Update on ACR TI-RADS: Successes, Challenges, and Future Directions, From the *AJR* Special Series on Radiology Reporting and Data Systems

Jenny K. Hoang, MBBS, MHS, MBA[1], William D. Middleton, MD[2], Franklin N. Tessler, MD, CM[3]

Neuroradiology/Head and Neck Imaging · Review

Keywords

thyroid, thyroid cancer, thyroid nodule, TI-RADS, ultrasound

Submitted: Sep 14, 2020
Revision requested: Sep 25, 2020
Revision received: Oct 12, 2020
Accepted: Oct 17, 2020
First published online: Oct 28, 2020

The authors declare that they have no disclosures relevant to the subject matter of this article.

The American College of Radiology (ACR) Thyroid Imaging Reporting and Data System (TI-RADS) is an ultrasound-based risk stratification system (RSS) for thyroid nodules that was released in 2017. Since publication, research has shown that ACR TI-RADS has a higher specificity than other RSSs and reduces the number of unnecessary biopsies of benign nodules compared with other systems by 19.9–46.5%. The risk of missing significant cancers using ACR TI-RADS is mitigated by the follow-up recommendations for nodules that do not meet criteria for biopsy. In practice, after a nodule's ultrasound features have been enumerated, the ACR TI-RADS points-based approach leads to clear management recommendations. Practices seeking to implement ACR TI-RADS must engage their radiologists in understanding how the system addresses the problems of thyroid cancer overdiagnosis and unnecessary surgeries by reducing unnecessary biopsies. This review compares ACR TI-RADS to other RSSs and explores key clinical questions faced by practices considering its implementation. We also address the challenge of reducing interobserver variability in assigning ultrasound features. Finally, we highlight emerging imaging techniques and recognize the ongoing international effort to develop a system that harmonizes multiple RSSs, including ACR TI-RADS.

The American College of Radiology (ACR) Thyroid Imaging Reporting and Data System (TI-RADS) is an ultrasound-based risk stratification system (RSS) for thyroid nodules that was published by the ACR TI-RADS Committee in 2017 [1]. Unlike other RSSs, ACR TI-RADS uses a points-based system instead of a pattern-based or intermediate approach. This system ensures that all thyroid nodules can be classified, an outcome that is not possible with other RSSs, such as the American Thyroid Association (ATA) guidelines, which do not categorize some nodules. ACR TI-RADS also uses higher size thresholds for fine-needle aspiration (FNA), which reduces biopsy recommendation rates by 19.9–46.5% [2]. This critical advantage addresses the significant morbidity and cost associated with workup of thyroid nodules that are either benign or unlikely to cause harm during the patient's lifetime even if malignant, termed "overdiagnosis" [3–8]. Moreover, specific follow-up recommendations for nodules smaller that the biopsy size thresholds in ACR TI-RADS lessen the probability that cancers will be missed in the long term [9].

Since ACR TI-RADS was published, articles have provided guidance on its application [10, 11]. This review expands on such articles by summarizing recent evidence on how ACR TI-RADS compares to other RSSs and by exploring key clinical questions faced by practices considering its implementation. We also address the challenge of reducing interobserver variability and highlight future directions for ultrasound-based RSSs for thyroid nodules.

ACR TI-RADS Basics and Synopsis

ACR TI-RADS seeks to increase the consistency with which thyroid nodules are evaluated and classified with ultrasound. To do so, the system provides specific recommendations for management of every nodule, which may otherwise be absent from ultrasound reports. In ACR TI-RADS, nodules are evaluated in five sonographic feature categories: composition, echogenicity, shape, margin, and echogenic foci. The radiologist assigns one ultrasound finding from each of the first four categories and all features from the fifth category [12]. Features are associated with 0, 1, 2, or 3 points that are related to their malignancy risk.

doi.org/10.2214/AJR.20.24608
AJR 2021; 216:570–578
ISSN-L 0361–803X/21/2163–570

[1]Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, 601 N Caroline St, Baltimore, MD 21231. Address correspondence to J. K. Hoang (jennykh@gmail.com).

[2]Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, MO.

[3]Department of Radiology, University of Alabama at Birmingham, Birmingham, AL.

The points are summed to place the nodule in one of five ascending risk levels: TR1 to TR5. The higher risk levels have lower size thresholds for FNA and follow-up with ultrasound (Fig. 1). Follow-up intervals for nonbiopsied nodules are smaller for nodules at higher TR levels. Followed nodules that grow to exceed the biopsy cutoff for their TR level are recommended for FNA. Also, nodules whose TR level increases over time are followed more closely.

Thyroid nodules are exceedingly common, being found in as many as 68% of high-resolution ultrasound scans [13]. To avoid the need to formally classify every nodule, ACR TI-RADS advises reporting no more than four nodules with the highest scores that warrant follow-up. Additionally, no more than two nodules with the highest scores that meet criteria for FNA should be recommended for biopsy. Again, the goal is to foster greater consistency in thyroid nodule reporting and management.

HIGHLIGHTS

- *Unlike other risk stratification systems (RSSs), ACR TI-RADS provides unambiguous management recommendations (biopsy, follow-up, no action) for all thyroid nodules, to include in all reports.*
- *ACR TI-RADS reduces the number of unnecessary biopsies of benign nodules by 19.9% to 46.5% compared with other RSSs.*
- *Variability in management recommendations can persist after adoption of ACR TI-RADS, primarily due to interobserver variability in assigning sonographic signs such as punctate echogenic foci.*

Do We Need a Risk Stratification System?

Ultrasound has long been recognized as the most effective method to detect and characterize thyroid nodules [13]. Over the past 2 decades, professional organizations and other groups have developed a multitude of independent RSSs that often resulted in different management recommendations for a given nodule, leading to confusion for practitioners and patients alike

