

ORIGINAL ARTICLE

Thyroid cytology: The reality before and after the introduction of ultrasound classification systems for thyroid nodules

Sara Campos Lopes^{a,*}, Bijal Shah^b, Catarina Eloy^{c,d}^a Endocrinology Department, Hospital de Braga, Braga, Portugal^b Histopathology Department, St. James's Hospital, Dublin, Ireland^c Pathology Laboratory, Institute of Molecular Pathology and Immunology of University of Porto, Porto, Portugal^d Pathology Department of Medical Faculty, University of Porto, Porto, Portugal

Received 12 April 2022; accepted 30 June 2022

Available online 18 October 2022

KEYWORDS

Thyroid nodule;
Fine-needle;
Biopsy;
Ultrasonography

Abstract

Background: Several ultrasound-based systems for classification of thyroid nodules are available. They allow for a better triage of the nodules that require cytological assessment, and lead to standardized recommendations. Our aim was to compare patients and nodules referred to fine-needle aspiration (FNA) before and after the introduction of these systems.

Methods: A retrospective study comparing two cohorts of patients referred for FNA was performed (386 patients and 463 nodules in 2015; 220 patients and 263 nodules in 2021).

Results: The sex distribution (89.1% vs 85.9% females, $p=0.243$), number of nodules referred to FNA per patient (median of 1), and the distribution of the Bethesda categories ($p=0.082$) was similar in both years. In 2021, patients were older (53.4 ± 14.5 years vs 57.8 ± 13.2 years, $p < 0.001$) and nodules over one centimetre were larger (median 17.0 mm vs 19.0 mm, $p=0.002$), especially the ones categorized as Bethesda III (median size 11 mm vs 23 mm, $p=0.043$). In 2021, at least 23.1% of the nodules referred to FNA did not have any criteria, and 38.8% of the nodules were not categorized by any system.

Conclusion: This analysis draws attention to the importance of systematically applying ultrasound-based classification systems. It seems that, by not being focused mainly on size thresholds, they allow for longer surveillance periods, without aggravating the cytology results when FNA becomes indicated. Nevertheless, greater efforts are needed to ensure more standardized reports, and to increase adherence to the resulting recommendations to reduce clinical uncertainty, unnecessary FNA, and overtreatment.

© 2022 SEEN y SED. Published by Elsevier España, S.L.U. All rights reserved.

* Corresponding author.

E-mail address: saragclopes@gmail.com (S.C. Lopes).

PALABRAS CLAVE

Nódulo tiroideo;
Aguja fina;
Biopsia;
Ultrasonografía

Citología tiroidea: la realidad antes y después de la introducción de los sistemas de clasificación de los nódulos tiroideos basados en la ecografía

Resumen

Antecedentes: Existen varios sistemas basados en la ecografía para la clasificación de los nódulos tiroideos. Permiten un mejor triaje de los nódulos que requieren una evaluación citológica y conducen a recomendaciones estandarizadas. Nuestro objetivo fue comparar los pacientes y los nódulos remitidos para punción aspiración con aguja fina (PAAF) antes y después de la introducción de estos sistemas.

Métodos: Se realizó un estudio retrospectivo comparando 2 cohortes de pacientes remitidos para PAAF (386 pacientes y 463 nódulos en 2015; 220 pacientes y 263 nódulos en 2021).

Resultados: La distribución por género (89,1% vs. 85,9% mujeres, $p=0,243$), el número de nódulos remitidos para PAAF por paciente (mediana: 1) y la distribución de las categorías Bethesda ($p=0,082$) fue similar en ambos años. En 2021 los pacientes eran de mayor edad ($53,4 \pm 14,5$ años vs. $57,8 \pm 13,2$ años, $p < 0,001$) y los nódulos supracentimétricos eran mayores (mediana 17,0 mm vs. 19,0 mm, $p=0,002$), especialmente los Bethesda III (mediana 11 mm vs. 23 mm, $p=0,043$). En 2021 al menos el 23,1% de los nódulos remitidos a PAAF no tenían criterios, y el 38,8% de los nódulos no fueron categorizados por ningún sistema.

Conclusión: Este análisis llama la atención sobre la importancia de aplicar este sistemas de clasificación. Parece que, al no estar centrados principalmente en el tamaño, permiten períodos de vigilancia más prolongados, sin agravar la citología cuando se indica la PAAF. No obstante, es necesario realizar mayores esfuerzos para garantizar informes más estandarizados y aumentar la adherencia a las recomendaciones resultantes, con el fin de reducir la incertidumbre clínica y las PAAF innecesarias.

© 2022 SEEN y SED. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Introduction

Thyroid nodules are common, with a prevalence as high as has 68% in some ultrasonography series.^{1,2} The incidence of nodular thyroid disease and thyroid cancer grew considerably in the last two decades, mainly due to the widespread use of imaging techniques leading to the detection of thyroid incidentalomas.³ Nevertheless, only a minority of these nodules are malignant (7–15%).⁴

Thyroid ultrasonography (US) is the gold standard for identifying and evaluating thyroid nodules via imaging.⁴ US helps to differentiate the nodules that need to be investigated with fine-needle aspiration (FNA) biopsy from the ones that only require surveillance. In order to improve diagnostic accuracy, decrease unwarranted FNA biopsies, and prevent overdiagnosis and overtreatment, several risk stratification systems based on the US features of thyroid nodules have been developed worldwide in the last few years.^{4–9} In general, these stratification systems focus on similar ultrasonographic features suggestive of malignancy, namely, hypoechogenicity, ill-defined margins, solid appearance, and the presence of calcifications.¹⁰

These updated and commonly used US classification systems include the system proposed in the American Thyroid Association (ATA) guidelines from 2015 (published in January 2016),⁴ the Thyroid Imaging Reporting and Data System (TIRADS) developed by the European Thyroid Association in 2017 (EU-TIRADS),⁶ and the TIRADS proposed by the American College of Radiology in 2017 (ACR-TIRADS).⁸ The impact of introducing these stratification US systems on the

thyroid nodules referred for FNA and their cytological results is unknown.

