



Retrospective analysis of the ultrasound features of resected thyroid nodules

Loredana Pagano¹ · Enrico Costantino Falco² · Alessandro Bisceglia¹ · Alessandro Gambella² · Ruth Rossetto¹ · Sara Garberoglio¹ · Francesca Maletta² · Donatella Pacchioni² · Roberto Garberoglio¹ · Ezio Ghigo¹ · Mauro Giulio Papotti^{2,3}

Received: 8 May 2020 / Accepted: 7 September 2020 / Published online: 2 October 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Purpose Several ultrasound (US) risk stratification systems (US-RSSs) have been proposed to stratify the risk of malignancy (ROM) of thyroid nodules. This risk might be overestimated due to selection bias and comparison with the cytological report alone. Our study aimed to compare ROM and diagnostic performance of three guidelines (ATA, AACE/ACE/AME, EUTIRADS) and evaluate the changes in unnecessary biopsy according to the nodule size cutoff for biopsy, using histology as gold standard.

Methods This retrospective observational study included 146 consecutive patients who underwent surgery after US and cytological characterization. We analyzed the effectiveness and accuracy of three US-RSSs.

Results 46.6% of nodules were diagnosed as malignant. Applying US-RSS, the percentage of nodules that should have been analyzed by biopsy was 84.25% with ATA, 69.86% with EUTIRADS and 64.38% with AACE/ACE/AME systems. The ROM was 94.9%, 86.0%, 87.0% for high-risk category, 36.4%, 32.0%, 35.4% for intermediate-risk category and 22.9%, 0.0%, 22.9% for low-risk category by ATA, AACE/ACE/AME and EUTIRADS systems, respectively. EUTIRADS and AACE/ACE/AME systems were more accurate in differentiating malignant from benign cases. ATA score was the more sensitive US-RSS to identify malignant tumors within the high-risk category. About the unnecessary biopsies, in the intermediate-risk category, the application of the size criterion helps to increase specificity in all systems.

Conclusions The US categorization of low and high-risk thyroid nodules using current US-RSSs helps alone to determine the optimal treatment option. Nodule size remains relevant to recommend biopsy for the intermediate-risk category.

Keywords Thyroid nodules · Ultrasound risk stratification · Cytology aspiration · Histology · Diagnosis

Introduction

Thyroid nodules are common gland abnormalities that can be found in up to 65% of the general population and are mostly benign and clinically silent. This high prevalence reflects a diagnostic detection bias associated with the growing use of US for the evaluation and follow-up of neck

structures, including the thyroid. In the presence of an (incidental) thyroid nodule, a main objective related to its management is to identify the subgroup of clinically significant thyroid tumors, accounting for 5–10% of cases [1].

Cause most thyroid nodules are asymptomatic or incidentally diagnosed by thyroid ultrasound, several scientific societies have proposed diagnostic algorithms for an initial stratification of the malignancy risk [2–6], but none of them has been sufficiently standardized [7–9]. As a consequence, US-guided fine-needle aspiration (FNA) biopsy of thyroid nodules is currently an overused practice resulting in non-selective extensive sampling of benign lesions that might not require this procedure or further clinical attention [10, 11].

In this retrospective study on patients with thyroid nodules who underwent surgical resection, we aimed to evaluate the accuracy in predicting malignancy of three different international US stratification systems, e.g., the

✉ Ezio Ghigo
ezio.ghigo@unito.it

¹ Division of Endocrinology, Diabetology and Metabolism, Department of Medical Sciences, University of Turin, Turin, Italy

² Pathology Unit, City of Health and Science University Hospital, Turin, Italy

³ Department of Oncology, University of Turin, Turin, Italy

2015 American Thyroid Association (ATA) guidelines, the American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi Medical Guidelines (AACE/ACE/AME) guidelines, and the European Thyroid Association guidelines (EUTIRADS) [2, 5, 6], highlighting the main critical issues of thyroid nodule diagnostic algorithms.

Materials and methods

This retrospective observational study included 146 consecutive patients who were referred to our Center for FNA cytology for suspected thyroid nodules and then underwent thyroid surgery.

All data were retrieved from the Città della Salute e della Scienza University Hospital in Turin from January 2015 to September 2018. The Institutional Review Boards of our hospital approved this study. Since the purpose of this study was to evaluate the reliability of US scores in predicting malignancy of thyroid nodules by comparing US scores outcomes with both cytological and histological diagnoses, cases undergoing surgery with a non-diagnostic cytology at FNA were excluded. Moreover, the small number of nodules with this cytological report would have been too low (7 cases) for an accurate analysis.

To minimize the detection bias, histological and cytological materials, as well as the clinical charts and ultrasonographic records, were anonymized by a staff person not involved in this project, and only coded data were used for microscopic review and statistical analyses.

Ultrasound features

Both transverse and longitudinal sonograms were obtained by real-time imaging of the thyroid nodules using an Esaote MyLab Twice real-time US system with a linear multi-frequency (7–14 MHz) probe. The still sonographic images were independently reviewed by two board-certified radiologists (R.G. and S.G.) and two endocrinologists (RR and LP) with >10 years of experience. In case of discordance, a mutual agreement was achieved after discussion and review of the video clips filmed during the FNA procedure.

