Surgery 166 (2019) 1148–1153

Contents lists available at ScienceDirect

Surgery

journal homepage: www.elsevier.com/locate/surg

Endocrine

Surgeon-Performed Ultrasonographic Evaluation and Predication for Large Thyroid Nodules—A Case-Control Study

Oded Cohen, MD[a,*], Yonatan Lahav, MD[a], Doron Halperin, MD, MPH[a], Moshe Yehuda, MD[a,b]

[a] Department of Otolaryngology Head and Neck Surgery, Kaplan Medical Center, Hadassah School of Medicine, Hebrew University in Jerusalem, Rehovot, Israel
[b] Department of Otolaryngology Head and Neck Surgery and Maxillofacial Surgery, Tel-Aviv Sourasky Medical Center, Sackler School of Medicine, Tel-Aviv University, Tel-Aviv, Israel

ARTICLE INFO

Article history:
Accepted 10 June 2019
Available online 31 July 2019

ABSTRACT

Background: The management of large thyroid nodules remains unclear. Ultrasonography is a central tool in the assessment of thyroid nodules, yet its role in risk stratification of large thyroid nodules has been studied only seldomly.
Objective: The aim of this study was to determine the utility of ultrasonography in characterizing and risk-stratifying thyroid nodules ≥3.0 cm.
Methods: We performed a retrospective, case-control study of all thyroid nodules aspirated between January 2010 and May 2014. Sonographic features of nodules ≥3.0 cm were compared with nodules <3.0 cm. All nodules were assessed by a single high-volume thyroid surgeon. Data collected included size (cm), texture, echogenicity, shape, calcifications, border, spongiform appearance, and vascularity.
Results: A total of 537 nodules were included in the study, with 137 nodules (25.5%) ≥3.0 cm comprising the study group, and 400 nodules (74.5%) as the control group. No differences were found between the 2 groups regarding age, sex, and risk factors. Nodules ≥3.0 cm were associated with an increased risk for malignancy (odds ratio 2.41 [1.08–5.38]). Microcalcifications (26.3% vs 17.5%, $P = .039$), hypoechogenicity (40.8% vs 23.4%, $P < .001$), and irregular borders (14.3% vs 3.6%, $P = .001$) were more prevalent in nodules <3.0 cm. Among the large nodules, hypoechogenicity (50% vs 22.8%, $P = .043$) and irregular margins (28.6% vs 0%, $P < .001$) were associated with malignancy. The specificity of irregular borders was greater in nodules ≥3.0 cm (100% vs 89.6%, $P = .011$).
Conclusion: Sonographic appearances of large nodules differ from those of smaller nodules. Hypoechogenicity and irregular borders were associated with malignancy and should direct management towards a more aggressive policy of fine needle aspiration and consideration of operative resections in large thyroid nodules.

Introduction

The management of large thyroid nodules (≥3.0 cm) has yet to be determined. The literature is divided between those who recommend resection regardless of preoperative workup (mostly sonographic description and fine needle aspiration [FNA] results),[1,2] and others who find preoperative assessment reliable and recommend that management should not differ from smaller nodules.[3,4] The controversy lies mostly in the reliability of benign aspirations of these nodular subgroups. Numerous publications have found low false-negative rates for benign aspirations ranging between 0.7% and 7%,[2–5] whereas others have found rates to be substantially greater.[6,7]

Thyroid ultrasonography (US) has become a pivotal decision-making tool in the workup and management of thyroid nodules, and subsequently, US has been incorporated into different thyroid guidelines.[8,9] Recent meta-analyses on US performance found that although no individual US feature is capable of predicting malignancy with acceptable diagnostic accuracy, a combination of known features on US can identify nodules with an increased risk for malignancy.[10,11] Yet, despite its important role and numerous

* Reprint requests: Oded Cohen, MD, Department of Otolaryngology Head and Neck Surgery, Kaplan Medical Center, POB 1 Rehovot 76100, Israel.
E-mail address: odedco@clalit.org.il (O. Cohen).

https://doi.org/10.1016/j.surg.2019.06.011
0039-6060/

publications, only a few studies have focused on the ability of sonographic features to stratify the risk of malignancy specifically in large thyroid nodules[6,12,13] and in 2 of those studies, US had a very limited role in differentiating malignancy versus no malignancy in large thyroid nodules.[6,13]

The main pitfall of thyroid US and its ability to predict malignancy is the interobserver variability, which is poor to moderate for most features.[14,15] This variability undermines the more generalized reproducibility across institutions. Surgeon-performed US has been recognized in several publications as an important and reliable tool,[16,17] which may be superior to community-based US[18,19] and may change the course of management for patients.[17,18,20] When conducted by a single high-volume surgeon, it can eliminate the pitfalls of interobserver variability as well.

