ORIGINAL ARTICLE

Validation and comparison of three newly-released Thyroid Imaging Reporting and Data Systems for cancer risk determination

Ting Xu[1] · Ya Wu[2] · Run-Xin Wu[3] · Yu-Zhi Zhang[4] · Jing-Yu Gu[2] · Xin-Hua Ye[5] · Wei Tang[1] · Shu-Hang Xu[6] · Chao Liu[6] · Xiao-Hong Wu[2]

Abstract

Purpose To validate and compare diagnostic value of three newly-released Thyroid Imaging Reporting and Data Systems (TIRADS) for cancer risk determination.

Methods Total 2031 patients with 2465 thyroid nodules were recruited for this study. Ultrasound (US) images were categorized based on three TIRADS editions established by Korean Society of Thyroid Radiology (KSThR), European Thyroid Association (ETA) and American College of Radiology (ACR). ROC curves were established to compare diagnostic value.

Results Total 1460 benign and 1005 malignant nodules were enrolled. The malignancy rates of each category in KSThR-TIRADS were 2.8%, 5.1%, 33.7% and 79.6%, respectively. For European-TIRADS, 0, 3.1, 22.8, and 73.5% of nodules categorized as 2 to 5 were malignant. Distribution of carcinomas among ACR-TIRADS categories was 0%, 2.3%, 7.5%, 40.1% and 81.4%, respectively. In terms of diagnostic value, KSThR-TIRADS had highest AUC (0.855) and specificity (87.4%), while lowest (71.4%) sensitivity. ACR-TIRADS showed best sensitivity (96.6%) with lowest specificity (52.9%) and the AUC (0.846) was slightly lower than KSThR-TIRADS. Total 56.1, 45.4, and 37.4% fine-needle aspiration biopsy (FNAB) were recommended by KSThR, ETA and ACR, revealing 42.8%, 44.5% and 53.6% malignant lesions, respectively. The rate of unnecessary FNAB was lowest with the ACR (17.3%), followed by ETA (25.2%) and KSThR (32.1%).

Conclusion All these US models showed great value in predicting thyroid malignancy. Among them, KSThR-TIRADS showed the most effective diagnostic performance in specificity, while ACR-TIRADS yielded best sensitivity. As for FNAB criteria, ACR-TIRADS showed the lowest rate of unnecessary FNAB and highest rate of malignancy in FNAB.

Keywords Thyroid nodule · Malignancy · Ultrasound · Diagnostic value

Introduction

Ultrasonography (US) has been used as an essential diagnostic tool in assessing thyroid nodule characteristics and the selection of nodules for fine-needle aspiration biopsy (FNAB) [1]. Specific US characteristics have been proposed to be associated with thyroid malignancy such as micro-calcifications, hypoechogenecity, irregular margins, taller-than-wide shape, and intranodular vascularity. However, none of them allows to reliably distinguish malignancy from benign nodules [2]. Thus, efforts have been made to develop US-based malignancy risk stratification systems for thyroid nodule. In 2009, Horvath was the first expert to propose Thyroid Imaging Reporting and Data System

---

These authors contributed equally: Ting Xu, Ya Wu, Run-Xin Wu.

✉ Xiao-Hong Wu
drxhwu@njmu.edu.cn

[1] Department of Endocrinology, Jiangsu Province Official Hospital, Nanjing, China

[2] Department of Endocrinology, the First Affiliated Hospital with Nanjing Medical University, Nanjing, China

[3] Nanjing Foreign Languages School, Nanjing, China

[4] Department of Ultrasound, Jiangsu Province Hospital on Integration of Chinese and Western Medicine, Nanjing, China

[5] Department of Ultrasound, the First Affiliated Hospital with Nanjing Medical University, Nanjing, China

[6] Department of Endocrinology, Jiangsu Province Hospital on Integration of Chinese and Western Medicine, Nanjing, China

Fig. 1 Diagram of the study group
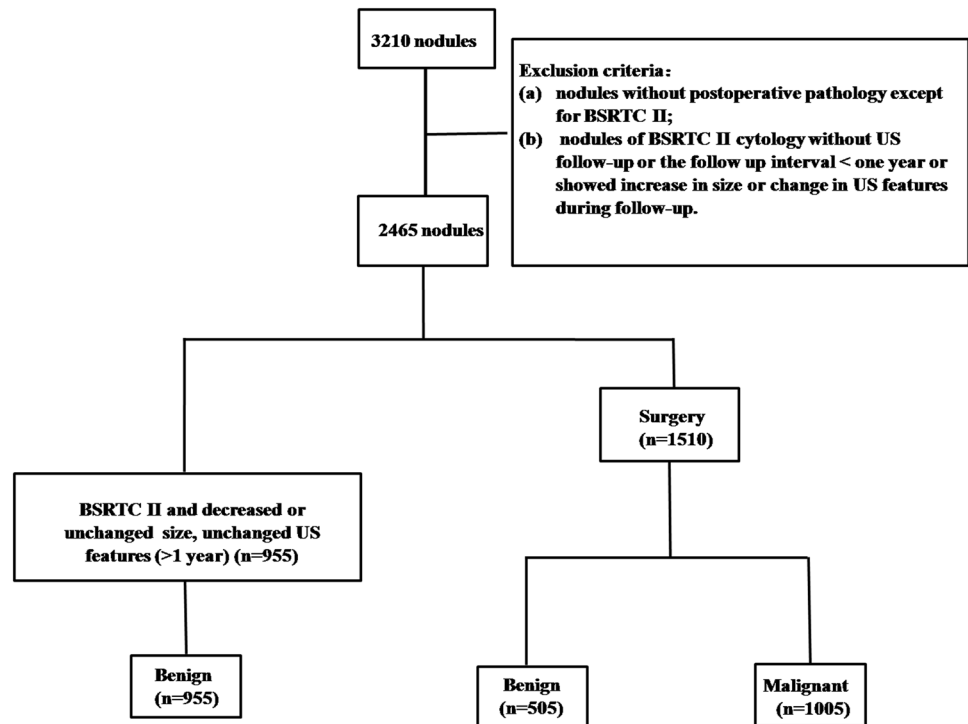
(TIRADS) based on ten US patterns, taking Breast Imaging Reporting and Data System (BIRADS) as a model [3]. However, these sonographic patterns are not applicable to all types of thyroid nodules [4]. Recently, Korean Society of Thyroid Radiology (KSThR), American College of Radiology (ACR), and European Thyroid Association (ETA) have published three editions of TIRADS in succession [5–7]. In 2016, KSThR has revised the recommendations for the US diagnosis and imaging-based management of thyroid nodules on the basis of the original edition published in 2011, stressing the risk stratifications based on solidity and echogenecity of thyroid nodules [5]. Then in 2017, ACR developed ACR-TIRADS by the sum of points of every characteristic feature and the total point determines the nodule's ACR TI-RADS level, which ranges from TR1 (benign) to TR5 (high suspicion of malignancy) [6]. In the same year, ETA created the novel European Thyroid Imaging and Reporting Data System (EU-TIRADS) consisting of five categories based on different patterns and US features [7]. However, as the representatives of thyroid association worldwide, the diagnostic value of the three newly released international US risk stratification systems has not been well validated. The present study aimed to evaluate the three newly-published TIRADS from KSThR, ETA, and ACR based on our large sample database and compare their diagnostic efficiency for better application in clinical practice.

