

OPEN

# Grey-Scale Analysis Improves the Ultrasonographic Evaluation of Thyroid Nodules

Giorgio Grani, MD, Mimma D'Alessandri, MD, Giovanni Carbotta, MD, Angela Nesca, MD, Marianna Del Sordo, MD, Stefania Alessandrini, MD, Carmela Coccaro, MD, Roberta Rendina, MD, Marta Bianchini, MD, Natalie Prinzi, MD, and Angela Fumarola, MD

**Abstract:** Ultrasonography is the main imaging method for the workup of thyroid nodules. However, interobserver agreement reported for echogenicity and echotexture is quite low. The aim of this study was to perform quantitative measurements of the degree of echogenicity and heterogeneity of thyroid nodules, to develop an objective and reproducible method to stratify these features to predict malignancy.

A retrospective study of patients undergoing ultrasonography-guided fine-needle aspiration was performed in a University hospital thyroid center. From January 2010 to October 2012, 839 consecutive patients (908 nodules) underwent US-guided fine-needle aspiration. In a single ultrasound image, 3 regions of interest (ROIs) were drawn: the first including the nodule; the second including a portion of the adjacent thyroid parenchyma; the third, the strap muscle. Histogram analysis was performed, expressing the median, mean, and SD of the gray levels of the pixels comprising each region. Echogenicity was expressed as a ratio: the nodule/parenchyma, the nodule/muscle, and parenchyma/muscle median gray ratios were calculated. The heterogeneity index (HI) was calculated as the coefficient of variation of gray histogram for each of the 3 ROIs. Cytology and histology reports were recorded.

Nodule/parenchyma median gray ratio was significantly lower (more hypoechoic) in nodules found to be malignant (0.45 vs 0.61;  $P=0.002$ ) and can be used as a continuous measure of hypoechoogenicity (odds ratio [OR] 0.12; 95% confidence interval [CI] 0.03–0.49). Using a cutoff derived from ROC curve analysis ( $<0.46$ ), it showed a substantial inter-rater agreement ( $k=0.74$ ), sensitivity of 56.7% (95% CI 37.4–74.5%), specificity of 72.0% (67.8–75.9%), positive likelihood ratio (LR) of 2.023 (1.434–2.852), and negative LR of 0.602 (0.398–0.910) in predicting malignancy (diagnostic odds ratio 3.36; 1.59–7.10). Parenchymal HI was associated with anti-thyroperoxidase positivity (OR 19.69; 3.69–105.23). The nodule HI was significantly higher in malignant nodules (0.73 vs 0.63;  $P=0.03$ ) and, if above the 0.60 cutoff, showed sensitivity of 76.7% (57.7–90.1%), specificity of 46.8% (42.3–51.4%), positive LR of 1.442 (1.164–1.786), and negative LR of 0.498 (0.259–0.960).

Evaluation of nodule echogenicity and echotexture according to a numerical estimate (nodule/parenchyma median gray ratio and nodule

HI) allows for an objective stratification of nodule echogenicity and internal structure.

(*Medicine* 94(27):e1129)

**Abbreviations:** CI = confidence interval, CV = coefficient of variation, DOR = diagnostic odds ratio, FNAC = fine-needle aspiration cytology, HI = heterogeneity index, OR = odds ratio, ROI = region of interest, US = ultrasonography.

## INTRODUCTION

Ultrasonography (US) is the most widely accepted imaging method for the diagnosis and follow-up of diffuse thyroid diseases and for the workup of focal lesions. Since its introduction, it has proven to be a cornerstone in the diagnostic algorithm of thyroid nodules because of its low cost and widespread availability. Thyroid echogenicity reflects the follicular structure because the interface between thyrocytes and colloid has a high acoustic impedance. Diffuse or focal variations of normal gland structure cause an alteration of this pattern. Nodule echogenicity is dependent on the composition and cellularity of the nodule.<sup>1</sup> Subjective visual analysis is the first-line method to assess thyroid nodules, although it is limited by inter-observer variability and observer experience, with a consequently low reproducibility of results. Echogenicity of the gland is usually compared with the strap muscles, and echogenicity of the nodule is compared with the surrounding parenchyma, allowing definitions like “hypo,” “hyper,” or “isoechogenic” tissue. Marked hypoechoogenicity is usually defined as relative hypoechoogenicity compared with the surrounding muscles (mm. sternohyoidei and sternothyroidei).<sup>2,3</sup> However, neither the AIUM Guidance<sup>4</sup> nor a recent multidisciplinary consensus statement<sup>5</sup> describes how to assess echogenicity (with respect to muscles or surrounding thyroid tissue). According to a recent report,<sup>6</sup> hypoechoogenicity and marked hypoechoogenicity showed an odds ratio (OR) for thyroid malignancy of 2.60 and 6.81, respectively. However, the comparison of different tissues is difficult and leads to inevitable variability among operators. Furthermore, thyroid US is performed and interpreted by a number of different specialists in various settings and with different training (endocrinologists, radiologists, otolaryngologist, and head and neck surgeons).<sup>5</sup> In fact, fair inter-observer agreement was recently reported for hypoechoogenicity ( $k=0.34$ ).<sup>7</sup> These findings were indirectly confirmed by a meta-analysis, showing that hypoechoogenicity has a diagnostic odds ratio (DOR) for malignancy of 4.5, with a significant heterogeneity between studies (inconsistency index  $I^2$  88%) and no distinction between various degrees of echogenicity.<sup>8</sup> Quantitative analysis is the second-line method, which aims to provide more objective and reproducible results. Several years ago, US gray-scale analysis was proposed to

