This copy is for personal use only. To order printed copies, contact *reprints@rsna.org*

Radiology

Reduction in Thyroid Nodule Biopsies and Improved Accuracy with American College of Radiology Thyroid Imaging Reporting and Data System[1]

Jenny K. Hoang, MBBS, MHS
William D. Middleton, MD
Alfredo E. Farjat, PhD
Jill E. Langer, MD
Carl C. Reading, MD
Sharlene A. Teefey, MD
Nicole Abinanti, MD
Fernando J. Boschini, MD
Abraham J. Bronner, MD
Nirvikar Dahiya, MD
Barbara S. Hertzberg, MD
Justin R. Newman, MD
Daniel Scanga, MD
Robert C. Vogler, MD
Franklin N. Tessler, MD

Purpose: To compare the biopsy rate and diagnostic accuracy before and after applying the American College of Radiology (ACR) Thyroid Imaging Reporting and Data System (TI-RADS) criteria for thyroid nodule evaluation.

Materials and Methods: In this retrospective study, eight radiologists with 3–32 years experience in thyroid ultrasonography (US) reviewed US features of 100 thyroid nodules that were cytologically proven, pathologically proven, or both in December 2016. The radiologists evaluated nodule features in five US categories and provided biopsy recommendations based on their own practice patterns without knowledge of ACR TI-RADS criteria. Another three expert radiologists served as the reference standard readers for the imaging findings. ACR TI-RADS criteria were retrospectively applied to the features assigned by the eight radiologists to produce biopsy recommendations. Comparison was made for biopsy rate, sensitivity, specificity, and accuracy.

Results: Fifteen of the 100 nodules (15%) were malignant. The mean number of nodules recommended for biopsy by the eight radiologists was 80 ± 16 (standard deviation) (range, 38–95 nodules) based on their own practice patterns and 57 ± 11 (range, 37–73 nodules) with retrospective application of ACR TI-RADS criteria. Without ACR TI-RADS criteria, readers had an overall sensitivity, specificity, and accuracy of 95% (95% confidence interval [CI]: 83%, 99%), 20% (95% CI: 16%, 25%), and 28% (95% CI: 21%, 37%), respectively. After applying ACR TI-RADS criteria, overall sensitivity, specificity, and accuracy were 92% (95% CI: 68%, 98%), 44% (95% CI: 33%, 56%), and 52% (95% CI: 40%, 63%), respectively. Although fewer malignancies were recommended for biopsy with ACR TI-RADS criteria, the majority met the criteria for follow-up US, with only three of 120 (2.5%) malignancy encounters requiring no follow-up or biopsy. Expert consensus recommended biopsy in 55 of 100 nodules with ACR TI-RADS criteria. Their sensitivity, specificity, and accuracy were 87% (95% CI: 48%, 98%), 51% (95% CI: 40%, 62%), and 56% (95% CI: 46%, 66%), respectively.

Conclusion: ACR TI-RADS criteria offer a meaningful reduction in the number of thyroid nodules recommended for biopsy and significantly improve the accuracy of recommendations for nodule management.

©RSNA, 2018

Online supplemental material is available for this article.

[1] From the Department of Radiology, Duke University Medical Center, Box 3808, Erwin Rd, Durham, NC 27710 (J.K.H., B.S.H.); Mallinckrodt Institute of Radiology, Washington University School of Medicine, St Louis, Mo (W.D.M., S.A.T.); Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC (A.E.F.); Department of Radiology, The Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pa (J.E.L.); Department of Radiology, Mayo Clinic College of Medicine, Rochester Minn (C.C.R.); Mecklenburg Radiology Associates, Charlotte, NC (N.A., D.S.); Rocky Mountain Radiologists, Denver, Colo (F.J.B.); Radiology Partners Research Institute, El Segundo, Calif (A.J.B.); Department of Radiology, Division of Ultrasound, Mayo Clinic, Phoenix, Ariz (N.D.); Memphis Radiological Professional Corporation, Methodist Le Bonheur Healthcare Memphis, Germantown, Tenn (J.R.H.); Duke Radiology of Raleigh, Duke University School of Medicine, Duke Raleigh Hospital, Raleigh, NC (R.C.V.); and Department of Radiology, University of Alabama at Birmingham, Birmingham, Ala (F.N.T.). Received November 3, 2017; revision requested December 12; revision received December 20; accepted January 2, 2018. **Address correspondence to** J.K.H. (e-mail: *jennykh@gmail.com*).

©RSNA, 2018

Ultrasonography (US) is the most accurate imaging modality with which to evaluate thyroid nodules (1). Some nodules are referred for biopsy or follow-up because their US features are associated with a higher malignancy risk. Although there are multiple risk stratification systems that can be used to guide management, they are not applied consistently by all radiologists, who may base recommendations on personal experience or local practice patterns (2–5). Recently, the American College of Radiology (ACR) Thyroid Imaging Reporting and Data System (TI-RADS) was published to provide radiologists with an easy-to-use US-based method to classify thyroid nodules (6). In the ACR TI-RADS, a nodule is assigned points in five US categories to determine its TI-RADS level, with the decision to perform biopsy or follow-up based on the maximum nodule diameter.