The sonographic findings were analyzed by assigning the examined features to each category of the Ultrasound Risk Stratification Systems (US-RSS) based on the ATA [2], EUTIRADS [6], and 2016 AACE/ACE/AME [5] guidelines. Diameters (anteroposterior, transverse, and longitudinal) of each thyroid nodule were measured in millimeters (mm). Regarding the echostructure/composition of the nodules, they were classified as solid -predominantly solid (cyst \leq 10% and cystic \leq 50%), cystic- predominantly cystic (cyst > 50%) and spongiform (nodules containing

multiple small cysts smaller than 5 mm interspersed within the solid tissue component [1, 3, 12].

The echogenicity of the nodules was classified as markedly hypoechoic, hypoechoic, isoechoic, anechoic or hyperechoic. A marked hypoechoic lesion was defined as a thyroid nodule that showed a relatively hypoechoic pattern compared to the adjacent strap muscles of the neck. Nodular margins were categorized and defined as ill-defined or smooth. Calcifications were subdivided into microcalcifications (defined as calcifications that were equal to or less than 1 mm and visualized as tiny punctate hyperechoic foci, either with or without acoustic shadows) and macrocalcifications (defined as hyperechoic foci larger than 1 mm). Nodule shape-also referred as nodule orientation along the longitudinal axis-was divided into parallel or not parallel (taller than wide). In this study, the taller-than-wide shape, highlighted in only two cases, was assessed by means of measurements and was defined in cases where anteroposterior (AP) diameter exceeded the transverse (T) diameter [13].

According to the US-RSS of the AACE/ACE/AME 2016 system only, we evaluated vascularization by color and power Doppler examination and stiffness by qualitative elastography. In our study, we have carried an elastographic strain evaluation. The pressure is exerted freehand through the ultrasound transducer. An elastographic image (elastogram) is then produced, represented as a color-coded image superimposed on the image in mode B; the two images are displayed side by side on the screen.

The elastogram provides a mapping of the stiffness of the nodule considered in each position and allows for a qualitative assessment [14, 15].

By evaluating the color pattern prevalent within the nodule, it is possible to qualitatively compare the result obtained with a progressive reference score, among those proposed by the literature. In our study, we used the classification proposed by Rago et al. [16]. We defined the nodules with patterns 1 and 2 proposed by Rago as soft; the nodules corresponding to pattern 3 were classified at intermediate elasticity and finally the nodules with presentation patterns 4 and 5 were considered hard. If an intra-nodular vascularization pattern or elevated stiffness were documented, the nodule was classified as at intermediate risk.

All patients underwent total thyroidectomy or hemithyroidectomy, with an ultimate histological definition of the nodule nature.

According to recently published data [17], in our study, we interpreted the ultrasound data of each US-RSS into the following categories:

Low risk nodule: low risk according to the 2016 AACE/ACE/AME guidelines; benign, very low suspicion and low suspicion according to the 2015 ATA guidelines; and classes 2–3 according to 2017 EUTIRADS.

Intermediate risk nodule: intermediate risk according to the 2016 AACE/ACE/AME guidelines; intermediate suspicion according to the 2015 ATA guidelines; and class 4 according to 2017 EUTIRADS.

High risk nodule: high risk according to the 2016 AACE/ACE/AME guidelines; high suspicion according to the 2015 ATA guidelines; and class 5 according to 2017 EUTIRADS.

US-guided FNA

After the ultrasound characterization, the FNA procedure followed. Samples were obtained with 21/23-gauge needles, and a range of 1–4 passes was performed depending on the location of the lesion; most passes were performed together with a rapid on-site evaluation (ROSE). The aspirated material was smeared onto two slides (one fixed in 95% ethanol and stained with a rapid hematoxylin-eosin (HE) stain for ROSE and the other air dried for Giemsa stain), and the excess material was placed in a 95% alcoholic solution for cell block preparation. For each cell block, two sections were prepared for routine (HE) and Papanicolaou (PAP) staining. Each case was then classified according to the criteria published in the Italian Consensus for Thyroid Cytopathology (SIAPEC-IAP) as follows: Tir1—non-diagnostic; Tir1c—nondiagnostic/cystic; Tir2—benign; Tir3A—indeterminate (low-risk lesion); Tir3B—indeterminate (high-risk lesion); Tir4—suspicious of malignancy; and Tir5—malignant [18].

Specimen interpretation

All cytological and histological specimens were reviewed by three pathologists trained in thyroid histo- and cytopathology (F.M., D.P., and M.P.). Each reviewer was blinded to the other judgments. Discordant cases were jointly discussed under a multi-head microscope, and a consensus was reached.