Editor: Gaurav Malhotra.

Received: September 17, 2014; revised: June 2, 2015; accepted: June 17, 2015.

From the Department Of Experimental Medicine, Unit of Endocrinology, “Sapienza” Università di Roma, Rome, Italy.

Correspondence: Prof. Angela Fumarola, Department of Experimental Medicine, Sapienza Università di Roma, V.le Regina Elena, 324, 00161 Rome, Italy (e-mail: angela.fumarola@uniroma1.it).

The authors report no conflicts of interest.

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

This is an open access article distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0, where it is permissible to download, share and reproduce the work in any medium, provided it is properly cited. The work cannot be changed in any way or used commercially.

ISSN: 0025-7974

DOI: 10.1097/MD.0000000000001129

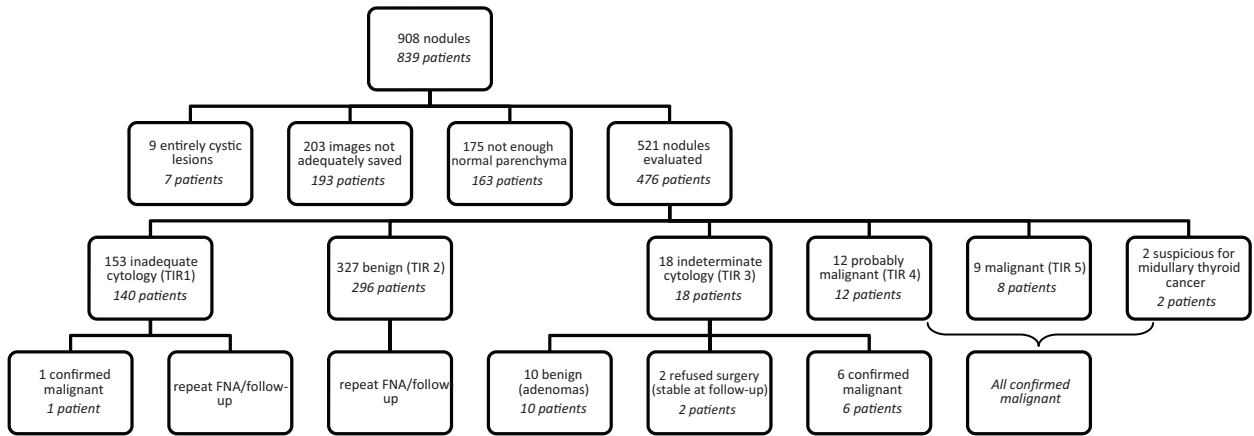


FIGURE 1. Flow chart of patient inclusion and data collection.

eliminate the subjective component of the evaluation, allowing quantitative determination of echogenicity, independently from the visual impression of the investigator, but it was applied only for the parenchymal assessment.<sup>2,3,9,10</sup> In the present study, we have performed a gray-scale histogram analysis of thyroid echogenicity in comparison with neck muscles, as well as nodule echogenicity in comparison with surrounding parenchyma. Through this procedure, we aim to perform – for the first time – a quantitative measurement of thyroid nodule echogenicity and echotexture (the distribution of echoes), obtaining a numerical and objective estimate of the degree of hypoechogenicity and homogeneity and evaluate their diagnostic accuracy in predicting the risk of nodule malignancy.

**MATERIALS AND METHODS**

The institutional review board approved this retrospective study and did not require patient approval or informed consent for the review of patient images and records. All patients gave written informed consent for the ultrasonography-guided fine-needle aspiration cytology (FNAC).

From January 2010 to October 2012, 839 consecutive patients (908 nodules) underwent US-guided FNAC at the Thyroid Center of our Department (Figure 1). The patients were aged  $55.92 \pm 13.55$  years (mean  $\pm$  SD) and all resided in Central Italy, an area of mild–moderate iodine deficiency. There was no difference in age between male and female patients: 161 men were aged  $57.12 \pm 13.35$  years, whereas 678 women were aged  $55.63 \pm 13.59$  years ( $P = 0.21$ ). All patients had at least 1 discrete nodular lesion of the thyroid or a multinodular goiter. Nodules were selected according to published international guidelines at the time of sampling,<sup>11,12</sup> but without any established volume cutoff.

