

# Differences between ATA, AACE/ACE/AME and ACR TI-RADS ultrasound classifications performance in identifying cytological high-risk thyroid nodules

A Lauria Pantano<sup>1,\*</sup>, E Maddaloni<sup>1,\*</sup>, S I Briganti<sup>1</sup>, G Beretta Anguissola<sup>1</sup>, E Perrella<sup>2</sup>, C Taffon<sup>2</sup>, A Palermo<sup>1</sup>, P Pozzilli<sup>1</sup>, S Manfrini<sup>1</sup> and A Crescenzi<sup>2</sup>

<sup>1</sup>Units of Endocrinology and Diabetes, Department of Medicine and <sup>2</sup>Pathology, Campus Bio-Medico University of Rome, Rome, Italy

\*(A Lauria Pantano and E Maddaloni contributed equally to this work)

Correspondence  
should be addressed  
to A Lauria Pantano  
**Email**  
a.lauria@unicampus.it

## Abstract

**Objective:** Thyroid ultrasound is crucial for clinical decision in the management of thyroid nodules. In this study, we aimed to estimate and compare the performance of ATA, AACE/ACE/AME and ACR TI-RADS ultrasound classifications in discriminating nodules with high-risk cytology.

**Design:** Cross-sectional study.

**Methods:** 1077 thyroid nodules undergoing fine-needle aspiration were classified according to ATA, AACE/ACE/AME and ACR TI-RADS ultrasound classifications by an automated algorithm. Odds ratios (ORs) and receiver operating characteristic (ROC) curves for high-risk cytology categories (TIR3b, TIR4 and TIR5) were calculated for the different US categories and compared.

**Results:** Cytological categories of risk increased together with all US classifications' sonographic patterns ( $P < 0.001$ ). The diagnostic performance (C-index) of ACR TI-RADS and AACE/ACE/AME significantly improved when adding clinical data as gender and age in the regression model ( $P < 0.001$ ). A significant difference in the final model C-index between the three US classification systems was found ( $P < 0.029$ ), with the ACR TI-RADS showing the highest nominal C-index value, significantly superior to ATA ( $P = 0.008$ ), but similar to AACE/ACE/AME ( $P = 0.287$ ). ATA classification was not able to classify 54 nodules, which showed a significant 7 times higher risk of high-risk cytology than the 'very low suspicion' nodules (OR: 7.20 (95% confidence interval: 2.44–21.24),  $P < 0.001$ ).

**Conclusions:** The ACR TI-RADS classification system has the highest area under the ROC curve for the identification of cytological high-risk nodules. ATA classification leaves 'unclassified' nodules at relatively high risk of malignancy.

European Journal of  
Endocrinology  
(2018) **178**, 595–603

## Introduction

Thyroid nodules are a common clinical problem affecting up to two-thirds of the general population (1, 2), with 7–15 percent that are definitely proved to be thyroid cancer depending on age, sex, radiation exposure history, family history and other factors (3, 4). Since about 90% of thyroid

nodules are benign, it is crucial to correctly stratify the malignancy risk of the nodules to avoid a huge number of unnecessary invasive procedures and/or surgery.

Thyroid ultrasound (US) examination is considered the gold standard for the initial stratification of thyroid

lesions, thus determining the need for fine-needle aspiration (FNA) (5). Cytological results from FNA are the main harbinger of clinical decisions, indicating whether surgery should be advised. However, an improper use of FNA has the risk of increasing healthcare expenditures and even of inappropriately referring patients to surgery in case of indeterminate cytology. Thus, a correct US screening is a crucial step for adequate management of such prevalent disease.

In order to achieve a reliable identification of thyroid nodules with cancer risk, specific US features suggestive of malignancy have been recognized and described in literature, namely hypoechogenicity, irregular or blurred margins, microcalcifications, taller than wider shape and vascular signals (6, 7). These features have been included in different US classifications developed over the years (8, 9, 10, 11). In 2016, the American Thyroid Association (ATA) and subsequently the American Association of Clinical Endocrinologists (AACE), American College of Endocrinology (ACE) and Associazione Medici Endocrinologi (AME) published US classification of thyroid nodules to facilitate the selection of nodules for FNA cytological analysis (12, 13). The revised ATA guidelines (12) propose five sonographic patterns related to increasing risk of malignancy: benign, very low, low, intermediate and high suspicion. Recently, the updated joint AACE/ACE/AME 'Medical Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules' (13) also proposed a different classification of thyroid nodules contemplating the use of three US malignancy risk categories: low (class 1), intermediate (class 2) and high (class 3). More recently, the American College of Radiology (ACR) has proposed a new US-based risk stratification system, the Thyroid Imaging Reporting and Data System (TI-RADS), to provide guidance regarding management of thyroid nodules on the basis of their US appearance (14). All three guidelines for the management of thyroid nodules recommend using the proposed US criteria together with nodule size thresholds to decide whether an FNA should be advised.

While the ATA system and the ACR TI-RADS have been widely used in clinical studies, the performance of the AACE/ACE/AME classification has barely been analyzed in comparison with the other two classifications. The purpose of this study is to estimate and compare the performance of ATA, ACR TI-RADS and AACE/ACE/AME US classification systems in discriminating nodules that are more likely to have a high-risk cytological result.