The overall aim of ACR TI-RADS is to reduce unnecessary biopsies and to detect the thyroid malignancies that are more likely to cause patient harm. When compared with other guidelines, such as those from the American Thyroid Association (ATA), ACR TI-RADS guidelines use higher size thresholds for biopsy of less suspicious nodules and do not recommend biopsy of nodules with benign characteristics, regardless of size (2–5).

The effect of ACR TI-RADS criteria on current practice in regard to diagnostic accuracy and missed malignancies is unknown. The aim of this study was to compare the biopsy rate and diagnostic accuracy before and after applying ACR TI-RADS risk stratification. We hypothesized that there would be a reduction in recommendations for biopsy among radiologists with application of ACR TI-RADS compared with their current practice patterns.

Implications for Patient Care

- Use of American College of Radiology (ACR) Thyroid Imaging Reporting and Data System (TI-RADS) criteria reduces the number of thyroid nodule biopsies recommended when compared with that recommended with radiologists' existing practice patterns and the American Thyroid Association (ATA), Korean TI-RADS, and French TI-RADS guidelines.
- Most of the nodules spared from biopsy with ACR TI-RADS criteria were benign; specificity improved from 20% (152 of 860; 95% confidence interval [CI]: 16%, 25%) to 44% (324 of 680; 95% CI: 33%, 56%) with ACR TI-RADS criteria.
- Only three of 120 (2.5%) malignancy encounters underwent no follow-up or biopsy with ACR TI-RADS criteria, primarily because of incorrect assignment of spongiform or mixed cystic and solid composition instead of solid composition; further education should be directed toward the assignment of thyroid nodule composition at US.

Materials and Methods

Study Population

The study cohort comprised 100 nodules in 92 patients who underwent fine-needle aspiration biopsy with definitive cytology results (Bethesda category II or VI) or surgical resection between April 2009 and May 2010 at one institution. These were nodules that the institution had contributed to a multi-institutional study (7) and that were selected as the first 100 consecutive nodules (newest to oldest). The only exclusion criterion was the absence of a dedicated video clip of the biopsied nodule. The images were anonymized and uploaded to an ACR Web-based portal. The image interpretation for this study was conducted in December 2016 before the publication of ACR TI-RADS guidelines. The study was approved by the institutional review board of the institution that provided the sonograms.

The US examinations were performed by using a variety of commercially available units equipped with 5–15-MHz linear array transducers. In all cases, images of the biopsied nodules were obtained in transverse and longitudinal planes. Video clips of the biopsied nodules were obtained in at least one plane.

Image Interpretation and Feature Assignments

A total of 11 radiologists from nine different institutions evaluated the nodules on the ACR portal. The readers were blinded to the pathology results. Three expert readers (J.E.L., C.C.R., F.N.T.) were on the ACR TI-RADS committee and had between 26 and 34 years of posttraining experience. They interpreted the sonograms independently, and their consensus was used as the "truth" for the nodule imaging features.

The other eight radiologists were test readers who had no knowledge of ACR TI-RADS management recommendations. All reported thyroid US in their clinical practice. This group comprised *(a)* two radiologists in academic practice with subspecialty training in US and 20 (N.D.) and 32 (B.S.H.) years of practical experience and *(b)* six radiologists in private practice (N.A., F.J.B, A.J.B, J.R.N., D.S., R.C.V.) with fellowship training in neuroradiology, women's imaging, and nuclear medicine, with a median of 13 years in practice (range, 3–32 years).

https://doi.org/10.1148/radiol.2018172572

Content code: HN

Radiology 2018; 287:185–193

Abbreviations:
ACR = American College of Radiology
ATA = American Thyroid Association
CI = confidence interval
TI-RADS = Thyroid Imaging Reporting and Data System

Author contributions:
Guarantor of integrity of entire study, J.K.H.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, J.K.H., W.D.M., A.E.F., J.E.L., A.J.B., J.R.N., R.C.V.; clinical studies, J.K.H., W.D.M., S.A.T., A.J.B., N.D., R.C.V., F.N.T.; statistical analysis, J.K.H., A.E.F., J.R.N., R.C.V.; and manuscript editing, J.K.H., W.D.M., A.E.F., J.E.L., S.A.T., F.J.B., A.J.B., N.D., B.S.H., J.R.N., R.C.V., F.N.T.

Conflicts of interest are listed at the end of this article.

See also the article by Tessler et al in this issue.