Anti-thyroperoxidase (anti-TPO) antibodies were measured using radioimmunoassay (RIA, BRAHMS Diagnostica, Berlin, Germany) and were considered positive if found to be above the cutoff point set by the laboratory ( $>50$  U/mL).

US was performed using a Toshiba Aplio XV device equipped with a linear high-frequency transducer. Examinations were performed by 1 of 10 endocrinologists with 3 to 5 years of experience in thyroid US. US features were prospectively recorded (diameters, echogenicity, echotexture, and vascularity) into an electronic database for clinical use. Echogenicity was subjectively classified as hyper-, iso-, hypoechogenic (when a nodule showed hyper-, iso-, or hypoechogenicity compared

with the normal thyroid gland), iso-hypoechogenic (slight hypoechogenicity), or heterogeneous (hypo- or iso-echogenicity with a marked heterogeneous echotexture).

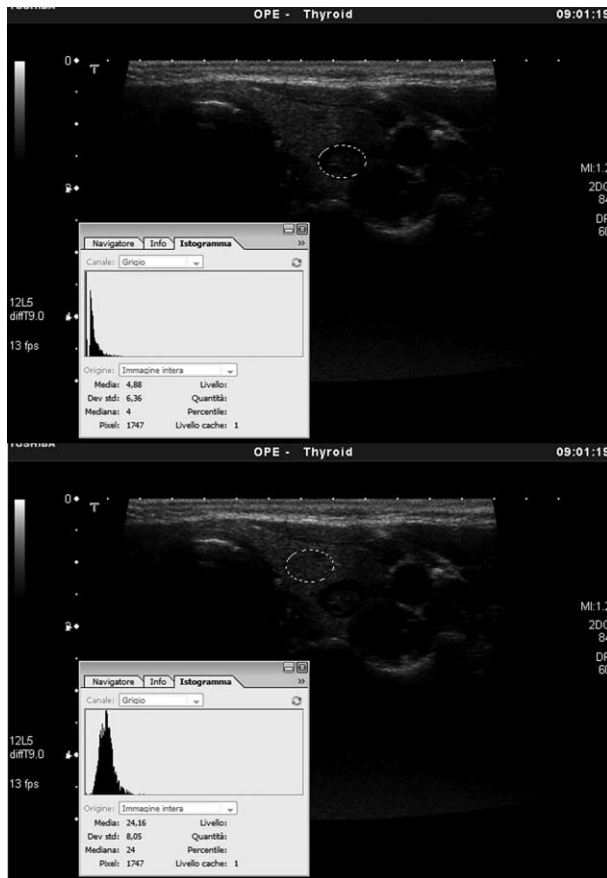
US-assisted FNAC was carried out through aspiration by using 23–25 gauge needles attached to a 10-mL syringe. The US transducer was placed to center the nodule with a transverse, free-hand approach, and the needle was inserted right in the middle of the transducer. The tip was localized as a bright point, when possible. One to 3 aspirations were performed on each nodule. In case of solid/cystic mixed nodule, a sample of the solid component was drawn.

The first smear was air-dried and stained using the May-Grunwald-Giemsa method. The remaining smears were fixed with Bio-fix (Bioptica, Milan, Italy), and stained with hematoxylin–eosin (HE). Cytology results were reported in 5 categories, as follows, according to the Thyroid Cytology Italian Consensus SIAPEC-IAP<sup>13</sup>: non-diagnostic, benign, indeterminate, probably malignant, and positive for malignant cells. Malignancy was confirmed by histology (30 cases; 2 medullary thyroid cancer, 28 papillary thyroid cancer) and ruled out by histology (20 cases), two consecutive FNACs (272), or a single FNAC with follow-up (197). Two patients had an indeterminate cytology report, but refused surgery and were excluded.

**Image Analysis**

The images were stored as 8-bit (28 = 0–255 gray-scale levels) JPEG images,  $716 \times 537$  pixels, and gray-scale mode. The grey histogram ranged from 0 (black pixels) to 255 (white pixels).

In each image, 2 regions of interest (ROIs) were manually drawn: the first included the entire nodule (excluding hypoechogenic halo, if present) and the second, a portion of the adjacent thyroid parenchyma with no focal lesions. The ROIs were drawn in the same image, having the same shape and number of pixels, avoiding blood vessels and artifacts. A third ROI included the strap muscle, having the same number of pixels, but different shape. Histogram analysis was then performed and echo intensity was measured in gray-scale levels, expressing the median, mean, and standard deviation (SD) of the gray-scale levels of the pixels included in each ROI (Figure 2). The echogenic appearance of the thyroid gland varies with the adjustment of various instrument settings (gain, depth range, dynamic range). To overcome the need of standard operating conditions, echogenicity was expressed as a ratio: the



**FIGURE 2.** Drawing regions of interest (ROIs) to include nodule (A) and parenchyma (B).

nodule/parenchyma median grey ratio, nodule/muscle median grey ratio, and parenchyma/muscle median grey ratio were then calculated. The heterogeneity index (HI) was calculated as the coefficient of variation (CV) of gray histogram (SD/mean) for each of the 3 ROIs (nodule HI, parenchyma HI, and muscle HI). Calcifications and echogenic foci of colloid were not excluded, to allow them to influence the HI. However, they unlikely influenced the median grey ratios because the median is slightly modified by a small number of bright pixels.