## Subjects and methods

The protocol of this cross-sectional study complies with the Declaration of Helsinki and was approved by the Ethic Committee of the Campus Bio-Medico University, Rome, Italy. Data about US features of all nodules undergoing FNA from January 2015 to May 2016 at the Unit of Endocrinology and Diabetes of the Campus Bio-Medico University of Rome were retrieved from medical records. As this is an observational real-world study, nodules were triaged for biopsy by clinicians in charge of the patients. Therefore, all FNAs were performed based on an impartial clinical indication, independent from the study. US scan of thyroid gland and neck area and US-guided FNA of thyroid focal nodules were performed by experienced physicians at the Unit of Endocrinology and Diabetes (ALP, GBA). Thyroid US was performed at a frequency range of 10–12 MHz on a MyLab 50 ( Esaote, Genova, Italy). Nodules were then classified according to ATA, AACE/ACE/AME US and ACR TI-RADS risk stratification criteria without prior knowledge of the cytological results. To allow an unbiased and blinded classification, each nodule was categorized by an automated algorithm. Briefly, based on the description retrieved from medical records, a yes or no answer to each of the following features derived from the ATA, AACE/ACE/AME and ACR guidelines (12, 13, 14) were input, for each nodule, into a Microsoft Excel worksheet: purely cystic, more than 50% cystic, eccentric solid area, spongiform, spongiform with internal vascularization, mixed cystic and solid, solid hypoechoic (or slightly hypoechoic according to the AACE/ACE/AME), solid marked (or very) hypoechoic, solid isoechoic, hyperechoic, macrocalcifications, microcalcifications, internal hyperechoic spots, calcified rim, irregular margins, taller than wide shape, rim calcifications with small extrusive soft tissue component, evidence of extrathyroidal extension/suspicious nodes. Then, by using a pre-specified coding developed according to the above-mentioned guidelines, the software combined all the yes or no answers and automatically assigned one ATA, one AACE/ACE/AME and one ACR TI-RADS category to each nodule (output). The investigator responsible for data input was blinded for cytological results.

FNA was performed by freehand technique under US guidance, using a 23- or 25-gauge needle. On average, 1–2 passes were performed for each nodule. Crossing thyroid vessels was avoided to prevent local bleeding; in mixed nodules, solid areas were chosen.

Cytology specimens were evaluated by expert cytopathologists (EP, CT) at the Unit of Pathology conforming to Italian Reporting System for Thyroid Cytology (15) as follows: TIR1 (non-diagnostic), TIR1C (non-diagnostic-cystic), TIR 2 non-malignant/benign, TIR3a (low-risk indeterminate lesion), TIR3b (high-risk indeterminate lesion), TIR 4 (suspicious of malignancy) or TIR 5 (malignant). Nodules with TIR1 cytology were excluded from the study. Nodules with TIR1c cytology were included in the study and were considered clinically non-malignant/benign. Non-diagnostic cytological specimens resulted from both solid and cystic lesions. Thyroid cancers are predominantly solid; in this setting, a scant cellular sample is deemed non-diagnostic. Cyst fluid may yield only macrophages, but the risk of malignancy is low for these lesions if they are simple and under 3 cm. For this reason, these cases are separated from inadequate specimens obtained from solid lesions and are reported as non-diagnostic followed by the subcategory 'cyst fluid only'. In the proper clinical setting (e.g., US evidence of a simple, unilocular cyst), these specimens may be considered clinically adequate, even though they are reported as non-diagnostic (16).

All cytologic samples were also reviewed by a third pathologist (AC). In case of disagreement, definitive reporting was achieved by mutual consensus. A preliminary analysis of 251 nodules with definitive pathology diagnosis served as internal validation to confirm the significant risk of malignancy associated to TIR3b, TIR4 and TIR5 categories (Supplementary material, see section on supplementary data given at the end of this article).

### Statistical analysis

Values are expressed as mean  $\pm$  s.d. or median (interquartile range (IQR)) for continuous variables and as proportions for categorical variables (%). Shapiro–Wilk normality test was used to assess the normality of continuous variables

distributions (variables with Shapiro–Wilk statistic  $<0.9$ ;  $P$  values  $<0.05$  were considered non-normally distributed). Groups were compared by analysis of variance, Kruskal–Wallis, chi-square and Fisher's exact test depending on distribution.

Binary logistic regression was used to evaluate the odds ratios (ORs) of malignant cytology based on one or more predictors, and the areas under the receiving operator characteristic (ROC) curves, or C-index, of regression models were tested for equality (17). Nonparametric variables were natural log-transformed before testing in the model. The binary dependent variable 'cytological high-risk nodules' was defined according to above-mentioned preliminary analysis (Supplementary data). In particular, the cytological categories TIR3B, TIR4 and TIR5 were considered as 'cytological high-risk' categories. Sensitivity, specificity, positive and negative predictive values for 'cytological high-risk' were calculated for each US category separately in the three classifications.

In the context of this study, sensitivity is the probability that each US category will include a lesion with high-risk cytology. Likewise, specificity is the likelihood that a US category will exclude a lesion with high-risk cytology. Positive predictive value (PPV) is the percentage of nodules in a given US category with a high-risk cytology diagnosis. Conversely, negative predictive value (NPV) is the percentage of nodules in a given US category without a high-risk cytology diagnosis.

Statistical significance was set for  $P$  values  $<0.05$ . Stata/IC 12.1 software (StataCorp) and Prism 7.0a Software (GraphPad Software) were used for data analysis and graphic representations.

### Results

A total of 1169 thyroid nodules in 946 patients (79% females) aged between 16 and 88 years (mean age  $\pm$  s.d. 56.0  $\pm$  13.3) were screened. After exclusion of

**Table 1** ATA US classification system in relation to cytology.

|       | ATA unclassified | Benign | Very low suspicion | Low suspicion | Intermediate suspicion | High suspicion | Tot   |
|-------|------------------|--------|--------------------|---------------|------------------------|----------------|-------|
| TIR1C | 1                | 4      | 9                  | 7             | 6                      | 2              | 29    |
| TIR2  | 35               | 2      | 175                | 132           | 323                    | 61             | 728   |
| TIR3A | 9                | 0      | 32                 | 33            | 111                    | 22             | 207   |
| TIR3B | 2                | 0      | 2                  | 8             | 19                     | 6              | 37    |
| TIR4  | 1                | 0      | 2                  | 3             | 5                      | 9              | 20    |
| TIR5  | 6                | 0      | 2                  | 0             | 16                     | 32             | 56    |
| Tot   | 54               | 6      | 222                | 183           | 480                    | 132            | 1,077 |

TIR1 nodules were excluded.  $P$  value for distribution of proportions among categories:  $<0.001$ .

