Comparison of different systems of ultrasound (US) risk stratification for malignancy in elderly patients with thyroid nodules. Real world experience

Fernando Di Fermo1 · Noelia Sforza1 · Melanie Rosmarin1 · Yanina Morosan Allo1 · Carina Parisi1 · Jimena Santamaria1 · Nestor Pacenza1 · Carlos Zuk2 · Cristina Faingold1 · Florencia Ferraro3 · Tomas Meroño3 · Gabriela Brenta1

Received: 10 December 2019 / Accepted: 28 March 2020 / Published online: 14 April 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Purpose To comparatively assess the performance of three sonographic classification systems, American Thyroid Association (ATA), the American College of Radiology Thyroid Imaging Reporting and Data System (ACR TI-RADS), and American Association of Clinical Endocrinologists (AACE)/American College of Endocrinology (ACE)/Associazione Medici Endocrinologi (AME) in identifying malignant nodules in an elderly population.

Methods Cross-sectional study of patients referred for fine needle aspiration biopsy in an academic center for the elderly. One nodule/patient was considered. Nodules classified Bethesda V/VI were considered malignant. Receiver operating characteristics (ROC) curves were established and compared to evaluate diagnostic performance. Malignancy among biopsies below the size cutoff for each ultrasound classification was also compared.

Results One thousand, eight hundred sixty-seven patients (92% females); median (Q1–Q3), age 71 (67–76) years, were studied showing 82.8% benign (Bethesda II) and 2.6% malignant cytology. The three classifications correctly identified malignancy (P < 0.01). Nonetheless, in the ATA and AACE/ACE/AME 16 and 2 malignant nodules, respectively, were unclassifiable. Including unclassifiable malignant nodules (n = 1234, malignant = 50), comparison of the ROC curves showed lower performance of ATA [area under the curve (AUC) = ATA (0.49) vs. ACR TI-RADS (0.62), p = 0.008 and ATA vs. AACE/ACE/AME (0.59), p = 0.022]. Proportion of below size cutoff biopsies for ATA, ACR TI-RADS, and AACE/ACE/AME was different (16, 42, and 29% (all p < 0.001)], but no differences in malignancy rate were observed in these nodules.

Conclusion The present study is the first to validate in elderly patients these classifications showing that AACE/ACE/AME and ACR TI-RADS can predict thyroid malignancy more accurately than the ATA when unclassifiable malignant nodules are considered. Moreover, in this aged segment of the population, the use of ACR TI-RADS avoided more invasive procedures.

Keywords Elderly patients · Ultrasound characteristics · Thyroid nodules · Malignant cytology

Introduction

The prevalence of thyroid nodular disease may vary if studied by palpation of the thyroid gland (detected nodules in about 5% of the population) or by ultrasound (US) (65% or more individuals with a nodule in the thyroid gland) [1].

Age is a risk factor for nodular disease since thyroid nodules are more frequently found in the elderly than in the general population [2]. In fact, it has been shown that multinodular thyroid disease is 30% more common in individuals over 70 years of age [3]. These data are in line with the presence of larger goiters associated with a retro
sternal presentation and deviation of the trachea at advanced age [4]. Interestingly, the risk of malignancy in thyroid nodules may decrease in elderly patients [3]. It has been previously described that between ages 20 and 60 years, with each passing year, there is a 2.2% reduction in the relative risk of a nodule to become malignant. In fact, at 20–29 years, each nodule harbors 14.8% risk of malignancy while that proportion drops to 5.6% after the age of 70 [3]. However, in the elderly, histological variants are more aggressive [5]. Thus the discovery of thyroid nodules with medullary, anaplastic, poorly-differentiated carcinoma, and distant metastasis is more frequent than in the young. Indeed, in patients over 40 years of age there is 7.0% increase per year in the relative risk of finding more aggressive cancer variants, a phenomenon partially explained by a delay in diagnosis [3].

Faced with this scenario it may be speculated that the indiscriminate use of thyroid US can result in overdiagnosis of thyroid nodules in the elderly patients, most of which result benign once biopsied. Nevertheless, a too stringent policy may neglect the early diagnosis of aggressive thyroid cancer present at this age. There are several US classifications that group echographic features of thyroid nodules into categories that stratify its malignant potential and may help to guide fine needle aspiration biopsy (FNAB). Nowadays, three internationally endorsed sonographic classification systems have been issued, the American Thyroid Association (ATA) [6], the American Association of Clinical Endocrinologists (AACE), American College of Endocrinology (ACE) and Associazione Medici Endocrinologi (AME) [7], and the American College of Radiology Thyroid Imaging Reporting and Data System (ACR TI-RADS) [8].

