# Endocrine

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

## --Manuscript Draft--

| | |
|---|---|
| Manuscript Number: | ENDO-D-19-01110R3 |
| Full Title: | Comparison of different systems of (US) risk stratification for malignancy in elderly patients with thyroid nodules. Real world experience. |
| Article Type: | Original Article |
| Corresponding Author: | Gabriela Brenta<br>Dr. Cesar Milstein Care Unit<br>ARGENTINA |
| Corresponding Author Secondary Information: | |
| Corresponding Author's Institution: | Dr. Cesar Milstein Care Unit |
| Corresponding Author's Secondary Institution: | |
| First Author: | Fernando Di Fermo |
| First Author Secondary Information: | |
| Order of Authors: | Fernando Di Fermo |
| | Noelia Sforza |
| | Melanie Rosmarin |
| | Morosan Allo |
| | Carina Parisi |
| | Jimena Santamaria |
| | Nestor Pacenza |
| | Carlos Zuk |
| | Cristina Faingold |
| | Florencia Ferraro |
| | Tomas Meroño |
| | Gabriela Brenta |
| Order of Authors Secondary Information: | |
| Funding Information: | |
| 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.<br>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 BethesdaV/VI were considered malignant. Receiver operating characteristics (ROC) curves were established and compared to evaluate diagnostic performance. Malignancy among biopsies below the size cut-off for each ultrasound classification was also compared.<br>Results: 1867 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 3 classifications correctly identified malignancy (P<0.01). Nonetheless, in the ATA and |

| | |
|---|---|
| | AACE/ACE/AME 16 and 2 malignant nodules, respectively, were unclassifiable. Including unclassified 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 cut-off 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. <br> 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. |
| Response to Reviewers: | We would like to thank the Reviewer for all of his/her constructive comments. We have tried to modify the manuscript accordingly with hope that it is now suitable for publication in Endocrine. <br> We have included a reponse below each comment. <br> Best regards <br> Gabriela Brenta <br><br> COMMENTS FOR THE AUTHOR: <br><br> Reviewer #1: The authors addressed all my previous comments. I have no further remarks about the contents. <br> However, I would suggest double-checking the style of the manuscript, which is very hard to read and follow. <br><br> Abbreviations should be used more sparingly. According to the Instructions for the Authors, abbreviations should be defined at first mention and used consistently thereafter. <br> R. Thank you for these suggestions. We edited the manuscript following these instructions. <br><br> It is valid for the Abstract, too: in the current version, the authors use five abbreviations in the Abstract without providing any definition (ATA, ACR, AACE/ACE/AME, US, AUC, M). It is true for TIRADS, also. <br> Furthermore, there is no reason to use (BII) instead of Bethesda II. <br> Please bear in mind that readers of Endocrine may not be familiar with thyroidologists' jargon. <br> R. Thank you for this suggestion. We edited the abstract anf the rest of the manuscript following these instructions. <br><br> In the main manuscript, I suggest avoiding the use of M (instead of malignant) and B instead of Bethesda. Also, the use of S for Sensitivity and SP for specificity makes the text harder to read. <br> R. Thank you for this suggestion. We edited the manuscript following these instructions. <br><br> At line 201, PPV is not defined (even if apparently obvious). <br> On the other hand, the title of Supplementary Table 1 is redundant, including twice the words Sensitivity and Specificity (Sensitivity: Sensitivity (S), Specificity (SP):specificity) <br> R. Thank you for this suggestion. We have defined PPV as positive predictive value and also corrected the title of supplementary table 1. <br><br> The use of spacing, semicolons, and "=" should be consistent throughout the text. For example, in line 170 there are both "n=" and "n:". <br><br> R. Thank you for this suggestion. We have unified this concept all along the manuscript. <br><br> What is the meaning of - NCBI at line 139? The reference provided has no connection with the National Center for Biotechnology Information. <br> R. Thank you for this suggestion. We have removed NCBI that stands for National Center for Biotechnology Information. |

| | |
|---|---|
| | Figure 1 has broken lines and asymmetric boxes.<br>R. Thank you for this suggestion. We changed the figure following the Reviewer´s instructions .<br><br><br>I would recommend careful and meticulous editing of the whole text.<br>R. Thank you for this suggestion. We edited the manuscript following these instructions. |

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

Di Fermo, Fernando[1] ; Sforza, Noelia[1] ; Rosmarin, Melanie[1] ; Morosan Allo, Yanina[1] ; Parisi, Carina[1] ; Santamaria, Jimena[1] ; Pacenza, Nestor[1] ; Zuk, Carlos[2]; Faingold, Cristina[1] ; Ferraro, Florencia[3] ; Meroño, Tomas[3]; Brenta, Gabriela[1].

