Published in final edited form as:

J Ultrasound Med. 2019 July ; 38(7): 1807–1813. doi:10.1002/jum.14870.

TIRADS Interobserver Variability Among Indeterminate Thyroid Nodules:

A Single-Institution Study

Zeyad T. Sahli, MD, Ashwyn K. Sharma, BS, Joseph K. Canner, MHS, Farah Karipineni, MD, Osama Ali, MD, Satomi Kawamoto, MD, Jen-Fan Hang, MD, Aarti Mathur, MD, Syed Z. Ali, MD, FRC Path, FIAC, Martha A. Zeiger, MD, FACS, Sheila Sheth, MD

Endocrine Surgery, Department of Surgery (Z.T.S., A.K.S., J.K.C., F.K., A.M.), Department of Radiology (O.A., S.K., S.Z.A.), and Department of Pathology (J.F.H., S.Z.A., S.S.), The Johns Hopkins University School of Medicine, Baltimore, Maryland; Department of Surgery, The University of Virginia School of Medicine, Charlottesville, VA (Z.T.S., M.A.Z.); and Department of Pathology and Laboratory Medicine, Taipei Veterans General Hospital, Taipei, Taiwan (J.F.H.).

Abstract

Objectives—A high proportion of cytologically indeterminate, Afirma Gene Expression Classifier “suspicious” thyroid nodules are benign. The Thyroid Imaging Reporting and Data System (TIRADS), was proposed by the American College of Radiology in 2017 to help classify thyroid nodules based on ultrasound characteristics in a standardized fashion to guide management. We aim to determine the interobserver variability of TIRADS classification among cytologically indeterminate and Afirma suspicious nodules.

Methods—We retrospectively queried cytopathology archives for thyroid fine-needle aspiration specimens obtained between February 2012 and September 2016 with associated (1) indeterminate diagnosis, (2) ultrasound imaging at our institution, (3) Afirma suspicious result, and (4) surgery at our institution. We compared the TIRADS variability of the 3 blinded radiologists using intraclass correlation coefficients.

Results—Our cohort consisted of 127 nodules. Intraclass correlation coefficients can be interpreted as follows: less than 0.4, poor; 0.4 to 0.59, fair; 0.6 to 0.74, good; 0.75 to 1.00, excellent. The intraclass correlation coefficients of the raw TIRADS score and category variability was 0.561 (95% confidence interval [CI]: 0.464–0.651) or fair and 0.547 (95% CI, 0.449–0.640) or fair, respectively. When analyzing composition, echogenicity, shape, margin, and echogenic foci, the ICCs were 0.552 (95% CI, 0.454–0.643), fair; 0.533 (95% CI, 0.432–0.627), fair; 0.359 (95% CI, 0.248–0.469), poor; 0.192 (95% CI, 0.084–0.308), poor; and 0.549 (95% CI, 0.451–0.641), fair, respectively.

Conclusions—Our results show that among the subset of cytologically indeterminate and Afirma suspicious nodules, TIRADS interobserver variability was fair. Shape and margin criteria

were the biggest sources of disagreement. Large prospective studies are needed to evaluate the interobserver variability of TIRADS in this subset of thyroid nodules.

Keywords

Afirma; endocrine surgery; indeterminate nodules; interobserver variability; thyroid; TIRADS

With the recent increase in the number of thyroidectomies performed, it has become more important to identify benign nodules so as to reduce the number of unnecessary thyroidectomies performed. However, thyroid nodules can be difficult to accurately diagnose, and it remains challenging to distinguish between benign versus potentially malignant nodules based on ultrasound characteristics. Even with the increased diagnostic value of fine-needle aspiration (FNA), over 30% of thyroid FNAs, with wide institutional variability,^{1,2} are labeled with an indeterminate cytopathologic category only. In addition, despite the development of molecular marker panels, such as Afirma Gene Expression Classifier (GEC), it is still difficult to identify benign nodules, with benign surgical pathology found in more than half of those identified as Afirma “suspicious.”³ Therefore, additional and more robust means are needed to assist in identifying benign nodules.