Therefore, the aim of this current study was to have the same surgeon (M.Y.) evaluate via US a cohort of large nodules (≥3.0 cm) compared to nodules <3.0 cm. To the best of our knowledge, there is no case-control study on thyroid US conducted by the same operator, and this is the first study directly comparing the sonographic differences between nodules ≥3.0 and <3.0 cm.

Patients and methods

The study was approved by our institutional review board. Inclusion criteria for the study were adult patients who underwent a surgeon-performed US and a surgeon-performed US-guided FNA of thyroid nodules at our dedicated thyroid outpatient clinic between January 2010 and May 2014. All thyroid biopsies were performed by the same high-volume head and neck surgeon (M.Y.) according to the guidelines of the American Thyroid Association (ATA).[21] Exclusion criteria were aspired nodules that did not comply with ATA guidelines, therapeutic aspirations for cystic nodules (>95% cystic component, and non-suspicious features of the solid component), and nodules with missing data that prevented data analysis.

In addition to medical histories and physical examinations, all clinic visits included a surgeon-performed US and surgeon-performed, US-guided FNA when indicated. Sonographic description of the nodule based on known features (detailed below) was documented on a designated form. All medical records were reviewed, and data on selected demographics (age, sex, history of any other malignancy) and risk factors for thyroid cancer (previous exposure to ionizing radiation during childhood or adolescence and a family history of thyroid cancer) were extracted. The following data of sonographic features were collected and recorded for each lesion: size (mm), solid texture (defined as cystic component <50%), echogenicity (defined by relation to the strap muscles: hypo-, iso-, hyper-, or mixed), shape (taller than wide in shape [anteroposterior dimension > transverse dimension on an axial image]), calcifications (microcalcifications were defined as ≤2 mm in diameter, tiny, punctuated hyperechoic foci; other calcifications were defined as macrocalcifications), margins (regular vs irregular), and spongiform appearance (similarly sized microcystic spaces separated by thin echogenic septa). Vascularity (intranodular vs perinodular) was assessed using a Doppler US. Malignant associated features (MAFs) according to the 2015 ATA guidelines were defined as a solid texture, marked hypoechogenicity, microlobulated or irregular borders, and lesions that were taller than wide in shape.[8] The score of the thyroid imaging reporting and data system (TI-RADS) was calculated based on the report from the American College Radiology[22] (Table I). As a rule, in patients with multinodular thyroids, the most suspicious nodule in each lobe was selected for FNA. For patients with more than 1 nodule, each nodule was analyzed separately. If such patients were operated on, the correlation between the descriptions of the surgeon-performed US and the final pathology was based on the side of the resection. Cases were excluded when both nodules were on the same side and a correlation with the final pathology was relevant yet impossible.

Table I
ACR TI-RADS for US features of nodules[22]

Category	Description	Points
Composition	Cystic/almost, completely cystic/spongiform	0
	Mixed cystic + solid	1
	Solid or almost/completely solid	2
Echogenicity	Anechoic	0
	Hyperechoic or isoechoic	1
	Hypoechoic	2
	Very hypoechoic	3
Shape	Wider than tall	0
	Taller than wide	3
Margins	Smooth/ill-defined	0
	Lobulated or irregular	2
	Extrathyroidal	3
Echogenic foci	None or large comet-tail artifacts	0
	Macrocalcifications	1
	Peripheral (rim) calcifications	2
	Punctate echogenic foci	3

TI-RADS level (sum of all total points above)

Total points	TI-RADS level
0	TR1
2	TR2
3	TR3
4-6	TR4
7+	TR5

All USs were performed with a high-resolution, portable, 8- to 12-MHz system (SonoSite, M-Turbo), with 35-mm linear array transducer (SonoSite Inc, Bothell WA). All US-guided FNA biopsies were performed with an on-site evaluation of adequacy by a thyroid cytopathologist. Cytologic diagnostic criteria were based on the Bethesda system for reporting thyroid cytopathology (BSRTC).[23]

