Journal of Biological Research

Bollettino della Società Italiana di Biologia Sperimentale

eISSN 2284-0230 - pISSN 1826-883

https://www.pagepressjournals.org/index.php/jbr/index

J Biol Res 2026 [Online ahead of print]

To cite this Article:

Al-Waeli DK, Mohammed AG, Odhaib SA, et al. **Thyroid nodules stratification by American Thyroid Association and American College of Radiology lexicons: Thi-qar study.** *J Biol Res* doi: 10.4081/jbr.2026.14148

Early access: 19 March 2026 Accepted: 29 January 2026 Submitted: 16 July 2025

Thyroid nodules stratification by American Thyroid Association and American College of Radiology lexicons: Thi-qar study

Dheyaa Kadhim Al-Waeli,[1,2] Adel Gassab Mohammed,[1,2] Samih Abed Odhaib,[1,3] Mahmood Thamer Altemimi,[1,2] Shakir Agla Khalifa,[1,2] Israa Ali Hasan,[1,2] Mohammed Uday Hatim,[1,2] Imad Hatim Tahir[1,2]

[1]Thi-Qar Specialized Diabetes, Endocrine and Metabolism Center (TDEMC), Thi-Qar Health Directorate, Thi-Qar, Iraq; [2]College of Medicine, University of Thi-Qar, Nasiriyah, Thi-Qar, Iraq; [3]College of Medicine, University of Sumer, Al-Rifa'i, Thi-Qar, Iraq

Corresponding Author

Mahmood Thamer Altemimi, Thi-Qar Specialized Diabetes, Endocrine and Metabolism Center (TDEMC), Thi-Qar Health Directorate, Thi-Qar, Iraq. Tel.: +9647807326088. E-mail:mahmoodaltimimi83@gmail.com

Key words: fine needle aspiration; thyroid cancer; thyroid nodule; ultrasound.

Abstract

The diagnostic performance of sonography to manage thyroid nodule (TN) may influence excess biopsy in different clinical settings. American College of Radiology (ACR) presented Thyroid Imaging Reporting and Data Systems (TIRADS) lator to American Thyroid Association (ATA) risk stratification guidelines, with variable diagnostic performance. The objective of this study is to evaluate the effectiveness of ACR-TIRADS versus ATA lexicons for thyroid cancer prediction among individuals with TN in Thi-Qar, Iraq. This cross-sectional observational study enrolled 1,007 individuals aged ≥ 14 years with 1,155 TNs attending a tertiary endocrine center. Malignancy risk was classified using both ACR-TIRADS and ATA lexicons with matched categories: TIRAD-1 (Benign<1%), TIRAD-2 (Very Low<3%), (TIRAD-3 (Low 5-10%), (TIRAD-4 (Intermediate 10-20%), and (TIRAD-5 (High 70-90%). Fine needle aspiration (FNAC) and histopathological diagnosis were made according to the Bethesda system and agreement between lexicons was assessed using

sensitivity, specificity, Cohen's Kappa Statistics, and McNemar test. Among participants, 85.2% were women, with mean age 46±14 years. Most individuals were euthyroid (71%), TNs were located in right lobe (51.3%), left lobe (42.5%), and isthmus (6.1%). FNAC was indicated for 54.6% of TNs by ATA and 44.6% by ACR-TIRADS criteria. High-risk cytology (Bethesda IV-V) was 7.1% and ATA system demonstrated higher sensitivity (74.3%) and accuracy (58.6%) than ACR-TIRADS (sensitivity 60%, accuracy 53%), though both gave modest diagnostic performance and low positive predictive values (<12%). Agreement between systems was weak (Cohen's κ=0.265±0.032, $p < 0.001$), with concordance in 50.1% of cases. Both risk stratification systems had comparable but modest diagnostic performance. ATA identified more FNAC-eligible TNs with slightly higher sensitivity, specificity, prediction, and overall accuracy. However, the weak agreement between raters highlights variability in nodule categorization and management, supporting the need to integrate sonographic findings with clinical and biochemical factors to optimize TNs evaluation and minimize unnecessary biopsies.

Introduction

Thyroid Nodules (TNs) are detected in up to 50-60% of healthy people. About 95% of TNs are asymptomatic and discovered incidentally during the neck Ultrasound (US) evaluation of non-thyroidal lesions. They appear in euthyroid people with no compressive or cosmetic concerns.[1-3] Still, 2-15% of TNs show malignant potentials even without symptoms.[4-6]

In most TNs, US findings are not predictive of malignancy.[1,7] Features that raise suspicion for thyroid cancer are marked hypoechogenicity, microcalcifications, irregular margins, and taller-than-wide shape.[8,9] The absence of suspicious sonographic signs by a single point US imaging may not be considered conclusively diagnostic for a benign lesion.[10,11]

According to the revised 2015 American Thyroid Association (ATA) management guidelines, TN is classified into five categories based on their qualitative US characteristics.[7] These categories were: high suspicious with an estimated Risk Of Malignancy (ROM) of 70-90%, intermediate suspicious (10-20%), low suspicious (5-10%), very low suspicious (<3%), and benign (cyst). The ATA recommends diagnostic fine needle aspiration (FNAC) for high

and intermediate suspicious TN ≥1 cm, low suspicious ≥1.5 cm, and very low suspicious ≥2 cm.[7,12]

In 2017, the American College of Radiology (ACR) presented the Thyroid Imaging Reporting and Data Systems (TIRADS), a different, five-tier approach, based on a quantitative scoring system and a higher size threshold for FNAC with superior diagnostic accuracy and specificity for TNs, helping reduce unnecessary FNACs for benign nodules which was proposed by the qualitative guideline of the ATA.[9]

However, risk stratification is complicated by the substantial harm of overdiagnosis and overtreatment of low-risk diseases. Imaging-based classification systems do not differentiate higher-risk from lower-risk cancers; rather, they are designed to assess the risk of a nodule being malignant.[7,9,13]

The main objective of this study was to compare the malignancy diagnostic performance and the FNAC recommendations using ATA malignancy risk stratification and ACR-TIRADS for TNs.

Materials and Methods

Study design and population

This is a cross-sectional, observational study which enrolled 1,007 individuals aged 14 years old or more harboring 1,155 TNs. They attended Thi-Qar Specialized Diabetes Endocrine and Metabolism Center (TDEMC), Thi-Qar, Southern Iraq, for a consultation about their thyroid concern from March 2019 to April 2022. The exclusion criteria involved individuals who did not consent to enrollment or refused examination or had subcentimetric TN (<1cm).

