

Thyroid Nodule: Alpha Score 2.0 Classification for FNAB Selection, Multicentric Study in Latin America

Glenn Mena^{1*}, Maria Cristina Chammas², Carlos Mario Gonzalez Vasquez³, Lylian Rocío Villagómez¹, Marco Alfredo Muñoz Pico¹, Patricio Alejandro Montalvo⁴, Santiago Mena-Bucheli⁵, Julio Olmedo¹, Elizabeth Quintero¹, Pedro Henrique de Marqui Moraes², Osmar Cassio Saito², Hubertino Diaz⁶, Denise Romero⁷, Gabriela Velalcazar⁸, Angel Ramón Sosa Fleitas⁹, Yamil Oliver Quevedo Ontaneda¹⁰, Victor Ricardo Chara¹¹

¹Alpha Imagen Radiología e Intervencionismo, Quito, Ecuador

²Hospital das Clínicas Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brasil

³Hospital Pablo Tobón Uribe, Medellín, Colombia

⁴Hospital del Instituto Ecuatoriano de Seguridad Social Quito Sur, Quito, Ecuador

⁵Hospital San Francisco de Quito, Quito, Ecuador

⁶Hospital Edgardo Rebagliati Martins, Lima, Perú

⁷Hospital Carlos Andrade Marín, Quito, Ecuador

⁸Hospital de Clínicas San José de San Martín, Buenos Aires, Argentina

⁹Hospital Universitario de Los Andes, Mérida, Venezuela

¹⁰Hospital de la Policía, Quito, Ecuador

¹¹Hospital Sisol, Tacna, Perú

Email: *glennmena@hotmail.com, mcchammas@hotmail.com, carlosmariogonzalesvasquez@hotmail.com, rocio_villagomez2000@yahoo.com.ar, markognr@hotmail.com, alesum41@hotmail.com, santiagomenabucheli@gmail.com, catolmedo@gmail.com, elizabethquintero_413@hotmail.com, pedrohenrique.mmoraes@gmail.com, ocsaito@gmail.com, hubertinodiaz@gmail.com, ddena81@hotmail.com, gvelalcazar@gmail.com, angelsosa89@gmail.com, dryamilito@hotmail.com, victorick13@gmail.com

How to cite this paper: Mena, G., Chammas, M.C., Vasquez, C.M.G., Villagómez, L.R., Pico, M.A.M., Montalvo, P.A., Mena-Bucheli, S., Olmedo, J., Quintero, E., de Marqui Moraes, P.H., Saito, O.C., Diaz, H., Romero, D., Velalcazar, G., Fleitas, A.R.S., Ontaneda, Y.O.Q. and Chara, V.R. (2021) Paper Title. *Open Journal of Radiology*, 11, 160-174.

<https://doi.org/10.4236/ojrad.2021.114015>

Received: September 18, 2021

Accepted: December 6, 2021

Published: December 9, 2021

Abstract

Introduction: To perform a Latin-American multicentric study for the prediction of benign and malignant thyroid nodules using Alpha Score, and to compare it with ACR TIRADS[®] and Bethesda[®]. **Materials and Methods:** A prospective multicentric study in 10 radiological hospitals and institutions of Latin America was performed and 818 thyroid nodules were analyzed by ultrasound and classified by using both ACR TIRADS[®] and Alpha Score; fine-needle aspiration biopsy was performed when needed and classified with Bethesda. The relationships between predictors were analyzed by using binary logistic regression, statistical significance was defined by a p-value of 0.05, with an error margin of 4% and 95% confidence intervals. **Results:** Alpha Score 2.0 establishes five types of malignant predictors: microcalcifications,



irregular borders, taller-than-wide shape, predominant solid texture and hypoechogenicity; a diameter equal to or greater than 1.5 cm adds an extra point to the final score. Resulting classification divides TNs into 4 categories: benign (1.9%), low suspicion (8.7%), mild suspicion (13.6%) and high suspicion (75.7%) of malignancy probability; sensitivity of 82%, specificity of 74%, the positive predictive value of 94%, the negative predictive value of 51%, the statistical accuracy of 81%, odds ratio of 108.89 and correlation with ACR TIRADS of 0.77 and Bethesda of 0.91. **Conclusions:** Alpha Score 2.0 has superior diagnostic accuracy and performance compared to the previously published Alpha Score and is able to classify a benign TN in a precise, safe and accurate way, avoiding unnecessary FNABs or determining the necessity of FNAB in cases of moderate to high suspicion of malignancy.

Keywords

Thyroid Cancer-Clinical, Radiology-Imaging, Thyroid Nodule, Alpha Score, TIRADS[®], Thyroid Ultrasonography

1. Introduction

Current clinical practice prioritizes active surveillance of the suspicious or malignant Thyroid Nodules (TN) in order to define the most appropriate treatment on a case-by-case basis while minimizing unnecessary invasive procedures such as Fine Needle Aspiration Biopsy (FNAB) or surgery [1]. A characteristic example is Papillary Thyroid Microcarcinomas which are usually slowly growing tumors with a low percentage of malignant transformation that will benefit from a robust risk stratification tool to further assess the risk and benefits of active surveillance vs. definitive diagnostic and surgical/pharmacological management [2].

