Validity of the American College of Radiology Thyroid Imaging Reporting and Data System in Children

Kelly E. Daniels, MD,
Department of Otolaryngology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

Amber D. Shaffer, PhD,
Division of Pediatric Otolaryngology, UPMC Children's Hospital of Pittsburgh, Pittsburgh, Pennsylvania, USA

Steven Garbin, MD,
Department of Otolaryngology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

Department of Emergency Medicine, University of Virginia School of Medicine, Charlottesville, VA, USA

Judy H. Squires, MD,
Division of Radiology, UPMC Children's Hospital of Pittsburgh, Pittsburgh, Pennsylvania, USA

Kevin G. Vaughan, MD,
Division of Radiology, UPMC Children's Hospital of Pittsburgh, Pittsburgh, Pennsylvania, USA

Pushpa Viswanathan, MD,
Division of Endocrinology, UPMC Children's Hospital of Pittsburgh, Pittsburgh, Pennsylvania, USA

Selma F. Witchel, MD,
Division of Endocrinology, UPMC Children's Hospital of Pittsburgh, Pittsburgh, Pennsylvania, USA

Kevin P. Mollen, MD,
Department of Surgery, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

Linwah Yip, MD,
Department of Surgery, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

Sara E. Monaco, MD,

Send correspondence to Jeffrey P. Simons, UPMC Children's Hospital of Pittsburgh, 4401 Penn Avenue, Pittsburgh, PA 15224, USA. jeffrey.simons@chp.edu.

Level of Evidence: 3

Division of Pediatric Otolaryngology, UPMC Children's Hospital of Pittsburgh, Pittsburgh, Pennsylvania, USA

Department of Laboratory Medicine, Geisinger Medical Center, Danville, PA, USA.

Umamaheswar Duvvuri, MD,
Department of Otolaryngology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

Jeffrey P. Simons, MD
Department of Otolaryngology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

Division of Pediatric Otolaryngology, UPMC Children's Hospital of Pittsburgh, Pittsburgh, Pennsylvania, USA

Abstract

Objective: To assess the validity of the American College of Radiology Thyroid Imaging Reporting and Data System (ACR TI-RADS) for evaluating thyroid nodules in children.

Methods: Patients aged <19 years with thyroid nodule(s) evaluated by ultrasound (US) from 2007–2018 at a tertiary children's hospital were included. Two radiologists scored de-identified thyroid US images using ACR TI-RADS (from 1, "benign" to 5, "highly suspicious"). The radiologists recorded size and rated vascularity for each nodule. Ultrasound findings were compared to pathology results (operative cases, $n = 91$) and clinical follow-up without disease progression (non-operative cases, $n = 15$).

Results: Thyroid images from 115 patients were reviewed. Nine patients were excluded due to the absence of an evaluable nodule. Forty-seven benign and 59 malignant nodules were included. Median age at ultrasound was 15 years (range 0.9–18 years). Twenty (18.9%) patients were male. There was moderate agreement between TI-RADS levels assigned by the two raters (kappa = 0.57, $p < 0.001$). When the raters' levels were averaged, >3 as the threshold for malignancy correctly categorized the greatest percentage of nodules (68.9%). Eleven (18.6%) malignant nodules received a TI-RADS level of 2 ($n = 3$) or 3 ($n = 8$). Sensitivity, specificity, and positive and negative predictive values were 81.4%, 53.2%, 68.6%, and 69.4%, respectively. Although not part of TI-RADS, vascularity was similar between benign and malignant nodules ($p = 0.56$).

Conclusion: In a pediatric population, TI-RADS can help distinguish between benign and malignant nodules with comparable sensitivity and specificity to adults. However, the positive and negative predictive values suggest TI-RADS alone cannot eliminate the need for FNA.

Keywords

INTRODUCTION

Thyroid nodules present a diagnostic challenge in both adult and pediatric populations. Although thyroid nodules are common in adults, the reported prevalence of thyroid

nodules is 0.5–5% in children and adolescents. Different from adults, available data have demonstrated that between 19.1%–25% of pediatric thyroid nodules are malignant.[1–3] Hence, developing precise, non-invasive tools to discriminate malignant from benign nodules is essential.[4] Pediatric patients are more likely to present with occult bilateral disease or local metastases to regional lymph nodes when compared to adults.[4–6] Currently, clinicians caring for children utilize the American Thyroid Association (ATA) Management Guidelines for Children with Thyroid Nodules and Differentiated Thyroid Cancer. In children and adolescents, the decision to perform FNA is based on ultrasonographic features and clinical context, and criteria for biopsy are generally more aggressive compared to adults. Any nodule that has high-risk ultrasonographic features, independent of size, or any nodule that is solid or partially cystic ≥ 1 cm is recommended for biopsy in pediatrics.[7] This differs from the adult population, where biopsy is only recommended once the nodule is ≥ 1, ≥ 1.5, and ≥ 2.5 cm for highly suspicious, moderately suspicious, and mildly suspicious nodules, respectively.[8]

Risk stratification of thyroid nodules for both adult and pediatric populations using ultrasound remains problematic. Several risk stratification systems have been proposed.[9] Kwak et al. developed a risk stratification system that considered hypoechogenicity, microlobulated or irregular margins, microcalcifications, and taller-than-wide shape as independent risk markers for thyroid malignancy.[10] The ATA created a system to risk stratify thyroid nodules based on ultrasound patterns.[11] Most recently, the American College of Radiologists (ACR) proposed the Thyroid Imaging Reporting and Data System (TI-RADS), which is an evidence-based multi-institutional consensus of expert opinion. ACR TI-RADS provides standard terms (lexicon) with recommendations for the management of thyroid nodules based on characteristics of ultrasound imaging paired with maximum diameter. These guidelines, first introduced in 2015 and expanded to include clinical management recommendations in 2017, were initially intended for the assessment and management of thyroid nodules in adults.[12,13] Although the validity and performance of ACR TI-RADS have been studied across adult populations, the usefulness of the ACR TI-RADS in pediatric populations is unclear.