1. Endocrinology Department, Cesar Milstein Hospital, Buenos Aires CABA, Argentina.
2. Radiology Department, Cesar Milstein Hospital, CABA, Buenos Aires, Argentina.
3. Clinical Biochemistry, School of Pharmacy and Biochemistry. University of Buenos Aires, CABA, Argentina.

**Corresponding Author**       Gabriela Brenta, M.D.

Department of Endocrinology and Metabolism

Dr. Cesar Milstein Hospital

951 La Rioja Street

Buenos Aires, Argentina

Tel. 54-11-49591500

Email: gbrenta@gmail.com

Orcid number: 0000000155317652

**Key Words:** Elderly patients, ultrasound characteristics, thyroid nodules, malignant cytology.

1

1  **ABSTRACT**

2  **Purpose:** To comparatively assess the performance of three sonographic classification systems, American

3  Thyroid Association(ATA), the American College of Radiology Thyroid Imaging Reporting and Data

4  System(ACR TI-RADS), and American Association of Clinical Endocrinologists(AACE)/American College of

5  Endocrinology(ACE)/Associazione Medici Endocrinologi(AME) in identifying malignant nodules in an elderly

6  population.

7  **Methods:** Cross-sectional study of patients referred for fine needle aspiration biopsy in an academic center for

8  the elderly. One nodule/patient was considered. Nodules classified BethesdaV/VI were considered malignant.

9  Receiver operating characteristics (ROC) curves were established and compared to evaluate diagnostic

10  performance. Malignancy among biopsies below the size cut-off for each ultrasound classification was also

11  compared.

12  **Results:** 1867 patients (92%females); median(Q1-Q3) age 71(67-76) years, were studied showing 82.8%

13  benign (Bethesda II) and 2.6% malignant cytology. The 3 classifications correctly identified malignancy

14  (P<0.01). Nonetheless, in the ATA and AACE/ACE/AME 16 and 2 malignant nodules, respectively, were

15  unclassifiable. Including unclassified malignant nodules (n=1234, malignant=50), comparison of the ROC

16  curves showed lower performance of ATA [area under the curve (AUC)=ATA (0.49) vs ACR-TI-RADS (0.62),

17  $p$=0.008 and ATA vs AACE/ACE/AME (0.59), $p$=0.022]. Proportion of below size cut-off biopsies for ATA,

18  ACR-TI-RADS and AACE/ACE/AME was different [16%, 42% and 29% (all p<0.001)], but no differences in

19  malignancy rate were observed in these nodules.

20  **Conclusion:** The present study is the first to validate in elderly patients these classifications showing that

21  AACE/ACE/AME and ACR TI-RADS can predict thyroid malignancy more accurately than the ATA when

22  unclassifiable malignant nodules are considered. Moreover, in this aged segment of the population, the use of

23  ACR TI-RADS avoided more invasive procedures.

24

25

26

27

28

2

**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 to 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 medullar, 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, 3 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],.

3

54 Although the elderly population may benefit from the use of any of these 3 US classifications, there is no

55 available data comparing them. The aim of this study is to compare the efficacy of these 3 US classifications in

56 finding malignant cytology in an elderly cohort with thyroid nodules.

57

58 **METHODS**

59 **Study Cohort and Protocol**

60     This is a cross-sectional study of consecutive patients with thyroid nodules referred for US guided-

61 FNAB in an academic referral center for the elderly, Dr. Cesar Milstein Hospital, which receives all the patients

62 belonging to the National Institute of Social Services for Retirees and Pensioners (INSSJP) in the city of Buenos

63 Aires, an iodine sufficient metropolitan area. The US features to be analyzed were collected in a consecutive

64 way and their distribution into US guidelines categories was done retrospectively.

65     Since our hospital is a referral center for FNAB from other medical institutions, biopsy was performed

66 according to the indication of each referring physician. In case of multiple nodules, the presence of suspicious

67 US characteristics was used for the selection of the nodule to be biopsied. Clinical criteria to refer a patient to

68 US guided-FNAB were either a neck mass that was visible or palpable, or that had been found incidentally in

69 a previous imaging study. All US-guided FNABs were carried out by one of 3 operators of our institution, each

70 having more than 20 years experience with this procedure. In order to avoid a large inter-observer variability

71 as previously described for single suspicious features [9], only these 3 experienced clinicians were responsible

72 of describing all the individual US characteristics of the nodules. All this information was filled in by a

73 technician into a specific form online immediately before the FNAB. At the Endocrine Department, this form

74 was used to complete a database with all the information of each patient. Since 2018, the risk categories

75 according to each of the 3 US systems was calculated for each patient and registered in the database.