Figure 1

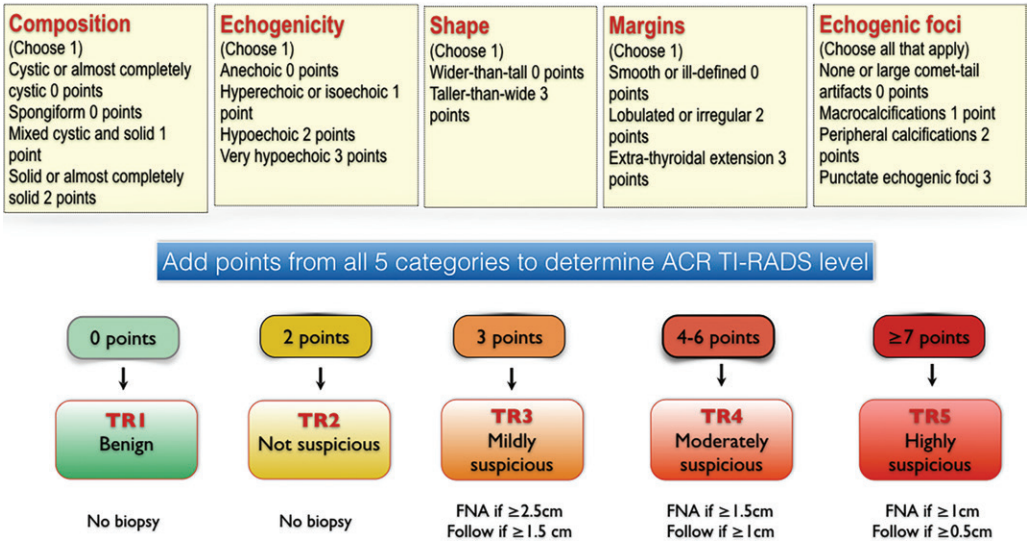

Figure 1: American College of Radiology Thoracic Imaging Reporting and Data System risk stratification system and management recommendations. *FNA* = fine-needle aspiration.

All the radiologists assessed the nodules for the five feature categories in the ACR TI-RADS lexicon (composition, echogenicity, shape, margin, and echogenic foci) (Fig 1) after reviewing two to four static US images and one or two video images of the same nodule. There were no Doppler US images. No clinical information about the patient was provided. The test readers also assigned a malignancy risk that matched the five risk stratification levels used in the ACR TI-RADS guidelines (highly suspicious, moderately suspicious, mildly suspicious, not suspicious, or benign) and stated whether they would recommend biopsy based on their own individual practice pattern. Test readers were asked for their reasons for biopsy after completing image interpretation. All readers stated that they followed a set of guidelines, which included those from the ATA (five readers), Society of Radiologists in Ultrasound (two readers), and guidelines developed by their institution (one reader). All readers also stated that they would deviate from guidelines on the basis of their personal experience, if needed.

Expert and test readers' feature assignments for nodules and maximum nodule size (irrespective of the acquisition plane) were then used to retrospectively assign an ACR TI-RADS risk stratification level and biopsy recommendation (Fig 1). ATA and Korean and French TI-RADS guidelines were retrospectively applied (2,4,5).

Statistical Analyses

An a priori power analysis was not performed. The cytology and pathology results served as the reference standard for malignancy. A comparison was made between the test readers' recommendations with and without ACR TI-RADS guidelines. Sensitivity, specificity, and accuracy with 95% confidence intervals were calculated. To account for the clustering effect of nodules in patients, a generalized linear mixed-effects model was used to estimate performance statistics along with corresponding measures of uncertainty.

ACR TI-RADS guidelines were also compared with other biopsy guidelines. Sensitivity, specificity, and accuracy of different biopsy guidelines were estimated by the expert readers in consensus. For the comparisons, generalized linear mixed-effect models were used with patients taken as random effect to account for clustering of nodules in patients. The models included a categorical variable as a fixed effect to indicate the type of biopsy guideline. Then, each of the guidelines was compared with ACR TI-RADS as the reference guideline with the *F* test.

The McNemar test was used to determine whether the ACR TI-RADS guidelines produced a significant change in the biopsy recommendations when compared with biopsy recommendations without ACR TI-RADS guidelines for test readers and when compared with other guidelines for expert consensus. Thus, the null hypothesis of marginal homogeneity between the pair of guidelines was tested.

A comparison was made for the distribution of assigned risk stratification levels with and without ACR TI-RADS guidelines for test readers and for test readers with ACR TI-RADS guidelines compared with expert consensus. Histograms with overlying bars were used to display results, and the shape of the histograms was compared with the Kolmogorov-Smirnov test.

In all cases, the threshold for statistical significance was set at $\alpha = .05$. Statistical analyses were performed by using R 2016 software (R Core Team, Vienna, Austria) and SAS 2015 statistical software (SAS Institute, Cary, NC).

Results

Study Population

Mean patient age was 52 years ± 14 (standard deviation) (range, 19–82 years). Mean nodule size was 2.7 cm ± 1.3 (range, 0.7–5.9 cm). There were 15 (15%) malignancies in 100 nodules, with a mean size of 2.2 cm ± 1.3, of which 11 (73%) were papillary carcinomas and four (27%) were the follicular variant of papillary carcinoma. One patient had two malignancies. The other 13 patients had one malignancy each. Table 1 summarizes patient demographics and US findings based on the expert consensus interpretation in the 100 nodules.

