Diagnostic performance of the modified Korean Thyroid Imaging Reporting and Data System for thyroid malignancy according to nodule size: a comparison with five society guidelines

Dong Gyu Na[1], Wooyul Paik[1], Jaehyung Cha[2], Hye Yun Gwon[1], Suh Young Kim[1], Roh-Eul Yoo[3]

[1]Department of Radiology, GangNeung Asan Hospital, University of Ulsan College of Medicine, Gangneung; [2]Medical Science Research Center, Korea University College of Medicine, Seoul; [3]Department of Radiology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea

Purpose: The aim of this study was to evaluate the diagnostic performance of the modified Korean Thyroid Imaging Reporting and Data System (K-TIRADS) compared with five society risk stratification systems (RSSs) according to nodule size.

Methods: In total, 3,826 consecutive thyroid nodules (≥1 cm) with final diagnoses in 3,088 patients were classified according to five RSSs. The K-TIRADS was modified by raising the biopsy size threshold for low-suspicion nodules and subcategorizing intermediate-suspicion nodules. We assessed the performance of the RSSs as triage tests and their diagnostic accuracy according to nodule size (with a threshold of 2 cm).

Results: Of all nodules, 3,277 (85.7%) were benign and 549 (14.3%) were malignant. In small thyroid nodules (≤2 cm), the American College of Radiology Thyroid Imaging Reporting and Data System (ACR TI-RADS) had the highest reduction rate of unnecessary biopsies (76.3%) and the lowest sensitivity (76.1%). The modified K-TIRADS had the second highest reduction rate of unnecessary biopsies (67.6%) and sensitivity (86.6%). The modified K-TIRADS and ACR TI-RADS had the highest diagnostic odds ratios (P=0.165) and the highest areas under the curve (P=0.315). In large nodules (>2 cm), the sensitivity of the ACR TI-RADS for malignancy was significantly lower (88.8%) than the sensitivities of the modified K-TIRADS and other RSSs, which were very high (98.7%–99.3%) (P<0.001).

Conclusion: The modified K-TIRADS allows a large proportion of unnecessary biopsies to be avoided, while maintaining high sensitivity and diagnostic accuracy for small malignant tumors and very high sensitivity for large malignant tumors.

Keywords: Thyroid nodule; Thyroid neoplasms; Ultrasonography; Diagnosis

Introduction

Ultrasonography (US) is a primary diagnostic tool for the evaluation of thyroid nodules [1] and many international societies have proposed widely used US risk stratification systems (RSSs) for thyroid nodules in clinical practice guidelines [2–7]. RSSs are used for triage to select patients for US-guided aspiration/biopsy and to rule out thyroid malignancy. As triage tests, RSSs play a role in reducing unnecessary nodule biopsies and require an appropriate sensitivity for thyroid malignancy [8]. Recent comparative studies [9–14] showed a wide spectrum of diagnostic performance for the biopsy criteria in the five US RSSs: the American Association of Clinical Endocrinologists (AACE)/American College of Endocrinology (ACE)/Associazione Medici Endocrinologi (AME) guideline, the American College of Radiology (ACR) Thyroid Imaging Reporting and Data System (TI-RADS) proposed by the ACR, the American Thyroid Association (ATA) Guideline, the European Thyroid Imaging Reporting and Data System (EU-TIRADS) proposed by the European Thyroid Association, and the Korean Thyroid Imaging Reporting and Data System (K-TIRADS) proposed by the Korean Society of Thyroid Radiology/Korean Thyroid Association. Previous comparative study results [9–14] have raised the need to find balanced optimal biopsy criteria within an RSS, and have also shown that the K-TIRADS had the highest sensitivity and highest rate of unnecessary biopsies. In light of this finding, it is necessary to modify the K-TIRADS to reduce the rate of unnecessary biopsies while maintaining an appropriate sensitivity for malignancy.

Tumor size is an important prognostic factor in papillary thyroid cancer (PTC) and follicular thyroid cancer (FTC) [15,16]. The risk of distant metastasis increases for tumors larger than 2 cm [17] and the risk of local tumor invasion, nodal metastasis, and distant metastasis becomes higher as tumor size increases [18]. Therefore, the diagnostic performance of RSSs needs to be evaluated depending on the nodule size, but this has rarely been investigated. The aim of this study was to develop a modified version of the K-TIRADS and to evaluate the diagnostic performance of the modified K-TIRADS compared with the five RSSs as triage tests for the detection of thyroid malignancy according to nodule size.

Materials and Methods

Compliance with Ethical Standards

This study was approved by the institutional review board of GangNeung Asan Hospital in Korea (2020-03-020) and informed consent was waived for this retrospective study.

Study Population

Overall, 4,359 consecutive patients underwent US-guided fine-needle aspiration (FNA) or core needle biopsy (CNB) for thyroid nodules between January 2011 and December 2019 at a single tertiary hospital. Among 3,905 patients with 4,832 nodules ≥1 cm, 998 nodules without final diagnoses confirmed through surgical or biopsy findings (non-diagnostic biopsy results [n=481], atypical or follicular lesion of undetermined clinical significance [n=424], follicular neoplasm or suspected follicular neoplasm [n=47], suspicious for malignancy [n=27], and one benign FNA result with a subsequent discordant biopsy result of follicular neoplasm or suspicious follicular neoplasm and suspicious for malignancy [n=19]), and eight nodules with US images of suboptimal quality were excluded. The remaining 3,088 patients with 3,826 nodules were included in the final study
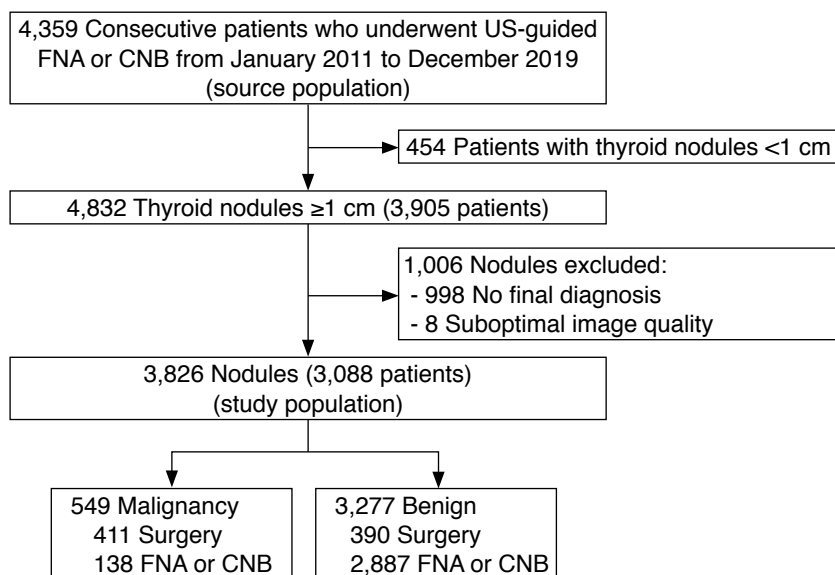
Fig. 1. Flowchart of study participants. US, ultrasonography; FNA, fine-needle aspiration; CNB, core needle biopsy.

population (2,497 women and 591 men; median age, 56 years; interquartile range [IQR], 47 to 64 years) (Fig. 1). Final diagnoses were determined by the definitive FNA or CNB results (benign or malignant) and surgical histologic diagnoses.