The unique imaging characteristics in pediatric patients, such as ectopic intrathyroidal thymic tissue resembling microcalcifications, are not addressed in the adult ACR TI-RADS system.[14] We hypothesize that ACR TI-RADS will provide actionable information regarding thyroid nodules in children and adolescents. To test our hypothesis, we compared ACR TI-RADS scores to clinical findings in 106 patients.

MATERIALS AND METHODS

Following University of Pittsburgh Institutional Review Board approval (STUDY19020295), our retrospective chart review identified 115 patients aged less than 19 years who had thyroid ultrasound studies performed between 2007–2018 and thyroidectomy or fine needle aspiration (FNA) biopsy. All US were performed in a single hospital network. A third party randomized and de-identified all thyroid ultrasound. Nine studies were excluded because an evaluable nodule could not be identified. Two pediatric radiologists (JHS and KGV) with 4+ and 11+ years, respectively, of post-fellowship experience and blinded to all clinical

information independently scored the 106 images using ACR TI-RADS points system ranging from 0 to 14, which corresponded to levels ranging from 1, "benign" to 5, "highly suspicious".[12,13] As recommended, the ACR TI-RADS considers composition, echogenicity, shape, margin, and echogenic foci to assign a score for each thyroid nodule. The reviewing radiologists also recorded size and rated vascularity for each nodule. Vascularity was assessed using color or power Doppler and assigned one of the following categories: no flow detected; similar to background thyroid; rim hypervascularity; or hypervascular. ACR TI-RADS scores were compared between independent radiologists for inter-rater agreement. The results of the thyroid ultrasound were compared to pathology findings ($n = 91$) for cases undergoing operations. For non-operative cases ($n = 15$), the ultrasound that was evaluated for this study was completed >6 years prior to data collection. These patients were judged to have benign disease based on clinical presentation and FNA biopsy and had no disease progression by the time of data collection. The follow-up period was based on date of data collection rather than the last clinic visit because of patients without concern for disease progression are less likely to return for additional clinic visits. Additionally, none of these patients were known to have moved or transferred care to another hospital system. Patients who presented after ultrasound more recently did not meet our inclusion criteria due to insufficient follow-up (2 years of clinical follow-up at the time of construction of the patient list in 2018) or absence of FNA.

Statistical Analysis

Categorical data were summarized as number and percent. Normally distributed data were presented as means and standard deviations. Non-normally distributed data (Shapiro–Wilk $p < 0.05$) were presented as median and ranges. Differences in characteristics between patients with benign and malignant nodules were assessed using exact or standard logistic regression for categorical characteristics, Wilcoxon rank-sum for non-normally distributed continuous characteristics, and t-tests for normally distributed continuous characteristics. Differences in total ACR TI-RADS points, ACR TI-RADS level, subscale scores, nodule largest dimension, and vascularity were assessed using exact logistic regression, linear regression, Wilcoxon rank-sum, and t-tests. Inter-rater agreement was assessed by weighted Cohen's kappa, except vascularity, which was assessed unweighted. $\kappa = 0$–0.20, 0.21–0.40, 0.41–0.60, 0.61–0.80, and 0.81–1.00 were interpreted as none to slight, fair, moderate, substantial, and almost perfect agreement, respectively.[15] The relationship between nodule largest dimension and ACR TI-RADS level was evaluated using Spearman rank correlation. The optimal threshold for ACR TI-RADS to discriminate benign from malignant nodules was determined by receiver operating characteristic (ROC) curve analysis. The area under the curve (AUC) to distinguish between benign and malignant nodules for total ACR TI-RADS points, subscale scores, nodule largest dimension, and vascularity were compared with those for ACR TI-RADS level using Chi-squared tests with Holm's adjustment for 7 multiple comparisons. Nodules were categorized as high-risk (with ACR TI-RADS 4) or lower risk (ACR TI-RADS <4) as described previously.[16] The AUC for ACR TI-RADS level to distinguish between benign and malignant nodules was compared between high-risk and lower-risk nodules using Chi-squared tests. Finally, the AUC to distinguish between benign and malignant nodules for ACR TI-RADS level was compared between indeterminate

and other nodules using Chi-squared tests. Analyses were performed using Stata/SE 16.1 (StataCorp, College Station, TX). $p < 0.05$ was considered significant.

