European Journal of Radiology Open 6 (2019) 122–127

Contents lists available at ScienceDirect

European Journal of Radiology Open

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

Diagnostic performance of 2015 American Thyroid Association guidelines and inter-observer variability in assigning risk category

Zhiyu Pang, Myles Margolis, Ravi J. Menezes, Hassaan Maan, Sangeet Ghai*

Toronto Joint Department of Medical Imaging, University Health Network - Mount Sinai Hospital - Women's College Hospital, University of Toronto, Toronto, ON, Canada

ARTICLE INFO

Keywords:
ATA guidelines
Interobserver agreement
Interobserver variability
Thyroid ultrasound
Fine needle aspiration biopsy

ABSTRACT

Objective: To determine the interobserver variability of the 2015 American Thyroid Association (ATA) thyroid guidelines and to evaluate the diagnostic accuracy of the guidelines in detecting thyroid cancer.
Materials and methods: Sonographic patterns of 189 thyroid lesions were retrospectively analyzed by two radiologists according to the 2015 guidelines. The risk of malignancy was calculated for each pattern and compared with the published expected risk of malignancy.
Results: The observed risk of malignancy for very low suspicion, low suspicion, intermediate suspicion and high suspicion patterns were 2%, 12.7%, 26.3% and 29.8% respectively. Interobserver agreement for final category assignment was moderate (κ 0.518).
Conclusion: The estimated risk of malignancy in the high suspicion pattern of the 2015 ATA thyroid biopsy guidelines appears to be less than stated. However, this needs further validation in a larger cohort study.

1. Introduction

Thyroid nodules are common. Studies have suggested a prevalence of 2–6% with palpation, 19–35% on ultrasound, and 8–65% on autopsy [1]. With increased use of ultrasound, improvement in detection techniques and screening in asymptomatic population, the incidence of thyroid cancer has increased by 4.5% per year over the last 10 years. This is a faster increase than for any other cancer, while the corresponding mortality rate from the disease has not changed [2,3]. This implies that the widespread rise in thyroid nodule detection and cancer incidence is from overdiagnosis of predominantly indolent subclinical lesions [4,5].

Several ultrasound (US) features have been described as potential predictors of thyroid malignancy, including presence of calcification, hypoechogenicity, irregular margins, taller than wide shape morphology, and a predominantly solid composition [6]. These features have been used to identify nodules that should be subjected to fine needle aspiration biopsy (FNAB). However, there is significant overlap between US features of benign and malignant nodules, and no single US feature has been shown to have a high sensitivity or a high positive predictive value for thyroid cancer detection. For this reason, guidelines with various combinations of US features using multivariate analysis have been put forward in an attempt to provide better diagnostic accuracy than to rely on a single sonographic feature [7–9]. In 2017, the American College of Radiology – Thyroid Imaging, Reporting and Data System (ACR-TIRADS) white paper for management of thyroid nodules was published with the aim to increase specificity of thyroid sampling in order to detect clinically significant malignancies [10]. However, debate continues as to which guidelines are the most accurate and effective.

The revised American Thyroid Association (ATA) guidelines were put forward in 2015 based on published evidence relating to the diagnosis and management of thyroid nodules and were implemented in our department soon after. The ATA guidelines categorize the nodules into 5 different categories – high risk, intermediate risk, low risk, very low risk and benign, with recommendations for biopsy based on risk category and size of the nodule [11]. In this study, we assess the accuracy and diagnostic performance of the ATA guidelines to predict the presence of histologically confirmed malignancy based on the Bethesda System for Reporting Cytopathology [6]. At the same time, evaluating inter-reader reproducibility of the ultrasound features of thyroid nodules is critical to confirm that any reporting guidelines are attaining their primary aim of standardization. Variability in interpretation of images can lead to a variable overall suspicion by readers, and therefore of the lexicon that underlies the assignment of such sonographic patterns. Thus, we included interobserver reproducibility of sonographic

* Corresponding author at: Toronto General Hospital, Department of Medical Imaging, 1PMB-283, 585 University Avenue, Toronto, M5G 2N2, ON, Canada.
E-mail addresses: zpang@insightimaging.ca (Z. Pang), myles.margolis@sinaihealthsystem.ca (M. Margolis), ravi.menezes@uhn.ca (R.J. Menezes), hmaan@uoguelph.ca (H. Maan), sangeet.ghai@uhn.ca (S. Ghai).

https://doi.org/10.1016/j.ejro.2019.03.002
Received 13 February 2019; Accepted 17 March 2019
Available online 27 March 2019

features of thyroid nodules in our assessment of the ATA thyroid biopsy guidelines.

2. Subjects and methods

2.1. Patients

Institutional Review Board approval was granted, and informed consent was waived for this retrospective study. Thyroid biopsy center database was set up in 2011 at our tertiary referral center. In the first year of inception, 897 nodules were entered into the IRB approved registry. Of these, 249 nodules were excluded because of lack of follow-up after results of initial nondiagnostic or benign results leaving 648 nodules with clinical or imaging follow up > 1 year, repeat biopsy or surgical pathology at our institute. The first 189 consecutive nodules from this cohort were retrospectively evaluated for sonographic features in a blinded manner by two staff radiologists and the results are presented here.

