Diagnostic Accuracy of Five Different Classification Systems for Thyroid Nodules

A Prospective, Comparative Study

Umut Percem Orhan Soylemez, MD, *Nesrin Gunduz, MD*

Supplemental material online at jultrasoundmed.org

Objective—To compare the diagnostic performance of five different thyroid ultrasound classification systems, and determine which system is optimal for evaluating thyroid nodules and reducing the unnecessary biopsy rate.

Methods—In this prospective study, 1,010 nodules referred for biopsy during a 2-year period were classified using five classification systems: the Kwak Thyroid Imaging Reporting and Data System (Kwak TI-RADS), the European TI-RADS (EU TI-RADS, the Korean TI-RADS (K TI-RADS), the American College of Radiology TI-RADS (ACR TI-RADS), and the American Thyroid Association (ATA) classification. After fine needle aspiration biopsy, all classifications were compared for all nodules and also particularly for nodules sized 1–3 cm. Sensitivity, specificity, and interobserver agreement were evaluated for each classification system.

Results—Of the 939 nodules (after exclusion of Bethesda 3 nodules) finally classified according to the surgical histopathology and cytology results, 73 (7.8%) were malignant and 866 nodules were benign (92.2%). The sensitivity was highest (94.5%) for the ACR TI-RADS and lowest for the Kwak TI-RADS (69%). After exclusion of small (<1 cm) and large nodules (>3 cm); while sensitivity was highest for ATA (97.8%), ACR TI-RADS was the second best classification (91.3%). There was substantial agreement among all classification systems except the Kwak TI-RADS (fair agreement).

Conclusions—The ACR TI-RADS was the most sensitive ultrasound risk stratification system for all nodules, while the Kwak TI-RADS was the most specific, ie, the most capable of excluding benign nodules based on the combined cytological and histopathological results. ATA and ACR-TIRADS were the most sensitive classification systems for nodules 1 to 3 cm in size. The ACR TI-RADS had higher sensitivity than the Bethesda classification system when compared according to the histopathological results.

Key Words—fine-needle aspiration biopsy; thyroid nodule; thyroid imaging reporting and data system; ultrasonography

Received May 6, 2021, from the Department of Radiology, Goztepe City Hospital, Istanbul Medeniyet University, Istanbul, Turkey (U.P.O.S.); and Department of Radiology, Goztepe City Hospital, Faculty of Medicine, Istanbul Medeniyet University, Istanbul, Turkey (N.G.). Manuscript accepted for publication July 7, 2021.

The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Address correspondence to Umut Perçem Orhan Söylemez, MD, Istanbul Medeniyet University, Goztepe City Hospital, Department of Radiology, Istanbul, Turkey.

E-mail: umutpercem@gmail.com

Abbreviations

ATA, American Thyroid Association; FNAB, fine needle aspiration biopsy; PET-CT, positron emission tomography-computed tomography; US, ultrasonography

doi:10.1002/jum.15802

The incidence of thyroid nodules has increased with more widespread use of ultrasonography (US). In fact, there has been an increase in the incidence of all cancers due to routine screening of high-risk groups.[1] Although US reportedly detects thyroid nodules in 68% of individuals, very few are malignant (1.6–12%).[1,2] This prompted clinicians and radiologists to develop more accurate classification systems for nodules.

Early and correct diagnosis of malignant lesions reduces the burden on healthcare systems by preventing unnecessary and continuous follow-up of patients with benign lesions.

US is the most appropriate technique to evaluate thyroid nodules. Several US classification systems have been proposed, developed and, ultimately, used or abandoned. However, US is a subjective imaging modality with low inter and intra-observer reliability, and must be performed by radiologists with a good knowledge of lesion characteristics.[2,3] The echogenicity, shape, margins, and calcifications of nodules are all considered by classification systems.[4] Although the features of high-risk nodules are defined similarly by different classification systems, there is no consensus on the features of intermediate-risk nodules. Also, the importance of nodule size in predicting the need for biopsy is controversial. Performing fine needle aspiration biopsy (FNAB) on all thyroid nodules is neither cost-effective nor necessary. FNAB is minimally invasive procedure and complications are very rare; however, it cannot be used for early diagnosis and follow-up.[5,6]

All of the classification systems for thyroid nodules have been subject to revisions. To guide the management of thyroid nodules, classification systems must have high sensitivity, specificity, reproducibility, and interobserver reliability. Several studies have evaluated the accuracy of different classification systems, but evaluations of newer versions of the systems are needed. Three different classification systems were compared recently, with the American College of Radiology Thyroid Imaging Reporting and Data System (ACR TI-RADS), which was revised in 2017, showing the best diagnostic performance.[7] However, in a study comparing three classification systems, the Kwak TI-RADS had the highest positive predictive value and the ACR TI-RADS had the highest negative predictive value.[8] In this prospective study, we applied five different ultrasound classification systems to a large series of patients. Diagnoses were verified based on the results of cytology and surgical histopathology. We compared the sensitivity, specificity, and inter-observer reliability of the classification systems.

Material and Methods

Institutional review board approval was obtained for this single-center prospective study (Istanbul Medeniyet University Ethics Committee; ID; 2019/0186). Study was conducted between December 2018 and December 2020 and performed in accordance with the revised Helsinki Declaration. US was performed on all patients referred to our Interventional Radiology Department for thyroid biopsy. The Vision Preirus system (Hitachi Medical Corp.) and Aplio 500 device (Toshiba Medical Systems, Co., Ltd.) were used for all biopsy and US procedures.