Statistical analyses

All nodules were retrospectively classified using the US criteria of three international guidelines [2, 5, 6]. The risks of malignancy (ROM) associated with each category were calculated as percentages. All nodules were dichotomized into two groups: those for which US-guided FNA was or was not indicated by the FNA criteria. Thyroid nodules were dichotomized into two groups based on the recommendations (including the dimensional criterion) for biopsy by each guideline, that is, US-guided biopsy indicated or not indicated. The diagnostic performances of all US criteria

in terms of identifying thyroid cancer were evaluated. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy of each set of guidelines were calculated. The agreement between each US-RSS was then evaluated by the Cohen kappa test. The kappa coefficient was interpreted as follows: 0.00–0.20, poor agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, good agreement; and 0.81–1.00, excellent agreement.

Continuous variables are presented as mean \pm standard deviation and were evaluated using the Student's *t*-test. Categorical variables were analyzed using the chi-square test. The results were considered statistically significant if the *P* value was less than 0.05. All statistical analyses were performed with STATA IC 10 (STACORP, LP, Texas, USA) analytic software.

Results

Demographic and sonographic characteristics of thyroid nodules

Of the 146 patients included in our study, 111 were females (76.0%); the mean age was 50.5 ± 14.8 . The mean nodule size was 26.2 ± 16.4 mm (for each patient, a suspicious nodule was evaluated).

According to the histological diagnosis, 78 lesions were benign and 68 malignant, considering also Non Invasive Follicular Tumor with Papillary-like nuclear features (NIFTP) and well differentiated tumor of uncertain malignant potential (WDT-UMP) as malignant lesions.

The patients with histological-confirmed malignant nodules were younger than those with benign nodules (47.8 vs 52.6 years, $P = 0.05$).

As summarized in Table 1, the malignant nodules were smaller than the benign nodules (19.8 ± 12.5 vs 31.8 ± 17.4 mm, $P < 0.001$) and overall had a solid pattern. Sonographically, malignant thyroid nodules showed a more marked hypoechogenicity ($P < 0.001$), irregular margins ($P < 0.001$), microcalcifications ($P < 0.001$), intranodular vascularization ($P = 0.05$) and elevated stiffness ($P < 0.001$) than the benign nodules.

Cytological and histological reports

In our series, 68 nodules (46.6%) were histologically proven to be malignant based on the resected specimens (54 papillary carcinomas, 10 follicular carcinomas, 1 medullary carcinoma, 1 NIFTP and 2 WDT-UMP). Benign lesions included 44 goiters, 1 thyroiditis and 33 follicular adenomas.

Regarding FNA cytological findings, 116 nodules (79.5%) were classified into high risk categories (Tir3B, 82

Table 1 Ultrasonographic features of 146 thyroid nodules

Ultrasonographic features	Total n = 146 (%)	Benign nodules n = 78 (%)	Malignant nodules n = 68 (%)	p-value
Echostructure/Composition <i>p</i> < 0.166				
Solid	142 (97.2)	74 (94.9)	68 (100%)	0.058
Spongiform	3 (2.1)	3 (3.8)	0 (0.0)	0.102
Cystic	1 (0.7)	1 (1.3)	0 (0.0)	0.348
Echogenicity <i>p</i> < 0.001				
Anechoic	1 (0.7)	1 (1.3)	0 (0.0)	0.348
Isoechoic	52 (35.6)	40 (51.3)	12 (17.7)	<0.001
Hypoechoic	62 (42.5)	32 (41.0)	30 (44.1)	0.706
Markedly Hypoechoic	31 (21.2)	5 (6.4)	26 (38.2)	<0.001
Hyperechoic	0 (0)	0 (0)	0 (0)	N.A.
Margins <i>p</i> < 0.001				
Smooth	122 (83.5)	78 (100)	44 (64.7)	<0.001
Ill-defined	24 (16.5)	0 (0.0)	24 (35.3)	<0.001
Calcifications <i>p</i> < 0.001				
Absent	116 (79.5)	76 (97.4)	40 (58.8)	<0.001
Macro	7 (4.8)	0 (0.0)	7 (10.3)	0.004
Micro	23 (15.7)	2 (2.6)	21 (30.9)	<0.001
Nodule shape <i>p</i> < 0.127				
Parallel	144 (98.6)	78 (100)	66 (97.1)	0.127
Not parallel (Taller than wide)	2 (1.4)	0 (0.0)	2 (2.9)	0.127
Vascularization <i>p</i> < 0.036				
Perinodular	62 (42.5)	31 (39.7)	31 (45.6)	0.476
Endonodular	16 (10.9)	5 (6.4)	11 (16.2)	0.050
Peri-Endonodular	63 (43.2)	37 (47.5)	26 (38.2)	0.274
Unknown	5 (3.4)	5 (6.4)	0 (0.0)	N.A.
Strain US Elastography <i>p</i> < 0.001				
Soft	55 (37.7)	36 (46.2)	19 (27.9)	0.023
Intermediate	27 (18.5)	13 (16.7)	14 (20.6)	0.543
Hard	32 (21.9)	8 (10.2)	24 (35.3)	<0.001
Unknown	32 (21.9)	21 (26.9)	11 (16.2)	N.A.
Size max (mm) <i>p</i> < 0.001				
1–10	19 (13.0)	3 (3.8)	16 (23.5)	<0.001
11–15	30 (20.6)	10 (12.8)	20 (29.4)	0.013
16–20	25 (17.1)	17 (21.8)	8 (11.8)	0.109
>20	72 (49.3)	48 (61.6)	24 (35.3)	0.002

N.A. not applicable

cases; Tir4, 9 cases; and Tir5, 25 cases), resulting in 65 malignant (56.0%) and 51 benign lesions (44.0%). The remaining 30 nodules were classified into low-risk cytological categories (Tir2, 25 cases, and Tir3A, 5 cases), 27 of them (90.0%) with a benign final histological report and 3 (10.0%) with a malignant one (Table 2).