The main aim of our study was to compare the patients and thyroid nodules referred to FNA and their cytological result before and after the introduction of these well-established US classification systems.

Materials and methods**Study design and patient selection**

A retrospective study was conducted at the Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal. The study included all the patients that underwent US-guided thyroid FNA in this institution from 1st October to 10th November 2015 and from 1st October to 10th November 2021 by the same experienced pathologist (CE). The same pathologist performed the cytology report for all the nodules. The system used for the classification of the cytopathological findings was the Bethesda system.¹¹

IPATIMUP is a reference centre for thyroid FNA with an active contract with the Administração Regional de Saúde do Norte (ARS Norte), where primary care providers can send patients for thyroid FNA biopsy and cytology. It has a pathology laboratory that is double accredited by the College of American Pathologists and by NP EN ISO 15189. The remaining patients are referred to IPATIMUP mainly from private practice, and occasionally from public hospitals after specific situations (consecutive Bethesda category

I results and/or technically difficult to perform FNA biopsy). To undergo FNA biopsy, the patient needs to have an US report of the thyroid performed within the previous six months.

We collected the most recent data from 2021 and the same corresponding period in 2015 to avoid having confounding factors, such as a bias in referral due to an eventual seasonal variation in seeking medical care that could invalidate possible explanations for differences that could emerge from the statistical analyses between the two samples. The year selected to perform the comparison with 2021 was 2015 since it was the last year before the new ATA guidelines⁴ were proposed (published in January 2016) and subsequently the EU-TIRADS⁶ and ACR-TIRADS⁸ scoring systems.

The data for characterizing, and then comparing, the patients and nodules from 2015 and 2021 were obtained from the available clinical information provided both by the US report and by the clinician that referred the patient to the IPATIMUP. The variables collected were: sex, age, number of nodules for which FNA was requested per patient, the nodules' size, the US classification system used and the number of nodules with US criteria to perform FNA. The latter variable was decided based on the guidelines that were in force at the time.

In 2015, we considered that the nodule had indication for FNA if it met the criteria presented in the 2009 ATA guidelines, as they were the ones most widely accepted at the time¹² and as other US classification systems with scoring systems, such as TIRADS (Horvath et al., 2009),⁵ were not commonly used in Portugal. To do this assessment, we only had access to the data provided by the clinician as, at the time, the US report was not added to the patients' file as it is now in 2021. The most common information given was the nodule size and, sporadically, the echogenicity and the nodule composition (solid, cyst, mixed). There was no reference in any of the patients from 2015 of the use of any US classification system, while in the information given by the clinician in 2021, the EU-TIRADS/ACR-TIRADS/ATA classification was frequently reported. Nevertheless, it is relevant to take into consideration that the clinical information given by the clinicians was scarce both in 2015 and in 2021. However, in 2021, we had access to more data, as we had an US report available for analysis for each patient from the previous six months. This was due to the COVID-19 pandemic. Since the first lockdown in Portugal in 2020, in IPATIMUP, the US reports started being scanned and attached to the clinical file of the patients to prevent additional physical contact. The nodule size and the US classification described in the report were collected and used to assess if the nodule had criteria for FNA. The US reports that did not have any reference to an US classification system were labelled as such and considered "missing data" in terms of whether these nodules had criteria for FNA or not. Additionally, the origin of the referral was collected (primary care providers vs other specialties).

As mentioned above, our main aim was to compare the cohort of patients and thyroid nodules referred to FNA and their cytological result from 2015 with the 2021 cohort from the same period. Our secondary aim was to characterize and compare the subset of patients and thyroid nodules with an US classification described in the US report (group 1) with the remaining 2021 cohort (group 2) and with the 2015 cohort. When significant differences between group 1 and 2 were

not detected on the comparative analysis, we considered the 2021 cohort as a homogenous cohort for that specific variable and did not describe the comparative analysis of group 1 versus the 2015 cohort.

Data analysis and statistics

Data were retrieved from the SISPAT[®] software and were exported to Microsoft Office Excel[®] 2013. The statistical analysis was conducted on Statistical Package for the Social Sciences (IBM[®] SPSS Statistics, version 26.0).

Categorical variables were presented as absolute value (*n*) and relative frequency (%) and summarized into tables.

To assess the normality of the distribution of the continuous variables of each sample, we performed the Shapiro–Wilk test. If normality was assumed, values were presented as mean ± standard deviation. If normality was not assumed, values were presented as median and 25–75th percentile (P25–P75).

Pearson's chi-square test was used to assess the associations between categorical variables. Student's *t*-test was used to compare means among normally distributed variables and the Mann–Whitney *U* test was used to analyze non-normally distributed variables.

A *p*-value below 0.05 was considered statistically significant.

