

Reducing the Number of Unnecessary Thyroid Biopsies While Improving Diagnostic Accuracy: Toward the “Right” TIRADS

Giorgio Grani,¹ Livia Lamartina,¹ Valeria Ascoli,² Daniela Bosco,² Marco Biffoni,³ Laura Giacomelli,³ Marianna Maranghi,¹ Rosa Falcone,¹ Valeria Ramundo,¹ Vito Cantisani,⁴ Sebastiano Filetti,¹ and Cosimo Durante¹

¹Dipartimento di Medicina Interna e Specialità Mediche, Università di Roma “Sapienza,” 00161 Rome, Italy;

²Dipartimento di Scienze Radiologiche, Oncologiche ed Anatomico Patologiche, Università di Roma “Sapienza,” 00161 Rome, Italy; ³Dipartimento di Scienze Chirurgiche, Università di Roma “Sapienza,” 00161 Rome, Italy; and ⁴UOS Innovazioni Diagnostiche e Ultrasonografiche, Azienda Ospedaliera Universitaria Policlinico Umberto I, Università di Roma “Sapienza,” 00161 Rome, Italy

ORCID numbers: 0000-0002-1791-5915 (C. Durante).

Context: The prevalence of thyroid nodules in the general population is increasingly high, and at least half of those biopsied prove to be benign. Sonographic risk-stratification systems are being proposed as “rule-out” tests that can identify nodules that do not require fine-needle aspiration (FNA) cytology.

Objective: To comparatively assess the performances of five internationally endorsed sonographic classification systems [those of the American Thyroid Association, the American Association of Clinical Endocrinologists, the American College of Radiology (ACR), the European Thyroid Association, and the Korean Society of Thyroid Radiology] in identifying nodules whose FNAs can be safely deferred and to estimate their negative predictive values (NPVs).

Design: Prospective study of thyroid nodules referred for FNA.

Setting: Single academic referral center.

Patients: Four hundred seventy-seven patients (358 females, 75.2%); mean (SD) age, 55.9 (13.9) years.

Main Outcome Measures: Number of biopsies classified as unnecessary, false-negative rate (FNR), sensitivity, specificity, predictive values, and diagnostic ORs for each system.

Results: Application of the systems’ FNA criteria would have reduced the number of biopsies performed by 17.1% to 53.4%. The ACR Thyroid Imaging Reporting and Data System (TIRADS) allowed the largest reduction (268 of 502) with the lowest FNR (NPV, 97.8%; 95% CI, 95.2% to 99.2%). Except for the Korean Society of Thyroid Radiology TIRADS, all other systems exhibited significant discriminatory performance but produced significantly smaller reductions in the number of procedures.

Conclusions: Internationally endorsed sonographic risk stratification systems vary widely in their ability to reduce the number of unnecessary thyroid nodule FNAs. The ACR TIRADS outperformed the others, classifying more than half the biopsies as unnecessary with a FNR of 2.2%. (*J Clin Endocrinol Metab* 104: 95–102, 2019)

ISSN Print 0021-972X ISSN Online 1945-7197

Printed in USA

Copyright © 2019 Endocrine Society

Received 4 August 2018. Accepted 3 October 2018.

First Published Online 8 October 2018

Abbreviations: AACE, American Association of Clinical Endocrinologists; ACE, American College of Endocrinology; ACR, American College of Radiology; AME, Associazione Medici Endocrinologi; ATA, American Thyroid Association; AUROC, area under the receiver operating characteristic; DOR, diagnostic OR; EU-TIRADS, European TIRADS; FNA, fine-needle aspiration; FNR, false-negative rate; K-TIRADS, Korean Society of Thyroid Radiology TIRADS; NPV, negative predictive value; TIRADS, Thyroid Imaging Reporting and Data System.

The number of individuals harboring sonographically detected thyroid nodules continues to rise, with an estimated 219 million in the United States alone. The challenge for clinicians is to identify those rare nodules harboring a clinically relevant malignancy (1). Fine-needle aspiration (FNA) has traditionally been used for this purpose (2–4). However, at least half of all biopsied nodules prove to be benign (5), and up to one third have cytological findings that are inconclusive (6). Strategies for the clinical management of thyroid nodule patients have therefore evolved (6): ultrasonography now plays a larger role (1), informing decisions on the need for FNA (7, 8) and plans for long-term follow-up (7, 8). To improve the accuracy of this guidance, ultrasound-based risk-stratification systems have now been developed by many national and international thyroid societies (7–10) and by the American College of Radiology (ACR) (11).

Robust evidence is lacking on the relative strengths and weaknesses of the various systems. Some systems have been validated in multicenter studies (12, 13). Independent validation and comparison studies (generally involving two to three of the systems) have mostly been retrospective (14–19). In a recent study involving central re-evaluation of nodules in a multi-institutional database, the 2017 ACR Thyroid Imaging Reporting and Data System (TIRADS) was found to compare favorably with the American Thyroid Association (ATA) guidelines and the Korean Society of Thyroid Radiology TIRADS (K-TIRADS), primarily because it more effectively reduced the number of biopsies performed on benign nodules (20). The largest prospective study of this type (21) compared the systems developed by the British Thyroid Association, the ATA, and the American Association of Clinical Endocrinologists (AACE)/American College of Endocrinology (ACE)/Associazione Medici Endocrinologi (AME) and found no significant differences between their overall diagnostic accuracy parameters.

