International Journal Dental and Medical Sciences Research
Volume 3, Issue 1, Jan-Feb 2021 pp 494-503 www.ijdmsrjournal.com ISSN: 2582-6018

Inter-Observer Variability of Sonographic Features Used In Thyroid Imaging Reporting and Data System

Major (Dr) Jitendra Kumar Tiwari, Dr (Maj) AparMathur, Col (Dr) R A George
Command Hospital Airforce, Bangalore, Karnatak
Command Hospital Airforce, Bangalore, Karnataka
, Command Hospital Airforce, Bangalore, Karnataka

Submitted: 10-01-2021 Revised: 18-01-2021 Accepted: 21-01-2021

ABSTRACTThis paper is to:

1. To assess inter-observer variability in assigning Thyroid Imaging Reporting and Data System (TI-RADS) score while performing USG of clinically suspected thyroid nodules.
2. To correlate the TIRADS score with FNAC results.

Methods: This was a prospective study conducted between November 2018 to March 2020 wherein ultrasound was performed by doctors specialized in field of radiology, on 100 patients with thyroid nodules and are analyzed in terms of five criteria namely-composition, echogenicity, shape, margin, and echogenic foci given by TIRADS system . The results will be compared with FNAC of thyroid nodule (as gold standard investigation for the thyroid nodule characterization).PPV will be calculated for all readers' combined assessment. Interobserver agreement will be calculated using appropriate statistical test as applicable.

Results: Interobserver agreement in interpretation was near-perfect in respect of shape (k=0.923) and echogenicity (k=0.912). Only substantial agreement observed for composition (k=0.698), moderate agreement for echogenic foci (k=0.479) whereas only fair agreement (k=0.270) was seen in respect of margins. Applying ACR TIRADS grading there is near perfect association between TIRADS score and FNA results.

Conclusion: Variability in analyzing thyroid nodule sonographic features was highest for margin and echogenic foci. Despite of this variability in assessing features of thyroid nodule, by meticulous application of ACR TIRADS, the probability of malignancy of thyroid nodule can be predicted, and diagnostic yield of targeted FNACs enhanced.

Keywords: Thyroid, Inter-observer, Sonography

I. INTRODUCTION

Thyroid nodules are very commonly found, which may occur in 50% or more of adults, though clinically thyroid cancer is rare affecting less than 1 in 10,000 persons.[1, 2] Thus there is high prevalence of thyroid nodules but with only few of them being cancerous.[3]

Since it is always not possible to determine if a thyroid nodule is cancerous by its general physical examination and blood tests, the evaluation of thyroid nodules thus requires specialized tests such as thyroid ultrasonography and fine needle aspiration cytology. The increased use and improved quality of ultrasonographic technique has lead to the detection of various morphologic characteristics in thyroid nodule. However, there is always an uncertainty about which nodules is cancerous, and there is a lack of evidence-based guidelines that has resulted in conflicting recommendations regarding which nodules requires biopsy. [4]

Although many studies have analyzed the association between the ultrasound imaging characteristics of thyroid nodules and the risk of thyroid cancer, most studies comprised of small group of patients and there was lack of overall agreement between the different observers. So a guideline to be followed universally so that these variations will not occur between different observers worldwide by adoption of uniform standards for the interpretation of thyroid sonograms was the first step in standardizing the diagnosis and treatment of the thyroid cancer and limiting unnecessary diagnostic evaluation and treatment.

Evolution of thyroid grading system with international consensus i.e. Thyroid Imaging Reporting and Data System (TIRADS) on ultrasound was proposed by the American College of Radiology (ACR) . ACR TI-RADS provides a standard scoring system for observers regarding recommendations for using fine needle aspiration (FNA) or ultrasound follow-up of any suspicious nodules, and when to safely leave the nodules that are benign or not suspicious.[4, 5, 6, 7] However substantial inter-observer variability has been documented in grading of thyroid nodule using TIRADS. Applicability of such a classification system in busy tertiary care hospitals need to be tested with varying experience of operators.

AIMS OF STUDY

1. To assess inter-observer variability in assigning Thyroid Imaging Reporting and Data System (TI-RADS) score while performing USG of clinically suspected thyroid nodules.
2. To correlate the TIRADS score with FNA results.