TN risk stratification has been dependent upon multiple classifications that use Thyroid Ultrasonographic Malignancy Predictors (TUMP), such as the Thyroid Imaging Reporting and Data System TIRADS[®], published first by Horvath *et al.*, 2009 [3] and followed by the American College of Radiology TIRADS[®] (ACRT) [4] with reliable and effective results. Every published classification is effective, such as ACRT [4], EUTIRADS [5], KTIRADS [6]; however, there are some differences such as the included TUMPs, the time it takes to properly assess each one, and the population that was used to validate each score. In contrast, less complex classifications have been published such as Fernandez Sanchez, 2014 [7], Bailey and Wallwork, 2018 [8], and Seo *et al.*, 2015 [9] presenting acceptable results but with certain difficulties regarding reliability, reproducibility, and correlation with the Bethesda cytological classification system [10] [11].

In 2018, we published the Alpha Score (AS), Mena *et al.* [12], which used 7 TUMPs: hypoechogenicity, Solid Consistency (SC), irregular borders, microcalcifications, Absence of Peripheral Halo (APH), intra-nodular vascularity (INV), and a size larger than 10 mm, with 4 categories based on the likelihood of malignancy and their need for FNAB (Figures 1-4). The AS has been used in Latin

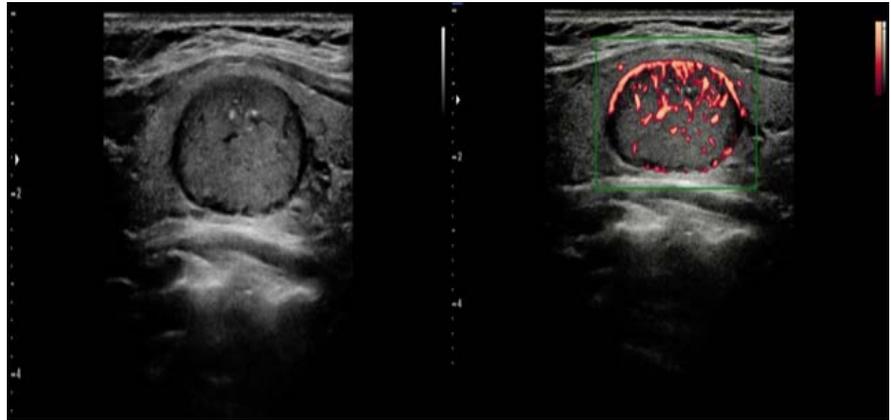


Figure 1. Solid Nodule (2 points), central and peripheral vascularization (0 points), peripheral microcalcifications (2 points), 1.8 cm diameter (1 point), absence of peripheral halo (0 points). Total AS 2.0: 5, moderate probability of malignancy, ACRT: 5 (T4), moderate suspicion. Conduct: FNAB, Diagnostic: Bethesda IV, Follicular neoplasia. Note: If observers don't consider the image as microcalcifications but rather as comet tail artifacts, AS 2.0 would have 3 points and ACRT 3 points (T3) resulting in active vigilance. Source: Hospital Edgardo Rebagliati Martins. Lima, Peru.

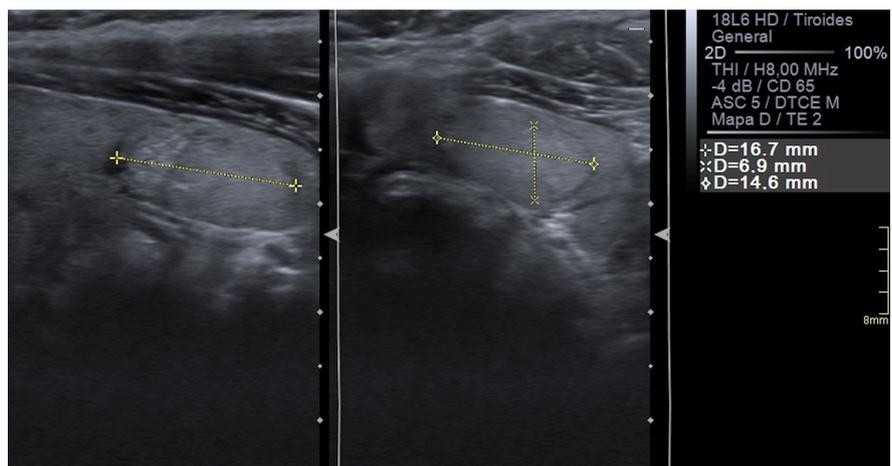


Figure 2. Solid nodule (2 points), isoechoic (0 point), regular borders (0 points), greater diameter above 1.5 cm (1 point). Total AS 2.0: 3, low probability of malignancy, ACRT: T2 + greater diameter superior to 1.5 cm. Conduct: Both classifications don't recommend FNAB. Diagnostic: Hyperplastic benign nodule, Bethesda II. Source: Hospital Pablo Tobón Uribe, Medellín, Colombia.

America but with concerns regarding its sensitivity and the use of two TUMPs, the APH and INV. AS was also conceived thanks to the feedback of radiologists and endocrinologists who prefer simpler classifications that avoid unnecessary FNABs, thus requiring adequate specificity and PPV for optimal reliability [1] [5].