Although the elderly population may benefit from the use of any of these three US classifications, there is no available data comparing them. The aim of this study is to compare the efficacy of these three US classifications in finding malignant cytology in an elderly cohort with thyroid nodules.

Methods

Study cohort and protocol

This is a cross-sectional study of consecutive patients with thyroid nodules referred for US guided-FNAB in an academic referral center for the elderly, Dr. Cesar Milstein Hospital, which receives all the patients belonging to the National Institute of Social Services for Retirees and Pensioners in the city of Buenos Aires, an iodine sufficient metropolitan area. The US features to be analyzed were collected in a consecutive way and their distribution into US guidelines categories was done retrospectively.

Since our hospital is a referral center for FNAB from other medical institutions, biopsy was performed according to the indication of each referring physician. In case of multiple nodules, the presence of suspicious US characteristics was used for the selection of the nodule to be biopsied. Clinical criteria to refer a patient to US guided-FNAB were either a neck mass that was visible or palpable, or that had been found incidentally in a previous imaging study. All US-guided FNABs were carried out by one of three operators of our institution, each having more than 20 years experience with this procedure. In order to avoid a large interobserver variability as previously described for single suspicious features [9], only these three experienced clinicians were responsible of describing all the individual US characteristics of the nodules. All this information was filled in by a technician into a specific form online immediately before the FNAB. At the Endocrine Department, this form was used to complete a database with all the information of each patient. Since 2018, the risk categories according to each of the three US systems was calculated for each patient and registered in the database.

Only one nodule/patient was considered for this analysis. In the case of two or more coexisting nodules, we selected for statistical analysis the one with a malignant cytological finding. If all nodules were benign, the one with the highest US category risk was selected. The decision of one nodule/patient was based on the idea that if several nodules were biopsied but only one was malignant; this patient would be referred to surgery based on this specific nodule. Furthermore, in the case of multiple nodules, solid nodules with suspicious US findings were the ones initially biopsied considering the other nodules as less relevant. Since this was the criteria chosen by physicians performing the FNAB, the inclusion of the second nodule would render the malignant cytology prevalence to a minimum. This common criteria shared by the sonographists is part of the study of only one-nodule per patient. Last, there were also nodules that were followed in time and were subjected to more than one FNAB during the study period. Given that the same nodule was analyzed at different time points, including every FNAB in the study would have introduced bias. All nodules with Bethesda V/VI cytology were considered malignant.

During a 10 year period, June 2008–June 2018, 1867 patients (92% females; aged, median (Q1–Q3), 71 (67–76) years) were consecutively included for the study. The total amount of biopsied nodules was 2400 but only one nodule per patient was considered. After exclusion of indeterminate (20% rate of malignancy at our Institution) and insufficient cytology results (n = 271), a subpopulation of 1596 nodules with benign and malignant cytological results was obtained (Fig. 1). Clinical and biochemical characteristics, such as age, sex, previous exposure to radiotherapy, family history...
of thyroid cancer, personal history of diabetes, thyroid peroxidase antibody (TPOab) positivity and TSH levels, as well as US details were prospectively collected.

The study was approved by the Ethical Committee of our Institution and all patients signed an informed consent form.

**Image analysis**

Prior to each biopsy all US characteristics were assessed with real-time US in each thyroid nodule. These included type of echostructure [solid, mixed (>25% cystic proportion), spongiform, and purely cystic], echogenic pattern [Hypoechoic = the nodule echogenicity was compared with normal thyroid (mild hypoechoic) and strap muscles (markedly hypoechoic), isoechoic, hyperechoic, and anechoic], margins (irregular or regular), presence of halo (yes or no), microcalcifications (defined as tiny, punctate hyperechoic foci, without comet-tail sign, and distinct of indeterminate hyperechoic spots), macrocalcifications (defined as coarse areas of calcification >1 mm in size), and the 3 diameters of the nodule in mm (taller than wide nodules were defined when the anterior–posterior dimension exceeded the axial dimension).

