BPR MCS
Bulletin of Pioneering Researches of Medical and Clinical Science

Bulletin of Pioneering Researches of Medical and Clinical Science

Available online: https://bprmcs.com
2022 | Volume 1 | Issue 2 | Page: 72-82

Comparative Diagnostic Performance of Six TIRADS Classifications in the Evaluation of Cytologically Indeterminate Thyroid Nodule

Da Yee Jeung[1], Su-Min Hong[2*]

[1] Department of Radiological Science, Jangpalpal Internal Medicine Clinic, 369, Haeundae-ro, Haeundae-gu, Busan 48062, Republic of Korea.
[2] Department of Radiological Science, College of Health Sciences, Catholic University of Pusan, Busan 46252, Republic of Korea.

Abstract

This study aimed to evaluate the diagnostic performance of several thyroid ultrasound risk-stratification systems in nodules with indeterminate or suspicious cytology, within a population with a history of iodine deficiency. The systems assessed included ACR-TIRADS (American College of Radiology), EU-TIRADS (European Thyroid Association), Korean-TIRADS, Kwak-TIRADS, AACE/ACE-AME guidelines (American Association of Clinical Endocrinologists/American College of Endocrinology–Associazione Medici Endocrinologi), and ATA guidelines (American Thyroid Association). A total of 1,000 nodules with confirmed histopathology were analyzed: 329 FLUS/AUS (10.6% malignant), 167 SFN/SHT (11.6% malignant), 44 SM (77.3% malignant), 298 benign lesions, and 162 malignant neoplasms. The proportion of papillary thyroid carcinoma (PTC) was highest in Bethesda MN (86.4%) and SM (91.2%) nodules compared to FLUS/AUS (57.1%, $p < 0.005$) and SFN/SHT (36.8%, $p < 0.001$). Diagnostic performance of TIRADS was superior for MN (AUC: 0.827–0.874) and SM nodules (AUC: 0.775–0.851), while lower for FLUS/AUS (AUC: 0.655–0.701) and SFN/SHT nodules (AUC: 0.593–0.621). Among FLUS/AUS nodules classified as high-risk by TIRADS, malignancy risk was 25%, whereas TIRADS categories did not alter malignancy risk in the SFN/SHT group. EU-TIRADS and AACE/ACE-AME guidelines identified the highest number of PTC, FTC, HTC, and MTC cases, while Kwak-TIRADS (OR = 12.6) and Korean-TIRADS (OR = 12.0) showed the strongest predictive value. In conclusion, TIRADS effectiveness is influenced by the prevalence of PTC. All systems aid in selecting FLUS/AUS nodules for surgical intervention but are less useful in guiding management of SFN/SHT nodules.

Keywords: TIRADS, Thyroid ultrasonography, FNA, Thyroid cancer

Corresponding author: Su-Min Hong
E-mail: Dayeejeung@gmail.com

How to Cite This Article: Jeung DY, Hong SM. Comparative Diagnostic Performance of Six TIRADS Classifications in the Evaluation of Cytologically Indeterminate Thyroid Nodule. Bull Pioneer Res Med Clin Sci. 2022;1(2):72-82. https://doi.org/10.51847/dcxP4Iwhgk

Introduction

Preoperative evaluation of thyroid nodules remains an active area of research, particularly regarding the utility of Thyroid Imaging Reporting and Data Systems (TIRADS) in guiding fine-needle aspiration (FNA) and estimating the malignancy risk (RoM) of cytologically indeterminate nodules. TIRADS are based on ultrasound (US) features associated with malignancy. Individually, these features

often lack both high sensitivity and specificity, prompting the development of multiple feature combinations. However, there is no consensus on the significance of specific features or their optimal integration. Although not all proposed risk-stratification systems are formally termed "TIRADS," this term is used throughout this study for simplicity.

The earliest TIRADS systems, proposed by Horvath *et al.* and Park *et al.* in 2009, were complex and challenging to apply in routine practice. Subsequently, simplified versions were developed in Asia, Europe, and the United States [1, 2]. In Korea, Kwak *et al.* introduced a simplified system based on several US malignancy features—hypoechogenicity, irregular or microlobulated margins, microcalcifications, taller-than-wide shape, and solid echostructure—assigning equal weight to each [3]. Later iterations incorporated weighted scores for each feature based on odds ratios for malignancy, culminating in the K-TIRADS recommended by the Korean Society of Thyroid Radiology (KSThR), which prioritizes evaluation of nodule structure followed by high-specificity features [4,5].

In Europe, the French Society of Endocrinology adapted Horvath's original system into a simplified five-point scale (French-TIRADS) [6], subsequently adopted by the European Thyroid Association (EU-TIRADS), which emphasizes marked hypoechogenicity, irregular shape or margins, and microcalcifications. In the USA, multiple systems were developed concurrently. The ATA recommends a five-category system emphasizing high-specificity US features in hypoechoic nodules [7, 8], while the AACE/ACE-AME guidelines use a three-tier scale, classifying nodules as high-risk when any high-specificity feature is present [9]. The ACR-TIRADS assigns points to individual features, with the total score determining the final category [10].

All systems relate US risk categories to size thresholds for biopsy. Nodules undergoing FNA are classified using the Bethesda System for Reporting Thyroid Cytopathology (BSRTC) [11, 12], which includes non-diagnostic, benign (BL), malignant (MN), and three indeterminate categories: FLUS/AUS, SFN/SHT, and SM. Management of indeterminate nodules depends on combined clinical, cytological, US, and sometimes molecular evaluation.

Epidemiological factors, such as iodine deficiency, influence the distribution of thyroid lesions and the relative frequency of papillary versus follicular thyroid carcinoma, potentially affecting the predictive value of US features [13,14]. Data on TIRADS performance in iodine-deficient populations, especially with histopathological verification, remain limited.

The present study aimed to evaluate the diagnostic performance of selected thyroid sonographic risk-stratification systems for nodules with indeterminate, suspicious, or unequivocal cytology in a population with a history of iodine deficiency.

Materials and Methods

Examined patients

Fine-needle aspiration (FNA) and ultrasound (US) examinations were conducted at a single center between 2010 and 2019 on patients referred by endocrinologists from outpatient clinics. The majority of the cohort had lifelong exposure to moderate iodine deficiency. In the 1990s, our country was classified as moderately iodine-deficient according to the International Council for Control of Iodine Deficiency Disorders. Mandatory household salt iodization was implemented in 1997, and its efficacy in reducing goiter prevalence among school-aged children to below 5% was confirmed by 2005 [15]. Nearly 90% of the patients had experienced moderate iodine deficiency for at least half of their lives, with only 10.4% being under 44 years of age and thus exposed to a longer period of sufficient iodine intake (maximum 22 years).