US Examinations and Image Analysis

All US examinations were performed using a 5- to 12-MHz linear probe and a real-time US system (IU22 or EPIQ7,Philips Medical Systems, Bothell, WA, USA). All US images of thyroid nodules between January 2011 and February 2017 were obtained according to the Korean Society of Thyroid Radiology guidelines [5,19] and the US images were retrospectively reviewed by one experienced radiologist (D.G.N.) with 22 years of experience in performing thyroid US, who had no previous knowledge of the FNA results or final diagnoses. The US images of the thyroid nodules obtained between March 2017 and December 2019 were prospectively evaluated before biopsy by two radiologists (D.G.N. and W.P.) with 22 years and 4 years of experience in performing thyroid US, respectively.

The US features of nodules were strictly assessed using the definitions in the US lexicons of the five RSSs [3–7] to minimize the misclassification of nodules in both the retrospective and prospective datasets (Supplementary Table 1). Extrathyroidal extension status was not evaluated in this study because of the absence of standardized specified US criteria. An isolated macrocalcification was defined as an entirely calcified nodule with posterior acoustic shadowing, in which no soft-tissue component was identified due to dense shadowing on the US image [20]. A nodule with this finding was categorized as a nodule of intermediate suspicion in the K-TIRADS [5], as a nodule of moderate suspicion (4 points) in the ACR TI-RADS [6,11], and as an unclassified nodule in other RSSs. Isoechoic nodules with an irregular margin, microcalcification, and a taller-than-wide shape were categorized as unclassified nodules in the ATA guideline. A reviewer (D.G.N.), who had no previous knowledge of the FNA results or final diagnoses, classified nodules based on the assessed US features, and determined the candidates for FNA based on the maximal diameter and category of each nodule according to the guideline of each RSS or TI-RADS (Supplementary Table 2).

Biopsy Size Thresholds According to Categories in the Five Risk Stratification Systems

Supplementary Table 2 lists the biopsy size thresholds and calculated malignancy risks according to the categories in the five RSSs. Nodules classified as low-risk by the AACE/ACE/AME guideline, not suspicious (TR2) or benign (TR1) by the ACR TI-RADS, or benign by the ATA guideline, EU-TIRADS, and K-TIRADS were considered not to

be indicated for biopsy in this study because they are not routinely indicated for biopsy for diagnostic purposes according to each RSS.

Development of the Modified Korean TI-RADS

The modified K-TIRADS was developed by revising the K-TIRADS (Table 1). Intermediate suspicion (K-TIRADS 4) nodules were subcategorized into K-TIRADS 4A and 4B based on the malignancy risk of the US patterns (Table 1). K-TIRADS category 4 includes solid hypoechoic nodules without any of three suspicious US features (microcalcification, nonparallel orientation [taller-than-wide], spiculated/microlobulated margin) and partially cystic or isoechoic and hyperechoic nodules with any of the three suspicious US features. Solid hypoechoic nodules without any of the three suspicious US features were subcategorized by the degree of hypoechogenicity (mild vs. marked hypoechogenicity) and macrocalcification based on the results of previous studies reporting that hypoechogenicity and macrocalcification increased the malignancy risk of solid hypoechoic nodules without any of the three suspicious US features (Shin HS, unpublished data) [21]. Marked hypoechogenicity was defined as similar echogenicity or hypoechogenicity relative to the anterior neck muscle [21]. Partially cystic or isoechoic and hyperechoic nodules with any of the three suspicious US features were subcategorized according to the number of coexisting three suspicious US features (one vs. two or three) because the presence of a higher number of suspicious US features in a nodule may indicate a higher malignancy risk [22]. The size threshold for biopsy was subdivided into 1 cm for K-TIRADS 4B and 1.5 cm for K-TIRADS 4A, and the size threshold for biopsy was raised from 1.5 cm to 2 cm for low suspicion (K-TIRADS 3) nodules.

Assessment of the Diagnostic Performance of Risk Stratification Systems for Thyroid Malignancy

All nodules were dichotomized into those for which a biopsy was indicated (test positivity) or was not indicated (test negativity) by the biopsy criteria of each RSS (Table 1, Supplementary Table 2). As primary measures of test performance, the reduction rate of unnecessary biopsies and sensitivity were used to assess the performance of each RSS as a triage test [8] and the negative likelihood ratio (LR−) was used to assess its performance as a rule-out test [23]. The LR indicates how much a positive or negative test result by the RSS raises or lowers the pretest probability of the target disorder (thyroid malignancy) [23,24]. The global discriminative performance was assessed as a secondary measure of the diagnostic performance of the RSSs by the diagnostic odds ratio (DOR) and the area under the receiver operating characteristic curve (AUC). The DOR is a single indicator of test performance and is independent of the prevalence of malignant tumors. The DOR is equal to the positive

likelihood ratio (LR+) divided by LR−, and is the ratio of the odds of a biopsy being indicated in a malignant nodule relative to the odds of a biopsy being indicated in a benign nodule [25].

Statistical Analyses

The chi-square or Fisher exact test was used to compare the frequency of categorical variables. Multivariable logistic regression analyses were performed to determine independent US predictors among US features in the subgroup of K-TIRADS 4 nodules. Sensitivity, specificity, positive predictive value, negative predictive value, LR+, LR−, DOR, and the AUC were calculated with 95% confidence intervals. All diagnostic values were compared among the RSSs in overall nodules and according to nodule size (with a size threshold of 2 cm). The statistical comparisons of the LR and DOR among the RSSs were performed using a regression model approach proposed by Gu and Pepe [26] and the Z test, respectively. The DeLong test was used to compare the AUC among the RSSs. Statistical analyses were performed using SPSS version 25 for Windows (IBM Corp., Armonk, NY, USA) and R 3.6.3 for Windows

(R Development Core Team, Vienna, Austria). A significant difference was defined as a P-value <0.05.

Results

Clinical Data

The median size (maximal diameter) of the nodules was 1.7 cm (IQR, 1.3 to 2.6 cm; range, 1 to 10 cm). The maximal diameter of nodules was small (1–2 cm) in 2,385 nodules (62.3%) (median size, 1.4 cm) and large (>2 cm) in 1,441 nodules (37.7%) (median size, 2.9 cm). Of the 3,826 nodules, 3,277 (85.7%) were benign and 549 (14.3%) were malignant. Malignant nodules were diagnosed based on histologic findings after surgery (n=411) or malignant FNA or CNB results (n=138). Benign nodules were diagnosed based on histologic findings after surgery (n=390), at least two benign FNA or CNB results (n=545), and one benign FNA (n=2,055) or CNB result (n=287) (Fig. 1).

The 549 malignant nodules included 494 PTCs (90.0%), 32 FTCs (5.8%), eight anaplastic carcinomas (1.5%), eight metastases

Table 1. Malignancy risk of classified nodules and biopsy size thresholds in the modified K-TIRADS

Category	US pattern	Benign nodules (n=3,277)	Malignant nodules (n=549)	Total (n=3,826)	Calculated risk of malignancy (%)	Size thresholds for biopsy (cm)
5 High suspicion	Solid hypoechoic nodule with any of 3 suspicious US features[a]	167	325	492 (12.9)	66.1 (61.9–70.2)	≥1.0
4 Intermediate suspicion	(1) Solid hypoechoic nodule without any of 3 suspicious US features[a] or (2) Partially cystic or isoechoic and hyperechoic nodule with any of 3 suspicious US features[a]	1,103	176	1,279 (33.4)	13.8 (11.9–15.6)	
4B	(1) Solid hypoechoic nodule with marked hypoechogenicity or macrocalcification[b] (2) Partially cystic or isoechoic and hyperechoic nodule with 2 or 3 suspicious US features[a]	189	83	272 (7.1)	30.5 (25.0–36.0)	≥1.0
4A	(1) Solid mildly hypoechoic nodule without macrocalcification[b] (2) Partially cystic or isoechoic and hyperechoic nodule with 1 suspicious US feature[a]	914	93	1,007 (26.3)	9.2 (7.4–11.0)	≥1.5
3 Low suspicion	Partially cystic or iso/hyperechoic nodule without any of the 3 suspicious US features[a]	1,881	46	1,927 (50.4)	2.4 (1.7–3.1)	≥2.0
2 Benign	(1) Spongiform (2) Partially cystic nodule with intracystic comet-tail artifact (3) Pure cyst	126	2	128 (3.3)	1.6 (0.0–3.7)	Not indicated[c]
1 No nodule	−	−	−	−	−	−

Values are presented as number (%) or risk (95% CI).

