

Thyroid Nodules with Indeterminate Cytology: Utility of the American Thyroid Association Sonographic Patterns for Cancer Risk Stratification

Pablo Valderrabano,¹ Melissa J. McGettigan,^{2,*} Cesar A. Lam,^{2,*} Laila Khazai,³
Zachary J. Thompson,⁴ Christine H. Chung,¹ Barbara A. Centeno,³ and Bryan McIver¹

Background: The 2015 American Thyroid Association (ATA) guidelines recommend using a classification based on sonographic patterns to set the size threshold for biopsies. Each pattern is associated with a distinct estimated rate of malignancy that it was hypothesized should stratify the risk of malignancy of cytologically indeterminate thyroid nodules (ITNs).

Methods: Ultrasound images of 463 ITNs (38% atypia/follicular lesions of undetermined significance; 62% follicular neoplasms) with histological follow-up consecutively evaluated between October 2008 and June 2015 at the authors' academic cancer center were independently evaluated by three observers and classified into one of the five sonographic patterns proposed by the ATA. Nodules with sonographic patterns not defined in the classification were grouped into a non-ATA pattern category. Differences in clinical and histological findings between the sonographic patterns were assessed. The prevalence of malignancy and odds ratio for malignancy were calculated for each sonographic pattern (low and intermediate patterns were collapsed for the analysis).

Results: The distribution of size and cytological diagnosis was significantly different between sonographic patterns ($p < 0.001$). The overall rate of malignancy was 27%. The rate of malignancy for the very low, low/intermediate, high, and non-ATA patterns were 0%, 19%, 56%, and 36%, respectively, and were all significantly different. Compared to the low/intermediate suspicion patterns, the odds ratios for malignancy were 2.35 for the non-ATA and 5.18 for the high suspicion patterns ($p < 0.001$). The odds ratio of the non-ATA pattern was 0.45 over the high suspicion pattern ($p = 0.04$). Results were similar in both cytological categories and for each observer separately. Sonographic patterns were associated with distinct histopathological profiles ($p < 0.001$).

Conclusions: ATA sonographic patterns are associated with distinct clinical features and pathological outcomes, and effectively stratify the cancer risk in ITNs. Thus, the ATA sonographic patterns should be used not only to set the size threshold for biopsy, but also to personalize management after the biopsy.

Keywords: thyroid ultrasound, thyroid cytology, thyroid nodules, thyroid cancer, noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP)

Introduction

THYROID ULTRASOUND HAS a central role in the evaluation and management of thyroid nodules (1–4). The 2015 American Thyroid Association (ATA) guidelines for the management of patients with thyroid nodules and differentiated thyroid cancer recommend using a classification based on the sonographic pattern to select nodules that need to be biopsied (1). This classification is integrated by five so-

nographic patterns—benign, very low, low, intermediate, and high suspicion—each of which is associated with a distinct estimated risk of malignancy (<1%, <3%, 5–10%, 10–20%, and 70–90%, respectively) and a different size threshold for biopsy (no biopsy, >2 cm, >1.5 cm, >1 cm, and >1 cm, respectively) (1).

Once a thyroid nodule is biopsied, management relies primarily on the cytological diagnosis. If cytology is benign, the nodule is usually observed; if it is malignant or suspicious

Departments of ¹Head and Neck—Endocrine Oncology, ²Diagnostic Imaging, ³Anatomic Pathology, and ⁴Biostatistics and Bioinformatics, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida.

*These authors contributed equally to this work.

for malignancy, the nodule is usually resected (1–3). However, 20% of the biopsies render an indeterminate cytology: either atypia/follicular lesion of undetermined significance (AUS/FLUS) or follicular/Hürthle cell neoplasm (FN/HCN) (5). Although indeterminate categories of thyroid cytology encompass a very heterogeneous group of nodules, they are usually clustered for management recommendations (1,6).

Because the pretest probability influences the posttest probability of any diagnostic test, it was hypothesized that different ATA sonographic patterns associated with different rates of malignancy should be able to stratify the risk of malignancy of cytologically indeterminate thyroid nodules (ITNs; AUS/FLUS and FN/HCN). If the hypothesis held true, the ATA sonographic patterns would be useful not only to set the size threshold for biopsy, but also to personalize management after an indeterminate cytological result.

Methods

Study cohort

The charts of the patients with 3325 thyroid nodules with consecutive cytological evaluation between October 2008 and April 2015 at the authors' institution were retrospectively reviewed. Of those, 861 were AUS/FLUS or FN/HCN. The study included all nodules resected before June 2015, with clear cytological–histological correlation (i.e., the biopsied nodule matched the resected specimen by tumor size and location) and presurgical thyroid ultrasound images available for review (Fig. 1). In this retrospective Institutional Review Board–approved study, a waiver of consent was granted.

Sonographic evaluation

Images of all thyroid ultrasounds were independently reviewed by three observers: one endocrinologist and two radiologists. A list with the medical record number, date of last thyroid ultrasound before surgery, nodule location, and tridimensional size was facilitated to ensure evaluation of the same nodule in the same exam by all observers. Presurgical neck ultrasound was also reviewed when available. Composition, echogenicity, margins, calcifications, extrathyroidal

extension, and presence of suspicious lymph nodes (when cervical ultrasound was available) were assessed in all nodules blinded to the final histological results. The following were considered suspicious sonographic features: hypoechoogenicity; irregular, microlobulated, or infiltrative margins; microcalcifications or interrupted rim calcifications; presence of extrathyroidal extension; presence of suspicious lymph nodes; and shape taller than wide in the transverse view. Echogenicity was classified into four categories: hypoechoogenicity, isoechogenicity, hyperechogenicity, and heterogeneous echogenicity (heteroechogenicity) with respect to the normal thyroid parenchyma. Composition was classified into four categories: solid (or predominantly solid), mixed, cystic, or spongiform. Margin was classified into four categories: regular, irregular, microlobulated, and infiltrative. Calcifications were classified into: none; microcalcifications; macrocalcifications; micro- and macrocalcifications; rim calcifications; interrupted rim calcifications; and comet tails. Extrathyroidal extension and suspicious lymph nodes were classified into two groups: present or absent. The shape of the nodule was classified as taller than wide or wider than tall in the transverse view based on the tridimensional measurements given to locate the nodule. Thus, the interpretation of this feature was the same for all observers. It was considered taller than wide in the transverse view when the measurement difference was ≥ 2 mm to avoid small artifactual differences.