A total of 1313 FNAB procedures were performed on 1516 patients referred to our clinic for biopsy. A total of 977 patients (mean age, 52 ± 13 years), including 801 (82%) females and 176 (18%) males with 1010 nodules, were included in the study. All nodules were evaluated before the biopsy procedure for the following features: echogenicity (isoechoic/hyperechoic/hypoechoic/markedly hypoechoic/anechoic), margins (well defined/ill-defined/irregular/extrathyroidal extension), three-dimensional size (in mm), localization (right lobe/left lobe/isthmus/anterior/posterior/cranial/caudal/central), shape (wider than tall/taller than wide/equal), internal and peripheral echogenicity (colloid crystals/macrocalcifications/punctate echogenicity/peripheral calcification), echostructure (solid/cystic/predominantly solid/predominantly cystic/mixed), vascularization (none/peripheral/intranodular/marked intranodular), and thyroid parenchyma features (normal, heterogeneous, chronic thyroiditis, multinodular goiter).[4,9-11]

Nodules with equal echogenicity to the thyroid gland were defined as isoechoic, those with higher echogenicity than the thyroid gland were defined as hyperechoic, and those with lower echogenicity were defined as hypoechoic. Nodules with less echogenicity than strap muscle were defined as markedly hypoechoic. Echogenicities ≤1 mm without a reverberation artifact were defined as microcalcifications (punctate echogenic foci), whereas those >1 mm causing an acoustic shadow were macrocalcifications.[12]

Using the set of features specified above, all thyroid nodules were classified using the European TI-RADS (EU TI-RADS),[13] Korean TI-RADS (K TI-RADS),[10] ACR TI-RADS,[11] Kwak TI-RADS,[9] and American Thyroid Association (ATA) system.[14] The criteria for high-risk malignancy and benign nodules, and the biopsy recommendations, are summarized in Table 1. Informed consent forms were signed by all patients before FNAB. Data including images and pathology results were collected.

Table 1. Nodule Classifications and Biopsy Recommendations of the Classification Systems

	Kwak TIRADS	ATA	EU-TIRADS	K-TIRADS	ACR TIRADS
	1-negative **2**-benign **3**-probably benign **4A**-low suspicion **4B**-intermediate suspicion **4C**-moderate concern but not classic for malignancy **5**-highly suggestive for malignancy	-Benign -Very low risk -Low risk -İntermediate risk -High risk	**1**-negative **2**-benign **3**-low risk **4**-intermediate risk **5**-high risk	**1**-negative **2**-benign **3**-low suspicion **4**-intermediate suspicion **5**-high suspicion	**1**-0 point-benign **2**-2 point-not suspicious **3**-3 points-mildly suspicious **4**-4 to 6 points-moderately suspicious **5**-≥7 points-highly suspicious
Benign		-Spongioform -Purely cystic	-Pure cystic -Entirely spongioform	-Pure cyst -Partially cystic with comet tail artifact -Spongioform	-Cystic or almost completely cystic -Spongioform -Wider than tall shape -With comet-tail artifact -Smooth margin -Anechoic
High risk	-hypoechogenity/marked hypoechogenity, -microlobulation -irregular margins -microcalcification -taller than wide shape	-Irregular margins -Microcalcifications in hypoechoic nodule -Taller than wide shape -Rim calcifications with small extrusive soft tissue component -Evidence of extrathyroidal extension	-Irregular shape -Irregular margins -Microcalcifications -Markedly hypoechoic and solid	-Microcalcification -Non-parallel orientation -Spiculated/ microlobulated margin	-Very hypoechoic -Taller than wide -Extrathyroidal extension -Punctate echogenic foci
Biopsy	For category 4 and 5	**≥2 cm**-very low suspicion and spongioform **≥1.5 cm**-low suspicion **≥1 cm**-intermediate and high suspicion	**≥2 cm**-low risk **≥1.5 cm**-intermediate risk **≥1 cm**-high risk	**≥2 cm**-benign **≥1.5 cm**-low suspicion **≥1 cm**-intermediate and high suspicion	**≥2.5 cm-**TR3 **≥1.5 cm-**TR4 **≥1 cm**-TR5

FNAB was performed on nodules ≥5 mm in diameter with at least one sonographic high-risk factor (taller than wide shape, irregular margins, markedly hypoechoic, microcalcification), and on nodules >1 cm with no sonographic high-risk factors in high-risk patients (history of thyroid cancer in patient or first-degree relative, history of radiotherapy to the neck), fluorodeoxyglucose update on a positron emission tomography-computed tomography (PET-CT) scan, multiple endocrine neoplasia 2/familial medullary thyroid cancer,[15] and an increase in volume of ≥50%/increase in at least two dimensions of ≥20% between two screenings. FNAB was performed in all nodules >1.5 cm in diameter.[15] Biopsies were not performed in patients with normal thyroid scans and nodules <5 mm, or in those with nodules <1 cm and no risk factors. Patients aged <18 years were not included in the study. Nodules with non-diagnostic (Bethesda 1) cytology results were also excluded.

US-guided FNAB was performed by a single radiologist with 10 years of experience using a 22 to 27-gauge needle and 5-mL syringe. If nodule vascularization was

Table 2. Diagnostic Performance of the Classification Systems, ie, Ability to Distinguish Malignant and Benign Nodules

Classification	Number of Nodules Recommended to Underwent the FNAB According to the Classifications ($n = 1010$)	Diagnostic Performance According to Cytology and Histopathology Results	
		Sensitivity % (CI)	Specificity % (CI)
ACR-TIRADS	586 (58%)	94.5 (86.5–98.4)	45.7 (42.3–49.1)
K-TIRADS	364 (36%)	64.3 (52.3–75.2)	67 (63.8–70.1)
EU-TIRADS	584 (57.8%)	79.4 (68.3–88)	43.7 (40.4–47.1)
ATA-TIRADS	776 (76.8%)	83.5 (73–91.2)	23.7 (20.9–26.7)
Kwak-TIRADS	346 (34.3%)	63 (50.9–74.3)	69.1 (65.9–72.2)

FNAB, fine needle aspiration biopsy; CI, confidence interval.

excessive, a thinner needle was chosen. Two to four passes were made on each nodule. In the cytological evaluation, all nodules were classified using the Bethesda classification.[16] The Bethesda system is a universal reporting system used in thyroid cytology. System evaluates the nodules in 6 categories such as 1: nondiagnostic-unsatisfactory, 2: benign, 3: atypia of undetermined significance or follicular lesion of undetermined significance, 4: follicular neoplasm or suspicious for a follicular neoplasm, 5: suspicious for malignancy and 6: malignant.[16] Bethesda 2 nodules were classified as negative for malignancy, whereas Bethesda 4 to 6 nodules were classified as positive. For thyroid nodules with an uncertain outcome (Bethesda 3), a repeat biopsy was performed, typically within 3 months. Bethesda 3 nodules that could not be definitively classified as malignant or benign after re-biopsy were excluded.