K-TIRADS, Korean Thyroid Imaging Reporting and Data System; US, ultrasonography; CI, confidence interval.

[a]Microcalcification, nonparallel orientation (taller-than-wide), spiculated/microlobulated (irregular) margin. [b]Includes isolated macrocalcifications. [c]Not routinely indicated, but may be performed for therapeutic aspiration of the cystic content and for diagnosis prior to ablation therapy or surgery.
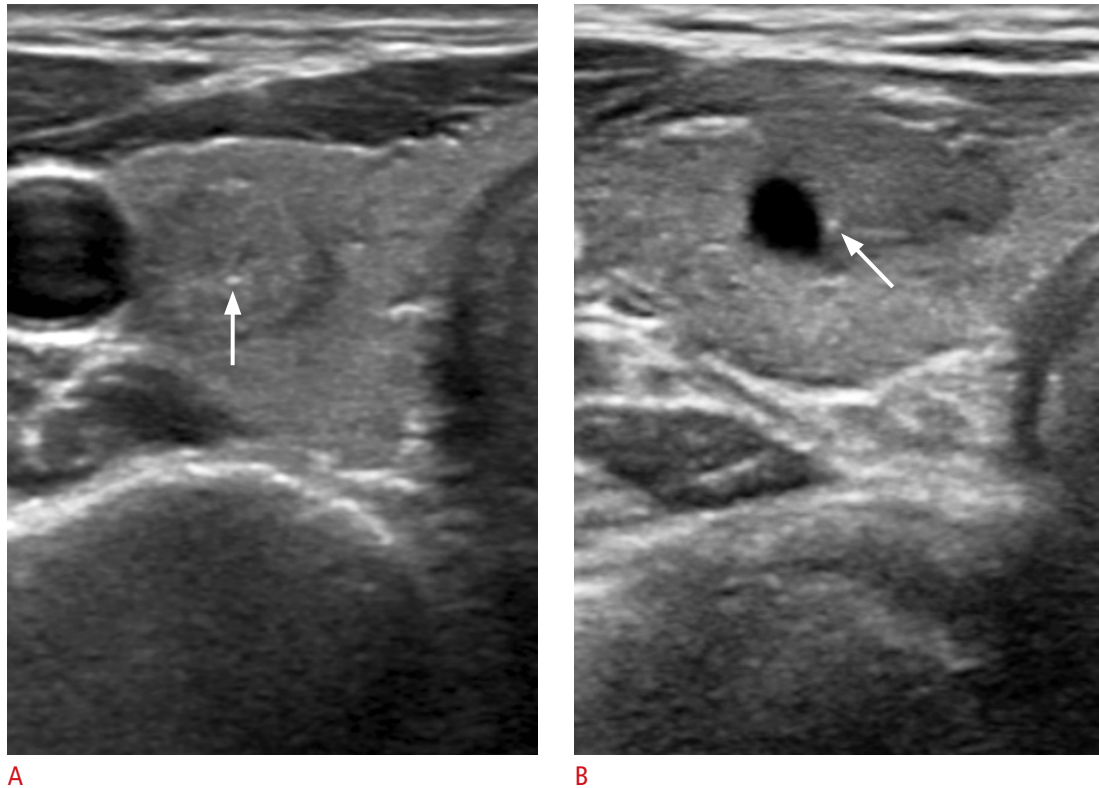
A   B
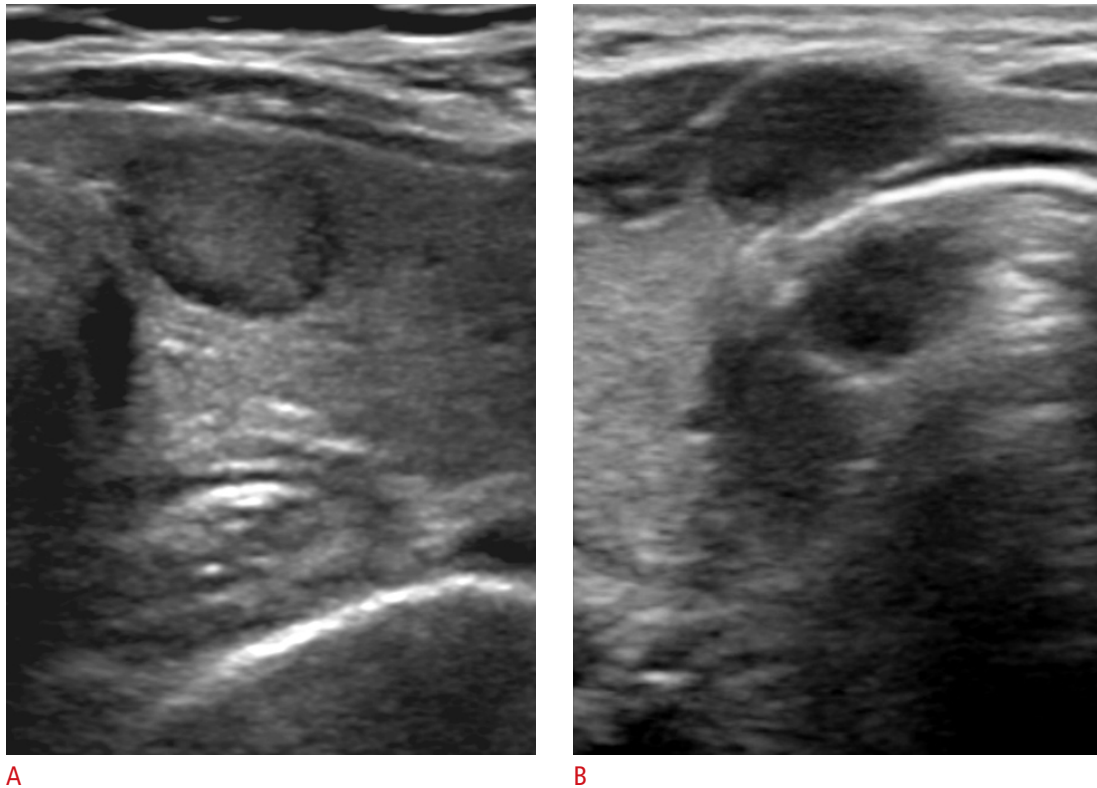
Fig. 2. Modified K-TIRADS 4A nodule with a solid hypoechoic US pattern in a 66-year-old woman (A) and modified K-TIRADS 4B nodule in a 29-year-old woman (B).