Nine nodules were excluded because of entirely cystic content, 203 were lacking a properly saved image, and 175 did

not have a sufficient amount of surrounding parenchyma. Five hundred twenty-one images of nodules were finally evaluated, coming from 476 patients (aged  $55.37 \pm 13.56$  years; 91 males aged  $55.67 \pm 14.27$  years, and 385 females aged  $55.30 \pm 13.40$  years;  $P = 0.81$ ).

Image analysis was performed retrospectively by a single operator (GG), without access to cytology and histology final data. A subsequent analysis was independently performed by another author (AF), to evaluate inter-rater agreement.

**Statistical Analysis**

The distributions of heterogeneity index and median gray ratios were not normal (Shapiro–Wilk test). Comparisons were performed using the Mann–Whitney *U* (between 2 groups) and Kruskal–Wallis tests (>2 groups). Categorical variables were compared using Fisher exact or Pearson chi-square test. Estimates of diagnostic accuracy were reported as sensitivity, specificity, likelihood ratio, and DOR with 95% confidence intervals (CIs). The DOR measures the discriminative power of a diagnostic test and reflects the test’s performance compared with the reference standard. The value ranges from 0 to infinity, with higher values indicating better performance. The sensitivities and specificities were compared using the exact McNemar test. The optimal cutoff value for the nodule/parenchyma ratio and nodule HI was calculated using receiver-operating characteristic (ROC) analysis. Odds ratios (ORs) and 95% CIs were also calculated. Interobserver variability for evaluation of echogenicity and for the measurement of the ratios (considering the nodule/parenchyma median gray ratio and nodule HI qualitatively as a binary variable) between the 2 observers was defined by using Cohen kappa statistic.<sup>14</sup> The Bland–Altman technique was used to determine the agreement between continuous measurements.<sup>15</sup> All tests used a 2-sided  $\alpha$  of 0.05. Data analysis was performed using IBM SPSS Statistics 20.0 (IBM Corp, Armonk, NY) and Microsoft Office Excel 2007.

**RESULTS**

**Age- and Sex-Related Differences**

No difference was recorded between males and females in relation to nodule/parenchyma median gray ratio, parenchyma HIs, and nodule HIs. Nodule/muscle median gray ratio, parenchyma/muscle median gray ratio, and muscle HIs were significantly higher in males than in females (Table 1). No significant difference was recorded between nodules in patients younger and older than 45 years (Table 2).

**TABLE 1.** Differences Between Males and Females (Mann–Whitney Test)

	Males		Females		<i>P</i>
	Median	Interquartile Range	Median	Interquartile Range	
Nodule/parenchyma median gray ratio	0.58	(0.43–0.97)	0.61	(0.44–0.86)	0.79
Parenchyma/muscle median gray ratio	2.14	(1.26–3.12)	1.47	(1.04–2.06)	<0.001
Nodule/muscle median gray ratio	1.20	(0.88–2.00)	0.96	(0.60–1.43)	<0.001
Nodule HI	0.61	(0.48–0.81)	0.65	(0.49–0.90)	0.40
Parenchyma HI	0.35	(0.29–0.43)	0.35	(0.30–0.42)	0.85
Muscle HI	0.57	(0.46–0.81)	0.53	(0.42–0.67)	0.02

HI = heterogeneity index.

**TABLE 2.** Age-related Differences (Mann–Whitney Test)

	Age ≤45 Years		Age >45 Years		P
	Median	Interquartile Range	Median	Interquartile Range	
Nodule/parenchyma median gray ratio	0.58	(0.43–0.86)	0.64	(0.43–0.89)	0.23
Parenchyma/muscle median gray ratio	1.63	(1.14–2.56)	1.53	(1.04–2.25)	0.58
Nodule/muscle median gray ratio	1.04	(0.58–1.64)	1.00	(0.64–1.50)	0.14
Nodule HI	0.63	(0.48–0.88)	0.64	(0.48–0.90)	0.85
Parenchyma HI	0.34	(0.29–0.42)	0.35	(0.30–0.42)	0.91
Muscle HI	0.55	(0.42–0.67)	0.54	(0.42–0.68)	0.32

HI = heterogeneity index.

