Diagnostic Performance of Gray-Scale US and Elastography in Solid Thyroid Nodules[1]

Hee Jung Moon, MD, PhD
Ji Min Sung, PhD
Eun-Kyung Kim, MD, PhD
Jung Hyun Yoon, MD
Ji Hyun Youk, MD, PhD
Jin Young Kwak, MD, PhD

Purpose: To evaluate the diagnostic performance of gray-scale ultrasonography (US) and elastography in differentiating benign and malignant thyroid nodules.

Materials and Methods: This was an institutional review board–approved retrospective study with waiver of informed consent. A total of 703 solid thyroid nodules in 676 patients (mean age, 49.7 years; range, 18–79 years) were included; there were 556 women (mean age, 49.5 years; range, 20–74 years) and 120 men (mean age, 50.7 years; range, 18–79 years). Nodules with marked hypoechogenicity, poorly defined margins, microcalcifications, and a taller-than-wide shape were classified as suspicious at grayscale US. Findings at elastography were classified according to the Rago criteria and the Asteria criteria. The diagnostic performances of gray-scale US and elastography were compared. For comparison between the diagnostic performances of gray-scale US and the combination of gray-scale US and elastography, three sets of criteria were assigned: criteria set 1, nodules with any suspicious grayscale US feature were assessed as suspicious; criteria set 2, Rago criteria were added as suspicious features to criteria set 1; and criteria set 3, Asteria criteria were added as suspicious features to criteria set 1. The diagnostic performances of gray-scale US, elastography with Rago criteria, and elastography with Asteria criteria, and odds ratios (ORs) with 95% confidence intervals for predicting thyroid malignancy were compared using generalized estimating equation analysis.

Results: Of 703 nodules, 217 were malignant and 486 were benign. Sensitivity, negative predictive value (NPV), and OR of gray-scale US for the 703 nodules were 91.7%, 94.7%, and 22.1, respectively, and these values were higher than the 15.7% and 65.4% sensitivity, 71.7% and 79.1% NPV, and 3.7 and 2.6 ORs found for elastography with Rago and Asteria criteria, respectively. Specificity, positive predictive value, and accuracy for criteria set 1 were significantly higher than those for criteria sets 2 and 3 for most of the nodule subgroups that were considered.

Conclusion: Elastography alone, as well as 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 features; elastography was not a useful tool in recommending fine-needle aspiration biopsy.

[1] From the Department of Radiology, Research Institute of Radiological Science (H.J.M., E.K.K., J. H. Yoon, J. H. Youk, J.Y.K.), and Department of Research Affairs (J.M.S.), Yonsei University College of Medicine, 250 Seongsanno, Seodaemun-gu, 120-752 Seoul, Korea; and Department of Radiology, Bundang CHA Medical Center, CHA University, College of Medicine, Seongnam, Korea (J. H. Yoon). Received April 24, 2011; revision requested June 15; revision received September 7; accepted September 27; final version accepted October 27. **Address correspondence to** J.Y.K. (e-mail: *docjin@yuhs.ac*).

A firm and hard thyroid nodule on palpation is associated with an increased risk of malignancy (1). Palpation is subjective and highly dependent on the examiner and on the size and location of the nodules (2–4). Elastography has been introduced to evaluate the tissue hardness objectively and to augment the diagnostic accuracy of gray-scale ultrasonography (US) (5,6). The stiffness of thyroid nodules is dependent on the composition and cellularity of the nodule (7). The basic concept of US elastography is that compression applied to the thyroid tissue produces the strain (tissue displacement in longitudinal direction) within the tissue, and the amount of strain is less in harder tissues than in softer ones (8). Elastography is useful in differentiating malignant from benign thyroid nodules because malignant nodules are harder than the surrounding adjacent parenchyma (2,3,7,9–16). Although many studies have been performed to evaluate the diagnostic performance of elastography (2,3,9–11,13–17), most were performed in small series (3,9–11,13–17). Moreover, although combinations of suspicious gray-scale US features, such as marked hypoechogenicity, microlobulated or irregular margins, microcalcifications, and a shape that is taller than wide, show higher diagnostic performances than each individual gray-scale US feature (18–24), elastography has been evaluated without comparison with gray-scale US (15), with each gray-scale US feature, or with combinations of a few suspicious gray-scale US features (2,3,9–11,13,14,16,17). To evaluate the diagnostic performance of elastography either as an adjunctive diagnostic tool to gray-scale US or as a separate diagnostic tool, comparison between elastography and gray-scale US features that are widely used in clinical practice is necessary.

Therefore, our purpose was to evaluate the diagnostic performances of gray-scale US and elastography in differentiating benign and malignant thyroid nodules.

Advances in Knowledge

- The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy of gray-scale US alone were 91.7%, 66.7%, 55.1%, 94.7%, and 74.4%, respectively; these values were superior to the 15.7% and 65.4% sensitivity, 95.3% and 58.2% specificity, 59.6% and 41.2% PPV, 71.7% and 79.1% NPV, and 70.7% and 60.5% accuracy of elastography with Rago and Asteria criteria, respectively.
- The sensitivity, specificity, PPV, NPV, and accuracy for the combination of gray-scale US and elastography with Rago (criteria set 2) and Asteria (criteria set 3) criteria were 92.2% and 94.5%, 65% and 47.5%, 54.1% and 44.6%, 94.9% and 95.1%, and 73.4% and 62%, respectively; these values were not superior to the 91.7%, 66.7%, 55.1%, 94.7%, and 74.4%, respectively, of grayscale US alone (criteria set 1).

Materials and Methods

The institutional review board approved this retrospective study and did not require patient approval or informed consent for the review of patient images and records.

Patients

From June to November 2009, 864 thyroid nodules were imaged at gray-scale US, elastography, and US-guided fine-needle aspiration (FNA). Of these, 101 nodules containing cystic components were excluded. Of 763 completely solid nodules, 60 were excluded for the following reasons: Elastography for 17 nodules had not been performed successfully because the nodules were bulging masses causing skin elevation (n = 7) or the pressure indicator at elastography displayed on the screen was 4 or more (n = 10). Forty-three nodules suspicious for papillary thyroid carcinoma (n = 3) or with indeterminate (n = 4) or inadequate results (n = 36) at cytologic evaluation were excluded because they had not undergone surgery or repeat US-guided FNA (Fig 1). Thyroid nodules that met the following criteria were included: *(a)* benign or malignant results at cytologic evaluation, *(b)* thyroid surgery was performed after obtaining cytologic results suspicious for papillary thyroid carcinoma or indeterminate results, such as follicular or Hürthle cell neoplasm, or *(c)* benign or malignant results at follow-up US-guided FNA or thyroid surgery after cytologic results of inadequate specimen (Fig 1). Finally, 703 solid thyroid nodules in 676 patients (mean age, 49.7 years; range, 18–79 years) were included in this study; patients included 556 women (mean age, 49.5 years; range, 20–74 years) and 120 men (mean age, 50.7 years; range, 18–79 years). Of the 703

Implication for Patient Care

- Because elastography alone, as well as the combination of elastography and gray-scale US, showed inferior performance in the differentiation of malignant and benign thyroid nodules in comparison with gray-scale US features, elastography was not a useful tool in recommending fine-needle aspiration biopsy.