Therefore, in order to improve and promote the use of our stratification tool, we deemed it necessary to perform a multicentric validation through a well-designed prospective study in various institutions and hospitals of Latin America using AS, ACRT and Bethesda in order to determine the reliability and diagnostic

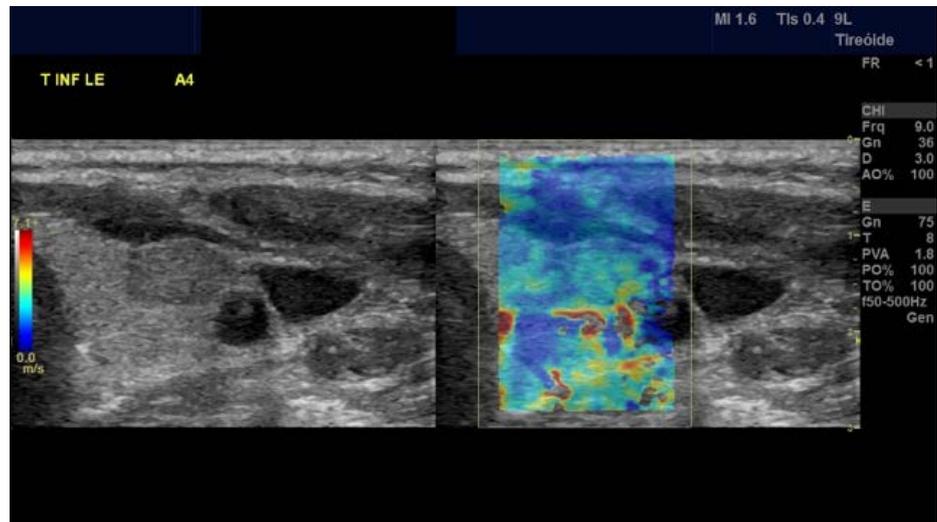


Figure 3. Solid nodule (2 points), hypoechoic (1 point), slightly elevated border on anterior thyroid capsule (2 points). Total AS 2.0: 5, moderate probability of malignancy, recommended action FNAB; ACRT: 6 points (T4), greater diameter less than 1.5 cm, conduct: vigilance. FNAB was performed, diagnostic: Follicular neoplasia, Bethesda IV. Note: If other observers consider that borders are smooth and regular then AS 2.0 would be 3 and would recommend active vigilance and ACRT: 4 points (T4) and with a greater diameter of less than 1.5 cm, also indicates vigilance. Right side of the image shows 2D Shear Wave quantitative elastography (see reference of m/s scale). Source: Hospital Das Clínicas, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brasil.

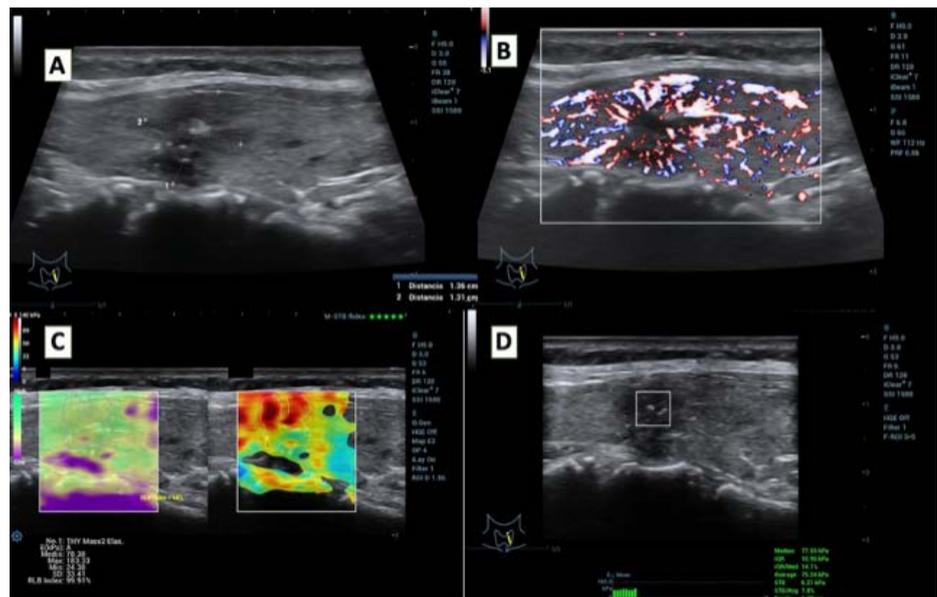


Figure 4. Panel A and B: Solid Nodule (2 points), Hypoechoic (1 point), irregular angulated borders (2 points), shape taller than wide (2 points), microcalcifications (2 points), peripheral abundant vascularity (0 points), Total AS 2.0: 9 points, high malignancy probability. ACRT 12 points (T5) high suspicion, conduct FNAB, diagnostic: Thyroid papillary carcinoma in a lymphocytic thyroiditis, Bethesda VI. Panel C Elastography, to the left: 2D Real time Shear Wave with an E max of 183 kPa, E mean of 78 kPa, with quality control mapping Panel D: Point Shear Wave multishot with a median of 79 kPa and IQR of 14%. Source: Alpha Imagen Radiología e Intervencionismo, Quito, Ecuador.

accuracy of our scoring system. Additionally, our study contributes to observations regarding to cytopathological data from Bethesda and the difficulties of applying elastography in these types of studies.