Results

Population characteristics

In this study, 606 adult patients were included (*n* = 386 in 2015; *n* = 220 in 2021). Within the 2021 cohort, group 1 included *n* = 133 patients (60.5%) and group 2 included *n* = 87 patients (39.5%). The number of patients undergoing FNA at IPATIMUP was higher in 2015 in comparison with 2021 due to the physician's availability to schedule the patients (57.0% more patients in 2015). Table 1 shows the general characteristics of the population.

In both years, the majority of patients were females (89.1% vs 85.9%, *p* = 0.243). The average age in 2015 was significantly lower when compared with 2021 (53.4 ± 14.5 years vs 57.8 ± 13.2 years, *p* < 0.001). The main source of referral was primary care providers. This source of referral was significantly more frequent in 2021 than in 2015 (89.1% vs 96.4%; *p* = 0.002).

No differences were observed between group 1 and group 2 regarding sex distribution (*p* = 0.138), mean age (*p* = 0.096) and source of referral (*p* = 0.904).

Characteristics of the thyroid nodules

A total of 471 nodules in 2015 and 273 nodules in 2021 were analyzed. Within the 2021 cohort, group 1 included 61.2% of the nodules (*n* = 165) and group 2 included 38.8% of the nodules (*n* = 106). These totals included all the nodules undergoing FNA biopsy (*n* = 463 in 2015 and *n* = 263 in 2021), plus the additional *n* = 8 nodules in 2015 and *n* = 10 nodules in 2021 for which FNA was requested but FNA was not performed. The reasons for not performing a biopsy in these

Table 1 General characteristics of the patients from 2015 and from 2021.

Variable	2015 (n = 386 patients)	2021 (n = 220 patients)	p
<i>Gender, n (%)</i>			
Male	42 (10.9)	31 (14.1)	0.243
Female	344 (89.1)	189 (85.9)	
<i>Age (years), mean ± SD</i>	53.4 ± 14.5	57.8 ± 13.2	<0.001
<i>Source of referral, n (%)</i>			
Primary care provider	344 (89.1)	212 (96.4)	0.002
Other specialties	42 (10.9)	8 (3.6)	
<i>Patients submitted to FNA, n (%)</i>			
1 nodule	316 (81.9)	180 (81.8)	0.993
2 nodules	63 (16.3)	37 (16.8)	
3 nodules	7 (1.8)	3 (1.4)	

FNA, fine needle aspiration; SD, standard deviation.

18 nodules were due to their small size (below 10 mm), and the fact the nodules did not have suspicious US features and were part of a multinodular disease.

The median nodule size in 2015 was 16.0 mm (P25–P75: 12.0–22.0 mm, ranging between 4 mm and 66 mm), while median nodule size in 2021 was 17.0 mm (P25–P75: 14.0–24.0 mm, ranging between 5 mm and 84 mm). This difference in the size of the nodules was statistically significant ($p=0.022$), seeming to be mainly due to an increase in nodules 10 mm or larger in 2021 (17.0 mm vs 19.0 mm, respectively, $p=0.002$). The nodules under 10 mm had similar medians (8.0 mm vs 8.05 mm, $p=0.873$). When comparing group 1 with the 2015 cohort, identical findings were obtained regarding median nodule size [17 mm (P25–P75: 14–24 mm) vs 16.0 mm (P25–P75: 12–22 mm), $p=0.016$], also due to the nodules larger than one centimetre [20 mm (P25–P75: 15–27 mm) 17 mm (P25–P75 13–22 mm), $p<0.001$], and not the ones under one centimetre in size (8.0 mm vs 8.05 mm, $p=0.873$).

Regarding the 2021 cohort, no significant differences regarding the median nodule size were detected between group 1 and group 2 [19.0 mm (P25–P75: 14.0–26.0 mm) vs 16.7 mm (P25–P75: 14.0–23.0 mm), $p=0.326$]; however, when dividing the nodules into under 10 mm and 10 mm or larger, similar results to those previously mentioned were detected. There were no differences regarding the nodules under one centimetre in size (8.0 mm vs 9.0 mm, $p=0.498$), but the nodules over one centimetre in size were significantly bigger in group 1 (20 mm vs 17 mm, $p=0.043$).

At least 3.2% of the nodules were under 10 mm in 2015 ($n=15$), as opposed to in 2021 which was almost double (6.6%; $n=18$), although this was not a statistically significant result ($p=0.155$). The size of the nodules was not recorded in 23.1% ($n=109$) of the nodules in 2015 and in 1.5% ($n=4$) of the nodules in 2021 ($n=3$ within group 1 – with an associated US categorization).

The median number of nodules for which FNA was requested per patient was one in both 2015 and 2021 (range in 2015: 1–5, range in 2021: 1–4). There was information regarding the number of nodules for which FNA biopsy was specifically requested in 428 nodules in 2015 (90.9%), and 271 nodules in 2021 (99.3%). There was no specific request

in $n=30$ patients from 2015 (7.8%) and in $n=2$ patients from 2021 (0.9%) that were excluded from this analysis.