Published online
10.1148/radiol.11110839 **Content code:** US

Radiology 2012; 262:1002–1013

Abbreviations:
FNA = fine-needle aspiration
NPV = negative predictive value
PPV = positive predictive value
OR = odds ratio

Author contributions:
Guarantors of integrity of entire study, H.J.M., J.M.S., J.Y.K.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; literature research, H.J.M., E.K.K.; clinical studies, H.J.M., E.K.K., J.Y.K.; statistical analysis, H.J.M., J.M.S.; and manuscript editing, H.J.M., E.K.K., J. H. Yoon, J. H. Youk

Potential conflicts of interest are listed at the end of this article.

Figure 1

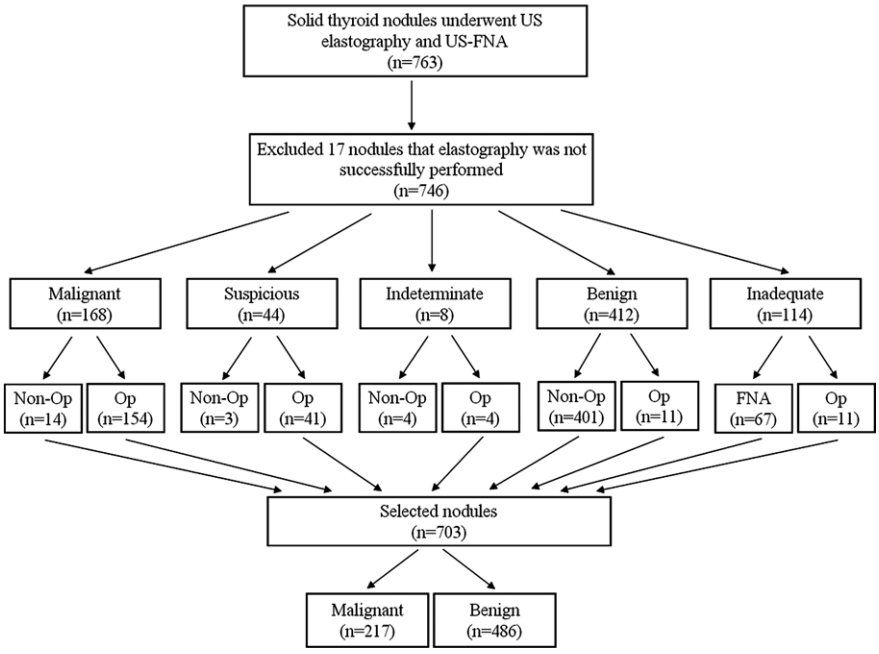


Figure 1: Inclusion criteria for the study. *FNA* = thyroid nodules with repeat aspirations, *Non-Op* = thyroid nodules in which no surgery was performed, *Op* = thyroid nodules in which surgery was performed.

nodules, 308 were larger than 10 mm and 395 nodules were 10 mm or smaller. There were 577 nodules larger than 5 mm and 126 nodules 5 mm or smaller (Table 1). There were 114 nodules in 112 patients that had macrocalcifications (n = 96) or eggshell calcifications (n = 18). Twenty-seven patients underwent US-guided FNA for two thyroid nodules and 649 patients underwent US-guided FNA for one thyroid nodule. In 221 nodules in 202 patients, surgery was performed after FNA.

Real-time Gray-Scale US

All gray-scale US images were obtained by using a 6–14-MHz linear array transducer (EUB-7500; Hitachi Medical, Tokyo, Japan). Real-time gray-scale US was performed by one of eight radiologists with 1 year to 15 years of experience in thyroid imaging.

Gray-scale US features of thyroid nodules that underwent US-guided FNA were prospectively recorded in the radiology reports according to the internal component, echogenicity, margin, calcifications, shape, and final assessment at the time of FNA by the radiologists who performed the US examination and FNA at our institution. The internal component of thyroid nodules which undergo US-guided FNA are routinely classified and recorded as completely solid, cystic portion greater than 50%, or cystic portion less than or equal to 50% at our institution. Based on our database, only completely solid thyroid nodules were included. Echogenicity was classified as hyper-, iso-, or hypoechogenicity (when a nodule showed hyper-, iso-, or hypoechogenicity compared with the normal thyroid gland) or marked hypoechogenicity (when a nodule showed relatively hypoechogenicity compared with the surrounding strap muscle). Margin was classified as well defined or poorly defined (microlobulated or irregular margin). Calcifications were classified as microcalcifications (less than or equal to 1 mm in diameter; tiny, punctate, hyperechoic foci, either with or without acoustic shadows), macrocalcification, eggshell calcification, or no calcification. Shape was classified as wider than tall or taller than wide. Suspicious malignant gray-scale US features included marked hypoechogenicity, poorly defined margin, microcalcifications, and taller than wide shape (20,21). When thyroid nodules showed one or more of these suspicious malignant gray-scale US features, they were assessed as "suspicious" (20,21). When thyroid nodules showed no suspicious features, they were assessed as "probably benign" (20,21).

Real-time Elastography

After gray-scale examination, elastography was routinely performed by the same radiologists who performed gray-scale US. Elastography was performed in thyroid nodules detected at gray-scale US and targeted for US-guided FNA by using the same US machine and probe. Freehand technique was used in obtaining elastography images. All elastography images were obtained in longitudinal planes. Prior to performing elastography, each of the eight radiologists had 2 months of experience with the machine and weekly thyroid imaging conferences regarding elastography images. During the training, each eight radiologist performed more than 205 US-guided FNA procedures for thyroid nodules, and elastography for thyroid nodules was performed for more than 100 nodules.

During elastography, the probe was positioned perpendicular to the skin while applying pressure. Images were obtained by applying light repetitive compression at the skin above the targeted thyroid nodule. A square region of interest positioning the target nodule at the center of the box was set for elastography acquisition. The superior margin was set to include subcutaneous fat, and the inferior margin was set to include longus colli muscle. Radiologists obtained optimal images showing both color homogeneity within the region of interest and pressure indicator displayed on the screen ranging between 2 and 3. Images were displayed in a split-screen mode, with the gray-scale images on the right and the translucent color-scale elastography images superimposed on the corresponding gray-scale US image in the left. Each pixel of the elasticity image was shown as one of

Table 1

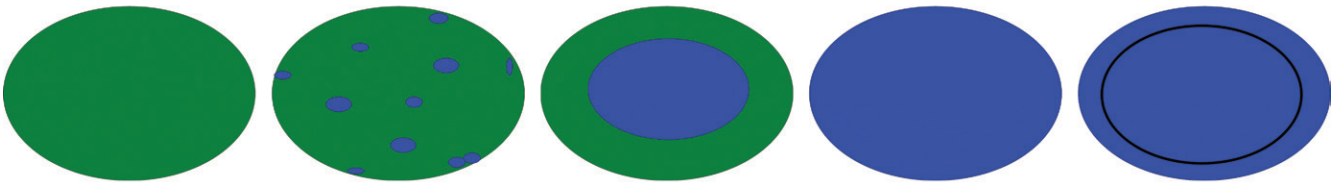
Nodule Size and Numbers of Nodules and Patients