The risk of malignancy (ROM) according to the different cytological classes was 0% for Tir2, 15.4% for Tir3A, 35.4% for Tir3B and 100% for Tir4 and Tir5. Using the chi-square test with Yates correction, the results of the cytological reports were statistically correlated with the histological diagnoses (χ^2 :21.1769; *P* < 0.001).

FNA cytology classified macrofollicular goiters (90.9% in Tir2 category) and classical papillary carcinomas (100% in TIR3B, 4 and 5 categories) with high accuracy. On the other hand, microfollicular goiters (especially those with an oxyphilic component) and variants of papillary carcinomas, as well as follicular adenomas and carcinomas caused major diagnostic problems since they were usually classified in the TIR3A-3B category. In case of indeterminate lesions (TIR3), the EUTIRADS and ATA scores performed similarly to the cytological classification in defining most TIR3A nodules (69.2%) as low risk and the majority of TIR3B nodules (70.7%) as intermediate/high risk. The AACE/ACE/AME performance seemed influenced by the relatively small number of lesions classified as low risk. Consequently, this system had a poor performance in defining TIR3A low risk lesions. About discordant cases, FNA cytological evaluations were unable to identify malignant lesions in two cases classified as TIR3A. These lesions turned out to be a follicular variant of papillary carcinoma and a Hurthle cell carcinoma, both characterized by minimally aggressive features. The first was classified as carcinoma because of the presence of papillary carcinoma type nuclei and of psammoma bodies, while the latter had only a single focus of capsular invasion corresponding to a diagnosis of minimally invasive Hurthle cell carcinoma.

Table 2 Stratification of cytologic FNA score based on surgery indication together with the comparison between cytologic FNA score and histology report

Italian consensus for thyroid cytopathology score	Benign histology diagnosis	Malignant histology diagnosis
Low risk (Tir 2–3A/Bethesda II–III)	27 Goiter: 25 Thyroiditis: 0 Adenoma: 2	3 Papillary carcinoma: 1 Follicular carcinoma: 1 Medullary carcinoma: 0 Others ^a : 1
High Risk (Tir 3B–4–5/Bethesda IV–V–VI)	51 Goiter: 19 Thyroiditis: 1 Adenoma: 31	65 Papillary carcinoma: 53 Follicular carcinoma: 9 Medullary carcinoma: 1 Others ^a : 2

^aUMP Tumors and NIFTP

When considering TIR3B nodules, the likelihood of a histologically malignant result in low ultrasound-risk categories (6/23, 26.1%) was similar to that observed in intermediate-risk lesions by the three US-scoring systems (EUTIRADS: 9/36, 25.0%; ATA: 11/42, 26.2%, AACE/ACE/AME: 17/60, 28.3%).

Ultrasound risk stratification systems (US-RSSs)

The percentage of nodules that would have undergone FNA biopsy was 84.25% according to ATA, 69.86% according to EUTIRADS, and 64.38% according to AACE/ACE/AME score system. In Table 3, the frequencies and malignancy risks of all categories of the three guidelines are reported.

In relation to the histological results, the ATA score was more sensitive as it missed fewer carcinomas than the other systems (18 with ATA vs 22 with EUTIRADS and 26 with AACE/ACE/AME). Conversely, the EUTIRADS achieved a higher specificity and would have prompted less biopsies (ATA score 84.25% versus EUTIRADS score 69.86%) but maintaining almost the same efficacy as the ATA score (Table 4).

To evaluate the agreement between the results of the three scoring systems in the thyroid nodules, we measured Cohen's kappa coefficient among the three systems, using two US-RSSs at the time. Each US-RSS demonstrated a good or moderate agreement with each other. However, AACE/ACE/AME and EU-TIRADS showed the highest US-RSS concordance ($\kappa = 0.75$, $P < 0.001$).

We then evaluated the ability of the three US-RSSs to discriminate between low suspicion (TIR 2, TIR3A) and high

suspicion nodules for malignancy (TIR3B, TIR4 and TIR5) in relation to histology. Low-risk nodules resulted histologically benign in most cases according to both the ATA and EUTIRADS systems (78.8%), while in all cases (100%) according to the AACE/ACE/AME system. However, the latter system classified as low risk only 6/146 lesion, likely explaining the high accuracy described. Remarkably, all lesions classified as very low risk by the ATA and low-risk by the AACE/ACE/AME or as class 2 by EUTIRADS systems were in fact histologically benign. The category of low-risk nodules by the ATA and class 3 by EUTIRADS contained a few false-negative cases, including 6 papillary thyroid cancers (all classified as a low risk of recurrence according to the 2015 ATA system), 2 follicular thyroid cancers (one minimally invasive FTCs with capsular invasion only, the other completely cystic), 1 NIFTP and 2 WDT-UMP at surgery.

