Comparison of Performance Characteristics of American College of Radiology TI-RADS, Korean Society of Thyroid Radiology TIRADS, and American Thyroid Association Guidelines

William D. Middleton[1]
Sharlene A. Teefey[1]
Carl C. Reading[2]
Jill E. Langer[3]
Michael D. Beland[4]
Margaret M. Szabunio[5]
Terry S. Desser[6]

OBJECTIVE. The American College of Radiology (ACR) Thyroid Imaging Reporting and Data System (TI-RADS) provides guidelines to practitioners who interpret sonographic examinations of thyroid nodules. The purpose of this study is to compare the ACR TI-RADS system with two other well-established guidelines.

MATERIALS AND METHODS. The ACR TI-RADS, the Korean Society of Thyroid Radiology (KSThR) Thyroid Imaging Reporting and Data System (TIRADS), and the American Thyroid Association guidelines were compared using 3422 thyroid nodules for which pathologic findings were available. The composition, echogenicity, margins, echogenic foci, and size of the nodules were assessed to determine whether a recommendation would be made for fine-needle aspiration or follow-up sonography when each system was used. The biopsy yield of malignant findings, the yield of follow-up, and the percentage of malignant and benign nodules that would be biopsied were determined for all nodules and for nodules 1 cm or larger.

RESULTS. The percentage of nodules that could not be classified was 0%, 3.9%, and 13.9% for the ACR TI-RADS, KSThR TIRADS, and ATA guidelines, respectively. The biopsy yield of malignancy was 14.2%, 10.2%, and 10.0% for nodules assessed by the ACR TI-RADS, KSThR TIRADS, and ATA guidelines, respectively. The percentage of malignant nodules that were biopsied was 68.2%, 78.7%, and 75.9% for the ACR TI-RADS, the KSThR TIRADS, and the ATA guidelines, respectively, whereas the percentage of malignant nodules that would be either biopsied or followed was 89.2% for the ACR TI-RADS. The percentage of benign nodules that would be biopsied was 47.1%, 79.7%, and 78.1% for the ACR TI-RADS, the KSThR TIRADS, and the ATA guidelines, respectively. The percentage of benign nodules that would be either biopsied or followed was 65.2% for the ACR TI-RADS.

CONCLUSION. The ACR TI-RADS performs well when compared with other well-established guidelines.

Keywords: thyroid cancer, thyroid guidelines, thyroid nodules, thyroid ultrasound, TI-RADS

doi.org/10.2214/AJR.17.18822

Received July 28, 2017; accepted after revision October 1, 2017.

M. D. Beland is a consultant to Hitachi Aloka America, Inc.

Supported by a grant from the Society of Radiologists in Ultrasound.

[1]Mallinckrodt Institute of Radiology, Washington University in St. Louis, 510 S Kingshighway Blvd, St. Louis, MO 63110. Address correspondence to W. D. Middleton (williammiddleton@wustl.edu).

[2]Mayo Clinic, Rochester, MN.

[3]Department of Radiology, The Perelman School of Medicine, Hospital of the University of Pennsylvania, Philadelphia, PA.

[4]Department of Diagnostic Imaging, Rhode Island Hospital, Brown University, Providence, RI.

[5]University of Kentucky, Lexington, KY.

[6]Department of Radiology, Stanford University, Stanford, CA.

AJR 2018; 210:1148–1154

0361–803X/18/2105–1148

In the past decade, numerous guidelines have been developed to assist physicians in deciding when it is appropriate to perform fine-needle aspiration (FNA) or follow-up of thyroid nodules in low-risk adult patients, many of whom have incidentally discovered thyroid nodules. The American College of Radiology (ACR) published the Thyroid Imaging Reporting and Data System (TI-RADS), which uses a point-based classification system [1]. The system was based on analysis of the literature, the Surveillance Epidemiology and End Results program of the National Cancer Institute, existing risk stratification systems, existing guidelines, and expert opinion. We applied the system to a multiinstitutional database of nodules with pathologically proven findings acquired from 2006 to 2010. The purpose of the present study was to gauge the performance of the ACR TI-RADS and compare it with well-established systems devised by the American Thyroid Association (ATA) and the Korean Society of Thyroid Radiology (KSThR) [2, 3]. Our hypothesis was that use of the ACR system would result in a decrease in the number of nodules that required biopsy.

Materials and Methods

Between 2006 and 2010, a total of 3315 patients from six tertiary academic institutions who were undergoing thyroid nodule FNA, were 18 years or older, and consented to participate in the study were enrolled in our HIPAA-compliant study, which was approved by the institutional review boards of each participating institution. During the time frame specified, all patients who were

undergoing ultrasound-guided FNA were asked to participate at each study site. Patients with thyroid nodules identified on sonography who did not undergo thyroid FNA were excluded.

The tissue sampling protocol, including the type of needle used, the number of passes, and the presence of on-site microscopic sample analysis, was left to the discretion of the physician performing the aspiration. The cytology reports from each institution were used to classify nodules into five categories: malignant, suspicious for malignancy, indeterminate (follicular or atypical), benign, and nondiagnostic. Nodules with FNA results that were interpreted as malignant or benign were considered diagnostic and were included in the final analysis. All other nodules were excluded unless either subsequent FNA was performed and had results that were interpreted as malignant or benign or the nodule was resected and histologic findings were available.