76     Only one nodule/patient was considered for this analysis. In the case of 2 or more coexisting nodules,

77 we selected for statistical analysis the one with a malignant cytological finding. If all nodules were benign, the

78 one with the highest US category risk was selected. The decision of one nodule/patient was based on the idea

79 that if several nodules were biopsied but only 1 was malignant; this patient would be referred to surgery based

80 on this specific nodule. Furthermore, in the case of multiple nodules, solid nodules with suspicious US findings

4

81 were the ones initially biopsied considering the other nodules as less relevant. Since this was the criteria chosen

82 by physicians performing the FNAB, the inclusion of the second nodule would render the malignant cytology

83 prevalence to a minimum. This common criteria shared by the sonographists is part of the study of only one-

84 nodule per patient. Last, there were also nodules that were followed in time and were subjected to more than

85 one FNAB during the study period. Given that the same nodule was analyzed at different time points, including

86 every FNAB in the study would have introduced bias. All nodules with Bethesda V/VI cytology were

87 considered malignant.

88 During a 10 year period, June 2008- June 2018, 1867 patients (92% females; aged, median (Q1-Q3),

89 71 (67-76) years) were consecutively included for the study. The total amount of biopsied nodules was 2400

90 but only one nodule per patient was considered. After exclusion of indeterminate (20% rate of malignancy at

91 our Institution) and insufficient cytology results (n=271), a subpopulation of 1596 nodules with benign and

92 malignant cytological results was obtained (Figure 1). Clinical and biochemical characteristics, such as age,

93 sex, previous exposure to radiotherapy, family history of thyroid cancer, personal history of diabetes, thyroid

94 peroxidase antibody (TPOab) positivity and TSH levels, as well as US details were prospectively collected..

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

96 consent form.

97 **Image Analysis**

98 Prior to each biopsy all US characteristics were assessed with real-time US in each thyroid nodule.

99 These included type of echostructure [solid, mixed (>25% cystic proportion), spongiform and purely cystic],

100 echogenic pattern [Hypoechoic= the nodule echogenicity was compared to normal thyroid (mild hypoechoic)

101 and strap muscles (markedly hypoechoic), isoechoic, hyperechoic, anechoic], margins (irregular or regular),

102 presence of halo (yes or no), microcalcifications (defined as tiny, punctate hyperechoic foci, without comet-

103 tail sign and distinct of indeterminate hyperechoic spots), macrocalcifications (defined as coarse areas of

104 calcification greater than 1 mm in size) and the 3 diameters of the nodule in mm. (taller than wide nodules were

105 defined when the anterior-posterior dimension exceeded the axial dimension).

106 The 3 US classifications: ATA [6], the AACE/ACE/AME [7] and the ACR TI-RADS [8] were applied

107 based on US findings. According to the suspicion of malignancy, the classification proposed by the ATA

5

108 divided nodules into 5 classes: 1 (benign), 2 (very low suspicion), 3 (low suspicion), 4 (intermediate suspicion)

109 and 5 (high suspicion). According to the score ACR TI-RADS, following a sum of points awarded according

110 to the ultrasound findings, nodules were divided into the following levels of suspicion of malignancy: TR 1

111 (benign), TR 2 (not suspect), TR 3 (very low suspicion), TR 4 (moderately suspect) and TR 5 (highly suspicious)

112 and according to the AACE/ACE/AME guide the risk of malignancy of the lesions was divided into 1 (low

113 risk), 2 (medium) and 3 (high).

114    FNABs are usually recommended by each US classification above a certain threshold of size. ATA

115 recommends diagnostic FNAB for thyroid nodules ≥1 cm of high suspicion and intermediate suspicion, low

116 suspicion ≥1.5 cm and very low suspicion ≥2 cm [6]. The AACE/ACE/AME proposes diagnostic FNAB for

117 thyroid nodules with Class 1 if ≥2.0 cm + increasing size or high-risk history, Class 2 if ≥2 cm and Class 3 if

118 ≥1.0 cm or ≥0.5 cm + subcapsular or paratracheal lesions, suspicious lymph nodes or extrathyroid spread,

119 positive personal or family history of thyroid cancer, history of head and neck irradiation, coexistent suspicious

120 clinical findings (e.g., dysphonia) [7]. The ACR TI-RADS system recommends diagnostic FNAB for thyroid

121 nodules with TR 3 ≥2.5 cm, TR 4 ≥1.5 cm, and TR 5 ≥1cm disregarding diagnostic FNAB for TR1 and TR2

122 nodules [8]. In order to investigate the value of these recommendations, all the nodules that were below the

123 recommended size cut-off of each classification were analyzed.