RESULTS

Patient Characteristics and Pathology

Of the 106 patients included in the study, 59 (55.7%) had malignant nodules, 20 (18.9%) were male, and 15 (14.2%) were diagnosed with chronic lymphocytic thyroiditis. The median age at ultrasound was 15.6 years (range 0.9–18.8 years) (Table I). Chronic lymphocytic thyroiditis was significantly more common in patients with malignant thyroid lesions compared with benign (OR: 6.36, 95% CI: 1.36–29.8, $p = 0.019$). The prevalence of benign thyroid nodules was comparable between males and females (OR: 0.755, 95% CI: 0.285–2.00, $p = 0.572$). There was no significant difference in age at ultrasound between those with benign and malignant lesions ($z = 0.397$, $p = 0.693$). FNA results were available for 91 patients; cytology results for 36 (39.6%) of these patients were indeterminate according to the Bethesda Classification, defined as Bethesda III or IV. Ninety-one patients underwent thyroid surgery. Of these patients, 49 had a total thyroidectomy, 34 had lobectomy with or without isthmusectomy, and 8 had a diagnostic lobectomy followed by subsequent completion thyroidectomy. The pathology results of these 91 patients showed benign findings ($n = 32$), papillary thyroid cancer (PTC; $n = 51$), follicular thyroid cancer ($n = 4$), NIFTP ($n = 2$), and medullary thyroid cancer ($n = 2$). (Table I).

Inter-Rater Reliability

Inter-rater reliability for each of the ACR TI-RADS measures, nodule largest dimension, and vascularity is presented in Table II. Overall, there was moderate agreement between raters for ACR TI-RADS level ($\kappa = 0.576$, SE = 0.066, $p < 0.001$). The composition subscale ($\kappa = 0.817$, SE = 0.081, $p < 0.001$) and nodule largest dimension ($\kappa = 0.870$, SE = 0.063, $p < 0.001$) exhibited almost perfect agreement. The shape subscale demonstrated the poorest agreement ($\kappa = 0.148$, SE = 0.075, $p = 0.025$). Agreement in ACR TI-RADS level between raters continued to be moderate when the subgroups with PTC ($\kappa = 0.600$, SE = 0.103, $p < 0.001$) and indeterminate nodules ($\kappa = 0.507$, SE = 0.106, $p < 0.001$) were examined independently. Agreement was only fair for nodules in patients with chronic lymphocytic thyroiditis ($\kappa = 0.379$, SE = 0.189, $p = 0.023$).

ACR TI-RADS Scores

Average ACR TI-RADS points, ACR TI-RADS level, subscale scores, nodule largest dimension, and vascularity for the two raters are summarized in Table III. Median ACR TI-RADS points was 7 (range 2–13) for malignant nodules and 3.5 (range 0–11) for benign nodules ($z = -4.56$, $p < 0.001$). This corresponded with median ACR TI-RADS levels of 4.5 (range 2–5) and 3 (range 1–5) for malignant and benign nodules, respectively ($z = -4.62$, $p < 0.001$); each unit increase in ACR TI-RADS level was associated with 2.48 increased odds of malignancy (95% CI: 1.62–3.98, $p < 0.001$). Scores for malignant lesions were significantly greater than those for benign lesions in the composition ($z = -5.75$, $p < 0.001$), margin ($z = -3.45$, $p < 0.001$), echogenic foci ($z = -3.00$, $p = 0.002$), and echogenicity ($t(104) = -3.95$, $p < 0.001$) subscales, but not for the shape subscale or largest dimension.

There was no significant difference in ACR TI-RADS levels between the two most common malignancies, PTC (median 5, range 2–5) and FTC (median 3.5, range 3.5–4.5). Vascularity, although not part of ACR TI-RADS, was also examined. For both raters, neither rim nor diffuse vascularity was significantly more common in malignant nodules. In multivariable regression analysis with malignancy and chronic lymphocytic thyroiditis as predictors and ACR TI-RADS level as the response variable, both malignancy ($\beta = 0.897$, 95% CI: 0.500–1.29, $p < 0.001$) and chronic lymphocytic thyroiditis ($\beta = 0.588$, 95% CI: 0.022–1.15, $p = 0.042$) were independently associated with greater ACR TI-RADS levels. Median (range) nodule largest dimension for each ACR TI-RADS level is shown in Table IV. There was no significant relationship between ACR TI-RADS level and nodule largest dimension ($\rho = 0.148$, $p = 0.131$).