When nodules belonging to the US-RSS high suspicion categories were analyzed, the ROM was 94.9% by ATA, 87.0% by EUTIRADS and 86.0% by AACE/ACE/AME systems. In this category, in relation to phenotypes, ATA misquoted 2 cases as benign (1 goiter, 1 follicular adenoma) out of the 39 histologically-proven malignant lesions, while AACE/ACE/AME and EUTIRADS each misquoted 6 benign cases (3 goiters and 3 follicular adenomas) out of 43 and 46 malignant lesions, respectively.

Our analysis of nodules with an intermediate risk showed that these were histologically proven as benign in 35/55 cases (63.6%) according to ATA, 66/97 (68.0%) according to AACE/ACE/AME, and 31/48 (64.6%) according to EUTIRADS systems. When this category in the ATA guideline modeled >20 mm instead of 10 mm as the size for biopsy, the

Table 3 Risk of malignancy of the three ultrasound risk stratification guidelines

Guidelines	Total <i>n</i> = 146 (%)	Benign nodules <i>n</i> = 78 (%)	Malignant nodules <i>n</i> = 68 (%)	Suggested risk of malignancy %	Calculated risk of malignancy %	<i>p</i> -value
ATA 2015						
Benign	0 (0.0)	0 (0.0)	0 (0.0)	<1	0.0	<i>p</i> < 0.001
Very Low suspicion	4 (2.7)	4 (5.1)	0 (0.0)	<3	0.0	N.A.
Low suspicion	48 (32.9)	37 (47.4)	11 (16.2)	5–10	22.9	0.058
Intermediate suspicion	55 (37.7)	35 (44.9)	20 (29.4)	10–20	36.4	<0.001
High suspicion	39 (26.7)	2 (2.6)	37 (54.4)	>70–90	94.9	0.054
AACE/ACE/AME 2016						
Low	6 (4.1)	6 (7.7)	0 (0.0)	1	0.0	0.000
Intermediate	97 (66.4)	66 (84.6)	31 (45.6)	5–15	32.0	<i>p</i> < 0.001
High	43 (29.5)	6 (7.7)	37 (54.4)	50–90	86.0	<0.001
EU-TiRADS 2017						
2 Benign	4 (2.7)	4 (5.1)	0 (0.0)	0	0.0	<i>p</i> < 0.001
3 Low risk	48 (32.9)	37 (47.4)	11 (16.2)	2–4	22.9	0.058
4 Intermediate risk	48 (32.9)	31 (39.8)	17 (25.0)	6–17	35.4	<0.001
5 High risk	46 (31.5)	6 (7.7)	40 (58.8)	26–87	87	0.059

N.A. not applicable

Table 4 Diagnostic performances of three ultrasound scoring systems compared with cytological and histological findings

FNAB and cytological exam					
Guidelines	Sensitivity	Specificity	PPV	NPV	Accuracy
All nodules (<i>n</i> = 146)					
ATA 2015	81%	3%	76%	4%	64%
ACEE/ AME 2016	58%	13%	71%	8%	49%
EU- TiRADS 2017	68%	23%	77%	16%	58%
FNAB and histological exam					
Guidelines	Sensitivity	Specificity	PPV	NPV	Accuracy
All nodules (<i>n</i> = 146)					
ATA 2015	74%	7%	42%	22%	38%
ACEE/ AME 2016	62%	34%	46%	50%	47%
EU- TiRADS 2017	68%	29%	46%	50%	47%

PPV positive predictive value, NPV negative predicting value

unnecessary biopsy rate decreased to 57.0% compared with 94.3%, increasing specificity from 5 to 43%. For the EUTIRADS and AACE/ACE/AME guidelines, the application of the dimensional criterion >20 mm would allow to have a specificity of 39 and 33%, respectively, with a sensitivity of 53 and 58%.

In the present series, a very small proportion of cases (1 NIFTP, 2 WDT-UMP) had ultrasound features consistent with a low malignancy suspicion (predominantly solid, isoechoic pattern, regular margins and no micro- or macrocalcifications) and were therefore included in the following categories: low-risk by ATA, intermediate by AACE/ACE/AME, and class 3 by EUTIRADS systems. In all cases, FNA was performed because of the dimensional criterion (>20 mm), and the cytological results were TIR3B for the histologically proven NIFTP and TIR2 and TIR3B for the two WDT-UMP tumors.

The ATA, EUTIRADS and AACE/ACE/AME sonographic stratification systems failed to identify 18, 22, and 26 histologically malignant nodules, respectively. As it can be seen in Table 5, the diagnostic performance, and false-negative rates of the three systems appeared to be influenced by the cutoff of the nodule size selected for biopsy recommendation.

Discussion

Our study compared the US-scoring systems, combined with cytological characterization and post-surgical histological diagnoses in a relatively large series of consecutive thyroid nodules undergoing surgery.