Classification schemes for better differentiating benign versus malignant thyroid nodules have been proposed by multiple organizations, including the American Thyroid Association, American Association of Clinical Endocrinologists/American College of Endocrinology/Associazione Medici Endocrinologi, European Thyroid Association, and the Korean Society of Thyroid Radiology.^{4–6} After developing a lexicon to describe thyroid nodules, in 2017, the American College of Radiology proposed the Thyroid Imaging Reporting and Data System (TIRADS), another system to classify thyroid nodules and help guide further management and to determine when FNA is appropriate over active surveillance of nodules.⁷ TIRADS provides clinicians with evidence-based recommendations defined by sonographic features of a given nodule in order to improve patient management and cost-effectiveness by avoiding unnecessary FNAs and decreasing variation in the reporting of thyroid nodules.² TIRADS includes a standardized scoring system using 5 ultrasound characteristics—composition, echogenicity, shape, margin, and echogenic foci—and has 5 different categories (TR1–TR5), each level associated with an increased likelihood of malignancy and therefore more aggressive recommended clinical management.⁷

With any new classification system, reproducibility studies provide valuable information regarding the applicability of the criteria. Recently, our group compared retrospectively applied TIRADS score and surgical pathology among cytologically indeterminate, Afirma suspicious nodules using 1 radiologist. We found a poor association among this subset of patients. However, this result may have been impacted by the use of only 1 radiologist to evaluate ultrasound images. In addition, the extent of variability among radiologists using the TIRADS system is currently unknown. For this reason, we sought to assess the interobserver variability of TIRADS in classifying this group of thyroid nodules among 3 experienced radiologists using the same cohort.

MATERIALS AND METHODS

The current study was approved by the Johns Hopkins' Institutional Review Board. We retrospectively searched a prospectively maintained cytopathology database for consecutive thyroid FNA specimens obtained at our institution between February 2012 and September 2016. Thyroid nodules with an indeterminate diagnosis of atypia of unknown significance or suspicious for follicular neoplasm, available ultrasound imaging results, and Afirma GEC testing with a suspicious result were included. Patients who did not undergo surgery or did not have sonography done at our institution were excluded. Onsite evaluation of FNA samples was routinely performed by either an experienced cytotechnologist or an attending cytopathologist for each case. Afirma GEC was performed after the final cytologic diagnosis of either atypia of unknown significance or suspicious for follicular neoplasm was made. Additional information collected and recorded included patient demographics (age, sex, and race), history of Hashimoto disease, sonography, cytology and pathology reports, and clinic and operative notes.

Our primary end point was to assess the interobserver variability of TIRADS in classifying this subset of thyroid nodules among 3 experienced radiologists. All 3 radiologists were body fellowship trained with expertise in sonography including thyroid sonography.

Pathology Slide Re-review

For cases diagnosed as follicular-variant papillary thyroid cancer on surgical resection specimens, all histopathologic slides were retrieved and re-reviewed separately by 2 pathologists (a fellow and experienced attending) to identify tumors that would now be classified as noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP).

The diagnostic criteria for NIFTP were based on the new consensus guidelines and 20 cases of follicular-variant papillary thyroid cancer met the criteria.⁸ For the purposes of our statistical analysis, pathology was considered either benign or malignant. Due to NIFTP's indolent nature and current debate about whether it should be considered benign or malignant, we performed 2 analyses, with NIFTP considered as either benign or malignant.

Ultrasonography Image Review

For all cytologically indeterminate and Afirma suspicious thyroid nodules, ultrasound imaging was retrieved and retrospectively reviewed in a blinded fashion by 3 experienced radiologists (S.S., O.A., S.K.) at our institution to retrospectively classify the nodules according to the TIRADS classification. The nodules of interest were matched carefully according to the information regarding location and size recorded in the FNA and histopathologic reports. A search of the surgical pathology database was performed to identify the corresponding histopathologic diagnosis and surgical procedure. Ultrasound images that were not evaluated by all 3 radiologists were excluded (n = 30 nodules).

