^{TM} Imaging, VTI) and measurement of shear wave velocity (SWV) (ie, Virtual Touch^{TM} Quantification, VTQ). VTI refers to measurement of tissue displacement or related physical quantity within the ARFI push region whereas VTQ refers to measurement of average SWV in a region of interest. Theoretically, ARFI can overcome the limitations of conventional strain elastography, including the different forces given by different examiners and lack of quantitative measurement of the tissue stiffness. Park et al[11] pointed out that conventional strain elastography did not show reliable interobserver agreement for the diagnosis of malignant thyroid nodules. Conversely, other studies showed that VTQ of ARFI had very good intraobserver and interobserver agreement.[12]

Fine-needle aspiration (FNA) biopsy is the most accurate method to distinguish benign thyroid nodules from malignant ones. However, due to high prevalence of thyroid nodule, it is not possible to perform FNA for all or even most of the thyroid nodules. Therefore, applying FNA to suspicious nodules and decreasing the risk of overlooking malignant nodules not submitted to FNA is a paramount issue in clinical practice. Although the diagnostic values of ARFI in differentiating benign from malignant thyroid nodules have been reported, the usefulness of ARFI in risk stratification of thyroid nodules and recommendation for FNA or just follow-up is still unknown.

The purpose of this study was to investigate the diagnostic performance of combination use of conventional US and elastography (including strain elastography and ARFI) in characterization of thyroid nodules and to evaluate their usefulness in decision making for the nodules.

MATERIALS AND METHODS

Patients

This prospective study was approved by the institutional review board and informed consent was obtained from all the patients to include their data for scientific analysis.

From May 2011 to March 2014, 867 consecutive patients with thyroid nodules were detected on conventional US and were referred to both conventional strain elastography and ARFI examinations in the institution. The inclusion criteria for the target nodules were as follows: solid or predominant solid (<25% cystic) nodules; nodules >0.5 cm in diameter; with enough thyroid tissue surrounding the nodule at the same depth. Finally, 377 patients were excluded for the following reasons: pathologic diagnoses were not available (n = 327); US or elastography images were unclear or incomplete (n = 50). The remaining 490 patients with 556 nodules underwent US, conventional strain elastography, ARFI (including VTI and VTQ) elastography examinations and the nature of all the target nodules were pathologically confirmed after surgery. They were 361 female and 129 male patients, with a mean age of 50.7 ± 11.8 years (range, 16–78 years). The mean age was 50.4 ± 12.0 years for female patients and 51.5 ± 11.1 years for male patients. The diameter of the nodules ranged from 0.5 to 6.2 cm (mean, 1.8 ± 1.1 cm). One hundred forty-six patients had single nodule and 344 had multiple nodules. For the patients with multiple nodules, the nodules that were suspicious of malignancy (based on conventional US findings) otherwise the largest solid ones were selected for analysis.

All the nodules in this study underwent surgery. The indications for thyroidectomy were as follows: confirmed malignancy by FNA (n = 105); highly suspicious of malignancy by conventional US (n = 25) or FNA (n = 78); nondiagnostic (n = 63) or indeterminate (n = 84) results by FNA; patient anxiety (n = 33); compressive symptom caused by the large nodules or the associated large nodules (n = 168).

On-Site US Examination

Conventional US, conventional strain elastography, VTI, and VTQ examinations were performed by 1 of 2 radiologists with at least 7 years of experience in thyroid US and at least 2 years of experience in thyroid elastography, with the same S2000 US machine (Siemens Medical Solutions, Mountain View, CA) and the 9L4-linear transducer (frequency range, 4–9 MHz).

All patients underwent conventional US examination firstly, including B mode US and color Doppler US. Afterwards, conventional strain elastography and ARFI elastography were applied. The following US features were evaluated on-site and were recorded: echogenicity, halo sign, calcifications, shape, and margin. Echogenicity was classified as hyper-, iso-, hypo-, or mixed echogenicity (compared with the normal thyroid gland). Margin was classified as well defined or poorly defined. Calcifications were classified as microcalcifications (≤1.5 mm in diameter; tiny, punctate, hyperechoic foci, without acoustic shadow), macrocalcification (>1.5 mm in diameter, with acoustic shadows), or no calcification. Shape was classified as ovoid to round, taller than wide, or irregular. Color-flow Doppler US patterns were defined as absence of blood flow (type I), perinodular and absent or slight intranodular blood flow (type II), and marked intranodular and absent or slight perinodular blood flow (type III).[13]

Conventional strain elastography (ie, elastic imaging, EI) was carried out after conventional US examination by the same operator on the longitudinal axis of the target nodule. EI image was displayed through color scale according to the degree of displacement for all pixels within the region of interest (ROI), with a scale from red (greatest strain, softest component) to green (average strain, intermediate component) and to blue (no strain, hardest component). The EI images were classified using the Itoh strain elastographic score.[14]

ARFI elastography was performed following EI examination. ARFI elastography involves VTI and VTQ. When VTI was performed, the ROI box was adjusted to include the nodule with adequate surrounding thyroid tissue. VTI is displayed as gray-scale image over the conventional B-mode image, in which the black means hard tissue whereas the white means soft tissue. The VTI images were scored according to Xu's VTI scoring system.[15] Score 1, the nodule is displayed as predominantly bright; Score 2, the nodule is displayed as predominantly bright with few dark portions; Score 3, the nodule is displayed as equally dark and bright; Score 4, the nodule is displayed as predominantly dark with a few bright spots; Score 5, the nodule is displayed as almost completely dark; and Score 6, the nodule is displayed as completely dark without bright spots.

VTQ was then performed following VTI. The sampling ROI for VTQ was placed on the solid portion of the nodule and the calcified or cystic portions were avoided. The size of the sampling ROI for VTQ is 6 mm × 5 mm, which cannot be altered and thus only suitable for nodules ≥5 mm in diameter. The adjacent thyroid tissue was not included in ROI. When VTQ of ARFI was initiated, the SWVs of the tissue in the ROI were calculated and shown on the screen. Seven consecutive measurements for each nodule were performed. The median of all 7 measurements was used for further analysis.

 Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

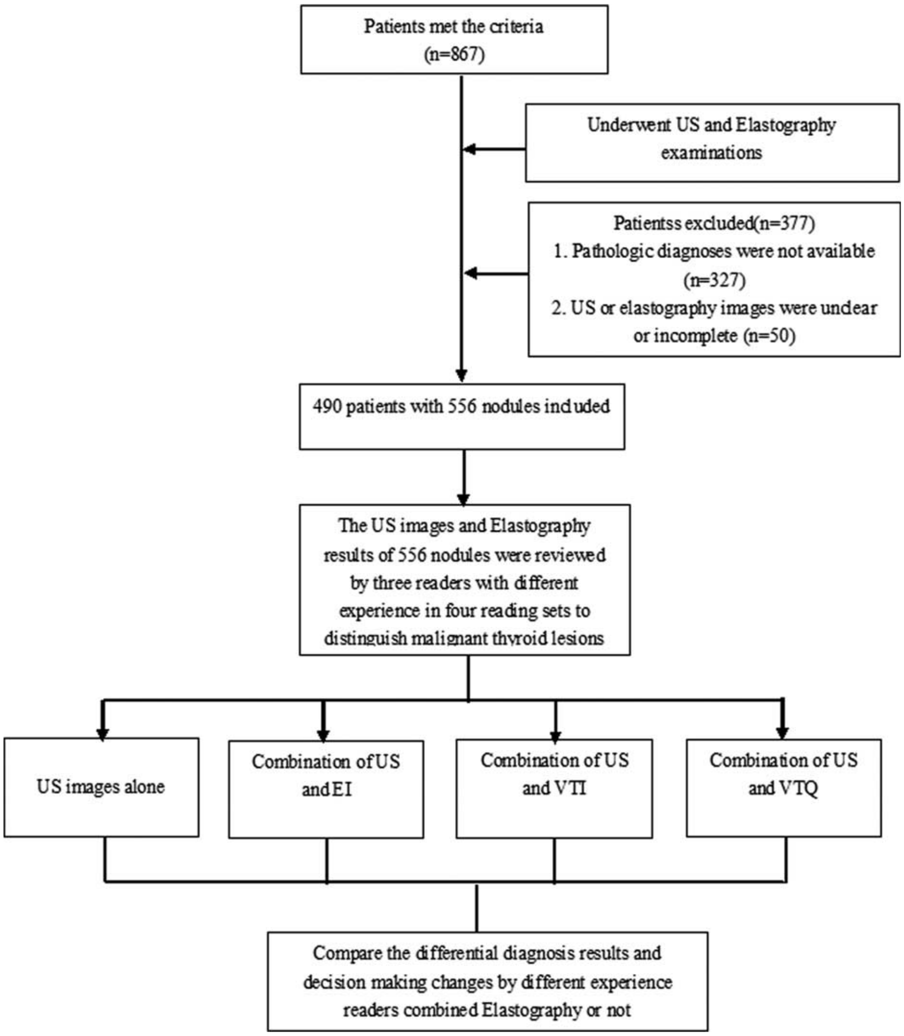

FIGURE 1. Flowchart of patient selection and the design of the study.

The imaging results of conventional US, EI, VTI, and the median values of VTQ measurements were evaluated on-site and recorded by the same radiologist who performed the examinations. At the same time, all the digital images and cine clips of conventional US, EI, VTI, and VTQ were stored in the machine and were transferred to a personal computer for further off-site analysis.

Off-Site Reading Procedures

Three readers who did not participate the US and elastography examinations analyzed the images independently without knowledge of patient identification, relevant clinical information, histopathology results, and other imaging results. Readers 1, 2, 3 had 17, 10, 6 years of experience in thyroid US and 5, 3, 1 years of experience in thyroid elastography, respectively, who represented readers with 3 different experience levels (ie, senior, medium, junior). Digital images and cine clips of conventional US, EI, VTI, and VTQ results were retrieved and presented randomly so that the benign and malignant cases were not grouped by diagnosis, and any identifying information (ie, the sequence number, name, sex, and age of each patient) was concealed.

Four datasets, including conventional US images alone, combination of conventional US and EI, combination of conventional US and VTI, and combination of conventional US and VTQ, were reviewed. The 4-step reading sessions were separated by at least 2 weeks, with cases in a random order to reduce memory effect. In each reading session, the 3 readers were asked to review the images of each nodule, then a confidence rating score for each nodule was assigned on the basis of a 5-point scale (1, definitely benign; 2, probably benign; 3, indeterminate; 4, probably malignant; 5, definitely malignant). If the score was 1 or 2, the nodule was regarded as needing follow-up; and if the score was 3 or higher, the nodule was regarded as needing FNA. The readers were not allowed to change the scores for the prior reading session.

The diagnostic criteria on conventional US were as follows: for benign nodules: no intranodular calcifications, hyper- or iso-echogenicity, a spongiform appearance, regular margin, halo sign, perinodular and absent or slight intranodular blood flow; for malignant nodules: solid lesion, intranodular

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

TABLE 1. Comparisons of Conventional Ultrasound and Elastography Features Between Benign and Malignant Nodules

Characteristics	Benign (n = 368)	Malignant (n = 188)	*P*
Gray-scale US characteristics			
Echogenicity			<0.001*
Hyperechoic	2 (0.5)	1 (0.5)	
Isoechoic	124 (33.7)	10 (5.3)	
Hypoechoic	149 (40.5)	171 (91.0)	
Mixed	93 (25.3)	6 (3.2)	
Calcifications			<0.001*
None	290 (78.8)	56 (29.8)	
Microcalifications	41 (11.1)	110 (58.5)	
Macrocalcifications	37 (10.1)	22 (11.7)	
Shape			<0.001*
Ovoid to round	335 (91.0)	101 (53.7)	
Taller than wide	17 (4.6)	68 (36.2)	
Irregular	16 (4.4)	19 (10.1)	
Margin			<0.001*
Well defined	289 (78.5)	103 (54.8)	
Ill defined	79 (21.5)	85 (45.2)	
Halo sign			<0.001*
Present	123 (33.4)	6 (3.2)	
Absent	245 (66.6)	182 (96.8)	
Vascularity			<0.001*
Type I	98 (26.6)	101 (53.7)	
Type II	213 (57.9)	23 (12.2)	
Type III	57 (15.5)	64 (34.1)	
Elastography features			
EI score			<0.001*
Score 1	9 (2.4)	1 (0.5)	
Score 2	35 (9.5)	7 (3.7)	
Score 3	231 (62.8)	101 (53.7)	
Score 4	80 (21.8)	49 (26.1)	
Score 5	13 (3.5)	30 (16.0)	
VTI score			<0.001*
Score 1	12 (3.3)	2 (1.1)	
Score 2	165 (44.8)	31 (16.5)	
Score 3	144 (39.1)	39 (20.7)	
Score 4	40 (10.9)	72 (38.3)	
Score 5	6 (1.6)	37 (19.7)	
Score 6	1 (0.3)	7 (3.7)	
VTQ			<0.001*
SWV (m/sec)	2.26 ± 1.06 (95% CI: 2.15–2.36)	3.92 ± 2.45 (95% CI: 3.60–4.28)	
Range of SWV(m/sec)	0.50–8.40	0.99–8.40	