**Median Gray Ratios**

Nodule/parenchyma median gray ratios were significantly different between mixed content and solid (iso-, or hypoechoic) nodules ( $P < 0.001$  and  $P = 0.004$  respectively) and, among solid nodules, between iso- and hypoechoic nodules ( $P < 0.001$ ), between heterogeneous and hypoechoic nodules ( $p = 0.02$ ), and between heterogeneous and isoechoic nodules ( $P = 0.001$ ) (Figure 3, Table 3). No significant difference was recorded between iso-hypoechoic nodules, isoechoic nodules, and heterogeneous nodules. Ratio was significantly lower (ie, more hypoechoic) in nodules found to be malignant (Table 4;  $P = 0.002$ ).

An ROC curve analysis was performed to establish a diagnostic cutoff (area under the curve [AUC] 0.67). A cutoff of 0.46 reaches sensitivity of 56.7% (95% CI 37.4%–74.5%) and specificity of 72.0% (95% CI 67.8%–75.9%) in predicting malignancy (DOR 3.36; 95% CI 1.59–7.10; Table 5). To rule

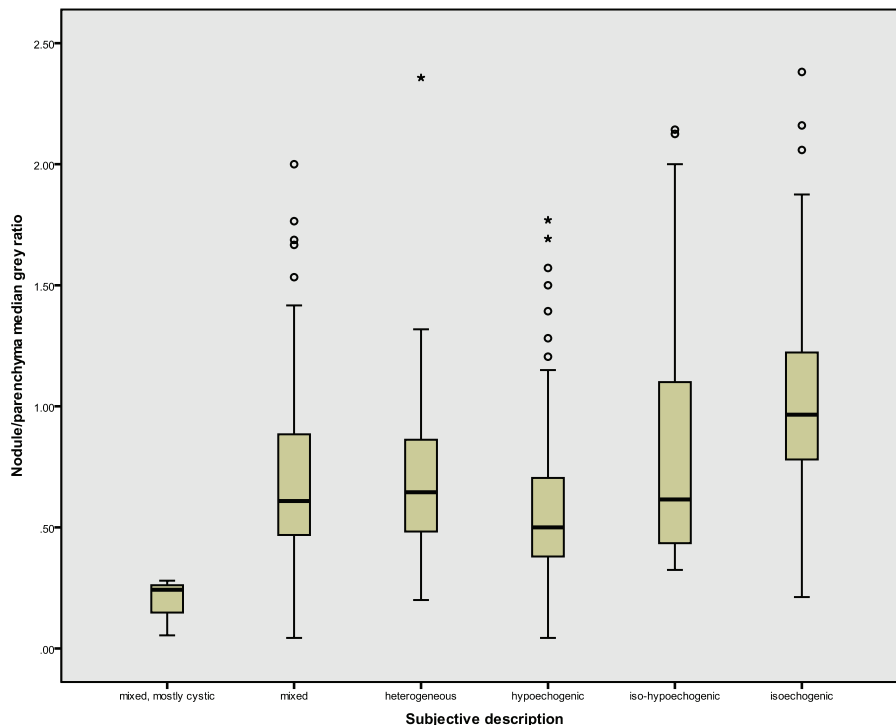
out liquid nodules, a double cutoff should be used (ratio  $>0.20$  and  $<0.46$ ), that achieves sensitivity of 53.3% (95% CI 34.3%–71.7%), specificity of 77.1% (95% CI 73.1%–80.7%), and DOR 3.85 (95% CI 1.82–8.12).

A logistic regression analysis was performed to demonstrate that nodule/parenchyma median gray ratio can be used as a continuous measure of hypoechoicity and – indirectly – of risk of malignancy (OR 0.12; 95% CI 0.03–0.49;  $P = 0.003$ ).

Nodule/muscle median gray ratio was analyzed stratified by sex, as it is influenced by sex (Table 1). Only in females, it turned out to be lower in malignant nodules (0.75, I.R. 0.42–0.83) than in benign ones (1.00, I.R. 0.62–1.44,  $P = 0.008$ ).

**HI**

To evaluate the ability of the proposed HI to evaluate the heterogeneous echotexture, the parenchymal HI was evaluated



**FIGURE 3.** Nodule/parenchyma median gray ratio according to subjective description of the nodules.

**TABLE 3.** Nodule/Parenchyma Median Gray Ratios and Nodule HI According to Subjective Description of the Nodules

Subjective Description	Nodule/Parenchyma Median Gray Ratio		Nodule HI	
	Median	Interquartile Range	Median	Interquartile Range
Solid				
Isoechogenic	0.97	(0.78–1.23)	0.47	(0.40–0.62)
Iso-hypoechoogenic	0.62	(0.43–1.11)	0.59	(0.46–0.69)
Heterogeneous	0.65	(0.48–0.87)	0.56	(0.49–0.83)
Hypoechoogenic	0.50	(0.38–0.71)	0.73	(0.52–1.13)
Mixed solid/liquid	0.61	(0.47–0.89)	0.69	(0.54–0.9)
Mostly cystic	0.24	(0.05–0.28)	2.09	(1.42–2.09)

HI = heterogeneity index.

in patients with or without anti-TPO antibodies. The median parenchymal HI was 0.38 (I.R. 0.31–0.44) in the positive anti-TPO group and 0.34 (I.R. 0.29–0.41) in the negative group ( $P = 0.007$ ). A logistic regression analysis showed that parenchymal HI is strongly associated with anti-TPO positivity (OR 19.69; 95% CI 3.69–105.23;  $P < 0.001$ ).

