European Journal of Radiology 117 (2019) 184–192

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Research article

The validity and reproducibility of the thyroid imaging reporting and data system (TI-RADS) in categorization of thyroid nodules: Multicentre prospective study

Mohammad Abd Alkhalik Basha[a,*], Ahmad Abdullah Alnaggar[a], Rania Refaat[b], Ahmed Mohamed El-Maghraby[a], Mona Mohammed Refaat[a], Marwa E. Abd Elhamed[a], Ahmed A. El-Hamid M. Abdalla[a], Sameh Abdelaziz Aly[c], Amr Shaaban Hanafy[d], Abd El Motaleb Mohamed[e], Amira Hamed Mohamed Afifi[f], Ola Harb[g]

[a] Department of Radio-diagnosis, Zagazig University, Zagazig, Egypt
[b] Department of Radio-diagnosis, Ain Shams University, Cairo, Egypt
[c] Department of Radio-diagnosis, Benha University, Benha, Egypt
[d] Department of Internal medicine, Zagazig University, Zagazig, Egypt
[e] Department of Clinical Oncology, Zagazig University, Zagazig, Egypt
[f] Department of Clinical Pathology, Zagazig University, Zagazig, Egypt
[g] Department of Pathology, Zagazig University, Zagazig, Egypt

ARTICLE INFO

Keywords:
Thyroid nodules
Thyroid imaging reporting and data system
Ultrasonography

ABSTRACT

Purpose: To assess diagnostic validity and reproducibility of Thyroid Imaging Reporting and Data System (TI-RADS) for interpretation of thyroid nodules by thyroid ultrasonography (US).

Method: A prospective multicentre study initially included 557 patients with clinically suspected thyroid nodules. After exclusion, a final cohort of 380 patients with 948 thyroid nodules detected by US were enrolled. Based on American College of Radiology (ACR) TI-RADS, three radiologists analysed all US examinations independently and assigned a TI-RADS category to each thyroid nodule. The final diagnosis was based on cytology which was used as reference standard for calculating diagnostic performance of TI-RADS for predicting malignant thyroid nodules. The Fleiss and weighted kappa (κ) statistics were applied to assess inter-observer agreement of morphological features and TI-RADS scoring results for thyroid nodules. Additionally, we made a simple screening among referring clinicians to assess the clinical response to application of TI-RADS.

Results: A total of 948 thyroid nodules were evaluated; 136 (14.3%) were malignant, and 812 (85.7%) were benign. The papillary carcinoma was the most common malignant thyroid nodules (81.6%). The best cut-off value for predicting malignant thyroid nodules was > TR3. On a lesion-based analysis, the TI-RADS had a sensitivity, specificity, and an accuracy of 98.3%, 90.9%, and 92.1%, respectively when regarding those thyroid nodules classified as > TR3 for predicting malignancy. The inter-observer agreement of the TI-RADS category was good (κ = 0.636). Ninety percent of referring clinicians accept TI-RADS.

Conclusions: TI-RADS improves diagnostic performance of US for predicting malignant thyroid nodules with high validity and high reproducibility.

1. Introduction

Thyroid nodules are a common clinical problem that increases in the general population [1]. The prevalence of thyroid nodules increases with age, and the female population is more affected [2,3]. A small percentage of thyroid nodules proves to be malignant, ranging between 1.6% and 12% [4]. Thyroid ultrasonography (US) examination is the best diagnostic tool available for the diagnosis of thyroid nodules, but

* Corresponding author.
E-mail addresses: Mohammad_basha76@yahoo.com (M.A.A. Basha), naggar2002us@yahoo.com (A.A. Alnaggar), raniarefaat_1977@hotmail.com (R. Refaat), dr_elmaghraby80@yahoo.com (A.M. El-Maghraby), refaatmona96@gmail.com (M.M. Refaat), dr.meroelsaed34@gmail.com (M.E. Abd Elhamed), ahmedzagazig@yahoo.com (A.A.E.-H.M. Abdalla), drsamehaly75@gmail.com (S.A. Aly), dr_amr_hanafy@yahoo.com (A.S. Hanafy), Abdelmotalabmohamed@yahoo.com (A.E.M. Mohamed), Amira_hamed79@hotmail.com (A.H.M. Afifi), olaharb2015@gmail.com (O. Harb).

https://doi.org/10.1016/j.ejrad.2019.06.015
Received 7 February 2019; Received in revised form 4 June 2019; Accepted 16 June 2019

the accuracy of US is low in differentiating benign and malignant thyroid nodules [5]. US is subjective and operator-dependent and has low reproducibility for its potential inter-observer and intra-observer agreement [6].

The need for a clinically valuable management guideline for thyroid nodules has led to the development of several classification systems. These systems use composite patterns of US findings to estimate the likelihood of malignancy and to identify thyroid nodules that need to be scheduled for fine needle aspiration biopsy (FNAB) [7,8]. Several of these systems have been endorsed by international scientific bodies [9–13], but their reproducibility is yet to be assessed. Horvath et al. [7] in 2009, developed the Thyroid Imaging Reporting and Data System (TI-RADS) similar to the Breast Imaging Reporting and Data System (BI-RADS) classification used for breast masses as an approach to enable structured reporting of thyroid nodules. This system was subjected to a modified recommendation by Kwak et al. [14]. In 2012, the American College of Radiology (ACR) organized committees to provide recommendations for reporting incidental thyroid nodules. The committees developed a set of standard terms (lexicon) for US reporting, proposed a TI-RADS based on the lexicon, and published the results of the first two efforts in 2015 [15,16]. The initial purpose of TI-RADS was to improve patient management and avoid unnecessary FNAB in patients with thyroid nodules. However, its clinical use is still limited, and its application in clinical practice is questioned [17].

Many studies have been performed to evaluate the validity of different TI-RADS classifications systems in the assessment of thyroid nodules [7,18–21]. However, these studies suffered from the limitations of being single-centre, retrospective, and lacked a rigorous reference standard. Accordingly, we conducted this multicentre prospective study to evaluate the diagnostic validity and reproducibility of ACR TI-RADS to assess malignancy risk and the need for biopsy of thyroid nodules. Additionally, we made a simple screening among referring clinicians involved in patient management to assess the clinical response to the application of ACR TI-RADS. We hypothesize that the application of ACR TI-RADS improves the diagnostic performance of US for predicting malignant thyroid nodules and reduce unnecessary FNAB.

2. Subjects and methods

2.1. Study population

This multicentre prospective study was conducted between May 2017 and December 2018. We initially collected 557 consecutive patients from three institutions.

Inclusion criteria: patients with clinically suspected thyroid nodules.

Exclusion criteria are listed in Fig. 1.

This yielded a final cohort of 380 patients (66 men and 314 women, with an age range of 18–71 years and a mean age of 45.3 ± 14.2 years). The patients' data are summarized in Table 1. All participants were subjected to a full history taking, complete general and local examinations, and US examination. The flow chart of our study is illustrated in Fig. 1.

2.2. Ethical considerations

The institutional review boards of the three participating institutions approved the study. All patients were notified about the study and written informed consent was obtained. The study was conducted following the ethical principles of the Declaration of Helsinki.