The histopathology results (of 68 patients undergoing thyroidectomy) were considered the gold standard for diagnosis. Lesions were classified as benign or malignant according to the surgical histopathology results (where available) and cytology results. After including the nodules in all sizes, a second group was described for the nodules 1 to 3 cm in size. Diagnostic performance of all classifications was also evaluated for these nodules. A second radiologist with 15 years of experience, who was blinded to the patient data, classified nodules based on the video loops recorded by the first observer, to evaluate the interobserver reliability of the classification systems. The final decisions on nodule classifications were made after a consensus meeting. Each classification system was compared with the Bethesda classification in terms of the cytopathology and surgical histopathology.

Statistical Analysis

The statistical analysis was performed using SPSS 19.0 software (IBM Corp.). The normality of the continuous variables was analyzed by the Shapiro–Wilk test. Descriptive statistics are reported as mean with standard deviation for continuous normally distributed variables, and as the median with interquartile range (25th–75th percentile) for non-normally distributed variables. Independent groups were compared using Pearson's chi-square test in cases where all assumptions were met. Otherwise, Fisher's exact test was used. Interobserver agreement was assessed using the Kappa statistic. The diagnostic performance of each classification system was analyzed in terms of sensitivity and specificity, and 95% confidence intervals were calculated. A P-value <.05 was considered significant.

Results

The sampled nodules were mostly in the right lobe ($n = 571$ (56.5%)). Nodules were most often in the posterior part of the lobe ($n = 350$ (34.7%)), followed by central ($n = 290$ (28.7%)), anterior ($n = 283$ (28%)), and whole-lobe ($n = 86$ (8.5%)) locations. One nodule was ectopically located in the pretracheal area in a patient with thyroidectomy. Sixty-four patients had malignant lesions according to the Bethesda classification (Bethesda 4–6). A surgical thyroidectomy had been performed in only 68 (6.7%) patients.

Ultrasonographic Features of the Nodules

Mixed solid and cystic components were present in 673 nodules (66.6%), of which 623 were predominantly solid cysts. In total, 479 nodules were isoechoic (47.4%). A wider than tall pattern was detected in

Table 3. Comparison of the Various Ultrasonographic and Bethesda Classification Systems in Terms of Their Ability to Predict Malignancy According to Surgical Histopathology Results

Classification	Confirmed Malignancy After Surgery ($n = 38$)	Validity of Classifications for Predicting Malignancy	
		Sensitivity % (CI)	Specificity % (CI)
ACR-TIRADS	36 (52.9%)	94.7 (82.2–99.4)	80 (61.4–92.3)
K-TIRADS	25 (36.8%)	65.8 (6.6–9.4)	43.3 (91.9–99.6)
EU-TIRADS	31 (45.6%)	81.6 (65.7–92.3)	16.7 (5.6–34.7)
ATA-TIRADS	32 (47.1%)	84.2 (68.8–94)	6.7 (0.8–22.1)
Kwak-TIRADS	23 (33.8%)	60.5 (43.4–76)	53.3 (34.3–71.7)
Bethesda	29 (42.6%)	76.3 (59.8–88.6)	90 (73.5–97.9)

974 nodules (96.4%). Smooth borders were present in 907 (89.8%) nodules. Among 313 (31%) nodules with an echogenic focus in the solid component, isolated single round macrocalcification was the most common pattern ($n = 172$ (17%)), while punctate calcification was present in only 68 (6.7%) nodules. Underlying thyroidopathy was identified in 95 (9.4%) cases, including multinodular goiter ($n = 62$), chronic thyroiditis ($n = 27$) and a history of operations for papillary thyroid cancer ($n = 6$). Cervical non-inflammatory lymphadenopathy was present in four cases. A hypoechoic peripheral rim was present in 85 (8.4%) nodules. Color Doppler US revealed vascularity in 670 (66.3%) nodules. (The sonographic features are detailed in Table S1.)

Sonographic Classification of the Nodules

The inter-observer agreement was substantial for all classification systems (ACR TI-RADS, kappa = 0.62 ($P < .001$); K T-IRADS, kappa = 0.59 ($P < .001$); EU TI-RADS, kappa = 0.65 ($P < .001$); ATA classification, kappa = 0.64 ($P < .001$)) except for the Kwak TI-RADS (fair agreement; kappa = 0.40, $P < .001$). (The distribution of the number of nodules for the classification systems is summarized in Table S2.)

Correlations Between Cytology and Surgical Histopathology

Of the 1010 nodules that underwent FNAB, 861 (85.2%) were Bethesda 2, 85 (8.4%) were Bethesda 3, 13 (1.3%) were Bethesda 4, 14 (1.4%) were Bethesda 5, and 37 (3.7%) were Bethesda 6. Of the thyroidectomies performed in 68 patients, 22 (2.6%) were Bethesda 2, 14 (16.5%) were Bethesda 3, 4 (31%) were Bethesda 4, 7 (50%) were Bethesda 5, and 21 (57%) were Bethesda 6. The histopathological examination of the resected material revealed benign nodule status in 30 patients. In 38 cases the nodules were malignant, including 31 papillary cancers, 6 follicular cancers, and 1 Hurthle cell cancer.

Table 4. Diagnostic Performance of the Classification Systems for 1–3 cm Nodules

CLASSIFICATION	Diagnostic Performance According to Cytology and Histopathology Results	
	Sensitivity % (CI)	Specificity % (CI)
ACR-TIRADS	91.3 (79.2–97.5)	45.7 (42.3–49.1)
K-TIRADS	51.1 (47.2–55)	64.2 (60.4–67.9)
EU-TIRADS	73.9 (58.8–85.7)	49.1 (45.2–53)
ATA-TIRADS	97.8 (88.4–99.9)	23.2 (20–26.6)
Kwak-TIRADS	84.7 (71.1–93.6)	64.4 (60.6–68.1)

CI, confidence interval.

Comparison of the Classification Systems in Terms of Predicting Malignancy

The final malignant-benign categorization of nodules was established in 939 nodules according to the combined surgical histopathology and cytology results; 72 of 85 Bethesda 3 nodules were excluded from the study. In total, 73 (7.8%) nodules were malignant and 866 were benign (92.2%). The sensitivity was highest (94.52%) for the ACR TI-RADS and lowest for the Kwak TI-RADS (63%); specificity was highest for the Kwak TI-RADS (69.17%) and lowest for the ATA (23.79%) classification (Table 2).