Nodules were classified into the 2015 ATA sonographic patterns using the description of the sonographic features (1). Spongiform or partly cystic nodules without eccentric solid areas or other suspicious features were classified as very low suspicion sonographic pattern. Nodules interpreted as pseudonodules during imaging review (one nodule overall) were also considered very low suspicion pattern and were grouped with this category for analysis. Iso- or hyperechoic solid or partly cystic nodules with eccentric solid areas without suspicious features were classified as low suspicion pattern. Hypoechoic solid nodules without other suspicious features were classified as intermediate suspicion pattern, whereas in the presence of at least one other suspicious feature they were classified as high suspicion pattern. Other scenarios not described in the 2015 ATA classification were grouped into a separate category that will be referred to as non-ATA patterns. These scenarios included heteroechoic nodules with or without other suspicious features and iso- or hyperechoic nodules with at least one suspicious feature.

The final interpretation of each sonographic feature and of the sonographic pattern was done by consensus of two or more observers. When there was no consensus (at least two observers) through the independent assessment, the image was reviewed jointly and a consensus reached.

Pathological evaluation

The cytological diagnosis of all specimens was done by board-certified cytologists at the authors' institutional Department of Anatomic Pathology. All cytological diagnosis followed the 2009 Bethesda System for Reporting Thyroid Cytopathology (7). Histological diagnosis was issued by board-certified pathologists at the Department of Anatomic Pathology with experience in head and neck pathology in 451 (97%) nodules, and was retrieved from external reports in another 12 (3%). Histology was reviewed blinded to

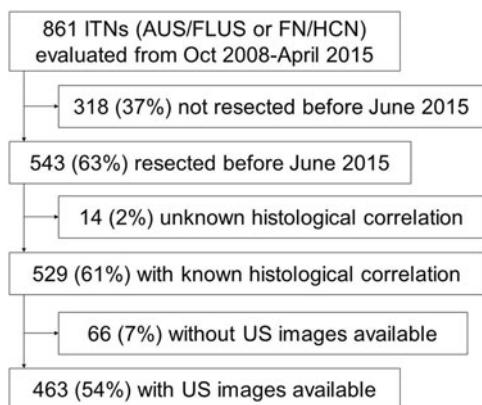


FIG. 1. Cohort selection process. AUS/FLUS, atypia/follicular lesion of undetermined significance; FN/HCN, follicular/Hürthle cell neoplasm; ITNs, cytologically indeterminate thyroid nodules (AUS/FLUS and FN/HCN); US, thyroid ultrasound.

ultrasound characteristics in all cases with malignant diagnosis, with slides available for review (87%; 110/126). Follicular variant papillary thyroid carcinomas (FVPTCs) were reclassified as noninvasive follicular thyroid neoplasms with papillary-like nuclear features (NIFTPs) or as conventional variant of papillary thyroid carcinoma (CVPTC) as appropriate (8). The last cytological diagnosis was used for cytological-histological correlation in nodules with multiple biopsies. Resected nodules with an unclear cytology-histology correlation were excluded from the study.

Malignancies were classified into three groups according to their histological features. Low-risk cancers were defined as completely excised intrathyroidal T1-T2 differentiated thyroid carcinomas, without vascular invasion (<4 foci for minimally invasive follicular thyroid carcinomas [FTCs]), clinical N0 (includes patients with ≤ 5 lymph node metastases all ≤ 2 mm), and without distant metastasis. High-risk cancers were differentiated thyroid carcinomas with any of the following: gross extrathyroidal extension, lymph node metastases >1 cm or in the lateral compartments, or distant metastases. Medullary thyroid carcinomas were also considered high-risk cancers. All other thyroid malignancies were considered intermediate-risk.

Statistical analysis

All analyses were performed using SAS v9.4 (SAS Institute, Cary, NC) and R v3.3.1 (The R Foundation for Statistical Computing, Vienna, Austria). Descriptive statistics and percentages are presented in the tables. Comparisons were performed using chi-square tests and Fisher's exact tests for categorical variables. The Van der Waerden one-way analysis test was used to compare means of continuous measures. Odds ratios (ORs) and associated confidence intervals (CIs) were calculated from the contingency tables and could not be calculated for the very low suspicion sonographic pattern due to absence of events. *p*-Values were adjusted for multiple comparisons with Holms' method. They are two-sided and were considered statistically significant at the 0.05 level.

NIFTPs are considered premalignant lesions and thus surgical disease. However, these tumors do not fit well into the binary benign versus malignant histology. Thus, rates of malignancy and ORs were calculated in duplicate, with NIFTPs considered benign and with NIFTPs considered malignant.

Results

Study cohort

A total of 463 ITNs in 415 patients were included in this study. The mean age at the time of biopsy was 53 years, and 76% ($n=317$) were female (Table 1). Thyroid function was unknown in 40 (10%) patients at the time of biopsy. In the other 375 patients, thyroid function was normal (thyrotropin [TSH] within reference range) in 288 (77%) patients; 69 (18%) were hypothyroid (TSH above reference range, or patient on levothyroxine treatment); and 18 (5%) were hyperthyroid (TSH below reference range, or patient on anti-thyroid drugs).