2.3. US examination and image analysis

All US examinations were performed using the same US machine (Philips Affiniti 70, Philips Ultrasound System, USA). The gray-scale US was done with the real-time sector scanner using a high-frequency probe (5/14 MHz). The US machine settings such as gain, time gain compensation, dynamic range, focus, depth, color gain, wall filter, were optimized until high quality US images were obtained. The patients lay supine, with their neck extended. The thyroid gland and adjacent tissues were scanned both transversely and longitudinally. Color Doppler US images were done for each target nodule.

Three highly experienced radiologists (with over 15 years of US experience and has performed > 1000 US examinations per year) independently performed all US examinations, blinded to patients' clinical data. Before the study started, several clinical sessions of lecture-based and hands-on instruction that explained ACR TI-RADS in detail were provided to the radiologists. The following morphological features were individually evaluated during US examination for each thyroid nodule:

1 Position (left lobe, right lobe, or isthmus).
2 The maximum diameter.
3 Composition.
4 Echogenicity.
5 Shape.
6 Margin.
7 Echogenic foci
8 Doppler evaluation of thyroid nodules was performed using the color Doppler gate to identify vascular color signals. Peripheral blood flow was defined if color signals in the wall or periphery of the mass; while central blood flow was defined if color signals in septa, papillary projections, solid areas, or the central part of the mass.
9 Suspicious neck lymph nodes

Each radiologist assigned points to each thyroid nodule for the separate five categories of composition, echogenicity, shape, margin, and echogenic foci, according to the TI-RADS protocols produced by ACR [10]. The sum of the points in each category indicated the TI-RADS category specified to each nodule, with 2 points denoting TR2 (not suspicious); 3 points, TR3 (mildly suspicious); 4–6 points, TR4 (moderately suspicious); and 7 or higher points, TR5 (highly suspicious). As per the instructions of the ACR TI-RADS committee, if margins, echogenicity, or composition could not be determined for any cause, they were assigned 0, 1, or 2 points, respectively. Additionally, the radiologists analyzed vascularity of each thyroid nodule and the suspicious lymph nodes, although they are not part of ACR TI-RADS.

The data from the three institutions were collected centrally for the analysis. For estimation of the diagnostic performance of ACR TI-RADS for predicting malignant thyroid nodules, the category for each thyroid nodule assigned by three independent radiologists was merged into a final category. During this procedure, any disagreement among radiologists in the assignment of ACR TI-RADS category of the thyroid nodule was reviewed until consensus was reached.

Before the start of the study, the meaning and aim of ACR TI-RADS were clarified to referring clinicians in several clinical sessions. Although we were not involved in the clinical decision-making process, we performed a simple screening among 30 of our referring clinicians involved in patient management, aimed to determine whether this reporting system could be helpful for deciding management protocol and in avoiding unnecessary FNAB.

2.4. Reference standard

The definite diagnoses of the thyroid nodules were based on the result of US-guided FNAB. The US-guided FNAB was performed by the same expert radiologist who performed the US. All US-guided FNAB were obtained using a 19 or 21-gauge needle fixed to a 10-cc syringe. Upon aspiration, negative pressure was sustained until blood came out in the syringe's hub. The sample of the FNAB was stained with hematoxylin and eosin by an on-site pathologist who provided an instant evaluation of sample adequacy, and the rest of the cytologic material

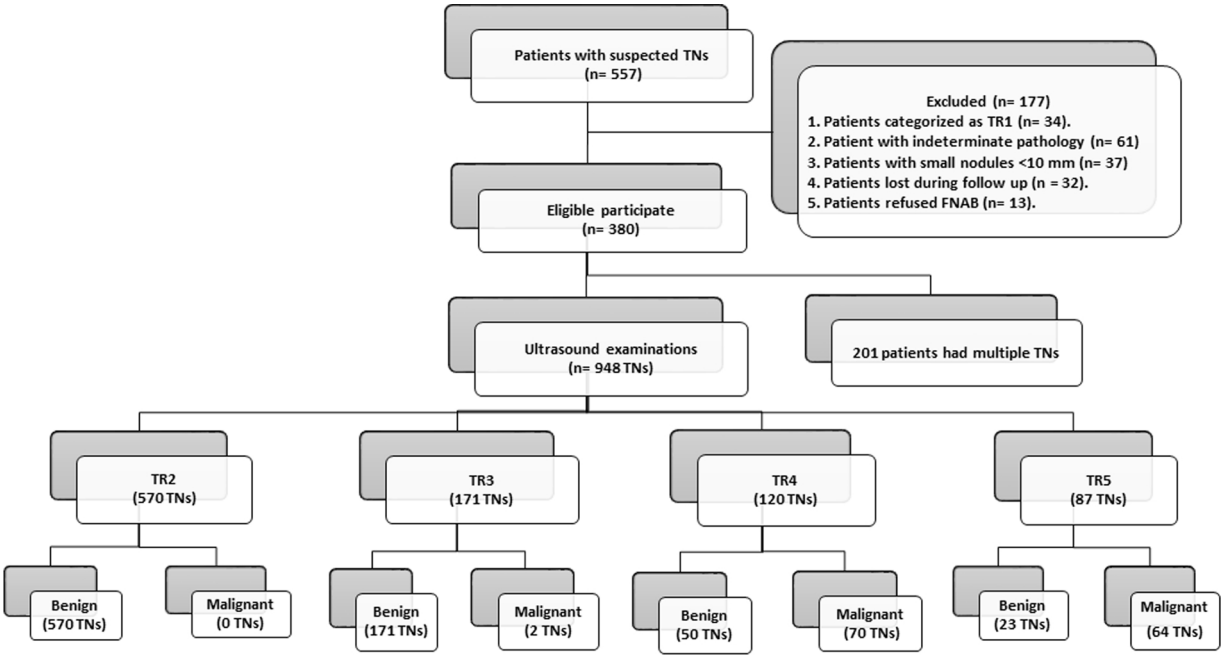

Fig. 1. Flow chart of the study population. TNs = Thyroid nodules.

Table 1
Patients' data.

Characteristic	Total	Benign nodules (on cytology)	Malignant nodules (on cytology)	P value
Patients, n (%)	380 (100)	325 (85.5)	55 (14.5)	
Age, y, mean ± SD (range)	45.3 ± 14.2 (18-71)	45.4 ± 12.9 (21-71)	44.7 ± 24.8 (18-67)	0.938
Patient sex, n (%)				0.869
Male	66 (17)	53 (16.3)	13 (23.6)	
Female	314 (83)	272 (83.7)	42 (76.4)	
Nodules, n (%)	948 (100)	812 (85.7)	136 (14.3)	
Maximum diameter of nodules, mm, mean ± SD, (range)	29.0 ± 13.7 (11-56)	28.9 ± 13.2 (11-56)	29.7 ± 20.6 (14-53)	0.930

Unless otherwise indicated, data are number with the percentage in parenthesis.
SD = Standard deviation.

was put in formalin. All samples were reviewed by two qualified pathologists, and the findings were reached by consensus. In patients with multiple nodules, up to two nodules were biopsied. The thyroid nodules were classified according to the criteria recommended by the Bethesda classification [22].