Comparison Between the Ultrasonographic and Bethesda Classification Systems for Predicting Malignancy According to the Surgical Histopathology Results

Surgical histopathology results were available for 68 patients. The ACR TI-RADS showed the best

diagnostic performance (sensitivity and specificity = 94.7 and 80%, respectively). The other classification systems had moderate sensitivity and poor specificity. Moreover, when the diagnostic performance of the Bethesda classification was compared with that of the ACR TI-RADS according to the surgical histopathology results, the sensitivity of the latter system was found to be lower (76.3% versus 94.7%), although the specificity was higher (90% versus 80%) (Table 3).

Nodules 1 to 3 cm in Size

After exclusion of nodules <1 cm and >3 cm in size diagnostic performance of the classifications were changed particularly for ATA classification. Sensitivity ranking was ATA (97.3%), ACR TI-RADS (91.3%), Kwak TI-RADS (84.7%), EU TI-RADS (73.9%), and K TI-RADS (51.1%). Kwak TI-RADS had the highest specificity (64.4%) in this group as well (Table 4).

Discussion

The malignancy rate of thyroid nodules that underwent FNAB was reported to range between 1.6 and 12%.[2] In line with this, 7.8% of the nodules in our series were malignant. The main aim of diagnostic tests is to prevent misdiagnosis, while also reducing unnecessary interventions as much as possible (corresponding to the sensitivity and specificity, respectively). In our study, sensitivity was highest for the ACR TI-RADS, followed in descending order by the ATA classification, EU TI-RADS, Kwak TI-RADS, and K TI-RADS. The specificity was highest for the Kwak TI-RADS, followed by the K TI-RADS, ACR TI-RADS, EU TI-RADS, and ATA classification. Interobserver agreement was highest for the EU TI-RADS although no significant difference in reliability was observed between this classification system and the others. Substantial agreement was observed for all of the classification systems, except the Kwak TI-RADS (fair agreement between two observers). In line with our findings, previous studies reported substantial agreement among classification systems.[17] One study reported that the ACR TI-RADS was the most reliable based on the interobserver agreement.[18] The fair interobserver agreement for the Kwak TI-RADs in this study was expected, as this classification system includes subgroups for category 4 nodules.

Kwak et al. included nodules >1 cm in diameter in their study, and considered category 4–5 nodules suitable for biopsy. To classify a nodule as Kwak TI-RADS 5, it must meet all of the high-risk criteria of TI-RADS 4 nodules plus at least one malignancy criterion.[9] Although category 4 is divided into three subgroups (a–c), there is no difference in diagnostic approach among these nodules (Table 1). According to our results, the Kwak TI-RADS had the highest specificity (69.1%) but relatively low sensitivity (63%). In a recent study evaluating the diagnostic performance of three classification systems for nodules <1 cm, the authors reported 97.4% sensitivity and 55.1% specificity for the Kwak TI-RADS 4c and 5 categorizations. In our study, we included nodules of all sizes and malignancy grades (ie, Kwak TI-RADS 4a–c and 5). This may explain the relatively low sensitivity and high specificity for the Kwak TI-RADS in our study. Overall, Gao et al.[19] reported 89.4% sensitivity and 77.4% specificity for the Kwak TI-RADS for nodules ≤1 cm, >1 cm, and both sizes. We found lower sensitivity from those studies. The results did not change when we exclude very small and big nodules. Kwak TI-RADS was the most specific classification in our study while its sensitivity was low. Kwak TI-RADS may decrease the number of FNABs, but the probability of misdiagnosing cancer increases as a result. This classification also has low inter observer agreement probably originating from large number of subgroups.

The ATA classification system is based on clinical and radiological findings, including for US, scintigraphy, and PET/CT, as well as serum hormone levels. At least one malignancy criteria must be met for a solid nodule to be considered highly suspicious.[14] In line with previous studies,[20] the sensitivity of the ATA system was high (83.5%) in this study, while its specificity (23.7%) was relatively low. Biopsy is recommended for low-risk nodules >1.5 cm in diameter and very low-risk nodules >2 cm. In previous studies, the sensitivity of the ATA system ranged from 61 to 92%, with specificity of 44 to 73%.[20] In our study, specificity was below this range. Particularly for nodules 1 to 3 cm in size ATA increases the number of FNABs but as expected diagnoses more cancers.

The ACR TI-RADS is a partly quantitative classification system recommending biopsy or follow-up based on specific nodule features.[4] Our results revealed that the ACR TI-RADS was the most sensitive classification system for whole nodules and the second for nodules which are 1 to 3 cm in size. The rate of unnecessary biopsy in a meta-analysis was lowest for the ACR TI-RADS among risk classification systems, while the sensitivity was lower than that of the ATA classification and K TI-RADS (although specificity was higher).[20] In contrast, the specificity of the ACR TI-RADS was lower (45.7%) than that of the Kwak and K TI-RADS in our study. This may be due to differences in study design and nodule inclusion criteria among the eight studies included in the meta-analysis.

A nodule in the EU TI-RADS system meeting at least one criterion for high risk is categorized as TI-RADS 5. Biopsy is recommended for nodules >1 cm in diameter.[13] According to our results, although the numbers of biopsies indicated by the ACR TI-RADS and EU TI-RADS systems were similar (586 and 584, respectively) both the sensitivity and specificity of the EU TI-RADS were lower than those of the ACR TI-RADS. A recent study reported 97.4% sensitivity and 49.3% specificity for the EU TI-RADS 5 category.[8] The discrepancies between the classification systems may be due to different high risk and size criteria for biopsy In our study, specificity was slightly lower, but sensitivity was significantly lower. Yoon et al. reported a sensitivity of 85.7% and specificity of 39.3% for the EU TI-RADS, similar to our study.[21]

In the K TI-RADS, solid nodules with one or more malignant features are considered highly suspicious, and FNAB is recommended for those ≥1 cm in diameter.[10] Although this classification had the lowest sensitivity, it had the second highest specificity after the Kwak TI-RADS in the current study. Thus, this classification can be considered useful for excluding benign nodules and decreases the biopsy rate.[22] In a large-scale study of systematic reviews and meta-analyses, the pooled sensitivity of the K TI-RADS was 85% and the specificity was 47%.[23] In our study, specificity was good, but both sensitivity and specificity were lower than these previously reported values.