OBJECTIVES OF STUDY

1. To familiarize the newly inducted trainees with the universal TIRADS scoring system for evaluation of thyroid nodule.
2. To check the feasibility of applying TIRADS in a tertiary care hospital with varied experience of the radiologists.

II. MATERIALS AND METHOD

The study cohort comprised of 100 patients who underwent fine-needle aspiration biopsy with definitive cytologic results or surgical resection between November 2018 to March 2020 at a single institution. This is a prospective study wherein ultrasound was performed by two radiologists on patients with thyroid nodules. PPV was calculated for all readers' combined assessment. Inter‑ observer agreement will be calculated using linear weighted kappa. The overall TIRADS grading given was then correlated with the FNA results. FNA done at same institution using 24g needle. The study was approved by the institutional review board of the institution. Informed consent was taken by all the patients who were undergoing ultrasound examination of thyroid nodule. The ultrasound examinations were performed with 5–15-MHz linear-array transducers on GE Logiq F8 USG machine. In all cases, images of the biopsied nodules were obtained in transverse and longitudinal planes.

Image Interpretation and Feature Assignments

The study was conducted from November 2018 to March 2020. Two radiologists one with 2 years of experience and other with 20 years of experience evaluated the nodules via the ACR portal. The readers were blinded to the biopsy outcomes. Assessment of the nodules was based on the five feature categories in ACR TI-RADS (composition, echogenicity, shape, margin, and echogenic foci), in which ultrasound findings are awarded 0–3 points that correspond to their association with malignancy. The size of the nodule was based on the maximum diameter measured on the static images, irrespective of the acquisition plane. The test readers were required to give the TIRADS stratification (TR1-TR5) and the results are then correlated with FNA results.

Statistical Analysis

The primary outcomes were the radiologists' interpretations of the thyroid nodule features and the recommendation for biopsy. Nodule findings in the same category with the same point values were grouped. Each of five possibilities for echogenic foci (none, large comet tail, macrocalcifications, peripheral calcifications, and punctate echogenic foci) were considered separately, given that a nodule could have more than one type of echogenic foci. Simple percentage of agreement for the same category was measured. The variability in interpretation for each of the features and for recommendation of biopsy was assessed with the Fleiss kappa statistic for categoric data. This is a measure of the difference between observed agreement and expected agreement. The kappa scale was as follows: < 0, less than chance agreement; 0.1–0.2, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; 0.81–0.99, almost perfect agreement. [13] In all cases, the threshold for assessing statistical significance was set at an alpha level of 0.05.

Statistical analyses were performed with R software (version 2016, R Core Team) and SAS statistical software (version 2015, SAS Institute).

Sample size estimation

$$\text{Sample Size (N)} = \frac{Z\,\alpha 2\ PQ}{d2}$$

Zα2 = Standard normal variate 1.96 @ 95% confidence limit

P = Prevalence (0.5)

Q = 1-P

d = Absolute error (0.1) @ Relative error 20%

$$N = \frac{(1.96)2\ (0.5)\ (0.5)}{(0.1)2}$$

N = 96

The sample size was calculated as per the study done By Hoang JK et al. [56]

III. RESULTS

Study Population

The mean (SD) of age (years) was 47.44 (14.04). The median (IQR) of age (years) was 48.50 (18.00). The age (years) ranged from 19 - 82. Of all the participants 25.0% of the participants are male and 75.0% are female. Distribution of the age and gender is given in table 1.

Table 1: Summary of Age/Gender

Age/Gender	Mean ± SD || Median (IQR) || Min-Max || Frequency (%)
Age (Years)	47.44 ± 14.04 || 48.50 (38.00-56.00) || 19.00 - 82.00
Gender	
Male	25%
Female	75%

There were 18 (18%) malignancies of which 11 (61%) were papillary carcinoma, 6(34%) were the follicular variant of papillary carcinoma and 1(5%) is adenomatoid carcinoma. There were 82 (82%) benign nodules of which 34 (41%) were colloid goiter, 21 (26%) were colloid cystic lesion, 12(15%) were follicular nodule, 7 (8%) were benign cystic lesion, 6 (7%) were thyroiditis (lymphocytic and Hoshimoto thyroiditis), 1(1%) follicular lesion of unknown significance, 1(1%) multinodular goiter.

Agreement for Ultrasound Features and Correlation of TIRADS grade with FNA results

Table 2 and 3 shows the frequency of sonographic features among the two readers for benign and malignant nodules.