The study analyzed 1000 nodules from 866 patients, all with complete US imaging data, diagnostic FNA results, and postoperative histopathological confirmation (see Figure S1, Supplementary Material). Patients with prior thyroid surgery, radioiodine therapy, or a history of neck irradiation were excluded. The analyzed nodules included all Bethesda categories III–VI and a selection of category II nodules that underwent biopsy to reach the total of 1000 nodules. Among these, 540 nodules were equivocal (EC), comprising 329 FLUS/AUS, 167 SFN/SHT, and 44 SM, while 460 nodules were unequivocal (UC), including 298 benign lesions (BL) and 162 malignant neoplasms (MN) **(Table 1)**.

Table 1. Demographic data of the patients and the percentage of cancers revealed in the nodules with unequivocal (UC) and equivocal (EC) FNA results

Parameter	Category of FNA					*p*
	UC (460)		EC (560)			
	BL	MN	FLUS/AUS	SFN/SHT	SM	
Number of nodules	298	162	329	167	44	
Number of patients	240	141	290	152	43	
Age—mean ± SD [years]	54.7 ± 11.6	50.3 ± 13.9	53.7 ± 13.6	54.1 ± 14.8	56.4 ± 14.4	$p < 0.01$ MN vs. others

No/% of males	18/7.5	20/14.2	32/11.0	15/9.9	5/11.6	NS
Volume of nodules mean ± SD [cm^3]	7.9 ± 15.4	4.6 ± 13.9	6.6 ± 13.6	5.9 ± 12.9	3.2 ± 5.6	NS
No of Ben/Mal nodules < 1 cm	16/0	0/47	13/1	22/2	1/11	
No/% of cancers	0/0.0	162/100.0	35/10.6	19/11.4	34/77.3	$p < 0.0001$ MN & SM vs. others
No/% of PTCs among cancers	0/0.0	140/86.4	20/57.1	7/36.8	31/91.2	$p < 0.005$ MN & SM vs. FLUS/AUS, SFN/SHT
Other cancers (No/%)	-	FTC (3/1.9) HTC (1/0.6) PDTC (2/1.2) AC (1/0.6) MTC (13/9.0) ST (2/1.2)	FTC (7/20.0) HTC (4/11.4) AC (1/2.8) MTC (2/5.7) ANG (1/2.8)	FTC (5/26.3) HTC (7/36.8)	FTC (1/2.9) HTC (1/2.9) MTC (1/2.9)	

BL, benign lesion; FLUS/AUS, follicular lesions of undetermined significance/atypia of undetermined significance; SFN/SHT, suspicion of follicular neoplasm/suspicion of Hürthle cell tumor; SM, suspicion of malignancy; MN, malignant neoplasm; PTC, papillary thyroid carcinoma; MTC, medullary thyroid carcinoma; FTC, follicular thyroid carcinoma; HTC, Hurthle cell thyroid carcionoma; PDTC, poorly differentiated thyroid carcinoma; AC, anaplastic carcinoma; ST, secondary tumor; ANG, angiosarcoma; Ben, benign lesion in histopathological outcome; Mal, thyroid malignancy in histopathological outcome.

Microscopic examination

FNA was performed on thyroid nodules measuring at least 5 mm (typically >1 cm) that exhibited one or more clinical or sonographic risk factors for malignancy. In most cases, two passes per nodule were obtained. Aspirates were fixed in 95% ethanol and stained with hematoxylin and eosin. Surgical thyroid specimens were processed following standard protocols. A detailed description of the classification of nodules according to the Bethesda System for Reporting Thyroid Cytopathology (BSRTC) was reported previously [16]. Notably, category IV excluded lesions displaying nuclear features of papillary thyroid carcinoma (PTC). Nodules exhibiting overlapping features of categories II and IV were assigned to category III. Rarely, specimens with otherwise benign morphology but focal nuclear atypia suggestive of PTC were classified as category III.

Patients with cytological diagnoses of SFN/SHT, SM, or MN were routinely referred for surgical intervention, whereas those with BL or FLUS/AUS underwent surgery based on clinical indications, nodule size, or patient preference. Histopathological evaluation adhered to the WHO classification of thyroid tumors in effect at the time. Reclassification to identify non-invasive follicular thyroid neoplasms with papillary-like nuclear features (NIFTP) was not performed; the sole post-NIFTP case was excluded. Histopathology confirmed all unequivocal FNA results (BL and MN) and revealed malignancy rates of 10.6% for FLUS/AUS, 11.6% for SFN/SHT, and 77.3% for SM nodules. The proportion of PTC among cancers was significantly higher in cytologically MN (86.4%) and SM (91.2%) nodules than in FLUS/AUS (57.1%, $p < 0.005$) or SFN/SHT nodules (36.8%, $p < 0.001$).

Analysis of ultrasound malignancy features

US malignancy features were assessed prospectively by experienced sonographers (three with over 20 years' experience and two with ten years' experience), immediately prior to FNA, following a standardized departmental protocol. Nodule measurements and the presence of specific features—including marked hypoechogenicity, hypoechogenicity, solid structure, taller-than-wide shape, irregular or suspicious margins, micro- and macrocalcifications, rim calcifications, and pathological intranodular vascularization—were recorded in a dedicated database. Additional features, such as predominantly cystic or spongiform echostructure, were also documented. Examinations were performed using the Aloka Prosound Alpha 7 system (7.5–14 MHz linear transducer, ALOKA Co. Ltd., Tokyo, Japan).

All nodules were classified according to six TIRADS systems: EU-TIRADS (EU-T) [7], K-TIRADS (K-T) [5], ACR-TIRADS (ACR-T) [10], Kwak-TIRADS (Kw-T) [3], ATA-T [8], and 3A-T (AACE/ACE/AME) [9]. Two independent researchers (KWK and DSK) assigned US features for TIRADS scoring; discrepancies (39 nodules) were resolved via joint reevaluation. Modifications were applied for the ATA-T system to account for iso- or hyperechoic nodules with high-risk features, resulting in 51 nodules (5.1%) being assigned to the highly suspicious category.

Statistical analysis

The distribution of US malignancy features was evaluated in relation to FNA categories and final histopathology.

Associations between individual features and malignancy were assessed using logistic regression, with odds ratios (OR) and 95% confidence intervals (CI) calculated separately for unequivocal (UC, categories II and VI) and equivocal cytology (EC, categories III–V) nodules.

