ORIGINAL ARTICLE

Journal of the ASEAN Federation of Endocrine Societies

Comparison of the Diagnostic Performance of Ultrasound-Based Thyroid Imaging Reporting and Data System (TIRADS) Classification with American Thyroid Association (ATA) Guidelines in the Prediction of Thyroid Malignancy in a Single Tertiary Center in Manila, Philippines

Ryan James Gacayan,[1] Ruben Kasala,[1] Ma. Patricia Puno-Ramos,[1] Dondee Jules Mojica,[2] Ma. Krisha Castro[2]

[1]*Section of Endocrinology, Diabetes and Metabolism, Department of Medicine, The Medical City, Philippines*
[2]*Department of Radiology, The Medical City, Philippines*

Abstract

Objective. To compare the diagnostic performance of American College of Radiology-Thyroid Image Reporting and Data Systems (ACR-TIRADS) and the American Thyroid Association (ATA) guidelines on screening for thyroid malignancy.

Methodology. A cross-sectional criterion-referenced study involving Filipino patients with thyroid nodules, 18-80 years old, who underwent ultrasound guided fine needle aspiration biopsy at the Thyroid Clinic of The Medical City from July to December 2019. The ACR-TIRADS and the ATA guidelines were compared for 197 nodules. Standard diagnostic parameters were calculated, namely sensitivity, specificity, PPV, NPV, positive and negative likelihood ratios and overall accuracies.

Results. The risks of malignancy were 15% and 22% for TIRADS 4 and 5 respectively. For ATA guidelines, it's 2%, 20%, and 15% for nodules with low, intermediate, and high suspicion respectively. The sensitivity, specificity, PPV, NPV, and accuracy of the American College of Radiology Thyroid Imaging Reporting and Data System (ACR TIRADS) in relation to Fine Needle Aspiration Cytology (FNAC) is 100%, 52.2%, 16.5%, 100%, and 56.4% respectively. For the American Thyroid Association (ATA) guidelines it is 88.2%, 57.8%, 16.5%, 98.1%, and 60.4% respectively.

Conclusion. The ACR TIRADS classifications appears to be more sensitive than the ATA classification. The ATA guidelines prove to be a more specific test. Each tool has its unique advantages and disadvantages. Therefore, clinicians must use these tools with utmost vigilance to avoid over or under diagnosis and to avoid unnecessary thyroid nodule biopsies.

Key words: Thyroid Imaging Reporting and Data System, American Thyroid Association, thyroid cancer, malignancy risk, thyroid nodules, ultrasound of thyroid

INTRODUCTION

Depending on the study population, the prevalence of thyroid nodules ranges from as low as 2% to as high as 35%.[1,2] In a 2012 nationwide study from Carlos-Raboca et al.,[3] involving 4,897 subjects, the estimated prevalence of nodular goiter in the Philippines is 8.9%.

The American Thyroid Association lists these sonographic findings suggestive of malignancy: solid nodules, nodule hypoechogenicity or marked hypoechogenicity, irregular margins, microcalcifications and a shape taller than wide on a transverse view. The varied spectrum of sizes and characteristics of thyroid nodules makes it difficult to select which nodule is a candidate for fine needle aspiration biopsy (FNAB).[4] Fine needle aspiration biopsy is a minimally invasive diagnostic procedure with published sensitivity and specificity that ranges between 65% to 98% and 73% to 100%, respectively.[5] The accuracy of ultrasound guided FNAB in different studies, both locally and abroad may range from 77.3% to as high 96.7%.[5,6] According to Cibas et al., only 3% to 7% of FNA cytology are malignant and most nodules are benign.[7] Given this, it is important to use an ultrasound classification that will help differentiate benign from malignant thyroid nodules to determine which nodule(s) will require FNAB and decrease unnecessary procedures.

Kwak et al., sought to implement a similar standardized model for thyroid nodules with the release of the Thyroid Imaging Reporting and Data System (TIRADS) in 2011.[8] In addition, in order to avoid the over usage of FNA for multiple benign thyroid nodules, several reports investigated the risk of malignant nodules for ultrasound-guided biopsy due to suspicious ultrasonographic features.[9-10] Park et al.,[11] and Horvath et al.,[12] established

ISSN 0857-1074 (Print) | eISSN 2308-118x (Online)
Printed in the Philippines

Received: May 3, 2020. Accepted: February 7, 2021.
Published online first: May 27, 2021.
https://doi.org/10.15605/jafes.036.01.14

Corresponding author: Ryan James B. Gacayan, MD
Section of Endocrinology, Diabetes and Metabolism
Department of Medicine, The Medical City
Ortigas Avenue, Pasig City, Metro Manila, Philippines
Tel. No. +632-8988-1000
E-mail: ryan_james8423@yahoo.com
ORCiD: https://orcid.org/0000-0002-9932-2602

a thyroid ultrasonographic system to stratify cancer risk and developed several categories based on 10 and 12 sonographic features, called the thyroid imaging reporting and data system (TIRADS).