From the 744 nodules analyzed, there was data available to assess criteria for undergoing FNA in 34.6% of the nodules from 2015 ($n=163$) and in 60.8% of the nodules from 2021 ($n=166$). Therefore, the availability of data enabling the criteria for FNA to be assessed was significantly better in 2021 ($p<0.001$). Within the 2021 cohort, there was also a significant difference between group 1 and group 2 regarding the available information to make this decision ($p<0.001$); the vast majority of the nodules where we could assess the criteria for FNA were from group 1 ($n=163$; 98.2%). The median nodules with an indication for FNA biopsy per patient was 1 in both years (range in 2015: 0–3, range in 2021: 0–4). FNA was requested for at least 23.1% ($n=63$) of all the nodules in 2021, although without US criteria [vs $n=158$ (33.5%) and $n=103$ (37.7%) nodules with inclusion criteria, respectively]. In both 2015 and 2021, none of the nodules reported to be under 10 mm and for which FNA was requested had criteria for FNA biopsy ($n=15$ and $n=18$, respectively).

The majority of the patients in both years underwent FNA of a single nodule (81.9% vs 81.8%, $p=0.993$). Therefore, the median nodules undergoing FNA per patient was one in both years (range in both years: 1–3).

US classification systems

None of the nodules from 2015 had any clinical information referring to the use of any risk stratification system. From the 2021 cohort, 60.5% of the patients had an US report with data regarding an US classification system (61.2% of the nodules; $n=167$) – group 1. More than one-third of the nodules from 2021 were not categorized using a US classification system (38.8%, $n=106$). The most frequently used US classification system in the sample from 2021 was the EU-TIRADS (39.2%, $n=107$) (Table 2). The majority of the nodules with a US classification were ranked as class 4 (46.5%, $n=74$) and 3 (35.6%, $n=57$). None were classified as class 1. Grading by US classification system used is reported on Table 2. There was a positive and significant correlation between having

Table 2 Distribution of the 273 nodules from 2021 as per the US classification system used and corresponding category.

US classification system used <i>n</i> (%)	EU-TIRADS	ACR-TIRADS	ATA	Not referred
	107 (39.2)	58 (21.2)	2 (0.7)	106 (38.8)
US classification system category	EU-TIRADS (<i>n</i> = 107)	ACR-TIRADS (<i>n</i> = 58)	ATA (<i>n</i> = 2)	Not known (<i>n</i> = 106)
Class 1, <i>n</i> (%)	0 (0)	0 (0)	0 (0)	NA
Class 2, <i>n</i> (%)	3 (2.8)	3 (5.2)	1 (50)	NA
Class 3, <i>n</i> (%)	29 (27.1)	27 (46.6)	1 (50)	NA
Class 4, <i>n</i> (%)	52 (48.6)	22 (37.9)	0 (0)	NA
Class 5, <i>n</i> (%)	16 (15.0)	5 (8.6)	0 (0)	NA
Not referred, <i>n</i> (%)	7 (6.5)	1 (1.7)	0 (0)	106 (100)

ACR, American College of Radiology; ATA, American Thyroid Association; EU, European; NA, not applicable; TIRADS, Thyroid Imaging Reporting and Data System; US, ultrasound.

criteria to undergo FNA and the US classification of the 2021 nodules ($p < 0.001$).

Cytology report

Table 3 shows the distribution of the 726 nodules submitted to FNA, as per the cytology report, according to the Bethesda classification¹¹ ($n = 463$ nodules in 2015; $n = 263$ nodules in 2021). There were no differences in terms of distribution of the Bethesda categories between 2015 and 2021 ($p = 0.082$). The same occurred when performing the analysis separately for the nodules that had criteria for FNA ($p = 0.185$) and the nodules that did not have criteria for FNA ($p = 0.974$). Within the 2021 cohort, there were no differences in the overall distribution of Bethesda categories ($p = 0.469$). When performing the analysis for the nodules without an indication for FNA, no differences were detected ($p = 0.988$). However, when evaluating exclusively nodules with indication for FNA, a significant difference was observed for the Bethesda III category nodules that was not taken into consideration due to the small sample size ($n = 5$; $n = 5$). In line with that, considering this small sample size effect for many of the Bethesda categories apart from Bethesda II in group 1, definite conclusions should not be drawn from comparisons between group 1 and the 2015 cohort for the Bethesda I and III to VI categories. When comparing the median nodule sizes for the Bethesda II category nodules, there were significant differences between groups (17 mm vs 19 mm, $p = 0.029$).

Regarding the size of the thyroid nodules, when performing the analysis separately for the nodules that were under 10 mm and the nodules that were 10 mm or larger, there were no differences in terms of Bethesda classification between 2015 and 2021 ($p = 0.586$ and $p = 0.091$, respectively). The most common Bethesda categories in any of these subgroups were Bethesda I and II. However, when comparing the median nodule sizes per Bethesda category, there were significant differences between 2015 and 2021. More specifically, between the Bethesda II and Bethesda III nodules from 2015 and 2021 (Table 4). The sample size of the nodules Bethesda V and VI was small [$n = 4$ ($n = 1$ in 2015, $n = 3$ in 2021) and $n = 10$ ($n = 9$ in 2015, $n = 1$ in 2021), respectively], with a wide range of nodule sizes (range 6.8–74 mm and 6.5–22 mm, respectively). Within the 2021 cohort, there were no differences between group 1 and group 2

regarding the median size of the nodules per Bethesda category, even when performing the analysis separately for the nodules under 10 mm and 10 mm or larger ($p > 0.05$). Table 5 summarizes the distribution of the nodules from 2021 that underwent FNA by US classification system category and the corresponding Bethesda category. In 41.4% of the 2021 nodules undergoing FNA ($n = 109$), the US classification system categorization was not applied.