The sample size was calculated according to the equation:[14]

$$\boldsymbol{N = \frac{Z(1-\alpha/2)^2\, P\,(1-P)}{d^2}}$$

Where:
N = minimum required size of the sample,
P = proportion of TN in the population which was (68%) according to known evidence,[4]
Z = confidence level that will be used (Z = 1.96 for 95),

d = desired margin of error (=0.05).

The minimum sample size required to conduct this study was N = 334 cases, but the real number of TNs cases in this study was 1,155 for more accuracy.

Ethical approval and informed consent

Ethical approval was obtained from the Ethical Committee of the Tertiary Endocrine Center and the Institutional Review Board (IRB) (3T/2/2019) for the study on February 2019, according to the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards, and an informed consent was obtained from all patients clearly after ensuring a full detailed description and intervention along the whole study timeline.

Baseline demographic and social characteristics were implemented for every enrolled patient including age, gender, weight, Body Mass Index (BMI), residency, occupation, smoking, educational level, and comorbidities like diabetes mellitus, hypertension, or ischemic heart disease. A thorough history and relevant clinical examination including vital signs were taken. A complete thyroid exam includes inspection and palpation of the thyroid gland and the lateral neck cervical lymph nodes.

Ultrasound evaluation

An expert radiologist did thyroid ultrasound for the enrolled individuals using Philips Affiniti 30. (Eindhoven. Netherlands) or LOGIQ E9. (GE Healthcare Wisconsin, USA). Sonographic features of every thyroid nodule were reported precisely like i) size (maximum diameter measures in millimeters; the nodular size threshold in this study was minimally 10 mm), ii) border (smooth, ill-defined, lobulated, or irregular), iii) echogenicity (hyperechoic, isoechoic, hypoechoic, or markedly hypoechoic related to surrounding thyroid parenchyma), iv) texture (solid, mixed cystic-solid, or cystic), v) taller than wide behaviour (defined as antero-posterior diameter greater than the transverse diameter on transverse imaging), vi) presence of foci of calcifications, such as microcalcification, macrocalcification, or peripheral (rim) calcifications, vii) any extra-thyroidal extension beyond the thyroid gland (identifying any distruption of the thyroid capsule or any invasion into adjacent soft tissues), viii) lateral neck

lymph nodes with abnormal features suggesting metastasis like (enlargement, loss of fatty hilum, round shape, microcalcification, cystic changes, or abnormal vascularity).

Malignancy risk assessment was done for each TN according to the ATA malignancy risk stratification and ACR-TIRADS stratification rater systems, i.e., in individuals with more than one thyroid nodule, we considered each TN alone. We adopted matching of the two raters: (TIRAD 1 = Benign <1%), (TIRAD 2 = Very Low <3%), (TIRAD 3 = Low 5-10%), and considered them collectively in subsequent measurements as minimal to moderate malignancy risk. On the other hand (TIRAD 4 = Intermediate 10-20%), and (TIRAD 5 = High 70-90%) were considered collectively also in the subsequent measurements as high malignancy risk.

Fine needle aspiration procedure

Each TN underwent risk stratification by using both the ATA and ACR-TIRAD systems. A full sterile aseptic method supported by US-guided FNAC was achieved from the indicated nodule when the neck slightly extended using a 21-23G 10 mL syringe. To obtain a good yield, two aspiration attempts were made. After visualization of the needle tip within the targeted thyroid nodule, FNA was performed by controlled forward movements of the needle by the endocrinologist, resulting in mechanical acquisition of tissue fragments that were retained within the needle lumen. Addititional cellular material can be collected by the simple backward motion of the needle for the next stroke forward.

For the first biopsy, the needle oscilated approximately three per second in the FNAC process with an intra-nodular dwell time of 3-5 seconds which begins when the needle reached the target lesion. This maneuver provided a suitable cellular field and less blood dilution.

Each nodule got at least two to four slides and each slide with satisfied clusters of cells (not less than 100 cells per field) were arranged to smear the aspirated material; all slides were fixed in 95% ethanol and transferred on the same day to a histopathologist.

The FNAC and histopathological diagnosis were made for 515 TNs according to each rater criteria and the individual acceptance for FNAC. Respondents favored a selective approach rather than sampling all nodules according to size criteria alone for patients with multinodular

thyroid glands containing several sonographically similar solid hypoechoic nodules. This study reported all cytologic interpretations using the Bethesda system.[6]

Biochemical and hormonal tests

For estimation of the thyroid function, our laboratory used cobas e411-Roche diagnostics (Mannheim, Germany), and electrochemiluminescence technology for immunoassay analysis. Thyroid function status interpretation in each enrolled individual with TN was made according to the current guidelines as Thyrotropin (TSH) (0.27–4.2 µIU/mL), free thyroxine (FT4) (0.93–1.7 ng/dL), subclinical hypothyroidism was defined as elevated serum TSH levels (4.5–10 mU/L) with normal FT4 concentrations, while subclinical hyperthyroidism was defined by suppressed TSH levels with normal FT4 and total triiodothyronine (T3) values (Table 1)

Statistical analysis

For statistical analysis, we used IBM-SPSS Statistics for Windows, Version 26.0. (IBM Corp., Armonk, NY, USA) to evaluate the different variables. The study used the mean ± Standard Deviation (SD) or frequency (%) for data expression.

We tested sensitivity, specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV), positive and negative likelihood ratio, and the accuracy of both raters for TNs that underwent FNAC after excluding the unclassified cases by either of the raters. The agreement between the two raters was measured using Kappa Statistics and the McNemar-Bowker Test. We considered a p value ≤0.05 as significant.

Results

Table 1 shows that more than 85% of the cohort were women (n=858), middle-aged (46 ± 14 years old), overweight to obese (BMI=31.77 ± 6.66 kg\m^2), and mostly nonsmokers (937 - 93%). Most participants were housewife 73.6%, employee 15.2%, free worker 4.9%, retired 3.5%, and students 2.8%. According to educational level, less than two third of the

participants were at low level of education (illiterate 24.1%, primary school 40.4%), while others were graduated from secondary to institutional degrees.