An updated version endorsed by the American College of Radiology (ACR) was released in 2017.[8,12] This does not include subcategories, nor does it include TIRADS 0 category which indicates a normal thyroid gland.[13] TIRADS categories range from TIRADS 1 to TIRADS 5.

A retrospective study by Middleton et al., comparing TIRADS with ATA and other scoring system showed that 13.9% of nodules could not be categorized using ATA guidelines and 9.4% of these non-categorized nodules were malignant.[14] The committee of ACR-TIRADS decided against the pattern-based approach used by ATA based on the results of a study by Yoon et al., which showed that using ATA guidelines, they were unable to classify 3.4% of 1,293 nodules, of which 18.2% were malignant. In this study, they only included nodules that were subjected to FNA or surgery.[8]

This study follows on the findings of a retrospective study done at The Medical City, Manila, Philippines by Dy and Kasala et al. The authors recommended a multicenter prospective study for the use of TIRADS as their research concluded that TIRADS was sensitive in recognizing patients with thyroid cancer and can be used as a guide in deciding the need for fine needle aspiration biopsy.[15]

GENERAL OBJECTIVES

To compare the diagnostic performance of Thyroid Imaging reporting and data System (TIRADS) and the American Thyroid Association (ATA) guidelines on screening for thyroid malignancy.

SPECIFIC OBJECTIVES

To determine the diagnostic performance of TIRADS in screening for thyroid malignancy in terms of:

a. Sensitivity and specificity
b. Positive Predictive Value and Negative Predictive Value
c. Positive Likelihood Ratio and Negative Likelihood Ratio
d. Diagnostic accuracy

To determine the diagnostic performance of ATA in screening for thyroid malignancy in terms of:

a. Sensitivity and specificity
b. Positive Predictive Value and Negative Predictive Value
c. Positive Likelihood Ratio and Negative Likelihood Ratio
d. Diagnostic accuracy

METHODOLOGY

Study design

This is a cross-sectional criterion-referenced study approved by our institutional review board. During the time frame specified, all patients who were undergoing ultrasound-guided FNA were asked to participate in the study with a signed informed consent.

Study population

This study enrolled Filipino patients with thyroid nodules aging 18 to 80 years old who underwent ultrasound guided fine needle aspiration biopsy of thyroid nodules at the Thyroid Clinic of The Medical City from July 2019 to December 2019. Patients are excluded if the cytology report of the FNAB is inadequate or non-diagnostic.

Sample size

A minimum of 90 patients were required for this study based on a level of significance of 5%, a prevalence of 33.56%, sensitivity of 98% with a marginal error of 0.05. The values for the prevalence of thyroid malignancy and sensitivity of TIRADS were based from the study by Dy and Kasala et al.[15]

Description of study procedure

All qualified subjects underwent a repeat ultrasound of thyroid gland using BK Flex Focus 800 ultrasound machine prior to their scheduled ultrasound guided fine needle aspiration biopsy. The scanning protocol in our study includes scanning of thyroid gland and cervical lymph nodes in both transverse and longitudinal planes by B–mode (brightness mode), CCDI (Color-coded Doppler imaging) and PDI (Power Doppler imaging). The ultrasonography of the thyroid gland was done by a second-year radiology resident. It was then reviewed and read by only 1 radiologist with more than ten years experience. The nodules were analyzed according to their type (solid, cystic, or mixed), echogenicity, margins, shape, echogenic foci, and evidence of calcification. The reports were categorized into two, ACR TIRADS and conventional ATA guidelines respectively.

American College of Radiology – Thyroid Imaging, Reporting and Data System (ACR TIRADS) described nodules according to composition, echogenicity, shape, margin, and echogenic foci and a corresponding point or points will be given. Points were added from all categories to determine the TIRADS level and nodules were classified into the following: TIRADS 1 benign, TIRADS 2 not suspicious, TIRADS 3 mildly suspicious, TIRADS 4 moderately suspicious, and TIRADS 5 highly suspicious (Appendix A).

A second report was provided and the nodules were described according to its size, location, composition (solid, cystic proportion, or spongiform), echogenicity, margins, presence and type of calcifications, and shape if taller than wide, and vascularity. The nodules were re-classified according to American Thyroid Association (ATA) Guidelines 2015 into the following based on the sonographic pattern: benign, very low suspicion, low suspicion, intermediate suspicion, and high suspicion (Appendix B).

The cytology reports were used to classify nodules into five categories using Bethesda Classification: I for non-diagnostic, II for benign, III for atypia of undetermined significance, IV for follicular neoplasm or suspicious for follicular neoplasm, V suspicious for malignancy, VI for malignant. Nodules with FNA results that were classified as Bethesda II to VI were considered diagnostic and included in the final analysis. The nodules classified as Bethesda cytology IV, V, and VI were considered as suspicious for malignancy and Bethesda cytology II and III were benign.

All other nodules were excluded unless the nodule was resected and histologic findings were available.