To obtain a broader picture of the merits and demerits of currently available sonographic risk-stratification systems, we conducted a prospective, observational study of 502 thyroid nodules referred to our center for FNA. During the real-time pre-FNA ultrasound examinations, each nodule was classified using five internationally endorsed systems, and the recommendation for FNA was analyzed in light of the nodule's pathologic diagnosis. Because a major aim of all of these systems is to eliminate unnecessary thyroid biopsies without jeopardizing the detection of clinically significant malignancies, our aims were to determine (1) the proportion of nodules whose biopsy would have been considered unnecessary by each system, and (2) the reliability of these exclusions, as reflected by multiple parameters of diagnostic accuracy.

Materials and Methods

Cases

The study was conducted in the Thyroid Cancer Unit of a large academic referral center. All patients consecutively referred to the unit for FNA cytology of a thyroid nodule between 1 November 2015 and 30 May 2018 were eligible for enrollment. The referring physicians included primary care physicians and secondary health care providers (*e.g.*, endocrinologists, surgeons, otolaryngologists, nuclear medicine specialists). The study was conducted with Institutional Review Board approval and written informed patient consent.

Pre-FNA ultrasound examination of the nodules

Prior to each biopsy, each nodule was carefully examined with a HI VISION Avius[®] ultrasound system (Hitachi Medical Corporation, Tokyo, Japan) and a 13-MHz linear-array transducer. During this examination, two clinicians experienced in thyroid sonography (G.G. and L.L.) recorded their consensus judgment on the sonographic features of each nodule on a standardized rating form (22), internally developed and based on published recommendations (23, 24). Judgments were made jointly to eliminate the problem of interobserver variability, which has been documented during assessments of the single sonographic features of thyroid nodules (25, 26). The nodule features recorded by the readers were: diameters (anteroposterior, transverse, and longitudinal); margins (peripheral halo, well defined, ill defined, microlobulated, or irregular); structure/composition (solid, cystic, or mixed); echogenicity (hyperechoic, isoechoic, hypoechoic—all relative to the perinodular parenchyma—or markedly hypoechoic, *i.e.*, less echoic than the adjacent strap muscle); calcifications (absent, microscopic, or macroscopic, with the latter including eggshell calcifications); other hyperechoic foci (comet-tail artifacts or indeterminate, with the latter including areas of fibrosis); and suspected extrathyroidal extension (loss of the echogenic thyroid border, abutment, or contour bulging) (see Table 1). For mixed-content nodules, the location of the solid component (nonnodular, eccentric, or central) was also rated. The shape was considered taller than wide when the anteroposterior diameter exceeded the transverse diameter.

Classification of nodules using five sonographic risk-stratification systems

For each nodule, the consensus ratings of each ultrasound feature were used to classify the risk of malignancy according to five widely used ultrasound risk-stratification systems—namely, those published by the AACE/ACE/AME (8), the ACR (ACR-TIRADS) (11), the ATA (7), the European Thyroid Association [European TIRADS (EU-TIRADS)] (9), and the K-TIRADS (10). With each system, the likelihood that a nodule is malignant is indicated by its risk class, which is in turn defined by a set of ultrasound features. Additionally, within each risk class, the advisability of FNA is indicated based on lesion size. For nodules in low-risk categories, the size thresholds for FNA range from 1.5 to 3.0 cm, depending on the system. For those in high-risk classes, FNA is usually indicated when the maximum diameter is ≥ 1 cm. None of the systems we tested routinely recommends the FNA of subcentimeter thyroid nodules, although some make exceptions in the presence of particular high-risk clinical features: for this reason, the nodules with a

Table 1. Clinical and Sonographic Features of Included Nodules (n = 502)