Materials and methods

Subjects

From January 2014 to October 2017, a consecutive of 3210 lesions that underwent thyroid US examination and FNA and/or surgery from three tertiary hospitals around JiangSu Province were enrolled in this study. Patients who met all the following criteria by reviewing US patterns and clinical data were included in this study: (a) nodules with definite histopathology results, (b) nodules with complete Bethesda system for reporting thyroid cytopathology (BSRTC) results. The exclusion criteria were that: (a) nodules without postoperative pathology except for BSRTC II cytology, (b) nodules of BSRTC II cytology whose US follow-up interval less than one year or during which increase in size or change in US features (Fig. 1). The increase in size was defined as more than a 50% change in volume or a 20% increase in at least two nodule dimensions with a minimal increase of 2 mm in solid nodules or in the solid portion of mixed cystic-solid nodule [8]. Finally, 2031 patients with 2465 thyroid nodules were enrolled in this study, which included 415 male and 1616 female patients. The mean age of the patients was 47.70 ± 13.38 years. The study was performed in accordance with the ethical guidelines of the Helsinki Declaration and approved by the local ethics review committee (2012-SR-058).
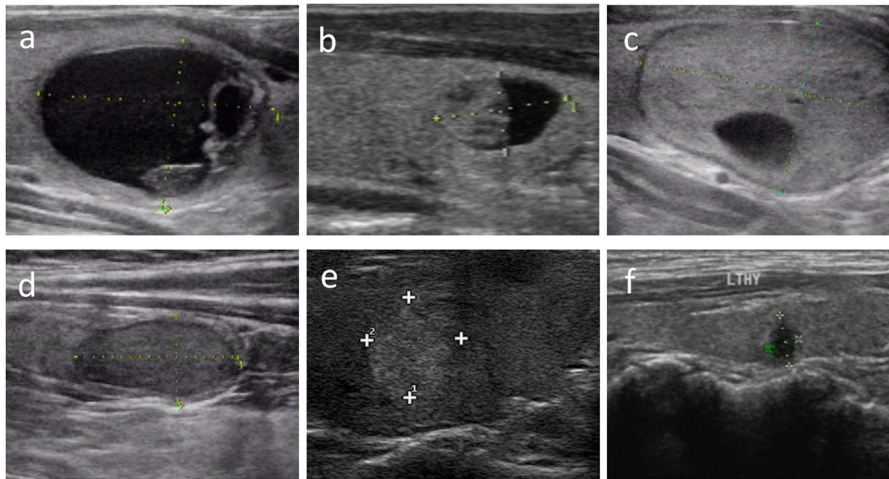
Fig. 2 Patterns categorized based on ACR-TIRADS, EU-TIRADS and KSThR-TIRADS: **a** TR1/EU-TIRADS 2/KSThR-TIRADS 2: almost completely cystic (nodular goiter with cystic degeneration). **b** TR2/EU-TIRADS 3/KSThR-TIRADS 3: mixed cystic and solid isoechoic without any suspicious feature (nodular goiter). **c** TR3/EU-TIRADS 3/KSThR-TIRADS 3: almost completely solid isoechoic without any suspicious feature (follicular adenoma). **d** TR4/EU-TIRADS 4/KSThR-TIRADS 4: solid hypoechoic without any suspicious feature (nodular goiter). **e** TR4/EU-TIRADS 5/KSThR-TIRADS 4: solid hyperechoic with taller-than-wide shape (papillary thyroid cancer). **f** TR5/EU-TIRADS 5/KSThR-TIRADS 5: solid hypoechoic with taller-than-wide shape (papillary thyroid cancer)

US examination technique

All US images were obtained by using a 4–13 MHz linear array transducer. The scanning protocol in all cases included both transverse and longitudinal real-time imaging of the thyroid nodules. Designated radiologists from three centers were asked to assess the thyroid nodules using one set of standards according to published literature [9]. The features used in analysis included size, composition (cystic/mixed cystic and solid/solid), echogenicity of solid portion (anechoic/isoechoic/hyperechoic/hypoechoic), echotexture (homogeneous/heterogeneous), vascularity (TypeI/II/III), shape (wider than tall/taller than wide), margin (well-defined/Ill-defined/irregular) and calcification (absent/comet-tail artifacts/macro/eggshell/micro). One specialist from each center extracted US features based on static US patterns and description of features and then input these features into database. Finally one experienced radiologist in thyroid imaging did all classifications according to the database. Clinical information and pathology results were blinded to the radiologist.