2. Materials and Methods

An analytical, observational, prospective, multicentric study was performed in 10 hospitals and radiological institutions of Latin America located in Argentina, Brazil, Colombia, Ecuador, Peru and Venezuela. A total of 1085 patients with FNABs for TN were studied in the period from January 2019 to September 2019. Database analysis was performed by the Research Department at Alpha Imagen in Quito, where data validation was also performed (duplications, integrity depletions, consistency). After depletion of the database and exclusion of Bethesda category I, 818 TN were included for final analysis. Ten radiologists, managers and coordinators of each participating institution, with a range of 5 - 20 years of experience in thyroid ultrasound, trained to use AS [12] and ACRT, analyzed the TNs by strictly using the lexicon and description of ACRT [4] and AS [12] for all TUMPs. The diameter of all TNs was determined by choosing the plane (sagittal, transverse or oblique) where the diameter was the greatest. All data extracted at each institution were registered in an on-line platform (Google Docs Version: DIC-AI. MOMeNTO. 2019. v2.1[®]) that was constantly monitored to ensure quality by the team from Alpha Imagen (Ecuador). Additionally, three institutions conducted elastography studies (Strain, 2D Shear wave and point Shear wave) on each TN.

Inclusion criteria included: 1) patients referred for thyroid FNAB from outpatient clinics or from the departments (endocrinology, internal medicine, or head and neck surgery) of the participating hospitals; 2) only one nodule per patient will be included in our analysis, if a patient has more than one, the TN with the highest score in the classifications will be selected. Exclusion criteria included: 1) patients with previous thyroid surgery; 2) patients with FNAB prior to the beginning of the study; 3) patients with previous iodine-based therapy [13].

Every TN was analyzed prospectively in the participating institutions, on real time ultrasound, with the following ultrasonography equipment: Resona 7 Mindray[®], Aplio 500 Canon Toshiba[®], Acuson S300 Siemens[®], RS80 Samsung[®], Aloka Arietta S70 Hitachi[®], Logic E9 GE[®], H60 Samsung[®], Voluson E8 GE[®], EPIQ5 Philips[®], Logic F8 GE[®] and Affiniti 50G Philips[®]. Only high definition linear transducers were used and there was no need for compatibility calculations between different equipment brands thanks to the standardization of tests performed to assure an adequate registration of B-Mode images; such tests are well documented in scientific literature (Sassaroli *et al.*, 2019) [14], as well as validated by the American Institute of Ultrasound in Medicine (AIUM) [15] [16]. FNABs were performed and sent for analysis by an expert thyroid pathologist who applied the Bethesda classification system.

After modeling tests (decision trees, factorial analysis), Binary Logistic Re-

gression (BLR) was determined to be the ideal method for analysis, by grouping 8 dichotomic variables, the 7 TUMPs of AS and the taller-than-wide shape (TTW) TUMP from ACRT, which is their only dichotomic variable (Figure 4 Panel A). BLR provides coefficients that correlate each variable with the likelihood of the TNs being malignant or benign, provides the associated p-value, odds ratio (OR), chi-square goodness of fit of Nagerkeke and classification index. The 8 TUMPs were statistically examined as predictors both individually (univariate analysis) and grouped (bivariable and multivariate analysis), depending on their mean values, standard deviations, medians, interquartile ranges, frequency, percentages. Kolmogorov (>50 TNs) was used to confirm that the data was non parametrical and different tests were applied, Kendall for dependency, Mann-Whitney for sum of ranks, McNemar for dichotomous contingency features and Spearman for statistical dependency, each one considering a p-value of 0.05 for a sample with 95% statistical confidence and an error margin of 4%. From the contingency table, Sen, Spe, PPV, NPV and statistical accuracy were calculated for each variable, data was validated with AUC values, and every calculation was made on EXCEL[®] and SPSS-22 of IBM[®].

Meetings took place between the co-authors to analyze distinct scenarios with resulting statistical values to decide, through medical radiological criteria, which variables would be used and validated. The β parameter was used to estimate values with a maximum verisimilitude method, which selects the coefficients that are more compatible with the observed results or that have the highest likelihood for the observed results to actually occur. The method assigns a value of 1 to the TUMP of least value, 0 cannot be assigned due to it representing absence of a characteristic. The data recorded in the β parameter for each TUMP were compared with each other to determine the real importance of each TUMP. To achieve this, the ratio/reason method is adopted, through which, each of the predictors gets assigned a specific value that results from the quotient with the variable assigned the value of 1, decimals are approximated when >0.5 [17].

Institutional Review Board approval was obtained at each individual institution. AS 2.0 was not used to inform or change the course of treatment and/or clinical management of the patients, it was only calculated in addition to a well validated prognostic tool (ACRT) that was the one used to inform patient management. This study was reported following the STROBE guidelines [18].

3. Results

818 TNs were included for final analysis, of those 83.7% were female and 16.3% were male between 14 and 88 years, subdivided by the median into 2 distinct age groups (14 - 52 years and ≥ 53 years). Cytopathological diagnoses, stratified with Bethesda, were as follows: 69.4% benign (Bethesda II), 18.1% uncertain (Bethesda III and IV), and 12.5% malignant (Bethesda V and VI), detailed in Table 1. Variables were excluded from final analysis based on the results outputted by the statistical model used, that calculates standard error, statistical significance and

Table 1. Multivariate cytopathological results of FNAB obtained according to Bethesda grades.