Also, Table 1 shows that 71% of the enrolled individuals were euthyroid (n=820), and the rest distributed as hypothyroidism 15.2%, hyperthyroidism 9.3%, subclinical hyperthyroidism 2.4%, and subclinical hypothyroidism 2.1%.

The full description of the nine TNs that underwent FNAC, although the raters advised against and revealed Bethesda 4 and 5, is shown in Supplementary Table 2S.

Table 2 shows around 46% of nodules were solitary (n=527), and another 46% were part of multinodular goiter (n=525). Only 6% of the nodules resided in the isthmus (n=71), with the rest being distributed between the right (n=593, 51.3%) and left (n=491, 42.5%) lobes. According to ATA categories, there were 631 (54.6%) nodules needing FNAC cytology while only 515 (44.6%) nodules needing FNAC cytology according to ACR-TIRAD categories (Table 2).

The majority of these TNs were isoechoic (78.9%), followed by hypoechoic (8.5%), hyperechoic (7.8%), and anechoic (4.8%). Most nodules showed no calcification or only large comet-tail artifacts (78.5%), while macrocalcifications were present in 9.3%, punctate calcifications in 10.6%, and peripheral rim calcifications in 1.6% of cases (Table 2).

Regarding nodule margins, nearly all were smooth (98.2%), with only a small proportion being ill-defined (1.1%) or lobulated/irregular (0.7%). Vascularity was predominantly absent (89.6% not vascular), while peripheral and intranodular vascularity were noted in 6.2% and 4.2%, respectively (Table 2).

Lymphadenopathy was uncommon and 95.9% showed no nodal involvement; when present, it was mostly ipsilateral (2.1%) or bilateral (1.7%). Among the 47 evaluated lymph nodes, 38 (3.3%) were inflammatory and 9 (0.8%) were considered suspicious (Table 2).

The majority of FNAC procedures were performed in 2021 (50.9%), followed by 2020 (40.5%) and 2019 (8.6%). Of the 80 patients with previous thyroid surgery, subtotal thyroidectomy was most common (67.5%), followed by lobectomy with or without isthmectomy (17.5%), total thyroidectomy (11.3%), and nodulectomy (3.8%) (Table 2).

Table 3 illustrates the categorization of the 1,155 TNs by the two comparing raters and the different US characteristics of these nodules. ACR-TIRADS lexicon could categorize all the TNs, while the ATA lexicon could not categorize 21 (1.8%) TNs. Most TNs were at low

suspicion (ATA 45.6%, ACR-TIRAD 37.7%), very low suspicion (ATA 26.8, ACR-TIRAD 30.3%), and intermediate suspicion (ATA 18.1%, ACR-TIRAD 20.3%). A smaller portion of TNs were high suspicion (ATA 6.9%, ACR-TIRAD 4.2%), or benign (ATA 3.5%, ACR-TIRAD 4.8%).

After full consent from the enrolled individuals, FNAC was done for 536 high suspicious TNs by either system. Bethesda 1-3 were considered low risk of malignancy and distributed as 27.1%, 60.8%, and 5% respectively, while Bethesda 4 (1.9%) and 5 (5.2%) were considered high risk (Table 5).

Table 3 shows the risk of malignancy according to the score of each rater when higher ACR-TIRAD and ATA categories carried a higher risk for malignancy in a positive correlation manner. In the ACR-TIRAD system, malignancy risk rose from 6.6% in TIRAD 2 to 11.4% in TIRAD 5, while in ATA system it rose from 4.8% in very low suspicion category to 43.8% in high-suspicion category. Furthermore, low-risk categories tend to have larger diameter (26-29 mm) than high-suspicion category (20 mm) (Table 3).

Table 4 summarizes the data in Table 3, which includes the results of 515 TNs only after removing the results of the 21 unclassified nodules and matching the raters.

After excluding unclassified nodules, 515 TNs were analyzed. Using the ACR-TIRADS system, malignancy was more frequent in higher categories (TIRADS 4–5: 8.4%) compared to lower categories (TIRADS 1–3: 5.3%). Similarly, under the ATA system, high-suspicion categories (ATA IV–V) showed a higher malignancy rate (11.3%) than low-suspicion categories (ATA I–III, 3.2%). These findings confirm that both lexicons demonstrate an increasing trend of malignancy risk with higher classification levels.

Table 4 shows both raters have a relatively similar diagnostic performance for malignant TNs. ATA system showed higher sensitivity (74.3%) and accuracy (58.6%) than ACR-TIRADS (sensitivity 60%, accuracy 53%), though both raters show lower sensitivity figures, gave modest diagnostic performance and low PPV (<12%). Other precision measures were similar between both raters and favored the ATA classification to an extent. The likelihood ratios of the nodules using both raters reflect the high possibility of over- and underdetection of high-risk TN.

The agreement between the two raters was estimated using Kappa Statistics and McNemar's test which was weak (Cohen's $\kappa=0.265\pm0.032$, $p < 0.001$), with concordance in 50.1% of cases. (Supplementary Table 1S). The agreement between the raters was significant

($p<0.001$) yet minimal, with a kappa value equal to 0.265, i.e., less than weak. The test shows a crude agreement in 50.10% of the cases only, representing 258 TNs out of 515 TNs (after matching the different raters' stages and after the exclusion of the 21 cases that were unclassified by the ATA Risk Stratification System (RSS). The McNemar-Bowker chi-square is significant ($p<0.001$) and equals 24.274, suggesting a statistically significant difference between the two raters (Supplementary Table 1S).

Table 5 illustrates the need for FNAC according to both raters. Medical suspicion was the base of performing FNAC for suspicious thyroid nodules, which did not fulfill FNAC requirements by either rater.

Discussion

The ATA and ACR-TIRADS lexicons for thyroid malignancy RSS provide an objective framework for clinicians to assign the ROM to thyroid nodules. No proven universal guideline has been proposed to reduce unnecessary FNACs and find as many thyroid cancers as possible. It has also been difficult to compare the risk stratification systems, as each of them uses a different size threshold to recommend FNAC. However, many studies have compared these guidelines' diagnostic performances and excessive FNAC rates.