Receiver Operating Characteristic Curve Comparisons

An ACR TI-RADS level >3 as the threshold for malignancy correctly categorized the greatest percentage of nodules (68.9%). Sensitivity, specificity, and positive and negative predictive values were 81.4%, 53.2%, 68.6%, and 69.4%, respectively, with an AUC of 0.758 (Figure 1). The AUC for predicting malignancy based on ACR TI-RADS scores did not significantly differ between the two reviewers (AUC 0.756 vs. 0.716, $p = 0.304$). In addition, the AUC for predicting malignancy based on ACR TI-RADS scores did not differ between high-risk (ACR TI-RADS TI-RADS 4) and lower-risk (ACR TI-RADS <4) nodules (AUC 0.824 vs. 0.750, $p = 0.395$) (Figure 2). Table V shows AUC for ACR TI-RADS total points, subscale scores, nodule largest dimension, and vascularity compared with the AUC for ACR TI-RADS level. No other measure significantly outperformed the ACR TI-RADS level based on AUC. The shape (AUC: 0.574, $p < 0.001$) and echogenic foci (AUC: 0.662, $p = 0.001$) subscales, largest dimension (AUC: 0.555, $p = 0.003$), and vascularity (AUC: 0.530, $p = 0.001$) demonstrated significantly poorer ability compared with ACR TI-RADS level to discriminate between benign and malignant nodules based on AUC. When excluding patients with non-papillary TC, an ACR TI-RADS level of 5 for malignancy correctly categorized the greatest percentage of nodules (73.5%), with a sensitivity of 52.9%, specificity of 95.7%, positive predictive value of 93.1%, and negative predictive value of 65.2% (AUC: 0.787) (Table V). The subset of patients with indeterminate nodules demonstrated a median ACR TI-RADS level of 3.5 (range 2–5). In this subset, an ACR TI-RADS level of 5 for malignancy correctly categorized the greatest percentage of nodules (58.3%), with a sensitivity of 30.0%, specificity of 93.8%, positive predictive value of 85.7%, and negative predictive value of 51.7% (AUC: 0.514) (Table V). ACR TI-RADS underperformed for classifying indeterminate nodules (AUC: 0.541) compared with other nodules (AUC: 0.882) ($p = 0.002$).

DISCUSSION

The current literature assessing the usefulness of the adult ACR TI-RADS in the pediatric population is inconsistent. Richman et. al, found that out of 404 nodules, 22.4% of the malignant nodules would be missed if the decision to perform FNA was based on imaging findings alone.[2] Based on their findings, they concluded that ACR TI-RADS is insufficient to risk stratify pediatric thyroid nodules. Gannon et. al, found that ultrasound alone was

insufficient to rule out cancer to defer FNA biopsy.[7] Importantly, ACR TI-RADS is intended to provide additional information to guide clinical management. In current practice with adults, ACR TI-RADS level provides one factor to be considered for the clinical management of a given nodule. Our study found that ACR TI-RADS generated statistically different points and levels for benign versus malignant nodules. In a patient population that is at baseline difficult to risk stratify, the ACR TI-RADS level would be a helpful adjunct as part of a comprehensive evaluation.

The goal of our study was to see whether ACR TI-RADS performance in pediatric patients is comparable to that in the adult population, where it has been validated and is being used with increasing frequency. In one of the first articles on ACR TI-RADS in pediatrics published in 2017, Martinez-Rios, et al. demonstrated a positive predictive value of 71.7% and a negative predictive value of 80.0% in a sample of 124 patients.[16] A recent systematic review and meta-analysis by Kim et al. published a pooled sensitivity and specificity of 0.84 (95% CI 0.78–0.89) and 0.64 (95% CI 0.41–0.81), respectively, for ACR-TI-RADS category 4 and 5 as malignant in pediatric cases.[17] Our test characteristics, computed for an ACR TI-RADS malignancy at >3, were consistent with the above-pooled analysis, with positive and negative predictive values of 68.6% and 69.4%, respectively, and a sensitivity of 81.4% and specificity of 53.2%. In comparison with a 2020 meta-analysis published by Li et. al on TI-RADS performance in the adult population, our test characteristics fall within the ranges published, demonstrating similar performance in both populations. Of note, a recent study out of Italy comparing ACR-TIRADS with three other risk stratification systems for 81 pediatric thyroid nodules found ACR-TIRADS test characteristics to differ from ours, with a sensitivity of 41.7% and specificity at 75.9%.[18] Their study had a larger proportion of benign nodules and excluded higher-risk patients including those with a family history of thyroid cancer and indeterminate nodules. Their test characteristics may be more reflective of generally lower-risk populations, with fewer true positives and relatively more true negatives (yielding lower sensitivity and higher specificity) when compared to our high-risk tertiary care population. With respect to inter-rater agreement, ACR TI-RADS performed with "moderate agreement" overall, with a kappa of 0.576 (0.447–0.705) in our study. This inter-rater agreement level is less than that found in a study by Basha et. al in adults, which yielded a kappa of 0.636 (0.507–0.766), considered one tier higher at "substantial agreement;" however, the confidence intervals do overlap between our pediatric and their adult cohorts indicating similar performance in adult and pediatric populations.[19]

For our study, surgical pathology results or clinical follow-up for non-surgical cases was used as the reference standard because cytology alone does not always accurately predict malignancy. One study looked at the predictive value of ACR TI-RADS in pediatric populations and used surgical pathology when available or cytology as the reference standard.[20] For cytology, these authors classified Bethesda I, II, III as benign and Bethesda IV, V, and VI as malignant. They found only one false negative in a < 1 cm nodule and defined TI-RADS 5 as their optimal cut-off for malignancy.[20] Notably, whereas there is no discreet guideline for adults with indeterminate cytology, those pediatric patients with Bethesda III thyroid nodules are recommended to undergo surgical resection of their nodules via lobectomy and isthmusectomy, as the risk of malignancy is between 8.3%–44%.[17,21,22]

The decision to classify Bethesda III nodules as benign, malignant, or neither will also influence the perceived performance of ACR TI-RADS.