124 **US-guided fine needle aspiration procedure**

125    A Mindray DC-3 (Shenzhen, China) Doppler-echo machine and a 7.5-10 MHz linear-array probe were

126 used to guide all FNABs in real time. Biopsies were performed using a 23-gauge needle, and visualization of

127 the tip of the needle inside the nodule helped to monitor the correct site for biopsy. At least 2-6 needle passages

128 were performed in each nodule. Material obtained from FNABs was smeared on glass slides, which were

129 immediately placed in 95% alcohol for Papanicolau stain and sent to the Pathology Department.

130 **Cytological analysis**

131    This study used the Bethesda System for Reporting Thyroid Cytopathology to describe the cytological

132 results [10]. The results of those nodules included in 2008 were adapted to the Bethesda System. Cytological

133 analysis was performed independently by two pathologists. Validation of this procedure by cyto-histological

134    correlation in our Institution was previously reported  [11]. Those patients that were referred to surgery at our

135    Institution were also considered for a descriptive analysis (n=31).

136    **Statistical analysis**

137            Chi2 and logistic regression were used to evaluate and compare malignant cytology within each US

138    classification. Receiver operating characteristics (ROC) curves were established to compare diagnostic

139    performance. The cut-off with the highest Youden Index was used to calculate the sensitivity and specificity.

140    Since some malignant nodules remained unclassifiable according to the ATA and AACE/ACE/AME US

141    classification systems, we performed two ROC curve comparison analyses. One after exclusion of

142    unclassifiable benign nodules and categorizing unclassifiable malignant nodules as the lowest risk category

143    (n=1234, 16 malignant nodules reclassified for ATA and AACE/ACE/AME combined) and the other excluding

144    all unclassifiable nodules whether benign or not (n=1218, malignant nodules=34). For paired-comparisons

145    between the area under the curve (AUC) of the US systems we used DeLong method [12].

146            The nodules that were biopsied even when their size was lower than the guidelines recommendations

147    were classified as "below size cut-offs" and pairs were compared using the McNemar test across the different

148    US systems. The false negative (malignant nodules in the "below size cut-offs" biopsies) rate (FNR) was also

149    calculated for each US classification system.

150            Normally distributed variables are presented as mean ± S.D. and skewed variables as median (Q1-Q3).

151    A *p* value <0.05 was considered significant. Statistical analyses were performed using SPSS 17.0 statistical

152    software (IBM, Chicago, Ill, USA) and R (R Foundation for Statistical Computing, Vienna, Austria) with the

153    'pROC' package [13].

154

155    **RESULTS**

156    **General and US characteristics of the population**

157            Out of 1867 patients (92% females); median (Q1-Q3), age 70 (67-76) years we obtained 82.8% Benign

158    (Bethesda II) and 2.6% malignant nodules. Three percent of the patients had previous exposure to radiotherapy,

7

159 4.2% family history of thyroid cancer, 14.6% personal history of diabetes, 31.5% were TPOab positive and

160 their TSH levels were 1.8 (0.9-3.3) mU/l.

161      With respect to the US characteristics considered suspicious, 75% were solid nodules (n=1400), 37.5%

162 mildly hypoechoic (n=702), 3.8% markedly hypoechoic (n=68), 29% had a halo (n=542), 20% had irregular

163 margins (n=368) and 23% had a taller than wide shape (n=428). Microcalcifications were found in 13.5%

164 (n=252) and macrocalcifications in 15% (n=285). The predominant pattern of central vascularization was found

165 in 15% of the nodules (n=286) (Table1).

166      Out of 50 malignant nodules 31 had surgical confirmation at our institution. Among these tumours 16

167 (51%) were classical papillary thyroid cancer and 15 (49%) other thyroid cancer histotypes= 11 were follicular

168 variants of papillary thyroid cancer, 1 follicular thyroid cancer, 1 medullary thyroid cancer, 2 lymphomas and

169 1 anaplastic thyroid cancer.

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

171      Malignant cytology within the categories of each US classification was compared. The 3 US

172 classifications correctly identified malignant cytology (P<0.01) (Table 2). According to the ACR TI-RADS,

173 the proportion of malignant cytology in nodules classified under category TR 3 was 1.9 %, under category TR

174 4, 3.1 % and under category TR 5, 5.8 %. Comparing between risk categories, those nodules classified in

175 category TR 5 were at significantly higher risk of being malignant than TR 3 [Odds Ratio (OR) (95%CI)= 3.21

176 (1.37-7.54), p= 0.007]. Nodules classified as ATA low suspicion had a risk of malignancy of 1.3 %, under

177 intermediate suspicion of 3.2 % [OR (95%CI)= 2.60 (1.02-6.62), p= 0.045] and under high suspicion of 6.7 %

178 [OR (95%CI)= 5.61 (2.47-12.75), p < 0.001 *vs.* ATA low suspicion category]. According to AACE/ACE/AME,

179 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

180 (1.63-6.40), p < 0.001].