**Table 2** AACE/ACE/AME US classification system in relation to cytology.

|       | AACE/AME/FNC<br>unclassified | Class I | Class II | Class III | Tot   |
|-------|------------------------------|---------|----------|-----------|-------|
| TIR1C | 7                            | 12      | 4        | 6         | 29    |
| TIR2  | 17                           | 120     | 347      | 244       | 728   |
| TIR3A | 3                            | 19      | 83       | 102       | 207   |
| TIR3B | 1                            | 1       | 13       | 22        | 37    |
| TIR4  | 0                            | 2       | 4        | 14        | 20    |
| TIR5  | 0                            | 0       | 5        | 51        | 56    |
| Tot   | 28                           | 154     | 456      | 439       | 1.077 |

TIR1 nodules were excluded. *P* value for distribution of proportions among categories: <0.001.

92 nodules with non-diagnostic cytology (TIR1), 1077 nodules (diameter range: 4–56 mm, median (25th–75th percentile): 14 mm (10–20)) were included in the study. Of these, 113 (10.5%) were classified as cytologically high risk (37 TIR3b, 20 TIR4 and 56 TIR5). Of the 964 cytologically benign nodules, 29 were TIR1c, 728 were TIR2 and 207 were TIR3a. The distribution of cytology categories in the different sonographic patterns according to the ATA, AACE/ACE/AME and ACR TI-RADS classification systems is reported in Tables 1, 2 and 3 respectively. Overall, the cytological categories of risk increased together with US classifications' sonographic patterns ( $P < 0.001$ ). Fifty-four nodules did not match any sonographic pattern proposed by the ATA and were categorized as 'ATA unclassified'. Nine (16.7%) of the 'ATA unclassified' nodules were cytologically high risk (2 were TIR3b, 1 was TIR4 and 6 were TIR5). A lower number of nodules ( $n = 28$ ) did not match the US categories proposed by the AACE/ACE/AME and were categorized as 'AACE/ACE/AME unclassified'. Of these, only one was cytologically high risk (TIR3b). All nodules match one US category proposed by the ACR.

### Malignancy risk associated with US categories

AACE/ACE/AME class 3 nodules were 12 times more likely of being cytological high risk than class 1 nodules

(OR: 12.44 (95% CI: 3.87–39.95),  $P < 0.001$ ). Class 2 nodules also showed an increased risk compared to class 1, even though this was not statistically significant (Fig. 1A). The few nodules that remained unclassified by using the AACE/ACE/AME system showed a comparable risk to class 1 nodules.

According to the sonographic patterns proposed by the ATA, all six nodules classified as 'benign' showed a cytology consistent with non-malignant nodules (four were TIR1c and two were TIR2). When compared to the 'very low suspicion' category, both the 'intermediate suspicion' (OR: 3.27 (95% CI: 1.37–7.83),  $P = 0.008$ ) and the 'high suspicion' categories (OR: 19.91 (95% CI: 8.21–48.29),  $P < 0.001$ ) showed a significant increased risk of malignant cytology. Of note, the nodules that remained unclassified by using the ATA US classification system had a significant 7 times higher risk than the 'very low suspicion' nodules (OR: 7.20 (95% CI: 2.44–21.24),  $P < 0.001$ ) (Fig. 1B).

Benign (TR1) nodules according to the ACR TI-RADS classification system had similar odds of malignancy when compared to not suspicious nodules (TR2). There was a stepwise increased risk of malignancy for nodules categorized within the TR3 (OR: 2.77 (95% CI: 0.97–7.92),  $P = 0.057$ ), TR4 (OR: 4.08 (95% CI: 1.60–10.42),  $P = 0.003$ ) and TR5 categories when compared to not suspicious nodules (OR: 24.63 (95% CI: 9.45–64.23),  $P < 0.001$ ) (Fig. 1C).

An inverse relationship between nodule diameter and risk of malignancy was found ( $\beta = -0.042$  (95% CI:  $-0.069$  to  $-0.015$ ),  $P = 0.002$ ) but disappeared after adjustment for US categories.

### Diagnostic accuracy of the three US classification systems

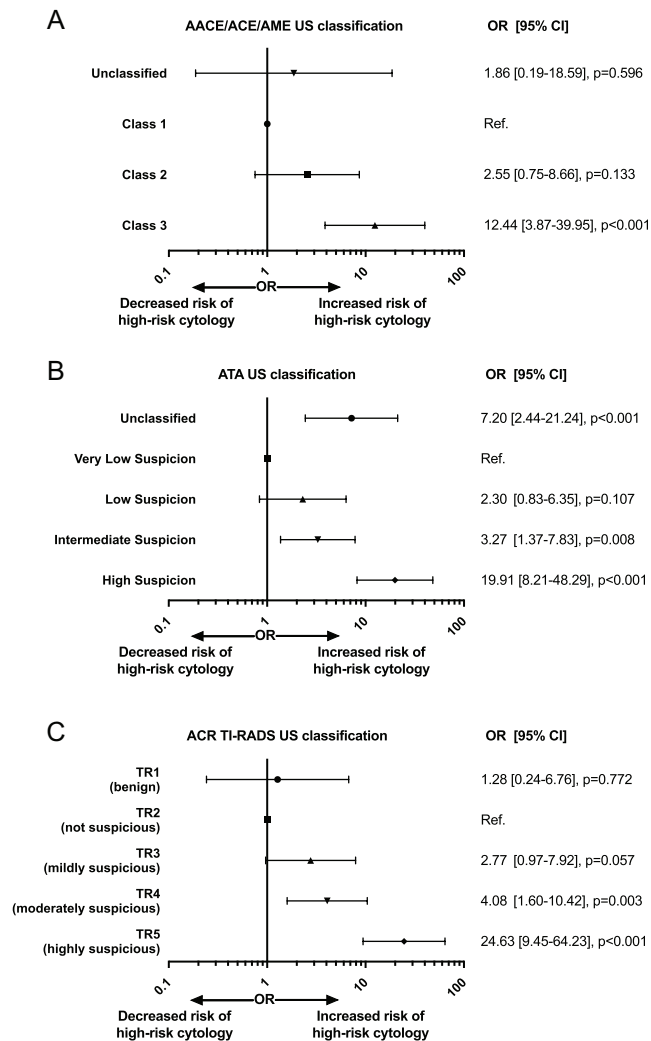
Sensitivity, specificity, PPV and NPV of all ATA AACE/ACE/AME and ACR TI-RADS US categories are reported in Tables 4, 5 and 6, respectively. The area under the ROC