Table 2: Summary of Interpretation (Rater 1)

Interpretation (Rater 1)	Frequency (%)
Composition	
Cystic	18
Spongiform	7
Mixed	43
Solid	32
Echogenicity	
Anechoic	19
Hyperechoic	61
Hypoechoic	20
Very hypoechoic	0
Shape	
Wider than tall	89
Taller than wide	11
Margins	
Smooth	59
Ill defined	20
Irregular/Lobulated	21
Extrathyroid extension	0
Echogenic Foci	
None	94
Macrocalcification	2
Peripheral	2
Punctate	2
TIRADS	
Grade: 1	15
Grade: 2	44
Grade: 3	22
Grade: 4	7
Grade: 5	12

Table 3: Summary of Interpretation (Rater 2)

Interpretation (Rater 2)	Frequency (%)
Composition	
Cystic	18
Spongiform	12
Mixed	37
Solid	33
Echogenicity	
Anechoic	22
Hyperechoic	57
Hypoechoic	21
Very hypoechoic	0
Shape	
Wider than tall	88
Taller than wide	12
Margins	
Smooth	57
Ill defined	20
Irregular/Lobulated	23
Extrathyroid extension	0
Echogenic Foci	
None	94
Macrocalcification	0
Peripheral	1
Punctate	5
TIRADS	
Grade: 1	13
Grade: 2	45
Grade: 3	24
Grade: 4	10
Grade: 5	8

Table 4 to 9 shows measures of interobserver agreement among the two readers in the study for the features and final recommendation of biopsy. The various agreements between two raters for various parameters are as detailed:

Composition (depicted in table 4 and figure 8): The two raters agreed in 79.0% of the cases and disagreed in 21.0% of the cases. There was substantial agreement between the two methods, and this agreement was statistically significant (Cohen's Kappa = 0.698, p = <0.001). The disagreements observed between the two raters were as follows: 2.0% of cases classified as mixed by rater 1 were classified by rater 2 as solid. 5.0% cases classified as mixed by rater 1 were classified as Spongiform by Rater 2. 4.0% cases classified as solid by rater 1 were classified by rater 2 as mixed. 9.0% cases classified as spongiform by Rater 1 were classified as mixed by Rater 2. 1.0%of cases classified as spongiform by Rater 1 were classified as Solid by Rater 2.

Table 4: Comparison of Composition (Rater 1) with Composition (Rater 2) (n = 100)

Composition		**Composition (Rater 2)**					Cohen’s Kappa	
		Cystic	**Spongiform**	**Mixed**	**Solid**	**Total**	**k**	**P Value**
Composition (Rater 1)	**Cystic**	18	0	0	0	18	0.698	<0.001
	Spongiform	0	2	5	0	7		
	Mixed	0	9	30	4	43		
	Solid	0	1	2	29	32		
	Total	18	12	37	33	100		

Echogenicity (depicted in table 5): The two raters agreed in 95.0% of the cases and disagreed in 5.0% of the cases.There was near perfect agreement between the two raters, and this agreement was statistically significant (Cohen's Kappa = 0.912, p = <0.001). The disagreements observed between the two raters were as follows: 9.0% cases classified as hyperechoic by Rater 1 were classified as anechoic by Rater 2. 1.0% cases classified as anechoic by Rater 1 were classified as hypoechoic by Rater 2.

Table 5: Comparison of Echogenicity (Rater 1) with Echogenicity (Rater 2) (n = 100)

Echogenicity		Echogenicity (Rater 2)					Cohen's Kappa	
		Anechoic	Hyperechoic	Hypoechoic	Very hypo echoic	Total	K	P Value
Echogenicity (Rater 1)	Anechoic	18	0	1	0	19	0.912	<0.001
	Hyperechoic	4	57	0	0	61		
	Hypoechoic	0	0	20	0	20		
	Very hypoechoic	0	0	0	0	0		
	Total	22	57	21	0	100		

Shape (depicted in table 6): The two raters agreed in 99.0% of the cases and disagreed in 1.0% of the cases. Perfect agreement found between the two raters and this agreement was statistically significant (Cohen's Kappa=0.8, p=<0.001).