2.2. Ultrasound and ultrasound guided FNA biopsy technique

Real-time US examinations of both thyroid glands and the cervical regions were performed by either 6–13-MHz (Toshiba Aplio 500) or 5–12-MHz linear transducer (iU22; Philips Medical Systems). US examinations and subsequent US-guided FNAB were performed by one of 4 sonographers (thyroid biopsy specialists) trained to perform US guided FNAB of thyroid nodules at our institute [12] each with > 5 years of experience in thyroid imaging at our institute.

The scanning protocol included both transverse and longitudinal real-time multiplane imaging of each thyroid nodule. Static grey scale as well as color doppler images were saved to PACS. For most thyroid nodules, saved multi-frame cine loops were used in adjunction to static images for sonographic feature characterization. 25 G needles were routinely used for FNA biopsy and 3–4 needle passes were typically made for each nodule. All FNAB were performed under ultrasound guidance using the capillary technique in most instances.

2.3. Image analysis

US features of the thyroid nodules that underwent biopsy were retrospectively recorded by two independent staff radiologists (SG, MM), both with > 15 year experience in thyroid ultrasound, and assigned a risk category as per the 2015 ATA guidelines. Initially, a training session was held to establish a baseline consensus in the lexicon for the US criteria. Both radiologists were blinded to cytopathology results. Table 1 shows the data points collected by each reader. Following independent reads, the data was analyzed by an author not involved in image reads (HM) and the discrepant assigned scores were recorded. A final sonographic pattern for the discrepant reads was decided based on a follow up convergent consensus decision by the two readers.

The sonographic patterns defined in the 2015 ATA guidelines [11] and the assigned patterns for each category are outlined below:

(1) Benign: Purely cystic nodules without solid component. (2) Very Low Suspicion: Spongiform or partially cystic nodules without any of the sonographic features described in low, intermediate or high suspicion patterns. (3) Low Suspicion: Isoechoic or hyperechoic solid nodule, or partially cystic nodule with eccentric solid areas, without microcalcification, irregular margin or extrathyroidal extension, or taller than wide shape. (4) Intermediate Suspicion: Hypoechoic solid nodule with smooth margins without microcalcifications, extrathyroidal extension, or taller than wide shape (Fig. 1). (5) High suspicion: solid hypoechoic nodule or solid hypoechoic component of a partially cystic nodule with one or more of the following features: irregular margins (infiltrative, microlobulated), microcalcifications, taller than wide shape, rim calcifications with small extrusive soft tissue component, evidence of extrathyroidal extension (Fig. 1).

During the initial reads, it was noted that some isoechoic or hyperechoic nodules occasionally had microcalcification or were taller than wide, a pattern not clearly defined in the 2015 ATA guidelines. We classified this pattern into category 5X/3 (Fig. 1), which indicated that it had both the high suspicion pattern (microcalcification) and the low suspicion pattern (isoechoic or hyperechoic). However, for final assessment we included these nodules into the higher risk category. The revised 2015 ATA guidelines also do not differentiate between hypoechogenicity (in comparison to the adjacent normal thyroid parenchyma) and markedly hypoechoic nodules (darker than the strap muscles), and thus for data collection we recorded these nodules as either category 4a (markedly hypoechoic) or 4b (hypoechoic to adjacent thyroid parenchyma) (Fig. 1). Nodules which were heterogeneous and had hypoechoic as well isoechoic areas were included in the intermediate risk category 4, the higher risk category for such nodules. We did not sub classify these nodules into a new category since one of the primary aims of the study was to look at the accuracy of the 2015 ATA guidelines, and also in a clinical setting one would have to decide if the nodule meets criteria for biopsy or follow up. We do however also present the results separately after excluding such nodules from the respective ATA categories [Table 2, columns 5–8].

2.4. Cytology reporting system

Cytology was reported based on the Bethesda System for Reporting Thyroid Cytology [13]. Basically, (I) nondiagnostic or unsatisfactory, (II) benign, (III) atypia of undetermined significance or follicular lesion of undetermined significance, (IV) follicular neoplasm or suspicious for follicular neoplasm, (V) suspicious for malignancy, and (VI) malignant. Categories III to V are considered cytologically indeterminate [14]. For the purpose of estimating the risk of malignancy for each ATA sonographic nodule sonographic pattern, we grouped diagnostic categories V and VI in this study. Malignancy rates based on surgical pathology follow-up at our institute are 99.3% for Category VI, 82% for category V, 26% for category IV, 13.5% for category III, 1.41% for category II and 3.6% for category I, estimated over a time period of 6 years (2009–2015).

2.5. Statistical analysis

Once individual readings for the sonographic features and patterns

Table 1
Sonographic features evaluated in our study, based on the ATA 2015 guidelines.