Nodules were subsequently categorized according to each TIRADS system, allowing the calculation of the proportion of cancers within each TIRADS category (T-RoM) and its impact on the malignancy risk associated with FNA category (FNA-RoM). Receiver operating characteristic (ROC) curves and area under the curve (AUC) values were used to identify optimal cut-off categories for distinguishing benign from malignant nodules. Sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), and the proportion of nodules meeting thresholds were calculated. Odds ratios for the established cut-offs were determined via logistic regression.

Statistical analyses were performed using Statistica version 10. Comparisons of categorical variables employed the χ^2 test, with adjustments for sample size as appropriate, and the Kruskal–Wallis test was applied for continuous variables. A p-value < 0.05 was considered statistically significant. The study was approved by the Local Bioethics Committee, and all patients provided informed consent.

Results and Discussion

The incidence of individual ultrasound (US) malignancy features in nodules with unequivocal cytology (UC) and in the subgroups of equivocal cytology (EC) nodules, stratified by final histopathological outcome (malignant vs. benign), is presented in Table S1 (Supplementary Material). In the UC group, all assessed US malignancy features were more frequently observed in malignant nodules compared to benign lesions, except for macrocalcifications without microcalcifications and isolated rim calcifications. Logistic regression analysis confirmed seven US features as independent predictors of malignancy in UC nodules: marked hypoechogenicity (OR: 9.8, 95% CI: 3.7–26.1, $p < 0.0001$), hypoechogenicity (OR: 4.0, 95% CI: 2.0–8.0, $p < 0.0001$), solid echostructure (OR: 3.3, 95% CI: 1.2–8.9, $p < 0.05$), suspicious shape (OR: 4.0, 95% CI: 1.6–9.8, $p < 0.005$), suspicious margins (OR: 6.8, 95% CI: 3.0–15.5, $p < 0.0001$), microcalcifications (OR: 14.9, 95% CI: 4.5–49.7, $p < 0.0001$), and pathological vascularization (OR: 2.3, 95% CI: 1.1–4.9, $p < 0.05$).

In the EC group, marked hypoechogenicity was the only feature consistently differentiating malignant from benign nodules across all subgroups. Suspicious margins were more frequent in cancers among FLUS/AUS and SM nodules, whereas microcalcifications were predictive only in the FLUS/AUS subgroup. Logistic regression confirmed microcalcifications (OR: 6.9, 95% CI: 2.2–21.6, $p < 0.005$) and suspicious margins (OR: 3.7, 95% CI: 1.1–11.8, $p < 0.05$) as independent predictors in FLUS/AUS nodules, and marked hypoechogenicity in SM nodules (OR: 4.4, 95% CI: 1.4–13.4, $p < 0.01$).

Table 2 presents the distribution of benign and malignant nodules across the categories of each TIRADS system, along with the category-specific risk of malignancy (T-RoM) and corresponding AUC values. Overall, T-RoM aligned with expected values, except in certain categories: high suspicion ATA-T (which increased to 62.3% after inclusion of iso- or hyperechoic nodules with high-risk features), low-risk EU-T, low-suspicion Kw-T, and mildly suspicious ACR-T, where T-RoM exceeded expectations. Diagnostic efficacy, as measured by AUC, ranged from 0.763 for 3A-T to 0.793 for Kw-T for the entire cohort. Efficacy was higher in groups with a high proportion of PTC among cancers (UC and SM) and lower in FLUS/AUS and SFN/SHT groups, where AUCs were not statistically significant. Excluding non-hypoechoic nodules from ATA-T category 5 decreased its AUC in FLUS/AUS and SFN/SHT groups but increased it in UC and SM nodules.

In UC nodules, classification into the highest-risk TIRADS category significantly increased the nodule's malignancy risk compared to its initial FNA-RoM. For Kw-T, even category 4c significantly elevated malignancy risk. In the EC group, only FLUS/AUS nodules demonstrated a significant increase in RoM across all TIRADS systems. In SFN nodules, no TIRADS significantly improved RoM estimation, though Kw-T at category 4c approached significance (11.2% increase). In SM nodules, RoM increased to 100% for all systems, but significance was reached only for ATA-T and EU-T. Conversely, assignment to the lowest-risk categories significantly reduced RoM in UC nodules across most TIRADS, whereas no similar effect was observed in EC subgroups regardless of threshold.

Table 4 summarizes diagnostic performance at threshold categories with maximal accuracy. EU-T and 3A-T demonstrated the highest sensitivity across all groups (UC: 77.8%, SM: 61.8%, FLUS/AUS: 51.4%, SFN/SHT: 52.6%), while ACR-T showed the lowest. Specificity exceeded 80% for all systems in UC, SM, and FLUS/AUS groups, reaching >90% for K-T, Kw-T, and ACR-T. In SFN/SHT nodules, only K-T, Kw-T, and ACR-T achieved >80% specificity. Across all groups, the highest combined sensitivity and specificity were observed for EU-T and 3A-T.

Lowering threshold categories by one grade (K-T, EU-T, ATA-T, ACR-T: category 4; 3A-T: category 2; Kw-T: 4b; Table S2) improved sensitivity to >90% in UC nodules (highest: 3A-T 100%, EU-T 96.3%), with specificity ranging from 54.4–62.8%, except 3A-T (16.4%). EC

subgroups achieved ≥80% sensitivity at these thresholds, with specificity varying from 29.9–50%, except for 3A-T, which showed very low specificity (SM: 10%, FLUS/AUS: 3.7%, SFN/SHT: 0.7%).

Table 5 details the number and type of cancers detected when threshold categories were set for maximal accuracy. EU-T and 3A-T would identify the greatest number of total cancers and each histologic subtype (PTC, FTC, HTC, MTC), though at the cost of the highest proportion of biopsied nodules (32.2%). ACR-T would reduce the number of FNAs by 13.6%, but sensitivity would decline by 22% compared to EU-T and 3A-T. The greatest increase in RoM at optimal thresholds was observed for Kw-T and K-T (OR 12.6 and 12.0, respectively), consistent even when 51 nodules outside ATA-T criteria were excluded.

Sensitivity was higher for PTC than for FTC or HTC across all systems. Most PTCs were classified into the highest-risk category (or Kw-T 4c), while FTC and HTC were generally assigned to categories one grade lower **(Figure 1)**.