A. Transverse US shows a solid and mildly hypoechoic nodule (12 mm) without any suspicious US features (microcalcification, spiculated or microlobulated margin, or nonparallel orientation) and macrocalcification in the right thyroid lobe. This nodule is classified as intermediate-risk by the AACE/ACE/AME guideline, moderately suspicious (TR4) by the ACR TI-RADS, intermediate suspicion by the ATA guideline, intermediate-risk (TIRADS 4) by the EU-TIRADS, and intermediate suspicion (TIRADS 4) by the K-TIRADS. Final diagnosis: nodular hyperplasia by surgery. B. Transverse US shows a solid and markedly hypoechoic nodule (11 mm) without any suspicious US features and macrocalcification at the isthmus of thyroid. This nodule is classified as high-risk by the AACE/ACE/AME guideline, moderately suspicious (TR4) by ACR TI-RADS, intermediate suspicion by the ATA guideline, high-risk (TIRADS 5) by the EU-TIRADS, and intermediate suspicion (TIRADS 4) by the K-TIRADS. Final diagnosis: papillary thyroid carcinoma by surgery. K-TIRADS, Korean Thyroid Imaging Reporting and Data System; US, ultrasonography; AACE, American Association of Clinical Endocrinologists; ACE, American College of Endocrinology; AME, Associazione Medici Endocrinologi; ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System; ATA, American Thyroid Association; EU-TIRADS, European Thyroid Imaging Reporting and Data System.

Fig. 3. Modified K-TIRADS 4B nodule with solid hypoechoic US pattern in a 76-year-old man. Transverse US shows a solid mildly hypoechoic nodule (11 mm) with macrocalcification and no suspicious features in the right thyroid lobe. This nodule is classified as intermediate-risk by the AACE/ACE/AME guideline, moderately suspicious (TR4) by the ACR TI-RADS, intermediate suspicion by the ATA guideline, intermediate-risk (TIRADS 4) by the EU-TIRADS, and intermediate suspicion (TIRADS 4) by the K-TIRADS. Final diagnosis: papillary thyroid carcinoma by surgery. K-TIRADS, Korean Thyroid Imaging Reporting and Data System; US, ultrasonography; AACE, American Association of Clinical Endocrinologists; ACE, American College of Endocrinology; AME, Associazione Medici Endocrinologi; ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System; ATA, American Thyroid Association; EU-TIRADS, European Thyroid Imaging Reporting and Data System.

Fig. 4. Modified K-TIRADS 4A nodules with isoechoic or partially cystic US pattern and one suspicious US feature in a 39-year-old woman (A) and in a 46-year-old woman (B).
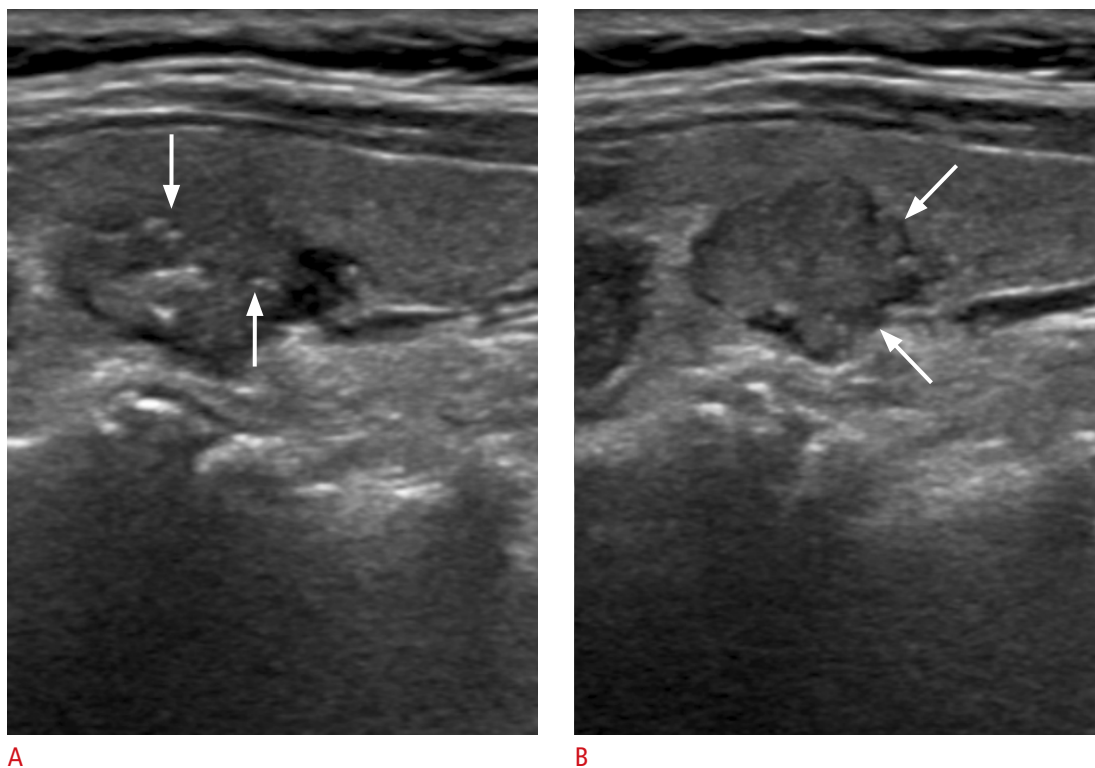A. Transverse US shows a solid and isoechoic nodule (14 mm) with punctate echogenic focus (microcalcification) (arrow) in the right thyroid lobe. This nodule is classified as high-risk by the AACE/ACE/AME guideline, moderately suspicious (TR4) by the ACR TI-RADS, an unclassified nodule by the ATA guideline, high-risk (TIRADS 5) by the EU-TIRADS, and intermediate suspicion (TIRADS 4) by the K-TIRADS. Final diagnosis: nodular hyperplasia by surgery. B. Transverse US shows a partially cystic and mildly hypoechoic nodule (12 mm) with an intrasolid punctate echogenic focus (microcalcification) (arrow) in the right thyroid lobe. This nodule is classified as high-risk by the AACE/ACE/AME guideline, moderately suspicious (TR4) by ACR TI-RADS, high suspicion by the ATA guideline, high-risk (TIRADS 5) by the EU-TIRADS, and intermediate suspicion (TIRADS 4) by the K-TIRADS. Final diagnosis: nodular hyperplasia by two benign findings on fine-needle aspiration and core needle biopsy. K-TIRADS, Korean Thyroid Imaging Reporting and Data System; US, ultrasonography; AACE, American Association of Clinical Endocrinologists; ACE, American College of Endocrinology; AME, Associazione Medici Endocrinologi; ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System; ATA, American Thyroid Association; EU-TIRADS, European Thyroid Imaging Reporting and Data System.

(1.5%), four lymphomas (0.7%), and three medullary thyroid carcinomas (0.5%). The proportion of PTCs among malignant tumors was significantly higher in small (1–2 cm) tumors than in large (>2 cm) tumors (96.5% vs. 73.0%, P<0.001), and the proportion of FTCs was significantly higher in large (>2 cm) malignant tumors than in small malignant tumors (1–2 cm) (15.8% vs. 2.0%, P<0.001).