2.5. Statistical analysis

Statistical analysis of the collected data was conducted using MedCalc (version 11.1; MedCalc, Mariakerke, Belgium). Categorical variables were described as number and percentage. An independent sample *t*-test was used to compare continuous variables. Categorical variables were compared using the Mann–Whitney *U* test The Fleiss and weighted kappa (κ) were used to assess the inter-observer agreement of morphological features and ACR TI-RADS scoring results for predicting malignancy of thyroid nodules. The κ values were interpreted as follows: 0.01–0.20 = poor agreement; 0.21–0.40 = fair agreement; 0.41–0.60 = moderate agreement; 0.61–0.80 = good agreement; and 0.81–1.0 = very good agreement. The ACR TI-RADS was compared with the final cytological diagnosis, and the diagnostic performance of ACR TI-RADS was estimated. On a lesion-based analysis, we calculated the accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the ACR TI-RADS system for identifying malignant thyroid nodules. Analysis of receiver operating characteristic (ROC) curve was done to determine the cut-off value, the area under the curve (AUC), and the 95% CI. A p value ≤ 0.05 indicated statistically significant results.

3. Results

3.1. Patients

We enrolled 380 patients with thyroid nodules in our study. Every included patient had at least one thyroid nodule on US examination. We detected a total of 948 thyroid nodules (201 patients (52.9%) had multiple thyroid nodules). The patients' data are presented in Table 1. We did not find a significant difference between benign and malignant thyroid nodules as regards patient age ($p = 0.938$). The mean presenting age for malignant thyroid nodules was 44.7 ± 24.8 years. Thyroid nodules were more common in females than males ($p < 0.0001$).

Definitive final diagnoses are shown in Table 2. The final diagnoses in 948 thyroid nodules were 136 (14.3%) malignant and 812 (85.7%) benign. We found that the benign follicular nodule was the most common benign thyroid nodules (621/812, 76.5%), and the papillary carcinoma was the most common malignant thyroid nodules (111/136, 81.6%).

3.2. Assignment of ACR TI-RADS categories and malignancy risk

Of the 948 thyroid nodules assessed, 570 (60.1%) were classified as TR2, 171 (18%) as TR3, 120 (12.7%) as TR4, and 87 (9.2%) as TR5. Benign and malignant thyroid nodules according to ACR TI-RADS are shown in Table 2. Of the 570 thyroid nodules categorized as TR2, none were malignant; of the 171 thyroid nodules categorized as TR3, two was malignant; Of the 120 thyroid nodules categorized as TR4, 70 were malignant; and of the 87 thyroid nodules categorized as TR5, 64 were malignant. The overall malignancy risk in TR 2, TR3, TR4, and TR5

Table 2
The ACR TI-RADS categorizations of thyroid nodules according to final diagnosis (n = 948).

cytological diagnosis	TR2	TR3	TR4	TR5	Total
Benign nodules	**570 (70.2)**	**169 (20.8)**	**50 (6.2)**	**23 (2.8)**	**812 (85.7)**
Benign follicular nodule	554	33	27	7	621 (65.5)
Granulomatous (subacute) thyroiditis	11	49	18	9	87 (9.2)
Lymphocytic (Hashimoto) thyroiditis	5	87	5	7	104 (11)
Malignant nodules	**0 (0)**	**2 (1.5)**	**70 (51.5)**	**64 (47)**	**136 (14.3)**
Papillary carcinoma	0	2	53	56	111 (11.7)
Poorly differentiated carcinoma	0	0	8	2	10 (1)
Medullary carcinoma	0	0	9	2	11 (1.2)
Undifferentiated (anaplastic) carcinoma Metastatic carcinoma	0	0	0	2	2 (0.2)
	0	0	0	2	2 (0.2)
Total	**570 (60.1)**	**171 (18)**	**120 (12.7)**	**87 (9.2)**	**948 (100)**

The data are number with the percentage in parenthesis.
ACR = American College of Radiology; TI-RADS = Thyroid Imaging Reporting and Data System.
Benign follicular nodule includes adenomatoid nodule, colloid nodule, etc.

were 0, 1.2, 58.3, and 73.6%, respectively.

3.3. Inter-observer agreement for morphological diagnostic features and ACR TI-RADS categorization of thyroid nodules

The US morphological features of thyroid nodules are reported in Table 3. The interobserver agreement for reproducibility of US features of thyroid nodules was very good for the suspicious neck lymph nodes (κ = 0.898), shape of nodule (κ = 0.868), maximum diameter of nodule (κ = 0.792), and echogenicity of nodule (κ = 0.750), while the agreement was good for composition of nodule (κ = 0.636), echogenic foci (κ = 0.598), and margins of nodule (κ = 0.524). There was a fair inter-observer agreement for the detection of color Doppler flow (κ = 0.211). Overall agreement among observers for assigning TI-RADS category was good (κ = 0.636).

3.4. Diagnostic performance of ACR TI-RADS for predicting malignant thyroid nodules

On a lesion-based analysis, the diagnostic performance of ACR TI-RADS for predicting malignant thyroid nodules is summarized in Table 4. Considering only those lesions classified as TR5 as predictors for malignancy, the TI-RADS had an accuracy, sensitivity, specificity, PPV, and NPV of 90%, 47.1%, 97.2%, 73.6%, and 91.6%, respectively. Considering TR4 and TR5 together as predictors for malignancy, the accuracy, sensitivity, specificity, PPV, and NPV were 92.1%, 98.5%, 91%, 64.7%, and 99.7%, respectively. Considering TR3, TR4 and TR5 together as predictors for malignancy, the accuracy, sensitivity, specificity, PPV, and NPV were 74.5%, 100%, 70.2%, 36%, and 100%, respectively.

Twenty-seven out of 30 referring clinicians involved in patient management accepted ACR TI-RADS to be a valuable reporting system for medical decision-making in thyroid nodules.

3.5. ROC analyses

The data set of ACR TI-RADS was analyzed to determine the AUC and the best cut-off value for predicting malignant thyroid nodules using ROC curve (Fig. 2, Table 4). The AUC was 0.96 (95% CI 0.94 – 0.98, $p < 0.0001$) and the best cut-off value for predicting malignant thyroid nodules was > TR3. Application of this cut-off value resulted in a sensitivity of 98.3% (95% CI 90.9–100.0), a specificity of 90.9% (95% CI 87.4–93.7), and a +LR of 10.84 (95% CI 7.8–15.1). Representative cases of our study are shown in (Figs. 3–6).

4. Discussion

Our study results indicated that if both TR4 and TR5 are combined as clearly confirming malignant thyroid nodules, the specificity is slightly reduced (91%), on the other hand, the sensitivity is considerably increased (98.5%). Whereas, the combination of TR3, TR4, and TR5 yielded a 100% sensitivity with a much lower specificity (70.2%). Our data are congruent with the overall sensitivity in

Table 3
The inter-observer agreement for morphological diagnostic features and ACR TI-RADS categorization for diagnosis of thyroid nodules.