Table 2. Distribution of benign and malignant nodules between particular categories of Thyroid Imaging Reporting and Data Systems (TIRADS), the comparison of expected T-ROM with calculated T-ROM for each TIRADS and diagnostic efficacy of evaluated TIRADS as measured with AUC (TIRADS categories corresponding to the lack of nodules have been omitted)

Category of TIRADS/Guideline		Expected T-RoM	Calculated T-RoM	Mal./Ben. Nodules	AUC (95%CI)
3A-T	1—low-risk thyroid lesion	1	1.6	1/62	0.763 (0.728–0.798) $p < 0.0001$
	2—intermediate-risk thyroid lesion	5–15	12.0	74/541	
	3—high-risk thyroid lesion	50–90	54.3	175/147	
K-T	2—benign	<3	0.0	0/51	0.788 [c] (0.755–0.821) $p < 0.0001$
	3—low suspicion	3–15	7.8	25/295	
	4—intermediate	15–50	21.5	93/340	
	5—high suspicion	>60	67.3	132/64	
EU-T	2—benign	0	0.0	0/47	0.784 [d] (0.752–0.816) $p < 0.0001$
	3—low risk	2–4	6.7	17/238	
	4—intermediate risk	6–17	15.4	58/318	
	5—high risk	26–87	54.3	175/147	
Kw-T	3—probably benign	0	2.53	3/116	0.793 [a,b] (0.760–0.825) $p < 0.0001$
	4a—low suspicion for malignancy	2–3	9.3	24/235	
	4b—intermediate suspicion for malignancy	7–38	20.5	86/333	
	4c—moderate concern, not classic for malignancy	21–92	66.7	128/64	
	5—highly suggestive of malignancy	89–98	81.8	9/2	
ACR-T	1—benign	–	0.0	0/48	0.771 (0.738–0.804) $p < 0.0001$
	2—not suspicious	<2	3.0	2/64	
	3—mildly suspicious	5	8.9	20/204	
	4—moderately suspicious	5–20	22.7	108/368	
	5—highly suspicious	>20	64.5	120/66	
ATA-T	1—benign	<1	0.0	0/1	0.778 (0.746–0.811) $p < 0.0001$
	2—very low suspicion	<3	1.2	1/81	
	3—low suspicion	5–10	8.4	24/260	
	4—intermediate suspicion	10–20	19.9	72/290	
	5—high suspicion	70–90	56.5	153/118	

a, $p < 0.05$ vs. 3A-T, ATA-T; b, $p < 0.005$ vs. ACR-T; c, $p < 0.05$ vs. 3A-T, ACR-T; d, $p < 0.0001$ vs. 3A-T.

Table 3. Diagnostic efficacy of evaluated TIRADS as measured with AUC in the UC group and subgroups of the EC group; the change from FNA-ROM of a nodule in relation to its TIRADS category

IRADS/Guideline Category	EC FLUS/AUS FNA-RoM: 10.6%			EC SFN/SHT FNA-RoM: 11.4%			SM FNA-RoM: 77.3%			UC BL & MN FNA-RoM: 35.2%		
	AUC	T-RoM	FNA-RoM vs. T-RoM	AUC	T-RoM	FNA-RoM vs. T-RoM	AUC	T-RoM	FNA-RoM vs. T-RoM	AUC	T-RoM	FNA-RoM vs. T-RoM

		p		*p*	*p*		*p*	*p*		*p*	*p*		*p*
3A-T	1	0.674	0.0	NS	0.613	0.0	NS	0.813	50.0	NS	0.827	0.0	<0.0001
	2	<0.005	6.9	NS	NS	8.1	NS	<0.0001	57.1	NS	<0.0001	15.2	<0.0001
	3		25.0	<0.005		18.2	NS		100.0	<0.05		72.4	<0.0001
K-T	2	0.692	0.0	NS	0.603	0.0	NS	0.803	0.0	NS	0.864 [d,e]	0.0	<0.0001
	3		5.2	NS		7.9	NS		60.0	NS		6.4	<0.0001
	4	<0.0001	10.2	NS	NS	9.9	NS	<0.0001	66.7	NS	<0.0001	37.3	NS
	5		34.3	<0.001		22.2	NS		100.0	NS		82.8	<0.0001
EU-T	2	0.693	0.0	NS	0.605	0.0	NS	0.851 [f]	0.0	NS	0.855 [d]	0.0	<0.0001
	3		4.9	NS		9.4	NS		50.0	NS		4.5	<0.0001
	4	<0.0001	7.9	NS	NS	7.6	NS	<0.0001	64.3	NS	<0.0001	25.4	<0.05
	5		25.0	<0.005		18.2	NS		100.0	<0.05		72.4	<0.0001
Kw-T	1	0.681	3.2	NS	0.621 [h]	0.0	NS	0.790	50.0	NS	0.874 [a,b,c]	0.0	<0.0001
	2		6.1	NS		8.8	NS		57.1	NS		9.2	<0.0001
	3		9.8	NS	NS	9.1	NS	<0.0001	66.7	NS	<0.0001	35.9	NS
	4	<0.0005	32.4	<0.001		25.0	NS		100.0	NS		82.3	<0.0001
	5		50.0	NS		0.0	NS		100.0	NS		100.0	<0.005
ACR-T	1	0.655 [g]	0.0	NS	0.593	0.0	NS	0.775	0.0	NS	0.857 [d]	0.0	<0.0001
	2		0.0	NS		0.0	NS		66.7	NS		0.0	<0.0001
	3	<0.005	6.8	NS	NS	10.0	NS	<0.0005	50.0	NS	<0.0001	8.0	<0.0001
	4		11.0	NS		9.6	NS		76.2	NS		36.5	NS
	5		27.3	<0.05		21.4	NS		100.0	NS		82.1	<0.0001
ATA-T	1		-	-		-	-		-	-		0.0	NS
	2	0.701	0.0	NS	0.589	0.0	NS	0.810	50.0	NS	0.843	0.0	<0.0001
	3	<0.005	5.6	NS	NS	8.1	NS	<0.0001	55.6	NS	<0.0001	7.6	<0.0001
	4	0.652 *	8.9	NS	0.554 *	10.2	NS	0.847 *	61.5	NS	0.862 *	36.5	NS
	5		28.1	<0.001		17.5	NS		100.0	NS		71.4	<0.0001

*, Value of AUC after the exclusion of non-hypoechoic nodules from the ATA-T category 5; a, $p < 0.001$ vs. ATA-T, 3A-T; b, $p < 0.005$ vs. K-T; c, $p < 0.05$ vs. ACR-T; d, $p < 0.005$ vs. 3A-T; e, $p < 0.05$ vs. ATA-T; f, $p < 0.05$ vs. ATA-T, ACR-T, K-T, Kw-T; g, $p < 0.05$ vs. K-T, ATA-T; h, $p < 0.005$ vs. ATA-T.