Decisions to refer a nodule for operative resection were based on the clinical judgment of the senior author (M.Y.) after considering all aspects of the nodule, including risk factors, suspicious features in the physical examination, its sonographic appearance, and most importantly, its cytology. Final histopathologic results were obtained after the operation. Non-benign cytology was defined as atypia of undetermined significance or follicular lesion with uncertain significance, follicular neoplasm, suspicious for malignancy, and malignant. For subgroup statistical analysis, we defined a "combined benign" group composed of patients with initial and repeated benign cytology (who were followed up with but not operated on) grouped together with patients operated on with benign histology. This combined group was compared with patients with malignant histologies. Incidental microcarcinomas were defined as nodules with a tumor <1.0 cm without histologic lymph node metastasis or extrathyroidal extension and were excluded from the statistical analysis for the purpose of malignancy.[8]

Categorical variables are reported as numbers (percentages), and continuous variables as medians and interquartile ranges. Continuous variables were tested for normal distribution using histograms and Q-Q plots. Categorical variables were compared using a chi-square test or Fisher exact test and continuous variables by a Mann Whitney test. Univariate and multivariate Cox regressions were performed to evaluate the crude and adjusted associations between various predictors and readmission. Age, sex, and variables with P values < .2 on the univariate analysis were included in the multivariate analysis. A 2-tailed P < .05 was considered statistically significant. Analyses were performed with SPSS version 22 (IBM Corp, Armonk, NY).

Table II
Multivariate analysis for factors associated with malignancy

Category	*P* value	OR (95% CI)
Sex	.256	0.63 (0.28–1.41)
Age	.218	0.99 (0.97–1.01)
Solid	<.001	3.98 (1.91–8.28)
Hypoechogenicity	.003	2.81 (1.41–5.61)
Irregular margins	<.001	5.59 (2.58–12.11)
Intranodular vascularity	.086	1.96 (0.91–4.24)
Macrocalcifications	.221	1.87 (0.69–5.13)
Microcalcifications	.034	2.17 (1.06–4.44)
Size ≥3.0 cm	.032	2.41 (1.08–5.38)

OR, odds ratio.

Results

A total of 537 nodules in 449 patients were included in the study; 137 nodules (25.5%) ≥3.0 cm in 107 patients comprised the study group, and 400 nodules (74.5%) in 342 patients comprised the control group. No differences ($P > .32$ each) were found between the study and control groups regarding age (57 [±18] vs 57 [±15] years), sex (males 19.6% vs 19.8%), and risk factors (13 % vs 13.6%).

First, a univariate analysis (Supplemental Table I) followed by a multivariate analysis (Table II) was conducted to evaluate the association of size and other sonographic features with malignancy within the study cohort ($n = 537$). A size ≥3 cm was found to be associated with malignancy independently of sonographic features (odds ratio 2.41 [1.08–5.38]; *P* value = .032). Except for taller than wide shape, all other MAFs were found to be independently associated with malignancy (Table II).

Comparisons of sonographic features between the groups are presented in Table III. The MAFs were statistically significantly more prevalent in nodules <3.0 cm. These sonographic features included microcalcifications (26.3% vs 17.5%, $P = .039$), hypoechogenicity (40.8% vs 23.4%, $P < .001$), and irregular borders (14.3% vs 3.6%, $P = .001$). Large nodules presented with greater rates of benign features, including cystic or mixed texture ($P < .007$), and greater rates of isoechogenicity ($P = .029$) and perinodular vascularity ($P = .015$). The distribution of the TI-RADS scores also differed between the groups ($P = .016$) with a greater prevalence of a TI-RADS score of ≥3 in the control group (14.8% vs 5.1%, respectively). Distribution of BSRTC after the initial FNA is presented in Figure 1. No differences were found in the BSRTC distribution between the groups ($P = 0.681$). The rate of malignancy within each BSRTC category is presented in Supplemental Table II.