Indeterminate cytology (Bethesda III and IV) continues to create diagnostic challenges, as these are the nodules with mixed features that are more likely to be misclassified. In adults, indeterminate nodules may undergo repeat FNA or molecular testing to assess for gene mutations prior to surgery. Molecular studies have not been as widely used in the pediatric population, and at the present time it is uncertain how the results of molecular testing will contribute to risk-stratification in children. Ultrasound characteristics may be helpful in further delineating surgical versus observation cases. Dissimilar from prior studies, we included nodules with Bethesda III or IV cytology in our analysis, and their inclusion can be expected to influence validity measures accordingly. In our study, we found that more than one-third (39.6%) of patients who underwent FNA had indeterminate cytology, representing a significant portion of patients. In the pediatric population, surgery is recommended for all Bethesda III nodules. Arora et. al found in a sample of 134 nodules that ACR TI-RADS was able to identify low-risk Bethesda III nodules that could instead be observed.[21] Likewise, Cherella et. al sub-classified indeterminate pediatric nodules and noted the nuclear atypia subtype was associated with an increased risk of malignancy; they proposed using the presence or absence of nuclear atypia in a decision tree to determine the need for repeat FNA prior to resection.[23] However, resection remains the standard of care. Because ACR TI-RADS underperformed for classifying indeterminate nodules (AUC: 0.541) compared with other nodules (AUC: 0.882) ($p = 0.002$), observation alone remains questionable, and this elucidates an area of necessary attention in future studies. At present, because surgery is recommended for Bethesda III and IV nodules, and ACR-TIRADS level performed poorest in this group, we would not recommend deviating from the current recommendations. In the current study, the reviewing radiologists were blinded to patient information, including FNA results. However, quantifying improvements in diagnostic accuracy when FNA classification is layered onto ACR TI-RADS level could be a target of future investigation.

There are a few limitations of our study. Although our study was conducted at a tertiary care children's hospital, there is still a relatively small sample size, with 115 nodules analyzed. The literature that exists about pediatric thyroid nodules is consistently limited by smaller sample sizes, generally ranging between 60–150 nodules, with only one article assessing greater than 400 nodules.[1–3,20–22,24] It is apparent that multi-institutional collaboration will be necessary to obtain a sizeable dataset that can match the statistical power of the adult literature. Our sample population is biased, as patients with benign or low-risk nodules are less likely to be referred to our institution and may instead follow with providers in the community closer to home. As a result, we have a higher representation of malignant or high-risk nodules in our study population. Our sample likely represents populations at comparable institutions but should not be considered representative of the population of pediatric thyroid nodules that a general or community otolaryngologist may evaluate. In addition, patients for this study were identified based on surgical and pathology records as opposed to screening all patients with thyroid ultrasounds during the time frame. This may have increased the bias toward surgical cases. Although AUC suggested slightly better performance for ACR TI-RADS when PTC nodules were examined separately from other TC, the small number of nodules with non-PTC malignancies limited comparisons between

pathologies. We are further limited by variability in ultrasound equipment over the 11-year timeframe of this study, as well as variable technique and operator skill in obtaining ultrasound images. However, all images were deemed adequate for interpretation at the time of acquisition and all images were acquired by dedicated pediatric sonographers.

The results of this study can be used to inform and guide future investigations. Prospective studies focused specifically on nodules with indeterminate cytology and nodules with intermediate risk sonographically in the pediatric population are warranted. The significance of additional features, including elastography and vascularization, is being studied in the adult literature, as well as consideration for a weighted point system for ACR TI-RADS.[25] Similarly, exploration of these sonographic features as well as those unique to pediatric patients, and consideration of adjusted weighted points, may contribute to a more accurate grading system specific to this patient population.

CONCLUSION

In our tertiary-care center pediatric population, we found that ACR TI-RADS can help distinguish between benign and malignant nodules with comparable sensitivity and specificity to the adult population. The predictive value suggests that TI-RADS cannot by itself direct clinical management, which is consistent with its utility in adult populations. ACR TI-RADS should not function alone to risk stratify nodules but should instead be used as an adjunct to history, physical exam, and cytology. In a pediatric population, which is historically difficult to accurately risk-stratify and to optimize patient selection for surgery, TI-RADS provides added value to ultrasound risk stratification for thyroid nodules.

FUNDING INFORMATION

BIBLIOGRAPHY

1. Polat Y, Öztürk V, Ersoz N, Anik A, Karaman C. Is thyroid imaging reporting and data system useful as an adult ultrasonographic malignancy risk stratification method in pediatric thyroid nodules? J Med Ultrasound. 2019;27(3):141–145. 10.4103/JMU.JMU_35_19. [PubMed: 31867177]

2. Richman DM, Benson CB, Doubilet PM, et al. Assessment of American college of radiology thyroid imaging reporting and data system (TI-RADS) for pediatric thyroid nodules. Radiology. 2020;294(2):415–420. 10.1148/radiol.2019191326. [PubMed: 31821121]

3. Uner C, Aydin S, Ucan B. Thyroid image reporting and data system categorization: effectiveness in pediatric thyroid nodule assessment. Ultrasound Q. 2020;36(1):15–19. 10.1097/RUQ.0000000000000476. [PubMed: 31725067]

4. Buryk MA, Simons JP, Picarsic J, et al. Can malignant thyroid nodules be distinguished from benign thyroid nodules in children and adolescents by clinical characteristics? A review of 89 pediatric patients with thyroid nodules. Thyroid. 2015;25(4):392–400. 10.1089/thy.2014.0312. [PubMed: 25627462]