**Table 3** ACR TI-RADS US classification system in relation to cytology.

|       | TR1 (benign) | TR2 (not suspicious) | TR3 (mildly suspicious) | TR4 (moderately suspicious) | TR5 (highly suspicious) | Tot   |
|-------|--------------|----------------------|-------------------------|-----------------------------|-------------------------|-------|
| TIR1C | 17           | 4                    | 1                       | 4                           | 3                       | 29    |
| TIR2  | 38           | 158                  | 149                     | 332                         | 51                      | 728   |
| TIR3A | 6            | 33                   | 33                      | 113                         | 22                      | 207   |
| TIR3B | 1            | 2                    | 10                      | 18                          | 6                       | 37    |
| TIR4  | 1            | 2                    | 1                       | 8                           | 8                       | 20    |
| TIR5  | 0            | 1                    | 2                       | 20                          | 33                      | 56    |
| Tot   | 63           | 200                  | 196                     | 495                         | 123                     | 1.077 |

TIR1 nodules were excluded. *P* value for distribution of proportions among categories: <0.001.

curves (ROC-AUC) for malignant cytology of regression models based on ACR TI-RADS or AACE/ACE/AME US categories significantly improved when adding age and gender in the model ( $P < 0.001$  for both, with younger age and male gender associated with increase malignancy risk).



**Figure 1**

Odds ratio for cytological high-risk nodules by AACE/ACE/AME (A), ATA (B) and ACR TI-RADS (C) US classification systems. (A) Class III nodules showed a significant increased risk for cytological malignancy as defined in the text compared to class I. (B) Intermediate- and high-suspicion nodules had increased risk for cytological malignancy as defined in the text. As well, also unclassified nodules were 7 times more likely to be cytologically malignant than very low-suspicion nodules. (C) A stepwise increased risk of malignancy was found for nodules categorized within the TR3, TR4 and TR5 categories when compared to not suspicious nodules.

The improvement was not significant when age and gender were added in the model based on ATA classification system ( $P=0.161$ ). Overall, there was a significant difference in the contribution of the addition of ATA vs AACE/ACE/AME vs ACR TI-RADS US categories in the model including age and gender for the prediction of high-risk cytology ( $P < 0.029$ ) (Fig. 2). In particular, the comparison of the C-indexes of models all accounting for age and gender, but differing for the US classification system, showed the model with ACR TI-RADS having the highest C-index (0.777 (95% CI: 0.729–0.825)), followed by AACE/ACE/AME (0.763 (95% CI: 0.718–0.808)), while the system proposed by the ATA showed the lowest C-index (0.711 (95% CI: 0.655–0.767)). While the difference in C-indexes between the models with ACR TI-RADS vs AACE/ACE/AME categories was not significant ( $P=0.287$ ), the C-indexes of both models were significantly higher than the C-index of the model accounting for the ATA categories ( $P=0.008$  vs ACR TI-RADS and  $P=0.036$  vs AACE/ACE/AME).

## Discussion

The use of classification systems largely improves communication among clinicians and helps in standardizing clinical practice. When different systems are available, the choice of the best instrument may be a challenge. In this study, we showed that all the ATA, AACE/ACE/AME and ACR TI-RADS US classifications provide an effective malignancy risk stratification for thyroid nodules, based on the FNA result. We also demonstrated that the ability of these schemes in recognizing malignancy can be further improved by considering age and gender. Finally, we found significant differences in the overall performance for the identification of nodules that will result in a high-risk cytology category, with the ACR TI-RADS having the highest C-index, similar to the AACE/ACE/AME, but significantly higher than the ATA scheme. Our data highlight some relevant dissimilarities between the three US systems. When considered separately, the AACE/ACE/AME highest risk category provided the highest sensitivity but low specificity, while the ATA and the ACR TI-RADS highest risk categories provided high specificity but low sensitivity. Furthermore, both the ATA and the AACE/ACE/AME, but not the ACR TI-RADS, tools did not allow the classification of nodules clinicians often deal with (up to 5% in the ATA and up to 2.6% in the AACE/ACE/AME systems). However, only the ATA classification system missed a significant proportion of nodules with a malignant cytology.