Feature	Inter-observer	
	κ coefficients	95% CI
Maximum diameter of nodules (mm) (mean ± SD)	**0.792**	**0.714 to 0.870**
Composition	**0.636**	**0.507 to 0.766**
Cystic	0.851	0.723 to 0.979
Spongiform	0.981	0.944 to 1.000
Cystic and solid	0.987	0.969 to 1.000
Solid	0.979	0.958 to 0.999
Echogenicity	**0.750**	**0.595 to 0.905**
Anechoic	1.000	1.000 to 1.000
Hyperechoic	0.955	0.919 to 0.991
Isoechoic	0.971	0.947 to 0.994
Hypoechoic	0.987	0.969 to 1.000
Very hypoechoic	0.798	0.524 to 1.000
Shape	**0.868**	**0.772 to 0.964**
Taller-than-wide	0.868	0.772 to 0.964
Widere-than-tall	0.868	0.772 to 0.964
Margins	**0.524**	**0.148 to 0.899**
Smooth	0.937	0.903 to 0.971
Ill-defined	0.954	0.921 to 0.988
Irregular or lobulated	0.959	0.927 to 0.992
Extrathyroidal extension	1.000	1.000 to 1.000
Echogenic foci	**0.598**	**0.413 to 0.783**
No echogenic foci	0.984	0.967 to 1.000
Large comet tail	0.885	0.757 to 1.000
Macrocalcifications	0.974	0.945 to 1.000
Peripheral	0.604	0.434 to 0.773
Punctate	0.957	0.915 to 0.999
Vascularity	**0.211**	**−0.192 to 0.613**
Peripheral	0.973	0.946 to 0.999
Central	0.853	0.803 to 0.903
none	0.736	0.669 to 0.803
Suspicious neck lymph nodes	**0.898**	**0.757 to 1.000**
Yes	0.898	0.757 to 1.000
No	0.898	0.757 to 1.000
TI-RADS categorization	**0.636**	**0.507 to 0.766**
TR2	0.990	0.975 to 1.000
TR3	0.986	0.966 to 1.000
TR4	0.985	0.956 to 1.000
TR5	0.945	0.884 to 1.000

ACR = American College of Radiology; TI-RADS = Thyroid Imaging Reporting and Data System; SD = Standard deviation; CI = Confidence interval.

Table 4
The diagnostic performance of ACR TI-RADS for predicting malignant thyroid nodules.

	TR5		TR4 + 5		TR3 + 4 + 5	
	%	95% CI	%	95% CI	%	95% CI
Accuracy	89.98		92.08		74.47	
Sensitivity	47.06	38.45 to 55.80	98.53	94.79 to 99.82	100.00	97.32 to 100.00
Specificity	97.17	95.78 to 98.20	91.01	88.83 to 92.89	70.20	66.92 to 73.33
AUC	0.72	0.69 to 0.75	0.95	0.93 to 0.96	0.85	0.83 to 0.87
Positive Likelihood Ratio	16.61	10.69 to 25.81	10.96	8.80 to 13.65	3.36	3.02 to 3.73
Negative Likelihood Ratio	0.54	0.46 to 0.64	0.02	0.00 to 0.06	0.00	
Disease prevalence	14.35	12.18 to 16.74	14.35	12.18 to 16.74	14.35	12.18 to 16.74
Positive Predictive Value	73.56	63.02 to 82.45	64.73	57.81 to 71.23	35.98	31.13 to 41.04
Negative Predictive Value	91.64	89.58 to 93.40	99.73	99.03 to 99.97	100.00	99.35 to 100.00

ACR = American College of Radiology; TI-RADS = Thyroid Imaging Reporting and Data System; CI = confidence interval.

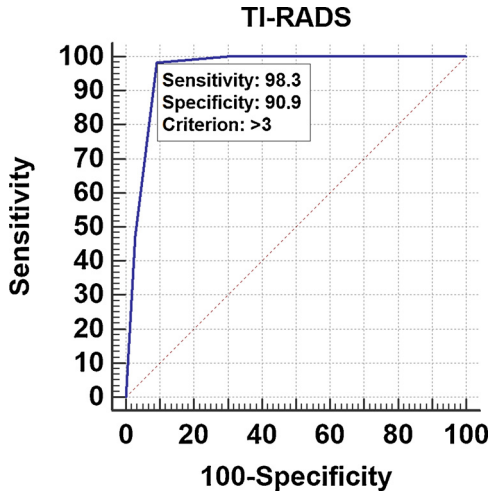

Fig. 2. The ROC analyses of the diagnostic performance of the TI-RADS for predicting malignancy of thyroid nodules as evidenced by FNAB as a reference standard. The best cut-off was > TR3, resulting in a 98.3% sensitivity and a 90.9% specificity.

previously published studies [23–29] that ranged from 72% to 98.5%. Moreover, our findings revealed that the combination of TR4 and TR5 provided remarkably good negative likelihood ratio (0.02), which support the value of ACR TI-RADS for avoiding unnecessary FNAB. Additionally, based on ROC curve analyses, our study revealed an AUC of 0.96 with the best cut-off value for predicting malignant thyroid nodules of < TR3. The application of this cut-off value resulted in a sensitivity of 98.3% and a specificity of 90.9%. Correspondingly, Delfim et al. [30] yielded an AUC of 0.92. Hence, consistent with the studies mentioned above, we should combine both TR4 and TR5 for the diagnosis of malignant thyroid nodules.

In this study, when we considered TR5 only as a predictor for malignancy, 23 pathologically confirmed benign thyroid nodules were falsely diagnosed as highly suspicious for malignancy (TR5) by the ACR TI-RADS. Whereas 72 pathologically confirmed malignant thyroid nodules were falsely diagnosed as mildly (TR3) to moderately (TR4) suspicious for malignancy.

Our findings revealed 73 false-positive nodules and two false-negative nodules when we considered TR4 and TR5 together as predictors for malignancy. The false-positive nodules were 34 benign follicular nodules (Fig. 3), 27 granulomatous (subacute) thyroiditis, and 12 lymphocytic (Hashimoto) thyroiditis (Fig. 6). On the other hand, the false-negative nodules were two papillary carcinomas (Fig. 7). This may indicate that subjective US assessment of thyroid nodules produces a high specificity among the false-negative. However, some malignant thyroid nodules can still be found among those considered benign thyroid nodules. Therefore, we recommend the combination of TR4 and TR5 as a predictor for malignancy of thyroid nodules because if we consider TR5 alone as conclusive for malignant thyroid nodules diagnosis, the ACR TI-RADS will miss a relevant number of malignant thyroid nodules

In our study, we had 111 (81.6%) papillary carcinoma, 11 (8.1%) medullary carcinoma, 10 (7.4%) poorly differentiated carcinoma, two (1.5%) undifferentiated (anaplastic) carcinoma, two (1.5%) metastatic carcinoma, and no follicular carcinoma. As we had a very high prevalence of papillary carcinoma and no follicular carcinoma, the ACR TI-RADS performed well in our cohort.

After using several sonological factors to decide the ACR TI-RADS scoring of the thyroid nodules, our study has a 0, 1.2, 58.3, and 73.6% malignancy risk for TR2, TR3, TR4, and TR5, respectively. These results are agreeing with those reported by Periakaruppan et al. [25], which showed that the risk of malignancy for TR2, TR3, TR4, and TR5 was 0, 2.2, 38.5, and 77.8%, respectively. Similarly, a prospective study using the TI-RADS showed the percentages of malignancy as follows: TR2 (0% malignancy), TR3 (< 5% malignancy), TR4 (5–80% malignancy), and TR5 (> 80% malignancy) [7]. In another retrospective validation study, the authors found a 7.3% and 8.3–96.6% risk of malignancy in TR3 and TR4-5 categories of thyroid nodules [31].