Statistical analysis

Descriptive statistics were used to summarize the data: frequency and proportion for nominal variables; median (range) and mean ± standard deviation (SD) for interval/ratio variables with and without normal distributions, respectively. Test on proportions was used to determine differences in proportions of nodules recommended for FNA between TIRADS and ATA.

Standard diagnostic parameters were calculated for the two sonographic criteria, namely sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and positive and negative likelihood ratios (LR), all with their corresponding 95% confidence intervals (CI). McNemar's test was used to compare overall accuracies of TIRADS and ATA recommendations for FNA. All valid data were included in the analysis. Missing variables were neither imputed nor estimated. Null hypotheses were rejected at 0.05 α-level of significance. STATA 15.0 was used for data analysis.

RESULTS

A total of 197 nodules from 121 patients (Figure 1), with median age of 53 (21–77) years and comprised mostly of females (85%), were included in the analysis (Table 1). Thyroid nodules were located almost equally on either side. Sonographically, half of the lesions measured 1.0–1.9 cm, 81% were solid or almost completely solid, 70% were hyperechoic or isoechoic, 88% were wider-than-tall, 92% possessed smooth margins, and 67% contained no echogenic focus or only a large comet-tail artifact.

Most patients were classified as moderately (41%) or mildly (35%) suspicious for malignancy by TIRADS classification. By ATA guidelines, 43% and 31% were of low and high suspicion, respectively. Cytologic analysis revealed most nodules (83%) to be benign, and only 8% were suspicious or obviously malignant. Of the latter nodules, 40% turned out benign on surgical histopathology.

The risks of malignancy were 15% for nodules considered moderately suspicious (TIRADS 4) and 22% for those that were highly suspicious (TIRADS 5). No malignant diagnoses were made among those with lower grade classification (TIRADS 1,2, and 3). The largest mean (± SD) nodular size was with mildly suspicious lesions, at 2.3±1.0 cm (Table 2).

On the other hand, the risks of malignancy were 2%, 20%, and 15% for nodules at low, intermediate, and high suspicion by ATA guidelines (Table 3). There were no malignant findings among the lower grade lesions. The largest lesion sizes were with low (2.3±1.0 cm) and high (2.2±1.1 cm) suspicion.