BETHESDA II:		
Benign colloid, benign follicular nodule, lymphocytic thyroiditis, granulomatous thyroiditis	568	69.4%
BETHESDA III:		
Atypia of undetermined significance, follicular lesion of undetermined significance	86	10.5%
BETHESDA IV:		
Follicular neoplasia, hurthle cell neoplasia	62	7.6%
BETHESDA V:		
Carcinomas: papillary, medullary and metastatic; lymphoma	31	3.8%
BETHESDA VI:		
Carcinomas: papillary, medullary, anaplastic, squamous cell, mixed and metastasis	71	8.7%
Total	818	100.0%

OR [19]. The reasons used for excluding variables were: statistical p-value > 0.05; standard deviation > 1 and OR value closer to 1, which points towards a lack of association with other variables or those variables have an exaggerated magnitude (Table 2). The TUMPs INV, APH and TN diameter 1.0 cm, that all had a p-value greater than 0.05 and an OR value with close proximity to 1, were modified from the original AS in order to improve the statistical power of the tool; APH and INV were eliminated, TN diameter acquired a special scoring system, and a new TUMP was introduced (TTW) resulting in the new Alpha Score 2.0 (AS 2.0) (Table 2, Figure 1, Figure 2 and Figure 4).

The data match between Bethesda and AS 2.0 shows expected results and percentage distribution points towards a direct correlation between AS 2.0 and Bethesda for classifying a TN as either benign or malignant; complete data is showed on Table 3. The results from the analysis of each of the included TUMPs was used to determine the final score given to each variable in order to calculate the final AS 2.0 as shown in Table 4 [20]. Follow up of TNs with SC showed higher malignant outcomes in around 60% of cases. Furthermore, when SC was associated with TN diameter, an additional variable chosen by measuring the highest association degree (X^2 Pearson, association and Cramer V for symmetry) between each individual variable compared with SC [21]. The association of these two variables with the likelihood of malignancy increased, which is why a bonification of 1 additional point was given to the SC TUMP when associated with a TN diameter equal or greater than 1.5 cm (Figure 1, Figure 2). Table 5 shows the statistic diagnostic tests (Se, Sp, PPV, NPV) for the TUMPs individually as well as for the whole AS 2.0. Regarding the correlation between AS 2.0 with ACRT and Bethesda, using Pearson's R (benign and malignant) the values were: 0.91 with Bethesda and 0.77 with ACRT, with a standard error of 0.027 and 0.04 respectively, both with a significance of 0.00. ROC of AS 2.0 TUMPs

Table 2. Regression equation variables and statistic results for the considered TUMPs.

Regression equation variables	Statistical measurements		
	Standard error	α statistical significance	Odds ratio occurrence probability/ non-occurrence probability
Vascularity	0.4530658	0.07076785	2.26749052
Absence of peripheral halo	0.41026163	0.93525034	1.03389171
Diameter 1.0 cm	0.48498926	0.69448993	0.82655497
Intranodular flow	0.44491332	0.12526605	1.97796753
Hypoechoic	0.45917851	0.04463	2.51448391
Solid consistency	0.43173031	0.00000	7.70073918
Irregular Borders	0.41341148	0.00000	9.60571352
Microcalcifications	0.40427569	0.00000	15.3969033
Shape taller than width	0.41508952	0.00000	7.94102587

Table 3. Cytopathological Bethesda results according to Alpha Score.

Alpha Score categories	Cytopathological Bethesda Results			
	Benign Bethesda II n and %	Uncertain Bethesda III. IV n and %	Malignant Bethesda V. VI n and %	Total n and %
Benign	451 79.5%	75 50.7%	2 1.9%	528 64.5%
Low suspicion of malignancy	75 13.2%	34 23.0%	9 8.7%	118 14.4%
Moderate suspicion of malignancy	28 4.9%	21 14.2%	14 13.6%	63 7.7%
High suspicion of malignancy	13 2.3%	18 12.2%	78 75.7%	109 13.3%
Total	567 100.0	148 100.0	103 100.0	818 100.0

Table 4. Score assignment for each of the selected predictors (TUMPs).

Thyroid Ultrasound Malignancy Predictors (TUMP)	B	Score	Assigned Score
Hypoechoic	1.075	1	1
Irregular Borders	2.068	1.92285415	2
Shape Taller than Wide	2.173	2.02063224	2
Solid Consistence	2.556	2.37669635	2
Microcalcifications	2.182	2.02924771	2

Note: Calculated with the formula: $1/1+e^{-f}$, 0.9909 probability of occurrence (OP), and 0.0091 probability of nonoccurrence and a 108.89 Odds ratio.

Table 5. Predictive validity metrics of each individual TUMP and AS 2.0.

	Hypoechoic	Solid	Microcalcifications	Irregular Borders	Shape Taller than Wide	Diameter = or More than 1.5 cm	Average Multicentric Alpha Score 2.0
Sensitivity	49%	90%	90%	89%	93%	50%	82%
Specificity	84%	57%	81%	84%	66%	51%	74%
PPV	95%	92%	96%	97%	94%	83%	95%
NPV	23%	52%	58%	58%	63%	17%	51%
Accuracy	54%	85%	88%	88%	89%	60%	81%

Abbreviations: PPV, positive predictive values; NPV, negative predictive values.

are detailed in **Table 6**. The elastography values in kPa, m/s and the Strain Ratio were not consistent between the institutions and had poor results on statistical testing which is why it was not possible to establish a benign/malignant cut-off point for this technique. **Table 7** shows a summary of the distribution of AS 2.0, Bethesda categories, assigned scores, additional scores and recommended conduct for TN management in terms of follow up, active vigilance or a FNAB.