KSThR-TIRADS classification

All nodules were scored based on patterns and US features of KSThR-TIRADS as followed (Fig. 2) [5]. Category 2: Spongiform; Partially cystic nodule with comet tail artifact; Pure cyst. Category 3: Partially cystic or iso/hyperechoic nodule without any of the three suspicious US features (Microcalcification, nonparallel orientation, spiculated/microlobulated margin). Category 4: Solid hypoechoic nodule without any of the three suspicious US features; Partially cystic or isohyperechoic nodule with any of the three suspicious US features. Category 5: Solid hypoechoic nodule with any of the three suspicious US features.

ACR -TIRADS classification

All nodules were scored based on ACR-TIRADS as follows (Fig. 2) [6]: Composition (0 points: Cystic or almost completely cystic/Spongiform; 1 point: Mixed cystic and solid; 2 points: Solid or almost completely solid); Echogenecity (0 points: Anechoic; 1 point: Hyperechoic or isoechoic; 2 points: Hypoechoic; 3 points: Very hypoechoic); Shape (0 points: Wider-than-tall; 3 points: Taller-than-wide); Margin (0 points: Smooth/ Ill-defined; 2 points: Lobulated or irregular; 3 points: Extra-thyroidal extension); Echogenic Foci (0 points: None or large comet-tail artifacts; 1 point: Macrocalcification; 2 points: Peripheral calcifications; 3 points: Punctate echogenic Foci). The total points were calculated to determine TI-RADS level. TR1 (0 points), TR2 (2 points), TR3 (3 points), TR4 (4 to 6 points), TR5 (7 points or more).

EU-TIRADS classification

All nodules were scored based on patterns and US features of EU-TIRADS as follows (Fig. 2) [7]. EU-TIRADS 2: Pure cyst; Entirely spongiform. EU-TIRADS 3: Ovoid, smooth isoechoic/hyperechoic and no features of high suspicion. EU-TIRADS 4: Ovoid, smooth, mildly hypoechoic and no features of high suspicion. EU-TIRADS 5: At least

one of the following features of high suspicion including irregular shape, irregular margins, microcalcifications, marked hypoechogenicity (and solid).

Statistical analysis

Statistical analysis was performed using SPSS 20.0 software (SPSS Inc., Chicago, USA). All quantitative values were expressed as means ± SD. Differences in the value of continuous variables were evaluated by non-parametric test. Differences in the distribution of categorical variables between groups were evaluated by the two-tailed Chi-square (χ^2) test or Fisher exact test. According to the final diagnosis (pathology or follow-up results), the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated for each method. Receiver Operating Characteristic (ROC) curve analysis and area under curves (AUC) with MedCalc 11.4.2.0 software (MedCalc Software, Ostend, Belgium) were used to compare the diagnostic value of the three models and to determine the optimal cut-off value between benign and malignant nodules. Paired Chi-square test was used to compare the sensitivity and specificity of each two models. AUC and P value were calculated. $P < 0.05$ was considered significant in all tests.

Results

Clinical and US profile

A total of 2465 thyroid nodules in 2031 patients were included in our study. The mean diameter of the nodules was 16.63 ± 11.78 mm. Among the nodules, 505 benign nodules and all 1005 malignant lesions were confirmed by histopathology. The remaining 955 benign lesions were diagnosed based on the benign cytology and follow-up ultrasound. The epidemiological, clinical data of studied cases between malignant and benign groups were shown in Table 1. The average age in benign group was older than that in malignant group ($P < 0.05$). Besides, malignant lesions tended to be single in nodularity and smaller in size ($P < 0.05$ for both). US features of benign and malignant nodules are illustrated in Table 2. As determined by χ^2 tests, the difference between echogenecity, composition, echotexture, margin, shape, vascularity, calcification was statistically significant ($P < 0.05$ for all).

Correlations between TIRADS classification and final diagnosis

All 2465 nodules could be sorted based on KSThR-TIRADS, EU-TIRADS, and ACR-TIRADS classification. As

Table 1 Clinical features of the study population and basic characters of the nodules

Parameter	Total	Malignant (%)	Benign (%)	P value
Sex				
Male	415	200 (48.2)	215 (51.8)	0.037
Female	1616	685 (42.4)	931 (57.6)	
Nodularity				
Single	854	420 (49.2)	434 (50.8)	0.000
Multiple	1177	465 (39.5)	712 (60.5)	
Age (year)[a]	47.68 ± 13.39	43.65 ± 12.53	50.82 ± 13.20	0.000
Size (mm)[a]	16.63 ± 11.78	13.64 ± 9.03	18.68 ± 12.96	0.000

[a]Data are means ± standard deviations

Table 2 Ultrasound features of the nodules according to final diagnosis

Features	Total	Malignant (%)	Benign (%)	P value
Echogenicity				
Anechoic	62	0 (0)	62 (100.0)	0.000
Isoechoic	593	37 (6.2)	556 (93.8)	
Hyperechoic	119	19 (16.0)	100 (84.0)	
Hypoechoic	1691	949 (56.1)	742 (43.9)	
Composition				0.000
Cystic				
Mixed cystic and solid	794	143 (18.0)	651 (82.0)	
Solid	1609	862 (53.6)	747 (46.4)	
Echotexture				
Homogeneous	547	201 (36.7)	346 (63.3)	0.030
Heterogeneous	1918	804 (41.9)	1114 (58.1)	
Vascularity				
Type I/II	1754	613 (34.9)	1141 (65.1)	0.000
Type III	711	392 (55.1)	319 (44.9)	
Margin				0.000
Well-defined				
Ill-defined	339	207 (61.1)	132 (38.9)	
Irregular	672	535 (79.6)	137 (20.4)	
Calcification				0.000
Absent				
Comet-tail artifacts				
Macrocalcification	111	35 (31.5)	76 (68.5)	
Eggshell calcification	16	5 (31.2)	11 (68.8)	
Microcalcification	746	584 (78.3)	162 (21.7)	
Shape				
Wider than tall	2152	750 (34.9)	1402 (65.1)	0.000
Taller than wide	313	255 (81.5)	58 (18.5)	

for KSThR-TIRADS based on histopathology or follow-up results, the malignancy rates of nodules categorized as 2 to 5 were found to be 2.8% (four of 141 nodules), 5.1% (35 of 687 nodules), 33.7% (248 of 735 nodules) and 79.6% (718