**Table 4** Sensitivity, specificity, positive (PPV) and negative (NPV) predictive values for malignant cytology of the ATA sonographic patterns. Data are presented as percentages.

|             | ATA unclassified | Benign | Very low suspicion | Low suspicion | Intermediate suspicion | High suspicion |
|-------------|------------------|--------|--------------------|---------------|------------------------|----------------|
| Sensitivity | 8.0              | 0.0    | 5.3                | 9.7           | 35.4                   | 41.6           |
| Specificity | 95.3             | 99.4   | 77.6               | 82.2          | 54.4                   | 91.2           |
| PPV         | 16.7             | 0.0    | 2.7                | 6.0           | 8.3                    | 35.6           |
| NPV         | 89.8             | 89.4   | 87.5               | 88.6          | 87.8                   | 93.0           |

We also found a lower rate of malignancy within the ATA high-suspicion category (35.6%) and within the AACE/ACE/AME class 3 nodules (19.8%) compared to what was expected based on the data reported by the societies in their guidelines (70–90% and 50–90% respectively). The lower PPV we found could be in part explained by the low proportion of malignant nodules (10.5%), anyhow aligned with the Italian thyroid cancer incidence (18). Before us, other authors also found lower than expected malignancy rates within the high-suspicion pattern compared with the range expected per ATA guidelines (19, 20), nevertheless higher than our data. Different from our study, Yoon *et al.* excluded nodules less than 10mm in the maximum diameter and nodules with indeterminate US-guided FNA, thus probably overestimating the malignancy risk (19). On the contrary, in the recently published paper by Persichetti *et al.* (20), the rate of malignancy of AACE/ACE/AME class III nodules was found in the expected range, as a result of an overall lower proportion of nodules categorized as class III (23.8% vs 40.8% in our series). This study had a retrospective design similar to our study and compared the ATA and the AACE/ACE/AME US stratification systems with that proposed by the British Thyroid Association, claiming for studies like ours evaluating the accuracy of the most recent classification proposed by the ACR TI-RADS (20).

Despite the difference in the C-index of the ATA and the ACR TI-RADS, the predictive values of the highest US categories of both classifications (TR 4 and 5 vs ATA intermediate and high suspicion) were similar. This suggests that the low C-index of the US system proposed

by the ATA is mostly due to the nodules this classification was not able to classify. Similar to previous reports (18), also in our study, nodules that were not classified into a specific ATA sonographic pattern had a relatively high risk of malignancy (OR 7.20), also higher than intermediate-suspicion US-pattern. In our dataset, ATA-unclassified nodules were isoechoic solid nodules with at least one of the following additional features: irregular margins, microcalcifications or mixed calcifications, or nonparallel shape (Fig. 3). While hyper- to isoechoic appearance has been associated to a benign behavior (9, 21, 22) the presence of additional suspicious US features such as microlobulated or irregular margins, microcalcifications or mixed calcifications, or nonparallel shape should suggest the needs for FNA with similar standards to those with indeterminate suspicion patterns (18). In this setting, some studies reported that the follicular variant of papillary thyroid cancer (FVPTC) shows a relatively benign sonographic appearance (23, 24, 25), in particular when larger than 1 cm (26). Of note, about 80% of our cytologically malignant ATA US-unclassified nodules were FVPTC at the definitive histological examination. However, recently, Trimboli *et al.* showed that the ATA classification may aid in the risk stratification of thyroid nodules with indeterminate FNA cytology (27), even though the US risk stratification systems have an overall sub-optimal diagnostic accuracy in discriminating malignant lesions in this setting (28).

A further significant improvement in the ROC areas under the curves was achieved by adding in the model also age and gender. This confirms that clinical decisions should never be made based only on US features, but a complete clinical assessment of patients with thyroid nodules is always needed. The literature indicates higher malignancy rates in individuals below 16 or above 45 years of age (29, 30), while contrasting data about a gender predominance have been published (29). Besides age and gender, other clinical parameters such as TSH levels and thyroid autoimmunity have been shown to improve US diagnostic accuracy in differentiating benign from malignant nodules (31, 32, 33). The AACE/ACE/AME societies also suggest considering elastography as an

**Table 5** Sensitivity, specificity, positive (PPV) and negative (NPV) predictive values for malignant cytology of the AACE/ACE/AME US categories. Data are presented as percent ages.

|             | AACE/ACE/AME unclassified | Class I | Class II | Class III |
|-------------|---------------------------|---------|----------|-----------|
| Sensitivity | 0.9                       | 2.7     | 19.5     | 77.0      |
| Specificity | 97.2                      | 84.3    | 55.0     | 63.5      |
| PPV         | 3.6                       | 1.95    | 4.8      | 19.8      |
| NPV         | 89.3                      | 88.1    | 85.3     | 95.9      |

**Table 6** Sensitivity, specificity, positive (PPV) and negative (NPV) predictive values for malignant cytology of the ACR TI-RADS US categories. Data are presented as percentages.

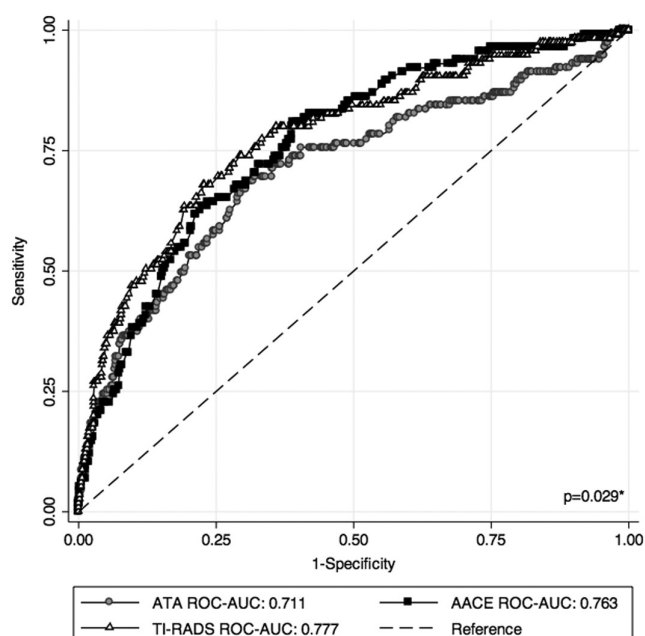
|             | TR1 (benign) | TR2 (not suspicious) | TR3 (mildly suspicious) | TR4 (moderately suspicious) | TR5 (highly suspicious) |
|-------------|--------------|----------------------|-------------------------|-----------------------------|-------------------------|
| Sensitivity | 1.7          | 4.4                  | 11.3                    | 40.9                        | 41.7                    |
| Specificity | 93.7         | 79.8                 | 81.0                    | 53.4                        | 92.1                    |
| PPV         | 3.2          | 2.5                  | 6.6                     | 9.5                         | 38.7                    |
| NPV         | 88.9         | 87.5                 | 88.4                    | 88.3                        | 93.0                    |