		N (%)
Hashimoto's thyroiditis	No	470 (93.6)
	Yes	32 (6.4)
Multinodular gland	No	157 (31.3)
	Yes	345 (68.7)
Location	Isthmus	40 (8)
	Right lobe	209 (41.6)
	Left lobe	253 (50.4)
Single sonographic features		
Margins	Smooth	172 (34.3)
	Ill-defined	60 (12)
	Irregular/microlobulated	32 (6.4)
	Infiltrative	4 (0.8)
	Halo	234 (46.6)
Extrathyroidal extension	No	498 (99.2)
Structure/composition	Suspicious	4 (0.8)
	Solid	163 (32.5)
	Cystic	6 (1.2)
	Mixed	333 (66.3)
Echogenicity	Including spongiform	13 (2.6)
	Anechoic	6 (1.2)
	Hyperechoic	2 (0.4)
	Isoechoic	357 (71.1)
	Hypoechoic	120 (24)
	Markedly hypoechoic	17 (3.4)
Foci	Comet tail	25 (5)
	Indeterminate	94 (18.7)
Calcifications	Macro	76 (15.1)
	Micro	37 (7.4)
Lymph nodes	Suspicious	6 (1.2)
Shape	Taller than wide	84 (16.7)
Sonographic classification systems		
ACR TIRADS	TR1: Benign	16 (3.2)
	TR2: Not suspicious	134 (26.7)
	TR3: Mildly suspicious	99 (19.7)
	TR4: Moderately suspicious	181 (36.1)
	TR5: Highly suspicious	72 (14.3)
AACE/ACE/AME	Low	24 (4.8)
	Intermediate	320 (63.7)
	High	158 (31.5)
ATA	Benign	6 (1.2)
	Very low suspicion	249 (49.6)
	Low suspicion	69 (13.7)
	Intermediate suspicion	28 (5.6)
	High suspicion	60 (12)
	Not classifiable	90 (17.9)
EU-TIRADS	Benign	4 (0.8)
	Low risk	270 (53.8)
	Intermediate risk	73 (14.5)
	High risk	155 (30.9)
K-TIRADS	Benign	9 (1.8)
	Low suspicion	316 (62.9)
	Intermediate suspicion	139 (27.7)
	High suspicion	38 (7.6)

maximum diameter <1 cm were excluded. Using each system, we identified the nodules for which FNA was suggested based on the size threshold for the assigned risk class.

Reference standard diagnosis

The biopsies were done under ultrasound guidance by clinicians (endocrinologists trained in thyroid sonography) (G.G. and L.L.) using 23- to 25-gauge needles. The nonaspiration

technique was used in most cases (one to four needle passes). Direct smears of each specimen were analyzed by experienced thyroid cytopathologists (V.A. and D.B.) and classified according to the criteria published in the Italian Consensus for Thyroid Cytopathology (27, 28). When surgery had been performed, the reference standard diagnosis (malignant vs benign) was based on histological examination of the resected nodule. When the nodule had been managed nonsurgically, the reference standard was FNA cytology: nodules were considered malignant when they had been classified as TIR4 or TIR5 [suspected malignancy or malignancy, corresponding to Bethesda classes V and VI (29)], and benign when they had been classified as TIR2, corresponding to Bethesda class II. Patients with cytologically benign nodules were advised to undergo repeat sonography in 12 to 36 months, depending on their baseline risk status, as suggested by the ATA guidelines (7). Nodules without histologic diagnoses that had been cytologically classified as TIR1 (nondiagnostic), TIR3A, or TIR3B (indeterminate) (similar to Bethesda classes I, III, and IV) were excluded from the final analysis, unless a repeat FNA had yielded conclusive results.

Statistical analysis

For each sonographic classification system, we calculated the number of nodules that did and did not meet the criteria for FNA (test positivity and test negativity, respectively). For the purposes of the study, the biopsies ordered for test-negative nodules were considered “unnecessary.” The sonographic recommendation regarding FNA was then compared with the reference-standard diagnosis (benign vs malignant) to estimate its sensitivity, specificity, positive predictive value, negative predictive values (NPVs), diagnostic OR (DOR), and area under the receiver operating characteristic (AUROC) curve (each with 95% CI). The proportions of biopsies that would have been considered unnecessary by the various systems were compared using the McNemar test, and the reliability of these indications (*i.e.*, whether the recommended deferral involved a nodule that was indeed benign) was assessed by calculation of the NPV and false-negative rate (FNR). Data were analyzed with the IBM SPSS Statistics package, version 25.0 (IBM Corp., Armonk, NY) and Microsoft Office Suite.

Results

As shown in Fig. 1, a total of 832 thyroid nodules were subjected to sonographic risk stratification prior to FNA. The ultrasound examination identified 79 subcentimeter nodules, which were excluded from our analysis because, in the absence of particular clinical features, FNA is not indicated by any of the five systems for nodules of this size. Of the 753 nodules measuring ≥ 1 cm, 251 (33.3%) were also excluded from the analysis because their reference standard diagnosis was inconclusive. The final cohort thus comprised 502 thyroid nodules identified in 477 patients (mean age, 55.9 ± 13.9 years; female/male ratio, 3.0) (Fig. 1). Thirty-six (7.2%) lesions met the reference standard criteria for malignancy. In 34 cases, the diagnosis was based on histologic findings: 27 papillary thyroid cancers, 1 follicular thyroid cancer, 2

