Canadian Association of Radiologists Journal 70 (2019) 68–73

CANADIAN ASSOCIATION OF RADIOLOGISTS JOURNAL

www.carjonline.org

Ultrasonography / Échographie

Retrospective Application of the 2015 American Thyroid Association Guidelines for Ultrasound Classification, Biopsy Indications, and Follow-up Imaging of Thyroid Nodules: Can Improved Reporting Decrease Testing?

Manijeh Mohammadi, MD[a,*], Carrie Betel, MD, BSc, FRCP[b], Kirsteen Rennie Burton, BSC, MBA, MSc, MD[b], Kevin McLughlin Higgins, MD, MSc, FRCRC[c], Zeina Ghorab, MD, MSc[d], Ilana Jaye Halperin, MD, MSC, FRCPC[e]

[a]*Department of Medicine, University of Toronto, Toronto, Ontario, Canada*
[b]*Department of Medical Imaging, Sunnybrook Health Sciences Center, Toronto, Ontario, Canada*
[c]*Department of Head and Neck Surgery, Sunnybrook Health Sciences Center, Toronto, Ontario, Canada*
[d]*Department of Laboratory Medicine and Pathobiology, Sunnybrook Health Sciences Center, Toronto, Ontario, Canada*
[e]*Division of Endocrinology, Department of Medicine, Sunnybrook Health Sciences Center, University of Toronto, Toronto, Ontario, Canada*

Abstract

Introduction: Thyroid ultrasound has been widely used to determine which nodules need further investigation. The goal of this study is to determine if using an ultrasonographic features checklist based on 2015 American Thyroid Association (ATA) guidelines can improve reporting and decrease unnecessary further testing.
Methods: In this retrospective study, ultrasonographic images of all nodules biopsied at our institution in 2014 and 2015 were reviewed by radiologists blinded to fine needle aspiration (FNA) biopsy result using a checklist. The checklist was prepared based on 2015 ATA guidelines. The ultrasonographic characteristics of thyroid nodules were compared with the result of biopsy to determine positive predictive value (PPV), negative predictive value (NPV), sensitivity, and specificity for predicting malignancy. Radiologists also made an overall recommendation on need for FNA.
Results: A total of 425 thyroid nodule ultrasound scans were reviewed by radiologists. Biopsy results of 31 nodules were malignant and 394 were non-malignant. Malignant nodules showed higher frequency of solid composition, hypoechoechogenicity, and cervical lymph node involvement compared to benign nodules. Solid nodule composition had the highest PPV (13%) and NPV (94.7%). Extra-thyroid extension had the highest specificity (90.1%). Lesion vascularity had the highest sensitivity (83.8%), followed by hypoechogenicity (65.6%). Overall, the checklist had a positive predictive value of 9%, negative predictive value of 97.5%, sensitivity of 96.8%, and specificity of 11.14%. Radiologists determined that 10% of the nodules were very low-risk and did not require FNA.
Conclusion: Using a checklist based on 2015 ATA guideline thyroid nodule ultrasonographic features is a sensitive tool with high NPV to predict benign thyroid nodule, thereby preventing unnecessary FNAs.

Résumé

Introduction : L'échographie thyroïdienne a été largement employée pour déterminer quels nodules nécessitent une évaluation plus approfondie. L'objectif de cette étude est de déterminer si l'utilisation d'une liste d'évaluation des caractéristiques à l'échographie fondée sur les recommandations de l'American Thyroid Association (ATA) 2015 peut améliorer le signalement et diminuer la conduite d'analyses supplémentaires inutiles.