Statistical Analysis

We compared the results from the 3 radiologists using intraclass correlation coefficients (ICCs).⁹ The ICC is a measure of the consistency of different measurements on the same target and is particularly appropriate for study designs where there are more than 2 raters. We calculated ICCs according to a 2-way random effects model. In this model, thyroid ultrasound image is a random effect and radiologist is a fixed effect, as only 3 radiologists were used and each radiologist reviewed each image. Four parameters were analyzed: (1) raw TIRADS score, (2) TIRADS classification, (3) recommended management as a result of this score (referral for FNA or not), and (4) individual components of the TIRADS criteria. To practically interpret ICCs, guidelines have been developed by biostatisticians for determining levels of practical, substantive, or clinical significance. Per these guidelines, ICCs can be interpreted as follows: less than 0.40, poor agreement; 0.40 to 0.59, fair agreement; 0.60 to 0.74, good agreement; and 0.75 to 1.00, excellent agreement.¹⁰

The diagnostic accuracy of the 3 radiologists were compared using sensitivity, specificity, negative predictive value, and positive predictive value (PPV). In these analyses, the recommended management based on the TIRADS score was compared with the final pathology to derive the measurements of diagnostic accuracy.

All analyses were performed using STATA 14.2/MP (StataCorp LLC, College Station, TX).

RESULTS

Patient Characteristics

Excluding the 30 nodules not evaluated by the 3 radiologists, our cohort consisted of 127 nodules. The mean patient age was 52 years (range, 17–80 years) and the majority of patients were white (71.7%) and female (72.4%). Ninety-nine (78.0%) nodules were Bethesda III on FNA, and 28 (22.1%) were Bethesda IV. Table 1 lists the patient demographics of our study population. Twenty nodules had a final diagnosis of NIFTP. Considering NIFTP as benign, the majority of nodules had a benign pathology (101; 79.5%) as depicted in Table 2. The most common benign pathology was follicular adenoma (36; 28.3%). Malignant pathology was found in 26 nodules (20.5%). Among the malignancies, the most common pathology was classic papillary thyroid cancer (13; 10.2%). Conversely, classifying NIFTP as malignant, 63.8% of thyroid nodules were benign and 36.2% of thyroid nodules were malignant. The mean nodule size was 2.3 [notdef] 1.2 cm with a range from 0.5 to 8.0 cm.

Interobserver Variability

Table 3 lists the TIRADS raw score and classification among the 3 radiologists. TR4 and TR1 were the most and least common classifications, respectively, among the 3 radiologists. Table 4 includes all the ICCs. The raw TIRADS score was first analyzed among the 3 radiologists. The ICC was 0.561 (95% confidence interval [CI], 0.464–0.651) or fair agreement. Similarly, when analyzing TIRADS category variability, the ICC was 0.547 (95% CI, 0.449–0.640) or fair agreement. We also analyzed the individual components of the TIRADS criteria to determine the source of disagreement. When we looked at each of

the 5 TIRADS categories, the results were as follows: composition, 0.552 (95% CI, 0.454–0.643), fair agreement; echogenicity, 0.533 (95% CI, 0.432–0.627), fair agreement; shape, 0.359 (95% CI, 0.248–0.469), poor agreement; margin, 0.192 (95% CI, 0.084–0.308), poor agreement; echogenic foci, 0.549 (95% CI, 0.451–0.641), fair agreement.

Finally, when analyzing patient management as a result of TIRADS score, the ICC was 0.672 (95% CI, 0.590–0.745) or good agreement. The diagnostic value of TIRADS among each radiologist is presented in Tables 5 and 6. Considering NIFTP as either benign or malignant, the TIRADS sensitivity, specificity, PPV, and negative predictive value among the 3 radiologists ranged from 64.58% to 75.00%, 36.36% to 41.77%, 25.00% to 41.77%, and 63.83% to 83.72%, respectively in predicting malignancy.

