Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.JournalofSurgicalResearch.com

Improving Diagnostic Performance for Thyroid Nodules Classified as Bethesda Category III or IV: How and by Whom Ultrasonography Should be Performed

Gregorio Scerrino, PhD, MD,[a,*] *Gianfranco Cocorullo, OP, MD,*[a] *Sergio Mazzola, MD,*[b] *Giuseppina Melfa, MD,*[a] *Giuseppina Orlando, MD,*[a] *Iole Laise, MD,*[c] *Alessandro Corigliano, MD,*[a] *Daniela Lo Brutto, MD,*[a] *Calogero Cipolla, AP, MD,*[c] *and Giuseppa Graceffa, PhD, MD*[c]

[a] *Department of Surgical Oncology and Oral Sciences, Unit of General and Emergency Surgery, University of Palermo, Palermo, Italy*

[b] *Unit of Clinical Epidemiology and Tumor Registry, Department of Laboratory Diagnostics, Policlinico "P. Giaccone", University of Palermo, Palermo, Italy*

[c] *Department of Surgical Oncology and Oral Sciences, Unit of Oncological Surgery, University of Palermo, Palermo, Italy*

ARTICLE INFO

Article history:
Received 4 May 2020
Received in revised form
30 September 2020
Accepted 4 December 2020
Available online xxx

Keywords:
Thyroid cytology
Thyroid ultrasound
Surgeon performed
thyroid ultrasonography

ABSTRACT

Background: The purpose of this prospective study is to evaluate if the association of Bethesda system and a 3-categories Ultrasonography (US) risk stratification system proposed by the American Association of Clinical Endocrinologists/American College of Endocrinology/Associazione Medici Endocrinologi improves the performance of cytology alone in III or IV categories and if further variables such as US provider (radiologist; endocrinologist, or endocrine surgeon both coming from a dedicated team) influence the accuracy of the diagnostic.

Methods: 570 consecutive patients with complete clinical records, affected by Bethesda III or IV nodules, have been addressed to two public referral surgical centers of Western Sicily. Age, sex, autoimmunity, nodule size, and US provider were recorded. Fisher's exact test was used for the univariate analysis; Odd's ratios were calculated for the multivariate analysis.

Results: 248 patients had malignancy at histology, 322 were benign. The mean age was 52 years for the malignancy group and 58 y for the benign group ($P < 0.001$). At univariate analysis, autoimmunity was correlated with benign group ($P < 0.001$), and US risk 2 and 3

Author's contribution: material preparation was performed by D.L.B. and J.L.; data collection was performed by G.M., G.O., and A.C.; data analysis was performed by S.M. The first draft of the manuscript was written by C.C.; G.S.. G.C. contributed in study conception and design, final draft revision for final approval. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Authors Gregorio Scerrino e Gianfranco Cocorullo have contributed equally to this work.

Grant support for the research: none.

Disclosure: the authors declare that they have no conflict of interest.

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

Ethical board approval: obtained by "Comitato etico del Policlinico Paolo Giaccone—Palermo".

* *Corresponding author.* Unit of General and Emergency Surgery, Policlinico "P. Giaccone", Via L. Giuffré, 5, 90100 Palermo, Italy. Tel.: +393427160673; fax: +390916552814.

E-mail address: gregorio.scerrino@tiscali.it (G. Scerrino).

0022-4804/$ – see front matter

https://doi.org/10.1016/j.jss.2020.12.009

were correlated with malignancy (nearly 10-folds, $P < 0.001$); In addition, no difference was found concerning nodule size. At multivariate analysis, US risk 2 and 3 were strong predictors of malignancy ($P < 0.0001$) especially if cytology was Bethesda IV; endocrinologist and surgeon were more accurate in predicting malignancy compared with the radiologist ($P < 0.01$).

Conclusions: In the context of indeterminate nodules, the American College of Endocrinology/American Association of Clinical Endocrinologists/Associazione Medici Endocrinologi US risk stratification system strongly improves the results of Bethesda system especially when performed from dedicated endocrinologist or endocrine surgeon.

Background

Thyroid ultrasonography is a sensitive and specific examination in evaluating thyroid nodules; it is the crucial imaging tool used to identify and classify thyroid lesions. Moreover, it addresses suspicious nodules for fine-needle aspiration biopsy (FNAB) for obtaining tissue samples and is useful to confirm or rule out malignancies.[1,2] Because its sensitivity could increase the number of unnecessary FNABs and diagnostic surgeries performed for “low-risk” lesions, statements for US risk and stratification have been suggested in several studies.[3,4]

Currently, the screening of lesions referred to surgery is performed as per the cytology results once US preselection has been performed earlier. Several scientific societies or individual research groups have proposed different systems for the assessment of US risk stratification of thyroid nodules.[5-7] A simple scoring system for US reporting has been proposed by the American Association of Clinical Endocrinologists (AACE), the American College of Endocrinology (ACE), and the (Italian) Associazione Medici Endocrinologi (AME). They proposed a new US rating system for the risk of malignancy for thyroid nodules, which is a three-category system that appeared easy to apply in practice.[8] Concerning cytology, the Bethesda system for reporting thyroid cytopathology had widespread diffusion and represented a step forward in systematizing the interpretation of cytology.[9] At the same time, it is the landmark for benign (Bethesda II), suspected (Bethesda V), or confirmed (Bethesda VI) malignant nature of a thyroid nodule. Both US and cytopathology separately have excellent diagnostic accuracy; however, no consensus has been reached to date in the treatment of undetermined (Bethesda III and IV) lesions.