Sonographic feature (Variable)	Categories
Calcification	Microcalcifications, interrupted rim calcification with soft tissue extrusion, any other forms of calcification
Echogenicity	marked hypoechogenicity, hypoechogenicity, isoechogenicity, hyperechogenicity
Margin	extra-thyroidal extrusion, extra-nodule soft tissue extrusion, irregular margin, regular margin, hypoechoic halo
Shape	taller than wide
Composition	complete solid, partial cystic with suspicious features, partially cystic with eccentric uniform solid area, partial cystic with no suspicious features, spongiform, complete cystic
Lymphadenopathy	Microcalcification, cystic appearance peripheral, vascularity, hyperechogenicity, round shape

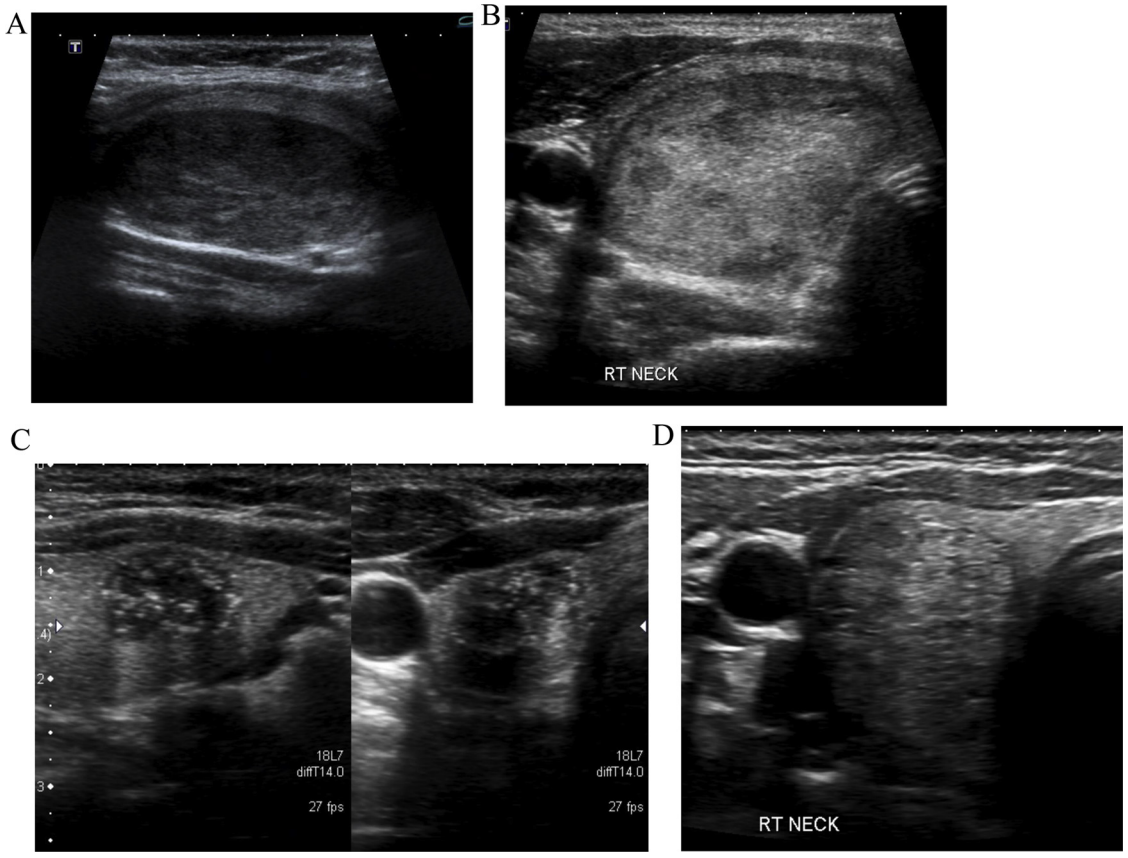

Fig. 1. Sonographic patterns of 4, 4b, 5 and 5x in the study a. Sonographic pattern of Intermediate Suspicion (4): Hypoechoic solid nodule with smooth margins without microcalcifications, extrathyroidal extension, or taller than wide shape. 40-year-old female with papillary thyroid carcinoma. b. Sonographic pattern 4b: Hypoechoic and isoechoic areas to adjacent thyroid parenchyma, solid nodule with smooth margins without microcalcifications, extrathyroidal extension, or taller than wide shape. 48-year-old female with benign thyroid nodule. c. Sonographic pattern of high suspicion (5): solid hypoechoic nodule with microcalcifications. 40-year-old female with papillary thyroid carcinoma. d. Sonographic pattern of 5x: isoechoic nodule with taller-than-wide shape. 36-year-old female with benign thyroid nodule.

were completed, the discordance between the two readers was identified, and consensus reading for those nodules was obtained. Interobserver agreement (kappa value) for each sonographic pattern, as well as those sonographic features that were responsible for the discordance were calculated using Cohen statistic. The guidelines of Landis and Koch were followed for the interpretation of values: 0.00 – 0.20 indicated slight agreement; 0.21– 0.40, fair agreement; 0.41– 0.60, moderate agreement; 0.61– 0.80, substantial and 0.80–1.00, almost perfect [15].

The consensus reports were then compared to surgical pathology reports available from the database, and final analysis of the risk of malignancy was calculated for each sonographic pattern.