Table 4. Data on the diagnostic efficacy of analyzed TIRADSs in examined groups of nodules—data for the thresholds that gave the highest ACC values

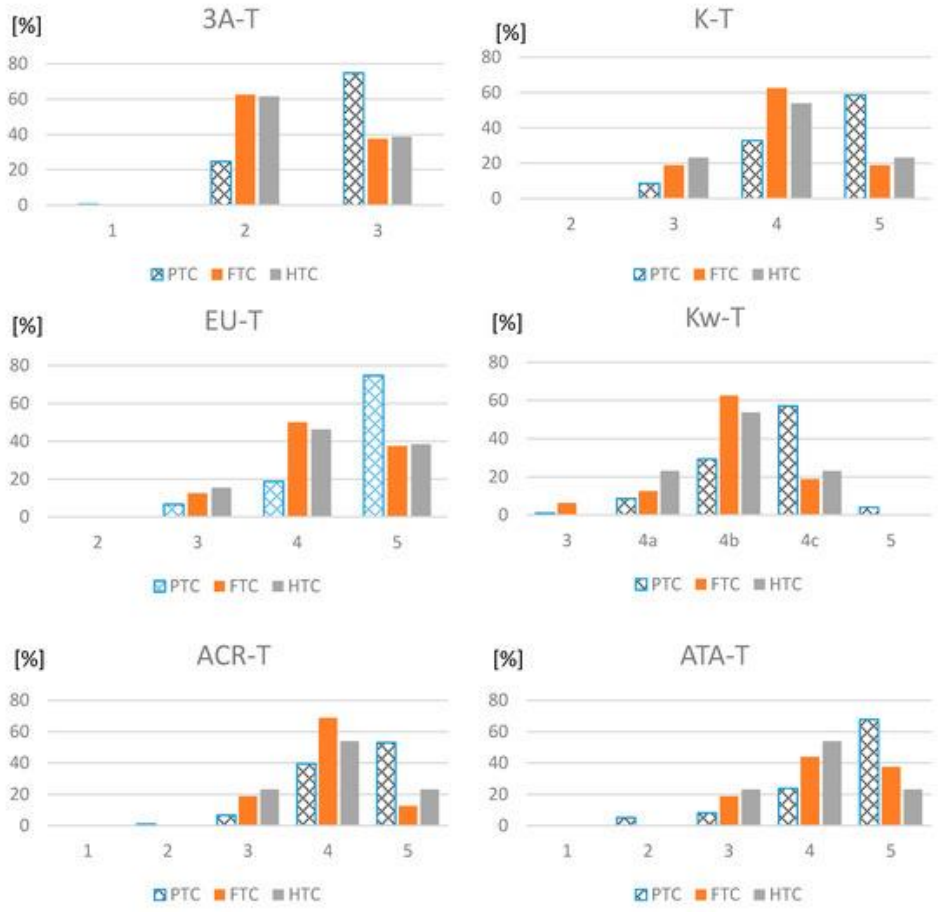
TIRADS/Guideline Threshold Category		SEN	SPC	ACC	PPV	NPV	% of Nodules	SEN	SPC	ACC	PPV	NPV	% of Nodules
					UC						SM		
3A-T	3	77.8	83.9	81.7	72.4	87.4	37.8	61.8	100.0	70.5	100.0	43.5	47.7
K-T	5	59.3	93.3	81.3	82.8	80.8	25.2	52.9	100.0	63.6	100.0	38.5	40.9
EU-T	5	77.8	83.9	81.7	72.4	87.4	37.8	61.8	100.0	70.5	100.0	43.5	47.7
Kw-T	4c	61.7	93.3	82.2	83.3	81.8	26.1	52.9	100.0	63.6	100.0	38.5	40.9
ACR-T	5	56.8	93.3	80.4	82.1	79.9	24.3	38.2	100.0	52.3	100.0	32.2	29.5
ATA-T	5	67.9	85.2	79.1	71.4	83.0	33.5	58.8	100.0	68.2	100.0	41.7	45.5
					FLUS/AUS						SFN/SHT		
3A-T	3	51.4	81.6	78.4	25.0	93.4	21.9	52.6	69.6	67.7	18.2	92.0	32.9
K-T	5	34.3	92.2	86.0	34.3	92.2	10.6	31.6	85.8	79.6	22.2	90.7	16.2
EU-T	5	51.4	81.6	78.4	25.0	93.4	21.9	52.6	69.6	67.7	18.2	92.0	32.9
Kw-T	4c	34.3	91.8	85.7	33.3	92.2	10.9	36.8	85.1	79.6	24.1	91.3	17.4
ACR-T	5	25.7	91.8	84.8	27.3	91.2	10.0	31.6	85.1	79.0	21.4	90.6	16.8
ATA-T	5	45.7	86.1	81.8	28.1	93.0	17.3	36.8	77.7	73.1	17.5	90.6	23.9

Table 5. Data on the number and percentage of detected cancers in the whole examined sample (for the threshold values that gave the maximum AUC)

0		No/% of nodules	No/% of cancers	No/% of Cancers ≥ 1 cm	No/% of PTC	No/% of FTC	No/% of HCT	No/% of MTC	OR 95%CI *
3A-T	3	322/32.2 [d,e]	175/70.0 [a,b,c,d]	124/65.6	148/74.7 [a,b,c]	6/37.5	5/38.5	12/75.0	9.6 (6.9–13.2)
K-T	5	196/19.6	132/52.8	88/46.6	116/58.6	3/18.8	3/23.1	8/50.0	12.0 (8.4–17.1)
EU-T	5	322/32.2 [d,e]	175/70.0 [a,b,c,d]	124/65.6	148/74.7 [a,b,c]	6/37.5	5/38.5	12/75.0	9.6

									(6.9–13.2)
Kw-T	4c	203/20.3	137/54.8	91/48.1	121/61.1	3/18.8	3/23.1	8/50.0	12.6 (8.8–17.9)
ACR-T	5	186/18.6	120/48.0	81/42.9	105/53.0	2/12.5	3/23.1	8/50.0	9.6 (6.7–13.6)
ATA-T	5	271/27.1	153/61.2 (16) #	107/56.6 (14) #	134/67.7 (14) #	6/37.5 (2) #	3/23.1	8/50.0	8.4 (6.1–11.7)

*, $p < 0.0001$ in all cases; #, cancers in nodules other than hypoechoic; a, $p < 0.0001$ vs. ACR-T; b, $p < 0.001$ vs. K-T; c, $p < 0.005$ vs. Kw-T; d, $p < 0.05$ vs. ATA-T; e, $p < 0.0001$ vs. Kw-T, K-T, ACR-T.