Test Reader Biopsy Recommendations with and without ACR TI-RADS

The mean number of nodules recommended for biopsy for the eight test readers was 80 ± 16 (standard deviation) (80%) with their own practice patterns and 57 ± 11 (57%) with ACR TI-RADS guidelines. This represents a 29% reduction in the number of biopsy recommendations with ACR TI-RADS guidelines ($P < .001$). Figure 2 shows an example of a benign nodule that had changed biopsy recommendations after application of ACR TI-RADS guidelines. Table 2 shows individual readers' biopsy recommendation rates and diagnostic accuracy. The test readers had a reduction in the number of biopsy recommendations after they applied ACR TI-RADS guidelines that ranged from 5% to 41%. Without ACR TI-RADS guidelines, they had overall sensitivity, specificity, and accuracy of 95% (95% confidence interval [CI]: 83%, 99%), 20% (95% CI: 16%, 25%), and 28% (95% CI: 21%, 37%), respectively. After applying ACR TI-RADS guidelines to the same feature assignments, overall sensitivity, specificity, and accuracy were 92% (95% CI: 68%, 98%), 44% (95% CI: 33%, 56%), and 52% (95% CI: 40%, 63%), respectively. The differences in mean specificity and accuracy for test readers were statistically significant, but 95% intervals did overlap for several individual readers (Table 2).

Table 1

Frequency of US Findings in 100 Thyroid Nodules Based on Expert Consensus

Findings	All Nodules ($n = 100$)	Benign ($n = 85$)	Malignant ($n = 15$)
Mean age (y)*	52 ± 14 (19–82)	54 ± 14 (26–82)	46 ± 13 (19–68)
Mean size (cm)*	2.7 ± 1.3 (0.7–5.9)	2.8 ± 1.3 (0.8–5.9)	2.2 ± 1.3 (0.7–5.8)
Composition			
Cystic	1 (1)	1 (1)	0
Spongiform	13 (13)	13 (15)	0
Cystic and solid	20 (20)	20 (24)	0
Solid	66 (66)	51 (60)	15 (100)
Echogenicity			
Anechoic	0	0	0
Hyperechoic	12 (14)	9 (11)	3 (20)
Isoechoic	52 (52)	51 (60)	1 (7)
Hypoechoic	33 (33)	24 (28)	9 (60)
Very hypoechoic	3 (3)	1 (1)	2 (13)
Shape			
Taller-than-wide	8 (8)	4 (5)	4 (27)
Not taller-than-wide	92 (92)	81 (95)	11 (77)
Margin			
Smooth	50 (50)	47 (55)	3 (20)
Ill defined	23 (23)	23 (27)	0
Irregular or lobulated	20 (20)	15 (18)	5 (33)
Extrathyroidal extension	7 (7)	0	7 (47)
Echogenic foci†			
No echogenic foci	62 (62)	60 (71)	2 (13)
Large comet tail	5 (5)	4 (5)	1 (1)
Macrocalcifications	19 (19)	14 (16)	5 (33)
Peripheral	3 (3)	0	3 (20)
Punctate	21 (21)	12 (14)	9 (60)
ACR TI-RADS level			
TR1	14 (14)	14 (16)	0
TR2	14 (14)	14 (16)	0
TR3	24 (24)	24 (28)	0
TR4	25 (25)	21 (25)	4 (27)
TR5	23 (23)	12 (14)	11 (73)

Note.—Unless otherwise indicated, data are number of nodules, and data in parentheses are percentages.

* Data are mean ± standard deviation, and data in parentheses are the range.

† Nodules could have more than one type of echogenic foci.

On the basis of the test readers' feature assignments, there were five malignancies that were not recommended for biopsy by some of the radiologists after ACR TI-RADS guidelines were retrospectively applied (Table 3). Although these malignancies did not meet the size criteria for biopsy, the majority of them met the size threshold for follow-up. Only three (2.5%) of 120 malignancy encounters (15 malignancies interpreted among the eight readers) underwent no follow-up or biopsy. The three malignancy encounters were for two malignancies and were based on incorrect categorization of signs by two of the eight test readers. One was a 1.9-cm TR5 solid nodule that was incorrectly categorized as spongiform in composition (TR1) by reader 2 and as an isoechoic mixed cystic and solid nodule without suspicious features (TR2) by reader 8. The expert consensus interpreted this nodule as solid, very hypoechoic, and taller than wide, with irregular margins and punctate echogenic foci (TR5). The other nodule was a solid 1.2-cm

Figure 2

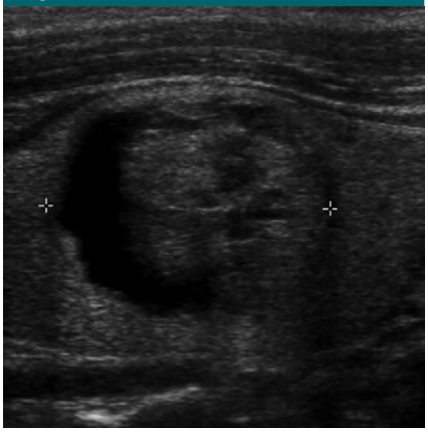

Figure 2: Transverse US image of a benign thyroid nodule with mixed solid cystic composition and a maximum diameter of 2.4 cm. Without ACR TI-RADS criteria, six of eight test readers recommended biopsy. After application of ACR TI-RADS criteria, three readers recommended biopsy and five recommended no biopsy or follow-up. Without ACR TI-RADS criteria, the risk stratification level was deemed benign by one reader, mildly suspicious by five readers, moderately suspicious by one reader, and highly suspicious by one reader. With ACR TI-RADS guidelines, the risk stratification level was deemed benign by five readers and moderately suspicious by three readers.