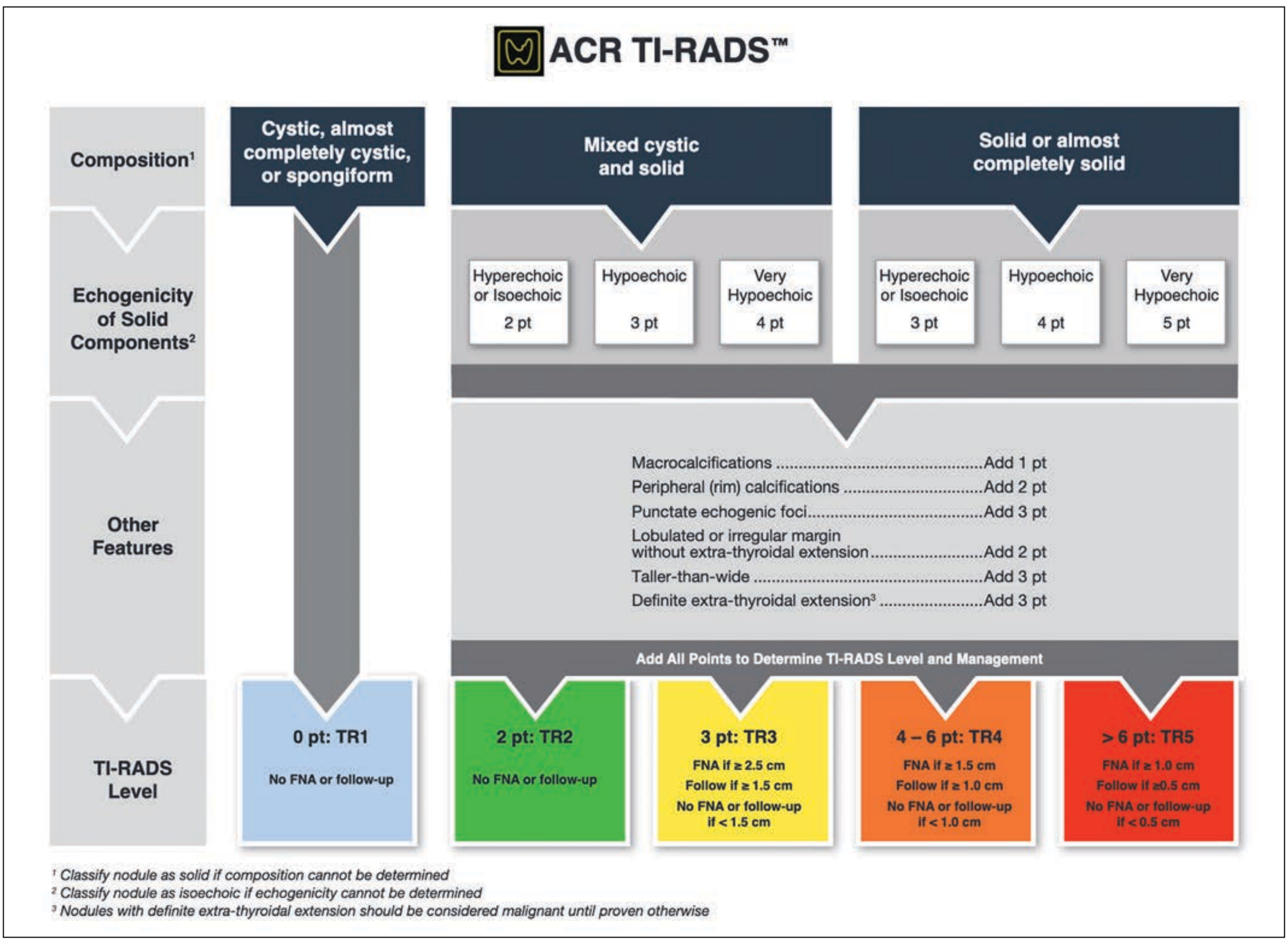

Fig. 1—American College of Radiology (ACR) TI-RADS Assessment Categories flowchart. FNA = fine-needle aspiration. This material is reprinted without modification with permission from American College of Radiology (© American College of Radiology; www.acr.org/-/media/ACR/Files/RADS/TIRADS/TIRADS-Alternative-chart.pdf) and pursuant to Creative Commons BY-NC-ND license and terms contained therein (creativecommons.org/licenses/by-nc-nd/4.0/), including disclaimer in Section 5.

[14]. Little to no consensus has been reached regarding which of the available RSSs to adopt. Some radiologists do not use any RSS at all, possibly because of perceived lack of importance of thyroid nodules [15].

Griffin et al. [16] performed a study on a two-phase introduction of ACR TI-RADS for thyroid ultrasound in a private practice. At baseline, the practice used free-text reports without specific management guidance. In the first phase, structured reporting using the ACR TI-RADS feature categories was implemented, still without management recommendations. The second phase used the same structured report but added guidance on management that was based on ACR TI-RADS. Adoption of an ACR TI-RADS structured reporting template improved the quality of thyroid ultrasound reports in two ways. First, it served as a checklist that helped radiologists to detect findings suspicious for malignancy. For example, using a free-text report, shape and margin were not reported for 100% and 92% of nodules, respectively [16]. No taller-than-wide nodules were reported at baseline, and an irregular margin was noted in only 1% of patients. When a structured template was introduced in the two-phase introduction of ACR TI-RADS, reporting of taller-than-wide shape rose to 4–8% and irregular margin increased to 6–12%, suggesting radiologists were paying closer attention to these features.

ACR TI-RADS also enhanced quality by increasing the number of reports with definitive management recommendations. Lack of management guidance is often cited by referring physicians as a shortfall [8, 17] and probably contributed to needless workup of many nodules in the study by Griffin et al. [16]. The proportion of nodules without such recommendations was 34% at baseline with free-text reports and 31% when risk stratification was introduced without management recommendations [16]. This figure decreased substantially to 6% in the last phase, when the practice fully implemented ACR TI-RADS.

Using an RSS also reduces unneeded thyroid biopsies compared with having no uniform guidelines. In a study by Hoang et al. [18], eight radiologists reviewed 100 nodules and provided biopsy recommendations. The mean number of biopsies was 80 of 100 (range, 38–95) based on radiologists' own practice patterns (some without an RSS). After retrospectively applying ACR TI-RADS on the basis of the radiologist's assignment of imaging features, the mean number of biopsies recommended was 57 of 100 (range, 37–73). In a prospective analysis of 502 nodules, Grani et al. [2] found that ACR TI-RADS avoided unnecessary biopsies in more nodules than other systems.

How Does ACR TI-RADS Compare With Other Risk Stratification Systems?

As mentioned, ACR TI-RADS places nodules in one of five risk levels by assigning specific point values to ultrasound features, with more suspicious findings garnering more points. At each level, ACR TI-RADS provides precise size cutoffs that determine management recommendations, whether FNA, ultrasound follow-up, or no further action.

The ATA guidelines [14], European Thyroid Association Thyroid Imaging Reporting and Data System (EU-TIRADS) [19], and the Korean Society of Thyroid Radiology Thyroid Imaging Reporting and Data System (K-TIRADS) [20] use different approaches. In these RSSs, the physician assigns a nodule to one of several predefined patterns or determines how many suspicious features are present without assigning numeric values to them (Table 1).

The ACR TI-RADS Committee members believed that a points-based system would be less ambiguous than matching nodules to predefined ultrasound patterns, as is done with the ATA guidelines [14]. In addition, they believed that associating every feature with a numeric score would result in fewer biopsies than assigning nodules with only one suspicious finding to the highest risk level, as is done in K-TIRADS and EU-TIRADS [19, 20] (Table 1). The ACR TI-RADS approach also does not leave any nodules unclassified, reflecting another design goal. Middleton et al. [21] found 3.9% and 13.9% of nodules could not be classified by K-TIRADS and the ATA guidelines, respectively (Fig. 2).