3. Results

3.1. General demographic information

189 nodules from 152 patients were included in this study, mean age 55.6 years (range 21.5–88.3), 118 (78.0%) females, 34 males. Of the 189 nodules, 31 were malignant on final cytology or histology (25 documented on surgical pathology, remaining on FNAB) [Table 2]. Of these 31, 10 nodules had diagnostic category VI on FNAB, 7 nodules had category V, 2 had diagnostic category IV, 7 had a benign category II on FNAB and a further 5 were classified into Bethesda category I (unsatisfactory) initially. Papillary carcinoma was identified in 12 category I and II nodules; the reason for thyroidectomy despite unsatisfactory or benign category on FNAB was malignancy on subsequent repeat FNAB (4 lesions), not clearly documented in chart review (3 lesions), incidental finding in ipsilateral lobe on hemithyroidectomy for another nodule (1 lesion), increase in size (1 lesion), microcalcification (1 lesion), high risk of occult malignancy not clearly explained in electronic patient records (1 lesion), and invasive carcinoma identified in contralateral lobe (1 lesion). All 31 malignant nodules in our study were papillary carcinomas.

3.2. Interobserver agreement of the 2015 ATA sonographic patterns

Of the 189 nodules included in the study, 123 nodules (65.1%) were assigned the same final ATA sonographic pattern by both readers [kappa 0.518, (95% confidence interval 0.427 − 0.610)]. The strength of agreement was considered to be moderate. Microcalcification was observed in 27 lesions, 8 of the them were malignant (7 category 5 lesions and one category 5x lesion).

Among the 66 lesions where there was disagreement in the assigned ATA sonographic pattern between the readers, the sonographic features for major discordant read was microcalcification (in 28 nodules, discrepancy 42%), followed by eccentric solid component for partial cystic lesion (17 nodules, 26%), spongiform (14 nodules, 21%), hypoechogenicity (10 nodules, 15.2%), and taller than wide (10 nodules, 15%). Thirty two nodules had a discordant read for more than 2 grades of the ATA sonographic patterns, for instance between high suspicion (5) and low suspicion (3). The most common cause of disagreement in these nodules was microcalcification (18 nodules, 56%), followed by taller than wide (7, 21.9%).

The most common individual sonographic feature for inter-observer discrepancy in our study irrespective of final ATA sonographic pattern,

Table 2
Performance of 2015 ATA sonographic patterns for detection of malignancy.

ATA 2015 sonographic pattern	Bethesda I	II	III	IV	V	VI	Surgical proven malignancy	Malignancy (Bethesda V/VI & surgical proven)	Final Consensus between readers 1 &2	Reader 1 Malignancy rates	Reader 1 Reads	Reader 2 Malignancy rates	Reader 2 Reads	Risk of malignancy (Bethesda V/VI& surgical proven): Final consensus reading	Reader 1	Reader 2	Expected risk of malignancy as per ATA guidelines
Benign 1	0	0	0	0	0	0	0	0	0	0	0	0	2		0.0%	0.0%	< 1%
Very low suspicion 2	12	37	0	0	1	0	1	1	50	2	49	4	51	2%	4.1%	7.8%	< 3%
Low suspicion 3	22	35	0	3	1	2	5	8	63	8	68	8	53	12.7%	11.8%	15.1%	5-10%
Intermediate suspicion 4	8	7	0	2	1	1	4	5	19	8	24	6	22	26.3%	33.3%	27.3%	10-20%
High suspicion 5	12	27	4	3	4	7	15	17	57	13	48	13	61	29.8%	27.1%	21.3%	70-90%
Intermediate suspicion 4 excluding 4b	5	5	0	1	1	1	4	5	13	8	17	1	2	38.5%	47.1%	50.0%	10-20%
Low suspicion 3 plus 4b	25	37	0	4	1	2	5	8	69	8	75	13	73	11.6%	10.7%	17.8%	5-10%
High suspicion 5 excluding 5x	12	27	4	2	5	6	14	16	51	8	28	12	43	31.4%	28.6%	27.9%	70-90%
Low suspicion 3 plus 5x	22	40	0	4	1	2	6	9	69	13	88	9	71	13.0%	14.8%	12.7%	5-10%
Total	54	106	4	8	7	10	25	31	189								

Table 3
Interobserver Variability in Thyroid Nodule Descriptors.

	% agreement between raters	Kappa	Interpretation
Microcalcifications	79% (151/190)	0.393	Fair
Interrupted rim calcification	98% (186/190)	0.327	Fair
Calcification (other types)	92% (174/189)	0.608	Substantial
Markedly hypoechoic	93% (175/189)	0.330	Fair
Hypoechoic nodule	78% (147/189)	0.529	Moderate
Isoechoic	60% (114/190)	0.242	Fair
Hyperechoic	80% (150/188)	0.307	Fair
Soft tissue extrusion beyond gland	97% (184/190)	0.243	Fair
Irregular nodule margin	78% (149/189)	0.072	Slight
Regular nodule margin	80% (152/190)	0.141	Slight
Hypoechoic halo around nodule	76% (145/190)	0.497	Moderate
Taller than wide	87% (165/189)	0.282	Fair
Complete solid	83% (158/190)	0.636	Substantial
Partially cystic with suspicious features	85% (159/187)	0.105	Slight
Partially cystic with eccentric uniform solid area	88% (168/190)	0.540	Moderate
Partially cystic no suspicious features	85% (162/190)	0.168	Slight
Spongiform	86% (163/190)	0.617	Substantial
Final ATA score		**0.518**	**Moderate**

were nodule margin and partially cystic nodules with or without suspicious features (slight agreement) [Table 3].