The inter-observer agreement is an essential factor of high clinical value. In the current study, the general agreement of each US feature was good to very good (κ = 0.524–868), which was higher than previous reports [6,26,32,33]. This may be explained by the use of the same US device in all examinations and the real-time images. Also, the higher experience of observers could account for these higher values; however, this is potentially affecting the diagnostic performance of ACR TI-RADS. Thus, further studies about the performance of this reporting system when applied by less experienced radiologists are needed. The overall agreement among observers for assigning ACR TI-RADS category was good (κ = 0.636).

The ACR TI-RADS is easy to understand and apply, as it does not has subcategories, nor does it has a TR0 category to indicate a normal thyroid gland [10]. However, one of the most important limitations of the present study was that TI-RADS is still uncommon and unfamiliar by many clinicians. Therefore, we have made several clinical sessions to all referring clinicians before the start of the study to explain the meaning and aim of ACR TI-RADS. Our screening program amongst referring clinicians involved in patient management showed that 90% of referring clinicians considered ACR TI-RADS a useful tool for clinical decision-making and referral. Moreover, the three clinicians who did not accept ACR TI-RADS in their management protocol attributed that to the novelty of this reporting system.

Although colour Doppler flow and suspicious neck lymph nodes are not part of the ACR TI-RADS, we analyzed these features in our study. We found that the inter-observer agreement was very good for the suspicious neck lymph nodes (κ = 0.898) and fair for the color Doppler flow (κ = 0.211). Our analysis goes in line with the recent European thyroid association guidelines [12] which reported that US assessment

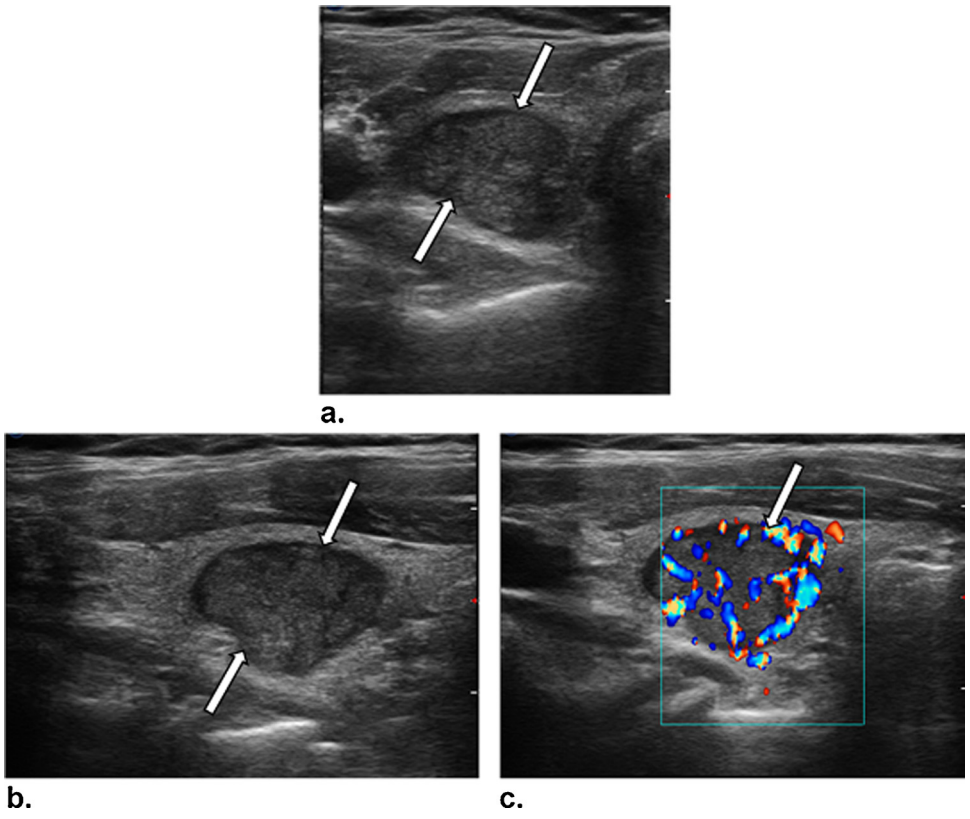

Fig. 3. A 44-year-old woman with a clinically suspected thyroid nodule. (a) Transverse and (b) longitudinal gray-scale ultrasound show a nodule at right thyroid lobe (arrows). The nodule is solid, very hypoechoic, wider-than-tall, and has a smooth margin. (c) Longitudinal color Doppler ultrasonography (not part of TI-RADS) shows marked intranodular vascularity (arrow). The nodule was categorized as TR4. Cytology shows follicular thyroid adenoma.

of the lymph nodes is advised for all thyroid nodules but is mandatory for intermediate- and high-risk ones, whereas the routine use of Doppler US is not recommended for US malignancy risk stratification. This is mainly because the sensitivity of Doppler is highly dependent on the US equipment and settings, and because the definition of central vascularity has a low interobserver agreement [34]. Alternatively, British Thyroid Association (BTA) guidelines reported that characteristic associated lymphadenopathy and mixed, and internal vascularity are classified as U5 (likely malignant), U3 (indeterminate), and U5, respectively [35]. However, we did not assess the added value of these

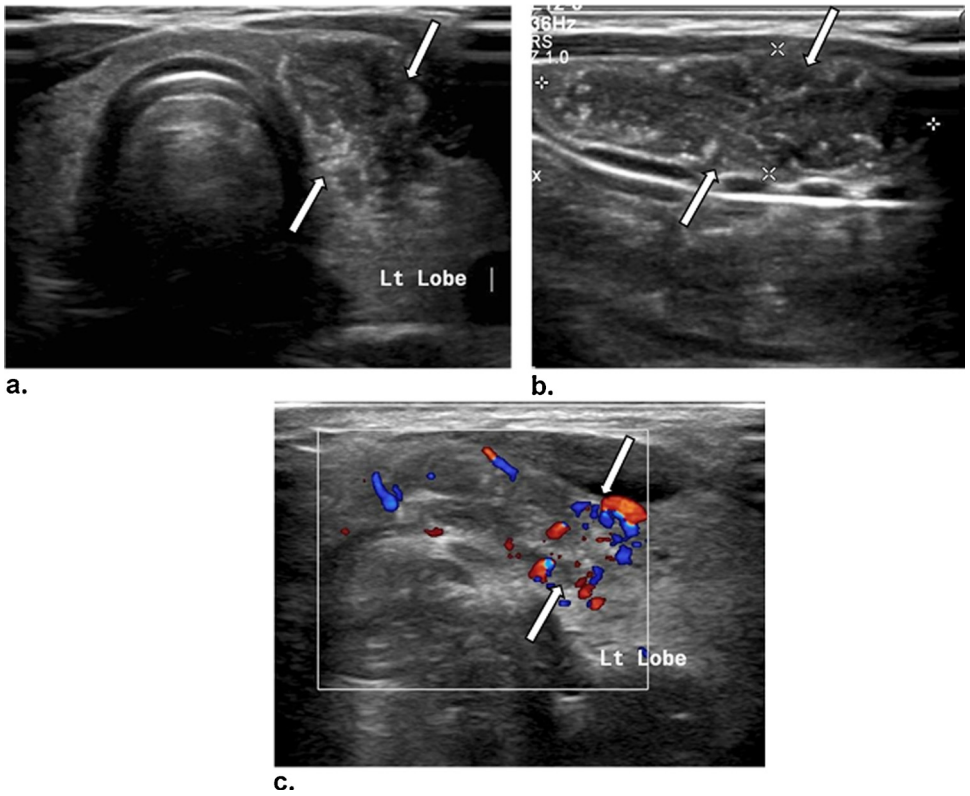

Fig. 4. A 67-year-old woman with a clinically suspected thyroid nodule. (a) Transverse and (b) longitudinal gray-scale ultrasound show a nodule at left thyroid lobe (arrows). The nodule is solid, very hypoechoic, taller-than-wide, has an irregular margin, and shows micro and macrocalcifications. (c) Transverse color Doppler ultrasonography (not part of TI-RADS) shows marked intranodular vascularity (arrow). The nodule was categorized as TR5. Cytology shows papillary thyroid carcinoma.