* Address for correspondence: Manijeh Mohammadi, MD, Department of Medicine, Sunnybrook Health Sciences center, University of Toronto, Toronto, Ontario, Canada

E-mail address: mohammadi.mjh@gmail.com (M. Mohammadi).

https://doi.org/10.1016/j.carj.2018.09.001

Méthodes : Lors de cette étude rétrospective, les clichés obtenus par échographie sur l'ensemble des nodules ayant fait l'objet d'une biopsie dans notre établissement en 2014 et en 2015 ont été analysés à l'aveugle concernant le résultat de biopsie par aspiration à l'aiguille fine (AAF) par des radiologistes utilisant la liste d'évaluation. La liste d'évaluation a été préparée selon les recommandations établies par l'ATA 2015. Une comparaison a été effectuée entre les caractéristiques des nodules thyroïdiens à l'échographie et le résultat de biopsie pour déterminer la valeur prédictive positive (VPP), la valeur prédictive négative (VPN), la sensibilité et la spécificité des facteurs prédictifs de malignité. Les radiologistes ont par ailleurs émis une recommandation générale sur les besoins en matière d'AAF.
Résultats : Les radiologistes ont analysé un total de 425 examens échographiques de nodules thyroïdiens. Les résultats de biopsie de 31 nodules démontraient une malignité, et 394 révélaient une non-malignité. Comparativement aux nodules bénins, les nodules malins présentaient une fréquence de composition solide supérieure et une hypoéchogénicité, et atteignaient le ganglion lymphatique cervical. La composition solide des nodules présentait la VPP (13 %) et la VPN (94,7 %) les plus élevées. La plus haute spécificité se situait au niveau de l'extension extrathyroïdienne (90,1 %). La vascularité lésionnelle présentait la plus haute sensibilité (83,8 %), suivie de l'hypoéchogénicité (65,6 %). Globalement, la liste d'évaluation était associée à une valeur prédictive positive de 9 %, une valeur prédictive négative de 97,5 %, une sensibilité de 96,8 % et une spécificité de 11,14 %. Les radiologistes ont déterminé que 10 % des nodules étaient à très faible risque et ne nécessitaient pas d'AAF.
Conclusion : La liste d'évaluation des caractéristiques des nodules thyroïdiens à l'échographie basée sur la recommandation de l'ATA 2015 est un outil sensible de prédiction de la présence de nodules thyroïdiens bénins et permet d'éviter les AAF inutiles.

Key Words: Thyroid nodule; Ultrasonographic features; 2015 American Thyroid Association guideline

Thyroid nodules are common; 4%−7% of adults in North America have palpable nodules; however, when informed by imaging studies, the prevalence rises to 30% [1]. Increased use of ultrasound (US) will likely lead to an increase in the number of nodules discovered, the need for fine needle aspiration (FNA) biopsies, the number of patients undergoing thyroidectomy and radioiodine treatment, and the number of people with benign nodules requiring ongoing follow-up, all of which have significant resource implications [2]. Thyroid US has been widely used to stratify the risk of malignancy in thyroid nodules and aid decision-making about whether FNA is indicated [3,4]. It has been proposed that thyroid US can be used to differentiate benign and malignant thyroid nodules on the basis of the size, internal texture, shape, echogenicity, margins, presence of calcification, and characteristics of adjacent structures [5]. Having a detailed US report which includes the presence or absence of all these characteristics would help the clinician to determine which patients are high risk requiring FNA and which ones can be observed. With more standardized reporting, it may be possible to decrease the frequency of FNA biopsies and follow-up US. For example, spongiform or partially cystic nodules without any concerning features can be followed with observation [3]. In the 2015 2015 American Thyroid Association (ATA guidelines), thyroid nodules are categorized as high-, intermediate-, and low-risk based on ultrasonographic features shown in Supplemental Appendix S1 (ATA guideline).

The primary goal of this project was to determine if the introduction of standardized reporting using the ATA [3] ultrasonographic criteria for thyroid nodules reduces the frequency of FNAs.