3.3. Accuracy of the 2015 ATA sonographic patterns

Diagnostic performances of the sonographic patterns are shown in Table 2. The observed risks of malignancy for the sonographic patterns were calculated. There was no benign ATA sonographic pattern observed in our biopsy cohort. The observed malignancy rate for the low suspicion and intermediate suspicion patterns were consistent with the expected risk of malignancy published in the 2015 ATA guidelines [11]. There were 57 high suspicion pattern lesions, of which 17 (29.8%) were malignant or suspicious for malignancy. The accuracy for the two readers independently was 27.1% and 21.3%, which increased to 29.8% following the consensus read.

There were 6 lesions associated with a 4b pattern (hypoechoic to adjacent thyroid parenchyma but not to strap muscle). One of them was a Hurthle cell lesion, no malignancy. Moving these 6 nodules to Category 3 would change accuracy for pattern 4 from 26.3%–38.5% and pattern 3 from 12.7%–11.6%. There were 6 lesions associated with a 5X pattern (questionable high suspicion feature in a low suspicion pattern), with one multifocal papillary carcinoma on surgery, observed risk of malignancy 16.7% in this pattern. Moving these lesions to Category 3, since they were not hypoechoic, would change accuracy of Category 5–31.4% and Category 3–13.0%. The accuracy of category 5 for the 2 readers independently would change to 28.6% and 27.9% respectively.

4. Discussion

Because of the high prevalence of thyroid nodules, it is essential to establish a reliable and cost-effective guideline for managing thyroid nodules based on their ultrasound features. In this retrospective study, we assessed the accuracy of the 2015 ATA guidelines for detection of malignancy. We found that the calculated risk of malignancy for each sonographic pattern was consistent with the estimated risk published in the 2015 ATA guidelines, except for the high suspicion pattern. Of the 57 lesions with high suspicion category ultrasound features, 17 (29.8%) were malignant; while the 2015 ATA guidelines suggest an expected malignancy rate of 70–90%. If we were to exclude nodules which were not hypoechoic from within this group, the diagnostic performance increases to 31.4%, but it is still lower than the expected risk of

malignancy. Similarly, a lower than expected risk of malignancy (54.7%) for the high suspicion pattern was recently published by Rosario et al. [16]. Ha et al. [17] recently reported a 76.6% calculated risk of malignancy for high suspicion nodules, which is in the expected range as per the 2015 ATA guidelines. A high percentage of unsatisfactory FNA from the lesions of high suspicion pattern in our database may have attributed to the finding. In our preliminary analysis of the entire cohort of 648 lesions, 193 (29.9%) nodules had an unsatisfactory FNA result. 26 of these lesions were proven malignant by surgical pathology. The risk of malignancy in this unsatisfactory (Bethesda I) group was 13.5%, which is higher than the published risk of 1–4% [13]. This high rate of unsatisfactory biopsy may be related to the experience of the FNA performer, absence of cytology technologist during biopsy, and stringent satisfactory criteria applied by reporting cytologists. Our general practice for such patients is to offer repeat FNA biopsy in 3–6 months.

Ultrasound evaluation is subjective and therefore can lead to variations in accuracy between examiners. There is limited data looking at the effect of inter-observer variations in the ultrasound assessment of thyroid nodules. In our study, the interobserver agreement for the 2015 ATA sonographic patterns between two experienced independent readers was moderate (kappa 0.518). Though good interobserver agreement (κ 0.72) for the 2015 ATA sonographic patterns for medullary thyroid carcinoma was published in 2016 [18], there has been no prior study looking at this only for papillary thyroid cancers which constitute over 90% of thyroid cancers. Grani G et al reported a wide variability between readers in description of individual sonographic features, but substantial agreement for indication to biopsy based on 5 different classification systems for biopsy (AACE/ACE/AME, ACR, ATA, EU-TIRADS and K-TIRADS [19]. Park et al [20] reported moderate interobserver agreement (κ 0.55) for the Thyroid Study Group of the Korean Society of Neuroradiology and Head and Neck Radiology guidelines. Choi et al also showed a moderate degree of agreement (κ 0.54) between readers for one of the four final assessment categories (benign, probably benign, low suspicious malignancy and suspicious malignancy) [21]. More recently, Hoang et al also reported moderate agreement among readers (κ 0.51) in final category assignment when applying ACR-TIRADS guidelines [22]. This degree of reproducibility is comparable to that reported for the Breast Imaging Reporting and Data System (BI-RADS). Reported k values for overall BI-RADS categories have included 0.45 for mammography [23,24], and 0.21–0.28 for breast US [24]. In addition, one study reported a k value of 0.45 amongst expert readers for the Liver Imaging Reporting and Data System (LI-RADS) [25], while for Thyroid Imaging Reporting and Data System (TI-RADS), reported k values have been around 0.61 [26]. Therefore, while our study shows considerable inter-reader variation using the ATA 2015 guidelines, it is consistent with the published data for other cancer imaging reporting lexicons that are currently in place.