Sixty-one nodules (44.5% of the large nodule cohort) were resected, of which 14 (22.9%) were found to be malignant. When integrating nodules with consecutive benign aspirations (the combined benign group), the rate of malignancy was 20% (14 of 69) among the large nodule groups compared with 13.7% (44 of 322) in the smaller nodules ($P = .223$). The distribution of malignant pathologies was as follows: papillary thyroid carcinoma (58%), follicular carcinoma (14%), anaplastic carcinoma (14%), poorly differentiated carcinoma (7%), and unknown (7%). A correlation of sonographic features with a malignant pathology in nodules ≥3.0 cm is presented in Table IV. Hypoechogenicity (50% vs 22.8%), irregular margins (28.6% vs 0%), and overall TI-RADS score were associated with malignancy ($P < .05$ each). Irregular margins had a 100% specificity and positive predictive value, whereas hyperechogenicity had the greatest negative predictive value of 86.3%. Intranodular vascularity was also associated with malignant nodules (50% vs 17.9%, $P = .012$). The odds ratio was 3.39 (CI 95% 1.01-11.43) for hypoechogenicity and 4.60 (CI 95% 1.32-16.07) for intranodular vascularity. Correlations of sonographic features with malignant pathologies in nodules <3.0 cm is presented in Supplemental Table III. Except for a taller than wide shape, all MAFs were associated with malignancy with $P < .001$ for each MAF and an overall TI-RADS score. Intranodular vascularity was not found to be associated with malignancy.

Next, we found that the sensitivity of surgeon-performed US between malignant nodules ≥3.0 cm and nodules <3.0 cm was no different in any of the MAFs in the 2015 ATA guidelines (Table V). The parameter of intranodular vascularity was more sensitive for detecting malignancy in nodules ≥3.0 cm (50% vs 20.5%, $P = .031$), whereas the specificity of irregular borders was greater in nodules ≥3.0 cm (100% vs 89.6%, $P = .011$).

Discussion

The aim of this current study was to compare a single operator's sonographic evaluation of large thyroid nodules (≥3 cm) compared with smaller nodules (<3 cm). To the best of our knowledge, we are the first to conduct such a study that eliminates the main drawback of US, that is, interobserver viability.[14] We found that sonographic features differ between the 2 groups. Large nodules had less

Table III
Distribution of sonographic features between the groups

Sonographic features		Control group (size <3.0 cm) n = 400 (%)	Study group (size ≥3.0 cm) n =137 (%)	Total n = 537 (%)	*P* value
2015 ATA MAF	Solid texture	194 (48.7)	68 (50.4)	262 (49.2)	.744
	Microcalcifications	105 (26.3)	24 (17.5)	129 (24.0)	.039
	Hypoechogenicity	163 (40.8)	32 (23.4)	195 (36.3)	<.001
	Irregular borders	57 (14.3)	5 (3.6)	62 (11.6)	.001
	Taller than wide	11 (2.8)	1 (0.7)	12 (2.2)	.166
Other features	Cystic texture	38 (9.5)	25 (18.2)	67 (11.8)	.016
	Mixed texture	124 (31.1)	63 (46)	187 (34.9)	.002
	Isoechogenicity	131 (32.8)	59 (43.1)	190 (35.4)	.029
	Hyperechogenicity	19 (4.8)	8 (5.8)	27 (5)	.615
	Intranodular vascularity	56 (14.0)	27 (19.9)	83 (15.5)	.106
	Perinodular vascularity	22 (5.5)	16 (11.8)	38 (7.1)	.015
	Macrocalcifications	46 (11.5)	11 (8.1)	57 (10.6)	.265
	Comet-tail sign	13 (3.3)	5 (3.6)	18 (3.4)	.823
	Spongiform appearance	20 (5.0)	7 (5.1)	27 (5.0)	.960
TI-RADS Score*	1	16 (5)	5 (7.2)	21 (5.4)	<.128
	2	134 (41.6)	25 (36.2)	159 (40.6)	
	3	97 (30.1)	30 (43.5)	127 (32.5)	
	4	75 (23.3)	9 (13)	84 (21.5)	

* Thyroid Imaging, Reporting and Data System.[22]