Although a good agreement was observed among the results of the three systems, our findings showed that each US-RSS has its own peculiarity. The ATA scoring system has higher sensitivity than the AACE/ACE/AME and EUTIRADS classifications, while specificity was lower. The ATA score could perform better because it requires a second parameter beyond the hypoechoic pattern, reducing overestimation of the other two US-RSS system. Our recent study has demonstrated the applicability of a score (U score) that combines the presence of at least two US features of suspected malignancy to proceed to FNA, regardless of the specific predictive value of each US feature [19]. When the accuracy of these scoring systems was assessed through the subsequent histological outcome, the ATA scoring system showed an accuracy of 38% in differentiating malignant from benign conditions as compared to the EUTIRADS classification, which retained a better accuracy of 47%, with a 29% specificity and 68% sensitivity. These EUTIRADS values appear like those obtained by the AACE/ACE/AME scoring system.

We showed that the categorization of low- and high-risk thyroid nodules using current US-RSSs, regardless of the lesion dimension, helps to determine the optimal treatment option. On the other hand, for intermediate-suspicion categories, a thyroid nodule size cutoff >20 mm for biopsy appeared to improve diagnostic accuracy, according to all the three guidelines. In fact, in our series US-RSS systems seem to provide a reliable stratification of low- and high-risk lesions; specifically, when the risk score of a nodule was low, the overall risk of malignancy varied from 0%, according to the AACE/ACE/AME system, to 21.2%

Table 5 Characterization of histologically malignant nodules not indicated for FNAB according to three guidelines

Guidelines	N° of malignant nodules not indicated for FNAB	Size max (mm)	Histology report		
			PTC	FTC	MTC
<i>ATA 2015</i>	18/146		PTC (n = 16)	FTC (n = 1)	MTC (n = 1)
Benign	0	N.A.	0	0	0
Very low suspicion	0	N.A.	0	0	0
Low suspicion	2	11–15	2	0	0
Intermediate suspicion	3	<10	3	0	0
High suspicion	13	<10	11	1	1
<i>AACE/ACE/AME 2016</i>	26/146		PTC (n = 22)	FTC (n = 3)	MTC (n = 1)
Low	0	N.A.	0	0	0
Intermediate	13	10–20	10	3	0
High	13	<10	12	0	1
<i>EU-TiRADS 2017</i>	22/146		PTC (n = 19)	FTC (n = 2)	MTC (n = 1)
2 Benign	0	N.A.	0	0	0
3 Low risk	2	10–15	2	0	0
4 Intermediate risk	6	10–15	5	1	0
5 High risk	14	<10	12	1	1

N.A. not applicable, PTC papillary thyroid cancer, FTC follicular thyroid cancer, MTC medullary thyroid cancer

(including NIFTP and WDT-UMP) according to the ATA and EUTIRADS. For these categories, ultrasound follow-up should be recommended because of the expected indolent behavior. Moreover, in the high-risk category, the risk of malignancy was higher than 85% based on US features only. The nodule size might affect the subsequent diagnostic-therapeutic decision. The surgical resection can be considered in case of nodules >10 mm. In the case of subcentimeter nodules with high-risk US features, although the most recent ATA guidelines do not recommend diagnosing low-risk papillary microcarcinoma based on cytology [2], recent data showed that it is better to cytologically diagnose suspicious nodules as papillary microcarcinoma and clearly disclose a diagnosis of carcinoma to patients [20, 21]. We also think that FNA should be more often performed. Without the diagnosis of microcarcinoma, it can be very difficult to persuade patients to undergo a regular checkup. After the diagnosis of microcarcinoma without any high-risk features (as pathological lymph nodes) we offered two management options, active surveillance and surgery, and we asked patients which option they would prefer. However, we recommend active surveillance as the first-line management because of the favorable data regarding active surveillance. These results prompted us to consider US parameters only, regardless of the lesion dimension, in the US ultrasonographic risk stratification, in order to identify nodules that require FNA.

The accuracy remains modest for lesions belonging to the intermediate risk class. Nodules scored as intermediate risk are challenging, and we showed that this is especially true for the AACE/ACE/AME system. This uncertainty possibly results from the inclusion in this category of follicular carcinomas (normally isoechoic or hypoechoic without other ultrasonographic criteria). In our study, about 63% of the intermediate-risk nodules were diagnosed as benign nodules by surgery. Intuitively, these patients should not be immediate candidate to surgery [22], and the dimensional cut-off size [23] may only help to discriminate the need for surgery in patients with potentially benign lesions. In such cases, clinical decisions on surgery can be made after considering the coexistent clinical factors and patient's preferences. Nevertheless, the risk of overdiagnosis of benign nodules and the delayed diagnosis of a follicular thyroid carcinoma remain to be considered.

Moreover, the proposed scoring parameters appeared to be more accurate for papillary cancers than other histological types. In fact, the unsuspecting ultrasonographic presentation of FTC, as well as the recognized limitations of cytological assessment to detect it, represent an important problem in intermediate risk classes; unfortunately, it is known that FTC can be underestimated and accounts for the majority of the false negatives of all stratification systems [24]. However, in our retrospective series, follicular carcinomas masked themselves also in low-risk ATA lesions or

in the EUTIRADS class 3. These nodules underwent surgery due to the large size (mean size was 29.8 mm versus 19.73 mm of all carcinomas).