Parameter	Nodules	Patients	Final Diagnosis		*P* Value*
			Benign	Malignant	
All Nodules	703	676	486	217	
Mean nodule size (mm)	11.6		12.6 (3–51)	9.3 (2–38)	<.001
Mean patient age (y)		48.9 (18–79)	50.9 (20–79)	47.3 (18–73)	<.001
Women	579	556	399	180	.809
Men	124	120	87	37	
Nodules >10 mm	308	296	239 (77.6%)	69 (22.4%)	
Mean nodule size (mm)	17.9		18.6 (11–51)	15.4 (11–38)	.006
Mean patient age (y)		49.5 (18–75)	50.9 (20–75)	44.6 (18–70)	<.001
Women	243	234	189	15	.917
Men	65	62	50	54	
Nodules ≤10 mm	395	380	247 (62.5%)	148 (37.5%)	
Mean nodule size (mm)	6.7 (2–10)		6.9 (3–10)	6.5 (2–10)	.055
Mean patient age (y)		50.0 (21–79)	50.8 (21–79)	48.6 (22–73)	.055
Women	336		210	126	.965
Men	59		37	22	
Nodules >5 mm	577	554	410 (71.1%)	167 (28.9%)	
Mean nodule size (mm)	13.2 (6–51)		14.2 (6–51)	10.8 (6–38)	<.001
Mean patient age (y)		50.4 (18–79)	51.4 (20–79)	47.9 (18–73)	.003
Women	468		333	135	.939
Men	109		77	32	
Nodules ≤5 mm	126	122	76 (60.3%)	50 (39.7%)	
Mean nodule size (mm)	4.3 (2–5)		4.4 (3–5)	4.3 (2–5)	.303
Mean patient age (y)		47.0 (21–72)	48.3 (21–72)	45.2 (22–67)	.103
Women	111		66	45	.663
Men	15		10	5	

Note.—Unless otherwise indicated, data are numbers of nodules or patients.

* *P* values were calculated by using generalized estimating equation analysis.

Figure 2

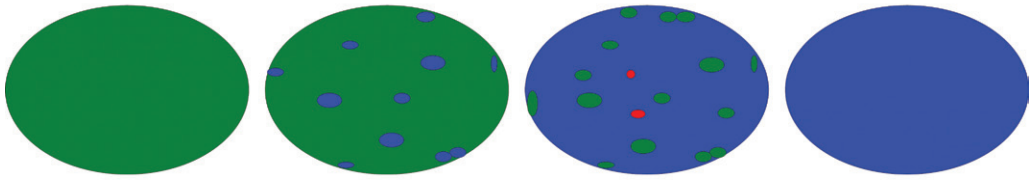


a. **b.** **c.** **d.** **e.**

Figure 2: Elastography scores according to Rago criteria. **(a)** A score of 1 indicated even elasticity in the whole nodule. **(b)** A score of 2 indicated elasticity in a large part of the nodule. **(c)** A score of 3 indicated elasticity only at the peripheral part of the nodule. **(d)** A score of 4 indicated no elasticity in the nodule. **(e)** A score of 5 indicated no elasticity in the nodule or in the area showing posterior shadowing.

the 256 specific colors, representing the extent of strain. The scale ranged from red, showing areas of greatest strain (ie, softest component), to blue, showing no strain (ie, hardest component). Elastography images were classified according to the scores by Rago et al (13) and Asteria et al (9). Elasticity according to Rago et al (hereafter, Rago criteria) originated from the elastography scale by Ueno et al (25) and was applied to thyroid nodules and elastography scores were classified on a scale of 1 to 5 (Fig 2). Nodules with Rago scores of 4 and 5 were classified as suspicious for malignancy (13). Elasticity according to Asteria et al (hereafter, Asteria criteria) originated from the elastography by Ito et al (8) and was applied to thyroid nodules and elastography scores were classified on a scale of 1 to 4 (Fig 3). Nodules with Asteria scores of 3 and 4 were classified as suspicious

Figure 3



a. b. c. d.

Figure 3: Elastography scores according to Asteria criteria. **(a)** A score of 1 indicated elasticity in the whole examined area. **(b)** A score of 2 indicated elasticity in a large part of the examined area. **(c)** A score of 3 indicated stiffness in a large part of the examined area. **(d)** A score of 4 indicated a nodule without elasticity.

Table 2

Grayscale US and Elastography Features according to Malignancy and Benignity

Feature	No. of Malignant Nodules (n = 217)	No. of Benign Nodules (n = 486)	P Value*
Echogenicity			<.001†
Hyperechogenicity (n = 3)	0 (0)	3 (100)	.999
Isoechogenicity (n = 200)	7 (3.5)	193 (96.5)	<.001
Hypoechogenicity (n = 433)	169 (39.0)	264 (61.0)	<.001
Marked hypoechogenicity (n = 67)	41 (61.2)	26 (30.8)	<.001
Margin			<.001
Well defined (n = 376)	29 (13.4)	347 (71.4)	
Poorly defined (n = 327)	188 (86.6)	139 (28.6)	
Calcification			<.001†
Microcalcification (n = 106)	79 (74.5)	27 (25.5)	<.001
Macrocalcification (n = 96)	26 (27.1)	70 (72.9)	.389
Eggshell calcification (n = 18)	2 (11.1)	16 (88.9)	.020
No calcification (n = 483)	110 (22.8)	373 (77.2)	<.001
Shape			<.001
Wider than tall (n = 456)	69 (15.1)	387 (84.9)	
Taller than wide (n = 247)	148 (59.9)	99 (40.1)	
Final assessment			<.001
Probably benign (n = 342)	18 (5.3)	324 (94.7)	
Suspicious (n = 361)	199 (55.1)	162 (44.9)	
Rago criteria			<.001
Score 1, 2, or 3 (n = 646)	183 (28.3)	463 (71.7)	
Score 4 or 5 (n = 57)	34 (59.6)	23 (40.4)	
Asteria criteria			<.001
Score 1 or 2 (n = 358)	75 (20.9)	283 (79.1)	
Score 3 or 4 (n = 345)	142 (41.2)	203 (581.8)	

Note.—Unless otherwise indicated, data are numbers of nodules, and numbers in parentheses are percentages.

* P values were calculated by using generalized estimating equation analysis.

† Fisher exact test.

for malignancy (9). The elastography scores were prospectively recorded along with the gray-scale US features in the radiology reports.

US-guided FNA Biopsy

US-guided FNA biopsy was performed by the same radiologists who performed gray-scale US and elastography. FNA was usually performed in either thyroid nodules with suspicious gray-scale US assessment or the largest thyroid nodule with probably benign assessment when no other lesion showed suspicious features at US. US-guided FNA was performed by using a 23-gauge needle and a 2-mL disposable plastic syringe with a freehand technique.