The most common sonographic feature for inter-observer discrepancy in our study was nodule margin and partially cystic nodules with or without suspicious features (slight agreement) followed by microcalcification and degree of hypoechogenicity (fair agreement). Most other studies have also shown only slight to fair agreement on nodule margin [20,22,27,28]. One of the reasons for the lack of agreement in assessment of nodule margin may be inaccurate classification of poorly defined or ill-defined margin as irregular. Differentiation between microcalcifications and echogenic microcystic wall interfaces may be difficult which can lead to misclassification of the grade of sonographic pattern between Spongiform (very low risk pattern) and High Suspicion (5 lesions in this study). Only few other studies have separately analyzed inter-observer agreement on microcalcification / echogenic punctate foci from other types of calcification in nodules. Similar to our results, Hoang et al also showed fair interobserver agreement for microcalcification [22]. For this reason, the ATA guidelines have provisions to sample nodules interpreted as spongiform when over 2 cm in size.

Unlike the 2015 ATA guidelines, the 2016 American College of Endocrinology (ACE) guidelines divides hypoechogenicity into two grades - mild hypoechogenicity relative to the surrounding thyroid parenchyma, considered an intermediate risk feature, while marked hypoechogenicity relative to strap muscles, a high-risk feature. Based on previously published literature, we divided the intermediate suspicion pattern within the 2015 ATA guidelines into two subcategories: 4a with marked hypoechogenicity and 4b with mild hypoechogenicity. There were six 4b lesions, none of which were malignant (Category VI) or suspicious (Category V) for malignancy on cytology. However, the low number of cases in this subgroup limited statistical analysis.

In the 2015 ATA guidelines, echogenicity was used as a major feature to classify nodules into different categories. Solid hypoechoic nodules with at least one associated malignant feature were classified as high suspicion pattern; while hypoechoic nodules without malignant feature(s) were classified as intermediate suspicion pattern. Microcalcification was considered as a high suspicion pattern, and was not included in any other patterns. However, we observed that microcalcification can occasionally co-exist with isoechoic or hyperechoic nodules. In our review, we subclassified this pattern within the high risk category as 5X (atypical for ATA pattern 5). One of the six nodules in this subcategory was malignant (risk 16.7%), which is similar to the expected risk of intermediate suspicion pattern (10–20%). Again, statistical analysis is limited by the small number of nodules in this review.

In the recently published 2016 AACE (American Association of Clinical Endocrinologists) guidelines, increased stiffness at elastography is included as an intermediate risk feature of malignancy (expected risk of malignancy 5–15%) [8]. However, elastography was not included for assessment of thyroid nodules in the ATA guidelines because of various reasons. While the technique holds promise to non-invasively assess cancer risk, it is variable and operator dependent, and requires an elastography module which often must be purchased separately. At the same time it can be effectively applied only to solid nodules, thus excluding its assessment for cystic or partially cystic nodules. The nodule must not overlap with other nodules in the anterioposterior plane for proper evaluation with elastography. Intranodular vascularity was also not assessed in this study. A recent study did not show intranodular vascularity to have an independent predictive value in detecting malignancy in thyroid nodules, while a meta-analysis has shown higher proportions of follicular thyroid cancer with intranodular vascularity [29,30]. Intranodular vascularity was considered a suspicious sonographic feature in the 2009 ATA guidelines, but excluded from the 2015 ATA guidelines [29,30]. We did not evaluate these two features in our study.

Macrocalcification is not included in the 2015 ATA guidelines as a risk of malignancy. Some publications have shown that macrocalcifications within a nodule, if combined with microcalcifications, confer the same malignancy risk as microcalcifications alone [31–33], and that macrocalcification alone is not consistently associated with thyroid cancer. On the other hand, macrocalcifications are commonly present in multinodular goiters; however, studies have shown that when found in solitary nodules, they may be associated with a malignancy rate of nearly 75% [34]. They are the most common type of calcification in medullary thyroid carcinomas [35]. In our study, macrocalcification was observed in 14 nodules, 4 of them were associated with microcalcification, and one with interrupted rim calcification. Of the remaining 9 nodules with macro-calcification, 3 were malignant or suspicious for malignancy on FNAB (33.3%), though none were medullary cancers.