Subcategorization of the Malignancy Risk of Intermediate-Suspicion Nodules in K-TIRADS

Table 1 lists the malignancy risk of nodules classified using the modified K-TIRADS, in which intermediate suspicion (K-TIRADS 4) nodules were subcategorized (Figs. 2–5). Among

the solid hypoechoic nodules without any suspicious US features (microcalcification, nonparallel orientation, spiculated/microlobulated margin) classified as K-TIRADS 4, marked hypoechogenicity and macrocalcification were independently predictive of malignancy (P<0.001 for both) in the multivariable analysis. Markedly hypoechoic nodules showed a significantly higher malignancy risk than mildly hypoechoic nodules (27.1% vs. 11.2%, P<0.001), and nodules with macrocalcifications showed a significantly higher malignancy risk than nodules without macrocalcifications (38.9% vs. 13.4%, P<0.001). Among the partially cystic or isoechoic and hyperechoic nodules with suspicious US features classified as K-TIRADS 4, the malignancy risk of nodules with two or three suspicious features was significantly higher than

Fig. 5. Modified K-TIRADS 4B nodule with a partially cystic US pattern and two suspicious US features in a 66-year-old woman. Longitudinal US shows a partially cystic and mildly hypoechoic nodule (12 mm) with multiple intrasolid punctate echogenic foci (microcalcifications) (arrows) (A) and a spiculated/microlobulated (irregular) margin (arrows) (B) in the left thyroid lobe. This nodule is classified as high-risk by the AACE/ACE/AME guideline, highly suspicious (TR5) by ACR TI-RADS, high suspicion by the ATA guideline, high-risk (TIRADS 5) by the EU-TIRADS, and intermediate suspicion (TIRADS 4) by the K-TIRADS. Final diagnosis: papillary thyroid carcinoma by surgery. K-TIRADS, Korean Thyroid Imaging Reporting and Data System; US, ultrasonography; AACE, American Association of Clinical Endocrinologists; ACE, American College of Endocrinology; AME, Associazione Medici Endocrinologi; ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System; ATA, American Thyroid Association; EU-TIRADS, European Thyroid Imaging Reporting and Data System.

that of nodules with one suspicious US feature (48.2% vs. 8.3%, P<0.001).

Diagnostic Performance of the Modified K-TIRADS and five RSSs in All Nodules

Table 2 presents the diagnostic performance of the biopsy criteria of the RSSs for malignancy in all thyroid nodules. The ACR TI-RADS had the highest reduction rate of unnecessary biopsies (65.2%), lowest sensitivity (79.6%), highest specificity (65.2%), and a relatively high LR− (0.31). The K-TIRADS had the lowest reduction rate of unnecessary biopsies (18.6%), highest sensitivity (96.9%), lowest specificity (18.6%), and lowest LR− (0.17). The modified K-TIRADS had the second highest reduction rate of unnecessary biopsies (43.7%) and sensitivity (90.0%), and the second lowest LR− (0.23). The ACR TI-RADS had the highest DOR (7.32) and AUC (0.724).

Diagnostic Performance of the Modified K-TIRADS and Five Risk Stratification Systems in Small Nodules (≤2 cm)

Table 3 shows the diagnostic performance of the biopsy criteria of the RSSs for malignancy in small thyroid nodules (≤2 cm). The reduction rate of unnecessary biopsies was the highest with the ACR TI-RADS (76.3%) and was the lowest with the K-TIRADS (27.4%). The modified K-TIRADS had the second highest reduction rate of unnecessary biopsies (67.6%), which was similar to that of the AACE/ACE/AME (P=0.343) and higher than those of the K-TIRADS, ATA guideline, and EU-TIRADS (P<0.001 for all). The highest sensitivity (96.2%) was found for the K-TIRADS and the lowest (76.1%) for the ACR TI-RADS. The modified K-TIRADS had the second highest sensitivity (86.6%), which was similar to the sensitivities of the ATA guideline (P=0.492) and the EU-TIRADS (P=0.174).

The K-TIRADS and the modified K-TIRADS had the lowest LR− values (0.14 and 0.20, respectively; P=0.106) which were

Table 2. Diagnostic performance of biopsy criteria by the modified K-TIRADS and five risk stratification systems in all nodules (n=3,826)

Ultrasound risk stratification systems	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR−	DOR	AUC	Reduction rate of unnecessary biopsies (%)[a]
Modified K-TIRADS	90.0 (494/549) [87.2−92.3]	43.7 (1,433/3,277) [42.0−45.4]	21.1 (494/2,338) [19.5−22.8]	96.3 (1,433/1,488) [95.2−97.2]	1.60 [1.53−1.67]	0.23 [0.18−0.30]	6.98 [5.24−9.30]	0.669 [0.653−0.683]	43.7 [42.0−45.4]
AACE/ACE/AME	84.7 (465/549) [81.4−87.6]	42.5 (1,392/3,277) [40.8−44.2]	19.8 (465/2,350) [18.2−21.5]	94.3 (1,392/1,476) [93.0−95.4]	1.47 [1.41−1.54]	0.36 [0.29−0.44]	4.09 [3.21−5.21]	0.636 [0.620−0.651]	42.5 [40.8−44.2]
ACR TI-RADS	79.6 (437/549) [76.0−82.9]	65.2 (2,138/3,277) [63.6−66.9]	27.7 (437/1,576) [25.5−30.0]	95.0 (2,138/2,250) [94.0−95.9]	2.29 [2.15−2.44]	0.31 [0.26−0.37]	7.32 [5.88−9.12]	0.724 [0.710−0.738]	65.2 [63.6−66.9]
ATA	84.0 (461/549) [80.6−86.9]	41.6 (1,362/3,277) [39.9−43.3]	19.4 (461/2,376) [17.8−21.1]	93.9 (1,362/1,450) [92.6−95.1]	1.44 [1.37−1.51]	0.39 [0.32−0.47]	3.73 [2.94−4.73]	0.628 [0.612−0.643]	41.6 [39.9−43.2]
EU-TIRADS	88.3 (485/549) [85.4−90.9]	33.4 (1,093/3,277) [31.7−35.0]	18.2 (485/2,669) [16.7−19.7]	94.5 (1,093/1,157) [93.0−95.7]	1.33 [1.28−1.38]	0.35 [0.28−0.44]	3.79 [2.89−4.97]	0.608 [0.593−0.624]	33.4 [31.7−35.0]
K-TIRADS	96.9 (532/549) [95.1−98.2]	18.6 (611/3,277) [17.3−20.0]	16.6 (532/3,198) [15.4−18.0]	97.3 (611/628) [95.7−98.4]	1.19 [1.16−1.22]	0.17 [0.10−0.27]	7.17 [4.39−11.72]	0.578 [0.562−0.593]	18.6 [17.3−20.0]

Values in parentheses are number of nodules; values in square brackets are 95% confidence intervals.
K-TIRADS, Korean Thyroid Imaging Reporting and Data System; PPV, positive predictive value; NPV, negative predictive value; LR+, likelihood ratio for positive results; LR−, likelihood ratio for negative results; DOR, diagnostic odds ratio; AUC, area under the receiver operating characteristic curve; AACE/ACE/AME, American Association of Clinical Endocrinologists/American College of Endocrinology/Associazione Medici Endocrinologi Medical Guidelines; ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System; ATA, American Thyroid Association Management Guideline; EU-TIRADS, European Thyroid Imaging Reporting and Data System.
[a]The reduction rate of unnecessary biopsies indicates the proportion of nodules not indicated for biopsy among all benign nodules.