Data Analysis

By using the indicator of malignancy as the dependent variable, the association between malignancy and patient age and sex for 703 nodules in 676 patients

Table 3

Diagnostic Performance of Gray-Scale US and Elastography in 703 Nodules in 676 Patients

Nodule Size and Criteria*	Sensitivity (%)	Specificity (%)	PPV (%)	Recalculated PPV (%)	NPV (%)	Recalculated NPV (%)	Accuracy (%)	OR†	QIC
Overall (n = 676)									
Gray-scale US	91.7	66.7	55.1	24.4	94.7	98.6	74.4	22.1 (13.2, 37.1)	641.7
Rago criteria	15.7 (<.001)	95.3 (<.001)	59.6 (.472)	28.0	71.7 (<.001)	90.6	70.7 (.147)	3.7 (2.1, 6.5)	850.9
Asteria criteria	65.4 (<.001)	58.2 (<.001)	41.2 (<.001)	15.5	79.1 (<.001)	93.5	60.5 (<.001)	2.6 (1.9, 3.7)	838.9
>10 mm (n = 296)									
Gray-scale US	88.4	84.5	62.2	40.1	96.2	98.4	85.4	41.6 (18.4, 94.1)	201.9
Rago criteria	23.2 (<.001)	97.5 (<.001)	72.7 (.319)	52.0	81.5 (<.001)	91.5	80.8 (.136)	11.7 (4.2, 31.4)	304.0
Asteria criteria	76.8 (.045)	64.9 (<.001)	38.7 (<.001)	20.4	90.6 (.007)	96.0	67.5 (<.001)	6.1 (3.3, 11.4)	293.1
≤10 mm (n = 380)									
Gray-scale US	93.2	49.4	52.5	17.8	92.4	98.4	65.8	13.5 (6.8, 26.8)	438.8
Rago criteria	12.2 (<.001)	93.1 (<.001)	52.4 (.896)	17.2	63.9 (<.001)	90.3	62.8 (.433)	1.9 (0.9, 3.8)	523.4
Asteria criteria	60.1 (<.001)	51.8 (.502)	42.8 (<.001)	12.8	68.4 (<.001)	91.7	54.9 (<.001)	1.6 (1.1, 2.5)	521.2
>5 mm (n = 554)									
Gray-scale US	90.4	74.1	58.8	29.1	95.0	98.5	78.9	27.1 (15.4, 47.4)	479.4
Rago criteria	17.4 (<.001)	96.3 (<.001)	65.9 (.323)	35.8	74.1 (<.001)	90.9	73.5 (.041)	5.5 (2.9, 10.6)	670.1
Asteria criteria	71.3 (<.001)	60.0 (<.001)	42.0 (<.001)	17.3	83.7 (<.001)	94.7	63.3 (<.001)	3.7 (2.5, 5.5)	650.8
≤5 mm (n = 122)									
Gray-scale US	96.0	26.3	46.2	13.3	90.9	98.2	54.0	8.6 (1.9, 38.6)	161.0
Rago criteria	10.0 (<.001)	89.5 (<.001)	38.5 (.555)	10.0	60.2 (.008)	89.4	57.9 (.600)	0.9 (0.3, 3.1)	173.3
Asteria criteria	46.0 (<.001)	48.7 (<.001)	37.1 (.043)	9.5	57.8 (.006)	88.5	47.6 (.247)	0.8 (0.4, 1.7)	172.9

Note.—Unless otherwise indicated, data in parentheses are *P* values from comparison with gray-scale US. Recalculated PPV = PPV recalculated as 10.5% of prevalence based on Bayes theorem, recalculated NPV = NPV recalculated as 10.5% of prevalence based on Bayes theorem, QIC = quasi–likelihood information criterion.

* Numbers in parentheses are numbers of patients.

† Numbers in parentheses are 95% confidence intervals.

were compared by using a generalized estimating equation analysis based on a binary logistic regression model. The correlation structure was modeled by assuming observations to be symmetrically correlated when derived from lesions within the same patient and otherwise independent. The final diagnosis was based on the cytologic and histopathologic results. Gray-scale US features and scores with Rago criteria and Asteria criteria showing association to thyroid malignancy were evaluated by using generalized estimating equation analysis. Diagnostic performances such as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy of gray-scale US, Rago criteria, and Asteria criteria were compared by using generalized estimating equation analysis. Odds ratios (ORs) with 95% confidence intervals of gray-scale US, Rago criteria, and Asteria criteria for predicting thyroid malignancy were also calculated and compared by using generalized estimating equation analysis. To investigate the suitable model in predicting thyroid malignancy, quasi–likelihood information criterion was calculated (26). The smaller quasi–likelihood information criterion means the suitable model for predicting thyroid malignancy. All diagnostic performances of gray-scale US, Rago criteria, and Asteria criteria were compared in all nodules, in nodules larger than 10 mm and smaller than or equal to 10 mm, and in nodules larger than 5 mm and smaller than or equal to 5 mm, respectively. Because macrocalcifications within the nodule may have effect on the stiffness and cause false-positive results at elastography (2,9–11,13,16), we also evaluated the diagnostic performances of 589 thyroid nodules in 564 patients after excluding 114 thyroid nodules with macro- or eggshell calcifications in 112 patients.

To validate whether adding Rago criteria (13) or Asteria criteria (9) elastography to gray-scale US improved the diagnostic performances for predicting thyroid malignancy, three sets of criteria were assigned as follows: criteria set 1, thyroid nodules with one or more suspicious gray-scale US features

Figure 4

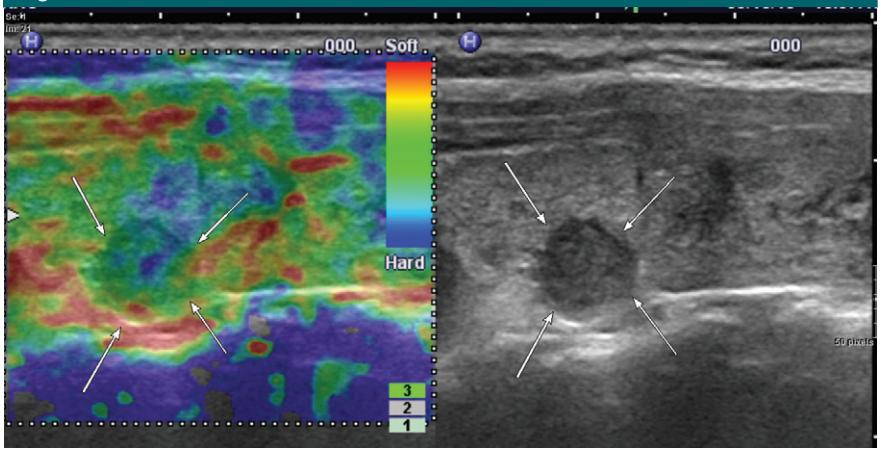


Figure 4: Images in a 46-year-old woman who underwent routine checkup. A 10-mm right thyroid nodule (arrows) with marked hypoechogenicity, poorly defined margins, microcalcifications, and a taller-than-wide shape was found at gray-scale US and assessed as suspicious. A score of 2, with both Rago and Asteria criteria, was assigned at elastography. This thyroid nodule was diagnosed as papillary thyroid carcinoma at cytologic evaluation and as a diffuse sclerosing variant of papillary thyroid carcinoma after surgery.