Our series showed a risk of malignancy in the cytological categories comparable with that reported in the 2017 Bethesda System for Reporting Thyroid Cytopathology and in the Italian consensus for the classification and reporting of thyroid cytology [18, 25], although it was slightly higher in indeterminate and high-risk lesions (TIR3 and higher). This discrepancy could be explained by the selection bias of our series, which only comprised surgically treated patients. A good concordance between cytological and histological reports was observed [18], either after considering or not NIFTP and WDT-UMP as malignant lesions. These are two relatively new entities sharing both benign and malignant cytomorphological features but are generally associated with indolent behavior and have an excellent prognosis, requiring no further treatment after surgery [26–28]. The performance of these systems in borderline conditions (WDTUMP and NIFTP) is unknown. In our hands, these lesions showed mild suspicious US features, but further cases need to be investigated to validate the data.

Focusing on indeterminate lesions, the risk of malignancy for TIR3A was 15.4%, while that for TIR3B was 37.8% if considering NIFTP as a malignant or 35.4% if considering NIFTP as benign lesions. The percentage of NIFTP reported in our series (2.4%) is lower than expected (mean 10–15%) [29–31], and we are inclined to speculate that this low figure may depend on our inclusion of NIFTPs as derived from clinically suspected nodules, while those incidentally found in histological specimens were not included. In all, this evidence suggests a good predictive ability of cytological classifications in defining the risk of malignancy for thyroid nodules.

With regard to nodules classified as TIR3B, our data fit perfectly into the 15–30% probability of malignancy reported by SIAPEC [18]. Moreover, the risk of US score increased towards high risk, (87.5%, 72.7%, 70.0% for the ATA, EU-TIRADS, and AACE/ACE/AME system, respectively). In summary, the use of US-RSS in combination with cytological results is expected to enable personalized treatment (US surveillance, repeat biopsy, diagnostic surgery) for cases of thyroid nodules with TIR3.

Our study has several limitations that must be considered when interpreting our findings. In particular, the retrospective nature, the analysis of US features conducted on static US images and the possibility of selection bias because only patients who underwent surgical resection were included. Therefore, the malignancy rate of the nodules in our study is higher than that of the general population.

In conclusion, clinicians should be aware of the strengths and weaknesses of each of the ultrasound score systems in the management of thyroid nodules [32]. The diagnostic

performance and unnecessary biopsy rates of various guidelines are influenced by nodule size cutoff for biopsy. Regardless of the scoring system that the clinician decides to use, the results of our study could help to simplify the decision-making process regarding the management of low-risk nodules. In addition, in high risk categories, the US-RSS performs better when evaluating at least two ultrasound risk features. Finally, the combined use of ultrasound features, a dimensional criterion and the cytological results should allow for personalized treatment (US surveillance, biopsy repetition, diagnostic surgery) for cases of intermediate-risk thyroid nodules.

Further validation of the malignancy risk of each category is required in a larger, prospective, longitudinal study that evaluates the outcomes and costs regarding adjustment in size thresholds in potential revisions of US-based risk stratification systems.

Funding This work did not receive any specific grant.

Author contributions All authors contributed to the study conception and design. Material preparation and data collection were performed by Loredana Pagano, Enrico Costantino Falco, Alessandro Gambella, Ruth Rossetto, Sara Garberoglio, Francesca Maletta, Donatella Pacchioni, Roberto Garberoglio and Mauro Giulio Papotti. Data analysis and tables designing were performed by Alessandro Bisceglia, Enrico Costantino Falco and Alessandro Gambella. The first draft of the manuscript was written by Loredana Pagano and all authors commented on previous versions of the paper. Ezio Ghigo e Mauro Giulio Papotti verified the analytical methods and supervised the manuscript drafting. All authors read and approved the final paper.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures were in accordance with 1964 Helsinki declaration and its later amendments; the Institutional Review Boards of our hospital approved this study.

Informed consent Patients had consented to data collection.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. C. Durante, et al. The diagnosis and management of thyroid nodules: a review. *JAMA* **319**(9), 914–924 (2018)
2. 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**(1), 1–133 (2016)
3. J.H. Shin, et al. Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean Society of