Table 6: Comparison of Shape (Rater 1) with Shape (Rater 2) (n = 100)

Shape		Shape (Rater 2)			Cohen's Kappa	
		Wider Than Tall	Taller Than Wide	Total	k	P Value
Shape (Rater 1)	Wider Than Tall	87	01	88	0.8	<0.001
	Taller Than Wide	00	11	12		
	Total	88	12	100		

Margins (depicted in table 7): The two raters agreed in 58.0% of the cases and disagreed in 42.0% of the cases.There was fair agreement between the two raters, and this agreement was statistically significant (Cohen's Kappa = 0.270, p = <0.001). The disagreements observed between the two raters were as follows: 1 (1.0%) cases classified as irregular/lobulated by rater 1 were classified as smooth by Rater 2. 10% cases classified as ill-defined by rater 1 were classified as smooth Rater 2. 5% cases classified as smooth by rater 1 were assigned as irregular/lobulated by rater 2. 8% cases classified as ill-defined by rater 1 were assigned as irregular/lobulated by rater 2. 8% of cases classified as smooth by rater 1 were classified as ill-defined by rater 2. 10% of cases classified as irregular/ lobulated by rater 1 were classified as ill-defined by rater 2.

Table 7: Comparison of Margins (Rater 1) with Margins (Rater 2) (n = 100)

Margins		Margins (Rater 2)				Cohen's Kappa	
		Smooth	Ill defined	Irregular/Lobulated	Total	k	P Value
Margins (Rater 1)	Smooth	46	8	5	59	0.270	<0.001
	Ill Defined	10	2	8	20		
	Irregular/ Lobulated	1	10	10	21		
	Extra-thyroidal extension	0	0	0	0		
	Total	57	20	23	100		

Echogenic foci (depicted in table 8): The two raters agreed in 94.0% of the cases and disagreed in 6.0% of the cases.There was moderate agreement between the two raters, and this agreement was statistically significant (Cohen's Kappa = 0.479, p = <0.001).The disagreements observed between the two raters were as follows:1.0% cases classified as peripheral calcification by Rater 1 was not having any echogenic foci as per rater 2. 1.0% cases classified as macrocalcification by Rater 1 were classified as Punctate by Foci Rater 2. 1.0% cases classified as peripheral echogenic foci by rater 1 were not having any echogenic foci by rater 1. 1% cases assigned to have punctate echogenic foci by rater 1 was not found to have any echogenic foci by rater 2. 1.0% cases classified as having macrocalcification by rater 1 were not found to have any echogenic foci (as this was not part of lesion as per rater 2.

Table 8: Comparison of Echogenic Foci (Rater 1) with Echogenic Foci (Rater 2) (n = 100)

Echogenic Foci		Echogenic Foci (Rater 2)					Cohen's Kappa	
		None	Macro-calcification	Peripheral	Punctate	Total	K	P Value
Echogenic Foci (Rater 1)	None	92	0	1	1	94	0.479	<0.001
	Macro-calcification	1	0	0	1	2		
	Peripheral	1	0	0	1	2		
	Punctate	0	0	0	2	2		
	Total	94	0	1	5	100		

TIRADS Grading (depicted in table 9): The two raters agreed in 76.0% of the cases and disagreed in 24.0% of the cases.There was Near Perfect agreement between the two raters, and this agreement was statistically significant (Weighted Kappa = 0.907, p = <0.001).The disagreements observed between the two raters were as follows: 5.0% cases classified as Grade: 2 by rater 1 were classified as Grade: 1 by rater 2. 7.0% cases classified as Grade: 1 by Rater 1 were classified as Grade: 2 by rater 2. 4.0% cases classified as Grade: 2 by rater 1 were classified as Grade: 3 by rater 2. 1.0% cases classified as Grade: 4 by Rater 1 were classified as Grade: 3 by rater 2. 4.0% cases classified as Grade: 5 by rater 1 were classified as Grade: 4 by rater 2.