DISCUSSION

Sonography is the best imaging modality to characterize thyroid nodules. While thyroid nodules are exceedingly common, most thyroid nodules are benign. It is well established that size alone does not allow differentiation between benign and malignant nodules, and ultrasound characteristics of nodules are more helpful in determining which nodules should undergo FNA. However, not only is there an overlap in the ultrasound appearance of benign and malignant nodules, but there has also been great variability in thyroid nodule reporting and recommendation for further workup. The aim of TIRADS was to decrease variation in the reporting of thyroid nodules and thereby improve patient management and cost-effectiveness by minimizing the number of unnecessary FNAs. Previous studies have demonstrated that TIRADS classification systems have good interobserver variability.^{11,12} However, some suggest that intermediate categories have poor interobserver variabilities compared to categories at the extremes of the spectrum.¹¹ To the best of the authors' knowledge, this study represents the first investigation in the literature evaluating interobserver variability among TIRADS classification in patients with cytologically indeterminate, Afirma suspicious thyroid nodules.

Our study found that overall, the interobserver variability among the 3 radiologists was fair when comparing classification designations and nodule characteristics in cytologically indeterminate and Afirma suspicious nodules. We also investigated the interobserver variability for individual ultrasound characteristics to better identify possible sources of variability in the interpretation of thyroid ultrasound images. We found the poorest interobserver agreement when evaluating for the shape and margin of thyroid nodules, with ICC values of 0.36 and 0.19, respectively. Park et al¹³ examined the interobserver variability among 3 radiologists for 400 thyroid nodules and found that composition had the highest agreement, with a kappa value of 0.818, and margins and shape demonstrated the lowest agreement, with kappa values of 0.420 and 0.330, respectively, indicating fair agreement. Similarly, another study reported moderate agreement with regard to shape, echogenicity, calcification, and diagnostic categories (kappa = 0.42, 0.57, 0.55, and 0.55, respectively), and fair agreement for margin, echotexture, and capsule invasion (kappa = 0.34, 0.26, and 0.32, respectively).¹⁴

The result of poor agreement among the 3 readers regarding nodule shape is somewhat surprising and differs from that reported by Park et al.¹⁴ This could be due to the presence of only 2 TIRADS categories, “wider than tall” and “taller than wide,” which leaves ambiguity for round nodules. Experienced radiologists know that suspicious nodules often have a combination of sonographic findings. Among individual nodules, features such as echogenicity or composition might have subconsciously swayed classification of round nodules in 1 of the 2 categories for shape. Another potential explanation is that the images were acquired before the TIRADS protocol was implemented and anterior-posterior measurements of the nodules were obtained on the sagittal rather than the transverse image. With future refinements of TIRADS classification, expanding the “wider than tall” category to include round nodules might help decrease variability with regard to shape.

Interestingly, previous reports have demonstrated that irregular margins have the highest PPV (94.45%) for malignancy.¹⁵ The results of our study, along with previous literature, suggest that margins can be difficult to differentiate, particularly between ill-defined margins and irregular/lobulated margins. One report demonstrated pairing trainees with more experienced radiologists to perform ultrasound imaging and interpret thyroid results together only slightly increased interobserver variability scores from moderate agreement (kappa range, 0.32–0.42) to substantial agreement (kappa range, 0.47–0.64) when interpreting thyroid nodule margins.¹⁶

Despite the differences in interpretation of specific ultrasound features, our results also demonstrated that there was “good” interobserver agreement when it pertained to clinical management and when the imaging results indicated whether an FNA biopsy was warranted. Similarly, Grani et al¹⁷ performed an analysis comparing the interobserver variability of different classification systems. The authors found that while the interobserver agreement values for individual ultrasound features were varied, with individual kappa values ranging from 0.11 to 0.89, the agreement on indications for FNA biopsy based on any particular guideline were substantially better, with individual kappa values ranging from 0.61 to 0.91, depending on the specific guideline used. Interestingly, the American College of Radiology TIRADS classification system demonstrated the lowest interobserver agreement, with a Krippendorff alpha of 0.61 (0.5–0.72). These results may be due to the fact that while radiologists may disagree on specific sonographic interpretation, the combination of suspicious features is more likely to warrant an FNA biopsy. Nodules that are likely to be scored higher on TIRADS are more likely to have more suspicious features, and therefore disagreements in specific features may not translate toward differences in final FNA biopsy recommendations. Furthermore, despite the variation in TIRADS scores and individual component scores, it is interesting to note that all 3 radiologists had remarkably low variation in specificity, negative predictive value, and PPV, regardless of whether NIFTP was considered benign or malignant, further illustrating that despite the disagreements in minor features, the overall impressions of radiologists were remarkably similar.