TR5 nodule that was misclassified by reader 8 as spongiform (TR1) (Fig 3).

Figure 4a shows the distribution of the five risk stratification levels with and without ACR TI-RADS guidelines for test readers. The majority of nodules were categorized by the test readers as mildly suspicious without ACR TI-RADS guidelines. After using the feature assignments to retrospectively assign ACR TI-RADS risk levels, most nodules were highly or moderately suspicious. The histogram shapes were significantly different from each other ($P < .001$). Figure 4b shows the five risk stratification levels for expert consensus, which was not significantly different from that of the test readers after application of ACR TI-RADS guidelines ($P = 0.50$).

Comparison with Other Biopsy Guidelines

If test readers had applied ATA, Korean TI-RADS, and French TI-RADS guidelines, the mean number of nodules recommended for biopsy would have been 77 (77%), 85 (85%), and 74 (74%), respectively, compared with 57 (57%) nodules with ACR TI-RADS guidelines. Five of eight test readers recommended more biopsies based on their experience than would have been recommended with the ATA biopsy guidelines.

Table 3 compares ACR TI-RADS and ATA recommendations for the five malignancies not recommended for biopsy by some of the test readers. As described previously, two malignancies would not have undergone biopsy or follow-up with ACR TI-RADS guidelines. Per the ATA guidelines, readers 2 and 8 also did not recommend biopsy of these two nodules on the basis of their feature assignments.

Expert consensus found that 55 of 100 nodules (55%) warranted fine-needle aspiration biopsy according to ACR TI-RADS criteria compared with 83 (83%), 84 (84%), and 70 (70%) nodules based on ATA, Korean TI-RADS, and French TI-RADS recommendations, respectively ($P < .001$) (Table E1 [online]). The two malignancies that did not meet the criteria for biopsy according to ACR TI-RADS guidelines also did not meet the criteria for biopsy with any of the other guidelines. Sensitivity was 87% (95% CI: 48, 98) for all guidelines ($P > .99$). The specificity of ACR TI-RADS guidelines (51%; 95% CI: 40, 62) was significantly higher than that with any other criteria. Specificities for the ATA, Korean TI-RADS, and French TI-RADS guidelines were 18% (95% CI: 11, 28), 16% (95% CI: 10, 26), and 33% (95% CI: 23, 44), respectively ($P < .03$). The accuracy rate for ACR TI-RADS (56%; 95% CI: 46, 66) was significantly higher than that for the ATA, Korean TI-RADS, or French TI-RADS guidelines, which had accuracies of 28% (95% CI: 20, 38), 27% (95% CI: 19, 37), and 41% (95% CI: 32, 51), respectively ($P < .04$).

Discussion

We retrospectively applied ACR TI-RADS guidelines to radiologists' interpretations of thyroid nodule US features. The interpretations were performed at an opportune time when the ACR TI-RADS white paper had not yet been published and was not known to the radiologists. When compared with radiologists' existing practice patterns, applying ACR TI-RADS guidelines reduces the number of thyroid nodule biopsies performed and improves specificity and accuracy in the detection of malignancy.

Reduction in the number of biopsy recommendations is highly desirable, especially if the majority of avoided biopsies are for benign nodules. Our study found that when compared with radiologists' own practice patterns, the use of ACR TI-RADS guidelines led to a reduction in the number of nodules recommended for biopsy of up to 41% in individuals, with an overall mean reduction of 29%. When compared with other biopsy guidelines, ACR TI-RADS guidelines also led to a reduction in the number of nodules referred for biopsy by test readers and expert consensus. Most of the nodules spared from biopsy with ACR TI-RADS guidelines were benign nodules, resulting in a marked improvement in test readers' specificity (from 20% based on individual practice patterns to 44% with ACR TI-RADS guidelines). One radiologist (reader 8) was an outlier, in that ACR TI-RADS guidelines did not improve his performance. This radiologist practiced in a health care system that established its own criteria 4 years ago. Those criteria were similar to ACR TI-RADS criteria.

There are two primary reasons why the other seven radiologists recommended so many more benign biopsies without ACR TI-RADS guidelines. First, they were following other guidelines, which used lower size thresholds for biopsy than did ACR TI-RADS guidelines. However, this is not the sole explanation, because five of the eight radiologists recommended more biopsies than they would have recommended with the ATA guidelines. The other reason for excess benign biopsy recommendations is that the test readers' decision to perform biopsy was based on their personal opinions, which may vary widely and may err on the side of overmanagement when there is less confidence or experience. Radiologist uncertainty

Table 2

Recommendation for Biopsy by Test Readers without and with ACR TI-RADS Guidelines