Discussion

It is crucial to accurately distinguish between benign and malignant nodules, to avoid unnecessary surgical procedures, as well as to alleviate potential patients' psychological distress and reduce healthcare costs.

US is the recommended imaging exam for guiding the initial management of thyroid nodules and several US-based classification systems for risk stratification are currently available.^{4–9} These systems allow for a standardized risk-stratification assessment of thyroid nodules, leading to more consistent clinical decisions and better agreement between clinicians.^{4–9,13–16} They are intended to reduce the number of nodules with an indication for cytological assessment, and consequently, decrease the workload and costs associated with the workup of nodules that would be either benign or malignancies unlikely to cause harm during the patient's lifespan.^{14,15,17} These systems include those used in this study: EU-TIRADS, ACR-TIRADS and ATA.^{4,6,8}

Our results suggest that a change in the management of thyroid nodules in the North region of Portugal might have happened since the US-based classification systems for thyroid nodules were published.^{4,6,8} In comparison with 2015, there was a significant adherence to US classification systems in 2021, with most of the nodules being categorized according to one of these systems (61.2%), mainly with the EU-TIRADS (39.2%) and ACR-TIRADS systems. Although developed by an American society, the ACR-TIRADS was used in 21% of the available US reports. This might be due to its well-validated performance and associated higher specificity in triaging thyroid nodules when compared to the EU-TIRADS.¹⁵ The ATA⁴ classification was rarely used (only described in one US report), possibly due to its pattern-based system, instead of the points-based system of ACR TI-RADS and EU-TIRADS, leading to some nodules not being categorized.

Table 3 Distribution of the nodules undergoing FNA, according to the Bethesda classification.

Bethesda classification	2015 (n = 463)	2021 (n = 263)	
		Group 1 (n = 102)	Group 2 (n = 161)
I – Non diagnostic, n (%)	58 (12.5)	7 (29.2)	24 (9.1)
II – Benign, n (%)	366 (79.1)	87 (40.5)	215 (81.7)
III – AUS/FLUS, n (%)	9 (1.9)	5 (50)	10 (3.8)
IV – Follicular tumour, n (%)	20 (4.3)	3 (30)	10 (3.8)
V – Suspicious for malignancy, n (%)	1 (0.2)	0 (0)	3 (1.1)
VI – Malignant, n (%)	9 (1.9)	0 (0)	1 (0.4)
		0 (0)	1 (100)

AUS/FLUS: Atypia of Undetermined Significance or Follicular Lesion of Undetermined Significance.

Table 4 Comparison of the median nodule size per Bethesda category between 2015 and 2021.

Bethesda classification	2015 (n = 463) median size, mm	2021 (n = 263) median size, mm	p
I – Non diagnostic	14	14	0.642
II – Benign	17	19	0.036
III – AUS/FLUS	11	23	0.043
IV – Follicular tumour	17	14.9	0.352
V – Suspicious for malignancy	15	30	0.655
VI – Malignant	11	22	0.377

AUS/FLUS: Atypia of Undetermined Significance or Follicular Lesion of Undetermined Significance.

Table 5 Distribution of the nodules from 2021 undergoing FNA, as per their US classification system category and Bethesda classification (n = 263).

US classification system category	Bethesda classification					
	Bethesda I n (%)	Bethesda II n (%)	Bethesda III n (%)	Bethesda IV n (%)	Bethesda V n (%)	Bethesda VI n (%)
EU-TIRADS 2 (n = 3)	0 (0)	3 (1.4)	0 (0)	0 (0)	0 (0)	0 (0)
EU-TIRADS 3 (n = 29)	2 (8.3)	24 (11.2)	1 (10)	1 (10)	1 (33.3)	0 (0)
EU-TIRADS 4 (n = 52)	5 (20.8)	42 (19.5)	1 (10)	3 (30)	1 (33.3)	0 (0)
EU-TIRADS 5 (n = 16)	2 (8.3)	13 (6.0)	0 (0)	1 (10)	0 (0)	0 (0)
ACR-TIRADS 2 (n = 2)	1 (4.2)	1 (0.5)	0 (0)	0 (0)	0 (0)	0 (0)
ACR-TIRADS 3 (n = 26)	3 (12.5)	21 (9.8)	2 (20)	0 (0)	0 (0)	0 (0)
ACR-TIRADS 4 (n = 21)	3 (12.5)	17 (7.9)	0 (0)	1 (10)	0 (0)	0 (0)
ACR-TIRADS 5 (n = 4)	1 (4.2)	2 (0.9)	0 (0)	1 (10)	0 (0)	0 (0)
ATA 3 (n = 1)	0 (0)	1 (0.5)	0 (0)	0 (0)	0 (0)	0 (0)
Unknown (n = 109)	7 (29.2)	91 (42.3)	6 (60)	3 (30)	1 (33.3)	1 (100)
TOTAL	24 (100)	215 (100)	10 (100)	10 (100)	3 (100)	1 (100)

ACR, American College of Radiology; ATA, American Thyroid Association; EU, European; NA, Not applicable; TIRADS, Thyroid Imaging Reporting and Data System; US, Ultrasound.

Nevertheless, some features of the population and respective nodules submitted to FNA biopsy were similar between 2015 and 2021, namely the sex distribution (mainly females, as expected according to the literature^{18,19}), the number of nodules for which FNA was requested per patient,

the number of nodules undergoing FNA and the Bethesda classification of the nodules.