Table 3 The malignancy rates of KSThR-TIRADS, ACR-TIRADS, and EU-TIRADS according to final diagnosis

	Total	Malignant (%)	Benign (%)	P value	Recommended malignancy risk
KSThR-TIRADS				0.000	
2 (Benign)	141	4 (2.8)	137 (97.2)		<3
3 (Low suspicion)	687	35 (5.1)	652 (94.9)		3–15
4 (Intermediate suspicion)	735	248 (33.7)	487 (66.3)		15–50
5 (High suspicion)	902	718 (79.6)	184 (20.4)		>60
Total	2465	1005 (40.8)	1460 (59.2)		
ACR-TIRADS				0.000	
TR1	62	0	62 (100.0)		≤2
TR2	426	10 (2.3)	416 (97.7)		≤2
TR3	319	24 (7.5)	295 (92.5)		≤5
TR4	917	368 (40.1)	549 (59.9)		5–20
TR5	741	603 (81.4)	138 (18.6)		≥20
Total	2465	1005 (40.8)	1460 (59.2)		
EU-TIRADS				0.000	
EU-TIRADS 2	62	0	62 (100.0)		0
EU-TIRADS 3	607	19 (3.1)	588 (96.9)		2–4
EU-TIRADS 4	659	150 (22.8)	509 (77.2)		6–17
EU-TIRADS 5	1137	836 (73.5)	301 (26.5)		26–87
Total	2465	1005 (40.8)	1460 (59.2)		

of 902 nodules), respectively, with significant differences among categories ($P < 0.001$) (Table 3).

For EU-TIRADS, the risk of malignancy significantly rose as the TIRADS category increased. The risk of malignancy was found to be 0 (0 of 62 nodules) in category 2, 3.1% (19 of 607 nodules) in category 3, 22.8% (150 of 659 nodules) in category 4 and 73.5% (836 of 1137 nodules) in category 5, respectively, with significant differences among categories ($P < 0.001$) (Table 3).

According to histopathology or follow-up results, distribution of carcinomas among ACR-TIRADS categories was 0% (0 of 62 nodules), 2.3% (10 of 426 nodules), 7.5% (24 of 319 nodules), 40.1% (368 of 917 nodules), and 81.4% (603 of 741 nodules), respectively (Table 3).

Comparison of the three TIRADS editions in diagnostic value

The ROC curves of three TIRADS editions were listed in Fig. 3. As shown in Table 4, among the three editions of TIRADS, KSThR-TIRADS had the highest AUC (0.855) and specificity (87.4%), while lowest (71.4%) sensitivity. ACR-TIRADS showed best sensitivity (96.6%) with lowest specificity (52.9%) and the AUC (0.846) was slightly lower than that of KSThR-TIRADS. Compared to ACR-TIRADS,
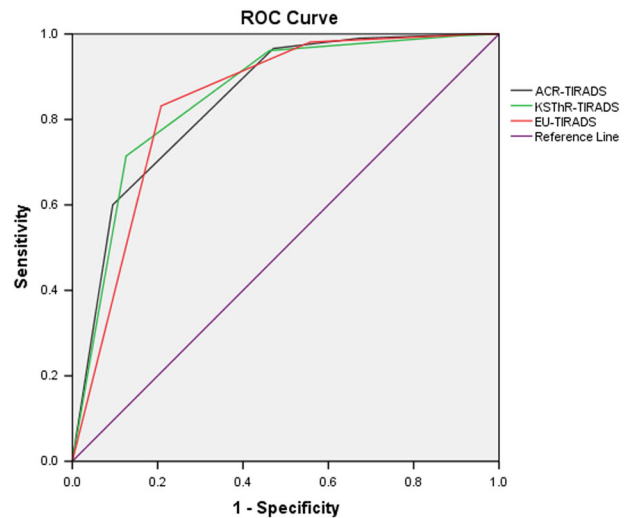

Fig. 3 ROC curves of KSThR-TIRADS, EU-TIRADS, and ACR-TIRADS based on final diagnosis

EU-TIRADS had relatively lower AUC (0.843), while the specificity was significantly higher (79.4%). The statistical differences of AUC, sensitivity and specificity between every two US risk stratification models were shown in Table 5. Difference between every two models in sensitivity and specificity was significant ($P < 0.05$). In terms of AUC, the differences between KSThR-TIRADS and EU-TIRADS were statistically significant.

Table 4 The comparison of three editions of TIRADS in diagnostic value

	Cut-off	AUC	Sensitivity	Specificity	PPV	NPV
KSThR-TIRADS	High suspicion	0.855 (0.840–0.870)	71.4 (68.5–74.2)	87.4 (85.6–89.1)	79.6 (76.8–82.2)	81.6 (79.6–83.5)
EU-TIRADS	5	0.843 (0.828–0.859)	83.2 (80.7–85.4)	79.4 (77.2–81.4)	73.5 (70.9–76.1)	87.3 (85.4–89.0)
ACR-TIRADS	TR4	0.846 (0.831–0.860)	96.6 (95.3–97.6)	52.9 (50.3–55.5)	58.6 (56.1–60.9)	95.8 (94.2–97.1)

AUC area under curve, *PPV* positive predictive value, *NPV* negative predictive value

Table 5 Statistical results of three TIRADS editions in diagnostic value between groups

	Z value	P value		
		AUC	Sensitivity	Specificity
ACR-TIRADS vs KSThR-TIRADS	1.581	0.114	<0.001	<0.001
ACR-TIRADS vs EU-TIRADS	0.522	0.601	<0.001	<0.001
KSThR-TIRADS vs EU-TIRADS	2.177	0.030	<0.001	<0.001