In this large pool study with TNs evaluation, more than 85% of the participants were overweight or obese women of middle-aged group supporting the high occurrence of TNs among women than men in a ratio 6:1 which was consistent with two local studies in Iraq (85.5%, 7:1) and global epidemiological distribution of nodular thyroid disease prevalence worldwide.[9,15,16]

Most of the participants were in euthyroid state (71%), while one-quarter of them were either overt or subclinical dysfunction (hypothyroidism 15.3%, hyperthyroidism 9.3%). This is concordant with the Middle- East Report for thyroid dysfunction assessment.[17] Sonographic features, especially vascularity pattern, echogenicity, taller than wide, and ill-defined borders are crucial to decide for FNAC in patients with hyperthyroid state and non-judged thyroid scintigraphy.

In this large cohort study, FNAC cytology was indicated for 54.6% and 44.6% of TNs by ATA and ACR-TIRAD criteria, respectively. Both ATA and ACR-TIRADS raters provided comparable, but modest diagnostic performance. The ATA lexicon showed slightly higher sensitivity, specificity, prediction, and overall accuracy, identifying a greater proportion of

TNs requiring FNAC, although we found that the 2017 ACR-TIRADS was found to be a less invasive and more conservative risk stratification system with fewer referrals for FNAC compared to 2015 ATA guidelines, *i.e.*, 515 versus 631 TNs, respectively, with 18.4% reduction of the referred TNs.

So, the data were comparable to Ahmadi *et al*. who demonstrated that the 2017 ACR-TIRADS and 2015 ATA had similar diagnostic values in detecting thyroid cancer.[18] Conversely, other studies showed that referrals for FNAC could be reduced by 35-53% using ACR-TIRADS, with very low false negative rates of only 2.2%[19-21] and Gao *et al*. showed the 2015 ATA to be more effective in detecting malignant nodules when compared to the 2017 ACR-TIRADS.[22]

In this cohort, the average dimensions of targeted TNs were Length=2±1 cm × Width=1.5±0.8 cm, which were considered optimal as recommended by most guidelines. Furthermore, Middleton *et al*. found that any modification of the size cut-offs to decrease the number of missed malignancies would mainly lead to an increase in the number of biopsied benign nodules without a significant impact on the prognosis of malignant ones.[23]

Many studies contributed to reducing the number of unnecessary biopsies by ACR-TIRADS to the higher size threshold or cut-offs for FNAC referral compared to ATA RSS.[21,25-28] For example, the ACR-TIRADS thresholds for biopsy are 1.5 cm for TIRAD4 and 2.5 cm for TIRAD3, while the ATA guidelines use 1 cm for similar nodules. The ATA guidelines recommend that FNAC be "considered" for spongiform nodules >2 cm, whereas ACR-TIRADS recommends that spongiform nodules receive no FNAC or follow-up at any size. ACR-TIRADS also considers TIRAD2 mixed cystic, solid nodules without suspicious features as benign and not requiring FNAC or follow-up at any size. This would, in the future, prevent the underdiagnosis of thyroid cancer and reduce unnecessary workup of benign nodules.[7,29]

By this study, ATA system showed higher sensitivity (74.3%) and accuracy (58.6%) than ACR-TIRADS (sensitivity 60%, accuracy 53%), though both raters show lower sensitivity figures, gave modest diagnostic performance and low positive predictive values (<12%). The alteration of these metrics by ACR-TIRADS may be true for very large cohorts like that of Hoang *et al*.,[29] who showed that the effect of reduced biopsy recommendations using ACR-TIRADS would lead to higher specificity and potentially lower sensitivity compared with ATA. Still, it is important to note that most studies compare RSS solely on biopsy

recommendations and do not account for ACR-TIRADS' follow-up guidance for suspicious nodules below the size cut-offs for FNAC.[29]

Our results may agree on some points with the Wu *et al*. study[25] who concluded that compared with the ACR-TIRADS guidelines, the ATA guidelines had a higher diagnostic sensitivity (93.4%, $p < 0.001$) and a larger Negative Predictive Value (NPV) (85.3%, $p = 0.034$). However, compared with the ATA guidelines, the ACR-TIRADS guidelines yielded a higher specificity (66.0%, $p < 0.001$), larger PPV (73.6%, $p = 0.001$), and greater accuracy (75.5%, $p = 0.017$).[25] These findings may be caused by the scoring of macrocalcification and very hypoechoic nodules in the ACR guidelines.

Thyroid malignancy on histopathology of the TNs that underwent FNAC (n=536) was 7.1% (N=38), and it favored ATA over ACR-TIRADS. Using ATA guidelines, we were able to diagnose 26/230 as malignant in highly suspicious groups, compared to 21/249 TNs for ACR-TIRADS 4 and 5.

Despite the high proportion of low suspicious TNs in both systems, n=269 (50.2%) in TIRADS and n=285 (53.2%) in ATA, malignant TNs TIRADS category 2 or 3 (n=269) and ATA very low and low risk (n=285) represented a significant proportion of patients, 50.2% and 53.2%, respectively. Malignant thyroid nodules were identified in these 15/269 and 9/285, according to ATA RSS. These findings encourage the clinical judgement of physcians in performing FNAC for low-risk TNs with some attractive precocious features. In comparison to our results, Mussoi de Macedo *et al*. and Rosario *et al*. did not show any malignant nodules in their large cohorts of low suspicious thyroid nodules by ATA.[30,31] Karaagac *et al*. demonstrated that the probability of malignancy in TIRADS 2, 3, 4, and 5 was 0%, 3.4%, 14%, and 87%, respectively.[32] Mussoi de Macedo *et al*. observed a malignancy risk of 100% of nodules TIRADS 5.[30] Therefore, proven high NPV of both systems could allow ultrasonographically (without FNAC) monitoring of these nodules categories unless they increase in volume or if there are new suspicious sonographic features.

The unclassified nodules that did not follow any pattern in ATA RSS were 21 TNs, comprising 2.1% of the total number of enrolled TNs, and 3.9% of the biopsied nodules by FNAC. These nodules were classified according to ACR-TIRADS but not ATA RSS. Our results confirm an applicable limit of the ATA classification, which leaves unclassified nodules to be reclassified by the ACR-TIRADS RSS, reflecting another design goal. The ATA RSS could not classify every TN due to factors in its design, which could not evaluate

the combination of US features. ATA guidelines did not include evaluating hyperechoic or isoechoic nodules with malignant signs (taller-than-wide, punctate foci, marginal irregularities, etc.).