There was no interobserver analysis due to the difficulties of a multicenter study in several developing countries. However, the magnitude of the differences for false positives and negatives of each country with respect to the total was calculated, resulting in a coincidence for false positives of 0.983 and 0.943 for false negatives, showing that there were no significant differences amongst institutions.

4. Discussion

To the best of our knowledge, in Latin America, the use of an ultrasound TN scoring system for malignancy prediction, applied exclusively in our populations, has not been performed until now. We have conducted this prospective study in 10 Latin-American institutions using the AS and validating it by comparing the results with ACRT and Bethesda with a correlation of 0.77 and 0.91, respectively. From the results obtained, we present a new version (AS 2.0), improved by robust statistical analysis in order to include only relevant TUMPs as explained in the methodology section. The new AS 2.0 was thus created by using the predictors that obtained the best statistical results HE, SC, IB, MC, TTW, achieving significant statistical values (Se 82%, Spe 74%, PPV 94%, NPV 51%) with an accuracy of 81% in comparison with original AS values: Se 47.6% and Sp of 98.1% (Mena *et al.* 2018) [12]. In comparison, ACRT reports a Se between 75% to 97% and a Spe of 53% to 67% [22] [23]. Furthermore, our model also grants an additional point to the final AS 2.0 score, in case of a solid TN with a diameter superior to 1.5 cm, in similar fashion as EUTIRADS [5] and KTIRADS [24], (**Figure 1, Figure 2**). This methodology has been used in previous studies showing that it results in a more efficient stratification tool [24] [25].

Table 6. Alpha Score 2.0 ROC of predictors.

Tested variables	ROC				
	Area	Deviation error ^a	Asymptotic significance ^b	95% Asymptotic confidence interval	
				Inferior limit	Superior limit
Hypoechoic	0.658	0.027	0.000	0.605	0.710
Solid consistency	0.728	0.031	0.000	0.668	0.789
Irregular borders	0.850	0.024	0.000	0.804	0.896
Microcalcifications	0.848	0.024	0.000	0.801	0.896
Shape taller than wide	0.779	0.029	0.000	0.721	0.837
Diameter = or >1.5 cm	0.500	0.046		0.411	0.589

Notes: ^aUnder the nonparametric assumption; ^bNull hypothesis: true area = 0.5.

Table 7. Alpha Score 2.0 categories compared with Bethesda and recommended score-based action.

Thyroid nodule: score and summation of the predictor	Benign	Low malignancy probability	Moderate malignancy probability	High malignancy probability
Hypoechoic	1			
Predominantly solid	2			
Irregular borders	2			
Shape taller than wide	2	0 - 2	3	4 - 6
Microcalcifications	2			
Diameter equal to or greater than 1.5 cm	1			
Malignant results	1.9%	8.7%	13.6%	75.7%
Recommended action	Habitual follow up	Active vigilance	Recommended FNAB	Mandatory FNAB

A recent guideline focusing on the state of the art of TN Ultrasound published in 2019 [22], mentions ACTR as the main scoring system; nonetheless, it appears to have no clear advantages of use when compared to other available scores such as EU-TIRADS, K-TIRADS, AACE/ACE/AME, [26] and ATA [27] [28]. Another investigation validated ACRT in comparison to the “Web-based malignancy risk stratification system (WMRS)” score and the Korean score KSThR, finding adequate predictive values between the 3, with WMRS having an advantage over the others [29]. Similar results were reported in another prospective study, published in 2019, comparing AACE/ACE/AME, ATA, ACRT and EU-TIRADS, with ACRT achieving the best prediction rates as it classified more than half of the FNABs as unnecessary with a NPV of 97.8% [30]. Furthermore, in May 2020, a meta-analysis comparing 5 ultrasound risk stratification systems also favored the use of ACRT as it demonstrated better performance by properly selecting TN that required FNAB [31]. Additionally, a paper reporting the institutional experience with ACRT, from Saudi Arabia, retrospectively estimated Se, Sp, PPV, and NPV of 75%, 62.35%, 15.7% and 96.3%, respectively; Sp and PPV are low,

which predisposes to performing less FNABs which might be ACRT's objective, but with the risk of misdiagnosing malignant TN in the process [32]. Finally, in March 2020, a systematic review found 66 articles, none of them reporting or using new scales, only ACRT or ATA [33]. Results of this analysis showed variability in the specificity of ACRT and concludes that the correct classification of TNs as negative (non-malignant) is greatly dependent upon operator skill [33].

In our experience, AS 2.0 offers an easy-to-use classification, quick to calculate and with similar effectiveness as ACRT which encompasses 5 evaluation groups conformed by 17 ultrasonographical parameters used to stratify the nodule in one of 5 categories (T1 - T5); this results in a very detailed classification, but that results in a steeper learning curve with more time invested in both training and performing the ultrasound classification. In comparison, AS 2.0 only includes 5 dichotomic TUMPs, which might result in faster training and quick stratification, something particularly desirable in the healthcare system of developing countries.

Our study shows male predominance of malignant TNs, which differs from reported data in the literature, however our general prevalence of TNs coincides with other studies that report a 4:1 and 2 - 3:1 ratio between females and males [34] [35] [36], we have a proportion of 5:1, similar to what is established internationally [37] [38].