Though not included in the 2015 ATA sonographic patterns, we collected data on presence or absence of hypoechoic halo at the margin of the nodule. The halo is produced by a pseudocapsule of fibrous connective tissue, compressed thyroid parenchyma, or chronic inflammatory infiltrates [36]. Studies have shown that a completely uniform halo around a nodule is suggestive of benignity. But it has a low specificity and may be absent in more than half of all benign

thyroid nodules [37]. In the 2016 AACE guidelines, a halo was considered as an indeterminate feature that was associated with isoechoic or hyperechoic nodules, a pattern that can be seen in low risk lesions [8]. Similarly, a halo with an isoechoic or mildly hyperechoic nodule was considered a benign pattern in the 2015 BTA guidelines [38]. The halo sign was present in 54 nodules in our study, of which 3 (5.5%) had malignant cytology.

The present study has several limitations. The main limitation of our study is its retrospective design and small cohort of patients from a single center. Thyroid biopsy registry was started immediately after the sonographers were trained to perform biopsy independently under supervision. This may have contributed to the relatively large number of unsatisfactory cytology results in the study. The unsatisfactory rates have significantly improved over the last few years, and have consistently been < 10% since 2014. In addition, not all malignant nodules in our study were surgically proven. This may have further lead to underestimation of malignant nodules.

All malignancies in our study were papillary carcinoma, without any follicular carcinoma. This can be secondary to the patient population and selection criteria for biopsy and thyroidectomy in our institute, or as a result of the small cohort of patients. A significant number of lesions with a single unsatisfactory or benign biopsy result did not have follow-up > 1 year and therefore were excluded from our study. Not all of these nodules may truly be benign.