Furthermore, our results in last points were also seen in a larger cohort (n=3,323) by Ha *et al*. who showed that 1.6% (54 of 3,323) of TNs did not meet the criteria for any pattern in the ATA guidelines. These nodules had a 25.9% (14 of 54) risk of malignancy and, therefore, could be categorized as an intermediate-suspicion group.[33] In addition, Yoon *et al*. and Ha *et al*. reported that ATA guidelines could not classify 3.4% (44/1,293) and 5.0% (100/2,000) nodules. The malignancy rates of these nodules were 18.2% and 19.0%, respectively.[34,35] Middleton *et al*. found that 13.9% of nodules could not be classified by the ATA guidelines.[23]

Ha *et al*. showed that the malignancy risk of the not suspicious and mildly suspicious categories in the ACR-TIRADS guidelines were 2.7% (16/590) and 5.4% (49/914), respectively.[35] Following the ACR-TIRADS guidelines, such nodules do not require biopsy or active surveillance.

A substantial proportion (65%) of thyroid nodules classified as TR4 or TR5 were determined to be benign, underscoring the critical role of ongoing surveillance and comprehensive clinical assessment for nodules within these categories.[18] Certain nodules met the established criteria for repeat ultrasonographic evaluation after a 12-month interval. Notably, the investigators reported that approximately 6% of thyroid malignancies identified in their cohort would not have been detected under these follow-up protocols.[18]

Especially for the false negative diagnosis by ATA RSS, it is important to consider the longitudinal follow-up even if they appeared first as benign. It is possible to avoid FNAC in very low-suspicion nodules if health care providers can ensure appropriate and timely surveillance. On the other hand, the size threshold for FNAC could cause some missing, as shown by Middleton *et al*., who showed that the majority of malignant nodules that do not receive FNAC recommendations based on the ACR-TIRADS receive a recommendation of follow-up USs for five years, are < 1 cm in maximum diameter, or both.[36]

In this study, ATA system showed higher sensitivity (74.3%) and specificity (578.5%) than ACR-TIRADS (sensitivity 60%, specificity 52.5%). Given the higher size threshold for the ACR-TIRADS compared to ATA RSS, we could expect that its diagnostic metrics are lower. Table 5 ensures such an assumption regarding the relatively low sensitivity of both RSS but not the specificity. Our result regarding the performance parameters may be different from

Castellana *et al*. meta-analysis,[37] which showed higher sensitivity and specificity due to reasons related to the cumulative sample size.

Other studies showed comparable performance parameters. Gao *et al*.[38] calculated that the sensitivity and specificity of ACR-TIRADS were around 80%. Ha *et al*. indicated that ACR-TIRADS had a sensitivity of 80% and a specificity of 69% in the differential diagnosis of thyroid nodules, confirming that ACR-TIRADS guidelines were significantly less sensitive and had a higher specificity than ATA guidelines.[39]

In our study, the ATA and ACR-TIRADS RSSs do not have a PPV to select thyroid malignancy, while the NPV to predict benign nodules or rule out malignancy in patients with higher Bethesda cytopathologic categories was around 95% for the ACR-TIRADS, and 97% for the ATA RSS. These results were similar to other studies.[28,40-43] Such results may be explained given that cytologically indeterminate nodules with TIRADS 2 or 3 and ATA very low or low risk consider similar benign patterns like spongiform nodules, isoechoic or hyperechoic solid nodules without signs of high suspicion. Both systems are feasible for clinical application and have an excellent negative predictive value allowing the selection of patients to FNAC-US. However, they may result in higher percentages of indeterminate cytology as Bethesda 2 and 3.

Contrary to our results, which showed higher diagnostic performances for ATA RSS compared to ACR-TIRADS, with lower FNAC referral using ACR-TIRADS, Ha *et al*. and Ruan *et al*. showed that the diagnostic performance of ACR-TIRADS was greater than that of ATA guidelines. The biopsy yield of malignant ACR-TIRADS was greater than that of ATA guidelines, and the unnecessary biopsy rate of ACR-TIRADS was lower than that of ATA guidelines. However, the biopsy rate of malignancy of ACR-TIRADS was lower than that of ATA guidelines.[33,43] The greater diagnostic performance of ACR-TIRADS might be attributed to the allocation of points for each US characteristic, reflecting the likelihood of malignancy.

Our results provided weak but significant evidence about the preference of ATA over the ACR-TIRADS RSSs, given that the two raters agreed to FNAC on about half of the referred TNs. Similarly, in Yoon *et al*., no significant differences were seen between the correlation coefficients between the two raters.[34] The use of head-to-head comparison to assess the performance of two US RSSs within the same population should not be impaired by recruitment, malignancy prevalence, or operator biases. Our results cannot be explained by

the size cut-offs for FNAC in intermediate and high risk-nodules, given that it is similar to the other US RSSs.

The major strength of this study is the largest real-world data about the risk stratification TNs in Southern Iraq and one of the few studies comparing directly between ATA and ACR-TIRAD raters in Middle East population. Secondly, the nodules for biopsy were examined in real-time ultrasonography before the sample was obtained. We are confident that the nodules being biopsied are sonographically classified compared to retrospective studies.

While the main limitation in this study is that the single-center nature of the study may limit generalizability and we did not include any TN in the subcentimetric ranges. However, in the guidelines, suspicious nodules could be as small as 5 mm. We used 10 mm as the minimum size threshold.

Conclusions

ATA and the ACR-TIRADS lexicons showed similar low sensitivity, specificity, prediction, and accuracy in detecting malignant thyroid nodules, although relatively in favor of ATA RSSs. There is a weak but significant agreement between the two raters, *i.e.*, we can miss cases by both raters. ACR-TIRADS can stratify the risk in the unclassified cases by ATA RSS. The two RSSs provide limited guidance to clinicians regarding follow-up plans and therapeutic options for TNs from different categories.

Updated and improved measures should be taken to revise the size and location criteria and the US point scoring system to increase the standardization of diagnosis and treatment of TNs and merge the characteristics of the different systems into an ultimate unified, cost-effective international standardized system without jeopardizing the detection of clinically significant malignancies.

Prospective and longitudinal follow-up studies are highly recommended to evaluate the immediately suspicious nodules and refine the noninvasive risk stratification of TNs in a way that emphasizes clinically important cancer. Replacement of 2015 revised ATA guidelines by ACR-TIRADS RSS is not justified now. Future research perspectives in future unified RSS may incorporate the number of TNs, age, gender, time since the discovery, and biochemical investigations like TSH and calcitonin.