Numbers in parentheses are percentages, otherwise 95% CI.
CI = confidence interval; EI = elasticity imaging; SWV = shear wave velocity; US = ultrasound; VTI = virtual touch tissue imaging; VTQ = virtual touch tissue quantification.
* There are significant differences between benign and malignant nodules ($P < 0.05$).

microcalcification, hypo- or marked hypo-echogenicity, irregular margin, absence of halo sign, taller-than-wide shape, intranodular and absent or slight perinodular blood flow.[16] The diagnostic criteria on EI were that score 4 and score 5 were regarded as malignant nodule, whereas score 1 to score 3 as benign according to the Itoh strain elastographic score system.[17] The diagnostic criteria on VTI were referred to Xu's VTI scoring system.[15] VTI score 1 to 3 was regarded as benign nodules whereas VTI score 4 to 6 as malignant nodules. As to VTQ, SWV value ≤2.87 m/sec was regarded as benign nodules whereas SWV value >2.87 m/sec as malignant.[12] The flow diagram of recruitment of patients and the protocol of this study was shown in Figure 1.

Impact of Elastography on Decision Making

In this study, the confidence rating score 3 or higher indicated malignancy and thus needing biopsy. On the other hand, the confidence rating score 2 or lower indicated benign and thus needing follow-up. Elastography techniques such as EI, VTI, and VTQ were applied to score 2 and 3 nodules to

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

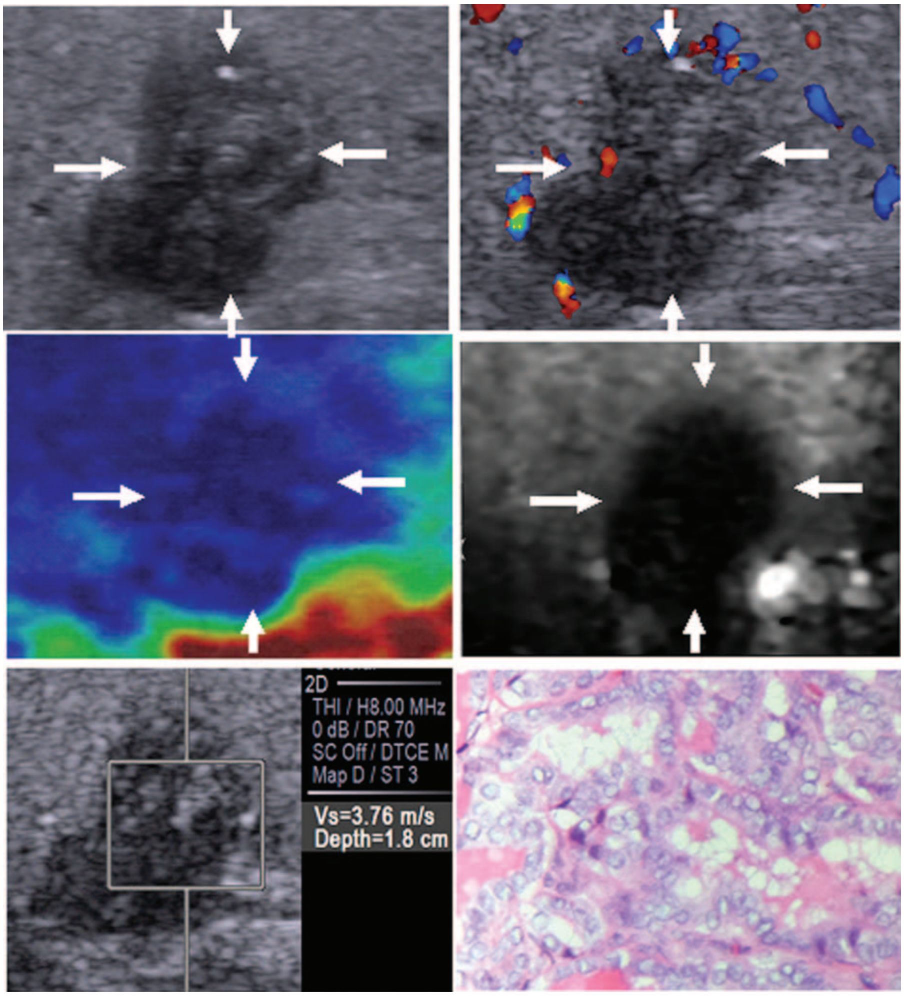

FIGURE 2. A 56-year-old man with papillary thyroid microcarcinoma. Arrows in each picture point out the contour of the lesions. (A) Conventional US shows hypoechogenicity, taller-than-wide shape, microcalifications, and ill-defined margin of the nodule. (B) Color Doppler flow imaging shows absence of blood flow in the nodule. (C) Score 4 is assigned at EI. (D) Score 5 is assigned at VTI. (E) The SWV of the nodule is 3.76 m/sec. (F) Histology of the lesion confirms the diagnosis of papillary thyroid microcarcinoma. Hematoxylin and eosin stain, ×100. EI = elasticity imaging; SWV = shear wave velocity; US = ultrasound; VTI = virtual touch tissue imaging.

upgrade or downgrade lesions. However, the upgrading or downgrading of nodules was not mandatory and was not applied to nodules with definitely benign (score 1) or probably and definitely malignant (score 4 or 5) since the treatment strategy for them has no controversy in clinical practice that follow-up is usually applied to score 1 nodules whereas FNA is applied to score 4 or 5 nodules.