The three US classifications: ATA [6], the AACE/ACE/AME [7], and the ACR TI-RADS [8] were applied based on US findings. According to the suspicion of malignancy, the classification proposed by the ATA divided nodules into five classes: 1 (benign), 2 (very low suspicion), 3 (low suspicion), 4 (intermediate suspicion), and 5 (high suspicion). According to the ACR TI-RADS score, following a sum of points awarded according to the US findings, nodules were divided into the following levels of suspicion of malignancy: TR 1 (benign), TR 2 (not suspect), TR 3 (very low suspicion), TR 4 (moderately suspect) and TR 5 (highly suspicious) and according to the AACE/ACE/AME guide the risk of malignancy of the lesions was divided into 1 (low risk), 2 (medium), and 3 (high).

FNABs are usually recommended by each US classification above a certain threshold of size. ATA recommends diagnostic FNAB for thyroid nodules ≥1 cm of high suspicion and intermediate suspicion, low suspicion ≥1.5 cm, and very low suspicion ≥2 cm [6]. The AACE/ACE/AME proposes diagnostic FNAB for thyroid nodules with Class 1 if ≥2.0 cm + increasing size or high-risk history, Class 2 if ≥2 cm and Class 3 if ≥1.0 or ≥0.5 cm + subcapsular or paratracheal lesions, suspicious lymph nodes or extrathyroid spread, positive personal or family history of thyroid cancer, history of head and neck irradiation, coexistent suspicious clinical findings (e.g., dysphonia) [7]. The ACR TI-RADS system recommends diagnostic FNAB for thyroid nodules with TR 3 ≥2.5 cm, TR 4 ≥1.5 cm, and TR 5 ≥1 cm disregarding diagnostic FNAB for TR 1 and TR 2 nodules [8]. In order to investigate the value of these recommendations, all the nodules that were below the recommended size cutoff of each classification were analyzed.

**US-guided fine needle aspiration procedure**

A Mindray DC-3 (Shenzhen, China) Doppler-echo machine and a 7.5–10 MHz linear-array probe were used to guide all FNABs in real time. Biopsies were performed using a 23-gauge needle, and visualization of the tip of the needle inside the nodule helped to monitor the correct site for biopsy. At least 2–6 needle passages were performed in each nodule. Material obtained from FNABs was smeared on glass slides, which were immediately placed in 95% alcohol for Papanicolaou stain and sent to the Pathology Department.

**Cytological analysis**

This study used the Bethesda System for Reporting Thyroid Cytopathology to describe the cytological results [10]. The results of those nodules included in 2008 were adapted to
the Bethesda System. Cytological analysis was performed independently by two pathologists. Validation of this procedure by cytohistological correlation in our Institution was previously reported [11]. Those patients with malignant cytology nodules that were referred to surgery at our Institution were also considered for a descriptive analysis (n = 31).

**Statistical analysis**

Chi2 and logistic regression were used to evaluate and compare malignant cytology within each US classification. Receiver operating characteristics (ROC) curves were established to compare diagnostic performance. The cutoff with the highest Youden Index was used to calculate the sensitivity and specificity. Since some malignant nodules remained unclassifiable according to the ATA and AACE/ACE/AME US classification systems, we performed two ROC curve comparison analyses. One after exclusion of unclassifiable benign nodules and categorizing unclassifiable malignant nodules as the lowest risk category (n = 1234, 16 malignant nodules reclassified for ATA and AACE/ACE/AME combined) and the other excluding all unclassifiable nodules whether benign or not (n = 1218, malignant nodules = 34). For paired comparisons between the area under the curve (AUC) of the US systems we used DeLong method [12].

The nodules that were biopsied even when their size was lower than the guidelines recommendations were classified as “below size cutoffs” and pairs were compared using the McNemar test across the different US systems. The false negative (malignant nodules in the “below size cutoffs” biopsies) rate (FNR) was also calculated for each US classification system.

Normally distributed variables are presented as mean ± S. D. and skewed variables as median (Q1–Q3). A p value <0.05 was considered significant. Statistical analyses were performed using SPSS 17.0 statistical software (IBM, Chicago, Ill, USA) and R (R Foundation for Statistical Computing, Vienna, Austria) with the “pROC” package [13].