It is difficult to evaluate nodules smaller than 1 cm and larger than 3 cm in size. In recent studies; it was reported that the false negativity rate is high for nodules larger than 3 cm in size and re-biopsy or surgery is recommended for these nodules even if cytology is benign.[24] In the isolated assessment by excluding these nodules, ATA outperforms ACR in sensitivity ranking. In the order of specificity, there was no difference except the displacement of ACR and EU TIRADS. In studies comparing ATA and ACR TI-RADS, it was reported that diagnostic performances and inter observer agreements were similar for these classifications, ACR TI-RADS suggested less FNAB but missed more cancers. In current study we found similar results for ATA and ACR TI-RADS. Considering that the clinical significance of thyroid papillary carcinoma is still controversial[8] therefore cost effectivity evaluations should be made.[25,26]

Laboratory surveys, scintigraphy, and molecular analyses must be performed in association with US to develop a strategy for managing nodules. US is a subjective diagnostic modality, in that different study designs and set-ups make it difficult to compare results among studies. Nevertheless, an exact diagnosis can be made through US follow-up, repeated biopsy, or surgery.[1] To decrease the rate of unnecessary biopsies and increase the sensitivity of US, a combination of classification systems differing in sensitivity and specificity

Figure 1. Isoechoic solid nodule in a 78 years old woman; with cystic areas and macrocalcifications (arrow). This nodule was categorized as follows: ACR TI-RADS 4, ATA low risk, EU TI-RADS 3, K TI-RADS 3, and Kwak TI-RADS 3. Biopsy was recommended because of the nodule size. Cytology: benign colloidal nodule, Bethesda 2.

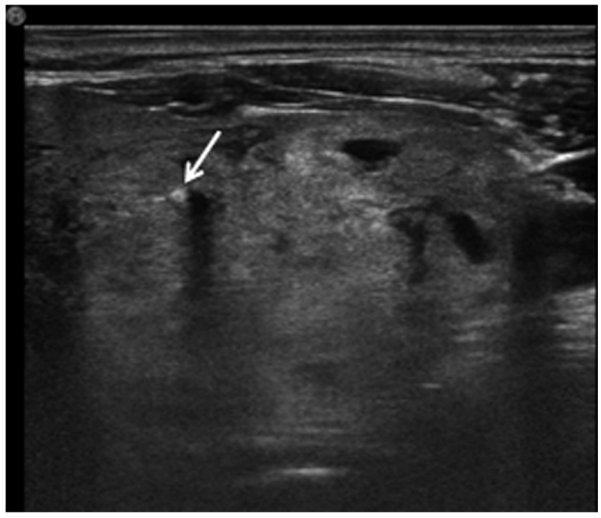


Figure 2. A and **B**, An isoechoic nodule with millimeter-sized cystic areas and peripheral calcifications (arrows) in a 55 years old woman. While the ATA classification system includes interrupted peripheral calcifications among the high-risk criteria, the ACR TI-RADS gives 2 points to a nodule with peripheral calcification. This nodule was categorized as follows: ACR TI-RADS 4, EU TI-RADS 3, K TI-RADS 3, ATA high risk, and Kwak TI-RADS 3. Cytology: benign colloidal nodule, Bethesda 2.

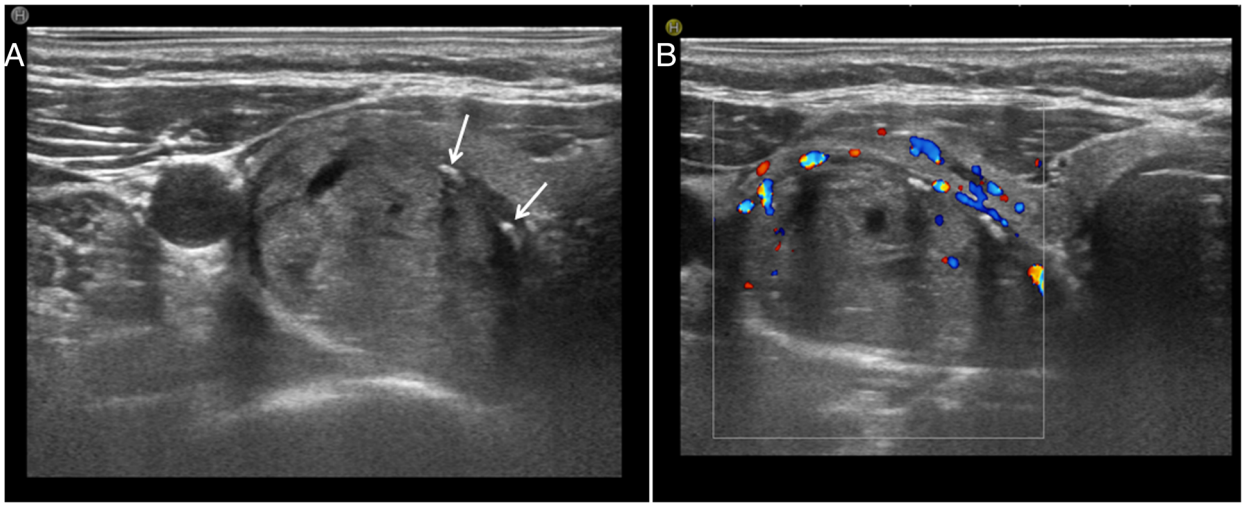
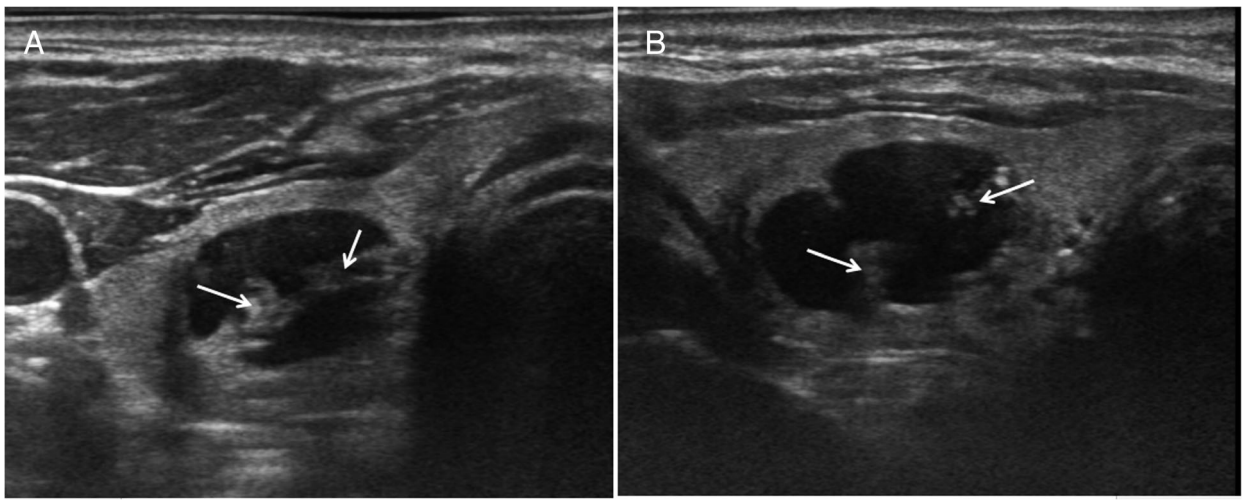