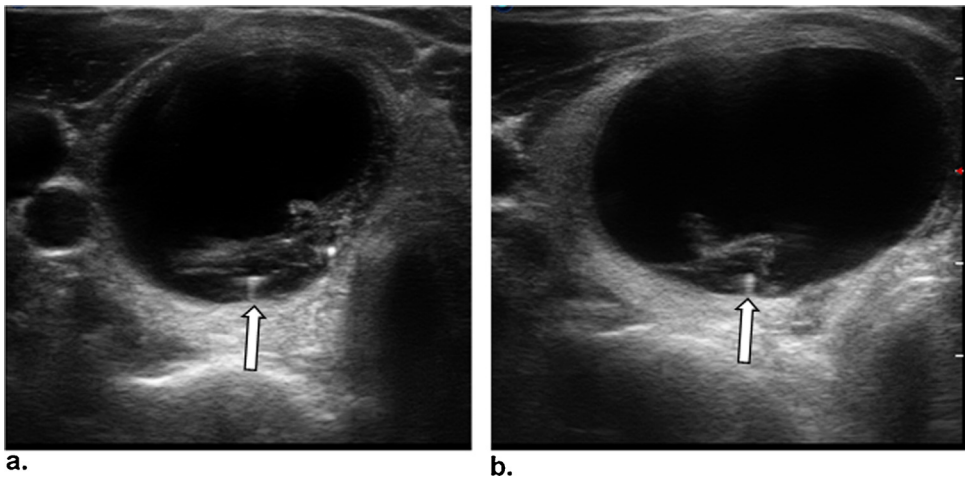

Fig. 5. A 38-year-old man with a clinically suspected thyroid nodule. (a) Transverse and (b) longitudinal gray-scale ultrasound show a nodule at right thyroid lobe (arrows). The nodule is well-defined, predominantly cystic, isoechoic, and shows comet tail artifact. The nodule was categorized as TR2. Cytology showed a colloid/hemorrhagic cyst.

features to ACR TI-RADS.

Based on our findings, which resembles those of previously published studies, the ACR TI-RADS is a categorized reporting of pattern recognition with several advantages: reduce mistakes of important data from US reports by standardizing report structure and content, decrease inconsistency in interpretation of thyroid nodules, increase contact with referring clinicians, help management procedures, and avoid unnecessary FNAB. Moreover, enhance monitoring of outcome, reviewing of performance, assuring of quality, and researching. So, we recommend the application of TI-RASDS in the management protocol of thyroid nodules. However, the ACR TI-RADS has disadvantages as some important sonographic features are not listed in the assignment of TI-RADS categories (e.g., the suspicious lymph node, the halo sign, and the color Doppler examination). Therefore, future large longitudinal studies on the long-term clinical outcomes are still needed to demonstrate the added value of these features to ACR TI-RADS. Additionally, the

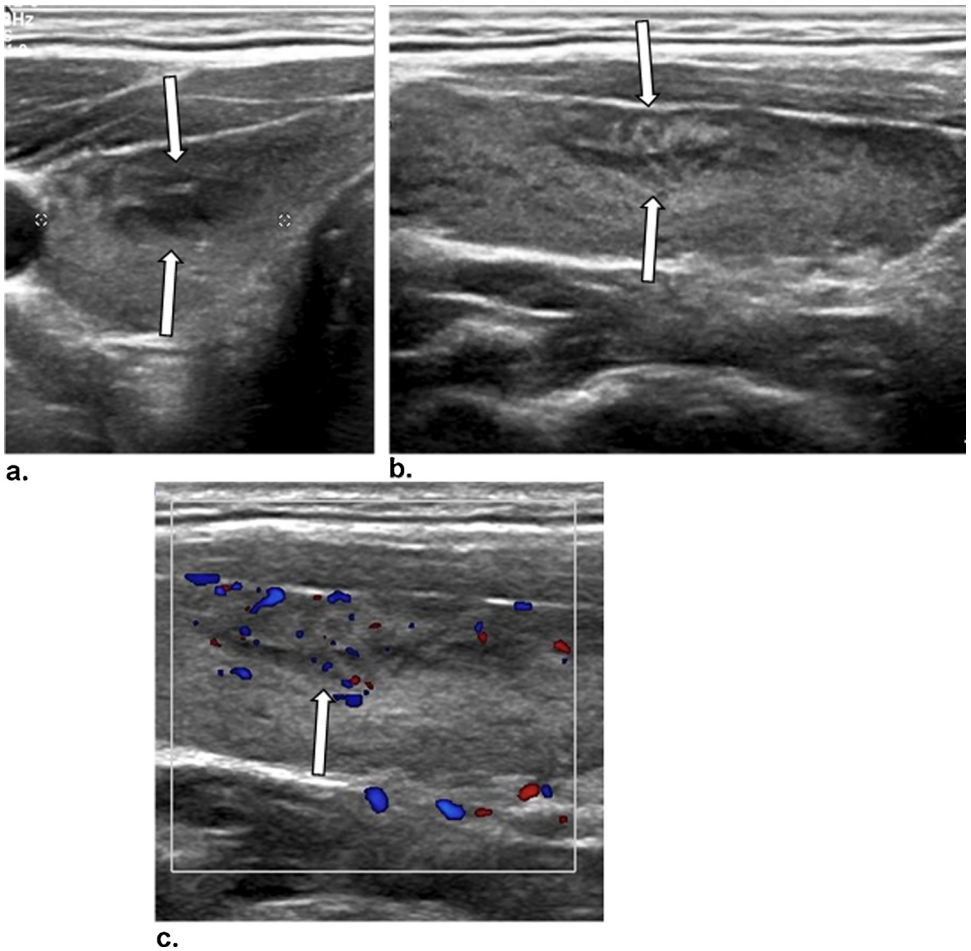

Fig. 6. A 29-year-old woman with a clinically suspected thyroid nodule. (a) Transverse and (b) longitudinal gray-scale ultrasound show a nodule at right thyroid lobe (arrows). The nodule is solid, hypoechoic, wider-than-tall, and has an ill-defined margin. (c) A longitudinal color Doppler ultrasonography (not part of TI-RADS) shows mild intranodular vascularity (arrows). The nodule was categorized as TR4. Cytology shows Hashimoto nodules on a background of Hashimoto thyroiditis.