Figure 5

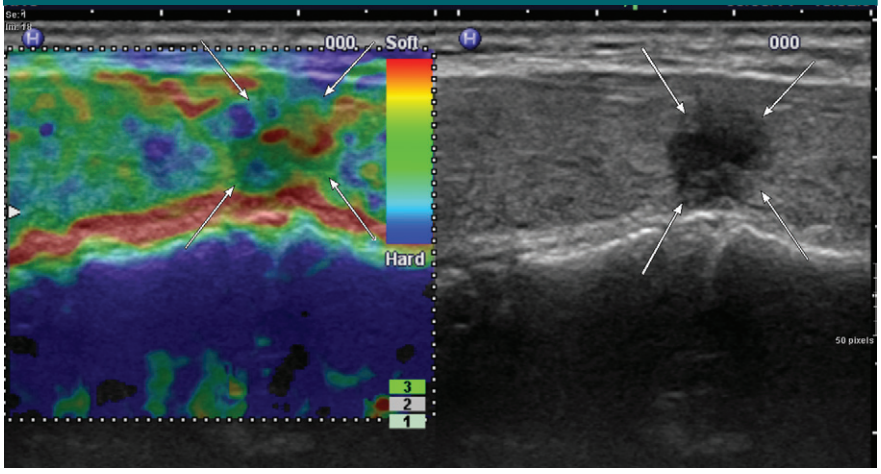


Figure 5: Images in a 49-year-old woman who underwent routine checkup. An 8-mm left thyroid nodule (arrows) with hypoechogenicity, poorly margins, and taller-than-wide shape was found at gray-scale US and assessed as suspicious. A score of 2, with both Rago and Asteria criteria, was assigned at elastography. This thyroid nodule was diagnosed as papillary thyroid carcinoma at cytologic evaluation and surgery.

that were assessed as suspicious; criteria set 2, Rago criteria scores of 4 or 5 were added as one of the suspicious features to criteria set 1; and criteria set 3, Asteria criteria scores of 3 or 4 were added as suspicious features to criteria set 1. We compared the sensitivity, specificity, PPV, NPV, accuracy, ORs with 95% confidence intervals, and quasi–likelihood information criterion (26) of three criteria in 589 nodules with macro- and eggshell calcifications in 564 patients, as well as all 703 nodules in all 676 patients.

The PPV and NPV were recalculated based on the Bayes theorem as follows: PPV = sensitivity × prevalence/(specificity × prevalence) + ([1−specificity] × [1−prevalence]), and NPV = specificity (1−prevalence)/(specificity × [1−prevalence]) + ([1−sensitivity] × prevalence).

Prevalence used in this equation was the reported prevalence of malignancy in thyroid nodules selected for US-guided FNA in the Korean population, 10.5% (27), which was consistent with the prevalence of other studies, 9.2%–13.0% (28).

Statistical significance was determined at a P value less than .05. Statistical analyses were performed with statistical software (SAS system for Windows, version 9.1.3; SAS Institute, Cary, NC).

Results

Demographic and Pathologic Characteristics

The mean age of patients with malignant nodules was younger than that of patients with benign nodules ($P <$ .001). Sex of patients was not associated with malignancy (P = .809) (Table 1).

Each Gray-Scale US and Elastography Feature Associated with Malignancy

Gray-scale US features of marked hypoechogenicity, poorly defined margin, microcalcifications, a shape that is taller than wide, and suspicious assessment were more significantly seen in malignant nodules than benign nodules (all $P <$ .001) (Table 2). Scores of 4 and 5 with Rago criteria and scores of 3 and 4 with Asteria criteria were also more significantly seen in malignant nodules than in benign nodules (all $P <$ .001).

Diagnostic Performance of Gray-Scale US, Rago Criteria, and Asteria Criteria

In the 703 nodules in 676 patients, sensitivities, NPVs, and recalculated NPVs of gray-scale US were significantly higher than those of elastography with Rago and Asteria criteria, regardless of nodule size (Table 3) (Figs 4, 5). Accuracy of gray-scale US was significantly higher than that of elastography with Asteria criteria except for in nodules 5 mm or smaller. Accuracy of gray-scale US features was mostly higher than that of elastography with Rago criteria, but it was not statistically significant. However, the ORs of gray-scale US were

Table 4

Diagnostic Performance of Gray-Scale US and Elastography in 589 Nodules in 564 Patients after Excluding 114 Thyroid Nodules with Macro- or Eggshell Calcifications

Nodule Size and Criteria*	Sensitivity (%)	Specificity (%)	PPV (%)	Recalculated PPV (%)	NPV (%)	Recalculated NPV (%)	Accuracy (%)	OR†	QIC
Overall (n = 564)									
Gray-scale US	91.5	74.0	62.5	29.2	94.9	98.7	79.6	30.8 (17.6, 53.8)	496.9
Rago criteria	12.2 (<.001)	97.5 (<.001)	69.7 (.379)	36.3	70.1 (<.001)	90.4	70.1 (<.001)	5.4 (2.5, 11.6)	722.4
Asteria criteria	63.0 (<.001)	64.3 (<.001)	45.4 (<.001)	17.1	78.6 (<.001)	93.7	63.8 (<.001)	3.1 (2.1, 4.4)	704.6
>10 mm (n = 249)									
Gray-scale US	88.1	89.6	71.2	49.7	96.3	98.5	89.2	63.7 (25.6, 158.1)	151.3
Rago criteria	18.6 (<.001)	98.5 (.001)	78.6 (.560)	59.4	80.5 (<.001)	91.2	80.4 (.005)	15.1 (4.1, 56.3)	261.4
Asteria criteria	76.3 (.051)	70.1 (<.001)	42.9 (<.001)	23.1	91.0 (.011)	96.2	71.5 (<.001)	7.6 (3.9, 14.8)	241.4
≤10 mm (n = 315)									
Gray-scale US	93.1	58.3	59.3	20.7	92.8	98.6	72.0	18.8 (9.0, 39.1)	344.4
Rago criteria	9.2 (<.001)	96.5 (<.001)	63.2 (.723)	23.5	61.9 (<.001)	90.1	62.0 (.016)	2.8 (1.1, 7.3)	440.9
Asteria criteria	56.9 (<.001)	58.3 (>.99)	47.1 (<.001)	13.8	67.4 (<.001)	92.0	57.8 (<.001)	1.8 (1.2, 2.9)	438.2
>5 mm (n = 460)									
Gray-scale US	89.4	81.7	67.2	36.5	94.9	98.5	84.0	37.8 (20.8, 69.0)	361.5
Rago criteria	13.4 (<.001)	97.9 (<.001)	73.1 (.518)	43.2	73.0 (<.001)	90.6	73.0 (<.001)	7.3 (3.08, 17.9)	565.4
Asteria criteria	69.0 (<.001)	66.4 (<.001)	46.2 (<.001)	19.4	83.6 (<.001)	94.8	67.2 (<.001)	4.4 (2.9, 6.7)	536.4
≤5 mm (n = 104)									
Gray-scale US	97.9	31.1	52.3	14.3	95	99.2	60.2	20.9 (2.7, 162.3)	133.8
Rago criteria	8.5 (<.001)	95.1 (<.001)	57.1 (.789)	16.9	57.4 (.009)	89.9	57.4 (.739)	1.8 (0.4, 8.5)	151.3
Asteria criteria	44.7 (<.001)	52.5 (<.001)	42.0 (.038)	9.9	55.2 (.006)	89.0	49.1 (.062)	0.9 (0.4, 1.9)	151.8

Note.—Unless otherwise indicated, data in parentheses are *P* values from comparison with gray-scale US. Recalculated PPV = PPV recalculated as 10.5% of prevalence based on Bayes theorem, recalculated NPV = NPV recalculated as 10.5% of prevalence based on Bayes theorem, QIC = quasi–likelihood information criterion.

* Numbers in parentheses are numbers of patients.

† Numbers in parentheses are 95% confidence intervals.

higher than those with Rago and Asteria criteria and the quasi–likelihood information criteria of gray-scale US were smaller regardless of nodule size.

In the 589 nodules in 564 patients selected after exclusion of 114 nodules with macro- or eggshell calcifications in 112 patients, sensitivities and NPVs of gray-scale US were significantly higher than those of elastography with Rago and Asteria criteria, regardless of nodule size, with an exception of sensitivity in Asteria criteria of nodules larger than 10 mm (Table 4). All accuracies of gray-scale US were significantly higher than those with Rago and Asteria criteria except in nodules 5 mm or smaller. All ORs of gray-scale US were higher than those with Rago and Asteria criteria and all quasi–likelihood information criteria of gray-scale US were smaller.