The study included 3315 patients. Three nodules were biopsied in 24 patients, two were biopsied in 459 patients, and one was biopsied in 2832 patients, resulting in a total of 3822 nodules (3056 from women and 766 from men). Patient age was 18–97 years, with a mean age of 54.4 years. A total of 173 nodules were excluded because of nondiagnostic cytologic analysis results with no further evaluation. In addition, 227 nodules were excluded because results of cytologic analysis were suspicious for (but not diagnostic of) malignancy or indicated indeterminate status with no further evaluation. Therefore, a total of 3422 nodules were included in the study.

Sonographic images and, in some cases, video clips of all biopsied nodules were obtained using a variety of commercially available ultrasound units. At the time of image acquisition, attention was focused on providing optimal imaging to allow assessment of nodule composition, echogenicity, margins, and echogenic foci. Static images and video clips of all sampled nodules were sent to a central reading site where they were retrospectively reviewed from 2006 to 2011 by two radiologists who were experienced in thyroid sonography, and a database was created. Both radiologists reviewed 50 cases jointly in conference at the beginning and in the middle of the study, to standardize the interpretations. The remaining cases were reviewed independently. At the time of their reviews, the radiologists had access to the original ultrasound reports but were blinded to the pathologic results.

The central reviewers categorized and recorded the imaging features of each nodule. Composition was categorized as follows: solid or almost entirely solid, mixed solid and cystic, spongiform, and cystic or almost entirely cystic. Margins were categorized as entirely smooth, lobulated or irregular, or ill-defined. Echogenic foci were categorized as macrocalcifications (calcifications with shadowing), peripheral calcifications (complete and partial), punctate echogenic foci (tiny bright reflectors without shadowing), or small nonshadowing echogenic foci with comet-tail artifacts. Echogenicity was assessed in the solid noncalcified portions of the nodule and was categorized as hyperechoic, isoechoic, mildly hypoechoic, or moderately to markedly hypoechoic. Moderately to markedly hypoechoic nodules included nodules that were less echogenic than strap muscles as well as nodules that were more echogenic than strap muscles but were substantially less echogenic than the thyroid parenchyma. Nodule features that could not be assessed were also recorded and were associated primarily with heavily calcified nodules.

TABLE 1: American College of Radiology Thyroid Imaging Reporting and Data System (TI-RADS) Classification and Recommendations

Feature (Choice)	Description and Points	TI-RADS Category (Sum of Points) Recommendation
Composition (choose 1)	Cystic or almost completely cystic	TR1 (0 points)
	0 points	No FNA
	Spongiform	
	0 points	
	Mixed cystic and solid	No follow-up
	1 point	
	Solid or almost completely solid	
	2 points	
Echogenicity (choose 1)	Anechoic	TR2 (2 points)
	0 points	No FNA
	Hyperechoic or isoechoic	
	1 point	
	Hypoechoic	No follow-up
	2 points	
	Very hypoechoic	
	3 points	
Margin (choose 1)	Smooth or ill-defined	TR3 (3 points)
	0 points	FNA for nodule ≥ 2.5 cm
	Lobulated or irregular	
	2 points	Follow-up for nodule ≥ 1.5 cm
	Extrathyroidal extension	
	3 points	
Echogenic foci (choose all that apply)	None or large comet-tail artifacts	TR4 (4–6 points)
	0 points	
	Macrocalcifications	FNA for nodule ≥ 1.5 cm
	1 point	
	Peripheral (rim) calcifications	
	2 points	Follow-up for nodule ≥ 1.0 cm
	Punctate echogenic foci	
	3 points	
Shape (choose 1)	Wider than tall	TR5 (≥ 7 points)
	0 points	FNA for nodule ≥ 1.0 cm
	Taller than wide	Follow-up for nodule ≥ 0.5 cm
	3 points	

Note—Data from [1]. FNA = fine-needle aspiration.

TABLE 2: Size Distribution of Benign and Malignant Nodules

Size (cm)	Malignant	Benign
0.0–0.4	2 (0.6)	5 (0.2)
0.5–0.9	62 (17.6)	174 (5.7)
1.0–1.4	81 (23.0)	548 (17.9)
1.5–1.9	65 (18.5)	601 (19.6)
2.0–2.4	43 (12.2)	475 (15.5)
2.5–2.9	31 (8.8)	382 (12.4)
≥ 3.0	68 (19.3)	885 (28.8)
Total	352	3070

Note—Data are number (%) of nodules.