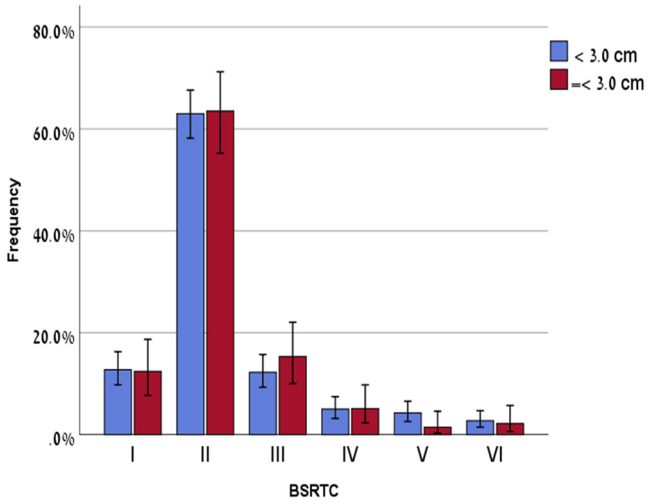

Fig 1. BSRTC: Bethesda System for Reporting Thyroid Cytopathology.[27] I, non-diagnostic; II, benign; III, atypia of undetermined significance/follicular lesion with uncertain significance; IV, follicular neoplasm; V, suspicious for malignancy; VI, malignant.

suspicious features compared with smaller nodules. Despite having only a small cohort of malignant large nodules (n = 14), which limits the impact of this current study, some important differences were found because some features were statistically significantly associated with malignancy (hypoechogenicity and irregular margins). The sonographic appearances of large nodule MAFs was poorer compared with those of smaller nodules, yet a direct comparison of sensitivity did not show any differences in sensitivity of known MAFs. Finally, a multivariate analysis of the cohort demonstrated that large thyroid nodules carry greater risks for malignancy in contrast to their generally benign appearance.

Ultrasonography is an important tool in the initial assessment and follow-up of thyroid nodules.[8] Ultrasonography together with FNA comprise the cornerstone of risk stratification for thyroid nodules. Although algorithms may change from guideline to guideline, the basic principle remains that the indications for FNA are based on the combination of the size of the nodule and the number of its suspicious features.[8,9,22] Yet, when the nodule is considered large (defined as nodules ≥3.0 cm in most studies[2,3,12,24,25]), the role of US and the reliability of a benign FNA result have been questioned, and the management of large thyroid nodules remains debated. This uncertainty is echoed in the recent guidelines of the ATA, which concluded that based on the evidence, it is still unclear if patients with large thyroid nodules should be managed differently than those with smaller nodules. Whether large nodules carry a greater risk for malignancy compared with smaller nodules has also been debated in the literature.[4,26] The results of our study support the finding of a recent systematic review, which found a greater rate of malignancy in nodules >3.0 cm.[26] Although there are many studies on the credibility of FNA results in large thyroid nodules,[2–5,25] only a few studies focused on the credibility of sonographic characterizations of large nodules.[6,12,13]

Wharry et al evaluated 382 nodules ≥4.0 cm.[6] They found that none of the MAFs were associated with malignancy, neither differentiated or undifferentiated thyroid cancer. Yet, the authors did not address the source of the US report (community/radiologist or surgeon-performed), which exposes the data for potential known biases because the quality of different US sources has been shown to differ considerably.[17,18,20] Conversely, Yoon et al reported their group's experience with 661 thyroid nodules ≥3.0 cm, all evaluated retrospectively by a single blinded observer. They found that all MAFs were associated with malignancy (all *P* values < .001), and a benign appearance was associated with benign nodules.[12] Another article by the same group assessed the ability of the TI-RADS score to predict false-negative cytology for thyroid nodules ≥3.0 cm (different cohorts)[13] as evaluated by 10 radiologists. Interestingly, in this cohort, only echogenicity and a solid composition were found to be different in benign and malignant large nodules. This finding is consistent with our findings in which echogenicity (together with irregular borders) were the only MAFs associated with malignancy. In summary, the only report from a surgical group before this publication that questioned the role of US in large thyroid nodule assessments found US to be reliable in 2 large-scale studies. Nevertheless, although the group of Yoon et al is world-renowned in the field of thyroid imaging with numerous important publications including the TI-RADS score,[27] the question of reproducibility and generalization of their results is important for everyday practice. In our opinion, the conflicting results of the aforementioned studies may not necessarily reflect the credibility of US as an imaging modality in evaluating large thyroid nodules but rather stresses the importance of operator experience in such evaluations.