In the 703 nodules, specificities, NPVs, accuracies, and ORs of gray-scale US, Rago criteria, and Asteria criteria in nodules larger than10 mm or those in nodules larger than 5 mm were superior to those in nodules 10 mm or smaller or 5 mm or smaller, respectively. In the 589 nodules without macro- or eggshell calcifications, specificities, accuracies, and ORs of gray-scale US, Rago criteria, and Asteria criteria in nodules larger than 10 mm or those in nodules larger than 5 mm were superior to those in nodules smaller than or equal to 10 mm or 5 mm, respectively. Diagnostic performances other than sensitivities of gray-scale US, Rago criteria, and Asteria criteria in nodules 5 mm or smaller were inferior to those of nodules 10 mm or smaller.

Diagnostic Performance of the Three Sets of Criteria

In the 703 nodules in 676 patients, specificities, PPVs, recalculated PPVs, and accuracies of criteria set 1 were significantly higher than those of criteria sets 2 and 3 (Table 5). ORs of criteria set 1 were equal to or higher than those of criteria sets 2 and 3 and the quasi–likelihood information criteria of criteria set 1 were equal to or smaller except for in nodules smaller than 5 mm. Sensitivities and NPVs of criteria

Table 5

Diagnostic Performances of Three Criteria in 703 Nodules in 676 Patients

Nodule Size and Criteria*	Sensitivity (%)	Specificity (%)	PPV (%)	Recalculated PPV (%)	NPV (%)	Recalculated NPV (%)	Accuracy (%)	OR†	QIC
Overall (n = 676)									
Criteria set 1	91.7	66.7	55.1	24.4	94.7	98.6	74.4	22.1 (13.2, 37.1)	641.7
Criteria set 2	92.2 (.316)	65.0 (.004)	54.1 (.015)	23.6	94.9 (.578)	98.6	73.4 (.019)	21.9 (12.9, 37.1)	648.8
Criteria set 3	94.5 (.014)	47.5 (<.001)	44.6 (<.001)	17.4	95.1 (.694)	98.7	62.0 (<.001)	15.5 (8.4, 28.4)	731.8
>10 mm (n = 296)									
Criteria set 1	88.4	84.5	62.2	40.1	96.2	98.4	85.4	41.6 (18.4, 94.1)	201.9
Criteria set 2	89.9 (.314)	82.4 (.024)	59.6 (.061)	37.5	96.6 (.419)	98.6	84.1 (.101)	41.5 (17.8, 97.2)	205.3
Criteria set 3	94.2 (.044)	61.1 (<.001)	41.1 (<.001)	22.1	97.3 (.301)	98.9	68.5 (<.001)	25.5 (9.0, 72.4)	254.9
≤10 mm (n = 380)									
Criteria set 1	93.2	49.4	52.5	17.8	92.4	98.4	65.8	13.5 (6.8, 26.8)	438.8
Criteria set 2	93.2 (>.99)	48.2 (.081)	51.9 (.083)	17.4	92.2 (.083)	98.4	65.1 (.082)	12.8 (6.4, 25.5)	442.7
Criteria set 3	94.6 (.156)	34.4 (<.001)	46.4 (<.001)	14.5	91.4 (.414)	98.2	57.0 (<.001)	9.2 (4.3, 19.6)	475.6
>5 mm (n = 554)									
Criteria set 1	90.4	74.1	58.8	29.1	95.0	98.5	78.9	27.1 (15.4, 47.4)	479.4
Criteria set 2	91.0 (.316)	72.2 (.004)	57.1 (.013)	27.7	95.2 (.562)	98.6	77.6 (.019)	26.3 (14.8, 46.7)	487.5
Criteria set 3	93.4 (.024)	52.0 (<.001)	44.2 (<.001)	18.6	95.1 (.913)	98.5	63.9 (<.001)	15.3 (8.1, 29.1)	576.3
≤5 mm (n = 122)									
Criteria set1	96.0	26.3	46.2	13.3	90.9	98.2	54.0	8.6 (1.9, 38.6)	161.0
Criteria set 2	96.0 (>.99)	26.3 (>.99)	46.2 (>.99)	13.3	90.9 (>.99)	98.2	54.0 (>.99)	8.6 (1.9, 38.6)	161.0
Criteria set 3	98.0 (.323)	23.7 (.152)	45.8 (.654)	13.1	94.7 (.408)	99.0	53.2 (.563)	15.2 (2.0, 118.0)	159.4

Note.—Unless otherwise indicated, data in parentheses are *P* values from comparison with criteria set 1. Recalculated PPV = PPV recalculated as 10.5% of prevalence based on Bayes theorem, recalculated NPV = NPV recalculated as 10.5% of prevalence based on Bayes theorem, QIC = quasi-likelihood information criterion.

* Numbers in parentheses are numbers of patients. Criteria 1 = thyroid nodules with any single or more suspicious gray-scale US feature were assessed as suspicious, criteria 2 = adding pattern 4 and 5 from Rago et al (13) on elastography images as one of the suspicious features to criteria 1, criteria 3 = adding pattern 3 and 4 from Asteria et al (9) on elastography images as one of the suspicious features to criteria 1.

† Numbers in parentheses are 95% confidence intervals.

sets 1, 2, and 3 did not show significant differences.

Excluding the nodules with macro- or eggshell calcifications, in the 589 nodules in 564 patients, specificities, PPVs, recalculated PPVs, and accuracies of criteria set 1 were equal to or higher than those of criteria sets 2 and 3 (Table 6). Specificities, PPVs, NPVs, and accuracies of criteria sets 2 and 3 were not better than those of criteria set 1. ORs of criteria set 1 were equal to or higher than those of criteria sets 2 and 3 regardless of nodule size, and quasi–likelihood information criteria of criteria set 1 were smaller than or equal to those of criteria sets 2 and 3.

In the 703 nodules in 676 patients, accuracies and ORs of the three sets of criteria in nodules larger than 10 mm or larger than 5 mm in size were higher than those of nodules smaller than or equal to 10 or 5 mm, respectively. Accuracies and ORs of the three sets of criteria in nodules less than or equal to 5 mm were the smallest in value. In the 589 nodules without macro- or eggshell calcifications in 564 patients, accuracies and ORs of the three criteria sets in thyroid nodules larger than 10 mm or larger than 5 mm were higher than those in nodules smaller than or equal to 10 or 5 mm, respectively. The accuracies of the three sets of criteria in nodules less than or equal to 5 mm were the smallest.

Discussion

Elastography has been introduced to evaluate hardness objectively, to improve the diagnostic performance of gray-scale US examination in differential diagnosis of thyroid nodules, and to eventually reduce unnecessary benign biopsies in thyroid nodules (2,3,5–16). Many previous studies have proved that elastography is useful in differentiating malignant from benign nodules (2,3,7,9–16). In our study, scores of 4 and 5 with Rago criteria and scores of 3 and 4 with Asteria criteria were more significantly seen in malignant nodules than benign; our findings were consistent with those other studies (2,3,7,9–16).