Figure

Comparing the performance of different TIRADS systems across populations or assessing a single TIRADS in various cohorts is challenging. In a meta-analysis by Kim *et al.* which included four systems—ACR-T, ATA-T, K-T, and EU-T—the overall diagnostic performance was considered comparable, with EU-T showing the highest pooled sensitivity and specificity [17]. Conversely, Castellana *et al.* reported markedly lower sensitivity for EU-T in selecting nodules for FNA compared to 3A-T, ATA-T, K-T, and ACR-T [18]. Differences among studies can arise from varying threshold levels, methods for confirming final diagnoses (histopathology, cytology, or clinical follow-up), and selection of nodules relative to FNA categories. Most studies, including ours, excluded non-diagnostic FNAs, yet some also excluded indeterminate or suspicious nodules [19]. This exclusion significantly affects results, as Bethesda indeterminate categories often include FTC and HTC, which have ultrasound characteristics distinct from PTC [14, 18, 20–22].

Our findings align with this observation: the diagnostic value of US malignancy features and TIRADS decreases as the proportion of PTC among cancers declines. The evaluated systems demonstrated good efficacy in cytological categories MN and SM, where PTC accounted for 86.4% and 91.2% of malignancies, respectively. However, performance was markedly lower in indeterminate cytology, particularly in SFN/SHT nodules, where PTC represented <40% of cancers and overall RoM was below 15%. Two main factors explain this: (1) the epidemiology of our population, long exposed to iodine deficiency, resulting in SFN/SHT nodules primarily representing non-neoplastic follicular lesions and a lower PTC-to-FTC ratio [23]; and (2) a conservative approach in

assigning smears to Bethesda category IV, with pathologists avoiding this classification for lesions exhibiting nuclear features of PTC. As a result, TIRADS were inefficient in distinguishing benign from malignant SFN/SHT nodules, with categorization into the highest- or lowest-risk categories failing to significantly alter RoM. Notably, no cancers were assigned to the lowest-risk category.

In Bethesda category III, US features and TIRADS demonstrated higher diagnostic utility. Although FNA-RoM was similar between FLUS/AUS and SFN/SHT nodules, PTC prevalence was over 20 percentage points higher in FLUS/AUS nodules. Consequently, classification into high-risk categories (e.g., Kw-T 4c or equivalent in other TIRADS) significantly increased RoM to levels justifying surgical intervention. This finding is particularly relevant for patients with repeated category III FNAs, where additional biopsies often do not clarify clinical management.

Reports on TIRADS utility in indeterminate nodules vary due to differences in selection criteria and baseline FNA-RoM, especially for FLUS/AUS nodules, whose malignancy risk ranges widely from a few percent to 70% [24]. Centers where category III is dominated by smears with nuclear atypia or PTC-like features report higher PTC prevalence and greater TIRADS efficacy compared to populations, like ours, where category III primarily includes nodules with borderline cytological changes between categories II and IV [25, 26]. Expectations for TIRADS use should consider these differences.

Several studies support selective utility of TIRADS in indeterminate nodules. Grani *et al.* reported ATA-T and the older K-T version effectively excluded malignancy in TIR3 nodules (Italian Consensus) [27]. Tang *et al.* found ATA-T predictive in FLUS/AUS nodules [28]. Kamaya *et al.* confirmed utility for Kw-T [29], while Lee *et al.* observed ATA-T useful only in the AUS subcategory [30]. Yoon JH *et al.* reported similar results for Kw-T [26]. Hong *et al.* consistent with our findings, noted high-suspicion K-T patterns significantly increased malignancy risk in FLUS/AUS but not SFN/SHT nodules [31]. Valderrabano *et al.* suggested ATA patterns could guide individualized management in both FLUS/AUS and SFN nodules, without differences in histological malignancy distribution [32]. Ahmadi *et al.* and Barbosa *et al.* (2019) reported comparable observations for ATA-T and ACR-T, though Barbosa *et al.* combined Bethesda IV and V categories (FNA-RoM 61.5%) [33, 34]. Yang *et al.*, like us, found ACR-T, ATA-T, and K-T unhelpful for RoM assessment in category IV nodules [35]. Chaigneau *et al.* observed French TIRADS (similar to EU-T) provided significant risk stratification only in Bethesda V, not in III or IV nodules [36].

In the entire cohort, including nodules with both unequivocal and equivocal cytology, the overall diagnostic performance of the evaluated TIRADS systems was comparable. Among them, Kw-T, K-T, and EU-T demonstrated slightly higher AUC values. When thresholds maximizing overall accuracy were applied (Kw-T category 4c and the high-risk category for other TIRADS), EU-T and 3A-T exhibited the highest sensitivity. These systems detected a larger number of cancers, including those >1 cm in diameter, and were effective in identifying both PTC and other common thyroid carcinomas (FTC, HTC, MTC). However, applying these thresholds as criteria for FNA would result in the highest number of biopsies. In contrast, ACR-T allowed a substantial reduction in the number of FNAs while maintaining high specificity, albeit at the cost of reduced sensitivity—a feature noted in previous studies [37–40]. The most favorable balance between the number of detected cancers and the number of FNAs was achieved with Kw-T and K-T, which also demonstrated the highest odds ratios (ORs).

In our dataset, the highest-risk categories of 3A-T and EU-T encompassed the same nodules, reflecting the similarity in their classification criteria. While 3A-T also considered extrathyroidal extension as an additional criterion, such cases without other high-risk features were not observed. Minimal differences between these categories were similarly reported by Grani *et al.* [37]. The advantage of EU-T lies in its four-grade scale, which allows finer stratification and flexibility in choosing thresholds depending on whether sensitivity or specificity optimization is desired. Other studies also support EU-T's discriminative value [41, 42]. Analogous flexibility can be found in K-T, ATA-T, ACR-T, and Kw-T. For 3A-T, lowering the threshold to category 2 nearly achieved 100% sensitivity but at the expense of very low specificity (<20%).

Comparative studies of TIRADS consistently highlight high sensitivity and AUC for K-T and Kw-T, and high specificity for ACR-T, though often with lower sensitivity [37–39]. Xu *et al.* reported the highest sensitivity for ACR-T, using a lower threshold than other systems [43]. Lauria Pantano *et al.* found ACR-T superior to ATA-T and 3A-T in detecting nodules with high cytological risk, though final diagnoses were not verified histologically [44]. In our study, no significant differences in AUC were observed among these systems, although excluding nodules that did not meet ATA-T criteria increased ATA-T's AUC above that of ACR-T, consistent with Gao *et al.* [38].