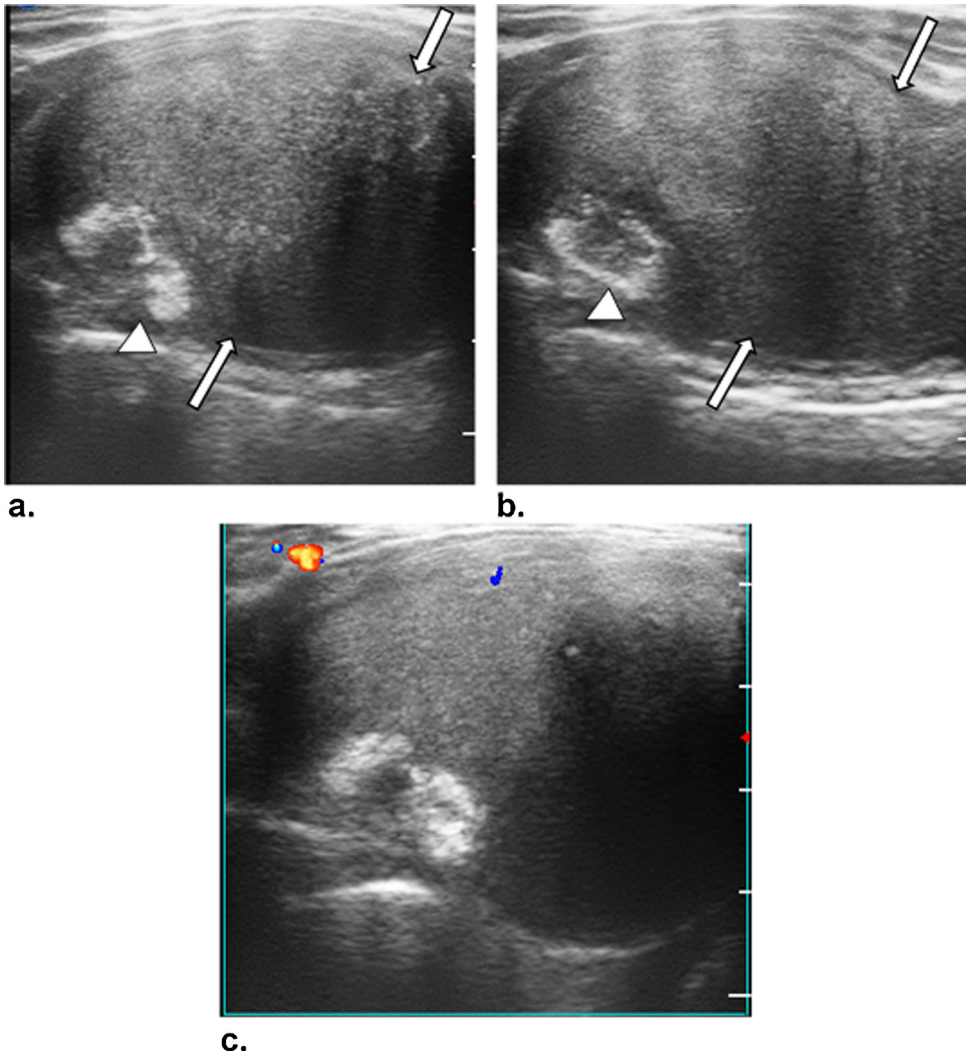

Fig. 7. A 46-year-old man with a clinically suspected thyroid nodule. (a) Transverse and (b) longitudinal gray-scale ultrasound show a large nodule occupying most of the right thyroid lobe (arrows). The nodule is mixed cystic and solid; the cystic component is iso to hyperechoic, wider-than-tall, has a well-defined margin, and shows macrocalcifications of solid component (arrowheads). (c) A transverse color Doppler ultrasonography (not part of TI-RADS) shows no intranodular vascularity. The nodule was categorized as TR3. Cytology shows papillary thyroid carcinoma.

incorporation of new techniques such as elastography is mandatory to increase the power of TI-RADS for detecting malignant thyroid nodules.

Our study has several strengths. It is a large, prospective, multicentre study, which avoids the selection bias of retrospective study. Moreover, real-time images were used for evaluation, which may be a better representative of daily practice. However, our study has limitations. First, the small nodules < 10 mm were excluded from the study. The recent guidelines published by the American Thyroid Association (ATA) [11] state that only thyroid nodules > 10 mm should be evaluated. Second, selection bias from exclusion criteria. This selection bias could have led to the lower malignancy risk of TR3 and TR4 nodules. So large multicentre randomised controlled study is needed to avoid selection bias and validate our findings. Third, the reference standards for benign and malignant diagnoses were based on FNAB only. Even though this may decrease selection bias and verification bias in the study population, it may certainly cause false-negative or false-positive results. Moreover, it is not possible to diagnose follicular carcinoma on cytology. Fourth, the malignant nodules in our study were heavily skewed towards papillary carcinoma, and there were no follicular carcinomas which often present with a different appearance on the US examination. This may contribute to bias and affect the performance of ACR TI-RADS. Therefore, further studies with validation based on histopathology whenever possible are required. Fifth, all US examinations were performed and analyzed by highly experienced radiologists; this is potentially affecting the diagnostic performance of ACR TI-RADS and explain the very good inter-observer agreement in our study. Thus, further studies about the performance of this reporting system when applied by less experienced radiologists are needed. Finally, we did not compare ACR TI-RADS to the other TI-RADS classification systems. However, to date, no TI-RADS has been widely accepted or recommended by any guidelines.

5. Conclusion

Our study established that the ACR TI-RADS improves diagnostic performance of US for predicting malignant thyroid nodules with high validity and high reproducibility.

Guarantor

The scientific guarantor of this publication is the corresponding author.

Funding

The authors state that this work has not received any funding.

Ethical approval

Institutional review boards' approval was obtained.

Informed consent

Written informed consent was obtained from all patients.

Statistics and biometry

The corresponding author has great statistical expertise

Methodology

- Prospective.
- Diagnostic or prognostic study.
- Performed at multiple centers.

Acknowledgements

The authors thank all staff members and colleagues in the Radiology Department-Zagazig University-Egypt for their helpful cooperation and all the study participants for their patience and support.