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

182      Since the ATA and AACE/ACE/AME classifications missed 14 and 2 cases of malignant cytology,

183 respectively. After discarding benign unclassifiable nodules, we proceeded to compare the ROC curves of the

184 3 US systems using two approaches. First, we considered malignant unclassifiable nodules in the lowest risk

8

185    category; and second without any imputation comparing only the nodules that could be classified by the 3 US

186    systems. When malignant unclassifiable nodules were included (n=1234, malignant=50), the AUC of ATA was

187    significantly lower than the two other US systems (Table 3). Excluding malignant unclassifiable nodules

188    (n=1218, malignant=34), ROC curve analysis showed the opposite. ATA US classification system had a

189    significantly higher AUC than the others (Table 3 and Supplementary Table 1). Most of the nodules with

190    malignant cytology unclassifiable by ATA were solid but with iso or hyperechogenicity.

191    When Bethesda III nodules were classified as benign and Bethesda IV as malignant, the results were

192    not different except for an increase in ATA sensitivity and AACE/ACE/AME specificity and an overall increase

193    of positive predictive values (data not shown).

**194    Analysis of FNABs below the recommended size cut-off and unclassifiable nodules**

195    When we estimated the nodules that were biopsied below size cut-offs we found 16% for ATA, 42%

196    for ACR TI-RADS and 29% for AACE/ACE/AME (p<0.001 for all paired comparisons). Considering the

197    proportion of malignant nodules among these groups as FNR (%), we obtained similar proportions the 3 US

198    systems [FNR ATA, ACR TI-RADS and AACE/ACE/AME: 1.4 % (3/210), 2.2% (15/674) and 1.8% (8/431),

199    respectively, *p*=ns]. On the other hand, both ATA and AACE/ACE/AME left some nodules outside of the

200    classification: 19% and 5.7% respectively.

**201    Age tertiles and individual US characteristics**

202    Patients were stratified into age tertiles.  The age in the first tertile was 65 (63-67) years, in the second

203    tertile was 71 (70-72) years and in the third tertile was 78 (75-81) years. When analyzed within each age tertile,

204    the proportion of solid echostructure and taller than wide shape was similar between benign and malignant

205    nodules. Hypoechogenicity instead, was significantly more frequent among malignant nodules both in the first

206    (*p*=0.024) and second age tertile (*p*= 0.009), but not in the third age tertile (*p*=0.224). The proportion of irregular

207    margins was significantly higher in malignant nodules along the age tertiles (*p*<0.01) while microcalcifications

208    were only significantly more frequent in malignant nodules in the third age tertile (*p*=0.001). With regards to

209    size, it was noted that only in the third age tertile malignant nodules were larger in size than benign ones

210    (32.6±19.6 mm *vs*. 21.9±9.5 mm; *p*=0.048).

211

9

**DISCUSSION**

The diagnostic performance of 3 (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 3 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 hyperechogenicity is of major relevance in aged patients.

Taking into consideration the proposed size cut-off each classification has for recommending FNAB, the ATA showed the lowest proportion of nodules referred to FNAB below the recommended size cut-off level. However, the 3 classifications found similar proportion of malignancy in nodules below the size cut-off level and in general, ACR TI-RADS spared more nodules from being biopsed.

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 co-morbidities 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 non classifiable nodules in the ATA classification. In fact, it was described that non classifiable nodules harbored 7 times higher risk than the "very low suspicion" nodules. In the present study, we also found that the 3 US classifications were reliable to stratify malignancy risk although when only those nodules that could be classified by the 3 systems were compared, the ATA performed better. One possible

10

240 explanation to the discrepancy between the 2 studies may include differences in both populations analyzed. In

241 particular, our study comprised only elderly patients in whom a very low frequency of malignant nodules was

242 detected, in line with previous reports [3] and in contrast to the suggested rate of malignancy proposed by the

243 guidelines for each US classification [6-8]. However, when we decided to include non classifiable nodules in

244 the total population for comparing among US classifications, AACE/ACE/AME and ACR TI-RADS resulted

245 better predictors of malignancy, in agreement with the results of the mentioned study [17].

246 Furthermore, it was revealed that younger subjects should be considered at higher risk than older ones

247 within the same US category [17]. These findings would also help explain the low specificity of individual

248 suspicious US features found in our study.  In fact, it was found that hypoechogenicity, one of the main US

249 characteristics, may lose its diagnostic value as age advances. Similarly, in older women we previously

250 described that for mixed nodules, none of the suspicious US characteristics were associated with malignancy

251 [18]. With regards to microcalcifications in particular, it has been reported that the associated malignant rate

252 differs between younger and older individuals, with a higher yield in patients < 45-years old compared with

253 older patients [19,20]. Since the literature in terms of US findings in the elderly is scarce, it could be argued

254 that the higher US risk categories of any classification would be expected to perform less efficiently in older

255 patients.