The second key difference is that ACR TI-RADS adopts larger size cutoffs for recommending FNA, with the intention of avoiding unnecessary interventions. For example, the ACR TI-RADS thresholds for biopsy are 1.5 cm for TR4 and 2.5 cm for TR3, whereas the ATA guidelines and K-TIRADS use 1 cm and 1.5 cm, respectively, for similar nodules [14, 20] (Figs. 2 and 3). The EU-TIRADS uses the same 1.5-cm threshold as ACR TI-RADS for TR4 but has a lower threshold of 2 cm for TR3 nodules [19].

Nodules are managed according to size thresholds not because size determines the likelihood of malignancy but because survival data show a poorer prognosis for larger primary tumors. However, thyroid cancer has excellent survival with a 10-year relative survival of 99.4% for a thyroid cancer smaller than 3 cm [22]. The highest size threshold in ACR TI-RADS, 2.5 cm for TR3 nodules, was supported by a study by Nguyen et al. [23]. Their study of 112,128 patients from the National Cancer Institute SEER database found that thyroid cancer tumor size did not increase the mortality rate above baseline until tumors were larger than 2.5 cm [23]. Furthermore, in a subgroup analysis of differentiated thyroid cancers smaller than 4 cm, the risk of local invasion, nodal metastases, or distant metastases was low, and no size threshold was associated with a sharp rise in adverse outcomes.

RSSs also differ in their management of benign nodules. The ATA guidelines and K-TIRADS recommend that FNA be "considered" for spongiform nodules larger than 2 cm [14, 20], whereas ACR TI-RADS recommends that spongiform nodules receive no FNA or follow-up at any size. ACR TI-RADS also considers TR2 mixed cystic-solid nodules without suspicious features as benign and not requiring FNA or follow-up at any size (Figs. 4 and 5).

Koseoglu Atilla et al. [24] quantified the impact of size thresholds and classification between the ATA guidelines and ACR TI-RADS. They applied ACR TI-RADS retrospectively to 2614 patients who had undergone biopsy on the basis of the ATA criteria. Of those, 1382 (52.9%) would have not undergone biopsy under ACR TI-RADS. This number included 508 TR2 nodules (19.4%) (mixed cystic-solid without suspicious features), which corresponded to the ATA low and very low suspicion categories (Figs. 4 and 5). These nodules would have been biopsied at 1.5 cm if they had an eccentric solid component and 2 cm without an eccentric solid component under the ATA guidelines (Table 1). The other group comprised 874 patients with TR3 nodules smaller than 25 mm (33.4%), which fell into the ATA low suspicion category. This group included mixed cystic-solid nodules with only macrocalcifications and iso- or hyperechoic solid nodules without macrocalcifications that are biopsied at 1.5 cm under the ATA system (Fig. 3).

TABLE 1: Comparison of ACR TI-RADS, ATA Guidelines, EU-TIRADS, and K-TIRADS

ACR TI-RADS Level	ATA	FNA Threshold (cm)	EU-TIRADS	FNA Threshold (cm)	K-TIRADS	FNA Threshold (cm)
High (TR5)	Solid hypoechoic nodule with any of five suspicious features[a]; solid hypoechoic component of partial cystic nodule with any suspicious features (both TR4–TR5)	≥ 1.0	At least one of four suspicious features[b] (TR4–TR5)	≥ 1.0	Solid hypoechoic nodule with any of three suspicious features[c] (TR4–TR5)	≥ 1.0
Intermediate (TR4)	Solid hypoechoic nodule with no suspicious features (TR4)	≥ 1.0	No suspicious features, but mildly hypoechoic (TR4)	≥ 1.5	Solid hypoechoic nodule with no suspicious features (TR4); partially cystic or isohyperechoic solid nodule with any of three suspicious features[c] (TR4–TR5)	≥ 1.0
Low (TR3)	Solid isoechoic or hyperechoic nodule with no suspicious features, only macrocalcifications (TR3–TR4); partial cystic nodule with eccentric solid areas with no suspicious features, only macrocalcifications (TR3); partial cystic nodule with eccentric solid areas with no suspicious features or echogenic foci (TR2)	≥ 1.5	No suspicious features, iso- or hyperechoic (TR2–TR3)	≥ 2.0	Partially cystic or isohyperechoic solid nodule with no suspicious features (TR2–TR3)	≥ 1.5
Very low (TR2)	Partially cystic nodule without high, intermediate or low suspicion features, or spongiform (TR1–TR2)	≥ 2.0	Anechoic or spongiform (TR1)	No FNA	Spongiform, partially cystic nodule with comet tail, pure cyst (TR1)	≥ 2.0 spongiform
Benign or normal (TR1)	Purely cystic (TR1)	No FNA	No nodule		No nodule	

Note—ACR TI-RADS level or levels given in parentheses represent the TR level or levels given the description provided by ATA, EU-TIRADS, and K-TIRADS. ACR = American College of Radiology, TI-RADS = Thyroid Imaging Reporting and Data System, ATA = American Thyroid Association, EU-TIRADS = European Thyroid Association Thyroid Imaging Reporting and Data System, K-TIRADS = Korean Society of Thyroid Radiology Thyroid Imaging Reporting and Data System, FNA = fine-needle aspiration.

[a]The five suspicious features are irregular margin, microcalcifications, taller-than-wide shape, rim calcifications with extrusive soft tissue, and extrathyroidal extension.

[b]The four suspicious features are irregular shape, irregular margin, microcalcifications, and marked hypoechogenicity.

[c]The three suspicious features are irregular or spiculated margin, microcalcifications, and taller-than-wide shape.

Does It Matter Which Risk Stratification System I Use?

Multiple studies have compared ACR TI-RADS to other RSSs [2, 21, 24, 25]. Given the higher size thresholds of ACR TI-RADS, its lower sensitivity and higher specificity are not surprising. However, the studies generally compared RSSs solely on biopsy recommendations and did not account for ACR TI-RADS follow-up guidance for TR3 to TR5 nodules below the size cutoff for FNA.

Middleton et al. [21] retrospectively applied ACR TI-RADS, K-TIRADS, and the ATA guidelines in 3422 thyroid nodules (352 malignant, 3070 benign) with pathologic proof. The percentage of benign nodules that would have received a recommendation for biopsy was lowest for ACR TI-RADS at 47.1%, compared with K-TIRADS (79.7%) and the ATA guidelines (78.1%). This represents a 39.7–40.9% reduction in biopsies of benign nodules. The yield of malignancy in nodules receiving a recommendation for biopsy was 14.2% for ACR TI-RADS, higher than K-TIRADS (10.2%) and the ATA guidelines (10.0%). The percentage of malignant nodules that would have been biopsied was lower for ACR TI-RADS at 68.2%, compared with 78.7% and 75.9% for K-TIRADS and the ATA guidelines, respectively, but the percentage of malignant nodules that would have been either biopsied or followed was 89.2% for ACR TI-RADS.