In conclusion, when applied in a clinical setting, our current data suggests that the risk of malignancy in the high suspicion sonographic pattern of the 2015 ATA guidelines may have been over estimated, though the risk of malignancy in the other categories appears to be accurate. The new guideline shows moderate inter-observer agreement, which is an important factor for diagnostic performance. This study provides important complementary data in assessing thyroid ultrasound features according to the ATA guidelines and may aid in revision of the ATA guidelines in the future.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] D.S. Dean, H. Gharib, Epidemiology of thyroid nodules, Best Pract. Res. Clin. Endocrinol. Metab. 22 (December 6) (2008) 901–911.
[2] B.Y. Cho, et al., Changes in the clinicopathological characteristics and outcomes of thyroid cancer in Korea over the past four decades, Thyroid 23 (July 7) (2013) 797–804.
[3] K. Bibbins-Domingo, et al., Screening for thyroid Cancer: us preventive services task force recommendation statement, JAMA 317 (May 18) (2017) 1882–1887.
[4] J.P. Brito, L. Davies, Is there really an increased incidence of thyroid cancer? Curr. Opin. Endocrinol. Diabetes Obes. 21 (October 5) (2014) 405–408.
[5] J.P. Brito, A. Al Nofal, V.M. Montori, I.D. Hay, J.C. Morris, The impact of subclinical disease and mechanism of detection on the rise in thyroid Cancer incidence: a population-based study in Olmsted County, Minnesota during 1935 through 2012, Thyroid 25 (September 9) (2015) 999–1007.
[6] J.K. Hoang, W.K. Lee, M. Lee, D. Johnson, S. Farrell, US Features of thyroid malignancy: pearls and pitfalls, Radiographics 27 (3) (2007) 847–860 discussion 861-5, 2007 May-Jun.
[7] E.K. Kim, et al., New sonographic criteria for recommending fine-needle aspiration biopsy of nonpalpable solid nodules of the thyroid, AJR Am. J. Roentgenol. 178 (March 3) (2002) 687–691.
[8] H. Gharib, et al., American association of clinical endocrinologists, American college of endocrinology, and Associazione Medici Endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules–2016 update, Endocr. Pract. 22 (May5) (2016) 622–639.
[9] M.C. Frates, et al., Management of thyroid nodules detected at US: society of Radiologists in Ultrasound consensus conference statement, Radiology 237 (December 3) (2005) 794–800.
[10] F.N. Tessler, et al., ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS committee, J. Am. Coll. Radiol. 14 (May 5) (2017) 587–595.
[11] B.R. Haugen, 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 26 (January 1) (2016) 1–133.
[12] S. Ghai, et al., Thyroid biopsy specialists: a quality initiative to reduce wait times and improve adequacy rates, Radiology 276 (September 3) (2015) 894–899.
[13] I.V. Renuka, G. Saila Bala, C. Aparna, R. Kumari, K. Sumalatha, The Bethesda system for reporting thyroid cytopathology: interpretation and guidelines in surgical treatment, Indian J. Otolaryngol. Head Neck. Surg. 64 (December 4) (2012) 305–311.
[14] C.G. Theoharis, K.M. Schofield, L. Hammers, R. Udelsman, D.C. Chhieng, The Bethesda thyroid fine-needle aspiration classification system: year 1 at an academic institution, Thyroid 19 (November 11) (2009) 1215–1223.
[15] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (March 1) (1977) 159–174.
[16] P.W. Rosario, A.L. da Silva, M.S. Nunes, M.A. Ribeiro Borges, G.F. Mourão, M.R. Calsolari, Risk of malignancy in 1502 solid thyroid nodules &1 cm using the new ultrasonographic classification of the American Thyroid Association, Endocrine 56 (May 2) (2017) 442–445.
[17] E.J. Ha, D.G. Na, J.H. Baek, J.Y. Sung, J.H. Kim, S.Y. Kang, US fine-needle aspiration biopsy for thyroid malignancy: diagnostic performance of seven society guidelines applied to 2000 thyroid nodules, Radiology (February) (2018) 171074.
[18] P. Valderrabano, et al., New American thyroid association sonographic patterns for thyroid nodules perform well in medullary thyroid carcinoma: institutional experience, systematic review, and meta-analysis, Thyroid 26 (August 8) (2016) 1093–1100.
[19] G. Grani, L. Lamartina, V. Cantisani, M. Maranghi, P. Lucia, C. Durante, Interobserver agreement of various thyroid imaging reporting and data systems, Endocr. Connect 7 (January 1) (2018) 1–7.
[20] C.S. Park, et al., Observer variability in the sonographic evaluation of thyroid nodules, J. Clin. Ultrasound 38 (July 6) (2010) 287–293.
[21] S.H. Choi, E.K. Kim, J.Y. Kwak, M.J. Kim, E.J. Son, Interobserver and intraobserver variations in ultrasound assessment of thyroid nodules, Thyroid 20 (February 2) (2010) 167–172.
[22] J.K. Hoang, et al., Interobserver variability of sonographic features used in the American college of radiology thyroid imaging reporting and data system, AJR Am. J. Roentgenol. 211 (July 1) (2018) 162–167.
[23] W.A. Berg, et al., Does training in the Breast Imaging Reporting and Data System (BI-RADS) improve biopsy recommendations or feature analysis agreement with experienced breast imagers at mammography? Radiology 224 (September 3) (2002) 871–880.
[24] N. Abdullah, B. Mesurolle, M. El-Khoury, E. Kao, Breast imaging reporting and data system lexicon for US: interobserver agreement for assessment of breast masses, Radiology 252 (September 3) (2009) 665–672.
[25] M.S. Davenport, et al., Repeatability of diagnostic features and scoring systems for hepatocellular carcinoma by using MR imaging, Radiology 272 (July 1) (2014) 132–142.
[26] S.P. Cheng, J.J. Lee, J.L. Lin, S.M. Chuang, M.N. Chien, C.L. Liu, Characterization of thyroid nodules using the proposed thyroid imaging reporting and data system (TI-RADS), Head Neck 35 (April 4) (2013) 541–547.
[27] S.J. Park, et al., Interobserver variability and diagnostic performance in US assessment of thyroid nodule according to size, Ultraschall Med. 33 (December 7) (2012) E186–E190.
[28] J.R. Wienke, W.K. Chong, J.R. Fielding, K.H. Zou, C.A. Mittelstaedt, Sonographic features of benign thyroid nodules: interobserver reliability and overlap with malignancy, J. Ultrasound Med. 22 (October 10) (2003) 1027–1031.
[29] J.P. Brito, et al., The accuracy of thyroid nodule ultrasound to predict thyroid cancer: systematic review and meta-analysis, J. Clin. Endocrinol. Metab. 99 (April 4) (2014) 1253–1263.
[30] H.J. Moon, J.Y. Kwak, M.J. Kim, E.J. Son, E.K. Kim, Can vascularity at power Doppler US help predict thyroid malignancy? Radiology 255 (April 1) (2010) 260–269.
[31] A. Kamaya, G.H. Lewis, Y. Liu, H. Akatsu, C. Kong, T.S. Desser, Atypia of undetermined significance and follicular lesions of undetermined significance: sonographic assessment for prediction of the final diagnosis, J. Ultrasound. Med. 34 (May 5) (2015) 767–774.
[32] E. Papini, et al., Risk of malignancy in nonpalpable thyroid nodules: predictive value of ultrasound and color-Doppler features, J. Clin. Endocrinol. Metab. 87 (May 5) (2002) 1941–1946.
[33] H.J. Moon, J.M. Sung, E.K. Kim, J.H. Yoon, J.H. Youk, J.Y. Kwak, Diagnostic performance of gray-scale US and elastography in solid thyroid nodules, Radiology 262 (March 3) (2012) 1002–1013.
[34] M.L. Khoo, et al., Underexpression of p27/Kip in thyroid papillary microcarcinomas with gross metastatic disease, Arch. Otolaryngol. Head Neck Surg. 128 (March 3) (2002) 253–257.
[35] S. Takashima, H. Fukuda, N. Nomura, H. Kishimoto, T. Kim, T. Kobayashi, Thyroid nodules: re-evaluation with ultrasound, J. Clin. Ultrasound 23 (3) (1995) 179–184 Mar-Apr 1995.
[36] R.A. Propper, M.L. Skolnick, B.J. Weinstein, A. Dekker, The nonspecificity of the thyroid halo sign, J. Clin. Ultrasound 8 (April 2) (1980) 129–132.
[37] D.A. Watters, et al., Role of ultrasound in the management of thyroid nodules, Am. J. Surg. 164 (December 6) (1992) 654–657.
[38] P. Perros, et al., Guidelines for the management of thyroid cancer, Clin. Endocrinol. (Oxf.) 81 (July Suppl 1) (2014) 1–122.