Since the initial reports, many publications have reported surgeon-performed US as a reliable tool for assessing thyroid nodules,[16,17,20] which may even change patient management compared with community and radiologist-performed US.[17,18,20] Surgeon-performed US can allow for a direct, single-operator assessment of thyroid nodules, which in turn increases the

Table IV
Association of US features with malignancy in nodules ≥3.0 cm*

		Benign† (n = 65)	Malignant (n = 14)	*P* value	Sen. (%)	Spe. (%)	PPV (%)	NPV (%)
2015	Solid	29 (53)	9 (64)	.438	64	47.3	24	84
ATA	Microcalcifications	8 (14)	4 (29)	.193	29	86.0	33	83
MAF	Hypoechogenicity	13 (233)	7 (50)	.043	50	77.2	35	86
	Irregular margins	0	4 (29)	<.001	289	100	100	85
	Taller > wide	1	0	.618	N/A	N/A	N/A	N/A
Intranodular vascularity		10 (18)	7 (50)	.012	50	82	41	87
TI-RAD score‡	1	5 (8.8) 100	0 (0)	.512	N/A			
	2	19 (33.3) 80	7 (50)					
	3	26 (45.6) 82	5 (35.7)					
	4	7 (12.3) 72	2 (14.3)					

Acc, accuracy; *PPV*, positive predictive value; *Sen.*, sensitivity; *Spe.*, specificity.
* Values are presented in *N* (%) unless written otherwise.
† The benign group is combined from benign pathology (*n* = 47) and patients with 2 consecutive benign cytologies (*n* = 18).
‡ Thyroid Imaging, Reporting and Data System[22] (*n* = 47).

Table V
Comparison of the prevalence of US features between malignant nodules ≥3.0 cm and malignant nodules <3.0 cm*

		Nodules ≥3.0 cm (n = 14)	Nodules <3.0 cm (n = 44)	*P* value
2015	Solid	9 (64)	36 (82)	.171
ATA	Microcalcifications	4 (29)	21 (48)	.207
MAF	Hypoechogenicity	7 (50)	32 (73)	.115
	Irregular margins	4 (29)	17 (40)	.460
	Taller > wide	0	2	.417
Intranodular vascularity		7 (50)	9 (21)	.031
TI-RAD score†	1	0(0)	1(2.3)	
	2	7 (50)	13(29.5)	.229
	3	5 (35.7)	9(20.5)	
	4	2(14.3)	21(47.7)	

The benign group is combined from benign pathology (n = 47) and patients with 2 consecutive benign cytologies (n = 18).
Acc, accuracy; *PPV*, positive predictive value; *Sen*., sensitivity; *Spe*., specificity.
* Values are presented in N (%) unless written otherwise.
† Thyroid Imaging, Reporting and Data System.[22]

consistency of the nodule description by eliminating interobserver variability, which was reported even among experienced radiologists.[14,15] Therefore, our group aimed at evaluating the ability of single surgeon-performed US to evaluate large thyroid nodules. By using a control group of smaller nodules (<3 cm) evaluated by the same high-volume thyroid surgeon, we were able to have direct comparisons between nodules, thereby eliminating many potential biases.

Our study shows some novel information regarding the utility of US in characterizing large nodules (≥3 cm) when compared with smaller nodules. Generally, large nodules seem to have a more benign appearance when compared to smaller nodules. In addition, similar to Wharry et al,[6] our study showed that known suspicious features are better used in smaller nodules when compared to large nodules, even if no difference was found in the sensitivity of each feature when comparing the 2 groups. Conversely, similar to the report by Nam et al,[13] we have shown that some but not all known sonographic features are associated with malignancy. As mentioned, hypoechogenicity was found to be associated with malignancy in both studies, suggesting that hypoechoic large nodules should be suspected to harbor malignancy regardless of the existence of other features. This finding is supported further by the fact that large nodules had greater rates of benign echogenicity (iso/hyper), which in turn make large, hypoechoic nodules suspicious. Each of the ATA MAFs is scored differently in the ACR TI-RADS.[22] Certain features, such as taller than wide and very hypoechoic nodules (Table I), which were prevalent at very low rates in our large nodule cohort, are given higher points. Other features, like hypoechogenicity and irregular borders, which were significantly more prevalent in our malignant large nodules, are given lower scores. Our results show that large thyroid nodules may present with more benign features (Table III). It is possible that the prediction scores of the ACR TI-RADS may not be as accurate for large thyroid nodules. This possibility is supported by the lack of differences in overall ACR TI-RADS scores between benign and malignant large nodules (Table IV), although this suggestion should be regarded with caution given our small sample size of large malignant nodules. Therefore, US can serve as more of a "rule in" to categorize suspicious large thyroid nodules, basing decisions primarily on hypoechogenicity and irregular margins. Yet, when considering their increased risk of malignancy, unlike nodules <3.0 cm, we suggest that benign-appearing large nodules must still be regarded with caution.