Table 3. Diagnostic performance of biopsy criteria by the modified K-TIRADS and five risk stratification systems in small nodules (≤2 cm) (n=2,385)

Ultrasound risk stratification systems	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR−	DOR	AUC	Reduction rate of unnecessary biopsies (%)[a]
Modified K-TIRADS	86.6 (344/397) [82.9−89.8]	67.6 (1,344/1,988) [65.5−69.7]	34.8 (344/988) [31.8−37.9]	96.2 (1,344/1,397) [95.1−97.1]	2.67 [2.48−2.88]	0.20 [0.15−0.25]	13.55 [9.99−18.36]	0.771 [0.754−0.788]	67.6 [65.5−69.7]
AACE/ACE/AME	79.6 (316/397) [75.3−83.5]	66.6 (1,325/1,988) [64.5−68.7]	32.3 (316/979) [29.4−35.3]	94.2 (1,325/1,406) [92.9−95.4]	2.39 [2.20−2.58]	0.31 [0.25−0.37]	7.80 [6.00−10.12]	0.731 [0.713−0.749]	66.6 [64.6−68.7]
ACR TI-RADS	76.1 (302/397) [71.6−80.2]	76.3 (1,517/1,988) [74.4−78.2]	39.1 (302/773) [35.6−42.6]	94.1 (1,517/1,612) [92.8−95.2]	3.21 [2.92−3.54]	0.31 [0.26−0.37]	10.24 [7.95−13.18]	0.762 [0.744−0.779]	76.3 [74.4−78.2]
ATA	85.4 (339/397) [81.5−88.7]	55.2 (1,098/1,988) [53.0−57.4]	27.6 (339/1,229) [25.1−30.2]	95.0 (1,098/1,156) [93.6−96.2]	1.91 [1.79−2.03]	0.26 [0.21−0.34]	7.21 [5.38−9.66]	0.703 [0.684−0.721]	55.2 [53.0−57.4]
EU-TIRADS	84.1 (334/397) [80.2−87.6]	54.2 (1,077/1,988) [52.0−56.4]	26.8 (334/1,245) [24.4−29.4]	94.5 (1,077/1,140) [93.0−95.7]	1.84 [1.72−1.96]	0.29 [0.23−0.37]	6.27 [4.72−8.32]	0.692 [0.673−0.710]	54.2 [52.0−56.4]
K-TIRADS	96.2 (382/397) [93.8−97.9]	27.4 (544/1,988) [25.4−29.4]	20.9 (382/1,826) [19.1−22.9]	97.3 (544/559) [95.6−98.5]	1.32 [1.28−1.37]	0.14 [0.08−0.23]	9.59 [5.67−16.22]	0.618 [0.598−0.637]	27.4 [25.4−29.4]

Values in parentheses are number of nodules; values in square brackets are 95% confidence intervals.
K-TIRADS, Korean Thyroid Imaging Reporting and Data System; PPV, positive predictive value; NPV, negative predictive value; LR+, likelihood ratio for positive results; LR−, likelihood ratio for negative results; DOR, diagnostic odds ratio; AUC, area under the receiver operating characteristic curve; AACE/ACE/AME, American Association of Clinical Endocrinologists/American College of Endocrinology/Associazione Medici Endocrinologi Medical Guidelines; ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System; ATA, American Thyroid Association Management Guideline; EU-TIRADS, European Thyroid Imaging Reporting and Data System.
[a]The reduction rate of unnecessary biopsies indicates the proportion of nodules not indicated for biopsy among all benign nodules.

significantly lower than those of the other RSSs (P<0.05). The modified K-TIRADS and the ACR TI-RADS had the highest DORs (13.55 and 10.24, respectively; P=0.165) and the highest AUCs (0.771 and 0.762, respectively; P=0.315). The ATA guideline had a reduction rate of unnecessary biopsies of 37.7%, a sensitivity of 95.7%, an LR− of 0.11, an AUC of 0.667, and a DOR of 13.54 when the unclassified nodules were categorized as intermediate-suspicion nodules.

Diagnostic Performance of the Modified K-TIRADS and Five Risk Stratification Systems in Large Nodules (>2 cm)

Table 4 lists the diagnostic performance of the biopsy criteria of the RSSs for malignancy in large thyroid nodules (>2 cm). The reduction rate of unnecessary biopsies was highest with the ACR TI-RADS (48.2%); this rate was significantly higher than those of other RSSs (1.2%−20.5%) (P<0.001 for all). The sensitivity of the ATA (80.3%) and the ACR TI-RADS (88.8%) was significantly lower than those of other RSSs, which had similarly very high sensitivities (98.7%−99.3%) (P≤0.001 for all). The LR- was lowest (0.19) with the modified K-TIRADS. The DOR was highest with the ACR TI-RADS (7.38) and second highest with the modified K-TIRADS (5.56) (P=0.712). The AUC was highest (0.685) with the ACR TI-RADS; this

AUC was significantly higher than the AUCs (0.503−0.528) of the other RSSs (P<0.001). The ATA guideline had a reduction rate of unnecessary biopsies of 0.9%, a sensitivity of 100.0%, a specificity of 0.9%, a LR− of 0.00, an AUC of 0.505, and no calculable DOR when the unclassified nodules were categorized as intermediate-suspicion nodules. Seventeen malignant tumors were missed by the ACR TI-RADS, of which 15 were classified as TR2 and 2 as TR3, and the histologic types of these tumors were PTC in 11 cases, including six follicular variant PTCs, and FTC in six cases. The malignant tumors classified as TR2 by the ACR TI-RADS accounted for 15 of the 152 malignant tumors larger than 2 cm (9.9%) (Fig. 6).

Discussion

The modified K-TIRADS substantially reduced the number of unnecessary biopsies compared to the K-TIRADS, while maintaining a relatively high sensitivity (86.6%) in small nodules (≤2 cm), by raising the size thresholds for biopsy in low suspicion (K-TIRADS 3) and subcategorizing intermediate suspicion (K-TIRADS 4A) nodules. Meanwhile, the modified K-TIRADS, K-TIRADS, EU-TIRADS, and AACE/ACE/AME guidelines had similarly very high sensitivities and very low reduction rates of unnecessary biopsies, whereas the ACR
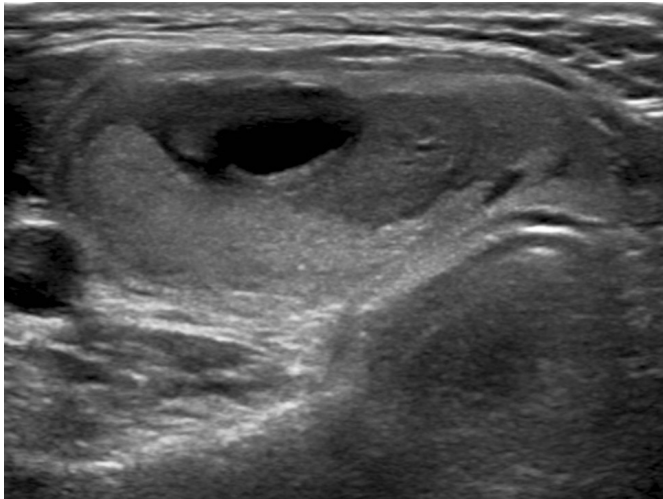
Table 4. Diagnostic performance of biopsy criteria by the modified K-TIRADS and five risk stratification systems in large nodules (>2 cm) (n=1,441)