256 As regards to the efficacy of a US classification to avoid irrelevant biopsies, Grani et al. [21] compared

257 the diagnostic yield of the ATA, AACE/ACE/AME, ACR TI-RADS and 2 other TI-RADS classification

258 systems established by Korean Society of Thyroid Radiology (KSThR) and European Thyroid Association

259 (ETA) and found ACR TI-RADS to have the lowest rate of unnecessary FNAB.  Similarly, Xu et al. [22]

260 compared the three newly-updated TI-RADS classification systems by KSThR, ETA and ACR and also found

261 ACR TI-RADS to have the lowest rate of unneeded FNAB. In agreement, in this study it was  also found that

262 ACR TI-RADS, due to its higher size threshold to recommend FNAB, yielded a larger proportion of

263 unnecessary biopsies than ATA and AACE/ACE/AME, a finding quite relevant when considering how to avoid

264 invasive procedures in older patients. This potential advantage observed in ACR TI-RADS was also supported

265 by the fact that malignancy rate was similar among nodules below the recommended size threshold in all 3

266 classifications.

11

267        In the present study the ATA classification performed better in diagnosis, however this advantage was

268        offset by the large number of nodules that could not fit in any category such as those that were solid and iso or

269        hyperechoic with at least one suspicious US finding. Moreover, the rate of malignancy in this unclassified group

270        almost reached the malignancy rate in the high risk ATA category. A new approach is therefore needed to unify

271        criteria and create a universal language to report on the US identification of each nodule that can facilitate the

272        implementation of guidelines [23]. Age could ideally be part of this project and larger size cut-offs might be

273        eventually considered according to each patient´s age.

274        One of the limitations of this study is the inclusion of only one nodule per patient thus creating a certain

275        bias. However, since benign thyroid nodular disease is more frequent in the elderly, the inclusion of all nodules

276        would have further reduced the low malignancy rate of this cohort.

277        Another questionable finding is the high number of false positive results in the high risk categories. A

278        plausible explanation could be the lack of a unified lexicon deemed necessary to avoid different interpretations

279        of the US features observed during the biopsy. Nevertheless, we relied in the vast experience of 3 high volume

280        operators to define US characteristics which would reduce this bias. An alternate explanation would be that

281        those suspicious US findings reported in the general population are not quite accurate in the elderly. Older

282        patients with long standing multinodular goiters may have nodules of different shapes and exhibit more

283        calcifications that could jeopardize US stratification. In fact, hypoechogenicity was not as specific in the oldest

284        old regarding the younger individuals of this cohort. Considering that only half of the patients with histological

285        confirmation had classical papillary thyroid cancer it could be argued that in the elderly, suspicious US findings

286        that typically are present in this classical variant may not be evident in other forms of thyroid cancer frequent

287        with advanced age. Furthermore, size can be critical since in the subgroup of oldest patients, malignant nodules

288        were larger than benign nodules. It also should be acknowledged as a limitation that all nodules classified as

289        Bethesda V and VI were considered malignant, but not all were submitted to surgery allowing for eventual false

290        positive cases.

291        Strengths of this study were its design with data collected prospectively and consecutively in a single

292        academic center. The number of patients was larger than most studies in the literature and it only included

293        elderly patients which makes it unique.

294

12

## CONCLUSION

295 CONCLUSION

296      The present study is the first to validate in elderly patients 3 US classifications which showed that

297 AACE/ACE/AME and ACR TI-RADS can predict thyroid malignancy more accurately than the ATA

298 classification when all nodules are considered. Moreover, in this aged segment of the population, the use of

299 ACR TI-RADS avoided more invasive procedures.

300      Also the fact that US suspicious characteristics of thyroid nodules in elderly patients did not result very

301 specific for malignancy might be considered in future guidelines.

302

303 COMPLIANCE WITH ETHICAL STANDARDS:

304 Funding: This study did not receive funding.

305 Conflict of Interest: None of the Authors has any conflict of interest.

306 Ethical approval: All procedures performed in studies involving human participants were in accordance with

307 the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki

308 declaration and its later amendments or comparable ethical standards.