However, no relation was found between subjective description of heterogeneity and nodular HI. In particular, no significant difference was recorded between nodules labeled “heterogeneous” and those identified in any of the other categories except “isoechogenic” ( $P < 0.001$ ). The nodular HI was not a significant predictor of malignancy according to logistic regression analysis. However, it was significantly higher in malignant nodules (Table 4) and, with a cutoff of 0.60, derived from the ROC curve analysis; it achieves a sensitivity of 76.7% (95% CI 57.7%–90.1%) and a specificity of 46.8% (95% CI 42.3%–51.4%).

**Inter-Rater Agreement**

The inter-rater agreement was fair ( $k = 0.40$ ) for subjective assessment of hypoechoogenicity, whereas it was substantial for the nodule/parenchyma grey ratio ( $k = 0.74$  with a single cutoff,  $k = 0.64$  with a double cutoff). A Bland-Altman plot is reported in Figure 4.

**DISCUSSION**

Thyroid nodules are very common and there is a large pool of asymptomatic people with nodules in the general population. Depending on the population studied and the method of detection used, the prevalence of thyroid nodules ranges from 5% to 60%.<sup>8</sup> The probability that these nodules will be discovered by a widely available, economic, and harmless method, such as

thyroid US, is therefore increasing. Although the majority of these thyroid nodules are benign, the frequency of malignancy is 8% to 15%.<sup>16</sup> US is the best diagnostic tool to date in the initial work-up of thyroid nodules, to avoid unnecessary surgical interventions. However, there is significant uncertainty surrounding the diagnostic accuracy of several of the features analyzed during the US evaluation.<sup>8,17,18</sup> This uncertainty is due, to some degree, to the operator-dependence of the US techniques and the lack of agreement at reporting these features.<sup>5</sup> The interobserver agreement seems to be relatively good, according to recent reports<sup>7,19</sup>; however, only a slight agreement was reported for echogenicity<sup>7</sup> and echotexture.<sup>19</sup> The former, in particular, is a major suspicious US feature. Furthermore, the terminology used by ultrasonologists is ambiguous at times and various nomenclatures are available for the same finding. For this reason, a consensus reporting protocol was recently proposed, including the description of internal architecture and echogenicity of the nodules.<sup>5</sup> However, no indication was given about the assessment and classification of echogenicity (eg, the terms “hypoechoic” and “profoundly hypoechoic” were suggested, without any guidance on consistent interpretation). In a recent population-based study, echogenicity did not turn out to be a significant predictor of malignancy, but it was quite arbitrarily classified in “hypoechoic to strap,” “isoechoic or hyperechoic to strap,” “isoechoic to thyroid,” and “hyperechoic to thyroid.”

In the absence of accurate predictors of malignancy, many of the nodules will still require FNAC, which is expensive and often challenging, as in the case of indeterminate and non-diagnostic results. On the contrary, US features should serve as a guide for determining the depth of nodule evaluation.

Some authors have proposed the recognition of complex patterns rather than single features to identify suspicious

**TABLE 4.** Nodule/Parenchyma Median Gray Ratio and Nodule HI According to Final Diagnosis

Final Diagnosis	Nodule/Parenchyma Median Gray Ratio		Nodule HI	
	Median	Interquartile Range	Median	Interquartile Range
Benign	0.61	(0.44–0.91)	0.63	(0.48–0.89)
Malignant	0.45	(0.36–0.68)	0.73	(0.59–1.16)
Mann–Whitney test		$P = 0.002$		$P = 0.033$

HI = heterogeneity index.

TABLE 5. Diagnostic Values of Subjective and Numerical US Parameters

	TP	FP	FN	TN	Sensitivity (%)	95% CI	Specificity (%)	95% CI	LR+	95% CI	LR-	95% CI	DOR	95% CI
Nodule/parenchyma median grey ratio <0.46	17	137	13	352	56.7*	37.4–74.5	72.0 <sup>†</sup>	67.8–75.9	2.023	1.434–2.852	0.602	0.398–0.910	3.360	1.589–7.103
Subjective hypochoic appearance	20	196	10	293	66.7*	47.2–82.7	59.9 <sup>†</sup>	55.4–64.3	1.663	1.263–2.190	0.556	0.334–0.928	2.990	1.370–6.524
Nodule HI >0.60	23	260	7	229	76.7	57.7–90.1	46.8	42.3–51.4	1.442	1.164–1.786	0.498	0.259–0.960	2.894	1.219–6.870
At least one positive parameter (median grey ratio <0.46 and/or nodule HI >0.60)	26	290	4	199	86.7	69.3–96.2	40.7	36.3–45.2	1.461	1.247–1.712	0.328	0.131–0.821	4.460	1.533–12.978
Nodule/parenchyma median grey ratio <0.46 and >0.20	16	112	14	377	53.3	34.3–71.7	77.1	73.1–80.7	2.329	1.605–3.378	0.605	0.412–0.890	3.847	1.821–8.125

CI = confidence interval, DOR = diagnostic odds ratio, FN = false negative, FP = false positive, HI = heterogeneity index, LR+ = positive likelihood ratio, LR- = negative likelihood ratio, TN = true negative, TP = true positive, US = ultrasonography.