Each nodule was then assigned an ACR TI-RADS category (TR1–TR5) on the basis of the point total obtained from each nodule feature (Table 1). Using the KSThR TIRADS, each nodule was categorized as high, intermediate, or low suspicion category, or as benign. For categorization of nodules according to the KSThR TIRADS, please see Table 2 in Shin et al. [3]. Using the ATA guidelines, each nodule was categorized as high, intermediate, low, or very low suspicion category or as benign. For categorization of nodules according to ATA guidelines, please see Table 6 in Haugen et al. [2]. Categorization was based on the sonographic features described in the relevant charts, tables, and accompanying figures for each of the systems. An additional category was created for nodules that could not be categorized. Once nodules were categorized, an assessment was made as to whether that system recommends FNA or follow-up on the basis of size thresholds. In cases for which recommendations listed more than one option for management, the option that appeared first in the charts was used for our analysis.

The ACR TI-RADS includes specific criteria for the follow-up of nodules that do not meet criteria for FNA, and the numbers of nodules that would and would not meet these criteria were determined. When possible, similar determinations were performed for the ATA guidelines. The KSThR TIRADS does not include specific recommendations for the follow-up of nodules that do not meet the criteria for biopsy.

To determine different performance characteristics for each system, calculations were made on the basis of the numbers of nodules in the database that would meet criteria for biopsy and would not meet criteria for biopsy. The equation used to determine the biopsy yield of malignancy was (number of malignant nodules biopsied / total number of nodules biopsied) × 100%, whereas that used to determine the percentage of malignant nodules biopsied was (number of malignant nodules biopsied / total number of malignant nodules) × 100% and that used to determine the percentage of benign nodules biopsied was (number of benign nodules biopsied / total number of benign nodules) × 100%.

The proportion and 95% CI of each characteristic were compared between the ACR TI-RADS and the KSThR TIRADS and ATA guidelines. The binomial test of proportions was conducted to test for differences in the proportions. All tests were two-sided, and a significance level of $p = 0.05$ was used. Statistical software (SAS, version 9.4, SAS Institute) was used to conduct data analyses.

Results

A total of 352 malignant nodules and 3070 benign nodules were included in the study. Of the 352 malignant nodules, 303 were resected and had histologic proof of malignancy. Of the resected malignant nodules, 179 were papillary cancer, 88 were follicular variant of papillary cancer, 16 were follicular cancer, six were medullary cancer, one was anaplastic cancer, and 13 were other malignancies. The sizes of the benign and malignant nodules are shown in Table 2.

When the ACR TI-RADS was applied to our database, nodule categorization was as follows: 299 nodules were TI-RADS category TR1; 548, category TR2; 775, category TR3; 1251, category TR4; and 549, category TR5. All nodules could be categorized using the ACR TI-RADS.

When the KSThR TIRADS was applied to our database, categorization was as follows: 420 nodules were benign, 1343 had low suspicion for malignancy, 1127 had intermediate suspicion for malignancy, and 398 had high suspicion for malignancy. A total of 3.9% of all nodules (134/3422) could not be distinctly categorized into a single category. Of the nodules that could not be categorized, 9.7% (13/134) were malignant. A total of 3.7% of all malignant nodules (13/352) could not be categorized. Most nodules that could not be categorized (115/134) were nodules for which composition, echogenicity, margins, or a combination of these features could not be determined, and the inability to make that determination precluded definitive classification into a single category (103 benign nodules and 12 malignant nodules). For instance, 25 solid hyperechoic or isoechoic nodules had no suspicious features (defined in this guideline as microcalcifications, a taller-than-wide shape, or a spiculated or microlobulated margin) that could not be categorized as intermediate or low suspicion because the margins could not be determined (24 benign nodules and one malignant nodule). On the other hand, 12 solid hypoechoic nodules with microcalcifications could be definitively categorized as having a high suspicion for malignancy despite the fact that the margins could not be determined to be smooth or either spiculated or microlobulated (seven benign nodules and five malignant nodules).

When the ATA guidelines were applied to the nodules in our database, the nodules were categorized as follows: 1045 had very low suspicion for malignancy, 648 had low suspicion, 640 had intermediate suspicion, and 612 had high suspicion. A total of 13.9% of nodules (477/3422) could not be categorized, and 9.4% of the nodules that could not be categorized (45/477) were malignant. A total of 12.8% of malignant nodules (45/352) could not be categorized. Most nodules that could not be categorized ($n = 210$) were mixed cystic and solid nodules that were hyperechoic or isoechoic and had malignant features (Fig. 1), which were defined in this set of guidelines as microcalcifications, irregular margins, taller-than-wide shape, and interrupted rim calcification with soft-tissue protrusion (201 benign nodules and nine malignant nodules). Solid hyperechoic or isoechoic nodules that had malignant features ($n = 127$), and nodules for which the composition, echogenicity, margins, or a combination of these features could not be determined ($n = 137$) (Fig. 2) were two other common groups that could not be categorized.