A common observation across studies is that AUC correlates with the proportion of PTC among cancers. Shen *et al.*, in a sample with 95.5% PTC, reported AUC values of 0.869–0.896 for ACR-T, ATA-T, EU-T, and

Kw-T [39], whereas Grani *et al.*, with 75% PTC, reported lower AUCs of 0.55–0.70 [37]. Similarly, Trimboli *et al.* noted low overall accuracy for ATA-T and 3A-T in a cohort with indeterminate cytology (101 nodules, 21% malignancy, 57% PTC) [45].

T-RoM values observed in our study largely matched expected ranges, with the exception of the high-risk category of ATA-T, which was lower than anticipated. This aligns with findings in Italian [46] and Brazilian populations [47] and partly reflects inclusion of non-hypoechoic nodules in this category. The T-RoM for these additional nodules was 31.4%, close to iso-/hyperechoic nodules with suspicious features reported by Gao *et al.* (25.9%) [48], but higher than for partially cystic nodules. Additionally, FTC and HTC accounted for 11.6% of all cancers, which are typically assigned lower TIRADS categories. Consequently, T-RoM was slightly elevated for low-risk categories of EU-T (6.7% vs. expected 2–4%) and mildly suspicious ACR-T (8.9% vs. 5%), with minor differences for other categories.

Study limitations include selection of nodules based on postoperative histopathology, which may affect generalizability, but also ensures diagnostic certainty. A major strength is the prospective evaluation of US malignancy features immediately before FNA, preventing bias from cytology results. Another limitation is the relatively small number of cancers in FLUS/AUS and SFN/SHT subgroups, reflecting the low malignancy risk in nodules from a population historically exposed to iodine deficiency.

Conclusion

The diagnostic performance of TIRADS is influenced by the proportion of PTC among cancers and is generally lower for nodules with indeterminate cytology compared to those with unequivocal cytology. All evaluated TIRADS systems are useful for selecting FLUS/AUS nodules for surgical management in populations with a low malignancy risk and a low prevalence of PTC. However, these systems are less effective for managing SFN/SHT nodules in such populations. While the overall diagnostic efficacy of the TIRADS systems is comparable, certain limitations exist: ATA-T does not cover all nodule patterns; 3A-T lacks a threshold that simultaneously optimizes sensitivity and specificity; and ACR-T prioritizes specificity at the expense of inadequate sensitivity at the highest-risk category. Among the evaluated systems, EU-T demonstrates the greatest versatility and reliability across different types of thyroid cancers.

Acknowledgments: None.

Conflict of interest: None.

Financial support: None.

Ethics statement: None.

References

1. Horvath E, Majlis S, Rossi R, Franco C, Niedmann JP, Castro A, et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. J Clin Endocrinol Metab. 2009;94:1748–51.
2. Park JY, Lee HJ, Jang HW, Kim HK, Yi JH, Lee W, et al. A proposal for a thyroid imaging reporting and data system for ultrasound features of thyroid carcinoma. Thyroid. 2009;19:1257–64.
3. Kwak JY, Han K, Yoon JH, Moon HJ, Son EJ, Park SH, et al. Thyroid imaging reporting and data system for us features of nodules: A step in establishing better stratification of cancer risk. Radiology. 2011;260:892–9.
4. Kwak JY, Jung I, Baek JH, Baek SM, Choi N, Choi YJ, et al. Image reporting and characterization system for ultrasound features of thyroid nodules: Multicentric Korean retrospective study. Korean J Radiol. 2013;14:110–7.
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;17:370–95.
6. Russ G, Royer B, Bigorgne C, Rouxel A, Bienvenu-Perrard M, Leenhardt L. Prospective evaluation of thyroid imaging reporting and data system on 4550 nodules with and without elastography. Eur J Endocrinol. 2013;168:649–55.
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;6:225–37.
8. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 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.
9. Gharib H, Papini E, Garber JR, Duick DS, Harrell M, Hegedüs 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;22:622–39.
10. 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;14:587–95.
11. Cibas ES, Ali SZ. The Bethesda system for reporting thyroid cytopathology. Thyroid. 2009;19:1159–65.
12. Cibas ES, Ali SZ. The 2017 Bethesda system for reporting thyroid cytopathology. Thyroid. 2017;27:1341–6.
13. Feldt-Rasmussen U. Iodine and cancer. Thyroid. 2001;11:483–6.
14. Jeh SK, Jung SL, Kim BS, Lee YS. Evaluating the degree of conformity of papillary carcinoma and follicular carcinoma to the reported ultrasonographic findings of malignant thyroid tumor. Korean J Radiol. 2007;8:192–7.
15. Szybinski Z, Gołkowski F, Buziak-Bereza M, Trofimiuk M, Przybylik-Mazurek E, Huszno B, et al. Effectiveness of the iodine prophylaxis model adopted in Poland. J Endocrinol Investig. 2008;31:309–13.
16. Słowin´ska-Klencka D, Wysocka-Konieczna K, Woz´niak-Oseła E, Sporny S, Popowicz B, Sopin´ski J, et al. Thyroid nodules with Hürthle cells: The malignancy risk in relation to the FNA outcome category. J Endocrinol Investig. 2019;42:1319–27.
17. 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;10.
18. Castellana M, Piccardo A, Virili C, Scappaticcio L, Grani G, Durante C, et al. Can ultrasound systems for risk stratification of thyroid nodules identify follicular carcinoma? Cancer Cytopathol. 2020;128:250–9.
19. 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;105:1659–69.
20. Słowin´ska-Klencka D, Wojtaszek-Nowicka M, Sporny S, Woz´niak-Oseła E, Popowicz B, Klencki M. The predictive value of sonographic images of follicular lesions-a comparison with nodules unequivocal in FNA-single centre prospective study. BMC Endocr Disord. 2016;16:69.
21. Remonti L, Kramer CK, Leitao C, Pinto LCF, Gross JL. Thyroid ultrasound features and risk of carcinoma: A systematic review and meta-analysis of observational studies. Thyroid. 2015;25:538–50.
22. Sillery JC, Reading CC, Charboneau JW, Henrichsen TL, Hay ID, Mandrekar JN. Thyroid follicular carcinoma: Sonographic features of 50 cases. Am J Roentgenol. 2010;194:44–54.
23. Słowin´ska-Klencka D, Woz´niak E, Wojtaszek M, Popowicz B, Sporny S, et al. Low malignancy risk of thyroid follicular lesion of undetermined significance in patients from post-endemic areas. Eur J Endocrinol. 2013;168:621–30.
24. Park VY, Kim EK, Kwak JY, Yoon JH, Moon HJ. Malignancy risk and characteristics of thyroid nodules with two consecutive results of atypia of undetermined significance or follicular lesion of undetermined significance on cytology. Eur Radiol. 2015;25:2601–7.
25. Eisa N, Khan A, Akhter M, Fensterwald M, Saleem S, Fananapazir G, et al. Both ultrasound features and nuclear atypia are associated with malignancy in thyroid nodules with atypia of undetermined significance. Ann Surg Oncol. 2018;25:3913–8.
26. Yoon JH, Kwon HJ, Kim EK, Moon HJ, Kwak JY. Subcategorization of atypia of undetermined significance/follicular lesion of undetermined significance (AUS/FLUS): A study applying thyroid imaging reporting and data system (TIRADS). Clin Endocrinol. 2016;85:275–82.
27. Grani G, Lamartina L, Ascoli V, Bosco D, Nardi F, D'Ambrosio F, et al. Ultrasonography scoring systems can rule out malignancy in cytologically indeterminate thyroid nodules. Endocrine. 2016;57:256–61.
28. Tang AL, Falciglia M, Yang H, Mark JR, Steward DL. Validation of American thyroid association ultrasound risk assessment of thyroid nodules selected for ultrasound fine-needle aspiration. Thyroid. 2017;27:1077–82.
29. Kamaya A, Lewis GH, Liu Y, Akatsu H, Kong C, Desser TS. Atypia of undetermined significance and follicular lesions of undetermined significance: Sonographic assessment for prediction of the final diagnosis. J Ultrasound Med. 2015;34:767–74.
30. Lee JH, Han K, Kim EK, Moon HJ, Yoon JH, Park VY, et al. Risk stratification of thyroid nodules with atypia of undetermined significance/follicular lesion of undetermined significance (AUS/FLUS) cytology using ultrasonography patterns defined by the 2015 ATA guidelines. Ann Otol Rhinol Laryngol. 2017;126:625–33.
31. Hong MJ, Na DG, Baek JH, Sung JY, Kim JH. Cytology-ultrasonography risk-stratification scoring system based on fine-needle aspiration cytology and the Korean-thyroid imaging reporting and data system. Thyroid. 2017;27:953–9.