STATISTICAL ANALYSIS

Reader agreement among the 3 radiologists in classifying conventional US, EI, and VTI images was estimated using the kappa analysis. The agreements were displayed as kappa value ± standard error. Agreement was graded as poor ($\kappa < 0.20$), moderate ($\kappa = 0.20$ to <0.40), fair ($\kappa = 0.40$ to <0.60), good ($\kappa = 0.60$ to <0.80), or very good ($\kappa = 0.80$–1.00). The ratios of gray-scale US features between benign and malignant thyroid nodules were compared by χ^2 statistics. If there were cells with fewer than 5 observations, the Fisher's exact test was used. Receiver operating characteristic (ROC) curve analysis was performed to evaluate the diagnostic performance (ie, area under the ROC curve, A_Z) regarding the 4 imaging datasets in distinguishing benign from malignant thyroid nodules. For each reader, the comparisons of A_Z among the 4 datasets were evaluated by univariate Z score test.

The nodules assigned a confidence score of 3, 4, or 5 were defined as positive results (ie, presence of malignancy or needing biopsy) and those assigned a confidence score of 2 or less were defined as negative results (ie, absence of malignancy or needing follow-up). Based on these decisions, differences in sensitivity, specificity, accuracy, PPV, and NPV were tested using the McNemar test or chi-squared test. The rates of decision changes were compared by χ^2 statistics. $P < 0.05$ were considered to be statistically significant. All the statistical analyses were carried out using the SPSS18.0 software package (version 18.0; SPSS, Inc., Chicago, IL).

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

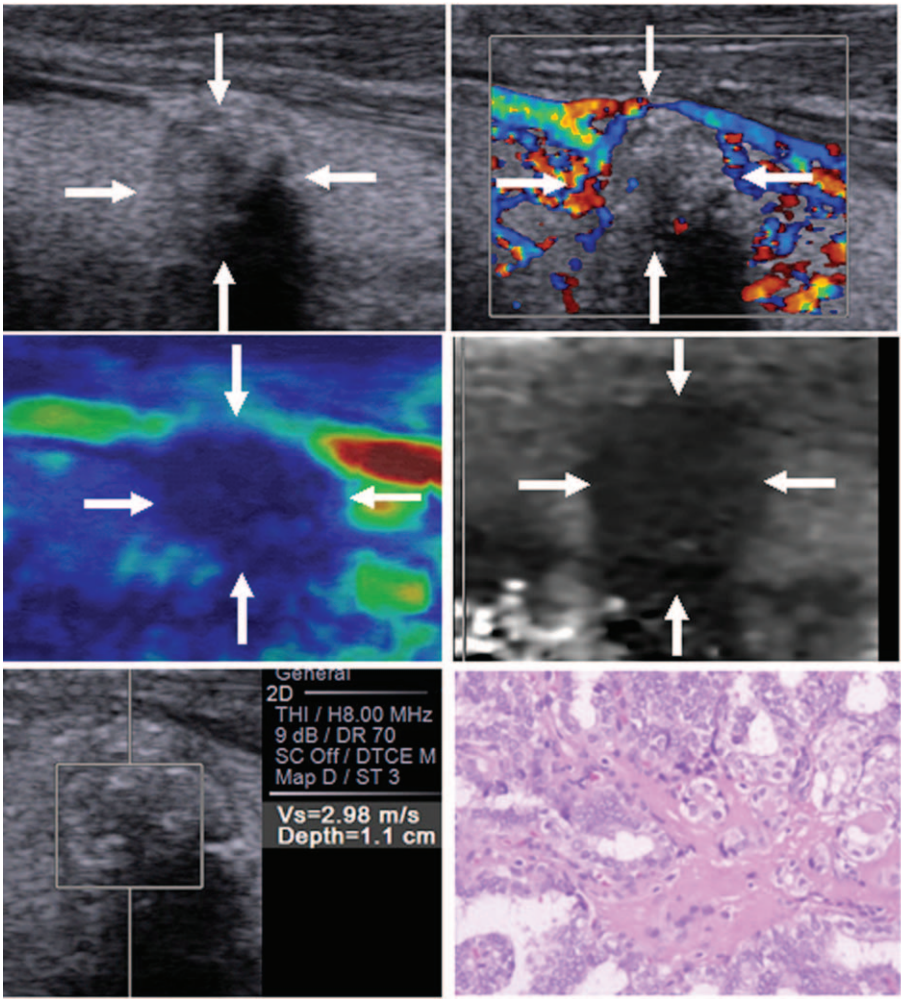

FIGURE 3. A 48-year-old man with papillary thyroid carcinoma. Arrows in each picture point out the contour of the lesions. (A) Conventional US shows isoechogenicity, ovoid to round shape, macrocalifications, and ill-defined margin of the nodule. (B) Color Doppler flow imaging shows slight intranodular flow. (C) Score 4 is assigned at EI. (D) Score 4 is assigned at VTI. (E) The SWV of the nodule is 2.98 m/sec. (F) Histology of the lesion confirms the diagnosis of papillary thyroid carcinoma. Hematoxylin and eosin stain, ×100. EI = elasticity imaging; SWV = shear wave velocity; US = ultrasound; VTI = virtual touch tissue imaging.

RESULTS

Histopathological Results

There were 368 benign nodules and 188 malignant nodules. Among the benign nodules, there were 294 nodular goiters, 39 Hashimoto's nodules, and 35 adenomas. Among the malignant nodules, there were 182 papillary thyroid carcinomas, 3 follicular thyroid carcinomas, 2 lymphomas, and 1 medullary thyroid carcinoma.

Reader Agreement

The interobserver agreement was good (reader 1 vs reader 2: $\kappa = 0.67 \pm 0.03$; reader 1 vs reader 3: $\kappa = 0.66 \pm 0.03$; reader 2 vs reader 3: $\kappa = 0.63 \pm 0.03$) for conventional US, whereas fair (reader 1 vs reader 2: $\kappa = 0.58 \pm 0.04$; reader 1 vs reader 3: $\kappa = 0.56 \pm 0.03$; reader 2 vs reader 3: $\kappa = 0.52 \pm 0.03$) for VTI. On the other hand, the interobserver agreement for EI was good for reader 1 vs reader 2 ($\kappa = 0.65 \pm 0.03$) and reader 1 vs reader 3 ($\kappa = 0.63 \pm 0.03$), whereas fair for reader 2 vs reader 3 ($\kappa = 0.59 \pm 0.03$).

Basic Characteristics and US and Elastography Features

Significant differences were found in patient gender (male/female) (96/226 for benign vs 33/135 for malignant, $P = 0.015$), age (52.4 ± 11.3 years for benign vs 47.4 ± 12.0 years for malignant, $P < 0.001$); however, the difference between the ratios of patients with solitary or multiple nodules (94/228 for benign vs 52/116 for malignant) was not significant ($P = 0.686$). Significant difference was found in nodule diameter between benign (20.7 ± 11.3 mm) and malignant nodules (13.3 ± 8.9 mm) ($P < 0.001$).