AUC area under curve

The comparison of three TIRADS editions in FNAB criteria

Among 2465 thyroid nodules, total 1383 (56.1%), 1120 (45.4%) and 921 (37.4%) FNABs were recommended according to FNAB criteria by KSThR, ETA and ACR, revealing 592 (42.8%), 498 (44.5%) and 494 (53.6%) malignant lesions, respectively (ACR vs ETA, ACR vs KSThR, $P < 0.001$, ETA vs KSThR, $P > 0.05$) (Table 6). Whereas total 413 (38.2%), 507 (37.7%) and 511 (33.1%) thyroid cancers would be missed among Non-FNABs (ACR vs ETA, ACR vs KSThR, $P < 0.001$, ETA vs KSThR, $P > 0.05$). The rate of unnecessary FNAB was lowest with the ACR guidelines (17.3%), followed by EU-TIRADS (25.2%) and KSThR-TIRADS (32.1%) ($P < 0.001$ for all). The false-positive rate was lowest with the ACR guidelines (29.2%), followed by EU-TIRADS (42.6%) and KSThR-TIRADS (54.2%) ($P < 0.001$ for all). The false-negative rate was highest with the ACR guidelines (50.8%), followed by EU-TIRADS (50.4%) and KSThR-TIRADS (41.1%) (ACR vs KSThR, ETA vs KSThR, $P < 0.001$, ETA vs ACR, $P > 0.05$).

Discussion

The present validation study has revealed that all these newly-released editions of TIRADS from KSThR, ETA, and ACR have shown great value in predicting thyroid malignancy. Among them, KSThR-TIRADS performed remarkably best in differentiating malignancy form benignity, while ACR-TIRADS and EU-TIRADS showed their own advantages in diagnostic sensitivity or specificity more or less. As for FNAB criteria, ACR-TIRADS showed the lowest rate of unnecessary FNAB and highest rate of malignancy in FNAB.

In 2009, Horvath was the first one to develop TIRADS for cancer risk determination, just as BIRADS did for breast lesions [3]. Then, Park et al. [10]. proposed an equation for predicting the probability of malignancy based on 12 ultrasound features, however, the complexity restricted its clinical practice. In 2011, Kwak et al. [11]. published a TIRADS classification based on five suspicious ultrasound features. However, the same weighting neglected their different contributions to a malignant lesion. In 2016, American Thyroid Association (ATA) and subsequently the American Association of Clinical Endocrinologists (AACE), American College of Endocrinology (ACE), Associazione Medici Endocrinologi (AME) constructed new ultrasound risk stratification models according to sonographic patterns [12, 13]. Our previous study revealed that thyroid nodule sizes influenced the diagnostic performance of Kwak-TIRADS and ATA ultrasound patterns [14].

Recently, KSThR, ACR, and ETA have published three editions of TIRADS in succession to help determine the nature of nodules and facilitate the selection of nodules for FNA cytological analysis. As for KSThR-TIRADS, the benign category included the pattern of pure cystic, spongiform, and partially cystic only when accompanied with comet tail, which were confirmed benign in other studies [15, 16]. The malignancy risk of category 4 in our study reached 33.7% once the partially cystic or isohyperechoic patterns in category 3 were accompanied by suspicious US features, or solid hypoechoic nodules without suspicious features. In Yoon's study, the malignancy risk of hyper- to isoechoic solid or partially cystic nodules with suspicious features was 18.2% and that of solid hypoechoic nodule with smooth regular margin was 16.7%, both within the range of the recommended risk of category 4 [17]. Besides, 79.6% of nodules classified as category 5 were malignant in our study, indicating that the pattern of solid hypoechoic

Table 6 Diagnostic performance of three TIRADS editions in FNAB criteria

TIRADS	No. of FNABs	No. of malignancy among FNABs	No. of benignity among FNABs (A)	No. of missed malignancy among Non-FNABs (B)	Unnecessary FNAB rate (A/2465)	False-positive rate (A/1460)	False-negative rate (B/1005)
KSThR	1383	592 (42.8)	791 (57.2)	413 (38.2)	32.1	54.2	41.1
EU	1120	498 (44.5)	622 (55.5)	507 (37.7)	25.2	42.6	50.4
ACR	921	494 (53.6)	427 (46.4)	511 (33.1)	17.3	29.2	50.8

with any of the three suspicious features was highly suspicious for malignancy.

In 2017, ACR presented a white paper of ACR-TIRADS which ranges from TR1 (benign) to TR5 (high suspicion of malignancy) [6]. Unlike KSThR, ACR-TIRADS was based on total scores of five US features including composition, echogenecity, shape, margin, and echogenic foci. The malignancy rates in our study were 0, 2.3, 7.5, 40.1, and 81.4% from TR1 to TR5, which were similar to that in William's recent multi-institutional analysis [18]. In ACR-TIRADS, features such as macrocalcification, rim calcifications and mixed cystic and solid were regarded as somewhat risk of malignancy. A nodule showing the pattern of solid hyperechoic accompanied with macrocalcification would be categorized as moderately suspicious by ACR-TIRADS, while it was classified as low suspicion by KSThR-TIRADS or EU-TIRADS. That's why ACR-TIRADS showed highest sensitivity, yet lowest specificity among these three systems. However, the diagnostic AUC of ACR-TIRADS was 0.846, which was close to Ha's results [19]. The AUC was relatively lower compared to KSThR-TIRADS.

In the same year, the ETA created the novel EU-TIRADS based on a review of the literature and on the AACE, ATA, and Korean guidelines [7]. Both based on US patterns, while compared to KSThR-TIRADS, EU-TIRADS put less emphasis on solidity. For example, one nodule showing the pattern of a combination of predominantly solid, mildly hypoechoic without suspicious features was categorized as low suspicion by KSThR-TIRADS, while it was classified as intermediate risk by EU-TIRADS. In a retrospective multi-center study, the malignancy rate of such pattern was 17.2%, which was close to the recommended risk of low suspicion by KSThR-TIRADS rather than intermediate risk by EU-TIRADS [20]. While the nodule with suspicious US features combined with solid hypoechoic would be regarded as high suspicion based on KSThR-TIRADS, it would be categorized as high risk by EU-TIRADS disregarding its echogenecity or internal content. Thus, KSThR-TIRADS showed a relatively lower sensitivity. In terms of diagnostic value, compared to EU-TIRADS and ACR-TIRADS, KSThR-TIRADS yielded a higher AUC and specificity at a cost of decreased sensitivity, indicating that KSThR-TIRADS may be diagnostically superior to the other two models in clinical practice in our population.