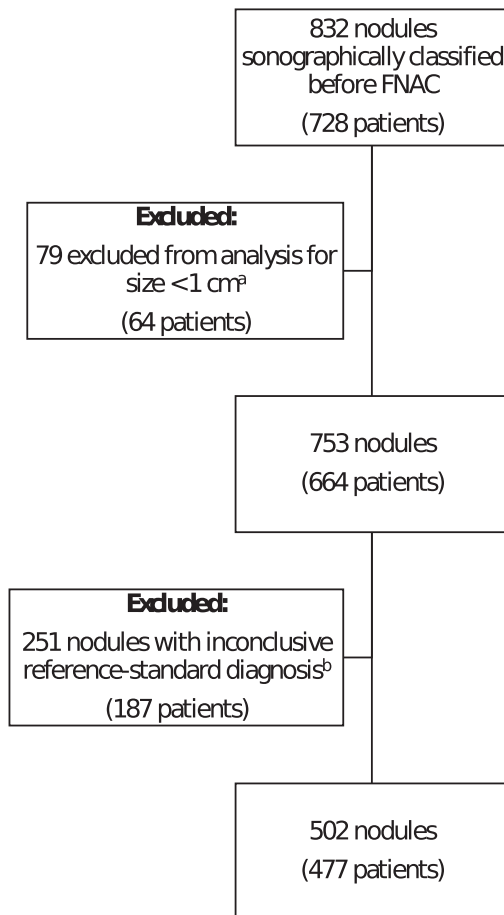


Figure 1. Diagram of the thyroid nodules included in the main analysis. ^aExcluded because fine-needle aspiration cytology (FNAC) is rarely indicated by any of the TIRADS approaches for subcentimeter nodules. ^bInconclusive reference standard diagnoses: nondiagnostic or indeterminate cytology without definitive histological diagnosis. Based on sonographic screening results, FNAC would not have been recommended for many of these 251 nodules: 126 (50.2%) based on ACR TIRADS, 98 (39%) based on ATA criteria, 77 (30.7%) based on AACE criteria, 67 (26.7%) based on EU-TIRADS criteria, and 34 (13.5%) based on K-TIRADS recommendations. Sixteen nodules whose initial FNA yielded nondiagnostic ($n = 8$) or indeterminate ($n = 8$) cytological findings were subsequently reclassified as benign based on a repeat biopsy and included in the final analysis.

medullary thyroid cancers, 1 anaplastic thyroid cancer, and 3 thyroid metastases from other malignancies (30). The remaining two were classified cytologically as TIR4/Bethesda V and managed (in accordance with patient preferences) with active surveillance alone.

Strict application of the indications furnished by the each of the five ultrasound systems would have appreciably reduced the number of FNAs performed (Table 2). The percentages of nodules identified as FNA deferrable varied widely (from 17.1% to 53.4%), and with each system, some of the exclusions were false negatives with reference standard diagnoses of malignancy (FNRs, 2.2% to 4.1%). The most effective system in our cohort was the ACR TIRADS, which would have eliminated

more than half of the biopsies ordered (268, 53.4%), with a FNR of only 2.2% (NPV, 97.8%; 95% CI, 95.2% to 99.2%). With the exception of the K-TIRADS, the other systems' discriminatory capacities (as reflected by their DOR and AUROC) were similar to that of the ACR TIRADS, but their impact on the number of procedures performed was significantly smaller (Table 2). Eleven nodules definitively diagnosed as malignant would have been misclassified as not requiring FNA by at least one of the TIRADS systems (Table 3). The three cancers missed by all five systems were either isoechoic or hyperechoic relative to the surrounding parenchyma and had no other features considered anomalous by any of the systems. (These were the only malignancies misclassified by the K-TIRADS system.) The ATA system also failed to identify six other malignant nodules, five of which emerged as false negatives with this system alone. These five nodules were isoechoic, but they also had irregular margins, which are considered suspicious by the other systems. Unfortunately, this "set of features" did not allow them to be allocated into any of the ATA system's risk classes.

Discussion

Thyroid nodule FNAs play key roles in ruling out the presence of thyroid cancer. However, the costs of sample collection and analysis are relatively high, and the aspiration itself can be a source of discomfort and anxiety for patients. Furthermore, in roughly one third of cases, cytomorphologic analysis of the aspirate yields inconclusive results (5) that prompt repeat biopsies or additional, more expensive testing (1). The sonographic risk-stratification tools we assessed are basically "rule-out" tests, designed to identify nodules with low risks of malignancy whose cytologic assessment can safely be deferred. Our goal was to compare the performances of five widely used systems of this type in achieving this goal. Each system assigns differential weights to the individual sonographic features evaluated to establish a nodule's risk of malignancy, and the weight assigned to a given feature varies substantially from one system to another. The systems also differ from one another in terms of the size thresholds for identifying nodules within a given risk class that require FNA.

In our cohort, the number of biopsies performed would have been reduced to some extent if the decision had been based on strict application of any of the five internationally endorsed systems tested. However, the safest and most substantial reduction would have been achieved with the ACR TIRADS: it classified more than half of the biopsies ordered as unnecessary (268 of 502, 53.4%) and had the lowest FNR of all systems tested

Table 2. Ability of the Five Sonographic Classification Systems to Identify Nodules Whose FNA Could Be Deferred and to Discriminate Between Benign and Malignant Nodules