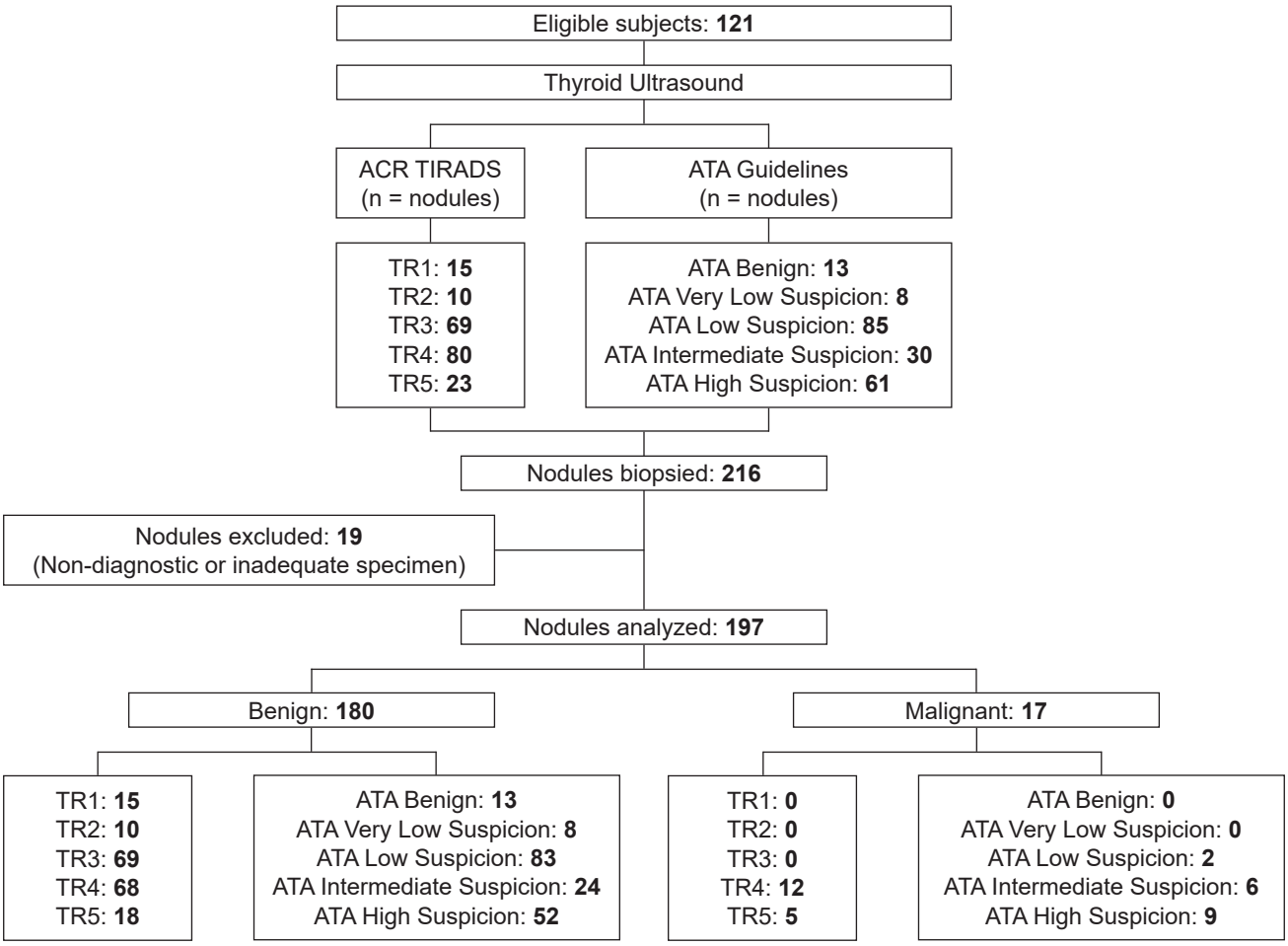

Figure 1. Flow chart of patients in the study. *TR – TIRADS, ATA – American Thyroid Association*

Table 1. Patient and nodule characteristics (n=197 nodules, 121 patients)

	Median (Range); Count (%)
Age (years)	53 (21–77)
Sex	
Male	18 (14.88)
Female	103 (85.12)
Solitary nodule	59 (48.76)
Nodule (cm)	
0–0.4	0
0.5–0.9	11 (5.58)
1.0–1.4	51 (25.89)
1.5–1.9	48 (24.37)
2.0–2.4	35 (17.77)
2.5–2.9	13 (6.60)
≥ 3.0	39 (19.80)
Location	
Left	96 (48.73)
Right	97 (49.24)
Isthmus	4 (2.03)
Composition	
Cystic	1 (0.51)
Almost completely cystic	10 (5.08)
Spongiform	4 (2.03)
Mixed cystic & solid	23 (11.68)
Solid or almost completely solid	159 (80.71)
Echogenicity	
Anechoic	14 (7.11)
Hyperechoic or isoechoic	138 (70.05)
Hypoechoic	40 (20.30)
Very Hypoechoic	5 (2.54)
Shape	
Wider-than-tall	174 (88.32)
Taller-than-wide	23 (11.68)
Margin	
Smooth	181 (91.88)
Ill-defined	11 (5.58)
Lobulated or irregular	5 (2.54)
Extra-thyroidal extension	0
Echogenic foci	
None or large comet-tail artifact	132 (67.01)
Macrocalcification	26 (13.20)
Peripheral or rim calcifications	9 (4.57)
Punctate echogenic foci	30 (15.23)
ACR TIRADS category	
Benign	15 (7.61)
Not suspicious	10 (5.08)
Mildly suspicious	69 (35.03)
Moderately suspicious	80 (40.61)
Highly suspicious	23 (11.68)
ATA classification	
Benign	13 (6.60)
Very low suspicion	8 (4.06)
Low suspicion	85 (43.15)
Intermediate suspicion	30 (15.23)
High suspicion	61 (30.96)
Bethesda system[a]	
I	0
II	164 (83.25)
III	16 (8.12)
IV	1 (0.51)
V	11 (5.58)
VI	5 (2.54)
Histopathology (n=15)	
Papillary thyroid carcinoma	8 (53.33)
Multinodular colloid goiter	6 (40.00)
Follicular thyroid carcinoma	1 (6.67)

[a] Bethesda system: I, non-diagnostic or unsatisfactory; II, benign; III, atypia of undetermined significance or follicular lesion of undetermined significance; IV, follicular neoplasm or suspicious for it; V, suspicious for malignancy; VI, malignant.

Table 2. Nodule size and risk of malignancy, by TIRADS category (n=197)

TIRADS	FNAC		Risk of Malignancy (%)	Nodule Size (cm)
	Suspicious for Malignancy (n=17)	Benign (n=180)		
Benign (TR1)	0/15	15/15	0	1.73 ± 0.60
Not suspicious (TR2)	0/10	10/10	0	1.91 ± 0.47
Mildly suspicious (TR3)	0/69	69/69	0	2.30 ± 1.03
Moderately suspicious (TR4)	12/80	68/80	15.00	2.06 ± 0.98
Highly suspicious (TR5)	5/23	18/23	21.74	1.94 ± 1.35

Table 3. Nodule size and risk of malignancy, by ATA classification (n=197)

ATA	FNAC		Risk of Malignancy (%)	Nodule Size (cm)
	Suspicious for Malignancy (n=17)	Benign (n=180)		
Benign	0/13	13/13	0	1.82 ± 0.59
Very low suspicion	0/8	8/8	0	1.77 ± 0.49
Low suspicion	2/85	83/85	2.35	2.30 ± 1.02
Intermediate suspicion	6/30	24/30	20.00	1.52 ± 0.71
High suspicion	9/61	52/61	14.75	2.20 ± 1.13

Using TIRADS

Using FNAC as the gold standard, TIRADS had a high sensitivity, with 100% (17/17) of the nodules that were suspicious for malignancy by FNAB having positive finding (IV-V). However, its ability to correctly rule out benign nodules was quite low, with only 52.2% of those who had negative findings (I-III) having benign nodules (specificity) (Table 4).