The most significant differences recorded between the two years were as follows: the patients in 2021 were 4.4 years older than in 2015, the size of the nodules was

significantly larger in 2021 than in 2015 mainly due to nodules larger than 10 mm, and the percentage of patients referred from primary care providers was higher in 2021 (89.1% vs 96.4%; $p=0.002$). The differences described here may have been due to at least two potential factors. First, the implementation of the US classification systems potentially allowed for extended surveillance periods, leading to older patients and larger nodules. This last observation is further corroborated by the results obtained by the comparison of group 1 with group 2 and with the 2015 cohort. The nodules categorized with an US system (group 1) were significantly larger, apparently at the cost of the nodules larger than one centimetre (3 mm larger in diameter when compared to the nodules from group 2 and the 2015 cohort). On the other hand, it may also represent the impact of the COVID-19 pandemic. It resulted in a delay in access to health care services and the emergence of financial-social difficulties, leading to fewer patients resorting to private practice and a higher referral rate from primary care providers.

Comparing the nodule size per Bethesda category, both Bethesda II and III category nodules were significantly larger in 2021 (17 mm vs 19 mm, $p=0.036$; 11 mm vs 23 mm, $p=0.043$, respectively). Bethesda V and VI nodules should not be compared regarding their dimensions due to the small sample size. The size difference between the Bethesda III nodules (12 mm) is clinically relevant, since the performing FNA in larger nodules is usually easier, more successful, and with a higher sensitivity.^{20,21} However, it is important to mention that the sample size of the Bethesda III nodules was small. It seems that the US classification systems allow delaying FNA, without aggravating the cytology results when FNA becomes indicated, leading to larger nodules in 2021 undergoing FNA. Even the detected size differences of 1 or 2 mm, although not clinically relevant, might be indicative of this positive trend. As the Bethesda IV-VI categories have a small sample size not detecting differences regarding the nodules' dimensions and Bethesda category distribution, it might be due to the low statistical power associated with small populations. According to Tables 3 and 4, we can perceive an apparent tendency towards bigger nodules in 2021 across the six Bethesda categories and a slight change in the distribution of the nodules throughout these categories from 2015 to 2021. Regarding the Bethesda category distribution, we could not make correlations with the US risk stratifying systems, as an US classification system categorization was not applied in 41.4% of the 2021 nodules undergoing FNA. However, the most reported US category was EU-TIRADS 4. We reported $n=16$ nodules EU-TIRADS 5 and $n=5$ nodules ACR-TIRADS 5, however they did not correspond to the Bethesda V and VI nodules of the 2021 cohort. As shown in Table 5, none of these Bethesda V-VI nodules were classified as EU-TIRADS 5 or ACR-TIRADS 5. The majority of the nodules EU-TIRADS 5 and ACR-TIRADS 5 undergoing FNA were Bethesda II. Nonetheless, in $n=1$ Bethesda V and $n=1$ Bethesda VI, no US classification system was applied so we cannot exclude they would not be classified as EU-TIRADS 5 or ACR-TIRADS 5. We acknowledge a very low percentage of Bethesda V and VI in our cohorts. Possible explanations for the scarcity of higher risk Bethesda categories might be the high percentage of nodules referred for FNA that did not have criteria for it; the fact that the US were performed by multiple radiologists with different degrees of

experience in thyroid US (being the majority non-exclusively dedicated to this type of US). Another important factor to considered is that there is a potential referral bias associated with the IPATIMUP itself, where the majority of the FNAs are requested by primary care providers, as opposed to the hospital setting where they have a higher percentage of referrals from hospital specialties, and nodules diagnosed in more complex contexts. Additionally, some more complicated situations are sent to IPATIMUP from private practice, and occasionally from public hospitals, namely, consecutive Bethesda category I or III and/or technically difficult to perform FNA biopsies. These nodules might not have initially had an indication for FNA. As the low rate of Bethesda V and VI was observed both in 2015 and 2021, a potential referral bias and high percentage of referred nodules without an indication for FNA are the most plausible justifications. A further topic of research in the IPATIMUP will be to assess in a larger population the potential variables responsible for this atypical low percentage of Bethesda V and VI categories in our sample and to have the statistical power to correlate the Bethesda categories with the US classification systems categories.

The most recent US-based classification systems^{4,6,8} (and the ones included in this study) are more conservative than the guidelines in force in 2015^{5,12} regarding nodule size thresholds that indicate cytological assessment. Nodules under one centimetre no longer indicate FNA, even if high-risk factors are present, as the course of papillary microcarcinoma is usually indolent.⁴ Consequently, this leads to fewer FNA. However, in 2021, nodules under 10 mm are still being referred to FNA. At least 6.6% of nodules referred in 2021 were under 10 mm, as opposed to 2015 (at least 3.2%). Nevertheless, in 2015 we did not have access to the nodule size in 33.5% of the nodules, and therefore a direct comparison between the two years should be interpreted with caution. Even with several risk-stratification systems available in 2021, nearly 40% of the nodules are still not being stratified with an ultrasound classification system, precluding the clinician to assess indication for FNA in virtually all of those nodules. This goes along the same lines of a recent study reporting inconsistent and incomplete descriptions of thyroid nodules, with some US features only being reported in as few as 9% of the nodules.²² Consequently, nearly all the 2021 nodules where we were able to assess the criteria for FNA were nodules classified with an US risk stratifying system (98.2%), and the $n=3$ nodules without this categorization that we were able to assess indication for FNA was due to the large dimensions of the nodules (more than 20 mm in diameter). Additionally, at least 38% of the nodules referred to FNA in 2021 did not have criteria to do it, testifying that the clinicians do not always follow the recommendations associated with each risk category. This might partly justify why the distribution of the Bethesda categories did not significantly change from 2015 to 2021.