In a prospective study of 477 patients with 502 thyroid nodules (36 malignant, 466 benign) referred for FNA, Grani et al. [2] found that various RSSs reduced the rate of unnecessary biopsies by 17.1–53.4%. The percentage of benign nodules that would have received a recommendation for biopsy was lowest for ACR TI-RADS at 43.8%, compared with K-TIRADS (82.2%), the ATA guidelines (54.7%), and EU-TIRADS (68.0%). This represents a 19.9–46.5% reduction in benign biopsies. ACR TI-RADS showed the greatest reduction (268 of 502 nodules) with the lowest false-negative rate of 2.2% (NPV, 97.8%; 95% CI, 95.2–99.2%).

Xu et al. [26] compared K-TIRADS, EU-TIRADS, and ACR TI-RADS for 1460 benign and 1005 malignant nodules. The rate of unnecessary FNA biopsies was lowest with ACR TI-RADS (17.3%), followed by EU-TIRADS (25.2%) and K-TIRADS (32.1%). This represents a 31.3–46.1% reduction in benign biopsies in this cohort with a high malignancy rate.

A meta-analysis by Li et al. [25] included 16 studies to obtain pooled sensitivity and specificity. In 10 of these studies, a direct comparison could be made between ACR TI-RADS and the ATA guidelines. The pooled sensitivity and specificity were not significantly different at 0.83 versus 0.87 (p = .50) and 0.69 versus 0.50 (p = .10), respectively. ACR TI-RADS could be compared with K-TIRADS in six studies. The pooled sensitivity was not significantly different at 0.85 for ACR TI-RADS versus 0.91 for K-TIRADS (p = .13), but the specificity difference (0.57 for ACR TI-RADS vs 0.24 for K-TIRADS, p < .001) was dramatic.

To date, studies comparing RSSs have typically been based on suspicious nodules that had been recommended for FNA, surgery, or both [2, 21, 24, 25]. However, thyroid RSSs are intended to be applied to all thyroid nodules seen on diagnostic ultrasound. An ACR registry studied this larger population to compare the biopsy recommendation rate (Hoang JK, et al., presented at the Radiological Society of North America [RSNA] 2020 annual meeting). The registry comprised thyroid nodules reported on diagnostic ultrasound from seven clinical sites. Evaluation of 32,746 nodules in 13,814 patients showed that ACR TI-RADS recommended biopsy in 8342 (26%) of nodules mentioned in the body of ultrasound reports (Hoang JK, et al., RSNA 2020 annual meeting). Applying the ATA guidelines, EU-TIRADS, and K-TIRADS to the same nodules resulted in biopsy recommendation rates of 51%, 34%, and 50%, respectively (Hoang JK, et al., RSNA 2020 annual meeting).

Will I Miss More Malignancies With ACR TI-RADS?

Although some physicians have expressed concern about the possibility of missing cancers, this risk is mitigated by the explicit recommendations of ACR TI-RADS for following nodules that do not merit FNA. Follow-up is a safe option for small suspicious nodules and even diagnosed small papillary thyroid cancers given the indolent behavior of thyroid cancer and the harms of thyroid cancer overdiagnosis [4–8].

Application of ACR TI-RADS results in a higher proportion of malignancies that do not receive a recommendation for FNA (17–32%) compared with the ATA guidelines (5–25%), K-TIRADS (0–21%), and EU-TIRADS (14%) [2, 27–29]. However, the likelihood of a negative outcome is low. Middleton et al. [9] compared RSSs and found that the ATA guidelines and K-TIRADS would recommend biopsy in 75.9% and 78.7% of 352 malignant nodules, respectively. ACR TI-RADS would recommend biopsy in 68.2% of nodules, but an additional 21% would be monitored [9]. The majority of malignant nodules that would not receive a recommendation for FNA based on the ACR TI-RADS would either receive follow-up ultrasounds for 5 years, be < 1 cm in maximum diameter, or both.

The previously mentioned study by Koseoglu Atilla et al. [24] found that ACR TI-RADS could prevent 1382 patients (52.9%) from receiving thyroid biopsies when retrospectively applied to a cohort of 2614 patients who were recommended to undergo biopsy under ATA guidelines. The specificity of ACR TI-RADS in this cohort of nodules was 98.8%. Thyroid malignancy was detected in 17 of 1382 patients (1.2%) who would not meet criteria for biopsy, some of whom would have met criteria for follow-up with ultrasound in 12 months.

The pursuit of workup for thyroid nodules has led to overdiagnosis of cancer. Davies and Welch [30] studied the epidemiology of thyroid cancer and reported that the incidence of thyroid cancer has nearly tripled in the last 40 years with no change in mortality. The analysis by Ahn et al. [6] of 8 years of data from a South Korean screening program reported a 15-fold rise in thyroid cancer incidence but no change in cancer-specific mortality. They also found an increase in surgery-related morbidity during this same period. Vaccarella et al. [4] described a worldwide epidemic of thyroid cancer and estimated that 66% of thyroid cancer diagnoses in the United States represented overdiagnosis, defined as identification of malignancies that would not have harmed the patient had they not been discovered. Active surveillance programs in the United States and abroad show no change in survival with delayed surgery even if there is growth in the malignancy or new nodal metastases [31–34]. These findings support the reduced biopsy rates recommended by ACR TI-RADS.

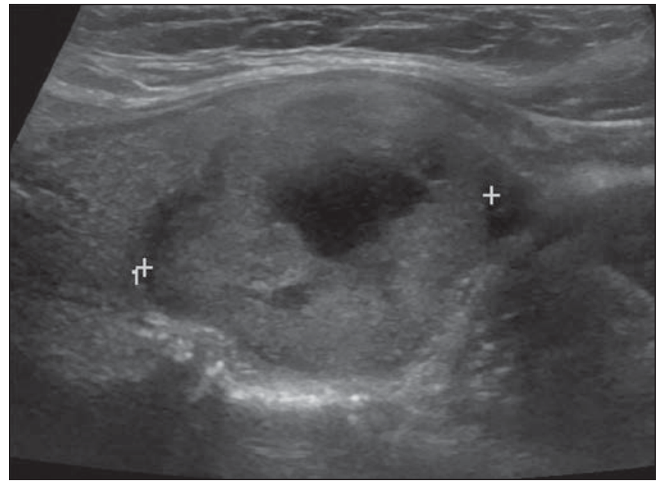

Fig. 2—Transverse ultrasound image in 52-year-old woman shows isoechoic mixed cystic-solid nodule with lobulated margins (*calipers*, *1*). Under American College of Radiology (ACR) TI-RADS Assessment Categories flowchart (Fig. 1), this nodule would receive category of TR4 (4 points) because of 1 point for composition, 1 point for isoechogenicity, and 2 points for lobulated margins. Under ACR TI-RADS, TR4 nodules are recommended for fine-needle aspiration at 1.5 cm and follow-up at 1.0 cm. This nodule could not be classified under American Thyroid Association guidelines because it does not match either partially cystic nodule with suspicious features in that system. European Thyroid Association TIRADS would consider this nodule to be TR5 because irregular shape is one of its four suspicious features. System recommends biopsy at 1.0 cm for TR5 lesions. Korean Society of Thyroid Radiology TIRADS (K-TIRADS) places partially cystic nodules with irregular margins in TR4, matching ACR TI-RADS, but recommends biopsy at 1.0 cm, which is same as K-TIRADS TR5.