Methods

A diagnostic accuracy study was done to determine the positive and negative predictive value of ATA US criteria to diagnose benign nodules. The pathology database provided the cytologic diagnosis and medical record numbers for all of the patients with a FNA during the study period (January 2014-December 2015). Two radiologists (1 staff with 12 years experience and 1 fourth-year radiology resident) blinded to the cytologic diagnosis, then reviewed the corresponding nodules using the checklist (Supplemental Appendix S1). The checklist assessed thyroid nodule characteristics including size, echogenicity (hypo-, hyper-, iso-, or mixed), composition (cystic, solid, or spongiform), calcification (microcalcification, macrocalcification, or mixed calcification), margin (regular, irregular), taller than wide shape, extrathyroid extension, interrupted peripheral calcification, and cervical chain node involvement. The first 35 nodules were reviewed in duplicate, and kappa for agreement was estimated. After that, the nodules were split between the 2 radiologists. Comparing the cytology result and ultrasonography report, we determined the positive and negative predictive value of the thyroid nodule US classification for the various cytology diagnoses [3]. Nodules were categorized based on cytology, and if available, pathology result, into 3 groups of benign, malignant and follicular lesion of unknown significance [3]. Nodules with follicular lesion of unknown significance cytology result were categorized as non-malignant. The study radiologists completed the checklist and gave a final recommendation for or against FNA. The radiologists made a recommendation based on sonographic patterns, estimated risk of malignancy, and the 2015 ATA guidelines [3]. A chart based on ATA guideline biopsy indications was provided to radiologists. We used this decision to assess the PPV, NPV, sensitivity, and specificity of the checklist based on the ATA ultrasonographic features. Patients with thyroid FNA and corresponding US images at Sunnybrook Health Science Centre between January 2014-December 2015 were sequentially entered in the study sample. Those nodules which had unsatisfactory or non-diagnostic biopsy results were excluded from the

diagnostic accuracy calculation. For those nodules which underwent thyroidectomy, we considered surgical pathology results instead of cytology results as the gold standard diagnostic tests. For patients with multiple nodules, we reviewed each nodule that had been biopsied separately. Data on characteristics of thyroid nodule and FNA and/or surgical pathology results were collected. Data were reported in percentage in case of categorical variables and means ± standard deviation (SD) for continuous variables. Descriptive statistics including frequencies and cross-tabs were used. Frequency was used to determine percentage of each variable, and cross-tabs was used to determine percentage of one variable in 2 groups (benign/malignant). IBM SPSS 20 was used to analyze data. Two-by-two tables for diagnostic accuracy were constructed of the ATA US classifications to determine the PPV, NPV, sensitivity, and specifity for the malignant thyroid nodule.

The study was approved by the Sunnybrook Hospital Research Ethics Board. The data collected was anonymized by removing patient identifiers. Enrolled patients were assigned a unique study identification number. The key code was stored on an encrypted USB stick. The data was stored on another encrypted USB stick. The USB sticks were stored in separate locked drawers.

Results

A total of 490 thyroid nodules were assessed. Of them, 65 nodules were excluded for incomplete data, such as no US or biopsy in our database (Figure 1). The kappa for agreement on the first 35 nodules for the need for FNA (yes/no) was 0.66.

Cytology results of included nodules showed 32 (7.6%) malignant and 393 (92.4%) non-malignant (Figure 1). Six nodules had cytologic diagnosis of follicular neoplasm. Of those, 5 had surgical pathology and were then re-categorized into either benign or malignant categories. One nodule labeled as follicular neoplasm with no corresponding surgical pathology was excluded from the study.

There were 340 (80%) nodules in female and 85 (20%) in male. Mean age of patients was 59.39 ± 14 years in benign thyroid nodule and 53.5 ± 13.8 years in malignant thyroid nodule ($P = .02$). Mean size of the largest dimension of thyroid nodule was 2.1 ± 1.04 in benign nodule and 2.2 ± 1 in malignant nodule ($P = .96$). Ultrasonographic characteristics of thyroid nodules in various cytologic classifications are summarized in Table 1.