Table 3 and **Table 7** summarizes the correlation between AS 2.0 and Bethesda, indicating that only 1.9% of malignant TNs are misclassified on the benign category, which translates to an optimal statistical certainty for choosing not to perform an FNAB (79.5% benignancy certainty). Furthermore, only 8.7% of malignant TNs are classified as low suspicion of malignancy which will result in active vigilance of the nodule, thus avoiding premature FNABs and diminishing patient's anxiety. In contrast, the moderate suspicion of malignancy category included malignant TNs in 13.6% of cases, and benign TNs in only 4.9%, which is why FNAB is recommended in this category as well as in the high suspicion of malignancy category where only 2.3% of nodules were benign and 75.7% were indeed malignant thus mandating confirmatory FNAB. These results are associated with a sensitivity of 82%, specificity of 74% and PPV of 94%, suggesting that AS 2.0 is a reliable classification system.

Although the use of elastography is not the focus of our research, there are many publications like the ones from the World Federation for Ultrasound in Medicine and Biology (WFUMB) [39], EU-TIRADS [5] and Remonti *et al.*, that report the use of ultrasound elastography in TNs, but variability in the results mainly due to ultrasound equipment differences, make it difficult to standardize exact values in order to include elastography as a TUMP with statistical validation [40]. EU-TIRADS includes elastography but without specifying the type of elastography or the cut-off values that should be used [31]. In our study the results of the three institutions that performed elastography on TNs were not statistically conclusive, so they were not included in this study. However, in our

experience, thyroid tissue has similar values between point shear wave and 2D shear wave (**Figure 3**, **Figure 4** Panel C and Panel D), we also use strain elastography with values of tissue deformity and strain ratios and we only use elastography as an additional informative tool and not as a proper ultrasonographic malignancy predictor classification system.

Limitations

AS 2.0 is not designed to be used in diffuse thyroid lesions nor to classify a nodule of mixed predominance or an atypical cystic or atypical calcified lesion in detail, for all of these lesions any current published classification such as ACRT that manages those cases, should be used. It was not possible to calculate interobserver agreement due to the difficulties inherent of multicenter study involving several developing countries. However, the differences of false positives and negatives of each institution were calculated, finding that there were no significant differences (see results). There were no histopathological studies performed on the TNs of this study because we used Bethesda classification and in some of those categories histopathological study is not required.

5. Conclusion

Results from this Latin-American multicentric study indicated that AS 2.0 has superior diagnostic accuracy and performance compared to the previously published AS. AS 2.0 is able to classify a benign TN in a precise, safe and accurate way, avoiding unnecessary FNABs, and aids in the decision of active vigilance without FNAB in those TNs classified as low suspicion for malignancy or determines the necessity of puncturing those nodes with moderate to high suspicion of malignancy. AS 2.0 has a good correlation with ACRT and Bethesda, two well-known TN stratification systems.

Acknowledgements

We would like to thank the following persons for their contributions to this study:

Kelly Tafur, Candy Valencia, Oscar Jaramillo, Christian Palacios, Karla Garay, Francisco Suárez, Lorena Moreno, Lucas Di Pasquale, Jenny Sosa, Sonia Araujo, Yusmary El Kantar, Carola Jácome, Clemencia Escobar, Jhoanna Arévalo, Karina Gangotena, Mayra Álava, Ximena Cuzco, Sergio Valencia, and Carlos Regalado.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Machens, A., Holzhausen, H.J. and Dralle, H. (2005) The Prognostic Value of Primary Tumor Size in Papillary and Follicular Thyroid Carcinoma: A Comparative