	Avoided Biopsies (%) ^a	FN (FNR)	TN (TNR)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	AUC	DOR ^b
ACR TIRADS	268 (53.4)	6/268 (2.2%)	262/268 (97.8%)	83.3 (67.2–93.6)	56.2 (51.6–60.8)	12.8 (8.8–17.8)	97.8 (95.2–99.2)	0.7 (0.62–0.78)	6.42 (2.62–15.72)
ATA	220 (43.8)	9/220 (4.1%)	211/220 (95.9%)	75 (57.8–87.9)	45.3 (40.7–49.9)	9.6 (6.4–13.6)	95.9 (92.4–98.1)	0.6 (0.51–0.69)	2.48 (1.14–5.39)
AACE/ACE/AME	175 (34.9)	5/175 (2.9%)	170/175 (97.1%)	86.1 (70.5–95.3)	36.5 (32.1–41.0)	9.5 (6.5–13.2)	97.1 (93.5–99.1)	0.61 (0.53–0.7)	3.56 (1.36–9.33)
EU-TIRADS	154 (30.7)	5/154 (3.2%)	149/154 (96.8%)	86.1 (70.5–95.3)	32 (27.8–36.4)	8.9 (6.1–12.4)	96.7 (92.6–98.9)	0.59 (0.5–0.68)	2.91 (1.11–7.64)
K-TIRADS	86 (17.1)	3/86 (3.5%)	83/86 (96.5%)	91.7 (77.5–98.2)	17.8 (14.4–21.6)	7.9 (5.5–11)	96.5 (90.2–99.3)	0.55 (0.46–0.64)	2.38 (0.71–7.96)

Abbreviations: AUC, area under the receiver operating characteristic curve; FN, false negative; PPV, positive predictive value.

^aThe rate of avoided biopsies is significantly different between the US classification systems (McNemar test: ACR TIRADS vs ATA, $P = 0.002$; ATA vs AACE/ACE/AME, $P < 0.001$; AACE/ACE/AME vs EU-TIRADS, $P < 0.001$; EU-TIRADS vs K-TIRADS, $P < 0.001$).

^bThe DOR measures the discriminatory power of a diagnostic test as compared with that of the reference standard. The value ranges from 0 to infinity, with higher values indicating better performance.

(6 of 268, 2.2%). Its abilities to exclude malignancy and to discriminate between benign and malignant nodules were substantially greater than those of its competitors (NPV, 97.8%; DOR, 6.42; 95% CI, 2.62 to 15.72). The high number of planned biopsies identified by this system as unnecessary reflects the higher size thresholds it sets for recommending biopsy of nodules classified as low risk (Table 4).

Importantly, the ACR TIRADS assigned all the thyroid nodules in the cohort to a risk class—a clear advantage over the widely used ATA system, which failed to classify a significant number of the nodules we studied (90, 17.9%) and of those analyzed by others (19, 31). In our hands, the K-TIRADS performance was disappointing: the number of biopsies it would have eliminated was quite modest (86 of 502, 17.1%), and its discriminatory capacity was not statistically significant, as reflected by the lower end of the confidence interval of the AUROC (<0.5)

and the DOR (<1). One small nodule classified as FNA deferrable by three of the five systems (the AACE/ACE/AME guidelines, the EU-TIRADS, and the ACR TIRADS) proved to be a medullary thyroid cancer. The sonographic features of these cancers are known to differ significantly from those of papillary thyroid cancer (32), which are the basis of sonographic risk stratification. One third of all medullary thyroid cancer have sonographic features (solid, round/ovoid shapes, smooth margins, mild hypoechogenicity) (33) considered “low suspicion” in many systems: these features are sufficient to classify the nodule at least in the ATA intermediate suspicion pattern, requiring FNA when it measured >1 cm (34).

Importantly, however, note that the choice of a TIRADS system cannot be based solely on the number of biopsies it flags as unnecessary and its diagnostic accuracy. Interobserver variability (25) and consistency over time (35) are also important considerations, as are the

Table 3. Ultrasound Features of the Malignancies Missed by One or More Sonographic Classification Systems

Pathological Diagnosis	US Description	Maximum Diameter (mm)	Missed by				
			ATA ^a	K-TIRADS	AACE	EU-TIRADS	ACR TIRADS
cPTC	Hyperechoic (hypoechoic parenchyma due to thyroiditis)	11.0	X	X	X	X	X
cPTC	Solid, mildly hypoechoic	13.0			X	X	X
MTC	Solid, mildly hypoechoic	12.1			X	X	X
cPTC	Solid, isoechoic, irregular margins	10.5	X				
cPTC	Solid, isoechoic, irregular margins	11.1	X				X
cPTC	Solid, isoechoic, irregular margins	20.3	X				
fvPTC	Solid, isoechoic, irregular margins	11.9	X				
fvPTC	Mixed, isoechoic nodule	13.6	X	X	X	X	X
Suspected PTC ^b	Mixed, isoechoic nodule	11.4	X	X	X	X	X
cPTC	Mixed, isoechoic nodule, irregular margins	15.0	X				
cPTC	Mixed, isoechoic nodule, irregular margins	17.2	X				

Abbreviations: cPTC, classic papillary thyroid cancer; fvPTC, follicular-variant papillary thyroid cancer; MTC, medullary thyroid cancer; PTC, papillary thyroid cancer.

^aIf irregular margins (or other worrisome sonographic features), even in the context of isoechoic nodules, were considered high risk, the number of malignancy missed by the ATA system would decrease.

^bClassified cytologically as TIR4/Bethesda V and managed (in accordance with patient preferences) with active surveillance alone.