additional US technique complimentary to gray-scale US in the evaluation of thyroid nodules. Elastography provides information about nodule stiffness, and it has shown high sensitivity for thyroid carcinoma. The combination of elastography with B-Mode US significantly improved

sensitivity and specificity of US features (34). Therefore, the accuracy of the AACE/ACE/AME categories might improve if using elastography.

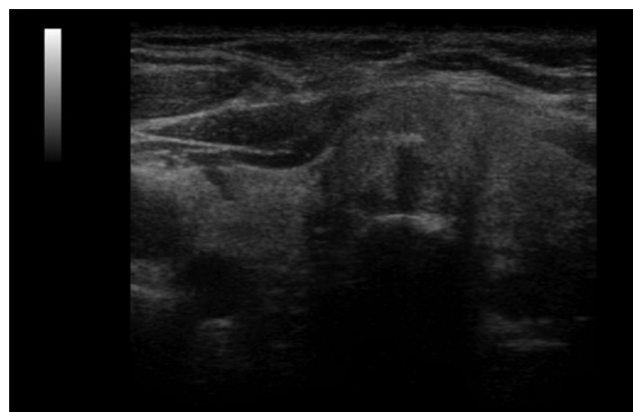
Finally, although larger size should be considered as a risk factor for malignancy (35), in our dataset, an inverse relationship between nodule diameter and risk of malignancy was found, similar to what was shown by Yon *et al.* This relationship, however, disappeared after adjustment for all AACE, ATA and ACR TI-RADS US categories. Overall, this suggests that US features should lead clinical decisions independently from size. In particular, bigger nodules without suspicious US features may probably avoid unnecessary FNA.

There were some limitations to this study. First, the final diagnoses were based on the cytopathology and not on surgical histology, which may cause false-negative and false-positive results. However, the probability of false diagnosis in TIR 2 and TIR 5 categories is low at <3 and <1%, respectively, as compared to histopathology (15). Moreover, the preliminary analysis conducted on nodules with definitive histology confirmed the significant risk of malignancy associated to TIR3b, TIR4 and TIR5 categories



**Figure 2**

Receiver operating characteristic (ROC) curves for the diagnosis of cytological high-risk malignant nodules. The areas under the ROC curve of regression models accounting for age and gender *plus* US categories from ATA classification (gray circles) or AACE/ACE/AME classification (black squares) or ACR TI-RADS classification (white triangles) are shown. The addition of ACR TI-RADS categories resulted in the highest nominal ROC-AUC value (0.777 (95% CI: 0.729–0.825)). This was similar to the ROC-AUC value obtained when AACE/ACE/AME categories were used (0.763 (95% CI: 0.718–0.808),  $P=0.287$  vs ACR TI-RADS ROC-AUC). The addition of categories from the ATA classification resulted in the lowest ROC-AUC value (0.711 (95% CI: 0.655–0.767),  $P=0.008$  vs ACR TI-RADS and  $P=0.036$  vs AACE/ACE/AME). \* $P$ -value for differences between the three models.



**Figure 3**

Ultrasound of a nodule unclassified according to the ATA US classification system. Transverse sonogram of thyroid isthmus shows a solid, isoechoic nodule with regular margins and microcalcification. The histology showed a papillary thyroid carcinoma follicular variant.

(Supplementary material), validating our cytopathology results and further supporting the validity of our data. Second, our report lacks information about some clinical parameters known to be associated with increased risk of malignancy, like thyroid autoimmunity, which deserve to be tested in the final model to evaluate their impact on the performance of US categories, as we did for age and gender. Furthermore, since this is an observational study, we only assessed nodules with a clinical indication to FNA as independently judged by clinicians referring patients to our clinic for the procedure. Therefore, we should assume that most nodules with low pre-test probability of high-risk cytology have not been included in the analysis, introducing a selection bias. Notably, our clinical records include a reasonable percentage of nodules in the lowest US categories of risk (21.2% in the ATA 'benign' and 'very low suspicion' categories, 14.3% in the AACE/ACE/AME 'class I' category and 24.4% in the ACR TI-RADS 'TR2' and 'TR3' categories), partially overcoming this bias. Finally, it is important to acknowledge that this study does not investigate the accuracy of the criteria proposed by the three societies for the final decision of performing or not an FNA. Indeed, all three guidelines suggest evaluating other clinical parameters together with US features, largest dimension in particular, before recommending FNA. In this regard, Xu *et al.* recently showed that nodules' size influence the diagnostic performance of US classification systems (36). Since size thresholds differ between the three guidelines, future studies should address whether the different criteria for FNA have different outcomes and whether size cutoffs should be changed to improve the accuracy of the proposed criteria.

In conclusion, our study shows the US classification systems proposed by the ATA, the AACE/ACE/AME and the ACR differ in their ability for the identification of nodules at high risk of malignancy. In particular, the ACR TI-RADS classification system has the highest ROC-AUC for the identification of cytological high-risk nodules and is the only US scheme able to classify all thyroid nodules. Our results confirm a relevant limit of the ATA classification which leaves 'unclassified' nodules at relatively high risk of malignancy. Finally, our analysis suggests that an improvement in the performance of all classifications is achieved by considering other clinical parameters such as age and gender.