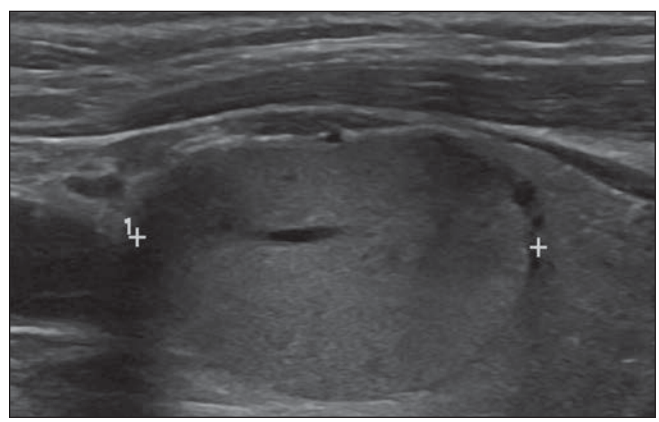

Fig. 3—Transverse ultrasound image in 65-year-old woman shows solid isoechoic nodule (*calipers*, *1*) without suspicious features. Under American College of Radiology (ACR) TI-RADS Assessment Categories flowchart (Fig. 1), this nodule is TR3 (3 points) because 2 points would be assigned to solid composition and 1 point would be assigned to isoechogenicity. Under ACR TI-RADS, TR3 nodules are recommended for fine-needle aspiration at 2.5 cm and follow-up at 1.5 cm. American Thyroid Association guidelines, European Thyroid Association TIRADS, and Korean Society of Thyroid Radiology TIRADS would consider this low suspicion and biopsy at 1.5 cm, 2.0 cm, and 1.5 cm, respectively.

What Are Challenges to Adoption of ACR TI-RADS?

A primary goal of the ACR's imaging, reporting, and data systems is to standardize reporting and provide consistent management recommendations across practices. However, although ACR TI-RADS has improved consistency in management recommendations, interobserver variability for biopsy recommendation remains a concern [35].

High interobserver variability between radiologists was seen in a study in which 100 nodules were interpreted by eight radiologists, of whom six were not subspecialty trained in ultrasound [35]. Interobserver variability was fair to moderate for all features except shape (κ = 0.61) and macrocalcifications (κ = 0.73), which exhibited substantial agreement. The features with the poorest agreement were margin and other types of echogenic foci, which had kappa values ranging from 0.25 to 0.39, indicating fair agreement. Applying ACR TI-RADS resulted in moderate agreement for management (κ = 0.51), which was superior to not applying any guidelines at all.

The ACR thyroid ultrasound registry provided insights into the high variability between practices. In 32,746 nodules in 13,309 patients, biopsy rates across seven practices (four academic, three private) varied by 4–35% (Hoang JK, et al., RSNA 2020 annual meeting). Although ACR TI-RADS was applied to nodules at all sites, the variability in recommendations was caused by significant differences in reported suspicious imaging findings, especially hypoechogenicity and punctate echogenic foci (PEF). Assignment of hypoechogenicity and PEF ranged from 19% to 69% and from 3% to 20%, respectively. As a result, the prevalence of TR4 and TR5 nodules varied from 25% to 58% and from 4% to 14%, respectively. Although these sites evaluated different nodules, the observed variation was likely due to differences in feature assignment.

Multiple other studies have evaluated interobserver agreement using sonographic criteria, with agreement ranging from poor to excellent [36–40]. Several factors contribute to interobserver variability between radiologists, between practices, and between research studies.

The first is readers' experience. Kim et al. [40] studied interobserver variability among five faculty radiologists and four residents who retrospectively analyzed 133 thyroid nodules for features. Agreement between the faculty was fair to good for all features but was only poor to fair for residents. This result suggested that residents misinterpreted findings. Chung et al. [41] came to the opposite conclusion. They studied the impact of experience on interpretation of 150 nodules by three experienced and three less experienced readers. The concordance for numerous characteristics was significantly higher for the less experienced readers for features such as margin and PEF. The authors suggested that this result was because the less experienced readers applied the ACR TI-RADS criteria more rigidly, whereas the radiologists who had been in practice longer partially used their accumulated experience to form impressions.

Another factor is the ability to confidently differentiate artifacts from pathology, a skill that comes with training and experience. Large comet tail artifacts or small areas of high echogenicity arising from the back wall of minute cysts may be misinterpreted as PEF (Figs. 4 and 5). Because PEF are assigned 3 points in ACR TI-RADS, their presence in otherwise low-suspicion nodules raises their classification to TR4. Also, subtle anechoic cystic components of mixed nodules may be misclassified as solid (Fig. 5). Assignment of hypoechoic or very hypoechoic echogenicity adds a further 2–3 points, moving a TR2 nodule that does not warrant FNA or follow-up to TR3 or even TR4.

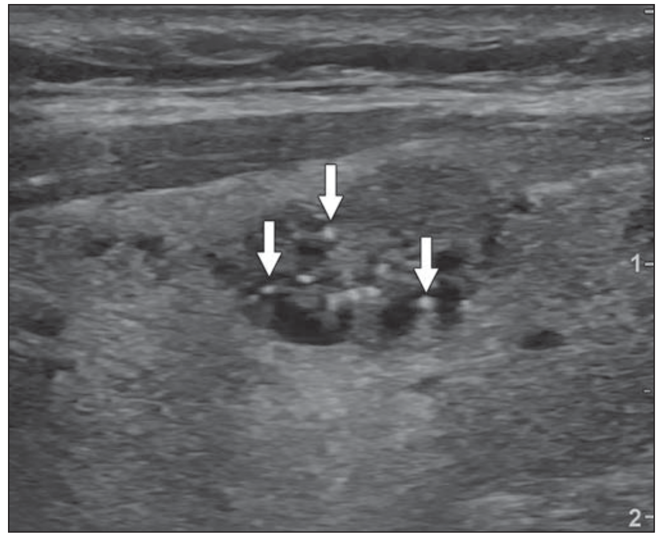

Fig. 4—Transverse ultrasound image in 46-year-old woman shows isoechoic, mixed cystic-solid nodule with category of TR2 (2 points). Comet tail artifacts (*arrows*) could be misinterpreted as punctate echogenic foci. This would result in categorization of nodule as TR4 (5 points) and biopsy with 1.5-cm threshold instead of no biopsy at any size.