As for conventional US, significant differences were found in US features such as echogenicity, calcification, shape, margin, halo sign, and vascularity type, between benign and

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

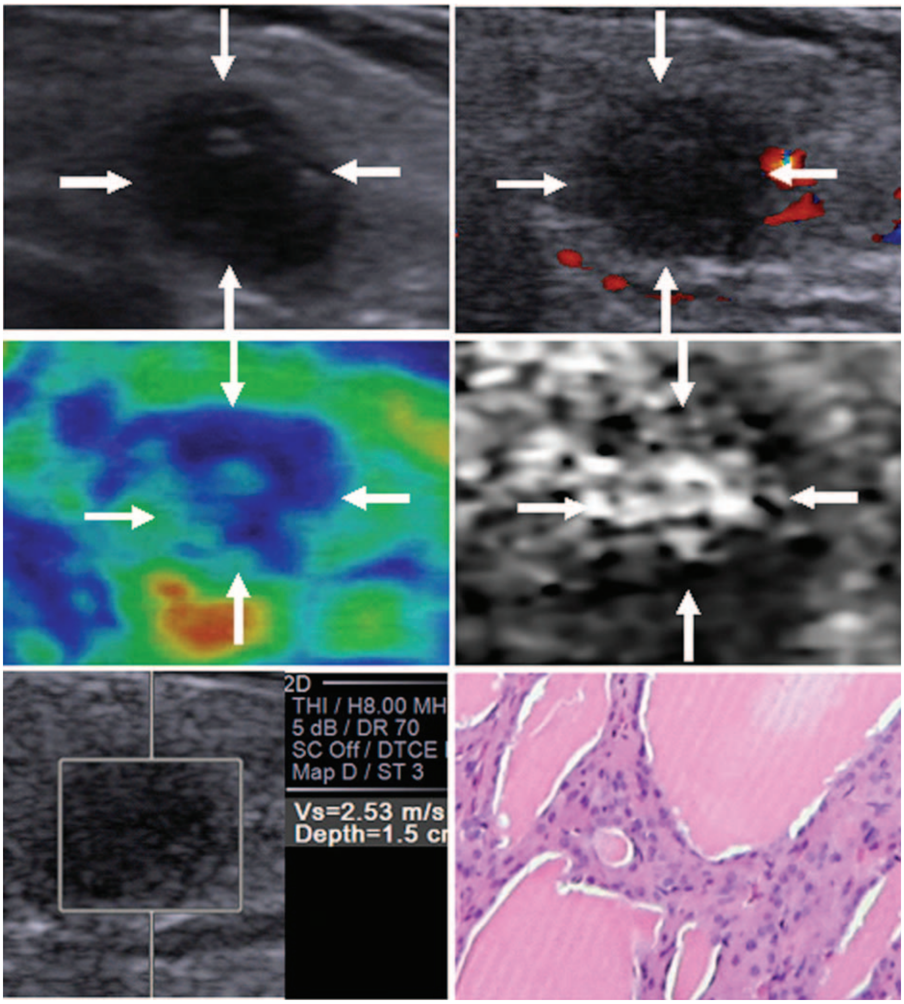

FIGURE 4. A 50-year-old woman with nodular goiter. Arrows in each picture point out the contour of the lesions. (A) Conventional US shows marked hypoechogenicity, ovoid to round shape, and ill-defined margin of the nodule. (B) Color Doppler flow imaging shows absent blood flow in the nodule. (C) Score 2 is assigned at EI. (D) Score 2 is assigned at VTI. (E) The SWV of the nodule is 2.53 m/sec. (F) Histology of the lesion confirms the diagnosis of nodular goiter. Hematoxylin and eosin stain, ×40. EI = elasticity imaging; SWV = shear wave velocity; US = ultrasound; VTI = virtual touch tissue imaging.

malignant nodules (all $P < 0.01$). With regard to elastography, significant differences were found in EI and VTI between benign and malignant nodules. The SWV of malignant nodules was significant higher than that of benign nodules ($P < 0.01$) (Table 1 and Figures 2–4). The US and elastography features were evaluated by the on-site investigators who performed the examinations.

Diagnostic Performance

The diagnostic performance in terms of Az did not show any change after adding EI, VTI, or VTQ for analysis; and no differences were found among different readers (Table2). With regard to the comparison among different datasets, the sensitivity and NPV did not show improvement after combined elastography. The sensitivity decreased in reader 1 (91.5% vs 79.8%) and reader 2 (88.8% vs 75.0%) significantly after adding the results of VTI, and in reader 3 after adding the results of VTQ (93.6% vs 86.7%) (all $P < 0.05$). Lower NPV was also found in reader 1 (94.3% vs 88.8%) and reader 2 (93.1% vs 85.4%) after adding the results of VTI (both $P < 0.05$).

The specificity achieved significant improvement after adding the results of VTI (82.1% vs 71.5%) or VTQ (92.4% vs 71.5%) for analysis in reader 1 (both $P < 0.05$) and after combined VTQ (69.6% vs 59.2%) in reader 3 ($P < 0.05$). But at the same time, the specificity decreased in reader 2 after adding the results of EI (77.2% vs 68.8%) ($P < 0.05$). PPV significantly increased after combined VTQ in reader 1 (85.1% vs 62.1%, $P < 0.001$). As for the comparison among different readers, the specificity, accuracy, and PPV of senior reader were found to be significantly higher than those of junior readers under combined datasets of conventional US and VTI or VTQ (all $P < 0.05$).