Table 3 provides a comparison of the results of the three systems. To account for the nodules that could not be classified, the results for the KSThR TIRADS and the ATA guidelines were calculated under the assumption that these nodules would be biopsied at a size threshold of 1.5 cm or greater. A sensitivity analysis of the ATA and KSThR

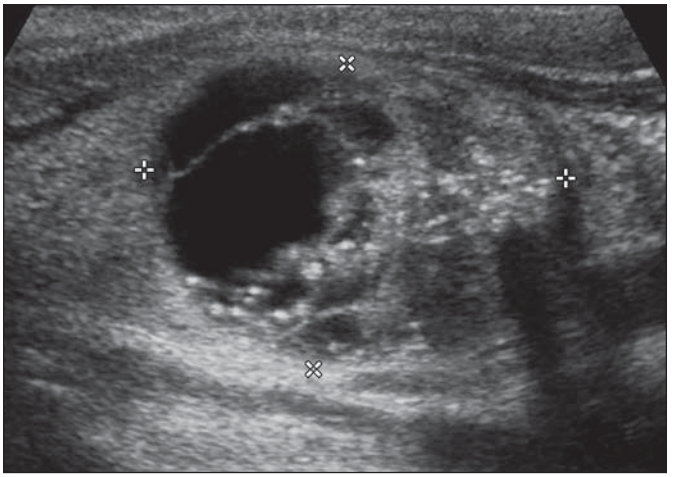

Fig. 1—46-year-old woman with papillary thyroid carcinoma. Longitudinal ultrasound view of thyroid shows mixed cystic and solid isoechoic nodule (within *calipers*) with smooth margins, microcalcifications (punctate echogenic foci), and wider-than-tall shape. This nodule is classified as category TR4 nodule according to American College of Radiology Thyroid Imaging Reporting and Data System, is of intermediate suspicion according to Korean Society of Thyroid Radiology Thyroid Imaging Reporting and Data System, and is not classifiable using American Thyroid Association guidelines. Distance between left and right calipers is 30 mm. Distance between top and bottom calipers is 21 mm.

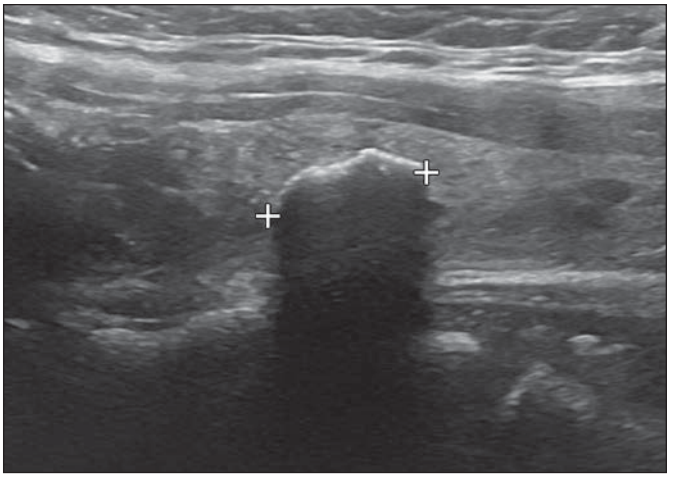

Fig. 2—46-year-old woman with papillary thyroid carcinoma. Longitudinal ultrasound view of thyroid shows nodule (within *calipers*) with dense peripheral calcification. Shadow from calcification precludes evaluation of composition and echogenicity. This nodule is classified as category TR4 in American College of Radiology Thyroid Imaging Reporting and Data System and is not classifiable in Korean Society of Thyroid Radiology Thyroid Imaging Reporting and Data System and American Thyroid Association guidelines. Distance between calipers is 1.21 cm.

systems was performed that calculated results if the unclassifiable nodules were excluded and, also, if the nodules were included under the assumption that they either would not be biopsied or would be biopsied at a size threshold of 1.0 cm or larger (Table 4).

The biopsy yield of malignancy was 14.2% (95% CI, 12.6–15.9%) for the ACR TI-RADS, 10.0% (95% CI, 8.9–11.2%) for the ATA guidelines, and 10.2% (95% CI, 9.1–11.4%) for the KSThR TIRADS. The ACR TI-RADS biopsy yield was significantly higher than that of the ATA guidelines ($p < 0.0001$) or the KSThR TIRADS ($p < 0.0001$).

The percentage of benign nodules that were biopsied was 47.1% (95% CI, 45.4–48.9%) for the ACR TI-RADS, 78.1% (95% CI, 76.7–79.6%) for the ATA guidelines, and 79.7% (95% CI, 78.3–81.1) for the KSThR TIRADS. The percentage of benign nodules biopsied when the ACR TI-RADS was used was significantly lower than the percentage biopsied when the ATA guidelines ($p < 0.001$) or the KSThR TIRADS ($p < 0.001$) was used. A total of 2891 benign nodules were 1 cm or larger. The percentage that would be biopsied using the ACR TI-RADS was 50.1% (95% CI, 48.1–51.9%). The percentage that would be biopsied using the ATA and KSThR systems was 83.0% (95% CI, 81.6–84.3%) and 84.6% (95% CI, 83.3–85.9%) respectively. The percentage of benign nodules 1 cm or larger biopsied when the ACR TI-RADS was used was significantly lower than that noted when the ATA guidelines ($p < 0.001$) or the KSThR guidelines ($p < 0.0001$) were used.