The sonographic pattern was classified as very low suspicion in 25 (5%), low suspicion in 159 (34%), intermediate

TABLE 1. PATIENT-LEVEL BASELINE CHARACTERISTICS, *N* (%)

<i>Patient level</i>	<i>All, n=415</i>
Age, mean \pm standard deviation, years	53 \pm 14
Female sex	317 (76)
Thyroid function ($n=375$)	
Euthyroid	288 (77)
Hypothyroid	69 (18)
Hyperthyroid	18 (5)
Multinodular goiter ($n=414$)	269 (65)
Contralateral nodules >1 cm ^a ($n=414$)	136 (33)
Other cancer diagnosis	126 (30)
Family history of thyroid cancer	23 (6)
Past radiation exposure to the neck	9 (2)
Syndrome associated with thyroid cancer	0

Percentages are calculated for 415 patients, unless a different number of patients with information available is specified in parentheses next to the variable.

^aIncludes patients with previous contralateral lobectomy.

suspicion in 74 (16%), high suspicion in 36 (8%), and non-ATA in 169 (37%). The mean nodule size, determined by the largest dimension in ultrasound, was 2.6 cm (Table 2). Five percent of the nodules were <1 cm, 34% were between 1 and 2 cm, 46% were between 2 and 4 cm, and 15% were ≥ 4 cm. There were significant differences ($p < 0.001$) in nodule size by sonographic pattern with smaller nodules among intermediate and high suspicion sonographic patterns, and larger nodules among non-ATA patterns. Cytological diagnosis was AUS/FLUS in 176 (38%) and FN/HCN in 287 (62%). There was also a significantly different ($p < 0.001$) distribution of cytological diagnosis between the patterns, with most (84%) very low suspicion pattern nodules being AUS/FLUS and most (76%) intermediate suspicion pattern nodules being FN/HCN.

Differences in prevalence and risk of malignancy between sonographic patterns

The prevalence of malignancy was 27% ($n=126$), falling to 17% ($n=78$) if NIFTPs were not considered malignant (Table 2). Differences in the prevalence of malignancy were not statistically significant between AUS/FLUS and FN/HCN specimens, when NIFTPs were considered either malignant (26% vs. 28%; $p=0.66$) or benign (18% vs. 16%; $p=0.67$). The prevalence of malignancy was not significantly different between the low and intermediate suspicion patterns with NIFTPs considered either malignant (20% vs. 18%; $p=0.65$) or benign (11% vs. 11%; $p=0.98$). For that reason, both patterns were collapsed for the comparisons with other groups. All other comparisons of the prevalence of malignancy between the sonographic patterns considering NIFTPs malignant were statistically significant (Table 3). Most NIFTPs had a low/intermediate suspicion or non-ATA sonographic pattern. Therefore, when considering NIFTPs benign, the prevalence of malignancy in those groups dropped by nearly half, whereas it remained almost unchanged in the very low and high suspicion sonographic patterns. In this scenario, all differences in the prevalence of malignancy remained significant except between the very low and low/intermediate sonographic

TABLE 2. NODULE LEVEL CHARACTERISTICS, N (%)

Sonographic pattern ^a	All, 463	Very low, 25 (5)	Low, 159 (34)	Intermediate, 74 (16)	High, 36 (8)	Non-ATA, 169 (37)	p-Value
US size, mean ± SD (range)	2.6 ± 1.5 (0.4–10)	2.2 ± 1 (0.5–6.1)	2.6 ± 1.3 (0.4–7.9)	1.8 ± 0.8 (0.5–4.4)	2.1 ± 0.9 (0.6–3.8)	3.2 ± 1.7 (0.8–10)	<0.001 ^b
<1 cm	23 (5)	1 (4)	5 (3)	8 (11)	4 (11)	5 (3)	
1–1.9 cm	158 (34)	10 (40)	57 (36)	41 (55)	12 (33)	38 (22)	
2–3.9 cm	213 (46)	13 (52)	73 (46)	24 (32)	20 (56)	83 (49)	
≥4 cm	69 (15)	1 (4)	24 (15)	1 (1)	0	43 (25)	
AUS/FLUS	176 (38)	21 (84)	55 (35)	18 (24)	16 (44)	66 (39)	<0.001
FN/HCN	287 (62)	4 (16)	104 (65)	56 (76)	20 (56)	103 (61)	
Cancer (w/ NIFTP)	126 (27)	0	32 (20)	13 (18)	20 (56)	61 (36)	<0.001
AUS/FLUS	45 (26)	0	9 (16)	3 (17)	11 (69)	22 (33)	<0.001
FN/HCN	81 (28)	0	23 (22)	10 (18)	9 (45)	39 (38)	<0.001
Cancer (w/o NIFTP)	78 (17)	0	17 (11)	8 (11)	17 (47)	36 (21)	<0.001
AUS/FLUS	31 (18)	0	5 (9)	3 (17)	11 (69)	12 (18)	0.003
FN/HCN	47 (16)	0	12 (12)	5 (9)	6 (30)	24 (23)	0.01

^aSonographic pattern as characterized by at least two of the observers.

^bVan der Waerden one-way analysis (normal scores test).

AUS/FLUS, atypia/follicular lesion of undetermined significance; FN/HCN, follicular/Hürthle cell neoplasm; Interm., Intermediate suspicion sonographic pattern; Non-ATA, see description in Methods section; SD, standard deviation; US size, largest dimension in presurgical ultrasound evaluation.

patterns (0% vs. 11%; $p=0.15$). Similar findings were observed for AUS/FLUS and FN/HCN specimens when analyzed separately, although some comparisons did not reach statistical significance.

Compared to the low/intermediate suspicion patterns, the risk of malignancy of nodules was 1.35 (OR=2.35) and 4.18 (OR=5.18) times higher for nodules in the non-ATA or high

suspicion patterns, respectively, whereas the risk of malignancy was 55% lower in the non-ATA pattern compared to the high suspicion pattern (OR=0.45). If NIFTPs were considered benign, the differences in the risk of malignancy with the high suspicion pattern would be even greater but would remain unchanged for the low/intermediate versus non-ATA patterns comparison (Table 3).