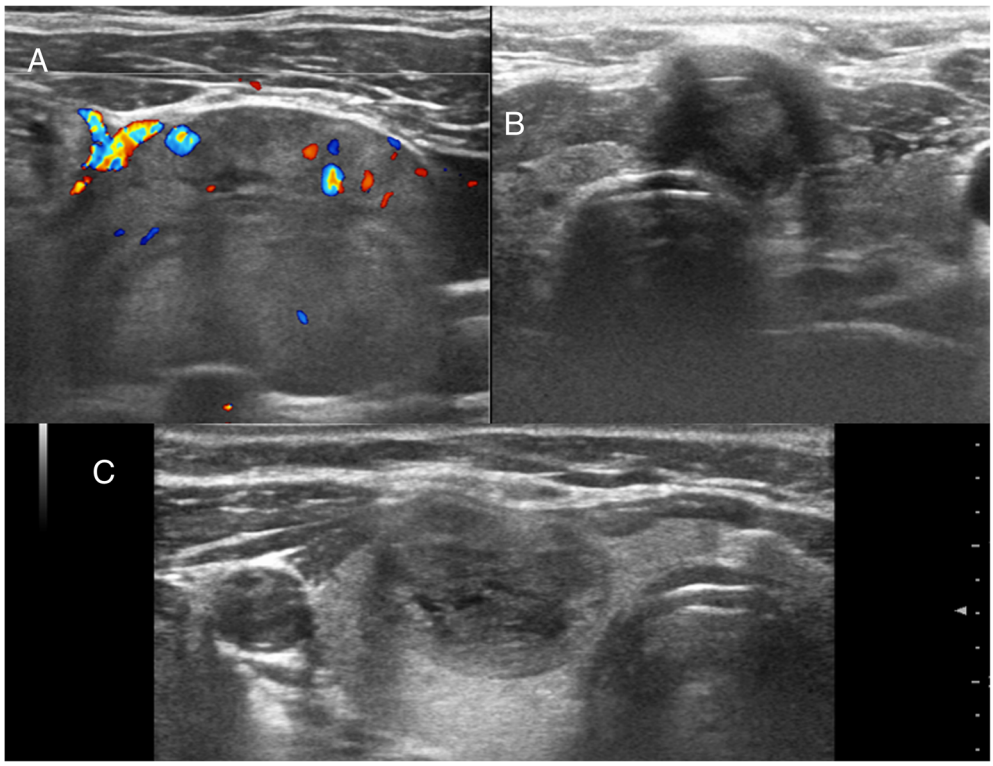
Figure 3. A and **B**, Right-sided cystic nodule with small isoechoic solid areas (arrows) in a 54 years old man; the EU TI-RADS, Kwak TI-RADS, and K TI-RADS, which focus on the solid part, classified this nodule as category 3. According to the ATA system, the nodule was classified as an eccentric solid area with low suspicion. According to the ACR TI-RADS, the nodule was a category 2 cystic and isoechoic solid, mixed nodule. FNAB was performed for this 54-year-old male patient because of a family history of thyroid cancer. Cytology: benign colloidal nodule with cystic changes.



has been recommended. For example, Barbosa et al. combined the ATA and ACR TI-RADS systems.[1]

The relative importance of certain nodule features differs among classification systems. Macrocalcifications are only evaluated by the ACR TI-RADS system, and score 1 point[11] (Figure 1). The ATA classification and ACR TI-RADS classifications evaluate peripheral calcifications. The ATA system classifies solid nodules with peripheral calcification as high risk,[14] while in the ACR TI-RADS these nodules score an extra 2 points (Figure 2).[11]

Figure 4. A, A nodule filling the entire lobe with an undetermined extrathyroidal extension in a 46 years old woman. This nodule was categorized as follows: Kwak TI-RADS 4A, EU TI-RADS 4, and K TI-RADS 4, ATA highly suspicious, and ACR TI-RADS 5. Cytology: benign colloidal nodule. **B**, A nodule located in the isthmus and extending into the gland in a 49 years old woman. This nodule was categorized as follows: Kwak TI-RADS 4A, EU TI-RADS 4, K TI-RADS 4, ATA highly suspicious, and ACR TI-RADS 5. Cytology: lymphocytic thyroiditis. **C**, A nodule located on the right lobe with capsular abutment in a 56 years old man. This nodule was categorized as follows: Kwak TI-RADS 3, EU TI-RADS 3, K TI-RADS 3, ATA low suspicion, and ACR TI-RADS 4. Cytology: benign follicular nodule.



The other classification systems do not consider macro and peripheral calcifications. Nodules with solid and cystic areas are classified as mixed nodules in the ACR TI-RADS and score 1 point (Figure 3).[11] "Eccentric solid component" is a nodule feature similar to "mixed", and is considered only in the ATA classification.[14] The other classification systems categorize nodules according to their dominant component (cystic or solid). Extrathyroidal extension is among the criteria for highly suspicious nodules in the ACR and ATA classification systems.[11,14] It is difficult to evaluate this feature, particularly in large nodules that cover the lobe and nodules located in the isthmus (Figure 4).