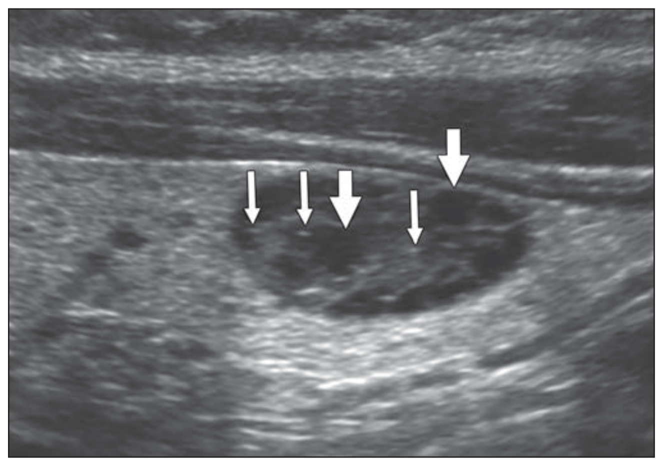

Fig. 5—Transverse ultrasound image in 78-year-old man showing isoechoic, mixed cystic-solid lesion with category of TR2 (2 points). Posterior acoustic enhancement could be missed and cystic regions (*thick arrows*) misinterpreted as hypoechoic components of heterogeneous solid nodule (TR4, 4 points). In addition, back wall of cysts (*thin arrows*) could be misinterpreted as punctate echogenic foci, which would result in categorization of nodule as TR5 (5 points) and biopsy with 1.0-cm threshold instead of no biopsy at any size.

Finally, scanning techniques, which differ by practice, can amplify interobserver variability. The appearance and conspicuity of features such as echogenicity vary considerably depending on scanning parameters, particularly gain, transmit frequency, compression, and pre- and postprocessing. Obtaining images at various gain settings helps to differentiate anechoic cystic areas from markedly hypoechoic nodules, and flow in hypoechoic areas differentiates solid tissue from debris or hemorrhage in fluid-containing regions. In larger nodules, inclusion of adjacent normal thyroid parenchyma and anterior musculature is essential to gauge relative echogenicity. These attributes must be clearly captured on recorded images to be apparent to interpreting radiologists, especially if they rely solely on images and clips captured by sonographers.

Online educational resources are available that instruct physicians about critical nuances of ACR TI-RADS and help address the problem of interobserver variability [42]. Practices may also conduct their own educational sessions and measure the rates in which radiologists describe suspicious features. Studies show that consensus discussion and training sessions improve interobserver variability [36, 40, 43].

Because all RSSs are based on the same sonographic signs, interobserver variability also affects the ATA guidelines, K-TIRADS, and EU-TIRADS. Seifert et al. [43] evaluated ACR TI-RADS and these three other commonly used RSSs and their respective ultrasound features for the assessment of thyroid nodules with respect to the rating agreement between four experienced observers. They found no difference in interobserver variability or diagnostic accuracy between the four systems. A study by Grani et al. [36] evaluated the interobserver variability of two experienced readers for five RSSs (American Association of Clinical Endocrinologists/American College of Endocrinology/Associazione Medici Endocrinologi, ACR TI-RADS, ATA, EU-TIRADS, and K-TIRADS). Concordance for sonographic signs was moderate to near perfect, and agreement for the indication to biopsy was substantial to near-perfect (0.73, 0.61, 0.75, 0.68, and 0.82, respectively) and not significantly different between systems. This outcome highlights the importance of accurate feature assignment regardless of which RSS is applied.

Transition to ACR TI-RADS after use of another RSS also poses a challenge. Tappouni et al. [10] describe issues in education, workflow, and adherence to interpretation and reporting standards when implementing such a practice-wide change. Resistance may come from radiologists who do not perceive the need for change and from providers who may prefer to follow the guidelines of their own professional societies. Research comparing ACR TI-RADS to other RSSs highlighted in this review may help justify the transition. Providers also may be reassured by a statement in the radiology report that gives an option to follow the ATA guidelines.

Future Directions

Modifications to ACR TI-RADS will be appropriate as new data emerge. Given the indolent behavior of thyroid cancer and known outcomes of active surveillance, a larger size threshold for recommending FNA of TR3–TR5 nodules may be warranted. Another potential modification would be to incorporate nodule location into the evaluation. A recent study reported an increased risk of malignancy in nodules located in the isthmus [44]. Another alteration may be to change point assignments for some features. For example, one study suggested that assigning fewer than 3 points to PEF may be appropriate when they occur in mixed cystic-solid nodules [45].

Other ultrasound techniques, such as elastography and contrast-enhanced ultrasound, may also be incorporated into future revisions of ACR TI-RADS. Elastography provides a reproducible assessment of tissue stiffness and shows increased firmness in malignant nodules, consistent with observations in other organs [46–50]. Meta-analyses of the sensitivity of elastography in thyroid nodules reveal sensitivities of 75–89% [51–56]. However, the sensitivity was more modest in a large prospective study by Friedrich-Rust et al. [57] of 657 nodules, which showed a sensitivity of 69% for a score of 2–3. Contrast-enhanced ultrasound evaluates tumor microvascularization and may be helpful given that angiogenesis is a basis for neoplastic growth [58]. Studies in thyroid nodules suggest this technique may add diagnostic value to conventional ultrasound [59–61]. However, both elastography and contrast-enhanced ultrasound add additional time and expense to thyroid nodule evaluation, and widespread adoption may not be feasible in the short term.

Artificial intelligence and machine learning show considerable promise in classifying nodules, whether as an adjunct to existing RSSs or independently [62–64]. Artificial intelligence can also suggest alterations in point allocation [65]. A recent study that used a genetic artificial intelligence algorithm to reassign points was simpler but more accurate than unmodified ACR TI-RADS [65]. One simplification was to assign points to composition only if the nodule was solid. In addition, points were only assigned to hypoechoic features in the echogenicity category and to punctate echogenic foci and peripheral calcifications in the echogenic foci category.

As noted, ACR TI-RADS joined a crowded field of ultrasound-based RSSs for thyroid nodules. Despite its adoption by many radiologists, other systems and guidelines remain in wide use, particularly by other specialists, leading to confusion. Although they differ in some respects, these RSSs have much in common, particularly their reliance on similar, if not identical, sonographic features. On this premise, the multidisciplinary International Thyroid Ultrasound Working Group recently formed to harmonize the leading systems into a single unified international guideline [66].

Conclusion

ACR TI-RADS is a points-based RSS that differs from the ATA guidelines, EU-TIRADS, and K-TIRADS by adopting higher size thresholds for biopsy and categorizing more nodules in a low-suspicion no-biopsy category. The fear of missing malignancy is mitigated by following nodules that do not meet the criteria for biopsy with ultrasound. The advantages of adopting ACR TI-RADS have been demonstrated by research showing a reduction in the number of unnecessary biopsies of benign nodules by 19.9–46.5% compared with other RSSs [2]. Nevertheless, variability in management recommendations due to interobserver variability in assigning sonographic signs remains a challenge, which will be met by educational efforts. Future directions include application of newer techniques, such as artificial intelligence, and development of an international RSS that incorporates the best of existing systems.