The percentage of malignant nodules that would be biopsied was 68.2% (95% CI, 63.1–72.8%) for the ACR TI-RADS, 75.9% (95% CI, 71.1–80.0%) for the ATA guidelines, and 78.7% (95% CI, 74.1–82.7%) for the KSThR TIRADS. For the ACR TI-RADS, the percentage of malignant nodules that would be biopsied was significantly lower than that noted for the ATA guidelines ($p = 0.0021$) or the KSThR TIRADS ($p < 0.0001$). A total of 288 malignant nodules were 1 cm or larger. The percentage of malignant nodules 1 cm or larger that would be biopsied when the ACR TI-RADS was used was 83.3% (95% CI, 78.6–87.2%). The percentage of malignant nodules 1 cm or larger that would be biopsied using the ATA and KSThR systems was 92.7% (95% CI, 89.1–95.2%) and 96.2%

TABLE 3: Results for Nodules Classified Using the Three Thyroid Classification Systems

Result	ACR TI-RADS		Nodules Biopsied Per ATA Guidelines	Nodules Biopsied Per KSThR TIRADS
	Biopsied Nodules	Monitored Nodules[a]		
Yield of malignancy	14.2 (240/1687)	11.8 (74/629)	10.0 (267/2666)	10.2 (277/2724)
Malignant nodules	68.2 (240/352)	21.0 (74/352)	75.9 (267/352)	78.7 (277/352)
Malignant nodules ≥ 1 cm	83.3 (240/288)	11.1 (32/288)	92.7 (267/288)	96.2 (277/288)
Benign nodules	47.1 (1447/3070)	18.1 (555/3070)	78.1 (2399/3070)	79.7 (2447/3070)
Benign nodules ≥ 1 cm	50.1 (1447/2891)	17.8 (515/2891)	83.0 (2399/2891)	84.6 (2447/2891)

Note—Data are percentage of nodules (number of nodules with feature/total number of nodules). Results were determined under the assumption that nonclassifiable nodules identified by the Korean Society of Thyroid Radiology (KSThR) Thyroid Imaging Reporting and Data System (TIRADS) and the American Thyroid Association (ATA) guidelines would be biopsied at a size threshold of 1.5 cm or greater.

[a]Nodules that do not meet the criteria for biopsy but do meet the criteria for sonographic follow-up.

TABLE 4: Results for Nodules Classified Using the Korean Society of Thyroid Radiology Thyroid Imaging Radiology and Data System (KSThR TIRADS) and American Thyroid Association (ATA) Guidelines

Result	ATA Guidelines			KSThR TIRADS		
	With Nonclassifiable Nodules Excluded	With No Biopsy of Nonclassifiable Nodules	With Biopsy of Nonclassifiable Nodules ≥ 1 cm	With Nonclassifiable Nodules Excluded	With No Biopsy of Nonclassifiable Nodules	With Biopsy of Nonclassifiable Nodules ≥ 1 cm
Yield of malignancy	10.3 (235/2289)	10.3 (235/2289)	10.1 (276/2733)	10.2 (267/2630)	10.2 (267/2630)	10.1 (278/2746)
Malignant nodules biopsied	76.5 (235/307)	66.8 (235/352)	78.4 (276/352)	78.8 (267/339)	75.9 (267/352)	79.0 (278/352)
Benign nodules biopsied	77.9 (2054/2638)	66.9 (2054/3070)	80.0 (2457/3070)	80.1 (2362/2949)	77.0 (2363/3070)	80.4 (2468/3070)

Note—Data are percentage of nodules (number of nodules with result/total number of nodules). Results were determined when nonclassifiable nodules were excluded altogether and, also, when they were included, under the assumption they would not be biopsied or would be biopsied at a size threshold of 1.0 cm or larger.

(95% CI, 93.2–97.9%), respectively. The percentage of malignant nodules 1 cm or greater biopsied when using the ACR TI-RADS was significantly lower than that noted when using the ATA guidelines ($p < 0.0001$) or the KSThR guidelines ($p < 0.0001$).

The ACR TI-RADS provides comprehensive guidelines for determining when sonographic follow-up is necessary for nodules that do not meet the criteria for biopsy (Table 1). The purpose of the follow-up examinations is to detect malignant nodules that do not initially meet criteria for biopsy. The KSThR TIRADS does not provide any guidelines for follow-up, and the ATA guidelines for nodules that do not meet the criteria for biopsy are not comprehensive. Because a direct comparison of follow-up recommendations of the three systems was not possible, we compared the nodules that would be either biopsied or monitored using the ACR TI-RADS with the nodules that would be biopsied using the KSThR TIRADS and the ATA guidelines.

The yield for nodules that would be either biopsied or followed using the ACR TI-RADS was 13.6% (95% CI, 12.2–15.0%), which was calculated as follows: (240 biopsied nodules + 74 monitored nodules) / (1686 biopsied nodules + 629 monitored nodules). This yield was significantly higher than the biopsy yield for the ATA guidelines (10.0%) ($p < 0.001$) or the KSThR TIRADS (10.2%) ($p < 0.0001$).