Our study has several limitations. First, our report is a single-institution study. An inherent weakness in this study also is its retrospective nature. Moreover, our results cannot be generalized to all TIRADS imaging, as our cohort is composed of only cytologically indeterminate, Afrima suspicious nodules.

CONCLUSION

In conclusion, our results show that among the subset of cytologically indeterminate and Afirma suspicious nodules, TIRADS interobserver variability was fair, which brings into question its efficacy. Larger prospective studies are needed to evaluate the interobserver variability of TIRADS in this subset of thyroid nodules.

Abbreviations

CI	confidence interval
FNA	fine-needle aspiration
GEC	Gene Expression Classifier
ICC	intraclass correlation coefficients
NIFTP	noninvasive follicular thyroid neoplasm with papillary-like nuclear features
PPV	positive predictive value
TIRADS	Thyroid Imaging Reporting and Data System

References

1. Kholová I, Ludvíková M. Thyroid atypia of undetermined significance or follicular lesion of undetermined significance: an indispensable Bethesda 2010 diagnostic category or waste garbage? *Acta Cytologica* 2014; 58:319–329. [PubMed: 25195864]
2. Straccia P, Rossi ED, Bizzarro T, et al. A meta-analytic review of the Bethesda System for Reporting Thyroid Cytopathology: Has the rate of malignancy in indeterminate lesions been underestimated? *Cancer Cytopathol* 2015; 123:713–722. [PubMed: 26355876]
3. Alexander EK, Kennedy GC, Baloch ZW, et al. Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. *N Engl J Med* 2012; 367:705–715. [PubMed: 22731672]
4. Gharib H, Papini E, Garber JR, et al. American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi Medical Guidelines for Clinical practice for the diagnosis and management of thyroid nodules–2016 update. *Endocr Pract* 2016; 22 (suppl 1):1–60.
5. 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. [PubMed: 26462967]
6. Shin JH, Baek JH, Chung J, et al. Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean Society of Thyroid Radiology consensus statement and recommendations. *Korean J Radiol* 2016; 17:370–395. [PubMed: 27134526]
7. 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. [PubMed: 28372962]
8. Nikiforov YE, Seethala RR, Tallini G, et al. Nomenclature revision for encapsulated follicular variant of papillary thyroid carcinoma: a paradigm shift to reduce overtreatment of indolent tumors. *JAMA Oncol* 2016; 2:1023–1029. [PubMed: 27078145]
9. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996; 1:30–46.

10. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 1994; 6:284–290.
11. Chandramohan A, Khurana A, Pushpa B, Manipadam MT, Naik D, Thomas N, et al. Is TIRADS a practical and accurate system for use in daily clinical practice? *Indian J Radiol Imaging* 2016; 26:145–152. [PubMed: 27081240]
12. Russ G, Royer B, Bigorgne C, Rouxel A, Bienvenu-Perrard M, Leenhardt L. Prospective evaluation of thyroid imaging reporting and data system on 4550 nodules with and without elastography. *Eur J Endocrinol* 2013; 168:649–655. [PubMed: 23416955]
13. Park S, Park S, Choi Y, et al. Interobserver variability and diagnostic performance in US assessment of thyroid nodule according to size. *Ultraschall Med–Eur J Ultrasound* 2012; 33:E186–E190.
14. Park CS, Kim SH, Jung SL, et al. Observer variability in the sonographic evaluation of thyroid nodules. *J Clin Ultrasound* 2010; 38:287–293. [PubMed: 20544863]
15. Srinivas MNS, Amogh V, 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. [PubMed: 27014501]
16. Kim HG, Kwak JY, Kim E-K, Choi SH, Moon HJ. Man to man training: can it help improve the diagnostic performances and interobserver variabilities of thyroid ultrasonography in residents? *Eur J Radiol* 2012; 81:e352–e356. [PubMed: 22137098]
17. 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. [PubMed: 29196301]