5. Cherella CE, Richman DM, Liu E, et al. Predictors of bilateral disease in pediatric differentiated thyroid cancer. J Clin Endocrinol Metab. 2021; 106(10):E4242–E4250. 10.1210/clinem/dgab210. [PubMed: 33780538]

6. Jain NK, Mostoufi-Moab S, Hawkes CP, et al. Extrathyroidal extension is an important predictor of regional lymph node metastasis in pediatric differentiated thyroid cancer. Thyroid. 2020;30(7):1037–1043. 10.1089/thy.2019.0229. [PubMed: 31573414]

7. Gannon AW, Langer JE, Bellah R, et al. Diagnostic accuracy of ultrasound with color flow Doppler in children with thyroid nodules. J Clin Endocrinol Metab. 2018;103(5):1958–1965. 10.1210/jc.2017-02464. [PubMed: 29546281]

8. Francis GL, Waguespack SG, Bauer AJ, et al. Management guidelines for children with thyroid nodules and differentiated thyroid cancer. Thyroid. 2015;25(7):716–759. 10.1089/thy.2014.0460. [PubMed: 25900731]

9. Horvath E, Majlis S, Rossi R, et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. J Clin Endocrinol Metab. 2009;94(5):1748–1751. 10.1210/jc.2008-1724. [PubMed: 19276237]

10. Kwak JY, Han KH, Yoon JH, 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(3):892–899. 10.1148/radiol.11110206. [PubMed: 21771959]

11. 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):1–133. 10.1089/thy.2015.0020. [PubMed: 26462967]

12. Grant EG, Tessler FN, Hoang JK, et al. Thyroid ultrasound reporting lexicon: white paper of the ACR thyroid imaging, reporting and data system (TIRADS) committee. J Am Coll Radiol. 2015;12(12):1272–1279. 10.1016/j.jacr.2015.07.011. [PubMed: 26419308]

13. 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(5):587–595. 10.1016/j.jacr.2017.01.046. [PubMed: 28372962]

14. Silva CT, Navarro OM. Pearls and pitfalls in pediatric thyroid imaging. Semin Ultrasound, CT MRI. 2020;41(5):421–432. 10.1053/j.sult.2020.05.007.

15. Landis J, Koch G. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–174. [PubMed: 843571]

16. Martinez-Rios C, Daneman A, Bajno L, van der Kaay DCM, Moineddin R, Wasserman JD. Utility of adult-based ultrasound malignancy risk stratifications in pediatric thyroid nodules. Pediatr Radiol. 2018;48(1):74–84. 10.1007/s00247-017-3974-y. [PubMed: 28983667]

17. Kim PH, Yoon HM, Hwang J, et al. Diagnostic performance of adult-based ATA and ACR-TIRADS ultrasound risk stratification systems in pediatric thyroid nodules: a systematic review and meta-analysis. Eur Radiol. 2021;31(10):7450–7463. 10.1007/s00330-021-07908-8. [PubMed: 33864505]

18. Scappaticcio L, Maiorino MI, Iorio S, et al. Exploring the performance of ultrasound risk stratification systems in thyroid nodules of pediatric patients. Cancers (Basel). 2021;13(21):1–13. 10.3390/cancers13215304.

19. Basha MAA, Alnaggar AA, Refaat R, et al. The validity and reproducibility of the thyroid imaging reporting and data system (TI-RADS) in categorization of thyroid nodules: multicentre prospective study. Eur J Radiol. 2019;117:184–192. 10.1016/j.ejrad.2019.06.015. [PubMed: 31307646]

20. Lim-dunham JE, Toslak IE, Reiter M, Martin B. Assessment of the American College of Radiology Thyroid Imaging Reporting and Data System for thyroid nodule malignancy risk stratification in a pediatric population. Am J Radiol. 2019;212(1):188–194. 10.2214/AJR.18.20099.

21. Arora S, Khoury J, Trout AT, Chuang J. Improving malignancy prediction in AUS/FLUS pediatric thyroid nodules with the aid of ultrasound. Horm Res Paediatr. 2020;93(4):239–244. 10.1159/000509118. [PubMed: 32894855]

22. Richman DM, Cherella CE, Smith JR, et al. Clinical Utility of Sonographic Features in Indeterminate Pediatric Thyroid Nodules. Eur J Endocrinol. 2021:184(5):657–665. 10.1530/EJE-20-1480. [PubMed: 33635830]

23. Cherella CE, Hollowell ML, Smith JR, et al. Subtype of atypia on cytology and risk of malignancy in pediatric thyroid nodules. Cancer Cytopathol. 2022;130(5):330–335. 10.1002/cncy.22544. [PubMed: 35119774]

24. Baumgarten H, Jenks CM, Isaza A, et al. Bilateral papillary thyroid cancer in children: risk factors and frequency of postoperative diagnosis. J Pediatr Surg. 2020;55(6):1117–1122. 10.1016/j.jpedsurg.2020.02.040. [PubMed: 32171533]

25. Hoang JK, Middleton WD, Tessler FN. Update on ACR TI-RADS: successes, challenges, and future directions, from the AJR special series on radiology reporting and data systems. Am J Roentgenol. 2020;216(3):570–578. 10.2214/AJR.20.24608.