**Results**

**General and US characteristics of the population**

Out of 1867 patients (92% females); median (Q1–Q3), age 70 (67–76) years we obtained 82.8% Benign (Bethesda II) and 2.6% malignant nodules. Three percent of the patients had previous exposure to radiotherapy, 4.2% family history of thyroid cancer, 14.6% personal history of diabetes, 31.5% were TPOab positive, and their TSH levels were 1.8 (0.9–3.3) mU/l.

| US characteristics | N (%) |
|--------------------|-------|
| Solid nodules      | 1400 (75%) |
| Echogenicity mildy hypoechoic | 702 (37.5%) |
| Echogenicity markedly hypoechoic | 68 (3.8%) |
| Presence of halo   | 542 (29%) |
| Microcalcifications| 252 (13.5%) |
| Macrocalcifications| 285 (15%) |
| Irregular margins  | 368 (20%) |
| Taller than wide   | 428 (23%) |
| Predominantly-central vascularization pattern | 286 (15%) |
| Bethesda System Classification |
| I                  | 134 (7.2%)  |
| II                 | 1546 (82.8%) |
| III                | 117 (6.2%)  |
| IV                 | 20 (1.1%)   |
| V                  | 24 (1.2%)   |
| VI                 | 26 (1.4%)   |

With respect to the US characteristics considered suspicious, 75% were solid nodules (n = 1400), 37.5% mildly hypoechoic (n = 702), 3.8% markedly hypoechoic (n = 68), 29% had a halo (n = 542), 20% had irregular margins (n = 368), and 23% had a taller than wide shape (n = 428). Microcalcifications were found in 13.5% (n = 252) and macrocalcifications in 15% (n = 285). The predominant pattern of central vascularization was found in 15% of the nodules (n = 286) (Table 1).

Out of 50 malignant nodules 31 had surgical confirmation at our institution. Among these tumors 16 (51%) were classical papillary thyroid cancer and 15 (49%) other thyroid cancer histotypes = 11 were follicular variants of papillary thyroid cancer, one follicular thyroid cancer, one medullary thyroid cancer, two lymphomas, and one anaplastic thyroid cancer.

**Diagnostic performance of each US-based risk-stratification system**

Malignant cytology within the categories of each US classification was compared. The three US classifications correctly identified malignant cytology (P < 0.01) (Table 2). According to the ACR TI-RADS, the proportion of malignant cytology in nodules classified under category TR 3 was 1.9%, under category TR 4, 3.1% and under category TR 5, 5.8%. Comparing between risk categories, those nodules classified in category TR 5 were at significantly higher risk of being malignant than TR 3 [Odds Ratio (OR) (95% CI) = 3.21 (1.37–7.54), p = 0.007]. Nodules classified as ATA low suspicion had a risk of malignancy of 1.3%, under
intermediate suspicion of 3.2% [OR (95% CI) = 2.60 (1.02–6.62), p = 0.045], and under high suspicion of 6.7% [OR (95% CI) = 5.61 (2.47–12.75), p < 0.001 vs. ATA low suspicion category]. According to AACE/ACE/AME, Class 2 nodules had risk of malignancy of 1.5% while the risk ascended to 4.7% in Class 3 [OR (95% CI) = 3.23 (1.63–6.40), p < 0.001].

Comparison of the ATA, ACR TI-RADS, and AACE/ACE/AME, in identifying malignant nodules

Since the ATA and AACE/ACE/AME classifications missed 14 and 2 cases of malignant cytology, respectively. After discarding benign unclassifiable nodules, we proceeded to compare the ROC curves of the three US systems using two approaches. First, we considered malignant unclassifiable nodules in the lowest risk category; and second without any imputation comparing only the nodules that could be classified by the three US systems. When malignant unclassifiable nodules were included (n = 1234, malignant = 50), the AUC of ATA was significantly lower than the two other US systems (Table 3). Excluding malignant unclassifiable nodules (n = 1218, malignant = 34), ROC curve analysis showed the opposite. ATA US classification system had a significantly higher AUC than the others (Table 3 and Supplementary Table 1). Most of the nodules with malignant cytology unclassifiable by ATA were solid but with iso or hyperechogenicity.

When Bethesda III nodules were classified as benign and Bethesda IV as malignant, the results were not different except for an increase in ATA sensitivity and AACE/ACE/AME specificity and an overall increase of positive predictive values (data not shown).