\* Exact McNemar test; *P* = 0.07.

<sup>†</sup> Exact McNemar test; *P* < 0.001.

nodules. To standardize this approach and to avoid unnecessary biopsies, several models were designed, for example, Thyroid Imaging Reporting and Data System (TIRADS),<sup>20–23</sup> similar to the Breast Imaging Reporting and Data System (BIRADS), used to standardize the interpretation of breast imaging. However, some of these are difficult to apply in clinical practice and rely on arbitrary evaluation of US features by the observer, even when employing complex equations.

In this study, we developed a method to perform a quantitative measurement of thyroid echogenicity (ie, the median gray ratio) and echotexture (ie, the heterogeneity index, calculated as CV of grey-levels histogram), using adimensional values and avoiding the need for fixed US operating conditions (such as depth and gain).

No significant difference was recorded (using numerical estimates of echogenicity) between iso-hypoechoic nodules and hypoechoic, iso-hypoechoic and isoechoic, and iso-hypoechoic and heterogeneous nodules, thereby showing a substantial overlap across these subjective descriptions, which are confusing for the clinician. However, a real, countable difference was observed between hypoechoic and isoechoic nodules. In particular, nodules with a nodule/parenchyma gray ratio <0.46 (ie, hypoechoic) have an OR for malignancy of 3.36. Moreover, our data showed that the appearance and echogenicity of neck muscles are different in males and females, thus demonstrating that the traditional definition of markedly hypoechoic nodules (hypoechoic with respect to strap muscles) is not easily reproducible. Indeed, a numerical evaluation of hypoechoic (nodule/parenchyma gray ratio) is able to precisely quantify the degree of hypoechoic and estimate the risk of malignancy related to this particular feature, without comparisons with different tissues, with a substantial interobserver agreement (*k* = 0.74) and diagnostic accuracy.

The heterogeneity index was recently proposed by Wakita et al<sup>10</sup>; however, they evaluated the US intensities along a straight line, a method not applicable for evaluation of focal lesions. We have calculated HI for thyroid parenchyma and nodules, with a different method. The parenchyma HI showed the ability to recognize thyroid autoimmunity and seems to be a promising approach, whereas the nodule HI was increased in malignant nodules.

Our study had several limitations. First, selection bias may have existed in recruiting patients to include in the study because all nodules were suspicious to a certain extent and submitted to FNAC. However, our institution policy is to perform FNAC also in case of a single risk factor, be it clinical (familiarity, history of irradiation) or sonographic (ill-defined margins, microcalcification, hypoechoic). Accordingly, the overall malignancy rate of this cohort is relatively low (5.76%) if compared with other reports. Furthermore, the sample was insufficient to evaluate the diagnostic value of this method in the different histotypes of thyroid cancer and the reproducibility of this technique with different US equipments is to be confirmed.

In conclusion, we demonstrated that gray-scale analysis is applicable to the US evaluation of thyroid nodules in clinical practice, rapidly and without the need for fixed operating conditions. Evaluation of nodule echogenicity and echotexture according to a numerical estimate (nodule/parenchyma median gray ratio and nodule HI) allows for a practical and objective stratification of nodule echogenicity and internal structure that could also be included in new TIRADS models. Adoption of uniform standards for the reporting of thyroid US could