The sensitivity of the Bethesda classification has been reported to be between 50 and 97%, while the specificity is between 74.9 and 100% based on surgical thyroidectomy.[27] In the current study, the sensitivity of the ACR TI-RADS (94.7%) was higher than that of the Bethesda classification (76.3%), whereas specificity was similar; thus, the ACR TI-RADS was better for predicting nodule malignancy. The clinical implication of this finding is that ACR TI-RADS category 5 nodules should undergo repeated biopsies or surgical resection despite a benign or indeterminate Bethesda categorization. Future prospective studies may be instructive regarding the management of cases with conflicting radiology and cytology results (Figures 5 and 6).

Since only up to three classification systems were compared in most of the studies, it was possible to evaluate the superiority of all classification systems with meta-analyses, which could lead us to drive inappropriate conclusions. In addition, since in previous studies only nodules above and below certain diameters were included in the studies, how

Figure 5. A, Markedly hypoechoic and solid taller than wide nodule located in the left lobe (arrow) in a 74 years old man. **B**, Lymphadenopathy in the ipsilateral cervical chain (arrows). The nodule was classified as high risk by all of the systems, but the cytology was benign. In the presence of such highly suspicious radiological findings, re-biopsy is essential. Metastatic papillary carcinoma was detected in this patient, who underwent thyroidectomy.

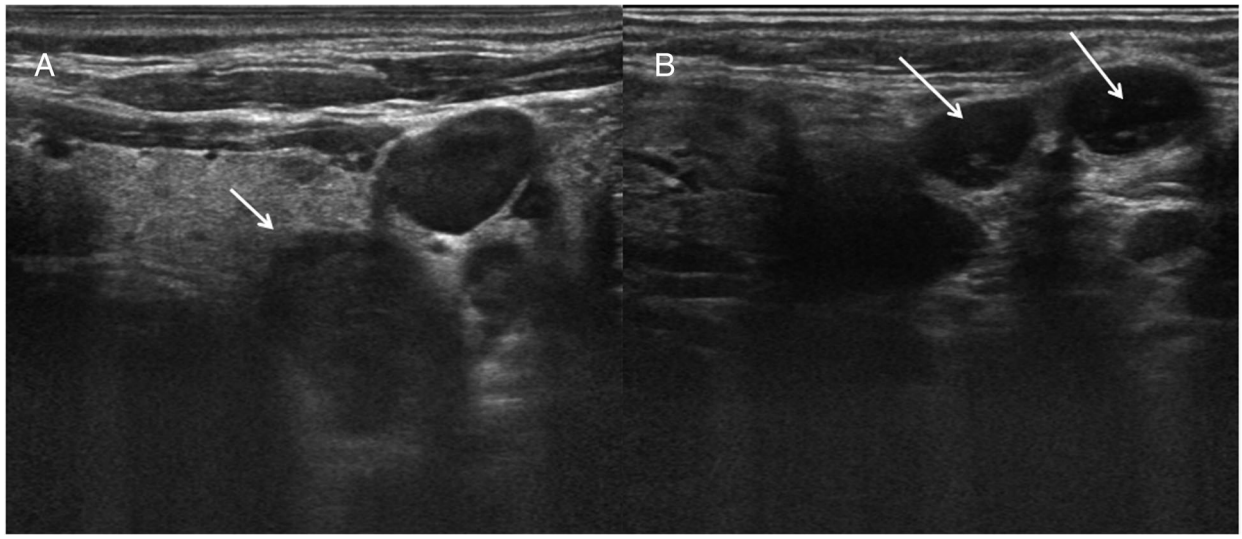
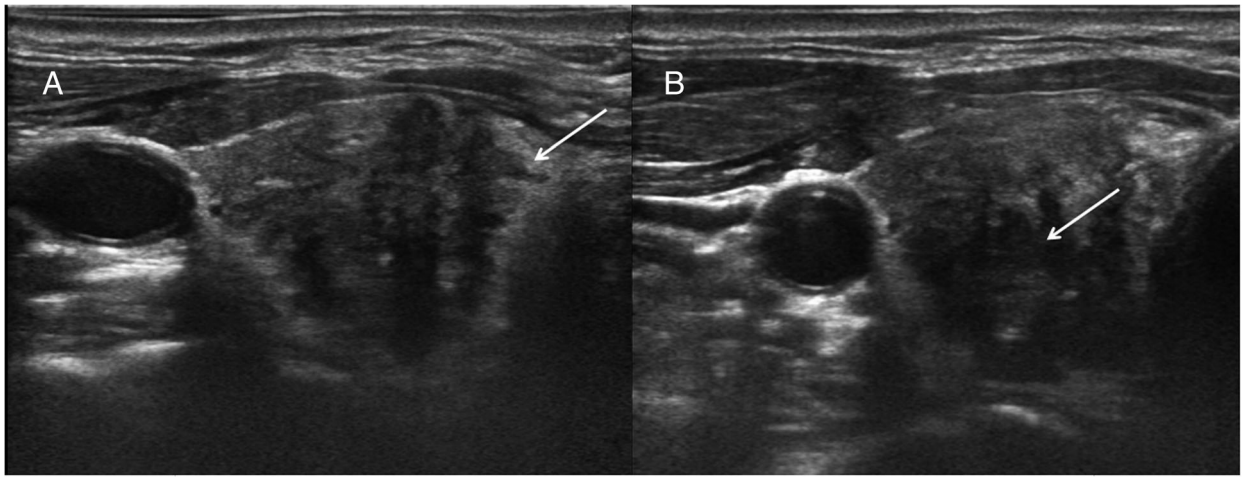


Figure 6. A and **B**, Hypoechoic areas with irregular borders (arrows), as seen in the right thyroid lobe of a 62 years old man, are difficult to interpret radiologically. This lesion meets high-risk criteria, ie, irregular borders, hypoechogenicity, and punctate echogenicities. The post-biopsy cytology was lymphocytic thyroiditis according to all classification systems.



these classification systems would perform in all diameter groups could only be evaluated with meta-analysis as well. The strongest point of our study was that our study was prospective and compared five different classification systems for nodules of all diameters, and also evaluated nodules between 1 and 3 cm. These factors increase the transparency of the study and adds to its reliability. As a limitation, the second observer had evaluated the nodules based on video clips of the cases.