This study aimed to verify the hypothesis that US, performed as per the ACE/AACE/AME US risk stratification system, is capable of increasing the diagnostic performance of cytology in Bethesda III and IV cytology categories assuming the same frequency of malignancy in both classes. Moreover, we evaluated whether providers performing US could influence the accuracy of the diagnosis, especially in terms of correct indication for surgery.

Methods

This institutional prospective study was performed in two public high-volume centers (>100 thyroidectomies/year) for thyroid surgery in Western Sicily: General and Emergency Surgery and General and Oncological Surgery both belong to the Department of Surgical, Oncological, and Oral Sciences of the University of Palermo. Together with the endocrinologists of the Endocrinology and Metabolic Diseases unit of Policlinico "P. Giaccone," University of Palermo, and two expert endocrinologists of Palermo district ("Azienda Ospedaliera Ospedali Riuniti Villa Sofia-Cervello" and "Ospedale Civico di Partinico"), two high-volume surgeons at high-volume centers who work on a multidisciplinary, dedicated team and share protocols, and guidelines and consensus meetings for the management of clinical cases participated in the study.

The study enrolled consecutive patients who underwent thyroidectomy (thyroid lobectomy or total thyroidectomy) from January 2012 to June 2019. Patients referred at one of the high-volume centers had US performed in one of the following settings: external radiologist/ultrasonographer (provider 1), one among expert endocrinologists on the dedicated team (provider 2), and one between endocrine surgeons of the dedicated team (provider 3). US was always performed by echographs endowed with a high-frequency (7,5-12 mHz) probe. Because the ACE/AACE/AME US risk score was used at our institution beginning on June 2016, and it is not diffusely applied to date, reports of external radiologists, or those performed before June 2016, were led back to the ACE/AACE/AME score system if possible. In both these circumstances, the risk score was assigned by the two thyroid surgeons after they had discussed the reinterpretation of the ultrasound reports and, when available, the archived images captured by the original provider. Patients with a lack of data in which this report’s translation was impossible or unclear were excluded from the study. All patients were recruited after an outpatient evaluation in one of the two surgery units or referral for surgery after endocrinology evaluation performed at our institution or elsewhere.

Criteria for surgery in the group of undetermined/suspicious lesions were Bethesda IV (all) or growing Bethesda III lesions, suspicious US findings (irregular/speculated margins, markedly hypoechoic nodules, intralesional irregular vascularization, microcalcifications, "taller than wide" shape, etc.) nodules > 4 cm in major diameter especially in the presence of compressive symptoms, and cosmetic and/or psychological concerns, including the patient’s intolerance to long-term follow-up. These criteria were consistent during the study period.

Patients suffering from any thyroid nodular disease (single nodule, multinodular goiter) with complete clinical records

from enrollment to discharge (preoperative and postoperative laboratory tests, operative report, and definitive histology) were included in the study.

We excluded US reports not in agreement with the AACE/ACE/AME standards, the absence of cytopathologic reports, or incomplete descriptions in accordance with the Bethesda system. Moreover, we excluded benign thyroid diseases (Grave's disease) and malignancies (medullary thyroid carcinomas) because their diagnoses are strongly connected with specific laboratory tests and all malignancies other than differentiated of follicular origin. We also excluded patients with family history of thyroid cancer, personal history of neck irradiation, or previous malignancy. Finally, we excluded patients with diagnosed (clinical, US, or intraoperative) gross extrathyroidal extension disease, including central and/or lateral metastases because we did not evaluate this specific topic.

Figure 1 summarizes in a flow chart the enrollment process highlighting the number of patients excluded and the reason for their exclusion.

Patients enrolled in this study had complete clinical records, including US and cytology. US risk stratification was classified into 3 categories (1 = low risk; 2 = intermediate risk; 3 = high risk) as per ACE/AACE/AME criteria (Table 1).

An endocrinologist or endocrine surgeon performed fine-needle aspiration, and slides obtained were examined by a dedicated pathologist at a university hospital. Cytology was reported as per Bethesda criteria. Only patients affected by nodules classified as Bethesda III and IV (thyroid nodules that showed atypia of uncertain significance or follicular lesion of undetermined significance and follicular neoplasm or suspicious for follicular neoplasms) were included.

The US provider (radiologist, endocrinologist, and endocrine surgeon) was also recorded. Because all nodules were evaluated in a team manner before they were referred to surgery, US performed by a radiologist was always repeated, but only the prereferral report was recorded for this study.

Age, sex, autoimmunity (assessed as per autoantibodies anti-TPO and anti-TG serum levels and confirmed at histological report), and nodule size (measured on US but definitively established as per histologic report) were also included in the recorded variables to evaluate their possible influence on the results. The "T" category was established as per the TNM staging system, AJCC/UICC, seventh edition, in which T1 size was subdivided into T1a (≤10 mm in diameter) and T1b (11-20 mm).[10,11] All data were related to definitive results of pathology, identifying two groups: benign (B) and malignant (M) at definitive histology.