Table 6

Diagnostic Performances of Three Criteria in 589 Nodules in 564 Patients after Excluding 114 Thyroid Nodules with Macro- or Eggshell Calcifications

Nodule Size and Criteria*	Sensitivity (%)	Specificity (%)	PPV (%)	Recalculated PPV (%)	NPV (%)	Recalculated NPV (%)	Accuracy (%)	OR†	QIC
Overall (n = 564)									
Criteria set 1	91.5	74.0	62.5	29.2	94.9	98.7	79.6	30.8 (17.6, 53.8)	496.9
Criteria set 2	91.5 (>.99)	73.0 (.044)	61.6 (.046)	28.5	94.8 (.465)	98.7	78.9 (.045)	29.2 (16.7, 51.1)	504.2
Criteria set 3	93.7 (.044)	54.0 (<.001)	49.0 (<.001)	19.3	94.7 (.854)	98.6	66.7 (<.001)	17.3 (9.3, 32.1)	598.3
>10 mm (n = 249)									
Criteria set 1	88.1	89.6	71.2	49.7	96.3	98.5	89.2	63.7 (25.7, 158.1)	151.3
Criteria set 2	88.1 (>.99)	88.1 (.081)	68.4 (.084)	46.4	96.2 (.083)	98.4	88.1 (.082)	54.8 (22.3, 134.3)	158.3
Criteria set 3	93.2 (.080)	66.2 (<.001)	44.7 (<.001)	24.4	97.1 (.437)	98.8	72.3 (<.001)	26.9 (9.4, 77.3)	209.3
≤10 mm (n = 315)									
Criteria set 1	93.1	58.3	59.3	20.7	92.8	98.6	72.0	18.8 (9.0, 39.1)	344.4
Criteria set 2	93.1 (>.99)	57.8 (.316)	59.02 (.317)	20.6	92.7 (.317)	98.6	71.7 (.317)	18.4 (8.8, 38.3)	346.0
Criteria set 3	93.8 (.316)	41.7 (<.001)	51.3 (<.001)	15.9	91.2 (.099)	98.3	62.3 (<.001)	10.9 (5.1, 23.5)	388.0
>5 mm (n = 460)									
Criteria set 1	89.4	81.7	67.2	36.5	94.9	98.5	84.0	37.9 (20.7, 69.0)	361.5
Criteria set 2	89.4 (>.99)	80.5 (.044)	65.8 (.046)	35.0	94.8 (.046)	98.5	83.2 (.045)	35.0 (19.2, 63.7)	369.8
Criteria set 3	92.3 (.044)	58.7 (<.001)	48.3 (<.001)	20.8	94.8 (.898)	98.5	68.7 (<.001)	16.9 (8.8, 32.5)	465.7
≤5 mm (n = 104)									
Criteria set 1	97.9	31.1	52.3	14.3	95.0	99.2	60.2	20.8 (2.7, 162.3)	133.8
Criteria set 2	97.9 (>.99)	31.1 (>.99)	52.3 (>.99)	14.3	95.0 (>.99)	99.2	60.2 (>.99)	20.8 (2.7, 162.3)	133.8
Criteria set 3	97.9 (>.99)	27.9 (.151)	51.1 (.157)	13.7	94.4 (.158)	99.1	58.3 (.154)	17.8 (2.3, 139.3)	136.4

Note.—Unless otherwise indicated, data in parentheses are *P* values from comparison with criteria set 1. Recalculated PPV = PPV recalculated as 10.5% of prevalence based on Bayes theorem, recalculated NPV = NPV recalculated as 10.5% of prevalence based on Bayes theorem, QIC = quasi-likelihood information criterion.

* Numbers in parentheses are numbers of patients. Criteria 1 = thyroid nodules with any single or more suspicious gray-scale US feature were assessed as suspicious, criteria 2 = adding pattern 4 and 5 from Rago et al (13) on elastography images as one of the suspicious features to criteria 1, criteria 3 = adding pattern 3 and 4 from Asteria et al (9) on elastography images as one of the suspicious features to criteria 1.

† Numbers in parentheses are 95% confidence intervals.

Elastography is usually performed in thyroid nodules detected at gray-scale US, and the high diagnostic performances of combinations of suspicious gray-scale US features are validated through many previous studies (18–24,29–31) as well as in our study. Accordingly, for elastography to be used widely as an adjunctive diagnostic tool supporting gray-scale US or to be used as a separate diagnostic tool, several conditions should be fulfilled. The diagnostic performance of elastography itself or a combination of elastography and gray-scale US should be superior to that of gray-scale US alone. In our study, sensitivities, NPVs, recalculated NPVs (recalculated by based on the Bayes theorem), and ORs of the combination of suspicious gray-scale US features were higher than those of the Rago and Asteria criteria, regardless of nodule size and presence of macro- or eggshell calcifications. The quasi-likelihood information criteria of gray-scale US were smaller than Rago and Asteria criteria, which meant that the gray-scale US was a suitable model for predicting malignancy. Most diagnostic performances of criteria set 1 were better than those of criteria sets 2 and 3 and all ORs of criteria set 1 were equal to or superior to those of criteria sets 2 or 3. Most of the quasi-likelihood information criteria of criteria set 1 were equal to or smaller than those of criteria sets 2 and 3. Therefore, elastography was not useful in predicting thyroid malignancy as an adjunctive diagnostic tool to gray-scale US or as a separate diagnostic tool.

Although Rago et al demonstrated that elastography with freehand technique shows high diagnostic performances even in nodules smaller than 10 mm (3), we found that specificities, PPVs, NPVs, accuracies, and ORs of both Rago and Asteria criteria in nodules larger than 10 mm and larger than 5 mm were inferior to those in nodules 10 mm or smaller and 5 mm or smaller, respectively. Nodules 5 mm or smaller, specifically, accuracies and ORs of gray-scale US, Rago criteria, Asteria criteria, criteria set 1, criteria set 2, and criteria set 3 were the smallest.

Suspicious gray-scale US features have high false-positive rates in nodules 5 mm or smaller (32). In our study, the false-positive rates of the three criteria sets were also high in nodules 5 mm or smaller. In nodules smaller than or equal to 5 mm, both gray-scale US and elastography have their limitations in predicting thyroid malignancy, although the clinical importance of these small cancers is not clear.

Our study had limitations. First, the reference standard was set as cytologic results for 482 nodules in 474 patients and as histopathologic results for 221 nodules in 202 patients, thus, including nodules without surgical confirmation. False-negative cytologic results may have existed. Second, cystic nodules were excluded from this study because elastography for cystic nodules does not give useful information (3). Third, most of 217 malignancies are papillary thyroid carcinomas and their variant form. There were no cases of follicular adenoma or follicular carcinoma and subgroups of follicular neoplasm have not been evaluated. There is a debate on the role of elastography in differentiating malignant from benign in nodules with indeterminate cytology (3,10,11,13,14). Studies to analyze elastography in follicular neoplasm are anticipated. Fourth, US examination, US-guided FNA, and elastography in this study were performed by eight radiologists, reflecting the expected variation among readers that inevitably exists in clinical practice. Studies on interobserver variability for US description and elastography at our institution have shown that US assessment showed a more than moderate degree of agreement (33) while elastography does not show reliable agreement (34). This may be due to the different US machines used and the fact that different radiologists were involved in obtaining elastography images. Fifth, this study did not focus on the diagnostic performances of elastography in subgroups such as benign nodules with suspicious assessment or malignant nodules with probably benign assessment at gray-scale US. To obtain more detail on the role of elastography regarding these subgroups, future studies are needed. Sixth, the malignancy rate of thyroid nodules with US-guided FNA was 30.9%. To compensate for the defect related to PPV and NPV, we have recalculated the PPN and NPV based on the Bayes theorem.