Reader	No. of Biopsy Recommendations without ACR TI-RADS	Sensitivity without ACR TI-RADS	Specificity without ACR TI-RADS	Accuracy without ACR TI-RADS	No. of Biopsy Recommendations with ACR TI-RADS	Sensitivity with ACR TI-RADS	Specificity with ACR TI-RADS	Accuracy with ACR TI-RADS
Reader 1	86	0.87 (13/15) [0.54, 0.97]	0.14 (12/85) [0.08, 0.23]	0.25 (25/100) [0.17, 0.35]	56	0.80 (12/15) [0.50, 0.94]	0.47 (40/85) [0.36, 0.58]	0.52 (52/100) [0.42, 0.61]
Reader 2	81	1.0 (15/15) [0.78, 1.00]	0.22 (19/85) [0.14, 0.33]	0.34 (34/100) [0.25, 0.44]	48	0.73 (11/15) [0.41, 0.91]	0.56 (48/85) [0.45, 0.67]	0.59 (59/100) [0.49, 0.69]
Reader 3	95	0.93 (14/15) [0.58, 0.99]	0.05 (4/85) [0.02, 0.12]	0.17 (18/100) [0.11, 0.27]	56	0.80 (12/15) [0.50, 0.94]	0.47 (40/85) [0.37, 0.58]	0.52 (52/100) [0.42, 0.62]
Reader 4	81	0.93 (14/15) [0.58, 0.99]	0.21 (18/85) [0.13, 0.31]	0.32 (32/100) [0.23, 0.42]	62	0.93 (14/15) [0.58, 0.99]	0.43 (37/85) [0.33, 0.55]	0.51 (51/100) [0.40, 0.61]
Reader 5	80	0.8 (12/15) [0.48, 0.95]	0.20 (17/85) [0.13, 0.30]	0.29 (29/100) [0.21, 0.39]	73	0.87 (13/15) [0.56, 0.97]	0.29 (25/85) [0.20, 0.40]	0.38 (38/100) [0.29, 0.48]
Reader 6	88	0.93 (14/15) [0.58, 0.99]	0.13 (11/85) [0.07, 0.22]	0.25 (25/100) [0.17, 0.35]	68	0.87 (13/15) [0.56, 0.97]	0.38 (32/85) [0.28, 0.49]	0.45 (45/100) [0.35, 0.55]
Reader 7 (academic)	89	0.93 (14/15) [0.58, 0.99]	0.12 (10/85) [0.06, 0.21]	0.24 (24/100) [0.16, 0.34]	56	0.80 (12/15) [0.50, 0.94]	0.49 (42/85) [0.39, 0.60]	0.54 (54/100) [0.44, 0.64]
Reader 8 (academic)	38	0.93 (14/15) [0.58, 0.99]	0.72 (61/85) [0.61, 0.80]	0.75 (75/100) [0.65, 0.83]	37	0.73 (11/15) [0.41, 0.91]	0.70 (60/85) [0.60, 0.79]	0.71 (71/100) [0.61, 0.79]
Overall test readers	80 ± 16	0.95 (110/120) [0.83, 0.99]	0.20 (152/680) [0.16, 0.25]	0.28 (262/800) [0.21, 0.37]	57 ± 11	0.92 (98/120) [0.68, 0.98]	0.44 (324/680) [0.33, 0.56]	0.52 (422/800) [0.40, 0.63]
Expert consensus	…	…	…	…	55	0.87 (13/15) [0.48, 0.98]	0.51 (43/85) [0.40, 0.62]	0.56 (56/100) [0.46, 0.66]

Note.—Data in parentheses are raw data and do not always match the estimated performance statistics because analysis takes into account the clustering effect of nodules within patients. Data in brackets are 95% confidence intervals.
* Data are mean ± standard deviation.

is reflected in the distribution of risk stratification levels for the nodules. The test readers classified the majority of nodules as mildly suspicious, with very few nodules being considered highly suspicious. After applying ACR TI-RADS guidelines, the majority of nodules were categorized as either highly suspicious or moderately suspicious, even though fewer nodules were recommended for biopsy.

Radiologists are concerned with the prospect of missing a malignancy when fewer nodules are recommended for biopsy. With expert consensus interpretation, the number of malignancies that did not meet the criteria for biopsy was the same for ACR TI-RADS guidelines, ATA guidelines, and Korean and French TI-RADS guidelines. Although the sensitivity of test readers was slightly lower with ACR TI-RADS guidelines (92%) than with readers' own practice patterns (95%), an important consideration is that many higher-suspicion nodules that did not meet the criteria for biopsy with ACR TI-RADS guidelines were recommended for follow-up US. ACR TI-RADS provides specific follow-up guidelines for TR3, TR4, and TR5 nodules based on a lower size threshold than that used for biopsy. Given the indolent behavior of small thyroid cancers, observation of small suspicious nodules is a safe strategy (8,9). Active surveillance of papillary microcarcinomas without treatment is performed in Japan (10) and has been acknowledged in the 2015 ATA guidelines as a management option in the United States (2). Active surveillance programs exist in the United States for papillary carcinomas up to 1.5 cm in diameter, and early evidence suggests it can reduce treatment-related morbidity and costs in patients with low-risk malignancies (11,12).