The nodule assessed with FNAB and ultrasound was indeed confirmed to be the nodule that was considered for the present study as histologically benign or malignant. As per prevailing literature trends,[12] the three non-invasive follicular thyroid neoplasms with papillary- like nuclear features diagnosed after thyroidectomy (one reported as Bethesda III in preoperative workup, two Bethesda IV) were classified as benign.

The present study was performed as per the declaration of Helsinki (1964) and its amendments, informed consent was obtained and the Ethics Committee of Azienda Ospedaliero—Universitaria Policlinico "P. Giaccone"—Palermo approved the study.

Statistical analysis

Data were included as discrete in an Excel sheet. Fisher's exact test was used for univariate analysis, and odds ratios were calculated in a logistic regression model for multivariate analysis. IDE RStudio software (version 3.4.1 of 2017-06-30) was used for univariate and multivariate analysis.

Results

A total of 570 patients (526 women, 92%; 44 men, 8%) matched the recruitment criteria. The mean age was 56 y (range: 20-86).

In 248 patients (43.5%), the definitive diagnosis was malignant nodule (M group) at histologic report, 322 patients (56.5%) had benign nodules (B group). The mean age was significantly different between the two groups: 52 years for the M group and 58 for the B group ($P < 0.001$). The distribution of sex in the two groups was homogeneous ($P = 0.16$). The univariate analysis investigated the relation between each variable and the definitive result of histology (Table 2). As expected, the relation of Bethesda category and histology had a *P*-value in the range of nonsignificance. In this analysis, the chi-square test measured the degree of relationship of each Bethesda value with histology, evaluated as a dichotomous variable (benign/malign). Therefore, both undetermined categories have an increased risk of malignancy although their distributions are statistically different.

The logistic regression showed that, at least in the series of patients examined, more advanced age and autoimmunity seem to be protective toward malignancy (OR = 0.98 and 0.56, respectively). In particular, the variable "age" was categorized by dividing it into 3 ranges: <45 y, 45-60 y, and > 60 y. This solution was preferred because it allows to identify for which age group the risk of malignancy was protective. The reference group of patients <45 y had an OR = 1 (not significant); therefore, the group of patients ranging 45-60 y in comparison with <45 y had an OR = 0.71, statistically not significant; the patients >60 y compared with the <45 y old showed an OR = 0.48 ($P < 0.01$). It means that age>60 y is protective against thyroid cancer.

Bethesda category IV is correlated with malignancy 1.5-fold compared with category III. In this model, the Bethesda category is considered as a qualitative variable. In this case, the significance should be interpreted as the excess risk of malignancy of Bethesda IV compared with Bethesda III. The correlation of advanced US risk category (2 and 3) with the M group was strong, and both categories were capable of revealing malignancy nearly 10-fold compared with US risk category 1.

Figure 2 shows the relationship between sensitivity and specificity of ACE/AACE/AME risk scale by means of a ROC curve. In our study, it has shown a moderate accuracy.

Gender and nodule size did not influence the risk of malignancy. On the contrary, the variable "autoimmunity," if present, seemed to be a protective factor against malignancy. At multivariate analysis, US risks 2 and 3 were strong independent predictors of malignancy ($P < 0.0001$), especially if the cytology was Bethesda IV; endocrinologists and surgeons