Decision-Making Change

As for the indeterminate nodules on US (score 3) that were pathologically benign, VTQ made correct decision-making changes (FNA biopsy to follow-up) in 82.6% nodules, which was significantly higher than those achieved by EI (46.8%) and

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

TABLE 2. Diagnostic Performance in Differentiating Malignant From Benign Thyroid Nodules Before and After Review of Elastography Results

Performance Parameters	Experience in Thyroid US, yr	Datasets						
		US	US + EI	$P^†$	US + VTI	$P^†$	US + VTQ	$P^†$
AUC								
Reader 1	17	0.887* (0.858–0.916)	0.793* (0.756–0.831)	>0.05	0.885* (0.857–0.913)	>0.05	0.907* (0.877–0.936)	>0.05
Reader 2	10	0.886* (0.856–0.916)	0.811* (0.773–0.80)	>0.05	0.788* (0.740–0.831)	>0.05	0.886* (0.856–0.915)	>0.05
Reader 3	6	0.823* (0.788–0.857)	0.789* (0.752–0.827)	>0.05	0.828* (0.794–0.823)	>0.05	0.849* (0.817–0.881)	>0.05
Sensitivity, %								
Reader 1	17	91.5 (172/188)	86.7 (163/188)	0.136	79.8 (150/188)	0.001	85.1 (160/188)	0.054
Reader 2	10	88.8 (167/188)	86.2 (162/188)	0.436	75.0 (141/188)	<0.001	89.9 (169/188)	0.738
Reader 3	6	93.6 (176/188)	93.1 (175/188)	0.836	93.6 (176/188)	1.000	86.7 (163/188)	0.024
Specificity, %								
Reader 1	17	71.5 (263/368)	65.5 (241/368)	0.081	82.1 (302/368)	0.001	92.4 (340/368)	0.000
Reader 2	10	77.2 (284/368)	68.8 (253/368)	0.010	74.5 (274/368)‡	0.389	76.9 (283/368)‡	0.930
Reader 3	6	59.2 (218/368)	57.6 (212/368)	0.654	57.1 (210/368)‡,§	0.550	69.6 (256/368)‡,§	0.003
PPV, %								
Reader 1	17	62.1 (172/277)	56.2 (163/290)	0.154	69.4 (150/216)	0.089	85.1 (160/188)†	<0.001
Reader 2	10	66.5 (167/251)	58.5 (162/277)	0.057	60.0 (141/235)‡	0.135	66.5 (169/254)‡	1.000
Reader 3	6	54.0 (176/326)	52.9 (175/331)	0.774	52.7 (176/334)‡	0.739	59.3 (163/275)‡	0.193
NPV, %								
Reader 1	17	94.3 (263/279)	90.6 (241/266)	0.105	88.8 (302/340)	0.017	92.4(340/368)	0.348
Reader 2	10	93.1 (284/305)	90.7 (253/279)	0.280	85.4 (274/321)	0.002	93.7 (283/302)	0.768
Reader 3	6	94.8 (218/230)	94.2 (212/225)	0.793	94.6 (210/222)	0.929	91.1 (256/281)	0.110

* Data in parenthesis indicate 95% confidence interval.
† P values indicate comparison between US only.
‡ Compared with Reader 1 in the same dataset, there are significant differences (all $P < 0.05$).
§ Compared with Reader 2 in the same dataset, there are significant differences (all $P < 0.05$).AUC = area under curve; EI = elasticity imaging; NPV = negative predictive value; PPV = positive predictive value; US = ultrasound; VTI = virtual touch tissue imaging; VTQ = virtual touch tissue quantification.

VTI (54.4%) (both $P < 0.05$). However, EI achieved the lowest incorrect decision-making changes (FNA biopsy to follow-up) in 19.0% indeterminate nodules (score 3) that were pathologically malignant, in comparison with VTI (30.3%) or VTQ (35.2%) (Table 3).

With regard to the probably benign lesions on US (score 2) that were pathologically malignant, EI made the highest correct decision-making changes (follow-up to biopsy) in 62.6% nodules, and VTQ made the lowest incorrect decision-making changes (follow-up to biopsy) in 9.7% probably benign lesions (score 2) that were pathologically benign (compared with EI and VTI, $P < 0.05$) (Table 4).

DISCUSSION

In this study, the diagnostic performance in terms of Az for each dataset of US combined elastography in all readers did not achieve significant improvement when compared with conventional US alone. The results indicated that although elastography, including conventional strain elastography and ARFI, was supposed to be useful for differentiating malignant from benign thyroid nodules, the overall performance of elastography was similar with that of conventional US. Therefore, US combined elastography did not improve the overall diagnostic performance significantly. Some authors even concluded that the combination of elastography and gray-scale US showed inferior performance in the differentiation of malignant and benign thyroid nodules compared with gray-scale US.[18] In their study, the sensitivity, NPV, and odd ratio (OR) for gray-scale US were 91.7%, 94.7%, and 22.1, respectively, higher than the 15.7% to 65.4% sensitivity, 71.7% to 79.1% NPV, and 2.6 to 3.7 ORs for conventional strain elastography.[18]

After further analysis, in reader 1 (ie, the senior reader), the specificity achieved significant improvement after adding the images of VTI or VTQ for analysis and the accuracy and PPV also increased significantly after adding VTQ images for analysis. The increase of specificity and PPV is meaningful, which indicates ARFI technique is helpful to make definite diagnosis for both malignant and benign nodules. As for the comparisons between different readers, there were no significant differences in overall diagnostic performance among the 3 readers after adding elastography for analysis; however, the specificity, accuracy, and PPV for senior reader were higher than those for junior readers after adding the VTI or VTQ images for analysis. These findings indicated that if radiologists received adequate training for ARFI, US combined ARFI would be helpful for differential diagnosis of thyroid nodules in comparison with conventional US only although a learning curve is present.

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

TABLE 3. Decision-Making Change for Score 3 Nodules on Conventional US After Review of Elastography

	No. of Score 3 Nodules on Conventional US			After Review of EI		After Review of VTI		After Review of VTQ	
	Total, n	Pathologically Benign, n	Pathologically Malignant, n	Correct Change for Pathologically Benign Nodules (B–F)	Incorrect Change for Pathologically Malignant Nodules (B–F)	Correct Change for Pathologically Benign Nodules (B–F)	Incorrect Change for Pathologically Malignant Nodules (B–F)	Correct Change for Pathologically Benign Nodules (B–F)	Incorrect Change for Pathologically Malignant Nodules (B–F)
Reader 1	124	77	47	55.8% (43/77)	25.5% (12/47)	68.8% (53/77)	44.7% (21/47)	100.0% (77/77)	27.7% (13/47)
Reader 2	74	50	24	48.0% (24/50)	16.7% (4/24)	50.0% (25/50)	29.2% (7/24)	72.0% (36/50)	33.3% (8/24)
Reader 3	126	79	47	36.7% (29/79)	14.9% (7/47)	44.3% (35/79)	17.0% (8/47)	75.9% (60/79)	44.7% (21/47)
Mean	108	69	39	46.8%	19.0%*	54.4%	30.3%	82.6%†	35.2%

Score 3 nodules on conventional US should be recommended to FNA biopsy, the number of which is depending on different readers. After review of elastography such as EI, VTI, or VTQ, however, the recommendation might be changed to follow-up. For pathologically benign score 3 nodules, the change from FNA biopsy to follow-up is correct. On the other hand, for pathologically malignant score 3 nodules, the change from FNA biopsy to follow-up is incorrect. The numerators in parentheses come from the number of score 3 lesions changed from FNA biopsy to follow-up after combined EI, VTI, or VTQ.