- Thyroid Radiology Consensus Statement and Recommendations. *Korean J. Radio.* **17**(3), 370–395 (2016)
4. 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. Radio.* **14**(5), 587–595 (2017)
 5. 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. Pr.* **22**(5), 622–39 (2016)
 6. G. Russ, et al. European Thyroid Association Guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: The EU-TIRADS. *Eur. Thyroid J.* **6**(5), 225–237 (2017)
 7. Y. Shen, et al. Comparison of different risk-stratification systems for the diagnosis of benign and malignant thyroid nodules. *Front Oncol.* **9**, 378 (2019)
 8. A. Pandya, et al. Limitations of the 2015 ATA Guidelines for prediction of thyroid cancer: a review of 1947 consecutive aspirations. *J. Clin. Endocrinol. Metab.* **103**(9), 3496–3502 (2018)
 9. A. Lauria Pantano, et al. Differences between ATA, AACE/ACE/AME and ACR TI-RADS ultrasound classifications performance in identifying cytological high-risk thyroid nodules. *Eur. J. Endocrinol.* **178**(6), 595–603 (2018)
 10. G. Grani, et al. Reducing the number of unnecessary thyroid biopsies while improving diagnostic accuracy: toward the “Right” TIRADS. *J. Clin. Endocrinol. Metab.* **104**(1), 95–102 (2019)
 11. M. Bongiovanni, et al. The Bethesda System For Reporting Thyroid Cytopathology: A Meta-analysis. *Acta Cytol.* **56**(4), 333–339 (2012)
 12. M. Andrioli, C. Carzaniga, L. Persani, Standardized Ultrasound Report for Thyroid Nodules: The Endocrinologist’s Viewpoint. *Eur. Thyroid J.* **2**(1), 37–48 (2013)
 13. G. Grani, et al. Taller-than-wide shape: a new definition improves the specificity of TIRADS systems. *Eur. Thyroid J.* **9**(2), 85–91 (2020)
 14. V. Cantisani, et al. Strain US elastography for the characterization of thyroid nodules: advantages and limitation. *Int J. Endocrinol.* **2015**, 908575 (2015)
 15. J.Y. Kwak, E.K. Kim, Ultrasound elastography for thyroid nodules: recent advances. *Ultrasonography* **33**(2), 75–82 (2014)
 16. T. Rago, et al. Elastography: new developments in ultrasound for predicting malignancy in thyroid nodules. *J. Clin. Endocrinol. Metab.* **92**(8), 2917–22 (2007)
 17. D. Tumino, et al. Nodular thyroid disease in the era of precision medicine. *Front Endocrinol. (Lausanne)* **10**, 907 (2019)
 18. F. Nardi, et al. Italian consensus for the classification and reporting of thyroid cytology. *J. Endocrinol. Invest* **37**(6), 593–599 (2014)
 19. L. Pagano, et al. A simplified ultrasonographic score for the prediction of cytologically suspicious thyroid nodules. *Ultraschall Med.* (2020). <https://doi.org/10.1055/a-1033-1351>. Epub ahead of print
 20. J.P. Brito, et al. A clinical framework to facilitate risk stratification when considering an active surveillance alternative to immediate biopsy and surgery in papillary microcarcinoma. *Thyroid* **26**(1), 144–149 (2016)
 21. Y. Ito, A. Miyauchi, H. Oda, Low-risk papillary microcarcinoma of the thyroid: a review of active surveillance trials. *Eur. J. Surg. Oncol.* **44**(3), 307–315 (2018)
 22. M.R. Haymart, et al. Thyroid ultrasound and the increase in diagnosis of low-risk thyroid cancer. *J. Clin. Endocrinol. Metab.* **104**(3), 785–792 (2019)
 23. S.M. Ha, et al. Diagnostic performance of practice guidelines for thyroid nodules: thyroid nodule size versus biopsy rates. *Radiology* **291**(1), 92–99 (2019)
 24. G. Russ, Risk stratification of thyroid nodules on ultrasonography with the French TI-RADS: description and reflections. *Ultrasonography* **35**(1), 25–38 (2016)
 25. E.S. Cibas, S.Z. Ali, The Bethesda system for reporting thyroid cytopathology. *Thyroid* **19**(11), 1159–65 (2009)
 26. Y.E. Nikiforov, et al. Nomenclature revision for encapsulated follicular variant of papillary thyroid carcinoma: a paradigm shift to reduce overtreatment of indolent tumors. *JAMA Oncol.* **2**(8), 1023–1029 (2016)
 27. F. Maletta, et al. Cytological features of “noninvasive follicular thyroid neoplasm with papillary-like nuclear features” and their correlation with tumor histology. *Hum. Pathol.* **54**, 134–42 (2016)
 28. S.Z. Ali, Thyroid cytopathology: Bethesda and beyond. *Acta Cytol.* **55**(1), 4–12 (2011)
 29. Z.W. Baloch, et al. Noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP): a changing paradigm in thyroid surgical pathology and implications for thyroid cytopathology. *Cancer Cytopathol.* **124**(9), 616–20 (2016)
 30. W.C. Faquin, et al. Impact of reclassifying noninvasive follicular variant of papillary thyroid carcinoma on the risk of malignancy in The Bethesda System for reporting thyroid cytopathology. *Cancer Cytopathol.* **124**(3), 181–187 (2016)
 31. K.C. Strickland, et al. The impact of noninvasive follicular variant of papillary thyroid carcinoma on rates of malignancy for fine-needle aspiration diagnostic categories. *Thyroid* **25**(9), 987–92 (2015)
 32. P. Trimboli, C. Durante, Ultrasound risk stratification systems for thyroid nodule: between lights and shadows, we are moving towards a new era. *Endocrine* **69**, 1–4 (2020)