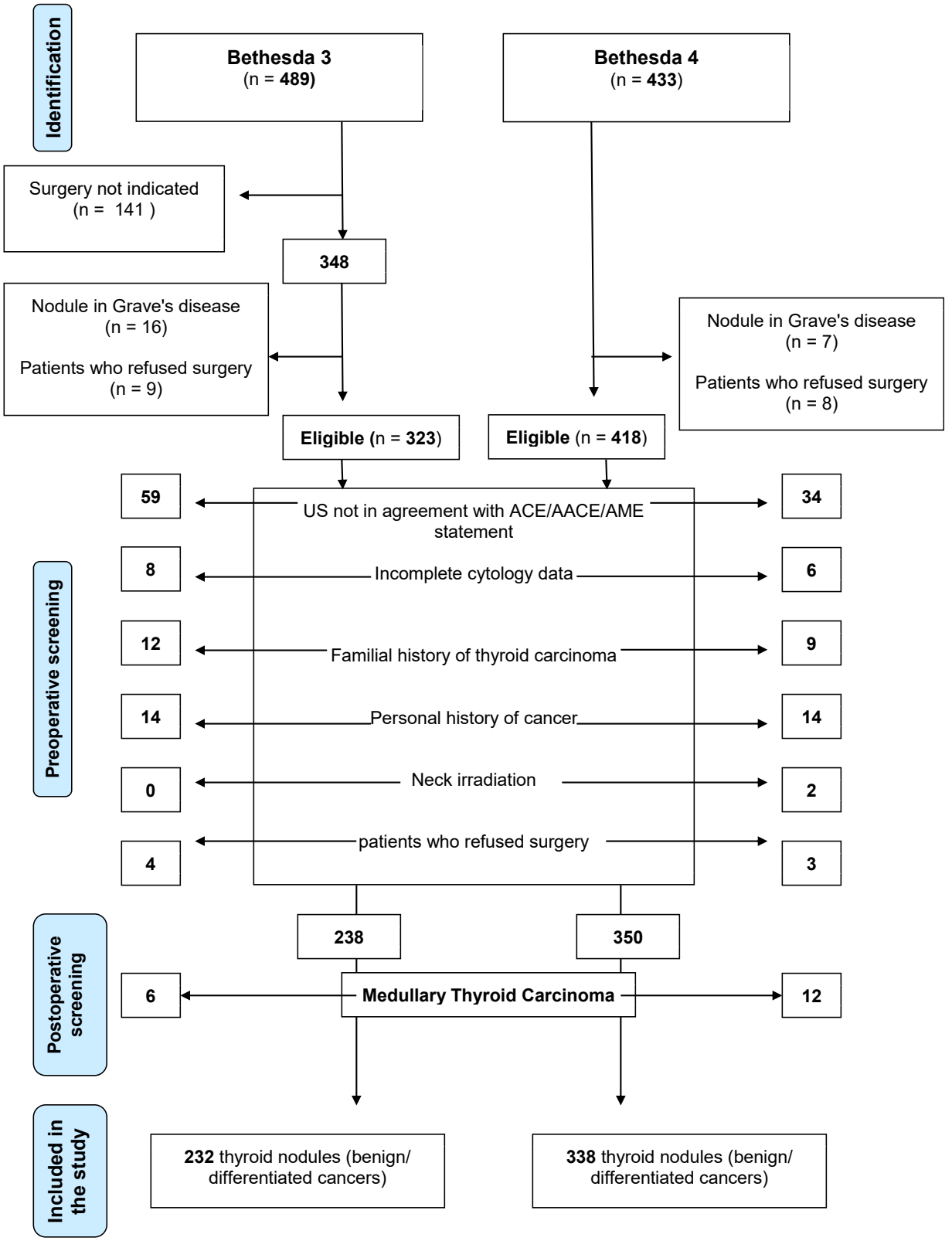

Fig. 1 – Patient recruitment process. Starting from the total number of patients identified during outpatient endocrine surgery visit of the units involved in the study, a description is given of how many patients were excluded, at what stage of the process and for what reasons. (Color version of figure is available online.)

Table 1 – ACE/AACE/AME US risk scores.

Class 1: Low-risk thyroid lesion (expected risk of malignancy: about 1%)
- Mostly cystic (>50%) nodules with reverberating artifacts that are not associated with suspicious US signs;
- Isoechoic spongiform nodules confluent or with regular halo
Class 2: Intermediate risk thyroid lesion (expected risk of malignancy: 5-15%)
- Slightly hypoechoic nodules (compared with surrounding thyroid tissue) and isoechoic nodules with ovoid to round shape and smooth or ill-defined margins
- Either intranodular vascularization, elevated stiffness at elastography, macrocalcifications or continuous rim calcifications, or hypoechoic spots of uncertain significance may be present
Class 3: High-risk thyroid lesion* (expected risk of malignancy: 50-90%)
- Marked hypoechogenicity (compared with prethyroid muscles)
- Spiculated or microlobulated margins
- Microcalcifications
- Taller-than-wide shape
- Evidence of extrathyroidal growth or pathologic adenopathy

*Nodules with at least one of the following suspicious features.
From: Gharib H, and coll. See reference.[8]

Table 2 – Demographics and univariate analysis.

Variable	Benign	Malignant	Total	P-value
Mean age	58	52		<0,0001
Gender				
F	302	224	526	0.1679
M	20	24	44	
Total	322	248	570	
Autoimmunity				
No	144	172	316	<0.0001
Yes	178	76	254	
Total	322	248	570	
Size				
T1a	61	54	115	0.8365
T1b	208	158	366	
T2	50	34	84	
T3	3	2	5	
Total	322	248	570	
US risk				
1	152	16	168	<0.0001
2	106	160	266	
3	64	72	136	
Total	322	248	570	
Bethesda				
3	137	95	232	0.3495
4	185	153	338	
Total	322	248	570	
Provider				
1	193	84	277	<0.0001
2	35	48	83	
3	94	116	210	
Total	322	248	570	

Univariate analysis: Age showed a significant difference in the two groups (malignancy group was younger); malignancy was more frequent in patients without autoimmunity; US risk 2 and 3 and in patients addressed to surgery from endocrinologist (provider 2) and surgeon (provider 3); malignancy was not significantly different in Bethesda III *versus* IV if this variable was analyzed in this univariate model; nodule size was not a predictor of malignancy.

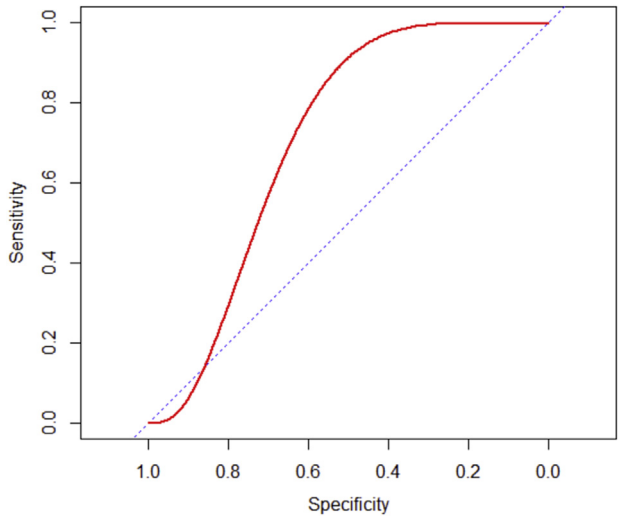

Fig. 2 – ROC curve representing the relationship between the sensitivity and specificity of the ACE/AACE/AME US risk score used in this study. Below are the criteria for evaluating the ability of the scale to detect malignancies. Sensitivity = 0.94 CI 95% (0.90-0.96); Specificity = 0.47 CI 95% (0.42-0.53); c-index AUC = 0.71 CI 95% (0.66-0.75). Based on this result, the ACE/AACE/AME US risk score is moderately accurate. (Color version of figure is available online.)

were more accurate than radiologists in predicting malignancy ($P < 0.01$). Finally, autoimmunity confirmed to be protective as well as age >60 years against malignancy (Table 3).