Ultrasound risk stratification systems	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR−	DOR	AUC	Reduction rate of unnecessary biopsies (%)[a]
Modified K-TIRADS	98.7 (150/152) [95.3−99.8]	6.9 (89/1,289) [5.6−8.4]	11.1 (150/1,350) [9.5−12.9]	97.8 (89/91) [92.3−99.7]	1.06 [1.04−1.09]	0.19 [0.05−0.77]	5.56 [1.36−22.82]	0.528 [0.502−0.554]	6.9 [5.5−8.4]
AACE/ACE/AME	98.0 (149/152) [94.3−99.6]	5.2 (67/1,289) [4.1−6.6]	10.9 (149/1,371) [9.3−12.6]	95.7 (67/70) [88.0−99.1]	1.03 [1.01−1.06]	0.38 [0.12−1.19]	2.72 [0.85−8.77]	0.516 [0.490−0.542]	5.2 [4.1−6.6]
ACR TI-RADS	88.8 (135/152) [82.7−93.3]	48.2 (621/1,289) [45.4−50.9]	16.8 (135/803) [14.3−19.6]	97.3 (621/638) [95.8−98.4]	1.71 [1.59−1.85]	0.23 [0.15−0.36]	7.38 [4.41−12.37]	0.685 [0.660−0.709]	48.2 [45.4−50.9]
ATA	80.3 (122/152) [73.0−86.3]	20.5 (264/1,289) [18.3−22.8]	10.6 (122/1,147) [8.9−12.6]	89.8 (264/294) [85.8−93.0]	1.01 [0.93−1.10]	0.96 [0.69−1.35]	1.05 [0.69−1.60]	0.504 [0.478−0.530]	20.5 [18.3−22.8]
EU-TIRADS	99.3 (151/152) [96.4−100.0]	1.2 (16/1,289) [0.7−2.0]	10.6 (151/1,424) [9.1−12.3]	94.1 (16/17) [71.3−99.9]	1.01 [0.99−1.02]	0.53 [0.07−3.97]	1.90 [0.25−14.41]	0.503 [0.477−0.529]	1.2 [0.7−2.0]
K-TIRADS	98.7 (150/152) [95.3−99.8]	5.2 (67/1,289) [4.1−6.6]	10.9 (150/1,372) [9.3−12.7]	97.1 (67/69) [89.9−99.6]	1.04 [1.02−1.06]	0.25 [0.06−1.02]	4.11 [1.00−16.95]	0.519 [0.493−0.545]	5.2 [4.1−6.6]

Values in parentheses are number of nodules; values in square brackets are 95% confidence intervals.
K-TIRADS, Korean Thyroid Imaging Reporting and Data System; PPV, positive predictive value; NPV, negative predictive value; LR+, likelihood ratio for positive results; LR−, likelihood ratio for negative results; DOR, diagnostic odds ratio; AUC, area under the receiver operating characteristic curve; AACE/ACE/AME, American Association of Clinical Endocrinologists/American College of Endocrinology/Associazione Medici Endocrinologi Medical Guidelines; ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System; ATA, American Thyroid Association Management Guideline; EU-TIRADS, European Thyroid Imaging Reporting and Data System.
[a]The reduction rate of unnecessary biopsies indicates the proportion of nodules not indicated for biopsy among all benign nodules.

Fig. 6. Modified K-TIRADS 3 nodule with partially cystic US pattern in a 44-year-old woman. Transverse US shows a predominantly solid isoechoic nodule (40 mm) without any suspicious US features in the right thyroid lobe. This nodule is classified as intermediate-risk by the AACE/ACE/AME guideline, not suspicious (TR2) by the ACR TIRADS, very low suspicion by the ATA guideline, intermediate-risk (TIRADS 4) by the EU-TIRADS, and low suspicion (TIRADS 4) by the K-TIRADS. Final diagnosis: minimally invasive follicular thyroid carcinoma by surgery. K-TIRADS, Korean Thyroid Imaging Reporting and Data System; US, ultrasonography; AACE, American Association of Clinical Endocrinologists; ACE, American College of Endocrinology; AME, Associazione Medici Endocrinologi; ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System; ATA, American Thyroid Association; EU-TIRADS, European Thyroid Imaging Reporting and Data System.

TI-RADS had a relatively low sensitivity and a high reduction rate of unnecessary biopsies in large nodules (>2 cm). The relatively low sensitivity of the ACR TI-RADS in large nodules was mainly due to "not suspicious" (TR2) nodules, which are not indicated for biopsy. The low sensitivity of the ATA guideline for malignancy in large thyroid nodules was caused by the unclassified nodules, and the diagnostic performance of the ATA guideline was similar to that of the K-TIRADS when the unclassified nodules were categorized as intermediate-suspicion nodules, as verified in a recent study [27]. The differences in diagnostic performance among RSSs in small nodules are mostly caused by differences in the size thresholds for biopsy and nodules not indicated for biopsy, rather than by differences in the structure (pattern-based versus point-based systems) or US criteria for nodule classification. The diagnostic performances of the RSSs were similar at the same size threshold for biopsy in simulation studies [13,27] and the diagnostic performance estimated by the classified categories was comparable among the RSSs [28].

Disagreements may exist regarding the most appropriate measure of test accuracy for evaluating the performance of an RSS as a triage

test. Although the DOR and AUC are effective measures of global diagnostic accuracy, two tests with an identical DOR and AUC can have very different sensitivities and specificities, with distinct clinical consequences [24]. Therefore, the DOR or AUC does not seem to be an appropriate primary measure for evaluating the performance of an RSS as a triage test. The most desirable RSS should be able to reduce unnecessary biopsies as much as possible, while maintaining an appropriate sensitivity for malignancy. Several points need to be considered regarding this issue. First, the diagnostic performance of the RSS needs to be stratified according to nodule size. The strategy of a higher reduction rate of unnecessary biopsies despite a lower sensitivity of the biopsy criteria may be appropriate for small nodules (1–2 cm), considering the favorable prognosis of most small thyroid cancers. Meanwhile, the strategy of a higher sensitivity despite a lower reduction rate of unnecessary biopsies may be appropriate in large nodules (>2 cm), considering the higher risk of aggressive behavior in large malignant tumors [17]. Second, the appropriate sensitivity of the biopsy criteria for malignancy should be determined based on a careful consideration of the risks and benefits to the patients. The hazard of false-negative results poses a potential risk of increased morbidity and mortality due to missing malignant tumors, which may be mitigated by US surveillance in small thyroid cancers. The hazard of false-positive results is a risk of potential complications and increased cost due to the increased number of biopsies. However, it should be considered that US-guided FNA is a very safe procedure and the cost-effectiveness of biopsy versus US surveillance may be controversial [29].

Although the strategy of using strict biopsy criteria and US monitoring of nodules that do not meet the biopsy criteria has been adopted for small thyroid nodules, it is still uncertain whether US monitoring of nodule growth can effectively prevent the potential risk of nodal or distant metastases because small PTCs may show macroscopic nodal metastases and small FTCs rarely show distant metastases. It should also be considered that there was no enlargement of the primary tumor in 11 of 12 low-risk papillary microcarcinomas (92%) that showed novel lymph node metastasis during active surveillance [30].

Our study has several limitations. First, our study included only nodules for which US-guided biopsy had been performed, which may inevitably induce selection bias and underestimate the actual reduction rate of unnecessary biopsies. Second, the reference standards for benign and malignant diagnoses were based on the biopsy results and surgical histologic findings, meaning that rare false-negative or false-positive results may have been present. The estimated malignancy risk of nodules might have been underestimated because many nodules were finally diagnosed based on one benign FNA or CNB result. Third, our cohort database

was generated at a single tertiary hospital. Further investigation in prospective multicenter studies will be necessary to validate the results of our study.

In conclusion, the modified K-TIRADS enables a high reduction rate of unnecessary biopsies, while maintaining a relatively high sensitivity and diagnostic accuracy for small malignant tumors compared to the K-TIRADS and other RSSs. Although the ACR TI-RADS has the strength of reducing unnecessary biopsies, it has a limitation of low sensitivity (less than 90%) for large malignant tumors, in contrast to the very high sensitivities of other RSSs. Further investigation and efforts should be made to reach a consensus on the appropriate sensitivity of the RSS for malignancy according to the nodule size.