#### Supplementary data

This is linked to the online version of the paper at <https://doi.org/10.1530/EJE-18-0083>.

#### Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of this study.

#### Funding

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

#### References

- 1 Tan GH & Gharib H. Thyroid incidentalomas: management approaches to nonpalpable nodules discovered incidentally on thyroid imaging. *Annals of Internal Medicine* 1997 **126** 226–231. (<https://doi.org/10.7326/0003-4819-126-3-199702010-00009>)
- 2 Guth S, Theune U, Aberle J, Galach A & Bamberger CM. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. *European Journal of Clinical Investigation* 2009 **39** 699–706. (<https://doi.org/10.1111/j.1365-2362.2009.02162.x>)
- 3 Hegedus L. Clinical practice. The thyroid nodule. *New England Journal of Medicine* 2004 **351** 1764–1771. (<https://doi.org/10.1056/NEJMc031436>)
- 4 Mandel SJ. A 64-year-old woman with a thyroid nodule. *JAMA* 2004 **292** 2632–2642. (<https://doi.org/10.1001/jama.292.21.2632>)
- 5 Gharib H, Papini E, Paschke R, Duick DS, Valcavi R, Hegedüs L & Vitti P. AACE/AME/ETA Task Force on Thyroid Nodules. American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and European Thyroid Association medical guidelines for clinical practice for the diagnosis and management of thyroid nodules. *Journal of Endocrinological Investigation* 2010 **33** 1–50. (<https://doi.org/10.1007/BF03346541>)
- 6 Solbiati L, Osti V, Cova L & Tonolini M. Ultrasound of thyroid, parathyroid glands and neck lymph nodes. *European Radiology* 2001 **11** 2411–2424. (<https://doi.org/10.1007/s00330-001-1163-7>)
- 7 Kim EK, Park CS, Chung WY, Oh KK, Kim DI, Lee JT & Yoo HS. New sonographic criteria for recommending fine-needle aspiration biopsy of nonpalpable solid nodules of the thyroid. *American Journal of Roentgenology* 2002 **178** 687–691. (<https://doi.org/10.2214/ajr.178.3.1780687>)
- 8 Frates MC, Benson CB, Charboneau JW, Cibas ES, Clark OH, Coleman BG, Cronan JJ, Doubilet PM, Evans DB, Goellner JR *et al.* Management of thyroid nodules detected at US: Society of Radiologists in Ultrasound consensus conference statement. *Radiology* 2005 **237** 794–800. (<https://doi.org/10.1148/radiol.2373050220>)
- 9 Horvath E, Majlis S, Rossi R, Franco C, Niedmann JP, Castro A & Dominguez M. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. *Journal of Clinical Endocrinology and Metabolism* 2009 **94** 1748–1751. (<https://doi.org/10.1210/jc.2008-1724>)
- 10 Ha EJ, Moon WJ, Na DG, Lee YH, Choi N, Kim SJ & Kim JK. A multicenter prospective validation study for the Korean thyroid imaging reporting and data system in patients with thyroid nodules. *Korean Journal of Radiology* 2016 **17** 811–821. (<https://doi.org/10.3348/kjr.2016.17.5.811>)
- 11 Perros P, Boelaert K, Colley S, Evans C, Evans RM, Gerrard Ba G, Gilbert J, Harrison B, Johnson SJ, Giles TE *et al.* Guidelines for the management of thyroid cancer. *Clinical Endocrinology* 2014 **81** (Supplement 1) 1–122. (<https://doi.org/10.1111/cen.12515>)
- 12 Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, Pacini F, Randolph GW, Sawka AM, Schlumberger M *et al.* 2015 American Thyroid Association Management Guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines task force on