The percentage of benign nodules that would be either biopsied or followed with use of the ACR TI-RADS was 65.2% (2002/3070; 95% CI, 63.5–66.9%). This percentage was significantly lower than the percentage of benign nodules biopsied using the ATA guidelines (78.1%) ($p < 0.0001$) or the KSThR TIRADS (79.7%) ($p < 0.0001$). The percentage of benign nodules 1 cm or larger that would be either biopsied or followed when the ACR TI-RADS was used was 67.9% (1962/2891; 95% CI, 66.1–69.5%). This percentage was also significantly lower than the percentage of benign nodules 1 cm or larger biopsied using the ATA guidelines (83.0%) ($p < 0.0001$) or the KSThR TIRADS (84.6%) ($p < 0.001$).

The percentage of malignant nodules that would be either biopsied or monitored using the ACR TI-RADS was 89.2% (314/352; 95% CI, 85.5–92.1%). This percentage was significantly higher than the percentage of nodules biopsied using the ATA guidelines (75.9%) ($p < 0.0001$) and the KSThR TIRADS (78.7%) ($p = 0.0001$). The percentage of malignant nodules 1 cm or larger that would be either biopsied or followed using the ACR TI-RADS was 94.4% (272/288; 95% CI 91.9–97.2%). This was not significantly different from the percentage biopsied using the ATA guidelines (92.7%) ($p = 0.2118$) or the KSThR TIRADS (96.2%) ($p = 0.1888$).

The ATA guidelines recommend that sonographic follow-up be performed, considered, or not be performed for some of the categories of nodules that do not meet criteria for biopsy. A total of 756 nodules (22.1%) did not meet ATA guidelines for biopsy. Of these nodules, 55.4% (419/756; 16 malignant nodules and 403 benign nodules) were classified into a category that did not have a recommendation for or against follow-up. A total of 10.8% of these nodules (82/756; 41 malignant nodules and 41 benign nodules) were classified into a category where follow-up was recommended in 6–12 months. Of these nodules, 27.1% (205/756; 26 malignant nodules and 179 benign nodules) were placed in a category where the recommendation was to consider follow-up in 12–24 months, and 6.6% (50/756; two malignant nodules and 48 benign nodules) were assigned to a category where the recommendation was to not perform follow-up examinations. If these guidelines were followed, 87.5% of malignant nodules (308/352) and 79.5% of benign nodules (2440/3070) would be either biopsied or monitored at 6–12 months, and 7.4% of malignant nodules (26/352) and 5.8% of benign nodules (177/3070) would be considered for follow-up at 12–24 months.

A definitive recommendation for no further evaluation (no FNA or follow-up) would be made for 1.5% of nodules (50/3422) with use of the ATA guidelines and for 32.3% of nodules (1106/3422) with use of the ACR TI-RADS. This difference was statistically significant ($p < 0.0001$).

Sensitivity analysis examined the robustness of the results for the ATA and KSThR systems when different assumptions on handling the noncategorized nodules were assumed (Table 4). The conclusions did not differ from the primary analysis.

Discussion

In 2006, the Society of Radiologists in Ultrasound convened a conference including radiologists, endocrinologists, thyroid surgeons, and pathologists to develop guidelines for determining which thyroid nodules should be biopsied and which should not be biopsied in low-risk adult patients [4]. Since then, several other guidelines have been proposed by individual institutions and professional societies. The ACR recently published guidelines for FNA and follow-up of thyroid nodules. This was based on existing risk stratification systems, data pertaining to prognosis of thyroid cancer, and the expert opinion of the committee members. It was also shaped by previous guidelines.

Although the risk stratification proposed by the ACR TI-RADS has been previously validated [5], the overall performance of the system, including the recommendations

for biopsy and follow-up, has not been tested. The present study was performed to determine the yield of the ACR TI-RADS when applied to a database of nodules collected before the development of the system. To place the results in perspective, we also performed a similar analysis based on recent, widely used recommendations from the KSThR and the ATA [2, 3].

All systems stratify nodules into risk categories on the basis of their sonographic features and then establish size thresholds for biopsy in each category. With use of the ACR TI-RADS, all nodules could be placed into one of five categories. A previous study showed that 3.4% of nodules cannot be classified using the ATA guidelines [6]. In the present study, 13.9% of nodules could not be categorized using the ATA guidelines, and 9.4% of these noncategorized nodules were malignant. For most noncategorized nodules, the important sonographic features could be assessed, but the combination of features was not included in the guidelines. An important additional class of nodules that could not be classified included nodules for which the composition, echogenicity, margins, or a combination of these features could not be determined. These nodules accounted for 28.7% of noncategorized nodules (137/477) assessed using the ATA guidelines and 85.8% of nodules (115/134) assessed using the KSThR TIRADS. The most common reason that features could not be determined was dense calcification with resulting shadowing. This problem was avoided in the ACR TI-RADS because specific instructions advised how to assign points if a feature could not be determined (composition, 2 points; echogenicity, 1 point; and margins, 0 points).