Positive findings in the TIRADS were about 2.09 times as likely to be observed in malignant nodules as compared to benign nodules (LR+). Negative findings were about 100% less likely to be seen in malignant nodules as opposed to benign nodules (LR-) (Table 4).

Using TIRADS, the probability that positive findings are malignant in the FNAC is 16.5% (PPV), whereas negative findings have 100% chance of having benign results (NPV) (Table 4).

Using ATA

ATA compared to TIRADS had a lower sensitivity with 88.2% (15/17) of nodules that were suspicious for malignancy by FNAB had a positive ATA category of IV-V. The system was unsatisfactory in correctly ruling out benign nodules, with only 57.8% of those who had negative findings (category I-III) having benign nodules (specificity) (Table 5).

Positive findings in the ATA were about 2.09 times as likely to be observed in malignant nodules as compared to benign nodules (LR+). Negative findings were about 80% less likely to be seen in malignant nodules as opposed to benign nodules (LR-) (Table 5).

Table 4. Accuracy of FNA recommendation to detect malignancy using TIRADS

TIRADS	Malignant (FNAC+)	Benign (FNAC-)	Total
	Frequency (%)		
TIRADS IV-V	17 (100)	86 (47.78)	103 (52.28)
TIRADS I-III	0	94 (52.22)	94 (47.72)
Total	17 (8.63)	180 (91.37)	197
Sensitivity		100% (80.5 to 100)	
Specificity		52.2% (44.7 to 59.7)	
Positive Predictive Value (PPV)		16.5% (14.5 to 18.7)	
Negative Predictive Value (NPV)		100%	
Positive Likelihood Ratio		2.09 (1.80 to 2.44)	
Negative Likelihood Ratio		0	
Accuracy		56.4% (49.1 to 63.4)	

Table 5. Accuracy of FNA recommendation to detect malignancy using ATA

ATA	Malignant (FNAC+)	Benign (FNAC-)	Total
	Frequency (%)		
ATA IV-V	15 (88.24)	76 (42.22)	91 (46.19)
ATA I-III	2 (11.76)	104 (57.78)	106 (53.81)
Total	17 (8.63)	180 (91.37)	197
Sensitivity		88.2% (63.6 to 98.5)	
Specificity		57.8% (50.2 to 65.1)	
Positive Predictive Value (PPV)		16.5% (13.4 to 20.1)	
Negative Predictive Value (NPV)		98.1% (93.4 to 99.4)	
Positive Likelihood Ratio		2.09 (1.64 to 2.67)	
Negative Likelihood Ratio		0.20 (0.06 to 0.75)	
Accuracy		60.4% (53.2 to 67.3)	

Table 6. Summary of diagnostic performance of TIRADS and ATA

	ATA	TIRADS
Sensitivity (%)	88.2 (63.6 to 98.5)	100% (80.5 to 100)
Specificity (%)	57.8% (50.2 to 65.1)	52.2% (44.7 to 59.7)
PPV (%)	16.5% (13.4 to 20.1)	16.5% (14.5 to 18.7)
NPV (%)	98.1% (93.4 to 99.4)	100%
Positive LR	2.09 (1.64 to 2.67)	2.09 (1.80 to 2.44)
Negative LR	0.20 (0.06 to 0.75)	0
Accuracy (%)	60.4% (53.2 to 67.3)	56.4% (49.1 to 63.4)
McNemar's test p-value	0.004	

Table 6.1. Comparison of ATA and TIRADS

	TIRADS		ATA	
	I-III	IV-V	I-III	IV-V
	Frequency (%)			
All nodules (n=197)				
Malignant nodules based on FNAC (n=17)	0	17 (100)	2 (11.76)	15 (88.24)
Benign nodules based on FNAC (n=180)	94 (52.22)	86 (47.78)	104 (57.78)	76 (42.22)
Solitary nodules (n=59)				
Malignant solitary nodules (n=6)	0	6 (100)	1 (16.67)	5 (83.33)
Benign solitary nodules (n=53)	17 (32.08)	36 (67.92)	21 (39.62)	32 (60.38)

Using ATA, the probability that negative findings are malignant in the FNAC is 16.5% (PPV), whereas positive findings have 98.1% chance of having benign results (NPV) (Table 5).

Overall accuracies of the FNA recommendation by TIRADS and ATA criteria were moderate (56.4% [95% CI 49.1–63.4) and 60.4% [95% CI 53.2–67.3], *P*=0.004). The former's sensitivity was high at 100% (95% CI 80.5–100), but the latter was inferior at 88.2% (95% CI 63.6–98.5). Both sonographic criteria had NPV's above 95% (Table 6).