Table 9: Comparison of TIRADS (Rater 1) with TIRADS (Rater 2) (n = 100)

TIRADS		TIRADS (Rater 2)						Weighted Kappa	
		Grade: 1	Grade: 2	Grade: 3	Grade: 4	Grade: 5	Total	k	P Value
TIRADS (Rater 1)	Grade: 1	8	7	0	0	0	15	0.907	<0.001
	Grade: 2	5	35	4	0	0	44		
	Grade: 3	0	3	19	0	0	22		
	Grade: 4	0	0	1	6	0	7		
	Grade: 5	0	0	0	4	8	12		
	Total	13	45	24	10	8	100		

Table 10 and 11 shows the correlation between TIRADS grading and FNAC results. These are as detailed:For rater 1 (depicted in table 10): 18.3% of the participants in the group [FNAC: Benign] had [TIRADS: Grade: 1]. 53.7% of the participants in the group [FNAC: Benign] had [TIRADS: Grade: 2]. 26.8% of the participants in the group [FNAC: Benign] had [TIRADS: Grade: 3]. 1.2% of the participants in the group [FNAC: Benign] had [TIRADS: Grade: 4]. 100.0% of the participants in the group [FNAC: Malignant] had [TIRADS: Grade: 4].

Table 10: Association Between FNAC and TIRADS (Rater 1) (n = 100)

TIRADS **(Rater 1)**	**FNAC**			**Fisher's Exact Test**	
	Benign	**Malignant**	**Total**	**χ2**	**P Value**
Grade: 1	15	0	15	94.193	<0.001
Grade: 2	44	0	44		
Grade: 3	22	0	22		
Grade: 4	1	6	7		
Grade: 5	0	12	12		
Total	82	18	100		

For rater 2 (depicted in table 11): 15.9% of the participants in the group [FNAC: Benign] had [TIRADS: Grade: 1]. 54.9% of the participants in the group [FNAC: Benign] had [TIRADS: Grade: 2]. 28% of the participants in the group [FNAC: Benign] had [TIRADS: Grade: 3]. 1.2% of the participants in the group [FNAC: Benign] had [TIRADS: Grade: 4]. 5.6% of the participants in the group [FNAC: Malignant] had [TIRADS: Grade: 3]. 50% of the participants in the group [FNAC: Malignant] had [TIRADS: Grade: 4]. 44.4% of the participants in the group [FNAC: Malignant] had [TIRADS: Grade: 5].

Table 11: Association Between FNAC and TIRADS (Rater 2) (n = 100)

TIRADS **(Rater 2)**	**FNAC**			**Fisher's Exact Test**	
	Benign	**Malignant**	**Total**	**χ2**	**P Value**
Grade: 1	13	0	13	87.410	<0.001
Grade: 2	45	0	45		
Grade: 3	23	1	24		
Grade: 4	1	9	10		
Grade: 5	0	8	8		
Total	82	18	100		

IV. DISCUSSION

Thyroid nodules are very commonly found during ultrasonography of neck in 50% or more of adults, though clinically thyroid cancer is rare, affecting less than 1 in 10,000 persons. Thus there is high prevalence of thyroid nodules in general population but with only few of them being cancerous. Since USG being the primary investigation for evaluation of thyroid, there was wide variability in reporting ultrasound features which lead to inconsistent management. ACR TI-RADS is a universally accepted reporting system for thyroid nodules developed with the aim to standardize reporting of thyroid nodule and recommend further management of thyroid nodules based on their size and ultrasound features. This study is targeted for dual purposes, firstly to analyze degree of inter-observer variability while characterizing the lesions in thyroid, and secondly to study correlation between TIRADS grading and FNAC results.

The five principle criteria governing TIRADS grading are composition, echogenicity, margins, echogenic foci and shape. Inter-observer variability in TIRADS scoring has been reported in scientific literature since year 2002. Most of the workers have attempted to corroborate observer variability as regards individual parameters noted in the evaluation of thyroid nodule.

Composition had substantial agreement in study conducted by Hoang JK et al (k=0.61), Park SJ et al (k=0.6) and Moon et al (k=0.62).[8,9,10] Good agreement (k=0.57) was documented in study conducted by Anuradha et al [11] and fair agreement (k=0.3) in study conducted by Middleton et al. [12] In our study also composition had substantial agreement (k=0.696) between the two participating observers, though widely varying in experience.

Echogenicity had substantial agreement in study conducted by Anuradha et al (K=0.61) and Park SJ et al (k=0.67), [9,11] and moderate agreement (k=0.54) in study conducted by Middleton et al and Hoang JK et al (k=0.47). [8,12] In our study echogenicity had near-perfect agreement (k=0.72).