References

1. Gharib H, Papini E, Garber JR, et al; AACE/ACE/AME Task Force on Thyroid Nodules. 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 2016;22:622-39.
2. Grani G, Lamartina L, Cantisani V, et al Interobserver agreement of various thyroid imaging reporting and data systems. Endocr Connect 2018;7:1-7.
3. Acar T, Ozbek SS, Acar S. Incidentally discovered thyroid nodules: frequency in an adult population during Doppler ultrasonographic evaluation of cervical vessels. Endocrine 2014;45:73-8.
4. Guth S, Theune U, Aberle J, et al. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. Eur J Clin Invest 2009;39:699-706.
5. American Thyroid Association (ATA) guidelines taskforce on thyroid nodules and differentiated thyroid cancer, Cooper DS, Doherty GM, Haugen BR, et al. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. Thyroid 2009;19:1167-214.
6. Cibas ES, Ali SZ. The 2017 Bethesda System for reporting thyroid cytopathology. Thyroid 2017;27:1341-6.
7. Haugen BR, Alexander EK, Bible KC, 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;26:1-133.
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;260:892-9.
9. Tessler FN, Middleton WD, Grant EG, et al, ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. J Am Coll Radiol 2017;14:587-95.
10. Smith-Bindman R, Lebda P, Feldstein VA, et al. Risk of thyroid cancer based on thyroid ultrasound imaging characteristics: results of a population-based study. JAMA Intern Med 2013;173:1788-96.

11. Pandya A, Caoili EM, Jawad-Makki F, et al. Retrospective cohort study of 1947 thyroid nodules: a comparison of the 2017 American College of Radiology TI-RADS and the 2015 American Thyroid Association Classifications. AJR Am J Roentgenol 2020;214:900-6.
12. Lim-Dunham JE, Toslak IE, Reiter MP, Martin B. Assessment of the American College of Radiology Thyroid Imaging Reporting and Data System for thyroid nodule malignancy risk stratification in a pediatric population. AJR Am J Roentgenol 2019;212:188-94.
13. Ahn HS, Kim HJ, Welch HG. Korea's thyroid-cancer "epidemic"--screening and overdiagnosis. N Engl J Med 2014;371:1765-7.
14. Charan J, Biswas T. How to calculate sample size for different study designs in medical research? Indian J Psychol Med 2013;35:121-6.
15. Hussein IH, Altemimi MT, Alidrisi HA, et al. The Performance of the American Thyroid Association (ATA) and American College of Radiology (ACR-TIRAD) thyroid nodule risk-stratification systems in determining high-risk nodules, and the correlation of site, size, and autoimmunity with high-risk features. Indian J Endocrinol Metab 2024;28:622-8.
16. Mansour AA, Abdullah Almomin AMS, et al. SUN-418 patterns of thyroid disease in Basrah, Iraq. Retrospective study. J Endocr Soc 2020;4:SUN-418.
17. Kargar S, Tabatabaei SM, Okati-Aliabad H, Izadi Rad H. Prevalence of thyroid dysfunction disorders among adult populations in the Middle–East: a systematic review and meta-analysis. Open Public Health J 2024;17:e187494452411270.
18. Ahmadi S, Oyekunle T, Jiang X', et al. A direct comparison of the ATA and TI-RADS ultrasound scoring systems. Endocr Pract 2019;25:413-22.
19. 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 2019;104:95-102.
20. Merhav G, Zolotov S, Mahagneh A, et al. Validation of TIRADS ACR risk assessment of thyroid nodules in comparison to the ATA Guidelines. J Clin Imaging Sci 2021;11:37.
21. Koseoglu Atilla FD, Ozgen Saydam B, Erarslan NA, et al. Does the ACR TI-RADS scoring allow us to safely avoid unnecessary thyroid biopsy? single center analysis in a large cohort. Endocrine 2018;61:398-402.

22. Gao L, Xi X, Jiang Y, et al. Comparison among TIRADS (ACR TI-RADS and KWAK- TI-RADS) and 2015 ATA guidelines in the diagnostic efficiency of thyroid nodules. Endocrine 2019;64:90-6.
23. 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;210:1148-54.
24. Kim PH, Suh CH, Baek JH, et al. Unnecessary thyroid nodule biopsy rates under four ultrasound risk stratification systems: a systematic review and meta-analysis. Eur Radiol 2021;31:2877-85.
25. Wu XL, Du JR, Wang H, et al. Comparison and preliminary discussion of the reasons for the differences in diagnostic performance and unnecessary FNAC biopsies between the ACR TIRADS and 2015 ATA guidelines. Endocrine 2019;65:121-31.
26. Chen H, Ye J, Song J, et al. Comparison of different ultrasound classification systems of thyroid nodules for identifying malignant potential: a cross-sectional study. Clinics (Sao Paulo) 2021;76:e2126.
27. Zhang WB, Xu HX, Zhang YF, et al. Comparisons of ACR TI-RADS, ATA guidelines, Kwak TI-RADS, and KTA/KSThR guidelines in malignancy risk stratification of thyroid nodules. Clin Hemorheol Microcirc 2020;75:219-32.
28. Botha M, Kisansa M, Greeff W. American College of Radiology Thyroid Imaging Reporting and Data System standardises reporting of thyroid ultrasounds. SA J Radiol 2020;24:1804.
29. Hoang JK, Middleton WD, Langer JE, et al. Comparison of thyroid risk categorization systems and fine-needle aspiration recommendations in a multi-institutional thyroid ultrasound registry. J Am Coll Radiol 2021;18:1605-13.
30. Mussoi de Macedo BM, Izquierdo RF, Golbert L, Souza Meyer EL. Reliability of Thyroid Imaging Reporting and Data System (TI-RADS), and ultrasonographic classification of the American Thyroid Association (ATA) in differentiating benign from malignant thyroid nodules. Arch Endocrinol Metab 2018;62:131-8.
31. Rosario PW, da Silva AL, Nunes MS, et al. Risk of malignancy in 1502 solid thyroid nodules >1 cm using the new ultrasonographic classification of the American Thyroid Association. Endocrine 2017;56:442-5.
32. Karaagac M, Sarigoz T, Ertan T, Topuz O. Evaluation of the Bethesda system and the ACR TIRADS in an endemic goiter region. Endocr Res 2020;45:226-32.