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

310

311

312

313

314

315

316

317

13

**REFERENCES**

1. Durante, C., Grani, G., Lamartina, L., Filetti, S., Mandel, S.J., Cooper, D.S.: The Diagnosis and Management of Thyroid Nodules: A Review. JAMA **319**(9), 914-924 (2018). doi:10.1001/jama.2018.0898

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

3. Kwong, N., Medici, M., Angell, T.E., Liu, X., Marqusee, E., Cibas, E.S., Krane, J.F., Barletta, J.A., Kim, M.I., Larsen, P.R., Alexander, E.K.: The Influence of Patient Age on Thyroid Nodule Formation, Multinodularity, and Thyroid Cancer Risk. J Clin Endocrinol Metab **100**(12), 4434-4440 (2015). doi:10.1210/jc.2015-3100

4. Diez, J.J.: 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). doi:60/7/920

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

6. Haugen, B.R., Alexander, E.K., Bible, K.C., Doherty, G.M., Mandel, S.J., Nikiforov, Y.E., Pacini, F., Randolph, G.W., Sawka, A.M., Schlumberger, M., Schuff, K.G., Sherman, S.I., Sosa, J.A., Steward, D.L., Tuttle, R.M., Wartofsky, L.: 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid **26**(1), 1-133 (2016). doi:10.1089/thy.2015.0020

7. Gharib, H., Papini, E., Garber, J.R., Duick, D.S., Harrell, R.M., Hegedus, L., Paschke, R., Valcavi, R., Vitti, P.: American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi Medical Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules--2016 Update. Endocr Pract **22**(5), 622-639 (2016). doi:10.4158/EP161208.GL

8. Tessler, F.N., Middleton, W.D., Grant, E.G., Hoang, J.K., Berland, L.L., Teefey, S.A., Cronan, J.J., Beland, M.D., Desser, T.S., Frates, M.C., Hammers, L.W., Hamper, U.M., Langer, J.E., Reading, C.C., Scoutt, L.M., Stavros, A.T.: ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. J Am Coll Radiol **14**(5), 587-595 (2017). doi:S1546-1440(17)30186-2

9. Grani, G., Lamartina, L., Cantisani, V., Maranghi, M., Lucia, P., Durante, C.: Interobserver agreement of various thyroid imaging reporting and data systems. Endocr Connect **7**(1), 1-7 (2018). doi:10.1530/EC-17-0336

10. Cibas, E.S., Ali, S.Z.: The Bethesda System for Reporting Thyroid Cytopathology. Thyroid **19**(11), 1159-1165 (2009). doi:10.1089/thy.2009.0274

11. Brenta, G., Schnitman, M., Bonnahon, L., Besuschio, S., Zuk, C., De Barrio, G., Peruzzotti, C., Saubidet, G.: Evaluation of innovative skin-marking technique performed before thyroid ultrasound-guided fine-needle aspiration biopsies. Endocr Pract **8**(1), 5-9 (2002). doi:EP01001.OR

14

359   12. DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L.: Comparing the areas under two or more
360        correlated receiver operating characteristic curves: a nonparametric approach. Biometrics
361        **44**(3), 837-845 (1988).

362   13. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., Muller, M.: pROC: an open-
363        source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics **12**,
364        77 (2011). doi:10.1186/1471-2105-12-77

365   14. Wang, Z., Vyas, C.M., Van Benschoten, O., Nehs, M.A., Moore, F.D., Jr., Marqusee, E., Krane, J.F.,
366        Kim, M.I., Heller, H.T., Gawande, A.A., Frates, M.C., Doubilet, P.M., Doherty, G.M., Cho, N.L.,
367        Cibas, E.S., Benson, C.B., Barletta, J.A., Zavacki, A.M., Larsen, P.R., Alexander, E.K., Angell,
368        T.E.: Quantitative Analysis of the Benefits and Risk of Thyroid Nodule Evaluation in Patients
369        >/=70 Years Old. Thyroid **28**(4), 465-471 (2018). doi:10.1089/thy.2017.0655

370   15. Yoon, J.H., Lee, H.S., Kim, E.K., Moon, H.J., Kwak, J.Y.: Malignancy Risk Stratification of Thyroid
371        Nodules: Comparison between the Thyroid Imaging Reporting and Data System and the
372        2014 American Thyroid Association Management Guidelines. Radiology **278**(3), 917-924
373        (2016). doi:10.1148/radiol.2015150056

374   16. Persichetti, A., Di Stasio, E., Guglielmi, R., Bizzarri, G., Taccogna, S., Misischi, I., Graziano, F.,
375        Petrucci, L., Bianchini, A., Papini, E.: Predictive Value of Malignancy of Thyroid Nodule
376        Ultrasound Classification Systems: A Prospective Study. J Clin Endocrinol Metab **103**(4),
377        1359-1368 (2018). doi:10.1210/jc.2017-01708

378   17. Lauria Pantano, A., Maddaloni, E., Briganti, S.I., Beretta Anguissola, G., Perrella, E., Taffon, C.,
379        Palermo, A., Pozzilli, P., Manfrini, S., Crescenzi, A.: Differences between ATA,
380        AACE/ACE/AME and ACR TI-RADS ultrasound classifications performance in identifying
381        cytological high-risk thyroid nodules. Eur J Endocrinol **178**(6), 595-603 (2018).
382        doi:10.1530/EJE-18-0083