TIRADS could aid in decision making about the use of FNAB. Compared with other US systems, ACR-TIRADS criteria offered the lowest rate of unnecessary FNAB [21], which could allow reduction in the percentage of benign nodules that are biopsied [22]. In consistent with these results, we found that the malignancy rate among recommended FNABs was highest and the missed malignancy rate among Non-FNABs was lowest with ACR-TIRADS compared to the other two systems. A recent study also found that ACR-TIRADS outperformed EU-TIRADS and KSThR-TIRADS in identifying nodules whose FNA can be safely deferred [23]. It could classify over half the biopsies as unnecessary and the malignancy rate among these non-FNABs was 2.2%, lower than that of its competitors. That was pretty lower than our study (33.1%), probably due to the difference in percentage of carcinomas (7.2% vs 40.8%).

Interobserver agreement and reproducibility were also indispensable for evaluating a diagnostic model. Recent study demonstrated that EU-TIRADS showed higher reproducibility than ACR-TIRADS and KSThR-TIRADS, possibly due to a lower number of high-suspicious features and more gradual scoring. However, KSThR-TIRADS had the highest reproducibility on FNAB because the majority of low-suspicion and intermediate-suspicion nodules are submitted to FNAB [24]. The interobserver variability of 2015 ATA classification was less reproducible compared with 2011 Korean-TIRADS, owing to higher overall complexity [25].

In addition to US, real-time elastography (RTE) and contrast-enhanced ultrasound (CEUS) had been applied to evaluate the tissue stiffness or blood perfusion of thyroid nodules. Both strain elastography and share wave elastography (SWE) were useful in malignancy evaluation, with the AUC ranged from 0.89 to 0.93 for strain ratio [26–29] and 0.91 to 0.94 for SWE [30–32]. The isthmic location, larger size, cystic component, inexperienced operator, follicular carcinoma pathology and inability to differentiate malignancy from thyroiditis influenced its efficiency [33–37]. As for CEUS, although no single qualitative or quantitative parameter showed sufficient advantage in the diagnosis of thyroid malignancy [38], it can record better nodule vasculature compared with color Doppler ultrasound. The pooled sensitivity, specificity of CEUS for differentiation of

malignant and benign nodules reached 0.9 and 0.86 [39]. Our previous study revealed that combined use of RTE, CEUS, and US could improve the diagnostic efficiency for solid thyroid nodules [40].

Several limitations of our study should also be addressed. Firstly, the extraction of ultrasound features was performed based on static images and reports of US operated by radiologists, which might possibly lead to unavoidable bias. Secondly, the percentage of carcinomas (40.8%) was high in the present study, which may be due to the tertiary referral hospitals enrolled in this study. Such high malignancy rate may cause relatively higher PPV [41]. Thirdly, some nodules were regarded as benign lesions based on cytology or US follow-up, which may cause false-negative results.

In summary, all the three newly-updated TIRADS showed good performance in predicting thyroid malignancy and enabled more personalized and optimized management decision for clinicians. This is the first study to evaluate the diagnostic value of US risk stratification models from KSThR, ETA and ACR and our findings still need to be further validated in a future prospective study and clinical practice.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This study has been approved by ethics committee at all participating centers and all procedures were in accordance with the ethical standards of the ethical committees and of the 1964 Declaration of Helsinki and its later amendments.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. H.J. Moon, E.K. Kim, J.Y. Kwak, Malignancy risk stratification in thyroid nodules with benign results on cytology: combination of thyroid imaging reporting and data system and Bethesda system. Ann. Surg. Oncol. **21**(6), 1898–1903 (2014)

2. K. Wolinski, M. Szkudlarek, E. Szczepanek-Parulska, M. Ruchala, Usefulness of different ultrasound features of malignancy in predicting the type of thyroid lesions: a meta-analysis of prospective studies. Pol. Arch. Med. Wewn. **124**(3), 97–104 (2014)

3. E. Horvath, S. Majlis, R. Rossi, C. Franco, J.P. Niedmann, A. Castro, M. Dominguez, An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. J. Clin. Endocrinol. Metab. **94**(5), 1748–1751 (2009)

4. S.Y. Xu, W.W. Zhan, W.H. Wang, Evaluation of thyroid nodules by a scoring and categorizing method based on sonographic features. J. Ultrasound Med. **34**(12), 2179–2185 (2015)

5. J.H. Shin, J.H. Baek, J. Chung, E.J. Ha, J.H. Kim, Y.H. Lee, H.K. Lim, W.J. Moon, D.G. Na, J.S. Park, Y.J. Choi, S.Y. Hahn, S.J. Jeon, S.L. Jung, D.W. Kim, E.K. Kim, J.Y. Kwak, C.Y. Lee, H.J. Lee, J.H. Lee, J.H. Lee, K.H. Lee, S.W. Park, J.Y. Sung, Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean Society of Thyroid Radiology Consensus Statement and Recommendations. Korean J. Radiol. **17**(3), 370–395 (2016)

6. F.N. Tessler, W.D. Middleton, E.G. Grant, J.K. Hoang, L.L. Berland, S.A. Teefey, J.J. Cronan, M.D. Beland, T.S. Desser, M.C. Frates, L.W. Hammers, U.M. Hamper, J.E. Langer, C.C. Reading, L.M. Scoutt, A.T. Stavros, ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. J. Am. Coll. Radiol. **14**(5), 587–595 (2017)

7. G. Russ, S.J. Bonnema, M.F. Erdogan, C. Durante, R. Ngu, L. Leenhardt, European Thyroid Association Guidelines for Ultrasound Malignancy Risk Stratification of Thyroid Nodules in Adults: the EU-TIRADS. Eur. Thyroid J. **6**(5), 225–237 (2017)

8. D.S. Cooper, G.M. Doherty, B.R. Haugen, R.T. Kloos, S.L. Lee, S.J. Mandel, E.L. Mazzaferri, B. McIver, F. Pacini, M. Schlumberger, S.I. Sherman, D.L. Steward, R.M. Tuttle, Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. Thyroid **19**(11), 1167–1214 (2009)