Margins had substantial agreement (k=0.61) in studies conducted by Anuradha et al and Moon et al, [58,60] but only fair agreement (k=0.37) was reported in studies conducted by Middleton et al, Hoang JK et al (0.27), Weinke et al (0.31) and Park CS et al (0.28). [8,12,13,14] In our study margins had only fair agreement (k=0.27) between the observers.

Echogenic foci had fair agreement in studies conducted by Middleton et al (k=0.4), Anuradha et al (k=0.4) and Hoang JK et al (k=0.47) [8,11,12] and moderate agreement (k=0.51) as documented by Park SJ et al (k=0.56) and Moon et al [9,10] was also the observation corroborated in our study (k=0.479).

Calcification had substantial agreement in studies conducted by Middleton et al (0.62), Anuradha et al (k=0.63), Hoang JK et al (0.62), Park SJ et al (k=0.65) and Moon et al (0.60). [8,9,10,11,12] In our study also calcification had substantial agreement (k=0.67).

Shape had substantial agreement in studies conducted by Middleton et al (k=0.63), Anuradha et al (k=0.67), Hoang JK et al (k=0.61), Park SJ et al (k=0.64), Weinke et al (k=0.6), Moon et al (k=0.64), and Park CS et al (k=0.61). [8,9,10,11,12,13,14] In our study, shape had near perfect agreement (k=0.89).

Overall, our study shows that the thyroid nodule ultrasound features with substantial agreement is composition, moderate agreement for echogenic foci and fair agreement for margins. Near-perfect agreement was noted in respect of shape and echogenicity.

Cheng et al study concluded that there was moderate to substantial inter-observer agreement for final assessment category (kappa value- 0.61) and concluded that TI-RADS is a helpful but not optimal reporting tool in characterizing thyroid lesions. [15] The correlation coefficient of TI-RADSs between category and malignancy rate was 0.712. Our study also had moderate inter-observer agreement for final assessment category (k=0.696). There is near-perfect correlation between TIRADS category and malignancy rate in our study.

Grani et al study showed that agreement for the indication to biopsy was substantial to near-perfect, being 0.73 (Cohen's kappa).[16] They concluded that despite the wide variability in the description of single ultrasonographic features, the classification systems may improve the inter-observer agreement that can further be enhanced after a specific training. Our study also had shown near-perfect agreement in assigning TIRADS grading and correlation with FNACs result.

Middleton et al study concluded that aggregate risk of malignancy for nodules assigned by TIRADS point level was within the TIRADS risk stratification thresholds. Overall, 86.1% of all nodules were within 1% of the TIRADS specified risk thresholds. [12] In our study also risk of malignancy for nodules assigned by TIRADS point level was within the TIRADS risk stratification thresholds with 94% of all nodules being within 1% of TIRADS specified risk thresholds.

William et al assigned points for each features and the risk of cancer associated with each point and final TIRADS level was determined. [8] The ACR TI-RADS biopsy yield was significantly higher than that of ATA guidelines (p<0.0001). The limitation of this study was inter-observer variability was not taken into account. In our study also ACR TI-RADS biopsy yield was significantly higher whereas our study had also included inter-observer variability for TIRADS grading and risk assessment.

Overall moderate inter-observer agreement for final assessment category and near-perfect correlation between TIRADS category and malignancy rate was observed in our study.

There are several potential sources of the greater variability regarding margin and echogenic foci in our study. The basic reason for the inter-observer variability is because of wide difference in experience of the two observers i.e. with junior observer having only 2 years of experience and senior observer having 20 years of experience in TIRADS application. Irregular or lobulated margin is the most likely criterion to be under-reported if not previously seen. Another possible explanation for variability is that a suspicious margin or echogenic foci may have been apparent in only a few parts of the thyroid nodule that can be easily missed by the observer with lesser experience. In contrast, composition and echogenicity, which had lower variability, are findings present throughout the nodule. Shape had near-perfect agreement, as is expected because it is based on objective measurements of antero-posterior and transverse diameters. The agreement on biopsy will further improve with targeted education about sonographic findings.

Our study had several limitations. TIRADS scoring was arrived at independently by both observers by mathematically aggregating scores for morphological criteria which were not simultaneously correlated/represented. Second, there was wide variation in experience with only two radiologists characterizing the thyroid nodule and directly reflects impact of expertise in evaluating thyroid ultrasound. This could have been improved with more numbers of observers having wider experience spectrum and increasing the number of nodules with low prevalence features. Finally, the kappa coefficient was not stratified on the basis of size of the nodules.