In our study, all malignancies were recommended for biopsy or follow-up, with only two exceptions. Two malignancies were incorrectly interpreted as spongiform and mixed cystic and solid nodules by two of the eight test readers but were regarded to have solid composition by all other readers and by the experts. This suggests that further

Table 3

Malignancies for which Not All Readers Recommended Biopsy or Follow-up with ACR TI-RADS and ATA Guidelines*

Characteristic	Nodule 1	Nodule 5	Nodule 13	Nodule 23	Nodule 58	Not Recommended for Biopsy with ACR-RADS Guidelines	Not Recommended for Biopsy or Follow-up with ACR-RADS Guidelines	Not Recommended for Biopsy with ATA Guidelines
Size (cm)	1.2	1.1	2.3	1.9	0.7			
Expert consensus	TR5, biopsy; ATA high, biopsy	TR4, follow-up; ATA low, no biopsy	TR5, biopsy; ATA high, biopsy	TR5, biopsy; ATA high, biopsy	TR5, follow-up; ATA high, no biopsy	2	0	2
Reader 1	TR4, follow-up; ATA intermediate, biopsy	TR4, follow up; ATA low, no biopsy	…	…	TR5, follow-up; ATA high, no biopsy	3	0	2
Reader 2	TR4, follow up; ATA intermediate, biopsy	…	TR3, follow-up; ATA low, biopsy	TR1, no follow-up; ATA very low, no biopsy	TR5, follow-up; ATA high, no biopsy	4	1	2
Reader 3	TR4, follow up; ATA high, biopsy	TR4, follow up; ATA low, no biopsy	…	…	TR5, follow up; ATA high, no biopsy	3	0	2
Reader 4	…	…	…	…	TR5, follow-up; ATA high, no biopsy	1	0	1
Reader 5	…	TR4, follow-up; ATA low, no biopsy	…	…	TR5, follow-up; ATA high, no biopsy	2	0	2
Reader 6	…	TR4, follow-up; ATA low, no biopsy	…	…	TR5, follow up; ATA high, no biopsy	2	0	2
Reader 7 (academic)	TR4, follow-up; ATA intermediate, biopsy	TR4, follow-up; ATA low, no biopsy	…	…	TR5, follow-up; ATA high, no biopsy	3	0	2
Reader 8 (academic)	TR1, no follow-up; ATA very low, no biopsy	TR4, follow-up; ATA low, no biopsy	…	TR2, no follow-up; ATA low, biopsy	TR5, follow up; ATA high, no biopsy	4	2	3

Note.—Risk categories are based on the readers' interpretation of the imaging features.
* ATA guidelines do not address solid or mixed solid and cystic nodules that are iso- to hyperechoic and have additional malignant features. For these nodules, we assumed fine-needle aspiration biopsy if a solid nodule was 1 cm or larger and if a mixed cystic and solid nodule was 1.5 cm or larger.

Figure 3

Figure 3: Transverse US image of a solid hypoechoic malignant thyroid nodule with a maximum diameter of 1.2 cm. Surgical pathology revealed a follicular variant of papillary thyroid carcinoma. Without ACR TI-RADS criteria, four of eight test readers recommended biopsy. After application of ACR TI-RADS criteria, three readers recommended biopsy, four recommended follow-up, and one reader recommended no biopsy or follow-up. Without ACR TI-RADS criteria, the risk stratification level was deemed not suspicious by three readers, mildly suspicious by two, moderately suspicious by two, and highly suspicious by one reader. With ACR TI-RADS criteria, the risk stratification level was judged as benign by one reader, moderately suspicious by four readers, and highly suspicious by three readers. The reason this malignancy was not recommended for biopsy or follow-up by one reader was because of misinterpretation of composition as spongiform, which placed the nodule in the benign category (TR1).

education should be directed toward the distinction between solid nodules (including those with a few cysts), spongiform nodules, and mixed cystic and solid nodules. Radiologists need to be cognizant that under the ACR TI-RADS recommendations, mixed solid and cystic nodules without suspicious features and all spongiform nodules are not recommended for biopsy or follow-up, regardless of size. Thus, careful consideration should be made before categorizing a nodule as spongiform if it has any suspicious features.

Our study had several limitations. The readers evaluated selected cases from clinical practice in which biopsy was performed because of the suspicious or indeterminate nature