B = FNA biopsy; EI = elasticity imaging; F = follow-up; FNA = fine-needle aspiration; US = ultrasound; VTI = virtual touch tissue imaging; VTQ = virtual touch tissue quantification.

* In comparison with VTQ, there is significant difference ($P < 0.05$).

† In comparison with EI or VTI, there are significant differences (both $P < 0.05$).

TABLE 4. Decision-Making Change for Score 2 Nodules on Conventional US After Review of Elastography

	No. of Score 2 Nodules on Conventional US			After Review of EI		After Review of VTI		After Review of VTQ	
	Total, n	Pathologically Benign, n	Pathologically Malignant, n	Incorrect Change for Pathologically Benign Nodule s (F–B)	Correct Change for Pathologically Malignant Nodules (F–B)	Incorrect Change for Pathologically Benign Nodules (F–B)	Correct Change for Pathologically Malignant Nodules (F–B)	Incorrect Change for Pathologically Benign Nodules (F–B)	Correct Change for Pathologically Malignant Nodules (F–B)
Reader 1	151	138	13	24.6% (34/138)	53.8% (7/13)	13.0% (18/138)	46.2% (6/13)	0% (0/138)	7.7% (1/13)
Reader 2	250	228	22	23.7% (54/228)	59.1% (13/22)	18.4% (42/228)	50.0% (11/22)	16.2% (37/228)	50.0% (11/22)
Reader 3	184	172	12	20.9% (36/172)	75.0% (9/12)	32.0% (55/172)	66.7% (8/12)	12.8% (22/172)	66.7% (8/12)
Mean	195	179	16	23.1%	62.6%*	21.1%	54.3%	9.7%†	41.5%

Score 2 nodules on conventional US should be recommended to follow-up, the number of which is depending on different readers. After review of elastography such as EI, VTI, or VTQ, however, the recommendation might be changed to FNA biopsy. For pathologically benign score 2 nodules, the change from follow-up to FNA biopsy is incorrect. On the other hand, for pathologically malignant score 2 nodules, the change from follow-up to FNA biopsy is correct. The numerators in parentheses come from the number of score 2 lesions changed from follow-up to FNA biopsy after combined EI, VTI, or VTQ.

F = follow-up; B = FNA biopsy; EI = elasticity imaging; FNA = fine-needle aspiration; US = ultrasound; VTI = virtual touch tissue imaging; VTQ = virtual touch tissue quantification.

* Compared with VTQ, there is significant difference ($P < 0.05$).

† Compared with EI and VTI, there are significant differences (both $P < 0.05$).

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

Elastography might be a useful tool for decision making in clinical practice. Cho et al[19] found that combined use of strain elastography and conventional US increased the specificity in decision making for breast masses biopsy recommendation at conventional US. In their study, 367 biopsy-proven breast masses were evaluated by 5 readers. When both elastography and Doppler scores were negative, leading to strict downgrading, the specificity increased for all readers without significant change in sensitivity. In another quantitative shear wave elastography study, it was reported that 46% of biopsies could have been avoided by modification of the BI-RADS category 4 to category 3 on the basis of the elasticity score.[20]

As for the management of thyroid nodules, the decision making for FNA biopsy is the first step since FNA is regarded as the most accurate and cost-effective method for evaluating thyroid nodules. All the nodules >5 mm with high-risk history and nodules with suspicious US features should be recommended to FNA.[21] Some recent studies showed that the majority of FNA results was benign (45.7–64.6%)[22,23] and the malignant rate after surgery was only 28%.[24] In other words, more than half of the patients with thyroid nodules received unnecessary FNA or surgery. Therefore, it is an important issue to stratify the thyroid nodules and to reduce unnecessary FNA or surgery.

In this study, whether elastography can help readers to make correct decision for FNA recommendation was assessed. As the results presented in Tables 3 and 4, after combined VTQ, unnecessary FNA can be avoided in 82.6% of the indeterminate lesions (based on US only) that were pathologically benign. On the other hand, EI made correct prediction in 62.6% of malignant nodules that were regarded as probably benign (score 2) on US. These results indicated that combined VTQ is helpful to avoid unnecessary FNA and combined EI can be used to detect the false negative lesions. Therefore, although shear wave elastography such as VTQ is a technical improvement in comparison with conventional strain elastography such as EI, both of them can find their place in clinical practice. Similar results were found in another study about breast lesions. If Breast Imaging Reporting and Data System (BI-RADS) category 4a lesions with no stiffness at US elastography had not undergone biopsy, 44% (22 of 50) of benign biopsy results could have been avoided.[25] With regard to thyroid nodules, however, Moon et al[18] had different findings in their study based on strain elastography. They concluded that conventional strain elastography was not a useful tool in recommending FNA biopsy. The possible reasons for the different findings were as follows: The elastography technique carried out in Moon's study was strain elastography only, while ARFI elastography was applied in the present study in addition to conventional strain elastography. Moon's study included 126 nodules <5 mm. However, in nodules ≤5 mm, both conventional US and elastography have limitations in predicting thyroid malignancy. As for the present study, the nodules <5 mm were excluded. The conclusion that elastography was not a useful tool in recommending FNA came from the evaluation of the diagnostic performance of elastography, as well as the combination of elastography and gray-scale US. Direct evaluation of the decision making for biopsy recommendation was not performed in their study.