References

1. Tessler FN, Middleton WD, Grant EG, et al. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): white paper of the ACR TI-RADS Committee. *J Am Coll Radiol* 2017; 14:587–595
2. Grani G, Lamartina L, Ascoli V, et al. Reducing the number of unnecessary thyroid biopsies while improving diagnostic accuracy: toward the "right"

TIRADS. *J Clin Endocrinol Metab* 2019; 104:95–102

3. Bahl M, Sosa JA, Nelson RC, Hoang JK. Imaging-detected incidental thyroid nodules that undergo surgery: a single-center experience over 1 year. *AJNR* 2014; 35:2176–2180
4. Vaccarella S, Franceschi S, Bray F, Wild CP, Plummer M, Dal Maso L. Worldwide thyroid-cancer epidemic? The increasing impact of overdiagnosis. *N Engl J Med* 2016; 375:614–617
5. Hoang JK, Nguyen XV, Davies L. Overdiagnosis of thyroid cancer: answers to five key questions. *Acad Radiol* 2015; 22:1024–1029
6. Ahn HS, Kim HJ, Welch HG. Korea's thyroid-cancer "epidemic": screening and overdiagnosis. *N Engl J Med* 2014; 371:1765–1767
7. Esserman LJ, Thompson IM Jr, Reid B. Overdiagnosis and overtreatment in cancer: an opportunity for improvement. *JAMA* 2013; 310:797–798
8. Hoang JK, Nguyen XV. Understanding the risks and harms of management of incidental thyroid nodules: a review. *JAMA Otolaryngol Head Neck Surg* 2017; 143:718–724
9. Middleton WD, Teefey SA, Tessler FN, et al. Analysis of malignant thyroid nodules that do not meet ACR TI-RADS criteria for fine needle aspiration. *AJR* 2020 Jun 30 [published online]
10. Tappouni RR, Itri JN, McQueen TS, Lalwani N, Ou JJ. ACR TI-RADS: pitfalls, solutions, and future directions. *RadioGraphics* 2019; 39:2040–2052
11. Tessler FN, Middleton WD, Grant EG. Thyroid Imaging Reporting and Data System (TI-RADS): a user's guide. *Radiology* 2018; 287:29–36
12. Grant EG, Tessler FN, Hoang JK, et al. Thyroid ultrasound reporting lexicon: white paper of the ACR Thyroid Imaging, Reporting and Data System (TI-RADS) Committee. *J Am Coll Radiol* 2015; 12(12 Pt A):1272–1279
13. Guth S, Theune U, Aberle J, Galach A, Bamberger CM. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. *Eur J Clin Invest* 2009; 39:699–706
14. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid* 2016; 26:1–133
15. Qadan L, Ahmed A, Kapila K. Thyroid ultrasound reports: deficiencies and recommendations. *Med Princ Pract* 2019; 28:280–283
16. Griffin AS, Mitsky J, Rawal U, Bronner AJ, Tessler FN, Hoang JK. Improved quality of thyroid ultrasound reports after implementation of the ACR Thyroid Imaging Reporting and Data System nodule lexicon and risk stratification system. *J Am Coll Radiol* 2018; 15:743–748
17. Davies L, Ouellette M, Hunter M, Welch HG. The increasing incidence of small thyroid cancers: where are the cases coming from? *Laryngoscope* 2010; 120:2446–2451
18. Hoang JK, Middleton WD, Farjat AJ, et al. Reduction in biopsies and improved accuracy for thyroid nodule evaluation: the American College of Radiology Thyroid Imaging Reporting and Data System (ACR TI-RADS). *Radiology* 2018; 287:185–193
19. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European Thyroid Association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: the EU-TIRADS. *Eur Thyroid J* 2017; 6:225–237
20. Shin JH, Baek JH, Chung J, et al.; 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; 17:370–395
21. Middleton WD, Teefey SA, Reading CC, et al. Comparison of performance characteristics of American College of Radiology TI-RADS, Korean Society of Thyroid Radiology TIRADS, and American Thyroid Association Guidelines. *AJR* 2018; 210:1148–1154
22. Nguyen XV, Choudhury KR, Eastwood JD, et al. Incidental thyroid nodules on CT: evaluation of 2 risk-categorization methods for work-up of nodules. *AJNR* 2013; 34:1812–1817
23. Nguyen XV, Roy Choudhury K, Tessler FN, Hoang JK. Effect of tumor size on risk of metastatic disease and survival for thyroid cancer: implications for biopsy guidelines. *Thyroid* 2018; 28:295–300
24. Koseoglu Atilla FD, Ozgen Saydam B, Erarslan NA, et al. Does the ACR TI-RADS scoring allow us to safely avoid unnecessary thyroid biopsy? Single center analysis in a large cohort. *Endocrine* 2018; 61:398–402
25. Li W, Wang Y, Wen J, Zhang L, Sun Y. Diagnostic performance of American College of Radiology TI-RADS: a systematic review and meta-analysis. *AJR* 2021; 216:38–47
26. Xu T, Wu Y, Wu RX, et al. Validation and comparison of three newly-released thyroid imaging reporting and data systems for cancer risk determination. *Endocrine* 2019; 64:299–307
27. Middleton WD, Teefey SA, Reading CC, et al. Multiinstitutional analysis of thyroid nodule risk stratification using the American College of Radiology Thyroid Imaging Reporting and Data System. *AJR* 2017; 208:1331–1341
28. Ha EJ, Na DG, Moon WJ, Lee YH, Choi N. Diagnostic performance of ultrasound-based risk-stratification systems for thyroid nodules: comparison of the 2015 American Thyroid Association Guidelines with the 2016 Korean Thyroid Association/Korean Society of Thyroid Radiology and 2017 American College of Radiology Guidelines. *Thyroid* 2018; 28:1532–1537 [Erratum in *Thyroid* 2019; 29:159]
29. 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; 278:917–924
30. Davies L, Welch HG. Current thyroid cancer trends in the United States. *JAMA Otolaryngol Head Neck Surg* 2014; 140:317–322
31. Oda H, Miyauchi A, Ito Y, et al. Incidences of unfavorable events in the management of low-risk papillary microcarcinoma of the thyroid by active surveillance versus immediate surgery. *Thyroid* 2016; 26:150–155
32. Oda H, Miyauchi A, Ito Y, et al. Comparison of the costs of active surveillance and immediate surgery in the management of low-risk papillary microcarcinoma of the thyroid. *Endocr J* 2017; 64:59–64
33. Brito JP, Ito Y, Miyauchi A, Tuttle RM. A clinical framework to facilitate risk stratification when considering an active surveillance alternative to immediate biopsy and surgery in papillary microcarcinoma. *Thyroid* 2016; 26:144–149
34. Ito Y, Miyauchi A, Inoue H, et al. An observational trial for papillary thyroid microcarcinoma in Japanese patients. *World J Surg* 2010; 34:28–35
35. Hoang JK, Middleton WD, Farjat AE, et al. Interobserver variability of sonographic features used in the American College of Radiology Thyroid Imaging Reporting and Data System. *AJR* 2018; 211:162–167
36. 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; 7:1–7
37. Bae JM, Hahn SY, Shin JH, Ko EY. Inter-exam agreement and diagnostic performance of the Korean Thyroid Imaging Reporting and Data System for thyroid nodule assessment: real-time versus static ultrasonography. *Eur J Radiol* 2018; 98:14–19
38. Srinivas MN, Amogh VN, Gautam MS, et al. A prospective study to evaluate the reliability of Thyroid Imaging Reporting and Data System in differentiation between benign and malignant thyroid lesions. *J Clin Imaging Sci* 2016; 6:5
39. Koh J, Moon HJ, Park JS, et al. Variability in interpretation of ultrasound