DISCUSSION

The pathological nature of thyroid nodules directly affects the therapeutic decisions and patient prognosis; therefore, the correct diagnosis of thyroid nodules at an early stage has important clinical significance. However, conventional sonographic diagnoses for thyroid nodules presents limitations related to overlapping boundaries, morphologies, internal blood streams, and echoes between malignant and benign nodules. In addition, subjective factors related to the diagnostician can also affect the accuracy of the diagnosis. Therefore, research by Kwak,[8] Park[11] and Horvath[12] indicates that the thyroid imaging reporting and data system (TIRADS) can be used to improve the diagnostic accuracy of thyroid nodules by ultrasound, which will provide improvements that can be used in clinical practice. This study was done to compare the diagnostic performance of ACR TIRADS and ATA guidelines for predicting risk of thyroid malignancy.

The suggested risk of malignancy for TIRADS is less than 2% for TIRADS 1 and TIRADS 2, 5% for TIRADS 3, 5-20% in TIRADS 4, and greater than 20% for TIRADS 5.[16] In our study, the risk of malignancy was 15% for nodules considered moderately suspicious or TIRADS 4 and 22% for those that were highly suspicious or TIRADS 5 which are well matched to the suggested risk of malignancy by ACR TIRADS. When compared to another local study done by Dy and Kasala et al.,[15] the malignancy risk for TIRADS 4 was 12.82% to 53% and is well matched with our result. The malignancy risk for TIRADS 5 in the former study was 66.67% which is higher than our result.[15] Selection bias may have contributed to the very high malignancy risk since it was a retrospective study.

The risk of malignancy recommended by the ATA is more than 70-90% for the high suspicion pattern, 10-20% for the intermediate suspicion pattern, 5-10% for the low suspicion pattern, less than 3% for the very low suspicion pattern and less than 1% for the benign pattern.[4] The risk of malignancy was 2%, 20%, and 15% for nodules at low, intermediate, and high suspicion respectively in our study by ATA guidelines. Only intermediate suspicion nodules matched the suggested risk of malignancy by ATA guidelines.

The diagnostic performance of both ACR TIRADS and the ATA guidelines are one of the most commonly compared sonographic classification of nodules in various studies. They both have outstanding performances with sensitivity ranging from 70% to 90% and specificity of 33% to 67%.[16] These are international studies and most of them are retrospective in nature.

In a local retrospective study by Dy and Kasala et al., TIRADS classification for predicting thyroid malignancy still maintained a very high sensitivity of 98%. However,

specificity is quite low at 7.07% which is also lower compared to other studies.[15] In our study, ACR TIRADS had a high sensitivity of 100% which is higher than Dy and Kasala et al., (98%)[15] Horvath et al., (88%)[11] Ha et al., (74.7%)[16] and Grani et al. (83%).[17]

On the other hand, ACR TIRADS in our study had a specificity of 52.2% which is higher than Dy and Kasala et al., (7.07%)[15] and Horvath et al., (49%)[11] but slightly lower than Grani et al., (56.2%)[17] and Ha et al., (67.3%).[16] The overall accuracy of ACR TIRADS is 56.4% which is higher than Dy and Kasala et al., (53%)[15] but lower than Ha et al., (69%)[16] and Horvath et al., (94%).[11] The low accuracy of ACR TIRADS in our study is probably due to high false positive rate (48%).

The prevalence of malignancy in our study is 8.63% which is similar to the malignancy rate according to Cibas et al.,[7] at 3% to 7%. When using ACR TIRADS, the probability that FNA recommended nodules are malignant in the FNAC is 16.5% (PPV) which is almost similar to Grani et al., (12.8%)[17] but lower than Ha et al,. (40.2%)[16] Horvath et al., (49%)[11] and Dy and Kasala et al., (34.75%)[15]. Thyroid nodules that are not recommended for biopsy when using ACR TIRADS in our study have 100% chance of having benign results (NPV) which is higher than other studies such as Dy and Kasala et al., (87.5%)[15] Ha et al., (90.1%)[16] Grani et al., (97.8%)[17] and Horvath et al., (88%).[11]

When using the ATA guidelines for predicting thyroid malignancy in our study, it had a sensitivity of 88.2% which is similar with Ha et al., (89.6%)[16] and higher than Grani et al., (75%)[17] specificity of 57.8% which is noted to be higher than Ha et al., (33.2%)[16] and Grani et al., (45.3%).[17] The overall accuracy of ATA guidelines in predicting thyroid malignancy in our study is 60.4% which is higher than Ha et al., (46%).[16] Using ATA, the probability that FNA recommended nodules are malignant in the FNAC is 16.5% (PPV) which is higher with Grani et al., (9.6%)[17] but lower than Ha et al., (28.3%).[16] Nodules that are not recommended for FNA have 98.1% chance of having benign results (NPV) which is higher than Ha et al., (91.6%)[16] and Grani et al., (95.9%).[17]

In summary, the ACR TIRADS classification when compared to the ATA guidelines had high sensitivity (100% vs 88.2%) in which the latter had more false negative results. The ACR TIRADS classification was less specific (52.2%% vs 57.8%) when compared to ATA in which the former had more false positive results. Both had equal PPV (10.9% vs 10.5%) and NPV (100% vs 98.1%) was slightly lower for the ATA. The ACR TIRADS had inferior overall accuracy (56.4% vs 60.4%) as compared to ATA in which the latter had more correctly identified nodules.