The present study had some limitations: Firstly, the transducer frequency might be low for optimal nodule evaluation with conventional US. However, we have only this transducer for thyroid examination in our institution and the ARFI function is only available in this transducer. Future studies with higher frequency might be necessary. Secondly, the population does not reflect a normal distribution of disease that the malignancy rate was as high as 34%. In order to ensure the credibility of the results, the nodules without histological diagnoses were excluded. Thirdly, cystic nodules were excluded from this study because VTQ measurement is not applicable in cystic nodules. Therefore, the results of the present study might be more suitable for the nodules that have suspicious features on US and should be carefully interpreted. In addition, the interobserver agreement of VTI and EI images classification was only fair to good. This is the common limitation of qualitative strain elastography, which indicates that more training for image scoring is needed. Finally, ARFI has some technical limitations. The elasticity features might be different in different parts of the nodule; VTQ can only reveal the local stiffness of the nodule, instead of the entire lesion. For a thyroid nodule close to the carotid artery and trachea, the pulsation of carotid artery and breathing might affect the VTQ measurement.

In conclusion, ARFI increases the specificity and PPV in diagnosing thyroid nodules, nevertheless, a strict training might be necessary as the increase is only seen in senior reader. US combined VTQ might be helpful in reducing unnecessary FNA for indeterminate nodules whereas US combined EI might be useful to detect the false negative nodules that are benign-looking on conventional US. Therefore, both shear wave elastography such as VTQ and strain elastography such as EI are needed in decision making for the thyroid nodules.

REFERENCES

1. Hundahl SA, Fleming ID, Fremgen AM, et al. A National Cancer Data Base report on 53,856 cases of thyroid carcinoma treated in the US, 1985–1995. *Cancer.* 1998;83:2638–2648.
2. Leenhardt L, Grosclaude P, Cherie-Challine L. Increased incidence of thyroid carcinoma in France: a true epidemic or thyroid nodule management effects? Report from the French Thyroid Cancer Committee. *Thyroid.* 2004;14:1056–1060.
3. Sprague BL, Warren AS, Trentham-Dietz A. Thyroid cancer incidence and socioeconomic indicators of health care access. *Cancer Causes Control.* 2008;19:585–593.
4. Frates MC, Benson CB, Charboneau JW, et al. Management of thyroid nodules detected at US: Society of Radiologists in ultrasound consensus conference statement. *Radiology.* 2005;237:794–800.
5. Shweel M, Mansour E. Diagnostic performance of combined elastosonography scoring and high-resolution ultrasonography for the differentiation of benign and malignant thyroid nodules. *Eur J Radiol.* 2013;82:995–1001.
6. Russ G, Royer B, Bigorgne C, et al. Prospective evaluation of thyroid imaging reporting and data system on 4550 nodules with and without elastography. *Eur J Endocrinol.* 2013;168: 649–655.
7. Trimboli P, Guglielmi R, Monti S, et al. Ultrasound sensitivity for thyroid malignancy is increased by real-time elastography: a prospective multicenter study. *J Clin Endocrinol Metab.* 2012;97:4524–4530.
8. Ragazzoni F, Deandrea M, Mormile A, et al. High diagnostic accuracy and interobserver reliability of real-time elastography in the evaluation of thyroid nodules. *Ultrasound Med Biol.* 2012;38:1154–1162.
9. Cappelli C, Pirola I, Gandossi E, et al. Real-time elastography: a useful tool for predicting malignancy in thyroid nodules with nondiagnostic cytologic findings. *J Ultrasound Med.* 2012;31:1777–1782.

 Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

10. Unluturk U, Erdogan MF, Demir O, et al. Ultrasound elastography is not superior to grayscale ultrasound in predicting malignancy in thyroid nodules. *Thyroid.* 2012;22:1031–1038.
11. Park SH, Kim SJ, Kim EK, et al. Interobserver agreement in assessing the sonographic and elastographic features of malignant thyroid nodules. *Am J Roentgenol.* 2009;193:W416–W423.
12. Zhang YF, Xu HX, He Y, et al. Virtual touch tissue quantification of acoustic radiation force impulse: a new ultrasound elastic imaging in the diagnosis of thyroid nodules. *PLoS ONE.* 2012doi:10.1371/journal.pone.0049094.
13. Rago T, Vitti P, Chiovato L, et al. Role of conventional ultrasonography and color flow-Doppler sonography in predicting malignancy in ''cold'' thyroid nodules. *Eur J Endocrinol.* 1998;138:41–46.
14. Itoh A, Ueno E, Tohno E, et al. Breast disease: clinical application of US elastography for diagnosis. *Radiology.* 2006;239:341–350.
15. Zhang YF, He Y, Xu HX, et al. Virtual touch tissue imaging of acoustic radiation force impulse: a new technique in the differential diagnosis between benign and malignant thyroid nodules. *J Ultrasound Med.* 2014;33:585–595.
16. Moon WJ, Jung SL, Lee JH, et al. Benign and malignant thyroid nodules: US differentiation-multicenter retrospective study. *Radiology.* 2008;247:762–770.
17. Rago T, Santini F, Scutari M, et al. Elastography: new developments in ultrasound for predicting malignancy in thyroid nodules. *J Clin Endocrinol Metab.* 2007;92:2917–2922.
18. Moon HJ, Sung JM, Kim EK, et al. Diagnostic performance of gray-scale US and elastography in solid thyroid nodules. *Radiology.* 2012;262:1002–1013.
19. Cho N, Jang M, Lyou CY, et al. Distinguishing benign from malignant masses at breast US: combined US elastography and color Doppler US—influence on radiologist accuracy. *Radiology.* 2012;262:80–90.
20. Athanasiou A, Tardivon A, Tanter M, et al. Breast lesions: quantitative elastography with supersonic shear imaging—preliminary results. *Radiology.* 2010;256:297–303.
21. Cooper DS, Doherty GM, Haugen BR, et al. American Thyroid Association Guidelines Taskforce 2006 Management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid.* 2006;16:109–142.
22. Yang J, Schnadig V, Logrono R, et al. Fine-needle aspiration of thyroid nodules: a study of 4703 patients with histologic and clinical correlations. *Cancer.* 2007;111:306–315.
23. Jo VY, Stelow EB, Dustin SM, et al. Malignancy risk for fine-needle aspiration of thyroid lesions according to the Bethesda system for reporting thyroid cytopathology. *Am J Clin Pathol.* 2010;134:450–456.
24. Williams BA, Bullock MJ, Trites JR, et al. Rates of thyroid malignancy by FNA diagnostic category. *J Otolaryngol Head Neck Surg.* 2013Online publication. doi:10.1186/1916-0216-42-61.
25. Cho N, Moon WK, Park JS, et al. Nonpalpable breast masses: evaluation by US elastography. *Korean J Radiol.* 2008;9:111–118.

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.