Table 1.**Clinical and Demographic Features of Patient Cohort**

Age	
Mean (SD, range)	52 (14, 17–80)
Race, no. (%)w	
Asian	2 (1.6)
Black	21(16.5)
Hispanic	1 (0.8)
White	91 (71.7)
Other	12 (9.4)
Sex, n (%)	
Female	92 (72.4)
Male	35 (27.6)
Hashimoto thyroiditis, n (%)	10 (7.9)
Nodules	
Size on ultrasonography, cm Mean (range)	2.3 (0.5 – 8.0)
Nodule size, number of nodules (%)	
<1 cm	3 (2.4)
1.0–1.9 cm	56 (44.1)
2.0–2.9 cm	36 (28.4)
3.0–3.9 cm	19 (15.0)
4.0 cm	13 (10.2)
Bethesda category	
III	99 (78.0)
IV	28 (22.1)

Table 2.**Surgical Pathology of Index Thyroid Nodule**

Surgical Pathology	Number of Nodules (%)
Nodular hyperplasia	17 (13.4)
Adenomatoid nodule	25 (19.7)
Follicular adenoma	36 (28.3)
Hurthle cell adenoma	1 (0.8)
Lymphocytic thyroiditis	2 (1.6)
NIFTP	20 (15.7)
Papillary carcinoma, conventional	13 (10.2)
Papillary carcinoma, follicular variant	7 (5.5)
Follicular carcinoma	5 (3.9)
Hürthle cell carcinoma	1 (0.8)

NIFTP indicates noninvasive follicular thyroid neoplasm with papillary-like nuclear features.

Table 3.

TIRADS Score and Classification According to Radiologist

TIRADS Classification	Radiologist 1	Radiologist 2	Radiologist 3
Raw Score			
0	3 (2.36)	0 (0.00)	1 (0.79)
1	1 (0.79)	0 (0.00)	1 (0.79)
2	4 (3.15)	7 (5.51)	10 (7.87)
3	25 (19.69)	33 (25.98)	23 (18.11)
4	38 (29.92)	42 (33.07)	59 (46.46)
5	12 (9.45)	14 (11.02)	6 (4.72)
6	13 (10.24)	10 (7.87)	11 (8.66)
7	18 (14.17)	9 (7.09)	11 (8.66)
8	3 (2.36)	2 (1.57)	2 (1.57)
9	5 (3.94)	3 (2.36)	1 (0.79)
10	5 (3.94)	6 (4.72)	1 (0.79)
11	0 (0.00)	0 (0.00)	0 (0.00)
12	0 (0.00)	0 (0.00)	1 (0.79)
13	0 (0.00)	1 (0.79)	0 (0.00)
Category			
TR 1	4 (3.15)	0 (0.00)	2 (1.57)
TR 2	4 (3.15)	7 (5.51)	10 (7.87)
TR 3	25 (19.69)	33 (25.98)	23 (18.11)
TR 4	63 (49.61)	66 (51.97)	76 (59.84)
TR 5	31 (24.41)	21 (16.54)	16 (12.60)

Table 4.

TIRADS Criteria Intraclass Correlations (ICC)

	Management Decision	Raw Score	TIRADS Classification	Composition	Echogenicity	Shape	Margin	Echogenic Foci
ICC	0.672	0.561	0.548	0.552	0.533	0.359	0.192	0.549

Table 5.

TIRADS Diagnostic Value of Each Radiologist (NIFTP Benign)

	Sensitivity	Specificity	PPV	NPV
Radiologist 1	75.00	36.36	25.00	83.72
Radiologist 2	71.43	39.39	25.00	82.98
Radiologist 3	71.43	40.40	25.32	83.33

NIFTP indicates noninvasive follicular thyroid neoplasm with papillary-like nuclear features; NPV, negative predictive value; and PPV, positive predictive value.

Table 6.

TIRADS Diagnostic Value of Each Radiologist (NIFTP Malignant)

	Sensitivity	Specificity	PPV	NPV
Radiologist 1	72.92	37.97	41.67	69.77
Radiologist 2	64.58	37.97	38.75	63.83
Radiologist 3	68.75	41.77	41.77	68.75

NIFTP indicates noninvasive follicular thyroid neoplasm with papillary-like nuclear features; NPV, negative predictive value; and PPV, positive predictive value.