Finally, intranodular vascularity has been found to be associated with malignancy and has greater rates of sensitivity compared with nodules <3.0 cm. The predictive value of intranodular vascularity has shown conflicting results in previous studies[16,28,29] and was omitted as an MAF in current guidelines of both the ATA[8] and the American College of Radiologists.[22] Nevertheless, our results suggest that intranodular vascularity has a greater value in assessing malignancy in the subgroup of large nodules ≥3 cm. Future prospective large-scale studies may help in establishing its role in this specific subgroup of nodules. Several important limitations of our study should be discussed. In addition to its retrospective nature, the major limitation of the study is its small sample of malignant thyroid nodules (n = 14). Such a small group exposes the results of the study to underpowered conclusions and questions its generalizability. This, in turn, should drive future larger-scale studies to confront some of the findings of the current study. This same consideration was also reflected in the prevalence of the feature of irregular margins, which was present in 4 of the 14 malignant nodules, thus limiting the clinical impact of the result interpretation. In addition, our study included various types of malignancies, including anaplastic and poorly differentiated carcinomas, which, in turn, influence the sonographic appearance of the nodule. Data on previous FNAs before the visit to our clinic were not collected, which may have influenced the surgeon's decision to refer a nodule for operative excision or follow-up. Finally, the study is exposed to a potential selection bias because our institution is a referral center for thyroid surgery. Some of the nodules may have had previous community-based US, which may have influenced the decision of whether and when to refer these nodules for our evaluation. Another selection bias that can be applied for studies on malignancy rates is the result of risk stratification and selection of suspicious nodules for operative resection. The authors have tried to overcome this limitation by defining a "combined benign" category including nodules with 2 benign aspirations indicating their benign nature with high probability.

We hope that this study leads to larger studies with the constant increase of newly diagnosed thyroid nodules and cancer[30] and the shift toward conservative management. All of this has made US into a crucial tool for decision-making in the management of thyroid nodules. Expanding our knowledge on the utility of US for the various nodule subgroups, such as large nodules and functional nodules, may contribute to better patient care.

In conclusion, large thyroid nodules ≥3 cm differ in their sonographic appearances when compared with nodules <3.0 cm and may carry an increased risk for malignancy. Sonographic evaluation of large nodules should emphasize the hypoechogenicity and irregular shape over other known MAFs. Finally, intranodular vascularity may have a role in risk stratification for large thyroid nodules and should be integrated into the holistic assessment of the malignant potential for these large nodules.

Funding/Support

None.

Conflict of interest/Disclosure

None of the authors have any conflicts of interest to disclose.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.1016/j.surg.2019.06.011.