TABLE 3. DIFFERENCES IN RISK OF MALIGNANCY BETWEEN SONOGRAPHIC PATTERNS

	Comparison		NIFTP cancer		NIFTP benign	
	Pattern 1	Pattern 2	OR [CI]	p-Value	OR [CI]	p-Value
All ITNs	Very low	Low/intermediate	NA	0.01	NA	0.15
	Very low	High	NA	<0.001	NA	<0.001
	Very low	Non-ATA	NA	<0.001	NA	0.005
	Low/intermediate	High	5.18 [2.34–11.6]	<0.001	7.36 [3.16–17.25]	<0.001
	Low/intermediate	Non-ATA	2.35 [1.46–3.80]	<0.001	2.24 [1.25–4.10]	0.005
AUS/FLUS	High	Non-ATA	0.45 [0.20–0.99]	0.04	0.30 [0.13–0.69]	0.003
	Very low	Low/intermediate	NA	0.06	NA	0.19
	Very low	High	NA	<0.001	NA	<0.001
	Very low	Non-ATA	NA	0.001	NA	0.03
	Low/intermediate	High	10.75 [2.84–47.30]	<0.001	16.93 [4.20–80.21]	<0.001
FN/HCN	Low/intermediate	Non-ATA	2.52 [1.07–6.24]	0.03	1.79 [0.62–5.47]	0.24
	High	Non-ATA	0.23 [0.06–0.83]	0.02	0.10 [0.02–0.40]	<0.001
	Very low	Low/intermediate	NA	0.58	NA	>0.99
	Very low	High	NA	0.26	NA	0.54
	Very low	Non-ATA	NA	0.29	NA	0.57
Low/intermediate	High	High	3.12 [1.05, 9.10]	0.02	3.57 [0.99–11.65]	0.03
	Low/intermediate	Non-ATA	2.33 [1.30, 4.23]	0.003	2.54 [1.23–5.38]	0.008
	High	Non-ATA	0.75 [0.25, 2.24]	0.62	0.71 [0.22–2.51]	0.57

Because no cancers were detected in the very low suspicion pattern, odds ratios could not be calculated for comparisons with this group. The low and intermediate sonographic patterns were grouped for comparisons with other patterns because their rates of malignancy were not significantly different.

CI, confidence interval; ITNs, cytologically indeterminate thyroid nodules (both AUS/FLUS and FN/HCN); Low/intermediate, low/intermediate suspicion sonographic patterns; NA, not applicable; NIFTP, noninvasive follicular thyroid neoplasm with papillary-like nuclear features; OR, odds ratio.

Correlation between histological diagnosis and sonographic patterns

The distribution of histological diagnoses was significantly different between AUS/FLUS and FN/HCN specimens ($p < 0.001$) and within different sonographic patterns ($p < 0.001$; Fig. 2A and Supplementary Table S1; Supplementary Data are available online at www.liebertpub.com/thy). Hyperplastic/adenomatous nodules were more prevalent among AUS/FLUS nodules, whereas adenomas were more likely classified as FN/HCN in cytology (Fig. 2B and Supplementary Table S1). Most benign nodules with a very low suspicion pattern were hyperplastic/adenomatous nodules, whereas they were mostly classified as adenomas in the other categories. Among malignancies with follicular architecture (NIFTPs, FVPTCs, and FTCs), the non-ATA pattern was the most prevalent. Other PTCs, however, were predominantly hypoechoic and thus more frequently had an intermediate or high suspicion sonographic pattern.

Of the 124 thyroid malignancies, 61% were low-risk cancers; 31% were intermediate-risk cancers, and 8% were high-risk cancers (Fig. 2C and Supplementary Table S2). Differences in the distribution of these groups of cancer aggressiveness by sonographic pattern did not reach statistical significance ($p = 0.13$). The rates of high-risk malignancies, however, were 2%, 8%, and 20% in the low/intermediate, non-ATA, and high suspicion sonographic patterns, respectively. The ORs of having a high-risk cancer in the non-ATA and high suspicion sonographic patterns compared to the low/intermediate sonographic patterns were 4.07 ([CI 0.46–36.17]; $p = 0.23$) and 11.0 ([CI 1.14–105.92]; $p = 0.028$).

FIG. 2. Distribution of histological diagnoses by sonographic pattern. (A) Distribution of histological diagnosis by Bethesda category and sonographic pattern. (B) Distribution of sonographic patterns by histological diagnosis. (C) Proportion of malignancies classified as low risk, intermediate risk, or high risk by sonographic pattern (excluded two intrathyroidal parathyroid carcinomas; see Methods section for definition of groups). FTC/HCC, follicular thyroid carcinoma/Hürthle cell carcinoma; FVPTC, follicular variant of papillary thyroid carcinoma; HP/AN/CLT, hyperplastic/adenomatous nodule or chronic lymphocytic thyroiditis; NIFTP, noninvasive follicular thyroid neoplasm with papillary-like nuclear features; Other PTC, other variants of papillary thyroid carcinoma different from FVPTC. ^aIncludes one intrathyroidal parathyroid adenoma. ^bIncludes two intrathyroidal parathyroid carcinomas, three medullary thyroid carcinomas, and one poorly differentiated thyroid carcinoma. Due to small cell counts, to compare the distribution of the histological diagnoses, NIFTPs and FVPTCs were collapsed into one group, and all other invasive cancers were collapsed into another group. Differences between AUS/FLUS and FN/HCN were statistically significant ($p < 0.001$). The distribution of histological diagnoses between sonographic patterns was also significantly different ($p < 0.001$ with Fisher's exact test for count data with a simulated p -value based on 2000 replicates). The low and intermediate suspicion patterns were collapsed into a single group for this analysis.