9. M.C. Frates, C.B. Benson, J.W. Charboneau, E.S. Cibas, O.H. Clark, B.G. Coleman, J.J. Cronan, P.M. Doubilet, D.B. Evans, J.R. Goellner, I.D. Hay, B.S. Hertzberg, C.M. Intenzo, R.B. Jeffrey, J.E. Langer, P.R. Larsen, S.J. Mandel, W.D. Middleton, C.C. Reading, S.I. Sherman, F.N. Tessler, Management of thyroid nodules detected at US: Society of Radiologists in Ultrasound consensus conference statement. Radiology **237**(3), 794–800 (2005)

10. J.Y. Park, H.J. Lee, H.W. Jang, H.K. Kim, J.H. Yi, W. Lee, S.H. Kim, A proposal for a thyroid imaging reporting and data system for ultrasound features of thyroid carcinoma. Thyroid **19**(11), 1257–1264 (2009)

11. J.Y. Kwak, K.H. Han, J.H. Yoon, H.J. Moon, E.J. Son, S.H. Park, H.K. Jung, J.S. Choi, B.M. Kim, E.K. Kim, Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk. Radiology **260**(3), 892–899 (2011)

12. B.R. Haugen, E.K. Alexander, K.C. Bible, G.M. Doherty, S.J. Mandel, Y.E. Nikiforov, F. Pacini, G.W. Randolph, A.M. Sawka, M. Schlumberger, K.G. Schuff, S.I. Sherman, J.A. Sosa, D.L. Steward, R.M. Tuttle, L. Wartofsky, 2015 AmericanThyroid 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 **26**(1), 1–133 (2016)

13. H. Gharib, E. Papini, J.R. Garber, D.S. Duick, R.M. Harrell, L. Hegedus, R. Paschke, R. Valcavi, P. Vitti, 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. **22**(5), 622–639 (2016)

14. T. Xu, J.Y. Gu, X.H. Ye, S.H. Xu, Y. Wu, X.Y. Shao, D.Z. Liu, W.P. Lu, F. Hua, B.M. Shi, J. Liang, L. Xu, W. Tang, C. Liu, X.H. Wu, Thyroid nodule sizes influence the diagnostic performance of TIRADS and ultrasound patterns of 2015 ATA guidelines: a multicenter retrospective study. Sci. Rep. **24**(7), 43183 (2017)

15. J.P. Brito, M.R. Gionfriddo, N.A. Al, K.R. Boehmer, A.L. Leppin, C. Reading, M. Callstrom, T.A. Elraiyah, L.J. Prokop, M.N. Stan, M.H. Murad, J.C. Morris, V.M. Montori, The accuracy of thyroid nodule ultrasound to predict thyroid cancer: systematic review and meta-analysis. J. Clin. Endocrinol. Metab. **99**(4), 1253–1263 (2014)

16. H. Malhi, M.D. Beland, S.Y. Cen, E. Allgood, K. Daley, S.E. Martin, J.J. Cronan, E.G. Grant, Echogenic foci in thyroid nodules: significance of posterior acoustic artifacts. AJR Am. J. Roentgenol. **203**(6), 1310–1316 (2014)

17. J.H. Yoon, H.S. Lee, E.K. Kim, H.J. Moon, J.Y. Kwak, Malignancy Risk Stratification of Thyroid Nodules: Comparison between the Thyroid Imaging Reporting and Data System and the 2014 American Thyroid Association Management Guidelines. Radiology **278**(3), 917–924 (2016)

18. W.D. Middleton, S.A. Teefey, C.C. Reading, J.E. Langer, M.D. Beland, M.M. Szabunio, T.S. Desser, Multiinstitutional Analysis of Thyroid Nodule Risk Stratification Using the American College of Radiology Thyroid Imaging Reporting and Data System. AJR Am. J. Roentgenol. **208**(6), 1331–1341 (2017)

19. S.M. Ha, H.S. Ahn, J.H. Baek, H.Y. Ahn, Y.J. Chung, B.Y. Cho, S.B. Park, Validation of three scoring risk-stratification models for thyroid nodules. Thyroid **27**(12), 1550–1557 (2017)

20. D.G. Na, J.H. Baek, J.Y. Sung, J.H. Kim, J.K. Kim, Y.J. Choi, H. Seo, Thyroid Imaging Reporting and Data System Risk Stratification of Thyroid Nodules: Categorization Based on Solidity and Echogenicity. Thyroid **26**(4), 562–572 (2016)

21. E.J. Ha, D.G. Na, J.H. Baek, J.Y. Sung, J.H. Kim, S.Y. Kang, US Fine-Needle Aspiration Biopsy for Thyroid Malignancy: Diagnostic Performance of Seven Society Guidelines Applied to 2000 Thyroid Nodules. Radiology **287**(3), 893–900 (2018)

22. W.D. Middleton, S.A. Teefey, C.C. Reading, J.E. Langer, M.D. Beland, M.M. Szabunio, T.S. Desser, 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. **210**(5), 1148–1154 (2018)

23. G. Grani, L. Lamartina, V. Ascoli, D. Bosco, M. Biffoni, L. Giacomelli, M. Maranghi, R. Falcone, V. Ramundo, V. Cantisani, S. Filetti, and C. Durante, Reducing the number of unnecessary thyroid biopsies while improving diagnostic accuracy: towards the "right" TIRADS. J. Clin. Endocrinol. Metab. (2018). [Epub ahead of print]

24. G. Grani, L. Lamartina, V. Cantisani, M. Maranghi, P. Lucia, C. Durante, Interobserver agreement of various thyroid imaging reporting and data systems. Endocr. Connect. **7**(1), 1–7 (2018)

25. G. Grani, L. Lamartina, V. Ascoli, D. Bosco, F. Nardi, F. D'Ambrosio, A. Rubini, L. Giacomelli, M. Biffoni, S. Filetti, C. Durante, V. Cantisani, Ultrasonography scoring systems can rule out malignancy in cytologically indeterminate thyroid nodules. Endocrine **57**(2), 256–261 (2017)