Table 4. Size Threshold for FNA Biopsy Proposed by the Five Sonographic Classification Systems

AACE/ACE/AME	ACR TIRADS	ATA	K-TIRADS	EU-TIRADS
High FNA >10 mm ROM 50%–90%	TR5 FNA \geq 10 mm ROM \geq 20%	High suspicion FNA \geq 10 mm ROM 70%–90%	TR5 FNA \geq 10 mm ROM >60%	TR5 FNA >10 mm ROM 26%–87%
Intermediate FNA >20 mm ROM 5%–15%	TR4 FNA \geq 15 mm ROM 5%–20%	Intermediate suspicion FNA \geq 10 mm ROM 10%–20%	TR4 FNA \geq 10 mm ROM 15%–50%	TR4 FNA >15 mm ROM 6%–17%
Low FNA >20 mm ROM \approx 1%	TR3 FNA \geq 25 mm ROM 5%	Low suspicion FNA \geq 15 mm ROM 5%–10%	TR3 FNA \geq 15 mm ROM 3%–15%	TR3 FNA > 20 mm ROM 2%–4%
	TR2 No FNA ROM 2%	Very low suspicion FNA \geq 20 mm ROM <3%	TR2 FNA \geq 20 mm ROM 1%–3%	TR2 No FNA ROM 0%
	TR1 No FNA ROM 2%	Benign No FNA ROM < 1%	—	—

Differences in the size threshold with respect to ACR TIRADS are highlighted in italic bold.

AACE/ACE/AME guidelines, including only three classes, are not directly comparable to the other four- or five-tiered systems.

Abbreviation: ROM, risk of malignancy.

setting in which it will be used (*e.g.*, equipment, operator experience). Additionally, the US features being evaluated must also be defined in a manner that is clear and unambiguous to the operators using the system, an outcome favored by specific training and experience (25). The system must also be easy to use in routine clinical practice. For example, the ACR TIRADS differs from the other four systems tested in that it is based on a point scale rather than on pattern recognition. Points are assigned for five individual ultrasound features, and their sum determines the nodule's risk class. This approach may be considered excessively time-consuming for use in daily practice.

Our study also has several limitations that must be considered when interpreting our findings. First, ours was a selected cohort of thyroid nodules, all of which had already been flagged for FNA by another physician (*e.g.*, endocrinologists, oncologists, general practitioners, clinicians from other fields, pathologists), and the criteria supporting these requests were not known. This cohort's 7.2% malignancy rate, however, was similar to those reported for unselected nodule series (6), and all sonographic risk classes were represented, including low-risk categories. Second, the composite reference standard used in our study is not error-free. For example, a benign cytology report was considered sufficient for classifying the nodule as benign. However, FNA cytology can yield false-negative results. Such outcomes are uncommon, with estimated frequencies of 3.7% emerging from a recent meta-analysis (5) and even lower frequencies (<1%) in prospective series of cytologically benign nodules with no high-suspicion ultrasound features (36). Second, as for our reference standard diagnosis of malignancy, the risk of error (false positivity) is limited to

two nodules whose classification was based on “suspicious” (TIR4, Bethesda class V) cytology alone: the other 34 reference standard diagnoses of malignancy were all histologically confirmed. It is also conceivable that our exclusion of 251 (33%) nodules with nondiagnostic or indeterminate cytology reports caused a selection bias; however, the proportion of nodules with these cytological reports is consistent with those reported in other cytological series (5). Additionally, as shown in Fig. 1, the rates of deferrable biopsies in this subgroup were similar to those in the study cohort itself.

The major strength of our study is its prospective nature: the US features of each nodule were evaluated during real-time examinations carried out before aspirates were collected. In this setting, each of the five internationally endorsed TIRADS approaches we evaluated identified multiple thyroid nodules for which the request for FNA was probably unnecessary. Four of the five (AACE/ACE/AME, ATA, ACR TIRADS, and EU-TIRADS) showed a significant diagnostic value. The best overall performance was that of the ACR TIRADS, which classified more than half of the requested biopsies as unnecessary, with a NPV of 97.8%.

Acknowledgments

Writing support was provided by Marian Everett Kent.

Financial Support: This work was supported by research starter grants of the University of Rome, Sapienza (AR216154C9B479B2) (to G.G.). G.G., L.L., R.F., and V.R. contributed to this paper as part of their PhD studies in Biotechnologies and Clinical Medicine at the University of Rome, Sapienza, which provided support for their studies. Writing support was funded by the Umberto Di Mario Foundation.

Correspondence and Reprint Requests: Cosimo Durante, MD, Dipartimento di Medicina Interna e Specialità Mediche,

Università di Roma “Sapienza,” Viale del Policlinico 155, 00161 Rome, Italy. E-mail: cosimo.durante@uniroma1.it.

Disclosure Summary: The authors have nothing to disclose.