References

1. McCoy KL, Jabbour N, Ogilvie JB, Ohori NP, Carty SE, Yim JH. The incidence of cancer and rate of false-negative cytology in thyroid nodules greater than or equal to 4 cm in size. *Surgery*. 2007;142:837–844.
2. Raguin T, Schneegans O, Rodier JF, et al. Value of fine-needle aspiration in evaluating large thyroid nodules. *Head Neck*. 2017;39:32–36.
3. Magister MJ, Chaikhoutdinov I, Schaefer E, Williams N, Saunders B, Goldenberg D. Association of thyroid nodule size and Bethesda class with rate of malignant disease. *JAMA Otolaryngol Head Neck Surg*. 2015;141:1089–1095.
4. Kamran SC, Marqusee E, Kim MI, et al. Thyroid nodule size and prediction of cancer. *J Clin Endocrinol Metab*. 2013;98:564–570.
5. Bohacek L, Milas M, Mitchell J, Siperstein A, Berber E. Diagnostic accuracy of surgeon-performed ultrasound-guided fine-needle aspiration of thyroid nodules. *Ann Surg Oncol*. 2012;19:45–51.
6. Wharry LI, McCoy KL, Stang MT, et al. Thyroid nodules (≥4 cm): can ultrasound and cytology reliably exclude cancer? *World J Surg*. 2014;38:614–621.
7. Carrillo JF, Frias-Mendivil M, Ochoa-Carrillo FJ, Ibarra M. Accuracy of fine-needle aspiration biopsy of the thyroid combined with an evaluation of clinical and radiologic factors. *Otolaryngol Head Neck Surg*. 2000;122:917–921.
8. 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.
9. Perros P, Colley S, Boelaert K, et al. Guidelines for the management of thyroid cancer. *Clin Endocrinol (Oxf)*. 2014;81(supplement 1):1–122.
10. Remonti LR, Kramer CK, Leitão CB, Pinto LCF, Gross JL. Thyroid ultrasound features and risk of carcinoma: a systematic review and meta-analysis of observational studies. *Thyroid*. 2015;25:538–550.
11. 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.
12. Yoon JH, Kwak JY, Moon HJ, Kim MJ, Kim E-K. The diagnostic accuracy of ultrasound-guided fine-needle aspiration biopsy and the sonographic differences between benign and malignant thyroid nodules 3 cm or larger. *Thyroid*. 2011;21:993–1000.
13. Nam SJ, Kwak JY, Moon HJ, Yoon JH, Kim EK, Koo JS. Large (≥3cm) thyroid nodules with benign cytology: can Thyroid Imaging Reporting and Data System (TIRADS) help predict false-negative cytology? *PLoS One*. 2017;12:1–14.
14. Choi SH, Kim EK, Kwak JY, Kim MJ, Son EJ. Interobserver and intraobserver variations in ultrasound assessment of thyroid nodules. *Thyroid*. 2010;20:167–172.
15. Park CS, Kim SH, Jung SL, et al. Observer variability in the sonographic evaluation of thyroid nodules. *J Clin Ultrasound*. 2010;38:287–293.
16. Cohen O, Raz Yarkoni T, Lahav Y, Azoulay O, Halperin D, Yehuda M. Surgeon-performed thyroid ultrasound—proving utility and credibility in selecting patients for fine needle aspiration according to the American thyroid association guidelines. A retrospective study of 500 patients. *Clin Otolaryngol*. 2018;43:267–273.
17. 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.
18. Carneiro-Pla D, Amin S. Comparison between preconsultation ultrasonography and office surgeon-performed ultrasound in patients with thyroid cancer. *World J Surg*. 2014;38:622–627.
19. Oltmann SC, Schneider DF, Chen H, Sippel RS. 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*. 2014;22:422–428.
20. 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.
21. American Thyroid Association (ATA) Guidelines Taskforce on Thyroid Nodules and Differentiated Thyroid Cancer, Cooper DS, Doherty GM, et al. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2010;20:674–675.
22. Tessler FN, Middleton WD, Grant EG, et al. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): white paper of the ACR TI-RADS Committee. *J Am Coll Radiol*. 2017;14:587–595.
23. Cibas ES, Ali SZ. The Bethesda System for reporting thyroid cytopathology. *Thyroid*. 2009;19:1159–1165.
24. Porterfield JR, Grant CS, Dean DS, et al. Reliability of benign fine needle aspiration cytology of large thyroid nodules. *Surgery*. 2008;144:963–969.
25. Mehanna R, Murphy M, McCarthy J, et al. False negatives in thyroid cytology: impact of large nodule size and follicular variant of papillary carcinoma. *Laryngoscope*. 2013;123:1305–1309.
26. Shin JJ, Caragacianu D, Randolph GW. Impact of thyroid nodule size on prevalence and post-test probability of malignancy: a systematic review. *Laryngoscope*. 2015;125:263–272.
27. Kwak JY, Han KH, Yoon JH, et al. Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk. *Radiology*. 2011;260:892–899.
28. Moon HJ, Kwak JY, Kim MJ, Son EJ, Kim EK. Can vascularity at power Doppler US help predict thyroid malignancy? *Radiology*. 2010;255:260–269.
29. Papini E, Guglielmi R, Bianchini A, et al. Risk of malignancy in nonpalpable thyroid nodules : predictive value of ultrasound and color-Doppler features. *J Clin Endocrinol Metab*. 2002;87:1941–1946.
30. 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.