Differences in frequency and malignancy rates of sonographic patterns between observers

The intraclass correlation coefficient for absolute agreement of the distribution of the sonographic patterns between the observers was poor ($r = 0.32$ [CI 0.26–0.38]). The frequency with which ITNs were classified in the very low, low, intermediate, high, and non-ATA pattern ranged between 2% and 9%, 19% and 37%, 11% and 34%, 6% and 12%, and 24% and 62%, respectively. The correlation of the patterns was moderate between observers #1 and #3 ($r = 0.46$; $p < 0.001$) and weak between observer #2 and observers #1 ($r = 0.23$;

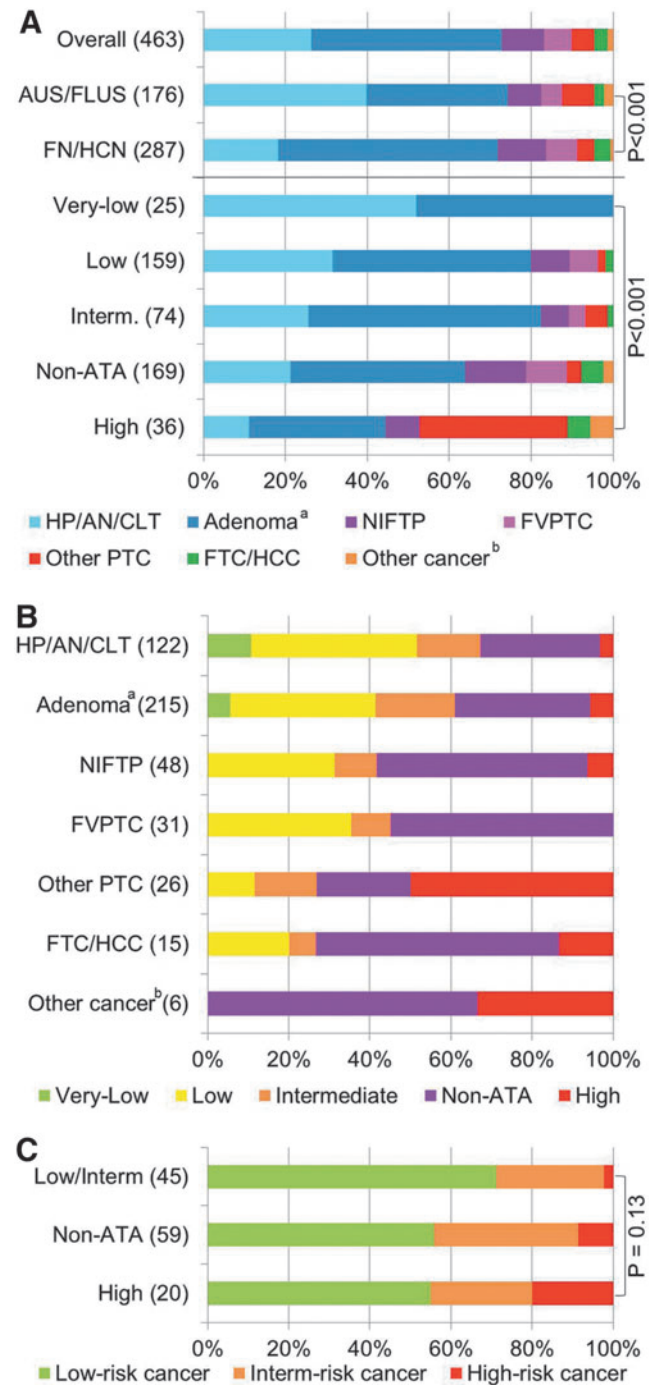


TABLE 4. DIFFERENCES IN FREQUENCY AND MALIGNANCY RATES OF SONOGRAPHIC PATTERNS BETWEEN OBSERVERS

<i>Sonographic pattern</i>		<i>Very low</i>	<i>Low</i>	<i>Intermediate</i>	<i>High</i>	<i>Non-ATA</i>
Frequency	Observer 1	41 (9)	171 (37)	52 (11)	57 (12)	142 (31)
	Observer 2	8 (2)	89 (19)	50 (11)	28 (6)	288 (62)
	Observer 3	22 (5)	135 (29)	159 (34)	37 (8)	110 (24)
	ICC			0.32 [0.26–0.38]		
PoM (with NIFTP)	Observer 1	1 (2)	34 (20)	9 (17)	29 (51)	53 (37)
	Observer 2	0	24 (27)	7 (14)	12 (43)	83 (29)
	Observer 3	1 (5)	30 (22)	33 (21)	17 (46)	45 (41)
	<i>p</i> -Value	0.82	0.43	0.54	0.76	0.04
PoM (without NIFTP)	Observer 1	0	18 (11)	5 (10)	25 (44)	30 (21)
	Observer 2	0	9 (10)	4 (8)	8 (29)	57 (20)
	Observer 3	1 (5)	16 (12)	22 (14)	15 (41)	24 (22)
	<i>p</i> -Value	0.35	0.40	0.46	0.39	0.89

ICC, intraclass correlation coefficient with confidence interval; PoM, prevalence of malignancy (considering NIFTPs as malignant or benign) for each sonographic pattern and observer.

$p < 0.001$) and #3 ($r = 0.20$; $p < 0.001$). The rates of malignancy of each sonographic pattern were, however, very consistent between all three observers, considering NIFTPs either malignant or benign (Table 4). Considering NIFTPs as malignant, the rate of malignancy ranged between 0% and 5% for the very low, 20% and 27% for the low, 14% and 21% for the intermediate, 43% and 51% for the high, and 29% and 41% for the non-ATA sonographic patterns. Only the rate of malignancy of non-ATA pattern nodules was significantly different between observers ($p = 0.04$). Considering NIFTPs as benign, differences were even smaller and not significantly different between observers, ranging between 0% and 5% for the very low, 10% and 12% for the low, 8% and 14% for the intermediate, 29% and 41% for the high, and 20% and 21% for the non-ATA sonographic patterns.

Discussion

In this study, the prevalence of malignancy of ITNs was effectively stratified by sonographic patterns. Moreover, the sonographic patterns were associated with distinct histopathological outcomes. They are therefore likely to improve the cytology–histology correlation. Differences in the proportion of these sonographic patterns could explain some of the observed interinstitutional variability in the risk of malignancy of the indeterminate categories of thyroid cytology and in the diagnostic performance of molecular marker tests.