26. V. Cantisani, P. Maceroni, V. D'Andrea, G. Patrizi, M. Di Segni, C. De Vito, H. Grazhdani, A.M. Isidori, E. Giannetta, A. Redler, F. Frattaroli, L. Giacomelli, G. Di Rocco, C. Catalano, F. D'Ambrosio, Strain ratio ultrasound elastography increases the accuracy of colour-Doppler ultrasound in the evaluation of Thy-3 nodules. A bi-centre university experience. Eur. Radiol. **26**(5), 1441–1449 (2016)

27. J. Sun, J. Cai, X. Wang, Real-time ultrasound elastography for differentiation of benign and malignant thyroid nodules: a meta-analysis. J. Ultrasound Med. **33**(3), 495–502 (2014)

28. S.A. Razavi, T.A. Hadduck, G. Sadigh, B.A. Dwamena, Comparative effectiveness of elastographic and B-mode ultrasound criteria for diagnostic discrimination of thyroid nodules: a meta-analysis. AJR Am. J. Roentgenol. **200**(6), 1317–1326 (2013)

29. M. Ghajarzadeh, F. Sodagari, M. Shakiba, Diagnostic accuracy of sonoelastography in detecting malignant thyroid nodules: a systematic review and meta-analysis. Am. J. Roentgenol. **202**(4), W379–W389 (2014)

30. F.J. Dong, M. Li, Y. Jiao, J.F. Xu, Y. Xiong, L. Zhang, H. Luo, Z. M. Ding, Acoustic Radiation Force Impulse imaging for detecting thyroid nodules: a systematic review and pooled meta-analysis. Med. Ultrason **17**(2), 192–199 (2015)

31. P. Lin, M. Chen, B. Liu, S. Wang, X. Li, Diagnostic performance of shear wave elastography in the identification of malignant thyroid nodules: a meta-analysis. Eur. Radiol. **24**(11), 2729–2738 (2014)

32. J. Zhan, J.M. Jin, X.H. Diao, Y. Chen, Acoustic radiation force impulse imaging (ARFI) for differentiation of benign and malignant thyroid nodules--A meta-analysis. Eur. J. Radiol. **84**(11), 2181–2186 (2015)

33. D. Cosgrove, R. Barr, J. Bojunga, V. Cantisani, M.C. Chammas, M. Dighe, S. Vinayak, J.M. Xu, C.F. Dietrich, WFUMB Guidelines and Recommendations on the Clinical Use of Ultrasound Elastography: Part 4. Thyroid. Ultrasound Med. Biol. **43**(1), 4–26 (2017)

34. V. Cantisani, H. Grazhdani, P. Ricci, K. Mortele, M. Di Segni, V. D'Andrea, A. Redler, G. Di Rocco, L. Giacomelli, E. Maggini, C. Chiesa, S.M. Erturk, S. Sorrenti, C. Catalano, F. D'Ambrosio, Q-elastosonography of solid thyroid nodules: assessment of diagnostic efficacy and interobserver variability in a large patient cohort. Eur. Radiol. **24**(1), 143–150 (2014)

35. H. Grazhdani, V. Cantisani, P. Lodise, G. Di Rocco, M.C. Proietto, E. Fioravanti, A. Rubini, A. Redler, Prospective evaluation of acoustic radiation force impulse technology in the differentiation of thyroid nodules: accuracy and interobserver variability assessment. J. Ultrasound **17**(1), 13–20 (2014)

36. Y. Wang, H.J. Dan, H.Y. Dan, T. Li, B. Hu, Differential diagnosis of small single solid thyroid nodules using real-time ultrasound elastography. J. Int. Med. Res. **38**(2), 466–472 (2010)

37. C. Oliver, J. Vaillant-Lombard, F. Albarel, J. Berbis, J.B. Veyrieres, F. Sebag, P. Petit, What is the contribution of elastography to thyroid nodules evaluation? Ann. Endocrinol. **72**(2), 120–124 (2011)

38. P.S. Sidhu, V. Cantisani, C.F. Dietrich, O.H. Gilja, A. Saftoiu, E. Bartels, M. Bertolotto, F. Calliada, D.A. Clevert, D. Cosgrove, A. Deganello, M. D'Onofrio, F.M. Drudi, S. Freeman, C. Harvey, C. Jenssen, E.M. Jung, A.S. Klauser, N. Lassau, M.F. Meloni, E. Leen, C. Nicolau, C. Nolsoe, F. Piscaglia, F. Prada, H. Prosch, M. Radzina, L. Savelli, H.P. Weskott, H. Wijkstra, The EFSUMB Guidelines and Recommendations for the Clinical Practice of Contrast-Enhanced Ultrasound (CEUS) in Non-Hepatic Applications: Update 2017 (Long Version). Ultraschall Med. **39**(2), e2–e44 (2018)

39. X. Ma, B. Zhang, W. Ling, R. Liu, H. Jia, F. Zhu, M. Wang, H. Liu, J. Huang, L. Liu, Contrast-enhanced sonography for the identification of benign and malignant thyroid nodules: Systematic review and meta-analysis. J. Clin. Ultrasound **44**(4), 199–209 (2016)

40. Y.Z. Zhang, T. Xu, H.Y. Gong, C.Y. Li, X.H. Ye, H.J. Lin, M.P. Shen, Y. Duan, T. Yang, X.H. Wu, Application of high-resolution ultrasound, real-time elastography, and contrast-enhanced ultrasound in differentiating solid thyroid nodules. Medicine **95**(45), e5329 (2016)

41. R.L. Ferris, Z. Baloch, V. Bernet, A. Chen, T.R. Fahey, I. Ganly, S.P. Hodak, E. Kebebew, K.N. Patel, A. Shaha, D.L. Steward, R. P. Tufano, S.M. Wiseman, S.E. Carty, American Thyroid Association Statement on Surgical Application of Molecular Profiling for Thyroid Nodules: Current Impact on Perioperative Decision Making. Thyroid **25**(7), 760–768 (2015)