Figure 4

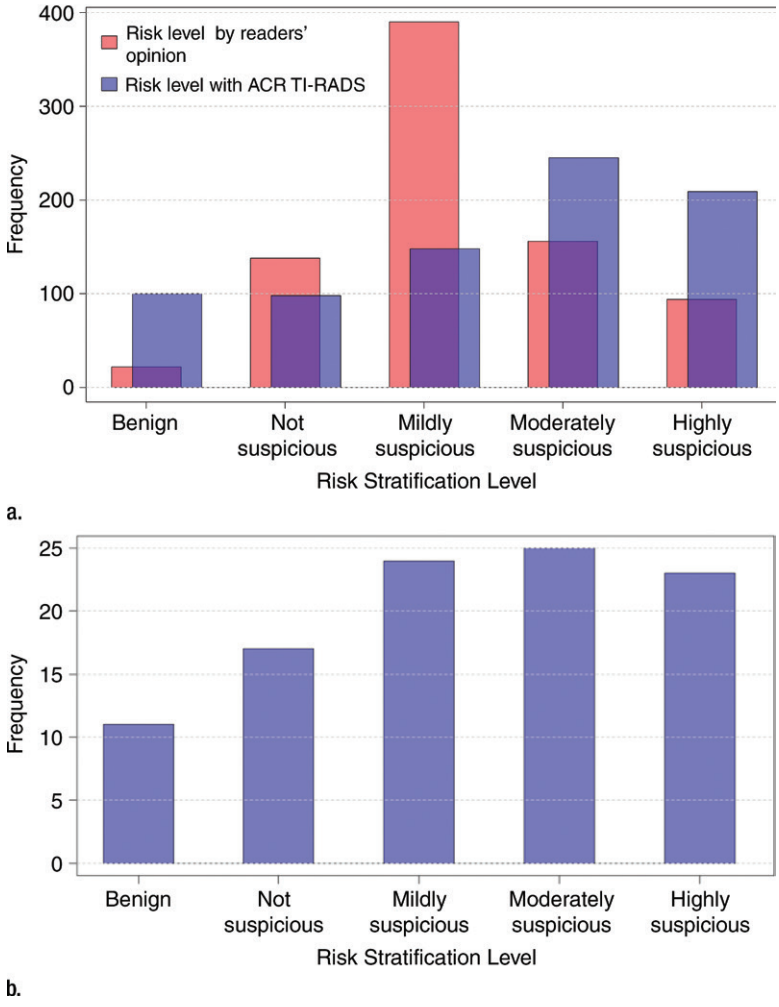

Figure 4: **(a)** Graph shows nodule risk stratification levels graded by all test readers based on their opinion without ACR TI-RADS criteria and after application of ACR TI-RADS criteria. **(b)** Graph shows nodule risk stratification levels graded by expert consensus.

of findings; therefore, our study may not be reflective of clinical practice. This sample included a higher number of malignancies and nodules that likely had a disproportionately higher frequency of suspicious US findings. Second, there was an unequal distribution of academic and private practice radiologists among the test readers. However, we believe that this reflects expertise in thyroid US among the community. Third, selected US images and video clips were evaluated with a Web-based portal by using standard computer monitors and not a picture archiving and communication system workstation. Although the quality of the images presented was thought to be adequate to evaluate US features, differences in monitor adjustments and other technical factors could have affected perception and led to variability in interpretation. Finally, comparison of the ACR TI-RADS guidelines with other guidelines was limited by the small number of nodules. Our results will need to be confirmed in larger studies.

In conclusion, use of ACR TI-RADS guidelines resulted in a meaningful reduction in the number of thyroid nodules recommended for biopsy and significantly improved the accuracy of recommendations for nodule management. The majority of malignancies not recommended for biopsy will be recommended for follow-up US.

Acknowledgments: We acknowledge the support of ACR staff members Lauren Hicks and Mythreyi Chatfield.

Disclosures of Conflicts of Interest: **J.K.H.** disclosed no relevant relationships. **W.D.M.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: receives royalties from Elsevier. Other relationships: disclosed no relevant relationships. **A.E.F.** disclosed no relevant relationships. **J.E.L.** disclosed no relevant relationships. **C.C.R.** disclosed no relevant relationships. **S.A.T.** disclosed no relevant relationships. **N.A.** disclosed no relevant relationships. **F.J.B.** disclosed no relevant relationships. **A.J.B.** disclosed no relevant relationships. **N.D.** disclosed no relevant relationships. **B.S.H.** disclosed no relevant relationships. **J.R.N.** disclosed no relevant relationships. **D.S.** disclosed no relevant relationships. **R.C.V.** disclosed no relevant relationships. **F.N.T.** disclosed no relevant relationships.

References

1. Hoang JK, Lee WK, Lee M, Johnson D, Farrell S. US Features of thyroid malignancy: pearls and pitfalls. RadioGraphics 2007;27(3):847–860; discussion 861–865.
2. 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.
3. Frates MC, Benson CB, Charboneau JW, et al. Management of thyroid nodules detected at US: Society of Radiologists in Ultrasound consensus conference statement. Radiology 2005;237(3):794–800.
4. 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(3):370–395.
5. Russ G, Bigorgne C, Royer B, Rouxel A, Bienvenu-Perrard M. The Thyroid Imaging Reporting and Data System (TIRADS) for ultrasound of the thyroid [in French]. J Radiol 2011;92(7-8):701–713.
6. 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.

7. 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 Am J Roentgenol 2017;208(6):1331–1341.
8. 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(1):59–64.
9. Hoang JK, Langer JE, Middleton WD, et al. Managing incidental thyroid nodules detected on imaging: white paper of the ACR Incidental Thyroid Findings Committee. J Am Coll Radiol 2015;12(2):143–150.
10. 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(1):150–155.
11. Pace MD, Fagin JA, Bach A, et al. Properly selected patients with papillary thyroid cancer readily accept active surveillance when offered as a standard of care alternative to immediate surgery. Thyroid 2013;23(A-116):8.
12. Griffin A, Brito JP, Bahl M, Hoang JK. Applying Criteria of Active Surveillance to Low-Risk Papillary Thyroid Cancer Over a Decade: How Many Surgeries and Complications Can Be Avoided? Thyroid 2017;27(4):518–523.