CONCLUSION

In conclusion, the ACR TIRADS and ATA guidelines provided the usefulness of ultrasound based risk stratifications of thyroid malignancy. The diagnostic performances of ultrasound-based risk stratification tools differed between ACR TIRADS and ATA guidelines. The ACR TIRADS classifications appears to be more sensitive than the ATA classification. The ATA guidelines on the other hand proves to be more specific test. Each tool has its unique advantages and disadvantages. Therefore, clinicians must use these tools with utmost vigilance to avoid over or under diagnosis and to avoid unnecessary thyroid nodule biopsies.

Limitations, Strengths and Recommendations

This study had several identified limitations. First, the gold standard used in this study is FNA cytology and can yield a false-negative result of up to 3.7% based on meta-analysis,[18] however, it would be unethical to surgically resect all nodules included in this study and confirm the diagnosis. Second, there might be an overestimation of the proportion of nodules with malignancy since this is based on Bethesda Class IV to VI, rather than Bethesda VI alone or the surgical pathology report since not all patients underwent surgery. Third, we had a small sample size as compared to other bigger studies because it was underestimated in the initial sample size calculation. Fourth, this study was done in a single institution which may reflect the relatively small sample size and might not be representative of the entire population. Fifth, the nodules for biopsy were already flagged by the referring physician, the criteria for classification of these nodules were not known.

The major strength of this study is that the nodules that were for biopsy were examined in real-time ultrasonography before sample is obtained and as compared to retrospective studies, we are confident that the nodules being biopsied are the nodules being sonographically classified.

As for our recommendations, a prospective multicenter study and a longer duration of study is highly recommended to achieve a greater number of subjects.

Acknowledgments
The authors would like to extend their deepest gratitude to Elizabeth Y. Arcellana-Nuqui, MD and Elizabeth Ann S. Alcazaren for providing the cytopathology and histopathology records of the patients in the study and to Brabim Giri, MD for being one of their research mentors.

Statement of Authorship
All authors certified fulfillment of ICMJE authorship criteria.

Author Disclosure
The authors declared no conflict of interest.

Funding Source
None.

References

1. Ezzat S, Sarti DA, Cain DR, Braunstein GD. Thyroid incidentalomas: Prevalence by palpation and ultrasonography. Arch Int Med. 1994;154(16):1838-40. PMID: 8053752. https://doi.org/10.1001/archinte.154.16.1838.
2. Dean DS, Gharib H. Epidemiology of thyroid nodules. Best Pract Res Clin Endocrinol Metab. 2008;22(6):901-11. PMID: 19041821. https://doi.org/10.1016/j.beem.2008.09.019.
3. Carlos-Raboca J, Jimeno CA, Kho SA, et al. The Philippine Thyroid Diseases Study (PhilTiDeS 1): Prevalence of thyroid disorders among adults in the Philippines. J ASEAN Fed Endocr Soc. 2012;27(1):27-33. https://doi.org/10.15605/jafes.032.01.05.
4. 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):1-33. PMID: 26462967. PMCID: PMC4739132. https://doi.org/10.1089/thy.2015.0020.
5. Muratli A, Erdogan N, Sevim S, Unal I, Akyuz S. Diagnostic efficacy and importance of fine-needle aspiration cytology of thyroid nodules. J Cytol. 2014;31(2):73-8. PMID: 25210233. PMCID: PMC4159900. https://doi.org/10.4103/0970-9371.138666.