Based on the 2015 ATA US risk patterns, 16 (50%) of the malignant nodules had high-risk US features, 14 (43.7%) of malignant nodules had intermediate features, and 2 (6.2%) malignant nodules with low-risk features. The cytology results based on US risk is shown in Table 2.

In our study, radiologists recommended to not biopsy 43 (10%) thyroid nodules that had very low-risk ultrasonographic features. The result showed that the checklist had a positive predictive value of 9%, negative predictive value of 97.5%, sensitivity of 96.8% and specificity of 11.14% for a malignant cytologic diagnosis. We also calculated NPV, PPV, sensitivity, and specificity for individual US features (Table 3). Review of the cytology result of the nodules that would not have been biopsied based on our study radiologist recommendations revealed that 36 of them were benign and 7 had an unsatisfactory result.

Discussion

In this study, we applied a structured checklist to nodules for which we already had the cytologic diagnosis. Comparing checklists results as a reference test with FNA or surgical pathology as a standard test, we found high NPV and

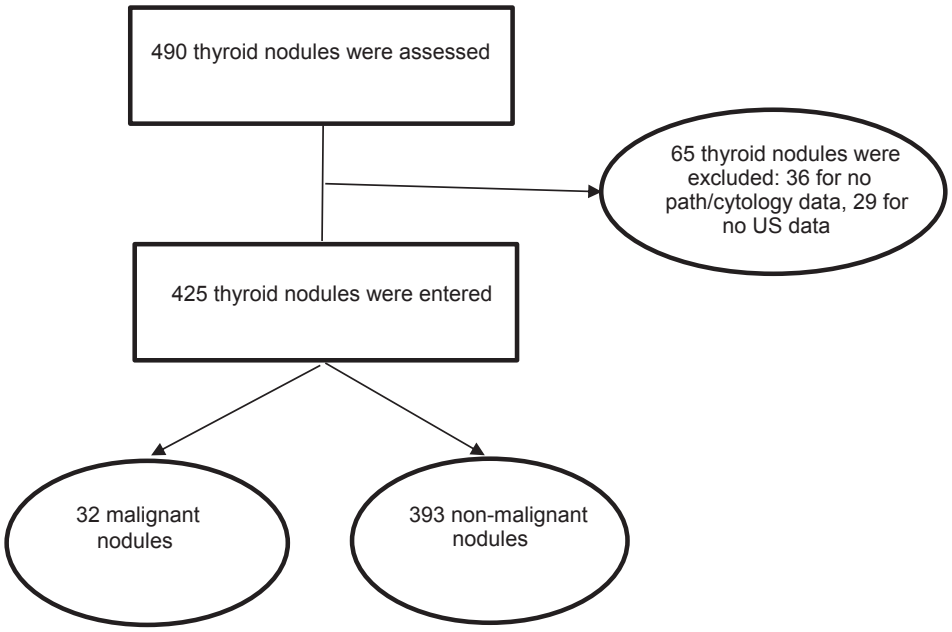

Figure 1. Flow chart of our study.