Strengths and limitations

This study was done on a large series of consecutively evaluated ITNs with histological correlation, and ultrasound images were independently assessed by three different observers blinded to the histological diagnosis, mitigating the limitations of a single-center retrospective design. Nonetheless, the authors acknowledge that the results could be different in other sites or with other observers. All cytological diagnoses and most (97%) histological diagnoses were made by board-certified, experienced pathologists at the authors' institution. Although it cannot be ascertained that the original histological diagnoses were blinded to the ultrasound features, this was likely the case. Moreover, the histology of most (87%) malignant diagnoses was reviewed blinded to the ultrasound characteristics in order to assure compliance with

current diagnostic criteria, particularly for FVPTC/NIFTP. Nonetheless, it is acknowledged that the expertise of the pathologists cannot overcome the limitations of light microscopy, particularly for follicular pattern lesions (9–12).

Sonographic patterns stratify the risk of malignancy of ITNs

Several studies in recent years have suggested the use of ultrasound patterns to stratify the risk of malignancy of ITNs (13–22). Most of them have used one of the Thyroid Imaging Reporting and Data System (TIRADS) classifications, which, unlike the ATA classification, are based on scoring systems. The ATA classification, however, has been used in some recent studies with smaller cohorts and conflicting results (18,21,23,24). In the present study, three observers independently assessed the ultrasound images of 463 thyroid nodules blinded to the histological outcomes, and the ATA sonographic pattern was decided by agreement or consensus of two or more observers. The correlation of the ATA sonographic patterns between observers was weak to moderate in this study. Whereas this might have significant implications for setting biopsy thresholds, it seems to have little impact for cancer risk stratification of nodules with indeterminate cytology, which was very consistently achieved by all observers, particularly if NIFTPs are not considered malignant. Multivariate backward elimination logistic regressions were run using all variables listed in Table 1, and the cytological category searching for the contribution of other variables to the differences observed in the risk of malignancy between different sonographic patterns (data not shown). Only sex and/or age seemed to contribute to these differences for some but not all the comparisons. This information was not added to the manuscript because the effects of age and sex were small and clinically insignificant, and could be driven by the small sample size in the high suspicion pattern. Furthermore, these findings could be artifactual, as some patients had several nodules biopsied, and thus the characteristics of the same patient could be present in two different sonographic patterns or twice in the same one.

Hypoechoogenicity alone does not seem to improve risk stratification of ITNs, as there were no differences in prevalence of malignancy between nodules with either low (iso/

hyperechoic) or intermediate (hypoechoic) suspicion patterns. However, any other suspicious sonographic feature significantly increases the risk of malignancy of ITNs. Moreover, non-ATA sonographic patterns seem to be very prevalent among ITNs (37% in our series) and have a prevalence of malignancy higher than the low and intermediate suspicion sonographic patterns. Given the heterogeneity of the group, it is possible that different non-ATA patterns have different histological findings. Nodules with non-ATA patterns were on average larger than nodules with other sonographic patterns, and had a higher proportion of nodules >4 cm. This could mean that nodules tend to get more heterogeneous in echotexture as they grow, or that heterogeneous nodules are more likely to grow, which could be relevant for clinical management. Observational studies are needed to clarify this. On the other hand, hypoechoic nodules (intermediate and high suspicion patterns) were smaller than the nodules in other groups. This finding might be related to the fact that hypoechogenicity has been recognized as a risk factor for many years, which may have triggered the biopsy earlier than in nodules with other echogenicity (25–27).

In this study, there could be an overrepresentation of more suspicious sonographic patterns, and the observed rates of malignancy could be overestimated because only resected ITNs were included. In the authors' opinion, this is unlikely because the sonographic features/patterns have been traditionally used to select nodules for biopsy but not to guide management once cytology is available. The overall rate of malignancy of ITNs in the present study is within the expected range and is consistent with previously published series (6,28), and it dropped approximately by half in the low, intermediate, and non-ATA patterns when NIFTPs were considered benign, whereas it was unchanged in the very low and high suspicion patterns.

Association of sonographic patterns with histological diagnoses and implications for molecular tests

This study found that the histological diagnoses associated with each pattern were significantly different. Although differences were also statistically significant between AUS/FLUS and FN/HCN nodules, larger differences were seen in the classification of benign nodules as either hyperplastic/adenomatous or adenomas, whereas the distribution of specific histological diagnoses of malignant nodules was less evident than when nodules were segregated by sonographic pattern. Although most malignancies were low-risk cancers in all sonographic patterns, the probability of aggressive histological features increased with the degree of sonographic suspicion. This indicates that sonographic patterns are likely to improve the cytology–histology correlation, which could impact the diagnostic performance and interpretation of molecular tests results. At the authors' institution, oncogene panels were used routinely after 2014 (29,30). As previously described, these were mainly used to decide the extent of surgery rather than to avoid resection (31). However, only 17% of this cohort was evaluated with oncogene panels, which is insufficient to evaluate the impact of sonographic patterns on the oncogene panels' results. Thus, future studies are needed to investigate this topic.

In the authors' opinion, molecular markers might not be useful for the presurgical evaluation of ITNs with very low or

high suspicion sonographic patterns, which represent 13% of all ITNs in this series. In nodules with very low suspicion pattern, the pretest probability of cancer was already $\leq 5\%$, which is similar to a benign cytology and, according to National Comprehensive Cancer Network and ATA guidelines, sufficient to elect observation (1,4). On the other hand, in nodules with high suspicion sonographic pattern, the pretest risk of malignancy is theoretically too high to achieve a low enough negative predictive value to avoid surgery with any of the currently available tests, and a positive result might be insufficient to modify the extent of surgery (32). Furthermore, other factors, such as the presence or absence of nuclear atypia, architectural atypia, or oncocyctic changes in the cytological specimen, seem to impact on molecular marker tests results (33,34). The correlation of these cytological scenarios and the sonographic patterns is currently unknown and needs to be further defined to optimize patient management in the era of personalized medicine.