Table 4 shows a cross tabulation of US risk by provider, Bethesda cytology category, and distribution of malignant versus benign nodules. Out of a total of 47 patients in US risk class 1 and Bethesda III category, 85% confirmed histologically their benign nature, with limited variation based on the provider that reported US. It can also be observed that provider 1 reports a lower percentage of malignancies than benign lesions in US risk classes 2 and especially 3. Conversely, in these US risk classes, providers 2 and 3 have higher rates of malignancy than benign nodules.

Discussion

Ultrasound criteria have long oriented thyroid nodules toward surgery or follow-up. The introduction of score systems dedicated to thyroid nodules has allowed a more precise selection between lesions for simple follow-up and nodules requiring FNAB. Only on the basis of the latter investigation, thyroid nodules could be selected for surgery. This research aimed to find the best method and the best provider for reporting thyroid US from the perspective of selecting thyroid nodules suspected of malignancy. For selecting these nodules, high accuracy is needed. At the same time, unnecessary diagnostics for lesions of little clinical significance might be avoided. Because thyroid nodules are common even in asymptomatic populations, we must consider that cytology should be performed only in nodules considered suspicious on US with the aim of preventing overdiagnosis and overtreatment and reducing costs.[13-15] Several scientific societies have published different US risk scores that have shown the capability to accurately identify benign and malignant nodules.[16-20] A recent systematic review and meta-analysis[21] showed that the Thyroid Imaging, Reporting and Data System proposed by the American College of Radiology (ACR TI-RADS) seems to be the most performant US score system compared with AACE/ACE/AME, ATA, EU-TIRADS, and K-TIRADS. The ACR TI-RADS score system evaluates five categories of US features: composition, echogenicity, shape, margin, and echogenic foci. For each feature, 0-3 points are assigned.[22] This meta-analysis emphasizes that the sensitivity and positive and negative values of ACR TI-RADS are not excellent and a high number of studies evaluated this system.

The AACE/ACE/AME score system was used only in a few studies, but in our practice, its use in three risk categories

Table 3 – Multivariate analysis.

Variable	OR	Inferior limit 95%	Superior limit 95%	P-value
Age (45-60]	0.71	0.43	1.17	0.179946
Age (>60]	0.48	0.28	0.80	<0.01
Male	1.14	0.57	2.31	0.717292
Autoimmunity	0.52	0.34	0.80	<0.01
Bethesda IV	2.53	1.67	3.90	<0.0001
US risk 2	9.81	5.52	18.40	<0.0001
US risk 3	9.94	5.32	19.50	<0.0001
Provider 2	2.74	1.47	5.18	<0.01
Provider 3	2.26	1.42	3.62	<0.001
T1b size	0.41	0.23	0.70	<0.01
T2 size	0.49	0.23	1.03	=0.059104
T3 size	1.98	0.19	17.5	=0.541647

In this logistic regression model, Bethesda IV (*versus* III) is a predictor of malignancy; US risk 2 and 3 (*versus* US risk 1) as well as providers working *in* the dedicated team (endocrinologist or surgeon *versus* external radiologist) are able of diagnosing malignancy more accurately.

Table 4 – Cross tabulation of US risk by provider, Bethesda cytology category, and distribution of malignant versus benign nodules.

US risk	Provider 1				Provider 2				Provider 3			
	Benign		Malignant		Benign		Malignant		Benign		Malignant	
	*B3	*B4	*B3	*B4	*B3	*B4	*B3	*B4	*B3	*B4	*B3	*B4
1	15 (83%)	86 (96%)	3 (17%)	4 (4%)	9 (100%)	9 (82%)	0 (0%)	2 (18%)	16 (80%)	17 (85%)	4 (20%)	3 (15%)
2	22 (59%)	27 (53%)	15 (41%)	24 (47%)	11 (31%)	3 (18%)	24 (69%)	14 (82%)	32 (47%)	11 (19%)	36 (53%)	47 (81%)
3	19 (73%)	24 (44%)	7 (27%)	31 (56%)	0 (0%)	3 (33%)	2 (100%)	6 (66%)	13 (76%)	5 (18%)	4 (24%)	22 (82%)

*B3-B4 = Bethesda 3-Bethesda 4. The percentages reported in each box refer to the total number of patients belonging to each Bethesda category in each US risk class for each provider.

appeared simple to apply, more reliable, and less time consuming than ACR TI-RADS. Sensitivity, specificity, and negative predictive value are performant compared with other risk assessment scales taken into consideration in this review.[21] Another review that compares different risk assessment systems shows that the prevalence of malignancy in high-risk categories is higher in the ACE/AACE/AME system than in the ATA, K-TIRADS, EU-TIRADS, ACR-TIRADS, and BTA systems. At the same time, its prevalence is very low (1%) in the low-risk category.[23] The ACE/AACE/AME US reporting system met the criteria of the reporting system proposed by Su *et al.* (2014), which was concerned with easy routine information workflow, optimization and automation of communication, standardization of documentation, reliability of collected data for comparison, and clinical research.[24]