References

[1] H. Gharib, E. Papini, Thyroid nodules: clinical importance, assessment, and treatment, Endocrinol. Metab. Clin. North Am. 36 (3) (2007) 707–735.
[2] L. Hegedus, Clinical practice. The thyroid nodule, New Engl. J. Med. 351 (17) (2004) 1764–1771.
[3] M.P. Vanderpump, The epidemiology of thyroid disease, Br. Med. Bull. 99 (1) (2011) 39–51.
[4] R. Smith-Bindman, P. Lebda, V.A. Feldstein, et al., Risk of thyroid cancer based on thyroid ultrasound imaging characteristics: results of a population-based study, JAMA Intern. Med. 173 (19) (2013) 1788–1795.
[5] J.D. Iannuccilli, J.J. Cronan, J.M. Monchik, Risk for malignancy of thyroid nodules as assessed by sonographic criteria, J. Ultrasound Med. 23 (11) (2004) 1455–1464.
[6] S.H. Choi, E.K. Kim, J.Y. Kwak, M.J. Kim, E.J. Son, Interobserver and intraobserver variations in ultrasound assessment of thyroid nodules, Thyroid 20 (2) (2010) 167–172.
[7] E. Horvath, S. Majlis, R. Rossi, et al., An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management, J. Clin. Endocrinol. Metab. 94 (5) (2009) 1748–1751.
[8] A. Ozel, S.M. Erturk, A. Ercan, et al., The diagnostic efficiency of ultrasound in characterization for thyroid nodules: how many criteria are required to predict malignancy? Med. Ultrason. 14 (1) (2012) 24–28.
[9] H. Gharib, E. Papini, J.R. Garber, et al., American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules–2016 update, Endocr. Pract. 22 (s1) (2016) 1–60.
[10] F.N. Tessler, W.D. Middleton, E.G. Grant, et al., ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS committee, J. Am. Coll. Radiol. 14 (5) (2017) 587–595.
[11] B.R. Haugen, E.K. Alexander, K.C. Bible, et al., 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) (2015) 1–33.
[12] G. Russ, S.J. Bonnema, M.F. Erdogan, C. Durante, R. Ngu, L. Leenhardt, European Thyroid Association Guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: the EU-TI-RADS, Eur. Thyroid J. 6 (5) (2017) 225–237.
[13] J.H. Shin, J.H. Baek, J. Chung, et al., Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean Society of Thyroid Radiology consensus statement and recommendations, Korean J. Radiol. 17 (3) (2016) 370–395.
[14] J.Y. Kwak, K.H. Han, J.H. Yoon, et al., Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk, Radiology 260 (3) (2011) 892–899.
[15] E.G. Grant, F.N. Tessler, J.K. Hoang, et al., Thyroid ultrasound reporting lexicon: white paper of the ACR thyroid imaging, reporting and data system (TI-RADS) committee, J. Am. Coll. Radiol. 12 (12) (2015) 1272–1279.
[16] J.K. Hoang, J.E. Langer, W.D. Middleton, et al., Managing incidental thyroid nodules detected on imaging: white paper of the ACR Incidental Thyroid Findings Committee, J. Am. Coll. Radiol. 12 (2) (2015) 143–150.
[17] H. Vargas-Uricoechea, I. Meza-Cabrera, J. Herrera-Chaparro, Concordance between the TI-RADS ultrasound criteria and the BETHESDA cytology criteria on the non-toxic thyroid nodule, Thyroid Res. 10 (1) (2017) 1.
[18] G. Grani, L. Lamartina, V. Cantisani, M. Maranghi, P. Lucia, C. Durante, Interobserver agreement of various thyroid imaging reporting and data systems, Endocr. Connect. 7 (1) (2018) 1–7.
[19] D.G. Na, J.H. Baek, J.Y. Sung, et al., Thyroid imaging reporting and data system risk stratification of thyroid nodules: categorization based on solidity and echogenicity, Thyroid 26 (4) (2016) 562–572.
[20] G. Russ, Risk stratification of thyroid nodules on ultrasonography with the French TI-RADS: description and reflections, Ultrasonography 35 (1) (2016) 25.
[21] A.R. Zayadeen, M. Abu-Yousef, K. Berbaum, JOURNAL CLUB: retrospective evaluation of ultrasound features of thyroid nodules to assess malignancy risk: a step toward TIRADS, Am. J. Roentgenol. 207 (3) (2016) 460–469.
[22] E.S. Cibas, S.Z. Ali, The Bethesda System for reporting thyroid cytopathology, Thyroid 19 (11) (2009) 1159–1165.
[23] J.K. Hoang, W.D. Middleton, A.E. Farjat, et al., Reduction in thyroid nodule biopsies and improved accuracy with american college of radiology thyroid imaging reporting and data system, Radiology 287 (1) (2018) 185–193.
[24] L. Gao, X. Xi, Y. Jiang, et al., Comparison among TIRADS (ACR TI-RADS and KWAK-TI-RADS) and 2015 ATA Guidelines in the diagnostic efficiency of thyroid nodules, Endocrine 64 (1) (2019) 90–96.
[25] G. Periakaruppan, K.G. Seshadri, G.M. Vignesh Krishna, et al., Correlation between ultrasound-based TI-RADS and Bethesda system for reporting thyroid-cytopathology: 2-year experience at a tertiary care center in India, Indian J. Endocrinol. Metab. 22 (5) (2018) 651–655.
[26] A. Chandramohan, A. Khurana, B.T. Pushpa, et al., Is TI-RADS a practical and accurate system for use in daily clinical practice? Indian J. Radiol. Imaging 26 (1) (2016) 145–152.
[27] F. Mao, H.X. Xu, H. Zhou, et al., S. Assessment of virtual touch tissue imaging quantification and the ultrasound thyroid imaging reporting and data system in patients with thyroid nodules referred for biopsy, J. Ultrasound Med. 37 (3) (2018) 725–736.
[28] J.H. Yoon, H.S. Lee, E.K. Kim, H.J. Moon, J.Y. Kwak, Malignancy risk stratification of thyroid nodules: comparison between the thyroid imaging reporting and data system and the 2014 American thyroid association management guidelines, Radiology 278 (3) (2015) 917–924.
[29] G. Russ, B. Royer, C. Bigorgne, A. Rouxel, M. Bienvenu-Perrard, L. Leenhardt, Prospective evaluation of thyroid imaging reporting and data system on 4550 nodules with and without elastography, Eur. J. Endocrinol. 168 (5) (2013) 649–655.
[30] R.L. Delfim, L.C. Veiga, A.P. Vidal, F.P. Lopes, M. Vaisman, P.D. Teixeira, Likelihood of malignancy in thyroid nodules according to a proposed Thyroid Imaging Reporting and Data System (TI-RADS) classification merging suspicious and benign ultrasound features, Arch. Endocrinol. Metab. 61 (3) (2017) 211–221.
[31] J.Y. Kwak, I. Jung, J.H. Baek, et al., Korean Society of Thyroid Radiology (KSThR) and Korean Society of Radiology. Image reporting and characterization system for ultrasound features of thyroid nodules: multicentric Korean retrospective study, Korean J. Radiol. 14 (1) (2013) 110–117.
[32] J. Koh, S.Y. Kim, H.S. Lee, et al., Diagnostic performances and interobserver agreement according to observer experience: a comparison study using three guidelines for management of thyroid nodules, Acta radiol. 59 (8) (2018) 917–923.
[33] C.S. Park, S.H. Kim, S.L. Jung, et al., Observer variability in the sonographic evaluation of thyroid nodules, J. Clin. Ultrasound 38 (6) (2010) 287–293.
[34] A. Lyshchik, R. Moses, S.L. Barnes, T. Higashi, R. Asato, M.I. Miga, J.C. Gore, Fleischer AC: quantitative analysis of tumor vascularity in benign and malignant solid thyroid nodules, J. Ultrasound Med. 26 (6) (2007) 837–846.
[35] P. Perros, K. Boelaert, S. Colley, C. Evans, R.M. Evans, B.A.G. Gerrard, J. Gilbert, B. Harrison, S.J. Johnson, T.E. Giles, L. Moss, Guidelines for the management of thyroid cancer, Clinical Endocrinol. 81 (2014) 1–22.