References

- Lamartina L, Deandreis D, Durante C, Filetti S. Endocrine tumours: imaging in the follow-up of differentiated thyroid cancer: current evidence and future perspectives for a risk-adapted approach. *Eur J Endocrinol*. 2016;175(5):R185–R202.
- Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, Mandel SJ, Mazzaferri EL, McIver B, Pacini F, Schlumberger M, Sherman SI, Steward DL, Tuttle RM; American Thyroid Association (ATA) Guidelines Taskforce on Thyroid Nodules and Differentiated Thyroid Cancer. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer [published corrections appear in *Thyroid*. 2010;20(6):674–675 and *Thyroid*. 2010;20(8):942]. *Thyroid*. 2009;19(11):1167–1214.
- Gharib H, Papini E, Paschke R, Duick DS, Valcavi R, Hegedüs L, Vitti P; AACE/AME/ETA Task Force on Thyroid Nodules. American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and European Thyroid Association medical guidelines for clinical practice for the diagnosis and management of thyroid nodules. *Endocr Pract*. 2010;16(Suppl 1):1–43.
- Filetti S, Durante C, Torlontano M. Nonsurgical approaches to the management of thyroid nodules. *Nat Clin Pract Endocrinol Metab*. 2006;2(7):384–394.
- Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda System for Reporting Thyroid Cytopathology: a meta-analysis. *Acta Cytol*. 2012;56(4):333–339.
- Durante C, Grani G, Lamartina L, Filetti S, Mandel SJ, Cooper DS. The diagnosis and management of thyroid nodules: a review. *JAMA*. 2018;319(9):914–924.
- Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, Pacini F, Randolph GW, Sawka AM, Schlumberger M, Schuff KG, Sherman SI, Sosa JA, Steward DL, Tuttle RM, Wartofsky L. 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):1–133.
- Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedüs L, Paschke R, Valcavi R, Vitti P; AACE/ACE/AME Task Force on Thyroid Nodules. 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(5):622–639.
- 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(5):225–237.
- Shin JH, Baek JH, Chung J, Ha EJ, Kim JH, Lee YH, Lim HK, Moon WJ, Na DG, Park JS, Choi YJ, Hahn SY, Jeon SJ, Jung SL, Kim DW, Kim EK, Kwak JY, Lee CY, Lee HJ, Lee JH, Lee KH, Park SW, Sung JY; Korean Society of Thyroid Radiology (KSThR) and Korean Society of Radiology. Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean Society of Thyroid Radiology consensus statement and recommendations. *Korean J Radiol*. 2016;17(3):370–395.
- Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, Cronan JJ, Beland MD, Desser TS, Frates MC, Hammers LW, Hamper UM, Langer JE, Reading CC, Scoutt LM, Stavros AT. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. *J Am Coll Radiol*. 2017;14(5):587–595.
- Middleton WD, Teefey SA, Reading CC, Langer JE, Beland MD, Szabunio MM, Desser TS. 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(6):1331–1341.
- Ha EJ, Moon WJ, Na DG, Lee YH, Choi N, Kim SJ, Kim JK. A multicenter prospective validation study for the Korean Thyroid Imaging Reporting and Data System in patients with thyroid nodules. *Korean J Radiol*. 2016;17(5):811–821.
- Nam SJ, Kwak JY, Moon HJ, Yoon JH, Kim EK, Koo JS. Large (≥ 3 cm) thyroid nodules with benign cytology: can Thyroid Imaging Reporting and Data System (TIRADS) help predict false-negative cytology? *PLoS One*. 2017;12(10):e0186242.
- Mendes GF, Garcia MR, Falsarella PM, Rahal A, Cavalcante Junior FA, Nery DR, Garcia RG. Fine needle aspiration biopsy of thyroid nodule smaller than 1.0 cm: accuracy of TIRADS classification system in more than 1000 nodules. *Br J Radiol*. 2018;91(1083):20170642.
- Zheng Y, Xu S, Kang H, Zhan W. A single-center retrospective validation study of the American College of Radiology Thyroid Imaging Reporting and Data System. *Ultrasound Q*. 2018;34(2):77–83.
- Koseoglu Atilla FD, Ozgen Saydam B, Erarslan NA, Diniz Unlu AG, Yilmaz Yasar H, Ozer M, Akinci B. Does the ACR TI-RADS scoring allow us to safely avoid unnecessary thyroid biopsy? single center analysis in a large cohort. *Endocrine*. 2018;61(3):398–402.
- Hoang JK, Middleton WD, Farjat AE, Langer JE, Reading CC, Teefey SA, Abinanti N, Boschini FJ, Bronner AJ, Dahiya N, Hertzberg BS, Newman JR, Scanga D, Vogler RC, Tessler FN. Reduction in thyroid nodule biopsies and improved accuracy with American College of Radiology Thyroid Imaging Reporting and Data System. *Radiology*. 2018;287(1):185–193.
- Yoon JH, Lee HS, Kim EK, Moon HJ, Kwak JY. Malignancy risk stratification of thyroid nodules: comparison between the Thyroid Imaging Reporting and Data System and the 2014 American Thyroid Association management guidelines. *Radiology*. 2016;278(3):917–924.
- Middleton WD, Teefey SA, Reading CC, Langer JE, Beland MD, Szabunio MM, Desser TS. 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(5):1148–1154.
- Persichetti A, Di Stasio E, Guglielmi R, Bizzarri G, Taccogna S, Misischi I, Graziano F, Petrucci L, Bianchini A, Papini E. Predictive value of malignancy of thyroid nodule ultrasound classification systems: a prospective study. *J Clin Endocrinol Metab*. 2018;103(4):1359–1368.
- Grani G, Lamartina L, Ascoli V, Bosco D, Nardi F, D’Ambrosio F, Rubini A, Giacomelli L, Biffoni M, Filetti S, Durante C, Cantisani V. Ultrasonography scoring systems can rule out malignancy in cytologically indeterminate thyroid nodules. *Endocrine*. 2017;57(2):256–261.
- Grant EG, Tessler FN, Hoang JK, Langer JE, Beland MD, Berland LL, Cronan JJ, Desser TS, Frates MC, Hamper UM, Middleton WD, Reading CC, Scoutt LM, Stavros AT, Teefey SA. Thyroid ultrasound reporting lexicon: White Paper of the ACR Thyroid Imaging, Reporting and Data System (TIRADS) Committee. *J Am Coll Radiol*. 2015;12(12 Pt A):1272–1279.
- Su HK, Dos Reis LL, Lupo MA, Milas M, Orloff LA, Langer JE, Brett EM, Kazam E, Lee SL, Minkowitz G, Alpert EH, Dewey EH, Urken ML. Striving toward standardization of reporting of ultrasound features of thyroid nodules and lymph nodes: a multidisciplinary consensus statement. *Thyroid*. 2014;24(9):1341–1349.