elastography and gray-scale ultrasound in assessing thyroid nodules. *Ultrasound Med Biol* 2016; 42:51–59
40. Kim SH, Park CS, Jung SL, et al. Observer variability and the performance between faculties and residents: US criteria for benign and malignant thyroid nodules. *Korean J Radiol* 2010; 11:149–155
41. Chung R, Rosenkrantz AB, Bennett GL, et al. Interreader concordance of the TI-RADS: impact of radiologist experience. *AJR* 2020; 214:1152–1157
42. American College of Radiology. ACR TI-RADS webinars. www.acr.org/clinical-resources/reporting-and-data-systems/ti-rads/webinars. Accessed October 17, 2020
43. Seifert P, Görges R, Zimny M, Kreissl MC, Schenke S. Interobserver agreement and efficacy of consensus reading in Kwak-, EU-, and ACR-thyroid imaging recording and data systems and ATA guidelines for the ultrasound risk stratification of thyroid nodules. *Endocrine* 2020; 67:143–154
44. Jasim S, Baranski TJ, Teefey SA, Middleton WD. Investigating the effect of thyroid nodule location on the risk of thyroid cancer. *Thyroid* 2020; 30:401–407
45. Teefey SA, Middleton WD, Reading CC, et al. Effect of decreasing the point assignment for punctate echogenic foci when they occur in mixed solid and cystic thyroid nodules. *AJR* 2020 Dec 9 [published online]
46. Zhao CK, Xu HX. Ultrasound elastography of the thyroid: principles and current status. *Ultrasonography* 2019; 38:106–124
47. Gerber L, Fitting D, Srikantharajah K, et al. Evaluation of 2D-shear wave elastography for characterisation of focal liver lesions. *J Gastrointestin Liver Dis* 2017; 26:283–290
48. Grgurevic I, Bokun T, Salkic NN, et al. Liver elastography malignancy prediction score for noninvasive characterization of focal liver lesions. *Liver Int* 2018; 38:1055–1063
49. Kuo YW, Chen YL, Wu HD, Chien YC, Huang CK, Wang HC. Application of transthoracic shear-wave ultrasound elastography in lung lesions. *Eur Respir J* 2020 Oct 8 [published online]
50. Yang H, Xu Y, Zhao Y, Yin J, Chen Z, Huang P. The role of tissue elasticity in the differential diagnosis of benign and malignant breast lesions using shear wave elastography. *BMC Cancer* 2020; 20:930
51. Hu X, Liu Y, Qian L. Diagnostic potential of real-time elastography (RTE) and shear wave elastography (SWE) to differentiate benign and malignant thyroid nodules: a systematic review and meta-analysis. *Medicine (Baltimore)* 2017; 96:e8282
52. Zhan J, Jin JM, Diao XH, Chen Y. Acoustic radiation force impulse imaging (ARFI) for differentiation of benign and malignant thyroid nodules: a meta-analysis. *Eur J Radiol* 2015; 84:2181–2186
53. Liu BJ, Li DD, Xu HX, et al. Quantitative shear wave velocity measurement on acoustic radiation force impulse elastography for differential diagnosis between benign and malignant thyroid nodules: a meta-analysis. *Ultrasound Med Biol* 2015; 41:3035–3043
54. Dong FJ, Li M, Jiao Y, et al. Acoustic radiation force impulse imaging for detecting thyroid nodules: a systematic review and pooled meta-analysis. *Med Ultrason* 2015; 17:192–199
55. Lin P, Chen M, Liu B, Wang S, Li X. Diagnostic performance of shear wave elastography in the identification of malignant thyroid nodules: a meta-analysis. *Eur Radiol* 2014; 24:2729–2738
56. Bojunga J, Herrmann E, Meyer G, Weber S, Zeuzem S, Friedrich-Rust M. Real-time elastography for the differentiation of benign and malignant thyroid nodules: a meta-analysis. *Thyroid* 2010; 20:1145–1150
57. Friedrich-Rust M, Vorlaender C, Dietrich CF, et al. Evaluation of strain elastography for differentiation of thyroid nodules: results of a prospective DEGUM multicenter study. *Ultraschall Med* 2016; 37:262–270
58. Zhan J, Ding H. Application of contrast-enhanced ultrasound for evaluation of thyroid nodules. *Ultrasonography* 2018; 37:288–297
59. Zhang YZ, Xu T, Gong HY, et al. Application of high-resolution ultrasound, real-time elastography, and contrast-enhanced ultrasound in differentiating solid thyroid nodules. *Medicine (Baltimore)* 2016; 95:e5329
60. Zhao RN, Zhang B, Yang X, Jiang YX, Lai XJ, Zhang XY. Logistic regression analysis of contrast-enhanced ultrasound and conventional ultrasound characteristics of sub-centimeter thyroid nodules. *Ultrasound Med Biol* 2015; 41:3102–3108
61. Li F, Luo H. Comparative study of thyroid puncture biopsy guided by contrast-enhanced ultrasonography and conventional ultrasound. *Exp Ther Med* 2013; 5:1381–1384
62. Park VY, Han K, Seong YK, et al. Diagnosis of thyroid nodules: performance of a deep learning convolutional neural network model vs. radiologists. *Sci Rep* 2019; 9:17843
63. Sun C, Zhang Y, Chang Q, et al. Evaluation of a deep learning-based computer-aided diagnosis system for distinguishing benign from malignant thyroid nodules in ultrasound images. *Med Phys* 2020 May 30 [published online]
64. Buda M, Wildman-Tobriner B, Hoang JK, et al. Management of thyroid nodules seen on US images: deep learning may match performance of radiologists. *Radiology* 2019; 292:695–701
65. Wildman-Tobriner B, Buda M, Hoang JK, et al. Using artificial intelligence to revise ACR TI-RADS Risk Stratification of Thyroid Nodules: Diagnostic Accuracy and Utility. *Radiology* 2019; 292:112–119
66. Tessler FN. Thyroid nodules and real estate: location matters. *Thyroid* 2020; 30:349–350