V. CONCLUSION

We evaluated inter-observer variability in interpreting thyroid nodules on ultrasound images among two radiologists of varied experience and found that agreement was substantial for margin, composition and shape of lesion. Inter-observer agreement will improve further with targeted training about sonographic findings, particularly in respect of lesional margin and echogenic foci. There is near-perfect correlation between the TIRADS grading and FNAC results in thyroid nodules.

By meticulous application of ACR TIRADS, the probability of malignancy of thyroid nodule can be predicted, and diagnostic yield of targeted FNACs enhanced, while avoiding unnecessary and unregulated interventions in patients having thyroid nodules with lower TIRADS score. This will help in optimal utilization of health-care resources and reduce frequency of hospital visits in such patients.

REFERENCES:

[1]. Rojeski MT, Gharib H. Nodular thyroid disease: evaluation and management. N Engl J Med 1985; 313:428-436.

[2]. Van Herle AJ, Rich P, Ljung BM. The thyroid nodule. Ann Intern Med 1992; 96:221-232.

[3]. Grebe SK, Hay ID. Follicular cell–derived thyroid carcinomas.Cancer Treat Res 1997; 89:91-140.

[4]. Guth S, Theune U, Aberle J, Galach A, and Bamberger, C.M. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. Eur J Clin Invest. 2009; 39: 699–706

[5]. Singh Ospina N, Brito JP, Maraka S. Diagnostic accuracy of ultrasound-guided fine needle aspiration biopsy for thyroid malignancy: systematic review and meta-analysis. Endocrine. 2016; 53: 651–661

[6]. Smith-Bindman R, Lebda P, Feldstein VA. Risk of thyroid cancer based on thyroid ultrasound imaging characteristics: results of a population-based study. JAMA Intern Med. 2013; 173: 1788–1796

[7]. Ito Y, Uruno T, Nakano K, Takamura Y, Miya A, Kobayashi K, Yokozawa T. An observation trial without surgical treatment in patients with papillary microcarcinoma of the thyroid.Thyroid. 2003; 13: 381–387

[8]. William, Franklin N, Edward G. Interobserver variability in assigning features in the American College of Radiology Thyroid Imaging Reporting and Data System (ACR TI-RADS) lexicon and in making recommendations for thyroid nodule biopsy. AJR 2018; 211: 1-6.

[9]. Park SJ, Kim SH, Jung SL. Observer variability and the performance between faculties and residents: US criteria for benign and malignant thyroid nodules. Korean J Radiol 2010; 11:149–155

[10]. Moon WJ, Jung SL, Lee JH, So LJ, Dong GN, JH Baek, YH Lee, J Kim. Benign and malignant thyroid nodules: US differentiation—multicenter retrospective study. Radiology 2008; 247:762–770

[11]. AnuradhaChandramohan, AbhishekKhurana, Pushpa BT, Marie Therese Manipadam, DukhabandhuNaik, Nihal Thomas, Deepak Abraham, Mazhuvanchary Jacob Paul. Is TIRADS a practical and accurate system for use in daily clinical?. IJRI 2016; 26:145-152.

[12]. Middleton WD, Teefey SA, Reading CC. Multiinstitutional analysis of thyroid nodule risk stratification using the American College of Radiology Thyroid Imaging Reporting and Data System. AJR 2017; 208: 1331–1341.

[13]. Wienke JR, Chong WK, Fielding JR, Zou KH, Mittelstaedt CA. Sonographic features of benign thyroid nodules: interobserver reliability and overlap with malignancy. J Ultrasound Med 2003; 22:1027–1031

[14]. Park CS, Kim SH, Jung SL, BJ Kang, J Kim, JJ Choi, SH Jeong. Observer variability in the sonographic evaluation of thyroid nodules. J Clin Ultrasound 2010; 38:287–293

[15]. Choi YJ, Baek JH, Hong MJ, Lee JH. Inter-observer variation in ultrasound measurement of the volume and diameter of thyroid nodules. Korean J Radiol 2015; 16:560–565

[16]. Giorgio Grani, LiviaLamartina, Vito Cantisani, Marianna Maranghi, Piernatale Lucia and Cosimo Durante.Interobserver agreement of various thyroid imaging reporting and data systems. Endocrine connections 2018; 7: 1-7