25. Grani G, Lamartina L, Cantisani V, Maranghi M, Lucia P, Durante C. Interobserver agreement of various thyroid imaging reporting and data systems. *Endocr Connect*. 2018;7(1):1–7.
26. Grani G, D'Alessandri M, Carbotta G, Nesca A, Del Sordo M, Alessandrini S, Coccaro C, Rendina R, Bianchini M, Prinzi N, Fumarola A. Grey-scale analysis improves the ultrasonographic evaluation of thyroid nodules. *Medicine (Baltimore)*. 2015;94(27):e1129.
27. Nardi F, Basolo F, Crescenzi A, Fadda G, Frasoldati A, Orlandi F, Palombini L, Papini E, Zini M, Pontecorvi A, Vittori P. Italian consensus for the classification and reporting of thyroid cytology. *J Endocrinol Invest*. 2014;37(6):593–599.
28. Fadda G, Basolo F, Bondi A, Bussolati G, Crescenzi A, Nappi O, Nardi F, Papotti M, Taddei G, Palombini L; SIAPEC-IAP Italian Consensus Working Group; Proposal of the SIAPEC-IAP Italian Consensus Working Group. Cytological classification of thyroid nodules. *Pathologica*. 2010;102(5):405–408.
29. Cibas ES, Ali SZ. The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid*. 2017;27(11):1341–1346.
30. Falcone R, Ramundo V, Lamartina L, Ascoli V, Bosco D, Di Gioia C, Montesano T, Biffoni M, Bononi M, Giacomelli L, Minni A, Segni M, Maranghi M, Cantisani V, Durante C, Grani G. Sonographic presentation of metastases to the thyroid gland: a case series. *J Endocr Soc*. 2018;2(8):855–859.
31. Rosario PW, da Silva AL, Nunes MS, Ribeiro Borges MA, Mourão GF, Calsolari MR. Risk of malignancy in 1502 solid thyroid nodules >1 cm using the new ultrasonographic classification of the American Thyroid Association. *Endocrine*. 2017;56(2):442–445.
32. Liu MJ, Liu ZF, Hou YY, Men YM, Zhang YX, Gao LY, Liu H. Ultrasonographic characteristics of medullary thyroid carcinoma: a comparison with papillary thyroid carcinoma. *Oncotarget*. 2017;8(16):27520–27528.
33. Kim C, Baek JH, Ha E, Lee JH, Choi YJ, Song DE, Kim JK, Chung KW, Kim WB, Shong YK. Ultrasonography features of medullary thyroid cancer as predictors of its biological behavior. *Acta Radiol*. 2017;58(4):414–422.
34. Valderrabano P, Klippenstein DL, Tourtelot JB, Ma Z, Thompson ZJ, Lilienfeld HS, McIver B. New American Thyroid Association sonographic patterns for thyroid nodules perform well in medullary thyroid carcinoma: institutional experience, systematic review, and meta-analysis. *Thyroid*. 2016;26(8):1093–1100.
35. Grani G, Lamartina L, Biffoni M, Giacomelli L, Maranghi M, Falcone R, Ramundo V, Cantisani V, Filetti S, Durante C. Sonographically estimated risks of malignancy for thyroid nodules computed with five standard classification systems: changes over time and their relation to malignancy. *Thyroid*. 2018;28(9):1190–1197.
36. Durante C, Costante G, Lucisano G, Bruno R, Meringolo D, Paciaroni A, Puxeddu E, Torlontano M, Tumino S, Attard M, Lamartina L, Nicolucci A, Filetti S. The natural history of benign thyroid nodules. *JAMA*. 2015;313(9):926–935.