Table 1
Ultrasonographic features of thyroid nodules

Cytology	Benign	Malignant	FLUS
Echogenicity			
Hyper	27 (7.7%)	3 (9.6%)	3 (7.1%)
Hypo	132 (37.6%)	17 (54.8%)	18 (42.9%)
Iso	96 (27.4%)	8 (25.8%)	11 (26.2%)
Mixed	93 (26.5%)	3 (9.6%)	10 (23.8%)
Composition			
Cystic			
Cystic < 50%	108 (31%)	7 (22.5%)	8 (18.6%)
Cystic > 50%	41 (11.7%)	1 (3.2%)	3 (7%)
Minimal cyst	48 (13.7%)	3 (9.6%)	5 (11.6%)
Solid	133 (37.9%)	20 (64.5%)	25 (58.1%)
Spongiform	17 (4.9%)	0 00%	1 (2.3%)
Mixed	2 (0.6%)	0 00%	0 00%
Margin			
Regular	190 (54.1%)	22 (70.9%)	24 (57.1%)
Irregular	157 (44.7%)	9 (29%)	16 (38.1%)
Calcification			
Yes	169 (48.1%)	12 (38.7%)	18 (42.9%)
No	175 (49.9%)	18 (58%)	23 (54.8%)
Taller than wide shape			
Yes	53 (15.1%)	5 (16.1%)	10 (23.8%)
No	290 (82.6%)	26 (83.8%)	30 (71.4%)
Extra thyroid extension			
Yes	32 (9.2%)	1 (3.2%)	2 (4.8%)
No	311 (88.6%)	30 (96.7%)	38 (90.5%)
Interrupted rim			
Yes	63 (17.9%)	5 (16.1%)	9 (20.9%)
No	286 (81.5%)	26 (83.8%)	34 (79.1%)
Vascularity			
Yes	256 (72.9%)	26 (83.8%)	33 (78.6%)
No	73 (20.8%)	5 (16.1%)	6 (14.3%)
Lymph node involvement			
Yes	18 (5.1%)	2 (6.5%)	0 00%
No	327 (93.2%)	29 (93.5%)	42 (100%)

FLUS = follicular lesion of unknown significance.

sensitivity and low PPV and specificity of the checklist as a diagnostic tool to diagnose malignant thyroid nodule. In addition, the use of the checklist provided support to the radiologists to recommend against biopsy in 43 low-risk nodules, none of which were malignant on confirmatory cytology. This confirms that biopsy of nodules with low-risk ultrasonographic features is unnecessary.

The risk of malignancy in thyroid nodules varies in previous reports from 4.8%–20% [6–9]. In our series, 7.6 percent of thyroid nodules showed malignancy. The lower frequency of malignancy in our studies could be due to the exclusion of patients who did not have a formal radiologic assessment. In our institution, the ear, nose and throat

Table 2
Cytology result of thyroid nodule with different ultrasonographic risk

US risk ▶ Cytology ▼	Low-risk	Intermediate-risk	High-risk
Benign	53 (15.1%)	77 (22%)	219 (62.7%)
FLUS	2 (4.6%)	17 (39.5%)	24 (55.8%)
Malignant	2 (6.2%)	14 (43.7%)	16 (50%)

FLUS = follicular lesion of unknown significance; US = ultrasound.

Table 3
Positive predictive value, negative predictive value, sensitivity, and specificity of US suspicious features in predicting malignant thyroid nodule

	PPV	NPV	Sensitivity	Specificity
Solid nodule	13%	94.7%	62.5%	62%
Hypoechogenicity	9.4%	93%	65.6%	42.4%
Irregular margin	5.9%	89.6%	31.2%	54.8%
Calcification	5.3%	90.4%	19.3%	68.9%
Taller than wide shape	8.6%	91.5%	15.6%	84.5%
Extra thyroid extension	2.8%	91.2%	3.2%	90.1%
Vascularity	9.2%	93.6%	83.8%	22.4%
Radiology recommendation for biopsy	9%	97.5%	96.8%	11.14%

NPV = negative predictive value; PPV = positive predictive value; US = ultrasound.

surgeons do office-based US and FNA that were excluded from our analysis, as our focus was on assessing and improving reporting and recommendations from radiologists. In addition, other studies may have over-represented malignant nodules due to selection bias [8,9].

In our study, specificity and PPV of US features were low; similar the findings of Gul et al [6]. In contrast to their study, sensitivity of individual US features was also low in our study [6]. Even though suspicious US features are more common in malignant nodules [10], not all malignant nodules showed suspicious US features. In our study, the radiologists' opinion on need for FNA has the highest sensitivity. Radiologists made this recommendation based on ATA criteria. Overall, this result supports that not a single US feature can be a predictor of malignancy as shown in previous reports [8,11,12].