The ongoing referral for biopsy of nodules under one centimetre in size and other nodules without indication for cytological assessment also shows us that other factors are accountable for this, independently of these classification systems. We could speculate some contributing factors are the patients' expectations, as research shows office consults' satisfaction suffers and health-related anxiety increases when wanted diagnostic interventions are not

obtained,^{23,24} and, somewhat related, the rise of defensive medicine in Europe.²⁵

It is pertinent to mention two important caveats concerning our data. We had a significant percentage of missing data regarding nodule size in 2015 (33.5% of the nodules) and regarding information on US criteria to undergo FNA in both 2015 and 2021 (65.4% vs 39.2%, respectively). To assess the indication for FNA of these nodules in 2015, we had to use that limited data (mainly, the nodule size), leading to a potential information bias without taking into consideration potential specific risk factors such as a history of thyroid cancer in first degree relatives, history of external beam radiation as a child, previous hemithyroidectomy with discovery of thyroid cancer, and ¹⁸F¹⁸FDG avidity on PET scans.¹² These factors are no longer considered in the updated systems,^{4,6,8} facilitating their application. This study includes other limitations. Thyroid US was performed by multiple radiologists introducing operator-driven variability that was not controllable. However, we can contemplate this as a strength of the study since the wide diversity of examiners with different levels of expertise performing the US led to a sample more representative of what is happening in daily clinical practice. Another limitation was the retrospective design of this study, leading to gaps in information. The higher number of patients who underwent FNA in 2015 (57% more patients) was due to institutional constraints associated with the availability of the pathologist performing this technique in the previous months, leading to a bigger waitlist in October 2015.

Conclusion

This study is the first study to compare the reality in Portugal of thyroid FNA and cytology before and after the publication of the most commonly used US-based classification systems for thyroid nodules. To the best of our knowledge, it is also the first published study worldwide making this comparison and assessment. This analysis draws attention to the importance of systematically applying US-based classification systems for thyroid nodules in our daily clinical practice. It seems that these systems, by not being focused mainly on size thresholds, allow for extended surveillance periods, without aggravating the cytology results when FNA becomes indicated. FNA potentially become safer as the nodules are permitted to have larger dimensions. Nevertheless, bigger efforts are needed to ensure more standardized US reports by systematically using these risk-stratification systems and, when applied, to increase adherence to the resulting recommendations. This will allow more consistent clinical decisions, with a decrease in clinical uncertainty, unnecessary FNA biopsies, and potential overtreatment.

Ethical considerations

All procedures were in accordance with the ethical standards of the institutional ethics committee and in line with the principles of the Declaration of Helsinki.

Funding sources

This research did not receive any specific grant from funding agencies.

Conflict of interest

No conflicts of interest in connection with this article.

Acknowledgments

Not applicable.

References

- Guth S, Theune U, Aberle J, Galach A, Bamberger CM. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. *Eur J Clin Invest.* 2009;39:699–706, <http://dx.doi.org/10.1111/j.1365-2362.2009.02162.x>.
- Dean DS, Gharib H. Epidemiology of thyroid nodules. *Best Pract Res Clin Endocrinol Metab.* 2008;22:901–11, <http://dx.doi.org/10.1016/j.beem.2008.09.019>.
- Vaccarella S, Dal Maso L, Laversanne M, Bray F, Plummer M, Franceschi S. The impact of diagnostic changes on the rise in thyroid cancer incidence: a population-based study in selected high-resource countries. *Thyroid.* 2015;25:1127–36, <http://dx.doi.org/10.1089/thy.2015.0116>.
- Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, 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, <http://dx.doi.org/10.1089/thy.2015.0020>.
- Horvath E, Majlis S, Rossi R, Franco C, Niedmann JP, Castro A, et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. *J Clin Endocrinol Metab.* 2009;94:1748–51, <http://dx.doi.org/10.1210/jc.2008-1724>.
- Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European thyroid association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: the EU-TIRADS. *Eur Thyroid J.* 2017;6:225–37, <http://dx.doi.org/10.1159/000478927>.
- Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedüs L, et al. 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. *Endocr Pract.* 2016;22:622–39, <http://dx.doi.org/10.4158/EP161208>.
- Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): white paper of the ACR TI-RADS committee. *J Am Coll Radiol.* 2017;14:587–95, <http://dx.doi.org/10.1016/j.jacr.2017.01.046>.
- Ha EJ, Chung SR, Na DG, Ahn HS, Chung J, Lee JY, et al. 2021 Korean thyroid imaging reporting and data system and imaging-based management of thyroid nodules: Korean Society of Thyroid Radiology Consensus Statement and Recommendations. *Korean J Radiol.* 2021;22:2094–123, <http://dx.doi.org/10.3348/kjr.2021.0713>.
- Remonti LR, Kramer CK, Leitão CB, Pinto LC, Gross JL. Thyroid ultrasound features and risk of carcinoma: a systematic