Fig. 1.
Receiver operating characteristic (ROC) curve for the sensitivity and specificity of American College of Radiology Thyroid Imaging Reporting and Data System (ACR TI-RADS) levels to predict malignancy.
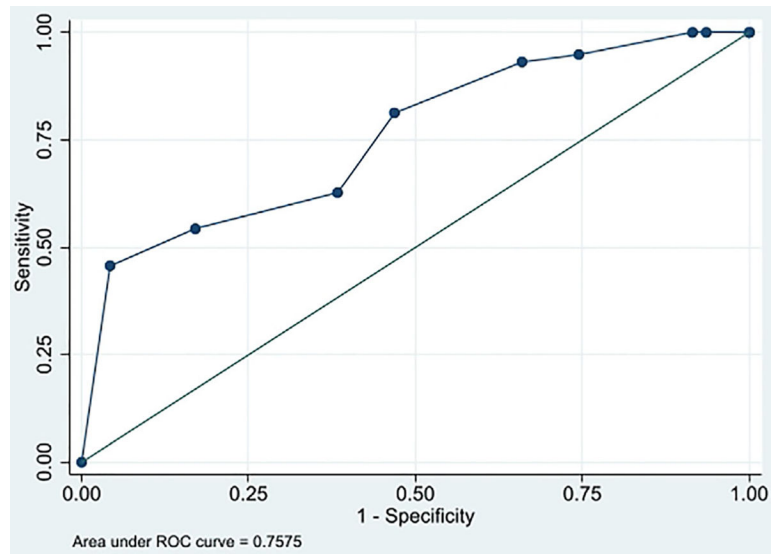
Fig. 2.
Receiver operating characteristic (ROC) curve for the sensitivity and specificity of American College of Radiology Thyroid Imaging Reportin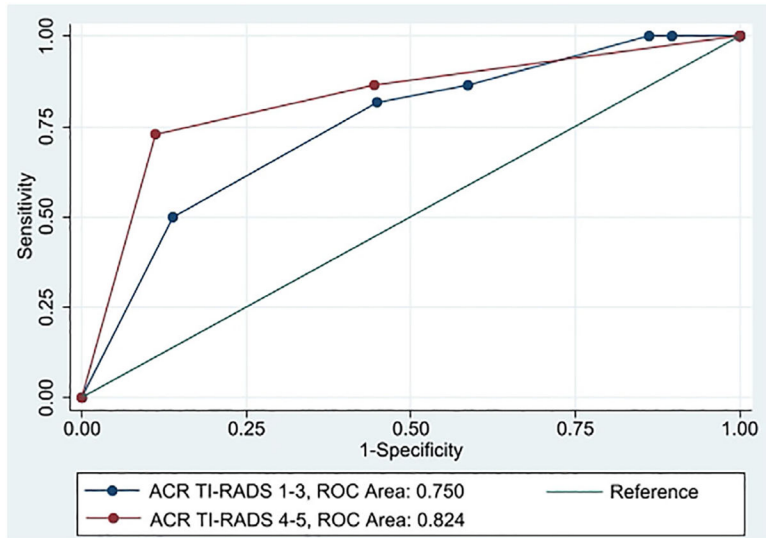g and Data System (ACR TI-RADS) levels to predict malignancy for nodules with ACR TI-RADS ≥4 (high-risk, red) compared with ACR TI-RADS <4 (lower risk, blue).

TABLE I.

Patient Demographics and Pathology.

	Overall (N = 106) n (%)	Benign (N = 47) n (%)	Malignant (N = 59) n (%)
Male	20 (18.9%)	10 (21.3%)	10 (16.9%)
Chronic lymphocytic thyroiditis	15 (14.2%)	2 (4.3%)	13 (22.0%)
FNA (Bethesda classification)	91	38	53
Inadequate (I)	1 (1.1%)	1 (2.6%)	0 (0.0%)
Benign (II)	22 (24.2%)	20 (52.6%)	2 (3.8%)
Indeterminate (III or IV)	36 (39.6%)	16 (42.1%)	20 (37.7%)
Suspicious or malignant (V or VI)	32 (35.2%)	1 (2.6%)	31 (58.5%)
Thyroidectomy[a]	91	32	59
Surgical pathology			
Benign	32 (35.2%)	32 (100%)	0 (0.0%)
FTC	4 (4.4%)	0 (0.0%)	4 (6.8%)
MTC	2 (2.2%)	0 (0.0%)	2 (3.4%)
NIFTP	2 (2.2%)	0 (0.0%)	2 (3.4%)
PTC	51 (56.0%)	0 (0.0%)	51 (86.4%)
Classic variant/not specified	28 (30.8%)	0 (0.0%)	28 (47.5%)
Follicular variant	19 (20.9%)	0 (0.0%)	19 (32.2%)
Other variant[b]	4 (4.4%)	0 (0.0%)	4 (6.8%)
	Median (range)	Median (range)	Median (range)
Age at ultrasound (years)	15.6 (0.9–18.8)	15.8 (6.4–18.8)	15.5 (0.9–18.7)

Abbreviations: FNA, Fine Needle Aspiration; FTC, Follicular Thyroid Carcinoma; MTC, Medullary Thyroid Carcinoma; NIFTP, Non-Invasive Follicular Thyroid neoplasm with Papillary-like nuclear features; PTC, Papillary Thyroid Carcinoma.