Concerning the US provider, the importance of surgeon-performed thyroid ultrasonography is reported by several studies who reported its effectiveness in characterizing suspected nodules as well as malignant lymph nodes.[25-27] A study by Mohanapriya[28] showed an improvement in the sensitivity of surgeons compared with radiology-reported US from 86.05% to 98.53%. The negative predictive value increased from 83.8% to 98%. Moreover, it is argued that surgeons performing US would be more advantageous, such as for the detection and evaluation of surgery contemporaneous parathyroid diseases[29] or suspicious nonrecurrent inferior laryngeal nerve.[30]

Our study showed the significant importance of evaluating thyroid nodules, not only taking into consideration the results of cytology but also integrating US reports in the overall judgment of lesions explored. In fact, US risk 2 and 3 classes increased the risk of cancer of undetermined (Bethesda III and IV) nodules nearly 10-fold. On the contrary, out of a total of 47 Bethesda III nodules classified in US risk class 1, only 7 (less than 15%) were malignant. In our opinion, therefore, as supported by these data, in this last context, a simple follow-up is enough. The other side of the coin is that US "low-risk" nodules were less than 30%, so over two-thirds of the sample is considered quite suspicious in accordance with our method because the difference between US risk 2 and 3 was slightly in favor of US risk 3 but not well defined. Notwithstanding, one undetermined nodule out of three is "low risk." Then, a simple follow-up is needed, especially for Bethesda III nodules.

Furthermore, the present study clarified the difference among reports coming from different providers. Endocrinologists or endocrine surgeons on the team responsible for patient management performed better in diagnosing thyroid cancer than radiologists. The role of surgeons (and/or endocrinologists) who performed US has recently been emphasized,[31] and advantages were found in predicting the benignity of nodules[32] as well as detecting malignancies.[33]

Some studies affirm that imaging performed by radiologists appears inadequate or incomplete.[34] This should be due to the experience of a dedicated sonographer that follows a standardization of reports and could obtain continuous feedback from clinical to ultrasound findings. On the other hand, the variations in ultrasound assessment of thyroid lesions have been demonstrated from an interobserver and intraobserver perspective.[35,36]

Although the available literature emphasizes the role of surgeon-performed US, our study showed that both surgeons and endocrinologists are able to improve the performance of US, provided that it is evaluated in the context of a dedicated team. Thus, we could suppose that additional clinical information available to dedicated teams appears crucial for adequate US reports.

A setting capable of overcoming the lack of clinical knowledge should be a dedicated radiologist integrated into the multidisciplinary team, but it is not standardized.[37]

The optimal scenario for the management of thyroid nodules as well as other neck endocrine diseases appears to be a multidisciplinary setting in which essential elements deemed to be important in each specialty of the team are well defined.[24]

The two main questions asked in the present study had clear answers. The ACE/AACE/AME reporting system appeared adequate in the characterization of thyroid nodules, and as deduced from the literature data, no substantial weakness was described.[8,16-22] In fact, both univariate and multivariate analyses showed that US risk 2 and 3 were strongly correlated with malignancy.

Concerning which provider should perform US of thyroid nodules referred for surgery, a member of the dedicated team (endocrinologist and endocrine surgeon) was much better at predicting malignancy. It do not seem related to specific points of US reporting, such as capsular irregularity, microcalcifications *versus* colloid crystals but rather to the overall "sense" of the entire image although this specific issue was not specifically evaluated.

The present study included only patients already referred to surgery, therefore with histologic confirmation. It excludes that some patients enrolled can hide a malignancy. This

advantage is counterbalanced by a selection bias leading to missing data, because a large number of patients are excluded before recruitment to a dedicated team; at the same time, our hospital is a referral center for thyroid surgery, so it can lead to a selection bias once again because previous US, performed in a community-based context, may have excluded some "low-risk" nodules. Probably this strongly affects the non-significance of the difference between the Bethesda III and IV categories in univariate analysis in their respective association with malignancy. In our practice, all Bethesda IV nodules, compared with 70% of Bethesda III nodules, have been taken into account for the surgery. This could lead to an additional selection bias which in turn could also explain, at least in part, why the risk of malignancy in our patients tends to be very similar in the ultrasound risk categories 2 and 3. In addition, it could be assumed that patients in the Bethesda III cytopathologic category, not all of whom have undergone surgery, were preselected as per some risk factors generally accepted by the scientific community (size, age of the patient, and increase in nodule volume) among which the ultrasound risk score may have played a predominant role. It is evident that among these nodules only 47 had a low risk score compared with 182 intermediate or high-risk score.

More weaknesses of the present study might be pointed out: the scarce number of nodules of greater size (>4 cm) included, the heterogeneity of echographs and echographers, the fact that endocrinologists and surgeons were not blinded to the results of the radiology performed ultrasound and report.

Conversely, the prospective nature of the study, the relatively large number of patients enrolled, belonging to a well-defined type (undetermined nodules), and the unequivocal characterization of outcomes (benign/malignant) due just to its biopsy are relevant strengths.