Analysis of FNABs below the recommended size cutoff and unclassifiable nodules

When we estimated the nodules that were biopsied below size cutoffs we found 16% for ATA, 42% for ACR TI-RADS, and 29% for AACE/ACE/AME (p < 0.001 for all paired comparisons). Considering the proportion of malignant nodules among these groups as FNR (%), we obtained similar proportions in the three US systems [FNR ATA, ACR TI-RADS, and AACE/ACE/AME: 1.4% (3/210), 2.6% (29/1218), and 3.1% (5/165), respectively, p = ns]. On the other hand, both ATA and AACE/ACE/AME left some nodules outside of the classification: 19% and 5.7% respectively.

Age tertiles and individual US characteristics

Patients were stratified into age tertiles. The age in the first tertile was 65 (63–67) years, in the second tertile was 71 (70–72) years, and in the third tertile was 78 (75–81) years. When analyzed within each age tertile, the proportion of solid echostructure and taller than wide shape was similar between benign and malignant nodules. Hypoechoogenicity instead, was significantly more frequent among malignant nodules both in the first (p = 0.024) and second age tertile (p = 0.009), but not in the third age tertile (p = 0.224). The proportion of irregular margins was significantly higher in malignant nodules along the age tertiles (p < 0.01), while microcalcifications were only significantly more frequent in malignant nodules in the third age tertile (p = 0.001). With regards to size, it was noted that only in the third age tertile malignant nodules were larger in size than benign ones (32.6 ± 19.6 mm vs. 21.9 ± 9.5 mm; p = 0.048).

### Table 2: Comparison of malignant cytology within the categories of each US classification

|                | n (%) | Malignant cytology n (%) | p for trend |
|----------------|-------|--------------------------|------------|
| **ACR TI-RADS** |       |                          |            |
| TR 1           | 39 (2.4%) | –                        |            |
| TR 2           | 105 (6.6%) | 2 (1.9%)                |            |
| TR 3           | 422 (26.6%) | 8 (1.9%)                |            |
| TR 4           | 730 (46.0%) | 23 (3.1%)               |            |
| TR 5           | 291 (18.3%) | 17 (5.8%)               | 0.028      |
| **ATA (n = 1285)** |   |                          |            |
| Benign         | 13 (1.0%) | –                        |            |
| Very low suspicion | 20 (1.6%) | 1 (5%)                  |            |
| Low suspicion  | 716 (55.7%) | 9 (1.3%)                |            |
| Intermediate suspicion | 281 (21.9%) | 9 (3.2%)                |            |
| High suspicion | 255 (19.8%) | 17 (6.7%)               | <0.001     |
| **AACE/ACE/AME (n = 1495)** |   |                          |            |
| Class 1        | 31 (2.1%) | 2 (6.5%)                |            |
| Class 2        | 725 (48.5%) | 11 (1.5%)               |            |
| Class 3        | 739 (49.4%) | 35 (4.7%)               | <0.001     |

### Table 3: Paired-comparison of ROC curves

|                | Z value | p value |
|----------------|---------|---------|
| Unclassifiable malignant nodules categorized in the lowest risk category (1234, M = 50) |         |         |
| ATA (0.49) vs. ACR TI-RADS (0.62) | 2.65 | 0.008   |
| ATA vs. AACE/ACE/AME (0.59) | 2.28 | 0.022   |
| AACE/ACE/AME vs. ACR TI-RADS | −0.58 | 0.56     |
| Excluding all unclassifiable nodules (n = 1218, M = 34) |         |         |
| ATA (0.68) vs. ACR TI-RADS (0.61) | −2.56 | 0.01    |
| ATA vs. AACE/ACE/AME (0.59) | −2.88 | <0.001  |
| AACE/ACE/AME vs. ACR TI-RADS | −0.73 | 0.46    |

AUC area under the curve
Discussion

The diagnostic performance of three (ATA, ACR TI-RADS, and AACE/ACE/AME) of the most widely used US classifications for malignancy detection, has been tested in an elderly cohort of patients for the first time. All three classifications were found to be useful for detection of malignant nodules. Nonetheless, 14 malignant nodules could not be classified by the ATA US system. This shortcoming made the ATA less convenient than the other two US systems. Only in head-to-head comparison using classifiable nodules, the ATA system was slightly superior to ACR TI-RADS and AACE/ACE/AME. Hence, the inability of the ATA classification to identify malignant nodules with iso or hypoechogenicity is of major relevance in aged patients.