32. Valderrabano P, Mcgettigan MJ, Lam CA, Khazai L, Thompson ZJ, Chung CH, et al. Thyroid nodules with indeterminate cytology: Utility of the American thyroid association sonographic patterns for cancer risk stratification. Thyroid. 2018;28:1004–12.
33. Ahmadi S, Herbst R, Oyekunle T, Jiang XS, Strickland K, Roman S, et al. Using the ATA and ACR TI-RADS sonographic classifications as adjunctive predictors of malignancy for indeterminate thyroid nodules. Endocr Pract. 2019;25:908–17.
34. Barbosa TLM, Junior COM, Graf H, Cavalvanti T, Trippia MA, Ugino RTDS, et al. ACR TI-RADS and ATA US scores are helpful for the management of thyroid nodules with indeterminate cytology. BMC Endocr Disord. 2019;19:112.
35. Yang W, Fananapazir G, Laroy J, Wilson M, Campbell MJ. Can the American thyroid association, K-TIRADS and ACR-TIRADS ultrasound classification systems be used to predict malignancy in bethesda category IV nodules? Endocr Pract. 2020;10:4158.
36. Chaigneau E, Russ G, Royer B, Bigorgne C, Bienvenu-Perrard M, Rouxel A, et al. TIRADS score is of limited clinical value for risk stratification of indeterminate cytological results. Eur J Endocrinol. 2018;179:13–20.
37. 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: Towards the “right” TIRADS. J Clin Endocrinol Metab. 2018;104.
38. Gao L, Xi X, Jiang Y, Yang X, Wang Y, Zhu S, et al. Comparison among TIRADS (ACR TI-RADS and KWAK-TI-RADS) and 2015 ATA guidelines in the diagnostic efficiency of thyroid nodules. Endocrine. 2019;64:90–6.
39. Shen Y, Liu M, He J, Wu S, Chen M, et al. Comparison of different risk-stratification systems for the diagnosis of benign and malignant thyroid nodules. Front Oncol. 2019;9:378.
40. 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 congress of radiology guidelines. Thyroid. 2018;28:1532–7.
41. Trimboli P, Ngu R, Royer B, Giovanella L, Bigorgne C, et al. A multicentre validation study for the EU-TIRADS using histological diagnosis as a gold standard. Clin Endocrinol. 2019;91:340–7.
42. Dobruch-Sobczak K, Adamczewski Z, Szczepanek-Parulska E, Migda B, Wolin´ski K, Krauze A, et al. Histopathological verification of the diagnostic performance of the EU-TIRADS classification of thyroid nodules—results of a multicenter study performed in a previously iodine-deficient region. J Clin Med. 2019;8:1781.
43. Xu T, Wu Y, Wu RX, Zhang YZ, Gu JY, et al. Validation and comparison of three newly-released thyroid imaging reporting and data systems for cancer risk determination. Endocrine. 2018;64:299–307.
44. Pantano AL, Maddaloni E, Briganti SI, Anguissola GB, Perrella E, et al. Differences between ATA, AACE/ACE/AME and ACR TI-RADS ultrasound classifications performance in identifying cytological high-risk thyroid nodules. Eur J Endocrinol. 2018;178:595–603.
45. Trimboli P, Fulciniti F, Zilioli V, Ceriani L, Giovanella L. Accuracy of international ultrasound risk stratification systems in thyroid lesions cytologically classified as indeterminate. Diagn Cytopathol. 2016;45:113–7.
46. Persichetti A, Di Stasio E, Guglielmi R, Bizzarri G, Taccogna S, et al. Predictive value of malignancy of thyroid nodule ultrasound classification systems: A prospective study. J Clin Endocrinol Metab. 2018;103:1359–68.
47. Rosario PW, Da Silva AL, Nunes MS, Borges MAR, Mourão GF, et al. Risk of malignancy in 1502 solid thyroid nodules >1 cm using the new ultrasonographic classification of the American thyroid association. Endocrine. 2017;56:442–5.
48. Gao L, Xi X, Wang J, Yang X, Wang Y, Zhu S, et al. Ultrasound risk evaluation of thyroid nodules that are “unspecified” in the 2015 American thyroid association management guidelines. Medicine. 2018;97:e13914.