33. Ha SM, Baek JH, Na DG, et al. Diagnostic performance of practice guidelines for thyroid nodules: thyroid nodule size versus biopsy rates. Radiology 2019;291:92-9.
34. Yoon JH, Lee HS, Kim EK, et al. Malignancy risk stratification of thyroid nodules: comparison between the Thyroid Imaging Reporting and Data System and the 2014 American Thyroid Association management guidelines. Radiology 2016;278:917-24.
35. Ha EJ, Na DG, Baek JH, et al. US fine-needle aspiration biopsy for thyroid malignancy: diagnostic performance of seven society guidelines applied to 2000 thyroid nodules. Radiology 2018;287:893-900.
36. Middleton WD, Teefey SA, Tessler FN, et al. Analysis of malignant thyroid nodules that do not meet ACR TI-RADS criteria for fine-needle aspiration. AJR Am J Roentgenol 2021;216:471-8.
37. Castellana M, Castellana C, Treglia G, et al. Performance of five ultrasound risk stratification systems in selecting thyroid nodules for FNAC. J Clin Endocrinol Metab 2020;105:dgz170.
38. Gao L, Xi X, Wang J, et al. Ultrasound risk evaluation of thyroid nodules that are "unspecified" in the 2015 American Thyroid Association management guidelines: A retrospective study. Medicine (Baltimore) 2018;97:e13914.
39. Ha EJ, Na DG, Moon WJ, et al. Diagnostic performance of ultrasound-based risk-stratification systems for thyroid nodules: comparison of the 2015 American Thyroid Association guidelines with the 2016 Korean Thyroid Association/Korean Society of Thyroid Radiology and 2017 American College of Radiology guidelines. Thyroid. 2018;28:1532-37.
40. Yang W, Fananapazir G, LaRoy J, et al. Can the American Thyroid Association, K-TIRADS and ACR-TIRADS ultrasound classification systems be used to predict malignancy in Bethesda category IV nodules? Endocr Pract 2020;26:945-52.
41. Grani G, Lamartina L, Ascoli V, et al. Ultrasonography scoring systems can rule out malignancy in cytologically indeterminate thyroid nodules. Endocrine 2017;57:256-61.
42. Goundan PN, Mamou J, Rohrbach D, et al, A preliminary study of quantitative ultrasound for cancer-risk assessment of thyroid nodules. Front Endocrinol (Lausanne) 2021;12:627698.
43. Ruan JL, Yang HY, Liu RB, et al. Fine needle aspiration biopsy indications for thyroid nodules: compare a point-based risk stratification system with a pattern-based risk stratification system. Eur Radiol 2019;29:4871-8.

Table 1. Baseline characteristics of the 1,007 enrolled individuals with a thyroid nodule.

Parameters		Value
Gender n (%)	Women	858 (85.2)
	Men	149 (14.8)
Age years mean ± Standard Deviation (SD)	Range (14 - 90)	45.88 ± 13.97
Body Mass Index (BMI) kg/m^2 mean ± SD	Range (13.39 - 64.52)	31.77 ± 6.66
Height cm mean ± SD	Range (139 - 194)	158.74 ± 7.83
Weight kg mean ± SD	Range (33 - 155)	79.92 ± 16.85
Smoking n (%)	Non-smoker	937 (93.0)
	Smoker	51 (5.1)
	Ex-smoker	19 (1.9)
Occupation n (%)	Employee	153 (15.2)
	Free worker	49 (4.9)
	Housewife	741 (73.6)
	Retired	36 (3.5)
	Student	28 (2.8)
Educational level n (%)	Illiterate	243 (24.1)
	Primary School	407 (40.4)
	Secondary School	139 (13.8)
	Preparatory School	77 (7.6)
	Diploma	55 (5.5)
	College	81 (8.0)
	Master's degree	3 (0.3)
	Doctorate degree	1 (0.1)
Comorbidities	Diabetes mellitus	227 (22.5)
	Hypertension	303 (30.1)
	Ischemic heart disease	28 (2.8)
Thyroid status n (%)	Euthyroid	820 (71.0)
	Subclinical hypothyroidism	24 (2.1)
	Subclinical hyperthyroidism	28 (2.4)
	Overt hypothyroidism	176 (15.2)
	Overt hyperthyroidism	107 (9.3)

; BMI, Body Mass Index; SD, Standard Deviation.

Table 2. Thyroid nodule basal characteristics in 1,155 thyroid nodules for the 1,007 individuals.

Parameters		**Value**
TSH mIU/l Median ± SE	Range (0.001 – 100.00)	1.50 ± 0.23
Thyroid nodules (TNs) mean ± SD cm	Length range (7.00 – 82.00)	19.58 ± 10.45
	Width range (7.00 – 50.00)	15.28 ± 7.81
Nodular status	Single	527 (45.6)
	Two nodules only	103 (8.9)
	Multinodular goitre	525 (45.5)
Position of the thyroid nodules	Right	593 (51.34)
	Left	491 (42.51)
	Isthmus	71 (6.15)
Need for FNAC according to ATA categories	Yes, need FNAC	631 (54.6)
Need for FNAC according to ACR-TIRADS categories	Yes, need FNAC	515 (44.6)
Echogenicity n (%)	Anechoic	56 (4.8)
	Isoechoic	911 (78.9)
	Hypoechoic	98 (8.5)
	Hyperechoeic	90 (7.8)
Calcification n (%)	none or large Comet-tail artifact	907 (78.5)
	Macrocalcification	107 (9.3)
	Peripheral calcification rim	18 (1.6)
	Punctate	123 (10.6)
Peripheral margin n (%)	Smooth	1134 (98.2)
	Ill-defined	13 (1.1)
	Lobulated or irregular	8 (0.7)
TN vascularity n (%)	Intranodular	48 (4.2)
	Peripheral	72 (6.2)
	Not vascular	1035 (89.6)
Lymphadenopathy n (%)	Bilateral	20 (1.7)

	Ipsilateral to thyroid nodule	24 (2.1)
	Contralateral	3 (0.3)
	No nodule	1108 (95.9)
Lymph nodes composition (n=47) n (%)	Inflammatory	38 (3.3)
	Suspicious	9 (0.8)
	Normal	1108 (95.9)
Year of FNAC (n=536)	2019	46 (8.58)
	2020	217 (40.49)
	2021	273 (50.93)
Type of previous thyroid surgery (n=80)	Total thyroidectomy	9 (11.3)
	Subtotal thyroidectomy	54 (67.5)
	Lobectomy +/- isthmectomy	14 (17.5)
	Nodulectomy	3 (3.8)

ACR-TIRADS, American College of Radiology-Thyroid Imaging Reporting and Data System; ATA, American Thyroid Association; FNAC, Fine Needle Aspiration Cytology; SD, Standard Deviation; SE, Standard Error of the Means; TSH, Thyrotropin; TNs, Thyroid Nodules.