The NPV of individual US features and overall radiologist opinion of thyroid nodules were high in our study; consistent with previous reports [6,10,11]. The high NPV in our study could be related to low proportion of malignant nodules. Other reports have demonstrated low PPV and specificity, but higher sensitivity of individual US features [6,11,13]. In our study, all of these values were low for individual suspicious US features, but when used to estimate an overall risk of malignancy, sensitivity was increased. These results showed this checklist can be used to determine nodules with low risk of malignancy. This is helpful to avoid unnecessary biopsy and reduce cost and patient anxiety associated with FNAs.

There are different reports in previous studies regarding the value of thyroid nodule US features to predict malignancy. Gulcelik et al [10] studied ultrasonographic features as a predictor of malignancy in the patients who underwent thyroidectomy. They assessed echogenicity, presence of calcification and nodule structure. The variable associated with the highest sensitivity was the presence of a solid nodule (88.5%), whereas the presence of microcalcifications had the highest specificity [10]. In our study, analyzing different US characteristics showed vascularity had the highest sensitivity and extra thyroid nodule extension had the highest specificity. Vascularity increases in malignant tissue due to cellular proliferation. Vascularity can be also increased in benign tissue with follicular proliferation and

granulation tissue. Because it can be seen both in benign and malignant nodules, the 2015 ATA guidelines no longer consider vascularity as a suspicious ultrasonographic feature to predict malignancy [3,7]. If we consider the same variables that were assessed in Gulcelik et al's study [10], which are echogenicity, presence of calcification, and nodule composition, then micro calcification would have the highest specificity; similar to other studies [8,10], and the variable with the highest sensitivity would be hypoechogenicity followed by solid nodule [10]. One possible reason for the difference in our results is that Gulcelik et al's study [10] included patients who had positive follicular neoplasm by FNA. Papini et al [8] who studied the same US features found solid nodule with highest sensitivity and microcalcification with the highest specificity [8]. They achieved this result from US features of non-palpable thyroid nodule. In our study, these values were obtained from the assessment of all thyroid nodules which underwent FNA regardless of pathology result or size.

In the current study, review of the malignant nodule composition showed 64.5% were solid or minimally cystic, 22.5% were classified as less than 50% cystic, and 12.8% were classified as more than 50% cystic. In contrast, a study by Henrichsen et al showed that 88% of malignant nodules were solid or minimally cystic 9% were less than 50% cystic and 2.5% were more than 50% cystic [14]. In our study, all but two of the malignant nodules with partially cystic composition had other suspicious ultrasonographic features or were hypo echoic. However, the sizes of malignant nodules without any suspicious ultrasonographic features were more than 1.5 cm in the largest dimension. Based on 2015 ATA guidelines, partially cystic nodules without suspicious ultrasonographic features had low risk of malignancy (5%−10%); therefore the recommendation is to only biopsy nodules greater than 1.5 cm and for nodule with very low suspicious for malignancy based on guideline can either biopsy for nodule above 2 cm or choose observation without biopsy [3]. Based on our study result, 6.2% of malignant nodule had low or very low-risk US features. In this study, FNA based on size was preferred compared to observation. Our approach was more conservative. Therefore, it is important to consider nodule size when determining which low-risk nodules require FNA.

In this study, agreement between the 2 radiologists regarding the need for FNA was moderate (kappa 0.66). Before introducing the checklist for wide use, we will consider choosing raters familiar with the checklists, and training the raters in systematic practice sessions to improve interrater reliability.

Our study has a number of limitations; as this was a single center retrospective study, we were unable to include US that were done outside of our center or in the ear, nose, and throat surgeon's offices. This may introduce a selection bias of nodules assessed by radiology at our site, but the blinding of radiologists to cytologic diagnosis should mitigate that bias. The review of ultrasonographic images was done by 2 radiologists; due to resource constraints, we were unable to have both radiologists review all the images, but a kappa for agreement on the need to biopsy was assessed at the outset and felt to be reasonable.