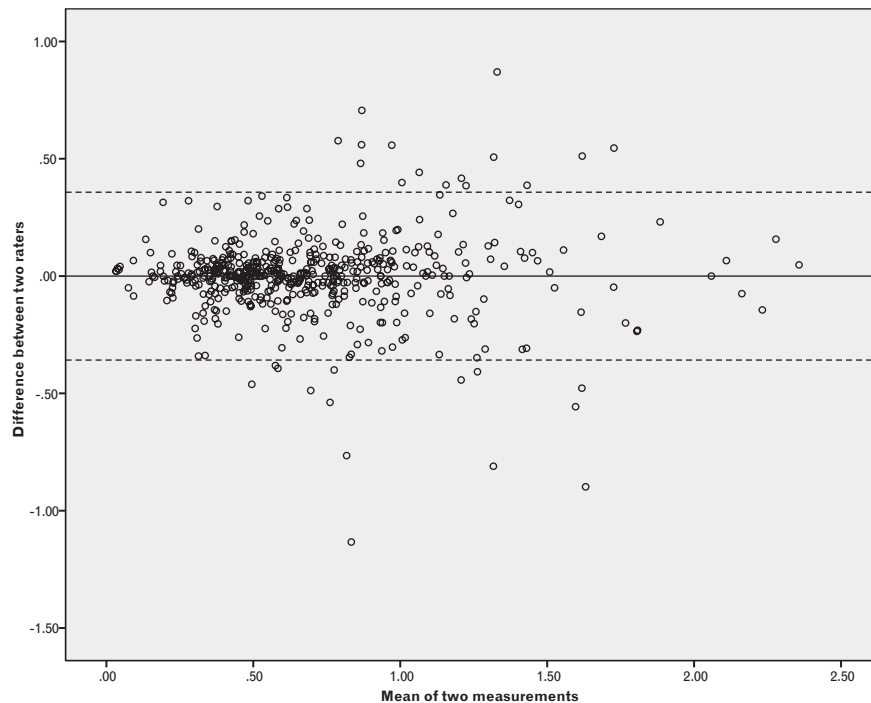


FIGURE 4. Bland-Altman plot for nodule/parenchyma median gray ratio.

probably be useful in limiting unnecessary diagnostic procedures (repeated FNAC, core-needle biopsy) or treatments.

## REFERENCES

- Chen SJ, Yu SN, Tzeng JE, et al. Characterization of the major histopathological components of thyroid nodules using sonographic textural features for clinical diagnosis and management. *Ultrasound Med Biol*. 2009;35:201–208.
- Schiemann U, Gellner R, Riemann B, et al. Standardized grey scale ultrasonography in Graves' disease: correlation to autoimmune activity. *Eur J Endocrinol*. 1999;141:332–336.
- Schiemann U, Avenhaus W, Konturek JW, et al. Relationship of clinical features and laboratory parameters to thyroid echogenicity measured by standardized grey scale ultrasonography in patients with Hashimoto's thyroiditis. *Med Sci Monit*. 2003;9:MT13–MT17.
- American Institute of Ultrasound in Medicine, American College of Radiology, Society for Pediatric Radiology, Society of Radiologists in Ultrasound. AIUM practice guideline for the performance of a thyroid and parathyroid ultrasound examination. *J Ultrasound Med*. 2013;32:1319–1329.
- 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.
- Kwak JY, Jung I, Baek JH, et al. Image reporting and characterization system for ultrasound features of thyroid nodules: multicentric Korean retrospective study. *Korean J Radiol*. 2013;14:110–117.
- Choi SH, Kim EK, Kwak JY, et al. Interobserver and intraobserver variations in ultrasound assessment of thyroid nodules. *Thyroid*. 2010;20:167–172.
- 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.
- Mazziotti G, Sorvillo F, Iorio S, et al. Grey-scale analysis allows a quantitative evaluation of thyroid echogenicity in the patients with Hashimoto's thyroiditis. *Clin Endocrinol (Oxf)*. 2003;59:223–229.
- Wakita Y, Nagasaki T, Nagata Y, et al. Thyroid heterogeneity, as indicated by the CV of ultrasonographic intensities, correlates with anti-thyroid peroxidase antibodies in euthyroid Hashimoto's thyroiditis. *Thyroid Res*. 2013;6:5.
- Gharib H, Papini E, Paschke R, et al. American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and European Thyroid Association Medical Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules. *Endocr Pract*. 2010;16(Suppl 1):1–43.
- Cooper DS, Doherty GM, Haugen BR, et al. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2009;19:1167–1214.
- Fadda G, Basolo F, Bondi A, et al. Cytological classification of thyroid nodules. Proposal of the SIAPEC-IAP Italian Consensus Working Group. *Pathologica*. 2010;102:405–408.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–310.
- Frates MC, Benson CB, Doubilet PM, et al. Prevalence and distribution of carcinoma in patients with solitary and multiple thyroid nodules on sonography. *J Clin Endocrinol Metab*. 2006;91:3411–3417.
- Smith-Bindman R, Lebda P, Feldstein VA, et al. Risk of thyroid cancer based on thyroid ultrasound imaging characteristics: results of a population-based study. *JAMA Intern Med*. 2013;173:1788–1796.
- Alexander EK, Cooper D. The importance, and important limitations, of ultrasound imaging for evaluating thyroid nodules. *JAMA Intern Med*. 2013;173:1796–1797.

19. Park CS, Kim SH, Jung SL, et al. Observer variability in the sonographic evaluation of thyroid nodules. *J Clin Ultrasound*. 2010;38:287–293.
20. 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.
21. 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.
22. 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.
23. Russ G, Royer B, Bigorgne C, et al. Prospective evaluation of thyroid imaging reporting and data system on 4550 nodules with and without elastography. *Eur J Endocrinol*. 2013;168:649–655.