The fact that the nodules addressed to surgery as "suspect" by providers 2 and 3 was in a much higher percentage malignant rather than benign (see Table 2), despite being burdened by the bias previously discussed supports the hypothesis that the US investigations carried out by these operators are much more performant in the detection of malignant lesions. The results of cross tabulation of US risk by provider, Bethesda cytology category and distribution of malignant *versus* benign nodules (Table 4) confirm that the ultrasound reports made by providers belonging to the dedicated team are much more reliable than the former.

Moreover, in addition to the improved diagnostic performance of multidisciplinary team-based management of thyroid nodules from a surgical perspective,[38] patient satisfaction, and cost savings should be taken into consideration due to the reduced number of visits.[39]

The present study investigated the relationship between the pretest probability of malignancy (relatively high in Bethesda III and higher in Bethesda IV cytology) and the performance of a prognostic test (ultrasound stratification).

The results of the study may not appear to be in line with those of the individual risk categories, both in terms of diagnostic cytopathology and US.[8,9] In this regard, it should be underlined once again that this is a study conducted on patients already directed to surgery, so it is to be expected a relatively higher number of malignancies, which, however, are concentrated in the more advanced risk classes (Bethesda IV *versus* III, US risk score 2 and 3 *versus* 1). This does not affect the validity of the respective score systems (Bethesda, ACE/AACE/AME) in which the risk of malignancy is assessed on studies conducted in patients before surgery was performed.

The novelty of these findings lies in the possibility of selecting in the most precise way possible and with simple and routine diagnostic tests the nodules to be addressed to surgery from those to be subjected to observation.

In conclusion, US performed within a dedicated team and reported with a simple, reliable, and efficacious scoring system achieves optimal results in terms of detecting malignancies, avoiding unnecessary procedures or restricting surgical aggressiveness.

We can affirm that in the context of indeterminate nodules (Bethesda III and IV), the ACE/AACE/AME US risk stratification system strongly improved the results of cytology alone. We are convinced that this reporting system assures the communication of minimum essential information among the members of the team to avoid underestimating malignancies or overrating benign nodules, although a comparison ACE/AACE/AME US risk scores with TI-RADS scores (assuming this as the reference score system) not performed in the present study, is needed to confirm the value and reliability of the first one.

From this point of view, a surgeon-performed thyroid US can lead to several advantages over those shown within the limits of the present study, such as a detailed overview of neck anatomy and a meticulous report of neck lymph nodes.

US can be repeated many times, even exchanging sensations and suggestions between surgeons and endocrinologists in real time, increasing the performance of US once again. It is a tool that cannot be given up in the diagnosis and staging of thyroid nodules, and we are resolutely convinced of its value.

Disclosure

The authors report no proprietary or commercial interest in any product mentioned or concept discussed in this article.

REFERENCES

1. Ianni F, Campanella P, Rota CA, et al. A meta-analysis-derived proposal for a clinical, ultrasonographic, and cytological scoring system to evaluate thyroid nodules: the "CUT" score. *Endocrine*. 2016;52:313–321.
2. Ha EJ, Baek JH, Na DG. Risk stratification of thyroid nodules on ultrasonography: current status and perspectives. *Thyroid*. 2017. https://doi.org/10.1089/thy.2016.0654.
3. 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.Thyr J*. 2017;6:225–237.
4. Brito JP, Gionfriddo MR, Al Nofal A, et al. The accuracy of thyroid nodule ultrasound to predict thyroid cancer: systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2014;99:1253–1263.