[a] Included lobectomy or total thyroidectomy.

[b] Two patients with cribriform morular variant of PTC, one with PTC with tall cell features, and one with diffuse sclerosing PTC.

TABLE II.

Inter-Rater Reliability.

	Overall			PTC			Indeterminate nodules		
	κ	SE	p	κ	SE	p	κ	SE	p
ACR TI-RADS points	0.547	0.057	**<0.001**	0.502	0.082	**<0.001**	0.494	0.097	**<0.001**
ACR TI-RADS level	0.576	0.066	**<0.001**	0.600	0.103	**<0.001**	0.507	0.106	**<0.001**
Subscale scores									
Composition	0.817	0.081	**<0.001**	0.778	0.140	**<0.001**	0.765	0.141	**<0.001**
Echogenicity	0.342	0.068	**<0.001**	0.314	0.106	**0.002**	0.180	0.095	**0.030**
Shape	0.148	0.075	**0.025**	0.135	0.071	**0.029**	0.000	0.000	0.500
Margin	0.466	0.085	**<0.001**	0.591	0.113	**<0.001**	0.321	0.147	**0.014**
Echogenic foci	0.682	0.078	**<0.001**	0.611	0.104	**<0.001**	0.705	0.132	**<0.001**
Largest dimension	0.870	0.063	**<0.001**	0.853	0.089	**<0.001**	0.923	0.108	**<0.001**
Vascularity	0.511	0.062	**<0.001**	0.481	0.085	**<0.001**	0.454	0.101	**<0.001**

Note: Bold indicates $p < 0.05$.

Abbreviations: ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System; PTC, Papillary Thyroid Carcinoma; SE, Standard Error.

TABLE III.

American College of Radiology Thyroid Imaging Reporting and Data System (ACR TI-RADS) Scores.

	Overall (N = 106) n (%)	Benign (N = 47) n (%)	Malignant (N = 59) n (%)	Malignant versus benign	
				OR (95% CI)	p
ACR TI-RADS level				2.48 (1.62–3.98)	**<0.001**
0.5–1	3 (2.8%)	3 (6.4%)	0 (0.0%)		
1.5–2	12 (11.3%)	9 (19.1%)	3 (5.1%)		
2.5–3	22 (20.8%)	14 (29.8%)	8 (13.6%)		
3.5–4	29 (27.4%)	13 (27.7%)	16 (27.1%)		
4.5–5	40 (37.7%)	8 (17.0%)	32 (54.2%)		
	Median (range)	Median (range)	Median (range)	z	p
ACR TI-RADS points	4.5 (0–13)	3.5 (0–11)	7 (2–13)	−4.56	**<0.001**
ACR TI-RADS level	4 (0–5)	3 (1–5)	4.5 (2–5)	−4.62	**<0.001**
Subscale scores					
Composition	2 (0–3)	1 (0–2)	2 (1–2)	−5.75	**<0.001**
Shape	0 (0–3)	0 (0–3)	0 (0–3)	−2.01	0.063
Margin	0 (0–3)	0 (0–1.5)	0 (0–3)	−3.45	**<0.001**
Echogenic foci	1 (1–5.5)	0 (0–4)	2 (0–5.5)	−3.00	**0.002**
Largest dimension (cm)	2.4 (0.3–8.5)	1.9 (0.3–5.7)	2.5 (0.3–8.5)	−0.963	0.338
	Mean (SD)	Mean (SD)	Mean (SD)	t (df)	p
Echogenicity	1.5 (0.6)	1.2 (0.6)	1.7 (0.6)	−3.95 (104)	**<0.001**

Note: Bold indicates *p* < 0.05.

Abbreviations: CI, Confidence interval; df, Degrees of Freedom; OR, Odds Ratio; SD, Standard Deviation.

TABLE IV.

Nodule Largest Dimension by Average American College of Radiology Thyroid Imaging Reporting and Data System (ACR TI-RADS) Level.

ACR TI-RADS level	Overall (*N* = 106) Nodule largest dimension (cm) Median (range)
0.5–1	1.1 (1.0–1.1)
1.5–2	3.0 (1.3–5.5)
2.5–3	2.1 (0.3–8.5)
3.5–4	2.3 (0.3–5.7)
4.5–5	2.6 (0.7–5.9)

TABLE V.

Receiver Operating Characteristic Curve Comparisons.

	Overall		PTC		Indeterminate nodules	
	AUC	p[a]	AUC	p[a]	AUC	p[a]
ACR TI-RADS points	0.758	0.970	0.784	0.692	0.534	0.799
ACR TI-RADS level	0.758		0.787		0.514	
Subscale scores						
Composition	0.782	0.539	0.792	0.903	0.67	0.124
Echogenicity	0.686	0.100	0.706	0.049	0.539	0.987
Shape	0.574	**<0.001**	0.564	**<0.001**	0.594	0.609
Margin	0.680	0.015	0.692	**0.008**	0.544	0.969
Echogenic foci	0.662	**0.001**	0.701	0.050	0.438	0.020
Largest dimension	0.555	**0.003**	0587	**0.005**	0.494	0.733

Note: Bold indicates significant after Holm's correction for multiple comparisons.

Abbreviations: ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System; AUC, Area Under the Curve.

[a] Compared with ACR TI-RADS leve.