6. Young J, Lumapas-Gonzalez CG, Mirasol R. The diagnostic accuracy of ultrasound guided fine-needle aspiration biopsy and intraoperative frozen section examination in nodular thyroid disease. J ASEAN Fed Endocr Soc. 2011;26(1):73-8. https://doi.org/10.15605/jafes.026.01.09.
7. Cibas ES, Ali SZ. The Bethesda system for reporting thyroid cytopathology. Thyroid. 2009;19(11):1159-65. PMID: 19888858. https://doi.org/10.1089/thy.2009.0274.
8. Kwak JY, Han KH, Yoon JH, et al. Thyroid imaging reporting and data system for US features of nodules: A step in establishing better stratification of cancer risk. Radiology. 2011;260(3):892-9. PMID: 21771959. https://doi.org/10.1148/radiol.11110206
9. American College of Radiology. Breast imaging reporting and data system (BI-RADS) ultrasound. Reston, VA: American College of Radiology; 2003.
10. Lazarus E, Mainiero MB, Schepps B, Koelliker SL, Livingston LS. BI-RADS lexicon for US and mammography: Interobserver variability and positive predictive value. Radiology. 2006;239(2):385-91. PMID: 16569780. https://doi.org/10.1148/radiol.2392042127.
11. Park JY, Lee HJ, Jang HW, et al. A proposal for a thyroid imaging reporting and data system for ultrasound features of thyroid carcinoma. Thyroid. 2009;19(11):1257-64. PMID: 19754280. https://doi.org/10.1089/thy.2008.0021.
12. Horvath E, Majlis S, Rossi R, et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. J Clin Endocrinol Metab. 2009;94(5):1748-51. PMID: 19276237. https://doi.org/10.1210/jc.2008-1724.
13. Tessler FN, Middleton WD, Grant EG, et al. ACR thyroid imaging, reporting and data system (TI-RADS): White paper of the ACR TI-RADS committee. J Am Coll Radiology. 2017;14(5):587-95. PMID: 28372962. https://doi.org/10.1016/j.jacr.2017.01.046.
14. Middleton WD, Teefey SA, Reading CC, et al. Comparison of performance characteristics of American College of Radiology TI-RADS, Korean Society of Thyroid Radiology TIRADS, and American Thyroid Association guidelines. AJR Am J Roentgenol. 2018;210(5):1148-54. PMID: 29629797. https://doi.org/10.2214/AJR. 17.18822.
15. Dy JG, Kasala R, Yao C, Ongoco R, Mojica DJ. Thyroid Imaging Reporting and Data System (TIRADS) in Stratifying Risk of Thyroid Malignancy at The Medical City. J ASEAN Fed Endocr Soc. 2017;32(2):108. PMID: 33442093. PMCID: PMC7784109. https://doi.org/10.15605/jafes.032.02.03.
16. Ha EJ, Na DG, Baek JH, Sung JY, Kim JH, Kang SY. US fine-needle aspiration biopsy for thyroid malignancy: Diagnostic performance of seven society guidelines applied to 2000 thyroid nodules. Radiology. 2018;287(3):893-900. PMID: 29465333. https://doi.org/10.1148/radiol.2018171074
17. Grani G, Lamartina L, Ascoli V, et al. Reducing the number of unnecessary thyroid biopsies while improving diagnostic accuracy: Toward the "right" TIRADS. J Clin Endocrinol Metab. 2018;104(1):95-102. PMID: 30299457. https://doi.org/10.1210/jc.2018-01674.
18. Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda System for reporting thyroid cytopathology: A meta-analysis. Acta Cytol. 2012;56(4):333–9. PMID: 22846422. https://doi.org/10.1159/000339959.

APPENDICES

Appendix A. American College of Radiology – Thyroid Imaging, Reporting and Data System (ACR- TIRADS)[14]

Composition		Echogenicity		Shape		Margin		Echogenic Foci	
Cystic or almost completely cystic	0	Anechoic	0	Wider-than-tall	0	Smooth	0	None or large comet tail artifacts	0
Spongiform	0	Hyperechoic or isoechoic	1	Taller-than-wide	3	Ill-defined	0	Macrocalcifications	1
Mixed Cystic or Solid	1	Hypoechoic	2			Lobulated or irregular	2	Peripheral (rim) calcifications	2
Solid or almost completely solid	2	Very Hypoechoic	3			Extra-thyroidal extension	3	Punctate echogenic foci	3

Points	0 points	2 points	3 points	4 to 6 points	7 points or more
TIRADS Scores	TIRADS 1	TIRADS 2	TIRADS 3	TIRADS 4	TIRADS 5
Interpretation	Benign	Not Suspicious	Mildly Suspicious	Moderately Suspicious	Highly Suspicious
Recommendation	No FNA	No FNA	FNA if ≥ 2.5 cm Follow up if ≥ 1.5 cm	FNA if ≥ 1.5 cm Follow up if ≥ 1cm	FNA if ≥ 1 cm Follow up if ≥ 0.5 cm

Appendix B. Sonographic patterns, estimated risk of malignancy, and fine-needle aspiration guidance for thyroid nodules based on American Thyroid Association management guidelines for adult patients with thyroid nodules[4]

Sonographic pattern	US features	Estimated risk of malignancy, %	FNA size cutoff
High suspicion	Solid hypoechoic nodule or solid hypoechoic component of a partially cystic nodule with one or more of the following features: irregular margins (infiltrative, microlobulated), microcalcifications, taller than wide shape, rim calcifications with small extrusive soft tissue component, evidence of extrathyroidal extension (ETE)	>70-90	Recommended FNA at ≥ 1cm
Intermediate suspicion	Hypoechoic solid nodule with smooth margins without microcalcifications, ETE, or taller than wide shape	10-20	Recommended FNA at ≥ 1 cm
Low suspicion	Isoechoic or hyperechoic solid nodule, or partially cystic nodule with eccentric solid areas, without microcalcification, irregular margin or ETE, or taller than wide shape	5-10	Recommended FNA at ≥ 1.5 cm
Very low suspicion	Spongiform or partially cystic nodules without any of the sonographic features described in low, intermediate, or high suspicion patterns	<3	Consider FNA at ≥ 2 cm
Benign	Purely cystic nodules (no solid component)	<1	No biopsy