The strength of our study lies in the creation and application of a checklist. As the checklist was based on evidence-based criteria from the 2015 ATA guidelines, it provided support to the radiologists to recommend against more FNAs than had actually been done in 2015. Future quality improvement efforts will focus on implementing standardized reporting and decision support in the busy workflow of our radiology department.

Conclusion

US has been widely used for thyroid nodule diagnosis. Based on US results, patients are referred for FNA. With a detailed US report, FNA can be avoided in very low-risk and low-risk nodules. Use of the checklist would result in a 10% theoretical reduction in previously ordered FNAs; however, we cannot comment if there would be an increase in FNAs ordered based on the 2015 ATA criteria. Given that the older guidelines suggested everything greater than 1 cm (unless it was a simple cyst) be biopsied; it is unlikely the new checklist would result in a high number of additional FNAs.

Supplementary Data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.carj.2018.09.001.

References

[1] Brito JP, Gionfriddo MR, Al Nofal A, et al. The accuracy of thyroid nodule ultrasound to predict thyroid cancer: systematic review and meta-analysis. J Clin Endocrinol Metab 2014;99:1253−63.
[2] Bastin S, Bolland MJ, Croxson MS. Role of ultrasound in the assessment of nodular thyroid disease. J Med Imaging Radiat Oncol 2009;53:177−87.
[3] Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid 2016;26:1−133.
[4] 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.
[5] Moon WJ, Baek JH, Jung SL, et al. Korean Society of Thyroid Radiology (KSThR), Korean Society of Radiology. Ultrasonography and the ultrasound-based management of thyroid nodules: consensus statement and recommendations. Korean J Radiol 2011;12:1−14.
[6] Gul K, Ersoy R, Dirikoc A, et al. Ultrasonographic evaluation of thyroid nodules: comparison of ultrasonographic, cytological, and histopathological findings. Endocrine 2009;36:464−72.
[7] Kim HG, Moon HJ, Kwak JY, Kim EK. Diagnostic accuracy of ultrasonographic features for subcentimeter thyroid nodules suggested by the revised American Thyroid Association guideline. Thyroid 2013;23:1583−9.
[8] Papini E, Guglielmi R, Bianchini A, et al. Risk of malignancy in nonpalpable thyroid nodules: predictive value of ultrasound and color-Doppler features. J Clin Endocrinol Metab 2002;87:1941−6.

[9] He LZ, Zeng TS, Pu L, et al. Thyroid hormones, autoantibodies, ultrasonography, and clinical parameters for prediciting thyroid cancer. Int J Endocrinol 2016. 8215834.

[10] Gulcelik NE, Gulcelik MA, Kuru B. Risk of malignancy in patients with follicular neoplasm. Predictive value of clinical and ultrasonographic features. Arch Otolaryngol Head Neck Surg 2008;134: 1312−5.

[11] Rago T, Vitti P, Chiovato L, et al. Role of conventional ultrasonography and color flow-Doppler sonography in predicting malignancy in 'cold' thyroid nodules. Eur J Endocrinol 1998;138:41−6.

[12] Ram N, Hafeez S, Qamar S, et al. Diagnostic validity of ultrasonography in thyroid nodules. J Pak Med Assoc 2015;65:875−8.

[13] Park VY, Kim E, Kwak JY, et al. Thyroid imaging and reporting system and ultrasound elastrography: diagnostic accuracy as tool in recommending fine needle aspiration for solid thyroid nodules with non diagnostic fine needle aspiration cytology. Ultrasound Med Biol 2016; 42:399−406.

[14] Henrichsen TL, Reading CC, Charboneau JW, et al. Cystic change in thyroid carcinoma: prevalence and estimated volume in 360 carcinomas. J Clin Ultrasound 2010;38:361−6.