Conclusion

The 2015 ATA sonographic patterns effectively stratify the risk of malignancy of thyroid nodules with indeterminate cytology and are associated with distinct clinical and histological features. These results have important implications for the evaluation and management of ITNs and suggest that the sonographic pattern should not only have a role in selecting nodules that need to be biopsied, but also in deciding management after cytological diagnosis. Sonographic patterns are likely to impact the performance of molecular marker tests. Future studies will need to study this interaction to restrict the use of these expensive tests to clinical scenarios in which they may be useful.

Acknowledgments

This work has been supported in part by the Biostatistics Core Facility at the H. Lee Moffitt Cancer Center & Research Institute; an NCI designated Comprehensive Cancer Center (P30-CA076292).

Author Disclosure Statement

No competing financial interests exist.

References

1. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, Pacini F, Randolph GW, Sawka AM, Schlumberger M, Schuff KG, Sherman SI, Sosa JA, Steward DL, Tuttle RM, Wartofsky L 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**:1–133.
2. Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedus L, Paschke R, Valcavi R, Vitti P; AACE/ACE/AME Task Force on Thyroid Nodules 2016 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**:622–639.

3. Perros P, Boelaert K, Colley S, Evans C, Evans RM, Gerrard Ba G, Gilbert J, Harrison B, Johnson SJ, Giles TE, Moss L, Lewington V, Newbold K, Taylor J, Thakker RV, Watkinson J, Williams GR 2014 Guidelines for the management of thyroid cancer. *Clin Endocrinol* **81**:1–122.
4. National Comprehensive Cancer Network 2017 NCCN Clinical Practice Guidelines in Oncology (NCCN guidelines). Thyroid Carcinoma (version 2.2017). Available at: www.nccn.org/professionals/physician_gls/f_guidelines.asp#thyroid (accessed June 30, 2017).
5. Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW 2012 The Bethesda System for Reporting Thyroid Cytopathology: a meta-analysis. *Acta Cytol* **56**:333–339.
6. Cibas ES, Ali SZ 2017 The 2017 Bethesda System for Reporting Thyroid Cytopathology. *Thyroid* **27**:1341–1346.
7. Cibas ES, Ali SZ 2009 The Bethesda System for Reporting Thyroid Cytopathology. *Thyroid* **19**:1159–1165.
8. Nikiforov YE, Seethala RR, Tallini G, Baloch ZW, Basolo F, Thompson LD, Barletta JA, Wenig BM, Al Ghuzlan A, Kakudo K, Giordano TJ, Alves VA, Khanafshar E, Asa SL, El-Naggar AK, Gooding WE, Hodak SP, Lloyd RV, Maytal G, Mete O, Nikiforova MN, Nose V, Papotti M, Poller DN, Sadow PM, Tischler AS, Tuttle RM, Wall KB, LiVolsi VA, Randolph GW, Ghossein RA 2016 Nomenclature revision for encapsulated follicular variant of papillary thyroid carcinoma: a paradigm shift to reduce overtreatment of indolent tumors. *JAMA Oncol* **2**:1023–1029.
9. Cibas ES, Baloch ZW, Fellegara G, LiVolsi VA, Raab SS, Rosai J, Diggans J, Friedman L, Kennedy GC, Kloos RT, Lanman RB, Mandel SJ, Sindy N, Steward DL, Zeiger MA, Haugen BR, Alexander EK 2013 A prospective assessment defining the limitations of thyroid nodule pathologic evaluation. *Ann Intern Med* **159**:325–332.
10. Elsheikh TM, Asa SL, Chan JK, DeLellis RA, Heffess CS, LiVolsi VA, Wenig BM 2008 Interobserver and intraobserver variation among experts in the diagnosis of thyroid follicular lesions with borderline nuclear features of papillary carcinoma. *Am J Clin Pathol* **130**:736–744.
11. Lloyd RV, Erickson LA, Casey MB, Lam KY, Lohse CM, Asa SL, Chan JK, DeLellis RA, Harach HR, Kakudo K, LiVolsi VA, Rosai J, Sebo TJ, Sobrinho-Simoes M, Wenig BM, Lae ME 2004 Observer variation in the diagnosis of follicular variant of papillary thyroid carcinoma. *Am J Surg Pathol* **28**:1336–1340.
12. Hirokawa M, Carney JA, Goellner JR, DeLellis RA, Heffess CS, Katoh R, Tsujimoto M, Kakudo K 2002 Observer variation of encapsulated follicular lesions of the thyroid gland. *Am J Surg Pathol* **26**:1508–1514.
13. Kihara M, Ito Y, Hirokawa M, Masuoka H, Yabuta T, Tomoda C, Higashiyama T, Inoue H, Fukushima M, Takamura Y, Kobayashi K, Miya A, Miyauchi A 2011 Role of ultrasonography in patients with cytologically follicular thyroid tumor. *Auris Nasus Larynx* **38**:508–511.
14. Jeong SH, Hong HS, Lee EH, Cha JG, Park JS, Kwak JJ 2013 Outcome of thyroid nodules characterized as atypia of undetermined significance or follicular lesion of undetermined significance and correlation with ultrasound features and *BRAF(V600E)* mutation analysis. *AJR A J Roentgenol* **201**:W854–860.
15. Yoon JH, Kwon HJ, Kim EK, Moon HJ, Kwak JY 2016 Subcategorization of atypia of undetermined significance/follicular lesion of undetermined significance (AUS/FLUS): a study applying Thyroid Imaging Reporting and Data System (TIRADS). *Clin Endocrinol* **85**:275–282.
16. Chng CL, Kurzawinski TR, Beale T 2015 Value of sonographic features in predicting malignancy in thyroid nodules diagnosed as follicular neoplasm on cytology. *Clin Endocrinol* **83**:711–716.
17. Russ G, Royer B, Bigorgne C, Rouxel A, Bienvenu-Perrard M, Leenhardt L 2013 Prospective evaluation of thyroid imaging reporting and data system on 4550 nodules with and without elastography. *Eur J Endocrinol* **168**:649–655.
18. Trimboli P, Deandrea M, Mormile A, Ceriani L, Garino F, Limone PP, Giovanella L 2018 American Thyroid Association ultrasound system for the initial assessment of thyroid nodules: use in stratifying the risk of malignancy of indeterminate lesions. *Head Neck* **40**:722–727.
19. Trimboli P, Fulciniti F, Zilioli V, Ceriani L, Giovanella L 2017 Accuracy of international ultrasound risk stratification systems in thyroid lesions cytologically classified as indeterminate. *Diagn Cytopathol* **45**:113–117.
20. Ulisse S, Bosco D, Nardi F, Nesca A, D'Armiento E, Guglielmino V, De Vito C, Sorrenti S, Pironi D, Tartaglia F, Arcieri S, Catania A, Monti M, Filippini A, Ascoli V 2017 Thyroid imaging reporting and data system score combined with the new italian classification for thyroid cytology improves the clinical management of indeterminate nodules. *Int J Endocrinol* **2017**:9692304.
21. Grani G, Lamartina L, Ascoli V, Bosco D, Nardi F, D'Ambrosio F, Rubini A, Giacomelli L, Biffoni M, Filetti S, Durante C, Cantisani V 2017 Ultrasonography scoring systems can rule out malignancy in cytologically indeterminate thyroid nodules. *Endocrine* **57**:256–261.
22. Maia FF, Matos PS, Pavin EJ, Vassallo J, Zantut-Wittmann DE 2011 Value of ultrasound and cytological classification system to predict the malignancy of thyroid nodules with indeterminate cytology. *Endocr Pathol* **22**:66–73.
23. Bardet S, Ciappuccini R, Pellot-Barakat C, Monpeyssen H, Michels JJ, Tissier F, Blanchard D, Menegaux F, de Raucourt D, Lefort M, Reznik Y, Rouxel A, Heutte N, Brenac F, Leconte A, Buffet C, Clarisse B, Leenhardt L 2017 Shear wave elastography in thyroid nodules with indeterminate cytology: results of a prospective bicentric study. *Thyroid* **27**:1441–1449.
24. Tang AL, Falciglia M, Yang H, Mark JR, Steward DL 2017 Validation of American Thyroid Association ultrasound risk assessment of thyroid nodules selected for ultrasound fine-needle aspiration. *Thyroid* **27**:1077–1082.
25. Brito JP, Gionfriddo MR, Al Nofal A, Boehmer KR, Lepin AL, Reading C, Callstrom M, Elraiyah TA, Prokop LJ, Stan MN, Murad MH, Morris JC, Montori VM 2014 The accuracy of thyroid nodule ultrasound to predict thyroid cancer: systematic review and meta-analysis. *J Clin Endocrinol Metab* **99**:1253–1263.
26. Campanella P, Ianni F, Rota CA, Corsello SM, Pontecorvi A 2014 Quantification of cancer risk of each clinical and ultrasonographic suspicious feature of thyroid nodules: a systematic review and meta-analysis. *Eur J Endocrinol* **170**:R203–211.
27. Remonti LR, Kramer CK, Leitao CB, Pinto LC, Gross JL 2015 Thyroid ultrasound features and risk of carcinoma: a systematic review and meta-analysis of observational studies. *Thyroid* **25**:538–550.
28. Straccia P, Rossi ED, Bizzarro T, Brunelli C, Cianfrini F, Damiani D, Fadda G 2015 A meta-analytic review of The Bethesda System for Reporting Thyroid Cytopathology:

- has the rate of malignancy in indeterminate lesions been underestimated? *Cancer Cytopathol* **123**:713–722.
29. Valderrabano P, Khazai L, Leon ME, Thompson ZJ, Ma Z, Chung CH, Hallanger-Johnson JE, Otto KJ, Rogers KD, Centeno BA, McIver B 2017 Evaluation of ThyroSeq v2 performance in thyroid nodules with indeterminate cytology. *Endocr Relat Cancer* **24**:127–136.
 30. Valderrabano P, Leon ME, Centeno BA, Otto KJ, Khazai L, McCaffrey JC, Russell JS, McIver B 2016 Institutional prevalence of malignancy of indeterminate thyroid cytology is necessary but insufficient to accurately interpret molecular marker tests. *Eur J Endocrinol* **174**:621–629.
 31. Valderrabano P, Khazai L, Thompson ZJ, Leon ME, Otto KJ, Hallanger-Johnson JE, Wadsworth JT, Chung CH, Centeno BA, McIver B 2018 Impact of oncogene panel results on surgical management of cytologically indeterminate thyroid nodules. *Head Neck* 2008 Apr 6 [Epub ahead of print]; DOI: 10.1002/hed.25165.
 32. Patel SG, Carty SE, McCoy KL, Ohori NP, LeBeau SO, Seethala RR, Nikiforova MN, Nikiforov YE, Yip L 2017 Preoperative detection of *RAS* mutation may guide extent of thyroidectomy. *Surgery* **161**:168–175.
 33. Baca SC, Wong KS, Strickland KC, Heller HT, Kim MI, Barletta JA, Cibas ES, Krane JF, Marqusee E, Angell TE 2017 Qualifiers of atypia in the cytologic diagnosis of thyroid nodules are associated with different Afirma gene expression classifier results and clinical outcomes. *Cancer Cytopathol* **125**:313–322.
 34. Brauner E, Holmes BJ, Krane JF, Nishino M, Zurakowski D, Hennessey JV, Faquin WC, Parangi S 2015 Performance of the Afirma gene expression classifier in Hurthle cell thyroid nodules differs from other indeterminate thyroid nodules. *Thyroid* **25**:789–796.

Address correspondence to:

Pablo Valderrabano, MD, PhD

Department of Head and Neck-Endocrine Oncology

H. Lee Moffitt Cancer Center and Research Institute

12902 Magnolia Drive

Tampa, FL 33612

E-mail: pvalderrabano@gmail.com