5. Park JY, Lee HJ, Jang HW, et al. A proposal for a thyroid imaging reporting and data system for ultrasound features of thyroid carcinoma. *Thyroid*. 2009;19:1257–1264.
6. Shin JH, Baek JH, Chung J, et al. Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean society of thyroid radiology consensus statement and recommendations. *Korean J Radiol*. 2016;17:370–395.
7. Hong MJ, Na DG, Baek JH, Sung JY, Kim JH. Cytology-ultrasonography risk-stratification scoring system based on fine-needle aspiration cytology and the Korean-thyroid imaging reporting and data system. *Thyroid*. 2017;27:953–959.
8. Gharib H, Papini E, Garber JR, et al. American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi Medical guidelines for clinical practice for the diagnosis and management OF thyroid nodules-2016 update. *Endocr Pract*. 2016;22:622–639. AACE/ACE/AME task force on thyroid nodules.
9. Cibas ES, Ali SZ. The Bethesda system for reporting thyroid cytopathology. *Thyroid*. 2009;19:1159–1165.
10. Baek HJ, Kim DW, Ryu JH. Association between TNM staging system and histopathological features in patients with papillary thyroid carcinoma. *Endocrine*. 2015;48:589–594.
11. Chereau N, Trésallet C, Noullet S, et al. Does the T1 subdivision correlate with the risk of recurrence of papillary thyroid cancer? *Langenbecks Arch Surg*. 2016;401:223–230.
12. Lindeman BM, Nehs MA, Angell TE, et al. Effect of noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP) on malignancy rates in thyroid nodules: how to counsel patients on extent of surgery. *Ann Surg Oncol*. 2019;26:93–97.
13. Haugen BR, Alexander EK, Bible KC, 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.
14. 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.
15. Vaccarella S, Franceschi S, Bray F, Wild CP, Plummer M, Dal Maso L. Worldwide thyroid-cancer epidemic? The increasing Impact of overdiagnosis. *N Engl J Med*. 2016;375:614–617.
16. Perros P, Colley S, Boelaert K, et al. British Thyroid Association Guidelines for the management of thyroid cancer. *Clin Endocrinol (Oxford)*. 2014;81(Suppl 1):1–122.
17. Horvath E, Majlis S, Rossi R, et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. *J Clin Endocrinol Metab*. 2009;94:1748–1751.
18. Ha EJ, Moon WJ, Na DG, et al. A multicenter prospective validation study for the Korean thyroid imaging reporting and data system in patients with thyroid nodules. *Korean J Radiol*. 2016;17:811–821.
19. Lee YH, Kim DW, In HS, et al. Differentiation between benign and malignant solid using the ATA and ACR-TIRADS sonographic classifications as adjunctive predictors of malignancy for indeterminate thyroid nodules. *Endocr Pract*. 2011;25:908–917.
20. Hamadi S, Herbst R, Oyekunle T, Jiang XS, Strickland K, Sanzi R, Sosa JA. Using the ATA and ACR-TIRADS sonographic classifications as adjunctive predictors of malignancy for indeterminate thyroid nodules. *Endocr Pract*. 2019;25(9):908–917.
21. Castellana M, Castellana C, Treglia G, et al. Performance of five ultrasound risk stratification systems in selecting thyroid nodules for FNA. A META-ANALYSIS. *J Clin Endocrinol Metab* 2019. https://doi.org/10.1210/clinem/dgz170.
22. Tessler FN, Middleton WD, Grant EG. Thyroid imaging reporting and data system (TI-RADS): a User's Guide. *Radiology*. 2018;287:29–37.
23. Ospina NS, Ariza NMI, Castro MR. Thyroid nodules: diagnostic evaluation based on thyroid cancer risk assessment. *BMJ*. 2020;368:l6670. https://doi.org/10.1136/bmj.16670.
24. Su HK, Dos Reis LL, Lupo MA, et al. Striving toward standardization of reporting of ultrasound features of thyroid nodules and lymph nodes: a multidisciplinary consensus statement. *Thyroid*. 2014;24:1341–1349.
25. Hamer PW, Aspinall SR, Malycha PL. Clinician-performed ultrasound in assessing potentially malignant thyroid nodules. *ANZ J Surg*. 2014;84:376–379.
26. Monteiro R, Han A, Etiwy M, et al. Importance of surgeon-performed ultrasound in the preoperative nodal assessment of patients with potential thyroid malignancy. *Surgery*. 2018;163:112–117.
27. Cohen O, Lahav Y, Halperin D, Yehuda M. Surgeon-performed ultrasonographic evaluation and predication for large thyroid nodules-A case-control study. *Surgery*. 2019;166:1148–1153.
28. Mohanapriya G, Chandrasekaran M. Is a surgeon-performed ultrasound good enough in diagnosing thyroid malignancy? *Indian J Endocr Metab*. 2018;22:181–184.
29. Sloan DA, Davenport DL, Eldridge RJ, Lee CY. Surgeon-driven thyroid Interrogation of patients presenting with primary hyperparathyroidism. *J Am Coll Surg*. 2014;218:674–683.
30. Citton M, Viel G, Iacobone M. Neck ultrasonography for detection of non-recurrent laryngeal nerve. *Gland Surg*. 2016;5:583–590.
31. Méndez W, Rodgers SE, Lew JI, Montano R, Solorzano CC. Role of surgeon-performed ultrasound in predicting malignancy in patients with indeterminate thyroid nodules. *Ann Surg Oncol*. 2008;15:2487–2492.
32. Goldfarb M, Gondek S, Solorzano C, Lew JI. Surgeon-performed ultrasound can predict benignity in thyroid nodules. *Surgery*. 2011;150:436–441.
33. Jabiev AA, Ikeda MH, Reis IM, Solorzano CC, Lew JI. Surgeon-performed ultrasound can predict differentiated thyroid cancer in patients with solitary thyroid nodules. *Ann Surg Oncol*. 2009;16:3140–3145.
34. Kumbhar SS, O'Malley RG, Robinson TF, et al. Why thyroid surgeons are frustrated with radiologists: lessons learned from pre- and postoperative US. *Radiographics RSNA*. 2016;36:2141–2153.
35. Choi SH1, Kim EK, Kwak JY, Kim MJ, Son EJ. Interobserver and intraobserver variations in ultrasound assessment of thyroid nodules. *Thyroid*. 2010;20:167–172.
36. Oltmann SC, Schneider DF, Chen H, Sippel RF. All thyroid ultrasound evaluations are not equal: sonographers specialized in thyroid cancer correctly label clinical N0 disease in well differentiated thyroid cancer. *Ann Surg Oncol*. 2015;22:422–428.
37. Carneiro-Pla D, Amin S. Comparison between preconsultation ultra- sonography and office surgeon-performed ultrasound in patients with thyroid cancer. *World J Surg*. 2014;38:622–627.
38. Hoang JK, Sosa JA, Nguyen XV, Galvin PL, Oldan JD. Imaging thyroid disease: updates, imaging approach, and management pearls. *Radiol Clin North Am*. 2015;53:145–161.
39. Mazzaglia PJ. Surgeon-Performed Ultrasound in patients referred for thyroid disease improves patient care by minimizing performance of unnecessary procedures and optimizing surgical treatment. *World J Surg*. 2010;34:1164–1170.