In conclusion, elastography, as well as the combination of elastography and gray-scale US, showed inferior performance in differentiating malignant and benign thyroid nodules in comparison with gray-scale US features. Elastography was not a useful tool in recommending FNA biopsy.

Disclosures of Potential Conflicts of Interest: H.J.M. No potential conflicts of interest to disclose. **J.M.S.** No potential conflicts of interest to disclose. **E.K.K.** No potential conflicts of interest to disclose. **J.H. Yoon** No potential conflicts of interest to disclose. **J.H. Youk** No potential conflicts of interest to disclose. **J.Y.K.** No potential conflicts of interest to disclose.

References

1. Gharib H, Papini E, Paschke R, et al. American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and EuropeanThyroid Association Medical Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules. Endocr Pract 2010;16(Suppl 1):1–43.
2. Bojunga J, Herrmann E, Meyer G, Weber S, Zeuzem S, Friedrich-Rust M. Real-time elastography for the differentiation of benign and malignant thyroid nodules: a meta-analysis. Thyroid 2010;20(10):1145–1150.
3. Rago T, Vitti P. Role of thyroid ultrasound in the diagnostic evaluation of thyroid nodules. Best Pract Res Clin Endocrinol Metab 2008;22(6):913–928.
4. Tan GH, Gharib H, Reading CC. Solitary thyroid nodule. Comparison between palpation and ultrasonography. Arch Intern Med 1995;155(22):2418–2423.
5. Gao L, Parker KJ, Lerner RM, Levinson SF. Imaging of the elastic properties of tissue—a review. Ultrasound Med Biol 1996;22(8):959–977.
6. Garra BS, Cespedes EI, Ophir J, et al. Elastography of breast lesions: initial clinical results. Radiology 1997;202(1):79–86.
7. Dighe M, Bae U, Richardson ML, Dubinsky TJ, Minoshima S, Kim Y. Differential diagnosis of thyroid nodules with US elastography using carotid artery pulsation. Radiology 2008;248(2):662–669.
8. Itoh A, Ueno E, Tohno E, et al. Breast disease: clinical application of US elastography for diagnosis. Radiology 2006;239(2):341–350.
9. Asteria C, Giovanardi A, Pizzocaro A, et al. US-elastography in the differential diagnosis of benign and malignant thyroid nodules. Thyroid 2008;18(5):523–531.
10. Friedrich-Rust M, Sperber A, Holzer K, et al. Real-time elastography and contrast-enhanced ultrasound for the assessment of thyroid nodules. Exp Clin Endocrinol Diabetes 2010;118(9):602–609.
11. Hong Y, Liu X, Li Z, Zhang X, Chen M, Luo Z. Real-time ultrasound elastography in the differential diagnosis of benign and malignant thyroid nodules. J Ultrasound Med 2009;28(7):861–867.
12. Lyshchik A, Higashi T, Asato R, et al. Thyroid gland tumor diagnosis at US elastography. Radiology 2005;237(1):202–211.
13. Rago T, Santini F, Scutari M, Pinchera A, Vitti P. Elastography: new developments in ultrasound for predicting malignancy in thyroid nodules. J Clin Endocrinol Metab 2007;92(8):2917–2922.
14. Rago T, Scutari M, Santini F, et al. Real-time elastosonography: useful tool for refining the presurgical diagnosis in thyroid nodules with indeterminate or nondiagnostic cytology. J Clin Endocrinol Metab 2010;95(12):5274–5280.
15. Rubaltelli L, Corradin S, Dorigo A, et al. Differential diagnosis of benign and malignant thyroid nodules at elastosonography. Ultraschall Med 2009;30(2):175–179.
16. Wang Y, Dan HJ, Dan HY, Li T, Hu B. Differential diagnosis of small single solid thyroid nodules using real-time ultrasound elastography. J Int Med Res 2010;38(2):466–472.
17. Kagoya R, Monobe H, Tojima H. Utility of elastography for differential diagnosis of benign and malignant thyroid nodules. Otolaryngol Head Neck Surg 2010;143(2):230–234.
18. Frates MC, Benson CB, Doubilet PM, Cibas ES, Marqusee E. Can color Doppler sonography aid in the prediction of malignancy of thyroid nodules? J Ultrasound Med 2003;22(2):127–131; quiz 132–134.
19. Khoo ML, Asa SL, Witterick IJ, Freeman JL. Thyroid calcification and its association with thyroid carcinoma. Head Neck 2002;24(7):651–655.

20. Kim EK, Park CS, Chung WY, et al. New sonographic criteria for recommending fine-needle aspiration biopsy of nonpalpable solid nodules of the thyroid. AJR Am J Roentgenol 2002;178(3):687–691.

21. Moon HJ, Kwak JY, Kim MJ, Son EJ, Kim EK. Can vascularity at power Doppler US help predict thyroid malignancy? Radiology 2010;255(1):260–269.

22. Moon WJ, Jung SL, Lee JH, et al. Benign and malignant thyroid nodules: US differentiation—multicenter retrospective study. Radiology 2008;247(3):762–770.

23. Papini E, Guglielmi R, Bianchini A, et al. Risk of malignancy in nonpalpable thyroid nodules: predictive value of ultrasound and color-Doppler features. J Clin Endocrinol Metab 2002;87(5):1941–1946.

24. Peccin S, de Castsro JA, Furlanetto TW, Furtado AP, Brasil BA, Czepielewski MA. Ultrasonography: is it useful in the diagnosis of cancer in thyroid nodules? J Endocrinol Invest 2002;25(1):39–43.

25. Ueno E, Itoh A. Diagnosis of breast cancer by elasticity imaging. Eizo Joho Med 2004;36(12):2–6.

26. Pan W. Akaike's information criterion in generalized estimating equations. Biometrics 2001;57(1):120–125.

27. Chung WY, Chang HS, Kim EK, Park CS. Ultrasonographic mass screening for thyroid carcinoma: a study in women scheduled to undergo a breast examination. Surg Today 2001;31(9):763–767.

28. 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(3):794–800.

29. Appetecchia M, Solivetti FM. The association of colour flow Doppler sonography and conventional ultrasonography improves the diagnosis of thyroid carcinoma. Horm Res 2006;66(5):249–256.

30. Cappelli C, Castellano M, Pirola I, et al. The predictive value of ultrasound findings in the management of thyroid nodules. QJM 2007;100(1):29–35.

31. 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(1):41–46.

32. Mazzaferri EL, Sipos J. Should all patients with subcentimeter thyroid nodules undergo fine-needle aspiration biopsy and preoperative neck ultrasonography to define the extent of tumor invasion? Thyroid 2008;18(6):597–602.

33. Choi SH, Kim EK, Kwak JY, Kim MJ, Son EJ. Interobserver and intraobserver variations in ultrasound assessment of thyroid nodules. Thyroid 2010;20(2):167–172.

34. Park SH, Kim SJ, Kim EK, Kim MJ, Son EJ, Kwak JY. Interobserver agreement in assessing the sonographic and elastographic features of malignant thyroid nodules. AJR Am J Roentgenol 2009;193(5):W416–W423.