Conclusion

In conclusion, ACR TI-RADS was the most sensitive US classification system for risk stratification of nodules (94.5%), while the Kwak TI-RADS was the most specific (69.1%), ie, the best able to exclude benign nodules according to the combined cytological and histopathologic results. The ACR TI-RADS had higher sensitivity than the Bethesda system based on the histopathological results. ATA classification was the most sensitive system for the nodules which are 1–3 cm in size. If we consider performing FNAB as simpler than misdiagnosing a cancer, it can be concluded that ATA and ACR TI-RADS are the best two classifications for thyroid nodules. Choice can be made according to the size of the nodule and the combination of these two classifications can lead us to near-perfect results.

References

1. Barbosa TLM, Junior COM, Graf H, et al. ACR TI-RADS and ATA US scores are helpful for the management of thyroid nodules with indeterminate cytology. *BMC Endocr Disord* 2019; 19:12902.
2. Basha MAA, Alnaggar AA, Refaat R, et al. The validity and reproducibility of the thyroid imaging reporting and data system (TI-RADS) in categorization of thyroid nodules: multicentre prospective study. *Eur J Radiol* 2019; 117:184–192.
3. Tappouni RR, Itri JN, McQueen TS, Lalwani N. Ou JJ. ACR TI-RADS: pitfalls, solutions, and future directions. *Radiographics* 2019; 39:2040–2052.
4. Huang BL, Ebner SA, Makkar JS, et al. A multidisciplinary head-to-head comparison of American College of Radiology Thyroid Imaging and Reporting Data System and American Thyroid Association ultrasound risk stratification systems. *Oncologist* 2020; 25: 398–403.
5. Fradin JM. ACR TI-RADS: an advance in the management of thyroid nodules or Pandora's box of surveillance? *J Clin Ultrasound* 2020; 48:3–6.
6. Periakaruppan G, Seshadri KG, Vignesh Krishna GM, Mandava R, Sai VPM, Rajendiran S. Correlation between ultrasound-based TIRADS and Bethesda system for reporting thyroid-cytopathology: 2-year experience at a tertiary Care Center in India. *Indian J Endocrinol Metab* 2018; 22:651–655.
7. 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; 210:1148–1154.
8. Schenke S, Klett R, Seifert P, Kreissl MC, Görges R, Zimny M. Diagnostic performance of different thyroid imaging reporting and data systems (Kwak-TIRADS, EU-TIRADS and ACR TI-RADS) for risk stratification of small thyroid nodules (≤10 mm). *J Clin Med* 2020; 9:236.
9. 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; 260:892–899.
10. Shin JH, Baek JH, Chung J, et al. Korean Society of Thyroid Radiology (KSThR) and Koreaan Society of Radiology. Ultrasonography diagnosis and imaging-based Management of Thyroid Nodules: revised Korean Society of Thyroid Radiology consensus statement and recommendations. *Korean J Radiol* 2016; 17:370–395.
11. Tessler FN, Middleton WD, Grant EG, Hoang JK. ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS Committee. *J Am Coll Radiol* 2018; 15:381–382.
12. Andrioli M, Carzaniga C, Persani L. Standardized ultrasound report for thyroid nodules: the Endocrinologist's viewpoint. *Eur Thyroid J* 2013; 2:37–48.
13. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European thyroid association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: the EU-TIRADS. *Eur Thyroid J* 2017; 6:225–237.
14. 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.
15. Kwak JY. Indications for fine needle aspiration in thyroid nodules. *Endocrinol Metab* 2013; 28:81–85.
16. Cibas ES, Ali SZ. The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid* 2017; 27:1341–1346.
17. Grani G, Lamartina L, Cantisani V, Maranghi M, Lucia P, Durante C. Interobserver agreement of various thyroid imaging reporting and data systems. *Endocr Connect.* 2018; 7:1–7.
18. Phuttharak W, Boonrod A, Klungboonkrong V, Witsawapaisan T. Interrater reliability of various thyroid imaging reporting and data system (TIRADS) classifications for differentiating benign from malignant thyroid nodules. *Asian Pac J Cancer Prev* 2018; 20:1283–1288.
19. 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: 90–96.
20. Kim PH, Suh CH, Baek JH, Chung SR, Choi YJ, Lee JH. Unnecessary thyroid nodule biopsy rates under four ultrasound risk stratification systems: a systematic review and meta-analysis. *Eur Radiol* 2020; 31:2877–2885.

21. Yoon SJ, Na DG, Gwon HY, et al. Similarities and differences between thyroid imaging reporting and data systems. *AJR Am J Roentgenol* 2019; 213:W76–W84.
22. Migda B, Migda M, Migda MS, Slapa RZ. Use of the Kwak thyroid image reporting and data system (K-TIRADS) in differential diagnosis of thyroid nodules: systematic review and meta-analysis. *Eur Radiol* 2018; 213:W76–W84.
23. Yang R, Zou X, Zeng H, Zhao Y, Ma X. Comparison of diagnostic performance of five different ultrasound TI-RADS classification guidelines for thyroid nodules. *Front Oncol* 2020; 10: 598225.
24. Nam SJ, Kwak JY, Moon HJ, Yoon JH, Kim EK, Koo JS. Large (≥3cm) thyroid nodules with benign cytology: can thyroid imaging reporting and data system (TIRADS) help predict false-negative cytology? *PLoS One* 2017; 10.
25. Ahmadi S, Oyekunle T, Jiang X, et al. A direct comparison of the ATA and TI-RADS ultrasound scoring systems. *Endocr Pract* 2019; 25:413–422.
26. Pandya A, Caoili EM, Jawad-Makki F, et al. Davenport MS. 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; 214:900–906.
27. Nandedkar SS, Dixit M, Malukani K, Varma AV, Gambhir S. Evaluation of thyroid lesions by fine-needle aspiration cytology according to Bethesda system and its histopathological correlation. *Int J Appl Basic Med Res* 2018; 8:76–82.