Taking into consideration the proposed size cutoff each classification has for recommending FNAB, the ATA showed the lowest proportion of nodules referred to FNAB below the recommended size cutoff level. However, the three classifications found similar proportion of malignancy in nodules below the size cutoff level and in general, ACR TI-RADS spared more nodules from being biopsied.

The clinical management of thyroid nodular disease in the elderly represents a challenge. It is known that nodules are more frequent as we age and although most of them are of benign nature, those that are found malignant may pertain to an aggressive variant [3]. Despite the fact that comorbidities that increase the surgical risk in an older person may discourage the study of nodular disease, the identification of sonographic findings suggesting aggressive malignant disease can help to take a decision [14]. Moreover, US-based classifications proposed by different scientific entities are now used to help in malignancy risk stratification. As regards to the ATA and ACR TI-RADS classifications, both retrospective [15] and prospective [16] studies that have compared their diagnostic performances in general population have found similar elevated predictive value of malignancy in high-risk categories.

Most recently, Lauria Pantano et al. [17] also confirmed that the highest risk categories of ATA, AACE/ACE/AME, and ACR TI-RADS classifications correctly identified cytologically high-risk thyroid nodules. However, when compared, ACR TI-RADS and AACE/ACE/AME performed better than ATA possibly due to the large amount of nonclassifiable nodules in the ATA classification. In fact, it was described that nonclassifiable nodules harbored 7 times higher risk than the “very low suspicion” nodules. In the present study, we also found that the three US classifications were reliable to stratify malignancy risk although when only those nodules that could be classified by the three systems were compared, the ATA performed better. One possible explanation to the discrepancy between the two studies may include differences in both populations analyzed. In particular, our study comprised only elderly patients in whom a very low frequency of malignant nodules was detected, in line with previous reports [3], and in contrast to the suggested rate of malignancy proposed by the guidelines for each US classification [6–8]. However, when we decided to include nonclassifiable nodules in the total population for comparing among US classifications, AACE/ACE/AME and ACR TI-RADS resulted better predictors of malignancy, in agreement with the results of the mentioned study [17].

Furthermore, Lauria Pantano et al. [17] revealed that younger subjects should be considered at higher risk than older ones within the same US category. These findings would also help explain the low specificity of individual suspicious US features found in our study. In fact, it was found that hypoechogenicity, one of the main US characteristics, may lose its diagnostic value as age advances. Similarly, in older women we previously described that for mixed nodules, none of the suspicious US characteristics were associated with malignancy [18]. With regards to microcalcifications in particular, it has been reported that the associated malignant rate differs between younger and older individuals, with a higher yield in patients <45 years old compared with older patients [19, 20]. Since the literature in terms of US findings in the elderly is scarce, it could be argued that the higher US risk categories of any classification would be expected to perform less efficiently in older patients.

As regards to the efficacy of a US classification to avoid irrelevant biopsies, Grani et al. [21] compared the diagnostic yield of the ATA, AACE/ACE/AME, ACR TI-RADS, and two other TI-RADS classification systems established by Korean Society of Thyroid Radiology (KSThR) and European Thyroid Association (ETA) and found ACR TI-RADS to have the lowest rate of unnecessary FNAB. Similarly, Xu et al. [22] compared the three newly-updated TI-RADS classification systems by KSThR, ETA, and ACR and also found ACR TI-RADS to have the lowest rate of unneeded FNAB. In agreement, in this study it was also found that ACR TI-RADS, due to its higher size threshold to recommend FNAB, yielded a larger proportion of unnecessary biopsies than ATA and AACE/ACE/AME, a finding quite relevant when considering how to avoid invasive procedures in older patients. This potential advantage observed in ACR TI-RADS was also supported by the fact that malignancy rate was similar among nodules below the recommended size threshold in all three classifications.

In the present study the ATA classification performed better in diagnosis, however this advantage was offset by the large number of nodules that could not fit in any category such as those that were solid and iso or hyperechoic with at least one suspicious US finding. Moreover, the rate
of malignancy in this unclassified group almost reached the malignancy rate in the high-risk ATA category. A new approach is therefore needed to unify criteria and create a universal language to report on the US identification of each nodule that can facilitate the implementation of guidelines [23]. Age could ideally be part of this project and larger size cutoffs might be eventually considered according to each patient’s age.

One of the limitations of this study is the inclusion of only one nodule per patient thus creating a certain bias. However, since benign thyroid nodular disease is more frequent in the elderly, the inclusion of all nodules would have further reduced the low malignancy rate of this cohort.