Two major differences between the ACR TI-RADS and the other systems is the elimination of a recommendation for FNA for nodules with certain features that are highly likely to be benign and the higher size thresholds for FNA of nodules with features of intermediate suspicion. For instance, the ACR TI-RADS does not recommend FNA of mixed solid and cystic nodules that are hyperechoic or isoechoic and have no malignant features, whereas the ATA and KSThR systems recommend FNA at size thresholds of 2 cm or larger and 1.5 cm or larger, respectively. Solid nodules that are hyperechoic or isoechoic and have no malignant features have an FNA size threshold of 2.5 cm or larger in the ACR TI-RADS, as compared with 1.5 cm or larger in the ATA and KSThR systems. The ACR TI-RADS does not recommend FNA of spongiform nodules of any size. The ATA guidelines recommend that FNA be performed on spongiform nodules 2 cm or larger in the evaluation and management algorithm (see Fig. 1 in [2]); the guidelines also recommend that FNA be considered as the first option for spongiform nodules with a size threshold of 2 cm or larger and then list observation without FNA as a second reasonable option in the risk stratification chart [2] and in the text itself. The KSThR TIRADS recommends in its risk stratification chart that FNA be performed on spongiform nodules at a threshold of 2 cm or larger [3], and it recommends in the text that FNA of spongiform nodules be selectively considered at a size of 2 cm or larger [3].

As we predicted, these differences produced fewer biopsies when the ACR TI-RADS was used. Table 3 shows the results of the three systems. Under the assumption that noncategorized nodules were biopsied at a size threshold of 1.5 cm or larger, the ACR TI-RADS would reduce the percentage of biopsies of benign nodules by more than twice as much as the other systems (52.9% for ACR TI-RADS, 21.9% for the ATA guidelines, and 20.3% for the KSThR TIRADS). This is one of the main strengths of the ACR TI-RADS and explains the higher biopsy yield compared with the other systems.

The ACR TI-RADS criteria that allow reduction in the percentage of benign nodules that are biopsied would also be expected to result in a lower percentage of malignant nodules that are biopsied. This is unavoidable because there are some malignancies that have benign sonographic features. As shown in Table 3, biopsy would not have been recommended for 31.8% of malignant nodules with the use of the ACR TI-RADS, compared with 24.1% and 21.3% with use of the ATA and KSThR systems, respectively, and under the assumption that noncategorized nodules were biopsied at a size threshold of 1.5 cm or larger.

With the knowledge that some cancers will be missed, the goal of all classification systems is to minimize the number of clinically significant cancers that are missed. Because localized papillary cancers smaller than 1 cm (i.e., papillary microcarcinoma) typically have minimal clinical significance [7, 8], a subanalysis of nodules 1 cm or larger was also performed. The percentage of malignant nodules that would be biopsied improved in all classification systems but was still lower when the ACR TI-RADS was used (83.3% for ACR TI-RADS vs 92.7% for the ATA guidelines and 96.2% for the KSThR TIRADS). The percentage of benign nodules that would be spared biopsy remained significantly higher with use of the ACR TI-RADS (49.9% for the ACR TI-RADS vs 17.0% for the ATA guidelines and 15.4% for the KSThR TIRADS).

Another important difference between the ACR TI-RADS and the other classification systems is the detailed and definitive recommendation for 5-year sonographic follow-up versus no further evaluation for all nodules that do not meet criteria for biopsy. When applied to the malignant nodules, these follow-up recommendations are advantageous and should result in detection of some of the cancers that otherwise would have been overlooked. In this database, the percentage of malignant nodules that would be either biopsied or followed using the ACR TI-RADS was 89.2% (68.2% biopsied and 21.0% monitored). This was actually significantly higher than the percentage of malignant nodules that would be biopsied using the ATA guidelines (75.9%) and the KSThR TIRADS (78.7%). The percentage of malignant nodules 1 cm or larger that would either be biopsied or monitored using the ACR system was 94.4% (83.3% biopsied and 11.1% monitored). This percentage was similar to the percentage that would be biopsied using the ATA and KSThR systems. When applied to the benign nodules, the ACR TI-RADS follow-up recommendations are detrimental because they lead to needless additional attention to and, in some cases, intervention of insignificant nodules. Nevertheless, the percentage of benign nodules that would either be biopsied or followed using the ACR recommendations (65.2%) was still lower than the percentage that would be biopsied using the ATA system (78.1%) or the KSThR system (79.7%).

The ATA guidelines provide follow-up recommendations for some categories of nodules that do not meet criteria for FNA. Repeat ultrasound in 6–12 months is recommended for nodules with high suspicion of malignancy. Consideration for repeat ultrasound in 12–24 months is recommended for nodules with low to intermediate suspicion for malignancy. For nodules with very low suspicion, no recommendation for or against follow-up of nodules larger than 1 cm exists, and the recommendation for nodules 1 cm or smaller is no follow-up. Unfortunately, the recommendations that were definitive did not apply to most nodules in this database. In

addition, the recommendations for nodules with intermediate or low suspicion of malignancy are not definitive. Because of these limitations, a definitive recommendation for no further evaluation (no FNA or follow-up) would be made for only 1.5% of nodules using the ATA guidelines, compared with 32.3% of nodules using the ACR TI-RADS. Follow-up recommendations are not included in the current KSThR TIRADS.