ORCID: Dong Gyu Na: https://orcid.org/0000-0001-6422-1652; Wooyul Paik: https://orcid.org/0000-0001-9617-6227; Jaehyung Cha: https://orcid.org/0000-0003-3072-4846; Hye Yun Gwon: https://orcid.org/0000-0002-6899-0309; Suh Young Kim: https://orcid.org/0000-0002-5101-0167; Roh-Eul Yoo: https://orcid.org/0000-0002-5625-5921

Author Contributions
Conceptualization: Na DG. Data acquisition: Na DG, Paik W. Data analysis or interpretation: Na DG, Cha J, Kim SY, Gwon HY. Drafting of the manuscript: Na DG, Cha J. Critical revision of the manuscript: Na DG, Cha J, Yoo RE. Approval of the final version of the manuscript: all authors.

Conflict of Interest
No potential conflict of interest relevant to this article was reported.

Acknowledgments

Supplementary Material
Supplementary Table 1. Definitions of major ultrasound descriptors used in the major risk stratification systems and this study (https://doi.org/10.14366/usg.20148).

Supplementary Table 2. Biopsy size thresholds and malignancy risk according to categories in the five risk stratification systems (https://doi.org/10.14366/usg.20148).

References

1. Ha EJ, Lim HK, Yoon JH, Baek JH, Do KH, Choi M, et al. Primary imaging test and appropriate biopsy methods for thyroid nodules: guidelines by Korean Society of Radiology and National Evidence-Based Healthcare Collaborating Agency. Korean J Radiol 2018;19:623-631.

2. Perros P, Boelaert K, Colley S, Evans C, Evans RM, Gerrard Ba G, et al. Guidelines for the management of thyroid cancer. Clin Endocrinol (Oxf) 2014;81 Suppl 1:1-122.

3. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid 2016;26:1-133.

4. Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedus L, et al. American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules--2016 update. Endocr Pract 2016;22:622-639.

5. Shin JH, Baek JH, Chung J, Ha EJ, Kim JH, Lee YH, et al. Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean Society of Thyroid Radiology consensus statement and recommendations. Korean J Radiol 2016;17:370-395.

6. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): white paper of the ACR TI-RADS Committee. J Am Coll Radiol 2017;14:587-595.

7. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European Thyroid Association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: the EU-TIRADS. Eur Thyroid J 2017;6:225-237.

8. Hayen A, Macaskill P, Irwig L, Bossuyt P. Appropriate statistical methods are required to assess diagnostic tests for replacement, add-on, and triage. J Clin Epidemiol 2010;63:883-891.

9. Ha EJ, Na DG, Baek JH, Sung JY, Kim JH, Kang SY. US Fine-needle aspiration biopsy for thyroid malignancy: diagnostic performance of seven society guidelines applied to 2000 thyroid ndules. Radiology 2018;287:893-900.

10. Ha EJ, Na DG, Moon WJ, Lee YH, Choi N. Diagnostic performance of ultrasound-based risk-stratification systems for thyroid nodules: comparison of the 2015 American Thyroid Association guidelines with the 2016 Korean Thyroid Association/Korean Society of Thyroid Radiology and 2017 American College of Radiology guidelines. Thyroid 2018;28:1532-1537.

11. Middleton WD, Teefey SA, Reading CC, Langer JE, Beland MD,

Szabunio MM, et al. Comparison of performance characteristics of American College of Radiology TI-RADS, Korean Society of Thyroid Radiology TIRADS, and American Thyroid Association guidelines. AJR Am J Roentgenol 2018;210:1148-1154.

12. Grani G, Lamartina L, Ascoli V, Bosco D, Biffoni M, Giacomelli L, et al. Reducing the number of unnecessary thyroid biopsies while improving diagnostic accuracy: toward the "Right" TIRADS. J Clin Endocrinol Metab 2019;104:95-102.

13. Ha SM, Baek JH, Na DG, Suh CH, Chung SR, Choi YJ, et al. Diagnostic performance of practice guidelines for thyroid nodules: thyroid nodule size versus biopsy rates. Radiology 2019;291:92-99.

14. Castellana M, Castellana C, Treglia G, Giorgino F, Giovanella L, Russ G, et al. Performance of five ultrasound risk stratification systems in selecting thyroid nodules for FNA. J Clin Endocrinol Metab 2020;105:dgz170.

15. Hay ID. Papillary thyroid carcinoma. Endocrinol Metab Clin North Am 1990;19:545-576.

16. Mazzaferri EL, Jhiang SM. Long-term impact of initial surgical and medical therapy on papillary and follicular thyroid cancer. Am J Med 1994;97:418-428.

17. Machens A, Holzhausen HJ, Dralle H. The prognostic value of primary tumor size in papillary and follicular thyroid carcinoma. Cancer 2005;103:2269-2273.

18. Nguyen XV, Roy Choudhury K, Tessler FN, Hoang JK. Effect of tumor size on risk of metastatic disease and survival for thyroid cancer: implications for biopsy guidelines. Thyroid 2018;28:295-300.

19. Moon WJ, Baek JH, Jung SL, Kim DW, Kim EK, Kim JY, et al. Ultrasonography and the ultrasound-based management of thyroid nodules: consensus statement and recommendations. Korean J Radiol 2011;12:1-14.

20. Paik W, Na DG, Gwon HY, Kim J. CT features of thyroid nodules with isolated macrocalcifications detected by ultrasonography. Ultrasonography 2020;39:130-136.

21. Lee JY, Na DG, Yoon SJ, Gwon HY, Paik W, Kim T, et al. Ultrasound malignancy risk stratification of thyroid nodules based on the degree of hypoechogenicity and echotexture. Eur Radiol 2020;30:1653-1663.

22. Middleton WD, Teefey SA, Reading CC, Langer JE, Beland MD, Szabunio MM, et al. Multiinstitutional analysis of thyroid nodule risk stratification using the American College of Radiology Thyroid Imaging Reporting and Data System. AJR Am J Roentgenol 2017;208:1331-1341.

23. Boyko EJ. Ruling out or ruling in disease with the most sensitive or specific diagnostic test: short cut or wrong turn? Med Decis Making 1994;14:175-179.

24. Simundic AM. Measures of diagnostic accuracy: basic definitions. EJIFCC 2009;19:203-211.

25. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 2003;56:1129-1135.

26. Gu W, Pepe MS. Estimating the capacity for improvement in risk prediction with a marker. Biostatistics 2009;10:172-186.

27. Yim Y, Na DG, Ha EJ, Baek JH, Sung JY, Kim JH, et al. Concordance of three international guidelines for thyroid nodules classified by ultrasonography and diagnostic performance of biopsy criteria. Korean J Radiol 2020;21:108-116.

28. Kim PH, Suh CH, Baek JH, Chung SR, Choi YJ, Lee JH. Diagnostic performance of four ultrasound risk stratification systems: a systematic review and meta-analysis. Thyroid 2020;30:1159-1168.

29. Alexander AA. US-based risk stratification "guidelines" for thyroid nodules: Quo Vadis? J Clin Ultrasound 2020;48:127-133.

30. Ito Y, Miyauchi A, Oda H. Low-risk papillary microcarcinoma of the thyroid: a review of active surveillance trials. Eur J Surg Oncol 2018;44:307-315.