Another questionable finding is the high number of false positive results in the high-risk categories. A plausible explanation could be the lack of a unified lexicon deemed necessary to avoid different interpretations of the US features observed during the biopsy. Nevertheless, we relied in the vast experience of three high volume operators to define US characteristics which would reduce this bias. An alternate explanation would be that those suspicious US findings reported in the general population are not quite accurate in the elderly. Older patients with long standing multinodular goiters may have nodules of different shapes and exhibit more calcifications that could jeopardize US stratification. In fact, hypoechogenicity was not as specific in the oldest old regarding the younger individuals of this cohort. Considering that only half of the patients with histological confirmation had classical papillary thyroid cancer it could be argued that in the elderly, suspicious US findings that typically are present in this classical variant may not be evident in other forms of thyroid cancer frequent with advanced age. Furthermore, size can be critical since in the subgroup of oldest patients, malignant nodules were larger than benign nodules. It also should be acknowledged as a limitation that all nodules classified as Bethesda V and VI were considered malignant, but not all were submitted to surgery allowing for eventual false positive cases.

Strengths of this study were its design with data collected prospectively and consecutively in a single academic center. The number of patients was larger than most studies in the literature and it only included elderly patients which makes it unique.

Conclusion

The present study is the first to validate in elderly patients three US classifications which showed that AACE/ACE/AME and ACR TI-RADS can predict thyroid malignancy more accurately than the ATA classification when all nodules are considered. Moreover, in this aged segment of the population, the use of ACR TI-RADS avoided more invasive procedures.

Also the fact that US suspicious characteristics of thyroid nodules in elderly patients did not result very specific for malignancy might be considered in future guidelines.

Acknowledgements

To the technicians and fellows of the Radiology Department who assisted in filling out the US characteristics forms. We thank Dr. Pol Castellano Escuder for his support on ROC curve comparison and other statistical analyses.

Compliance with ethical standards

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

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

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

References

1. C. Durante, G. Grani, L. Lamartina, S. Filetti, S.J. Mandel, D.S. Cooper, The diagnosis and management of thyroid nodules: a review. JAMA 319(9), 914–924 (2018). https://doi.org/10.1001/jama.2018.0898

2. C. Reiners, K. Wegscheider, H. Schicha, P. Theissen, R. Vaupel, R. Wrbitzky, P.M. Schumm-Draeger, Prevalence of thyroid disorders in the working population of Germany: ultrasonography screening in 96,278 unselected employees. Thyroid 14(11), 926–932 (2004). https://doi.org/10.1089/thy.2004.14.926

3. N. Kwong, M. Medici, T.E. Angell, X. Liu, E. Marqusee, E.S. Cibas, J.F. Krane, J.A. Barletta, M.I. Kim, P.R. Larsen, E.K. Alexander, The influence of patient age on thyroid nodule formation, multinodularity, and thyroid cancer risk. J. Clin. Endocrinol. Metab. 100(12), 4344–4440 (2015). https://doi.org/10.1210/jc.2015-3100

4. J.J. Diez., Goiter in adult patients aged 55 years and older: etiology and clinical features in 634 patients. J. Gerontol. A Biol. Sci. Med. Sci. 60(7), 920–923 (2005). https://doi.org/10.1093/gerona/60.7.920

5. S. Sorrenti, E. Baldini, F. Tartaglia, A. Catania, S. Arcieri, D. Pironi, P.G. Calo, A. Filippini, S. Ulisse, Nodular thyroid disease in the elderly: novel molecular approaches for the diagnosis of malignancy. Aging Clin. Exp. Res. 29(Suppl 1), 7–13 (2017). https://doi.org/10.1007/s40520-016-0654-y

6. B.R. Haugen, E.K. Alexander, K.C. Bible, G.M. Doherty, S.J. Mandel, Y.E. Nikiforov, F. Pacini, G.W. Randolph, A.M. Sawka, M. Schlumberger, K.G. Schuff, S.I. Sherman, J.A. Sosa, D.L. Steward, R.M. Tuttle, L. Wartofsky, 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). https://doi.org/10.1089/thy.2015.0020

7. H. Gharib, E. Papini, J.R. Garber, D.S. Duick, R.M. Harrell, L. Hegedus, R. Paschke, R. Valcavi, P. Vitti, American Association