Table 3. Risk of malignancy according to different raters in 5 aspiration cytology.

Classification system	All Thyroid Nodule (TN) n (%)	Fine Needle Aspiration Cytology (FNAC)	Risk of malignancy %	Nodule's highest dimension	
		Suspicious for malignancy (n=38)	Benign (n=498)		
American College of Radiology-Thyroid Imaging Reporting and Data System Risk Stratification System **(ACR-TIRADS RSS)**					
TIRAD-2 (n=61)	350 (30.3)	4	57	6.56	28.88 ± 9.64
TIRAD-3 (n=208)	435 (37.7)	11	197	5.29	25.20 ± 9.90
TIRAD-4 (n=188)	234 (20.3)	14	174	7.45	23.86 ± 8.44
TIRAD-5 (n=79)	80 (6.9)	9	70	11.39	21.91 ± 4.94
American Thyroid Association Malignancy Risk Stratification System **(ATA Malignancy RSS)**					
Very low suspicion <3% (n=63)	310 (26.8)	3	60	4.76	27.92 ± 9.32
Low suspicion 5 -10% (n=222)	527	6	216	2.70	26.68

	(45.6)				± 10.76
Intermediate suspicion 10 -20% (n=182)	209 (18.1)	5	177	2.75	22.27 ± 5.17
High suspicion (n=48)	48 (4.2)	21	27	43.75	20.38 ± 6.08
Unclassified (n=21)	21 (1.8)	3	18	14.29	24.14 ± 9.73
All TNs with FNACs					

Abbreviations: ACR-TIRADS RSS, American College of Radiology-Thyroid Imaging Reporting and Data System Risk Stratification System; ATA Malignancy RSS, American Thyroid Association Malignancy Risk Stratification System; FNAC, Fine Needle Aspiration Cytology; TN, Thyroid Nodule

Table 4. Fine Needle Aspiration Cytology (FNAC) outcomes and diagnostic performance of American College of Radiology–Thyroid Imaging Reporting and Data System (ACR-TIRADS) and American Thyroid Association (ATA) risk stratification systems (RSS) (n = 515 thyroid nodules).

Diagnostic performance metrics		ACR-TIRADS RSS	ATA malignancy RSS
Suspicious for malignancy (n)	Total	35	35
	True positives	21	26
	False negatives	14	9

Benign (n)	Total	480	480
	False positives	228	204
	True negatives	252	276
Sensitivity (%)		60.00	74.29
Specificity (%)		52.50	57.50
PPV (%)		8.43	11.30
NPV (%)		94.74	96.84
Positive Likelihood Ratio		1.26	1.75
Negative Likelihood Ratio		0.76	0.45
Accuracy (%)		53.01	58.64

ACR-TIRADS, American College of Radiology–Thyroid Imaging Reporting and Data System; ATA, American Thyroid Association; FNAC, Fine Needle Aspiration Cytology; RSS, Risk Stratification System; PPV, Positive Predictive Value; NPV, Negative Predictive Value.

Table 5. Estimated need for Fine Needle Aspiration Cytology (FNAC) according to the predetermined scores for each rater in 515 thyroid nodules; the 21 unclassified nodules by either rater were excluded.

Raters	**Parameters**	**Bethesda 1**	**Bethesda 2**	**Bethesda 3**	**Bethesda 4**	**Bethesda 5 (n=25)**	**Total (n=515**

		(n=137) (27.1%)	(n=318) (60.8%)	(n=25) (5%)	(n=10) (1.9%)	(5.2%)	)
ACR-TIRADS RSS	**Need FNAC**	105	262	16	5	21	409
	Follow up	32	56	9	5	4	106
ATA RSS	**Need FNAC**	136	314	24	10	25	509
	Follow up	1	4	1	0	0	6

ACR-TIRADS, American College of Radiology–Thyroid Imaging Reporting and Data System; ATA, American Thyroid Association; FNAC, Fine Needle Aspiration Cytology; RSS, Risk Stratification System

Contributions: DKA and AGM conceived the present idea of the study, developed the theory. DKA, AGM, SAO, MTA collected the data and performed the computations. MTA, and SAK conceived and designed the analysis, wrote the first draft of the manuscript. DKA, IHT and SAO investigated and supervised the findings of this work, Discuss the results. All authors commented on and approved the final manuscript.

Conflict of interest: the authors have no conflict of interest to declare.

Ethics approval: ethical approval was obtained from the ethical committee of the tertiary endocrine center and the Institutional Review Board (IRB) (3T/2/2019) for the study on February 2019, according to the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Informed consent: informed consent was obtained from all patients clearly along the whole study timeline.

Availability of data and material: the datasets generated and/or analyzed during the current study are not publicly available due to confidentiality but are available on reasonable request from the corresponding author.

Funding: this research did not receive any specific grant from funding agencies in the public, commercial, or non- profit sectors.

Acknowledgments: all authors are grateful to all participants for their cooperation and Thi-Qar Specialized Diabetes, Endocrine Center medical staff for approving this work. We also extend our appreciation to the efforts of the endocrinologist and radiologist in our center in helping us to perform the best imaging to make the work successful.

Online supplementary materials

Table 1S: Kappa Statistics and Cohen's Kappa test for 611 thyroid nodules that underwent fine needle aspiration cytology. The two raters only show agreement in 50.10% of the cases, i.e., 258 out of 515 thyroid nodules that underwent fine needle aspiration cytology. Cohen's kappa value equals 0.265±0.032, with $p<0.001$. McNemar-Bowker Test equals 24.274, with $p<0.001$.

Table 2S. Description of nine highly suspicious cases where the raters were inconclusive regarding referral to fine needle aspiration cytology. The mean thyrotropin (TSH) for these nine cases was (3.05±2.18 mIU/L). All cases were euthyroid at the time of the Fine Needle Aspiration Cytology (FNAC), with no lymphadenopathy. All nine thyroid nodules had smooth margins.