- review and meta-analysis of observational studies. *Thyroid*. 2015;25:538–50, <http://dx.doi.org/10.1089/thy.2014.0353>.
11. Cibas ES, Ali SZ. The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid*. 2017;27:1341–6, <http://dx.doi.org/10.1089/thy.2017.0500>.
 12. American Thyroid Association (ATA) Guidelines Taskforce on Thyroid Nodules and Differentiated Thyroid Cancer, Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, Mandel SJ, Mazzaferri EL, McIver B, Pacini F, Schlumberger M, Sherman SI, Steward DL, Tuttle RM. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2009;19:1167–214. [doi:10.1089/thy.2009.0110](https://doi.org/10.1089/thy.2009.0110). Erratum in: *Thyroid*. 2010 Jun;20(6):674–5.
 13. Grani G, Lamartina L, Cantisani V, Maranghi M, Lucia P, Durante C. Interobserver agreement of various thyroid imaging reporting and data systems. *Endocr Connect*. 2018;7:1–7, <http://dx.doi.org/10.1530/EC-17-0336>. Epub 2017 Dec 1. PMID: 29196301; PMCID: PMC5744624.
 14. Pang Z, Margolis M, Menezes RJ, Maan H, Ghai S. Diagnostic performance of 2015 American Thyroid Association guidelines and inter-observer variability in assigning risk category. *Eur J Radiol Open*. 2019;6:122–7, <http://dx.doi.org/10.1016/j.ejro.2019.03.002>. Erratum in: *Eur J Radiol Open*. 2020 Dec 17;8:100308. PMID: 30976628; PMCID: PMC6443900.
 15. Seminati D, Capitoli G, Leni D, Fior D, Vacirca F, Di Bella C, et al. Use of diagnostic criteria from ACR and EU-TIRADS systems to improve the performance of cytology in thyroid nodule triage. *Cancers (Basel)*. 2021;13:5439, <http://dx.doi.org/10.3390/cancers13215439>. PMID: 34771602; PMCID: PMC8582424.
 16. Hoang JK, Middleton WD, Farjat AE, Teefey SA, Abinanti N, Boschini FJ, et al. Interobserver variability of sonographic features used in the american college of radiology thyroid imaging reporting and data system. *AJR Am J Roentgenol*. 2018;211:162–7, <http://dx.doi.org/10.2214/AJR.17.19192>. Epub 2018 Apr 27. PMID: 29702015.
 17. Persichetti A, Di Stasio E, Guglielmi R, Bizzarri G, Tacogna S, Misischi I, et al. Predictive value of malignancy of thyroid nodule ultrasound classification systems: a prospective study. *J Clin Endocrinol Metab*. 2018;103:1359–68, <http://dx.doi.org/10.1210/jc.2017-01708>. PMID: 29408952.
 18. Ezzat S, Sarti DA, Cain DR, Braunstein GD. Thyroid incidentalomas. Prevalence by palpation and ultrasonography. *Arch Intern Med*. 1994;154:1838–40, <http://dx.doi.org/10.1001/archinte.154.16.1838>. PMID: 8053752.
 19. Xu L, Zeng F, Wang Y, Bai Y, Shan X, Kong L. Prevalence and associated metabolic factors for thyroid nodules: a cross-sectional study in Southwest of China with more than 120 thousand populations. *BMC Endocr Disord*. 2021;21:175, <http://dx.doi.org/10.1186/s12902-021-00842-2>. PMID: 34452638; PMCID: PMC8400405.
 20. Shrestha M, Crothers BA, Burch HB. The impact of thyroid nodule size on the risk of malignancy and accuracy of fine-needle aspiration: a 10-year study from a single institution. *Thyroid*. 2012;22:1251–6, <http://dx.doi.org/10.1089/thy.2012.0265>. Epub 2012 Oct 19. PMID: 22962940.
 21. Kim DW, Lee EJ, Kim SH, Kim TH, Lee SH, Kim DH, et al. Ultrasound-guided fine-needle aspiration biopsy of thyroid nodules: comparison in efficacy according to nodule size. *Thyroid*. 2009;19:27–31, <http://dx.doi.org/10.1089/thy.2008.0106>. PMID: 19021460.
 22. Qadan L, Ahmed A, Kapila K. Thyroid ultrasound reports: deficiencies and recommendations. *Med Princ Pract*. 2019;28:280–3, <http://dx.doi.org/10.1159/000497789>. Epub 2019 Feb 10. PMID: 30739108; PMCID: PMC6597922.
 23. Kravitz RL, Cope DW, Bhrany V, Leake B. Internal medicine patients' expectations for care during office visits. *J Gen Intern Med*. 1994;9:75–81, <http://dx.doi.org/10.1007/BF02600205>. PMID: 8164081.
 24. Sox HC Jr, Margulies I, Sox CH. Psychologically mediated effects of diagnostic tests. *Ann Intern Med*. 1981;95:680–5, <http://dx.doi.org/10.7326/0003-4819-95-6-680>. PMID: 7305144.
 25. Garattini L, Padula A. Defensive medicine in Europe: a 'full circle'? *Eur J Health Econ*. 2020;21:165–70, <http://dx.doi.org/10.1007/s10198-019-01144-0>. Epub 2019 Dec 26. Erratum in: *Eur J Health Econ*. 2020 Jul;21(5):815. PMID: 31879861.