All of the systems are expected to undergo periodic revisions. Some of the deficiencies described in the present study can be easily eliminated. For instance, the problem of nonclassifiable nodules can be eliminated by incorporating minimal additional details in the nodule categorization system of the ATA guidelines and the KSThR TIRADS. Creation or refinement of recommendations for follow-up of nodules that do not meet criteria for FNA is more difficult because such recommendations must balance the competing goals of reducing the number of clinically significant malignancies that are missed and minimizing the attention and resources committed to benign nodules. Adjustments in size thresholds can also produce significant changes in the percentage of nodules requiring FNA, although such adjustments will need to be guided by large, prospective, longitudinal studies that evaluate outcomes and costs.

The present study has several limitations. Most patients who undergo thyroid sonography do not undergo FNA. The database from which our results are calculated was derived only from the small number of patients who did undergo FNA and thus does not reflect the total population of patients who physicians encounter when they are deciding whether to recommend nodule FNA, follow-up, or no further evaluation [9]. In addition, 10.5% of nodules were excluded because no definitive cytologic or histologic diagnosis existed. The degree to which inclusion of such patients would affect our results is uncertain. A high-volume study that includes long-term clinical or sonographic follow-up or both types of follow-up of all patients who have sonographically detected thyroid nodules but do not undergo FNA or surgery would be beneficial to minimize this limitation. Nevertheless, we believe that this limitation would affect the three systems in a similar manner and would not negate the comparative results that we obtained.

The database was created at a time when the significance of very hypoechoic echogenicity was not fully understood. This may have altered the results for the ACR TI-RADS, because that system assigns a 1-point difference between hypoechoic and very hypoechoic nodules. If so, it is possible that the effect would lead to a lower TIRADS category for some nodules and lowered sensitivity for the detection of malignancy. It would not have changed the results of the other systems because they do not discriminate between hypoechoic and very hypoechoic nodules. The significance of a taller-than-wide shape also was not well recognized at the time that the database was created and therefore was not analyzed. This may have decreased the sensitivity for detection of malignant nodules. However, because all three guidelines treat this finding as a malignant feature, we do not believe that this would have caused significant differences in the comparison of the three systems.

Several options exist for accounting for nodules that were not classifiable using the ATA and KSThR systems. We did not think it would be reasonable to treat these nodules in a manner similar to the treatment of nodules with high suspicion for malignancy that have an FNA size threshold of 1.0 cm or greater. We also did not think that it would be reasonable to treat these nodules similar to nodules of very low suspicion or benign nodules, which have either a size threshold for FNA of 2.0 cm or larger or a recommendation for no FNA. Therefore, we chose to focus our comparison under the assumption that these nodules would be biopsied at a size threshold of 1.5 cm or larger. Our subanalysis, which calculated results using several different approaches to nonclassifiable nodules, showed that differences were minimal regardless of how these nodules were handled.

In conclusion, the 2017 ACR TI-RADS compares favorably with the 2015 ATA guidelines and the 2016 KSThR TIRADS. The ACR system has a higher biopsy yield of malignancy, primarily because of a reduced number of biopsies of benign nodules. It also eliminates further evaluation (biopsy or follow-up) of a much higher number of nodules. The ATA and KSThR systems result in biopsy of a higher percentage of malignant nodules, but this difference is at least partially mitigated by the follow-up recommendations included in the ACR system. These results, along with future prospective longitudinal studies that look at outcomes and costs, should help guide revisions of all systems and perhaps lead to a single unified multidisciplinary system that can be accepted internationally.

Acknowledgment

We thank Amber Salter for her statistical analysis of our data.

References

1. Tessler FN, Middleton WD, Grant EG, et al. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): white paper of the ACR TI-RADS committee. *J Am Coll Radiol* 2017; 14:587–595
2. 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
3. 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
4. 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; 237:794–800
5. Middleton WD, Teefey SA, Reading CC, et al. Multiinstitutional analysis of thyroid nodule risk stratification using the American College of Radiology Thyroid Imaging Reporting and Data System. *AJR* 2017; 208:1331–1341
6. Yoon JH, Lee HS, Kim EK, Moon HJ, Kwak JY. Malignancy risk stratification of thyroid nodules: comparison between the Thyroid Imaging Reporting and Data System and the 2014 American Thyroid Association management guidelines. *Radiology* 2016; 278:917–924
7. Harach HR, Franssila KO, Wasenius VM. Occult papillary carcinoma of the thyroid: a "normal" finding in Finland—a systematic autopsy study. *Cancer* 1985; 56:531–538
8. Davies L, Welch HG. Current thyroid cancer trends in the United States. *JAMA Otolaryngol Head Neck Surg* 2014; 140:317–322
9. Smith-Bindman R, Lebda P, Feldstein VA, et al. Risk of thyroid cancer based on thyroid ultrasound imaging characteristics: results of a population-based study. *JAMA Intern Med* 2013; 173:1788–1796