- thyroid nodules and differentiated thyroid cancer. *Thyroid* 2016 **26** 1–133. (<https://doi.org/10.1089/thy.2015.0020>)
- 13 Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedüs L, Paschke R, Valcavi R, Vitti P & AACE/ACE/AME Task Force on Thyroid Nodules. 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. *Endocrine Practice* 2016 **22** 622–639.
  - 14 Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, Cronan JJ, Beland MD, Desser TS, Frates MC *et al.* ACR Thyroid Imaging, Reporting and Data System (TI-RADS): white paper of the ACR TI-RADS Committee. *Journal of the American College of Radiology* 2017 **14** 587–595. (<https://doi.org/10.1016/j.jacr.2017.01.046>)
  - 15 Nardi F, Basolo F, Crescenzi A, Fadda G, Frasoldati A, Orlandi F, Palombini L, Papini E, Zini M, Pontecorvi A *et al.* Italian consensus for the classification and reporting of thyroidcytology. *Journal of Endocrinological Investigation* 2014 **37** 593–599. (<https://doi.org/10.1007/s40618-014-0062-0>)
  - 16 *The Bethesda System for Reporting Thyroid Cytopathology: Definitions, Criteria and Explanatory Notes*, 2 ed. Eds SZ Ali & ES Cibas. Springer International Publishing, 2017.
  - 17 DeLong ER, DeLong DM & Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988 **44** 837–845. (<https://doi.org/10.2307/2531595>)
  - 18 Papini E, Guglielmi R, Bianchini A, Crescenzi A, Taccogna S, Nardi F, Panunzi C, Rinaldi R, Toscano V & Pacella MC. Risk of malignancy in nonpalpable thyroid nodules: predictive value of ultrasound and Color-Doppler features. *Journal of Clinical Endocrinology and Metabolism* 2002 **87** 1941–1946
  - 19 Yoon JH, Lee HS, Kim EK, Moon HJ & Kwak JY. Malignancy risk stratification of thyroid nodules: comparison between the thyroid imaging reporting and data system and the 2014 American Thyroid Association Management Guidelines. *Radiology* 2016 **278** 917–924. (<https://doi.org/10.1148/radiol.2015150056>)
  - 20 Persichetti A, Di Stasio E, Guglielmi R, Bizzarri G, Taccogna S, Misischi I, Graziano F, Petrucci L, Bianchini A & Papini E. Predictive value of malignancy of thyroid nodule ultrasound classification systems. A prospective study. *Journal of Clinical Endocrinology and Metabolism* 2018 **103** 1359–1368. (<https://doi.org/10.1210/jc.2017-01708>)
  - 21 Park JY, Lee HJ, Jang HW *et al.* A proposal for a thyroid imaging reporting and data system for ultrasound features of thyroid carcinoma. *Thyroid* 2009 **19** 1257–1264. (<https://doi.org/10.1089/thy.2008.0021>)
  - 22 Moon WJ, Jung SL, Lee JH, Na DG, Baek JH, Lee YH, Kim J, Kim HS, Byun JS, Lee DH *et al.* Benign and malignant thyroid nodules: US differentiation multicenter retrospective study. *Radiology* 2008 **247** 762–770. (<https://doi.org/10.1148/radiol.2473070944>)
  - 23 Yoo WS, Choi HS, Cho SW, Moon JH, Kim KW, Park HJ, Park SY, Choi SI, Choi SH, Lim S *et al.* The role of ultrasound findings in the management of thyroid nodules with atypia or follicular lesions of undetermined significance. *Clinical Endocrinology* 2014 **80** 735–742. (<https://doi.org/10.1111/cen.12348>)
  - 24 Layfield IJ, Cibas ES & Baloch Z. Thyroid fine needle aspiration cytology: a review of the National Cancer Institute state of the science symposium. *Cytopathology* 2010 **21** 75–85. (<https://doi.org/10.1111/j.1365-2303.2010.00750.x>)
  - 25 Feldt-Rasmussen U. Iodine and cancer. *Thyroid* 2001 **11** 483–486. (<https://doi.org/10.1089/105072501300176435>)
  - 26 Jeon EJ, Jeong YJ, Park SH, Cho CH, Shon HS & Jung ED. Ultrasonographic characteristics of the follicular variant papillary thyroid cancer according to the tumor size. *Journal of Korean Medical Science* 2016 **31** 397–402. (<https://doi.org/10.3346/jkms.2016.31.3.397>)
  - 27 Trimboli P, Deandrea M, Mormile A, Ceriani L, Garino F, Limone PP & Giovanella L. American Thyroid Association ultrasound system for the initial assessment of thyroid nodules: use in stratifying the risk of malignancy of indeterminate lesions. *Head Neck* 2017.
  - 28 Trimboli P, Fulciniti F, Zilioli V, Ceriani L & Giovanella L. Accuracy of international ultrasound risk stratification systems in thyroid lesions cytologically classified as indeterminate. *Diagnostic Cytopathology* 2017 **45** 113–117. (<https://doi.org/10.1002/dc.23651>)
  - 29 Eng CY, Quraishi MS & Bradley PJ. Management of thyroid nodules in adult patients. *Head and Neck Oncology* 2010 **2** 1–5. (<https://doi.org/10.1186/1758-3284-2-1>)
  - 30 *British Thyroid Association, Royal College of Physicians: British Thyroid Association Guidelines for the management of thyroid cancer*, 2nd ed., 2007.
  - 31 Baser H, Topaloglu O, Tam AA, Evranos B, Alkan A, Sungu N, Dumlu EG, Ersoy R & Cakir B. Higher TSH can be used as an additional risk factor in prediction of malignancy in euthyroid thyroid nodules evaluated by cytology based on Bethesda system. *Endocrine* 2016 **53** 520–529. (<https://doi.org/10.1007/s12020-016-0919-4>)
  - 32 Vasileiadis I, Boutzios G, Charitoudis G, Koukouliti E & Karatzas T. Thyroglobulin antibodies could be a potential predictive marker for papillary thyroid carcinoma. *Annals of Surgical Oncology* 2014 **21** 2725–2732. (<https://doi.org/10.1245/s10434-014-3593-x>)
  - 33 Kim ES, Lim DJ, Baek KH, Lee JM, Kim MK, Kwon HS, Song KH, Kang MI, Cha BY, Lee KW *et al.* Thyroglobulin antibody is associated with increased cancer risk in thyroid nodules. *Thyroid* 2010 **20** 885–891. (<https://doi.org/10.1089/thy.2009.0384>)
  - 34 Trimboli P, Guglielmi R, Monti S, Misischi I, Graziano F, Nasrollah N, Amendola S, Morgante SN, Deiana MG, Valabrega S *et al.* Ultrasound sensitivity for thyroid malignancy is increased by real-time elastography: a prospective multicenter study. *Journal of Clinical Endocrinology and Metabolism* 2012 **97** 4524–4530. (<https://doi.org/10.1210/jc.2012-2951>)
  - 35 Trimboli P, Treglia G, Guidobaldi L, Saggiorato E, Nigri G, Crescenzi A, Romanelli F, Orlandi F, Valabrega S, Sadeghi R *et al.* Clinical characteristics as predictors of malignancy in patients with indeterminate thyroid cytology: a meta-analysis. *Endocrine* 2014 **46** 52–59. (<https://doi.org/10.1007/s12020-013-0057-1>)
  - 36 Xu T, Gu JY, Ye XH, Xu SH, Wu Y, Shao XY, Liu DZ, Lu WP, Hua F, Shi BM *et al.* Thyroid nodule sizes influence the diagnostic performance of TIRADS and ultrasound patterns of 2015 ATA guidelines: a multicenter retrospective study. *Scientific Reports* 2017 **7** 43183. (<https://doi.org/10.1038/srep43183>)

Received 30 January 2018

Revised version received 14 March 2018

Accepted 6 April 2018