383   18. Vera, M.I., Merono, T., Urrutia, M.A., Parisi, C., Morosan, Y., Rosmarin, M., Schnitman, M., Brites,
384        F., Grisendi, S., Serrano, M.S., Luciani, W., Serrano, L., Zuk, C., De Barrio, G., Cejas, C.,
385        Faingold, M.C., Brenta, G.: Differential profile of ultrasound findings associated with
386        malignancy in mixed and solid thyroid nodules in an elderly female population. J Thyroid
387        Res **2014**, 761653 (2014). doi:10.1155/2014/761653

388   19. Shi, C., Li, S., Shi, T., Liu, B., Ding, C., Qin, H.: Correlation between thyroid nodule calcification
389        morphology on ultrasound and thyroid carcinoma. J Int Med Res **40**(1), 350-357 (2012).
390        doi:10.1177/147323001204000136

391   20. Wang, Z., Zhang, H., Zhang, P., He, L., Dong, W.: Diagnostic value of ultrasound-detected
392        calcification in thyroid nodules. Ann Acad Med Singapore **43**(2), 102-106 (2014).

393   21. Grani, G., Lamartina, L., Ascoli, V., Bosco, D., Biffoni, M., Giacomelli, L., Maranghi, M., Falcone,
394        R., Ramundo, V., Cantisani, V., Filetti, S., Durante, C.: Reducing the Number of Unnecessary
395        Thyroid Biopsies While Improving Diagnostic Accuracy: Toward the "Right" TIRADS. J Clin
396        Endocrinol Metab **104**(1), 95-102 (2019). doi:10.1210/jc.2018-01674

397   22. Xu, T., Wu, Y., Wu, R.X., Zhang, Y.Z., Gu, J.Y., Ye, X.H., Tang, W., Xu, S.H., Liu, C., Wu, X.H.:
398        Validation and comparison of three newly-released Thyroid Imaging Reporting and Data
399        Systems for cancer risk determination. Endocrine **64**(2), 299-307 (2019).
400        doi:10.1007/s12020-018-1817-8

401   23. Tappouni, R.R., Itri, J.N., McQueen, T.S., Lalwani, N., Ou, J.J.: ACR TI-RADS: Pitfalls, Solutions, and
402        Future Directions. Radiographics **39**(7), 2040-2052 (2019). doi:10.1148/rg.2019190026

15

**FIGURE LEGENDS**

Figure 1: Selection of the studied population

16

**Table 1: US characteristics and cytological results of the total group of nodules (n=1867)**

| US characteristics | N (%) |
|---|---|
| Solid nodules | 1400 (75%) |
| Echogenicity (mildly 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% ) |

17

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

| ACR TI-RADS<br><br>(n= 1587) | n (%) | Malignant cytology<br>n (%) | *p* for trend |
|---|---|---|---|
| TR1 | 39 (2.4%) | - | |
| TR2 | 105 (6.6%) | 2 (1.9%) | |
| TR3 | 422 (26.6%) | 8 (1.9%) | |
| TR4 | 730 (46.0%) | 23 (3.1%) | |
| TR5 | 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 |

18

**Table 3. Paired-comparison of ROC curves**

| (AUC) | 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.


**Supplementary Table 1. Cut-off values in each US system as predictor of malignancy.**

| Score | Cut-off level | AUC | S | SP | PPV | NPV | OR (95% CI) | LR |
|---|---|---|---|---|---|---|---|---|
| ACR TI-RADS (n=1587, malignant=50) | TR5 | 0.61 (0.52-0.71) | 34% | 82% | 6% | 97% | 2.36 (1.3-4.3) | 1.90 |
| ATA (n=1285, malignant=36) | High suspicion | 0.68 (0.58-0.77) | 47% | 80% | 6.6 % | 98% | 3.8 (1.9-7.4) | 2.47 |
| AACE/ACE/AME (n=1495, malignant=48) | Class 3 | 0.59 (0.50-0.68) | 73% | 51% | 4.7% | 98% | 2.84 (1.4-5.4) | 1.49 |

**AUC= Area under the curve. S= Sensitivity. SP=Specificity. PPV= positive predictive value.**

**NPV= Negative predictive value. OR= odds ratio. LR= likelihood ratio.**

19

**Figure 1: Selection of the studied population**



| | | |
|---|---|---|
| **2400 nodules** | Only 1 nodule/patient was considered | Bethesda I/III and IV nodules were not considered for the statistical analysis (14.5%, n=271) |
| | 1867 nodules | Malignant group (Bethesda V/VI) (2.6%, n=50) |
| | | Benign group (Bethesda II) (82.8%, n=1546) |