- Analysis. *Cancer*, **103**, 2269-2273. <https://doi.org/10.1002/cncr.21055>
- [2] Miyauchi, A. and Ito, Y. (2019) Conservative Surveillance Management of Low-Risk Papillary Thyroid Microcarcinoma. *Endocrinology and Metabolism Clinics of North America*, **48**, 215-226. <https://doi.org/10.1016/j.ecl.2018.10.007>
- [3] Horvath, E., Majlis, S., Rossi, R., Franco, C., Niedmann, J.P., Castro, A. and Dominguez, M. (2009) An Ultrasonogram Reporting System for Thyroid Nodules Stratifying Cancer Risk for Clinical Management. *Journal of Clinical Endocrinology & Metabolism*, **94**, 1748-1751. <https://doi.org/10.1210/jc.2008-1724>
- [4] Tessler, F.N., Middleton, W.D., Grant, E.G., Hoang, J.K., Berland, L.L., Teefey, S.A., et al. (2017) ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. *Journal of the American College of Radiology*, **14**, 587-595. <https://doi.org/10.1016/j.jacr.2017.01.046>
- [5] Russ, G., Bonnema, S.J., Erdogan, M.F., Durante, C., Ngu, R. and Leenhardt, L. (2017) European Thyroid Association Guidelines for Ultrasound Malignancy Risk Stratification of Thyroid Nodules in Adults: The EU-TIRADS. *European Thyroid Journal*, **6**, 225-237. <https://doi.org/10.1159/000478927>
- [6] Migda, B., Migda, M., Migda, M.S. and Slapa, R.Z. (2018) Use of the Kwak Thyroid Image Reporting and Data System (K-TIRADS) in Differential Diagnosis of Thyroid Nodules: Systematic Review and Meta-Analysis. *European Radiology*, **28**, 2380-2388. <https://doi.org/10.1007/s00330-017-5230-0>
- [7] Sánchez, J.F. (2014) Clasificación TI-RADS de los nódulos tiroideos en base a una escala de puntuación modificada con respecto a los criterios ecográficos de malignidad. *Revista Argentina de Radiología*, **78**, 138-148. <https://doi.org/10.1016/j.rard.2014.07.015>
- [8] Bailey, S. and Wallwork, B. (2018) Differentiating between Benign and Malignant Thyroid Nodules. *Australian Journal for General Practitioners*, **47**, 770-774. <https://doi.org/10.31128/AJGP-03-18-4518>
- [9] Seo, H., Na, D.G., Kim, J.H., Kim, K.W. and Yoon, J.W. (2015) Ultrasound-Based Risk Stratification for Malignancy in Thyroid Nodules: A Four-Tier Categorization System. *European Radiology*, **25**, 2153-2162. <https://doi.org/10.1007/s00330-015-3621-7>
- [10] Cibas, E.S. and Ali, S.Z. (2017) The 2017 Bethesda System for Reporting Thyroid Cytopathology. *Thyroid*, **27**, 1341-1346. <https://doi.org/10.1089/thy.2017.0500>
- [11] Magister, M.J., Chaikhouidinov, I., Schaefer, E., Williams, N., Saunders, B. and Goldenberg, D. (2015) Association of Thyroid Nodule Size and Bethesda Class with Rate of Malignant Disease. *JAMA Otolaryngology—Head & Neck Surgery*, **141**, 1089-1095. <https://doi.org/10.1001/jamaoto.2015.1451>
- [12] Mena, G., Benavides, R., Villagomez, R., Muñoz, M., Bucheli, S.M., Mariela, M. and Rosa, G. (2018) Assessment of Malignancy Risk in Thyroid Nodules Using a Practical Ultrasound Predictor Model: “Alpha Score”. *Open Journal of Radiology*, **8**, 191-202. <https://doi.org/10.4236/ojrad.2018.84022>
- [13] Maurandi-López, A., Del Río Alonso, L., González-Vidal, A., Jaén, M.E.F. and Vicente, Á.H. (2019) Fundamentos Estadísticos para Investigación. Introducción a R y Modelos. *Zenodo*. <https://doi.org/10.5281/zenodo.2628915>
- [14] Sassaroli, E., Crake, C., Scorza, A., Kim, D.S. and Park, M.A. (2019) Image Quality Evaluation of Ultrasound Imaging Systems: Advanced B-Modes. *Journal of Applied Clinical Medical Physics*, **20**, 115-124. <https://doi.org/10.1002/acm2.12544>
- [15] Brendel, K., Filipczynski, L.S., Gerstner, R., Hill, C.R., Kossoff, G., Quentin, G., et al. (1977) Methods of Measuring the Performance of Ultrasonic Pulse-Echo Diagnostic Equipment. *Ultrasound in Medicine and Biology*, **2**, 343-350.

- [https://doi.org/10.1016/0301-5629\(77\)90040-0](https://doi.org/10.1016/0301-5629(77)90040-0)
- [16] AIUM Technical Standards Committee (1995) AIUM Quality Assurance Manual for Gray-scale Ultrasound Scanners: Stage 2—AIUM Ytechnical Standards Committee. American Institute of Ultrasound in Medicine, Laurel.
- [17] del Campo, N.M.S. and Matamoros, L.Z. (2019) Evolución histórica de las técnicas estadísticas y las metodologías para el estudio de la causalidad en ciencias médicas. *Medisan*, **23**, 534-556.
- [18] von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C. and Vandembroucke, J.P. (2008) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *Journal of Clinical Epidemiology*, **61**, 344-349.
<https://doi.org/10.1016/j.jclinepi.2007.11.008>
- [19] Valle, A.R. (2017) Curvas ROC (Receiver-Operating-Characteristic) y sus aplicaciones. (Trabajo Fin de Grado Inédito). Universidad de Sevilla, Sevilla.
- [20] Espino, C. (2017) Análisis predictivo: Técnicas y modelos utilizados y aplicaciones del mismo—herramientas Open Source que permiten su uso. Universidad Oberta de Catalunya, Barcelona.
- [21] López-Roldán, P. and Fachelli, S. (2015) Metodología de la investigación social cuantitativa. Universitat Autònoma de Barcelona, Barcelona.
- [22] Floridi, C., Cellina, M., Buccimazza, G., Arrichiello, A., Sacrini, A., Arrigoni, F., Pompili, G., Barile, A. and Carrafiello, G. (2019) Ultrasound Imaging Classifications of Thyroid Nodules for Malignancy Risk Stratification and Clinical Management: State of the Art. *Gland Surgery*, **8**, S233-S244.
<https://doi.org/10.21037/gs.2019.07.01>
- [23] Hoang, J.K., Langer, J.E., Middleton, W.D., Wu, C.C., Hammers, L.W., Cronan, J.J., Tessler, F.N., Grant, E.G. and Berland, L.L. (2015) Managing Incidental Thyroid Nodules Detected on Imaging: White Paper of the ACR Incidental Thyroid Findings Committee. *Journal of the American College of Radiology*, **12**, 143-150.
<https://doi.org/10.1016/j.jacr.2014.09.038>
- [24] Phuttharak, W., Boonrod, A., Klungboonkrong, V. and Witsawapaisan, T. (2019) Interrater Reliability of Various Thyroid Imaging Reporting and Data System (TIRADS) Classifications for Differentiating Benign from Malignant Thyroid Nodules. *Asian Pacific Journal of Cancer Prevention*, **20**, 1283-1288.
<https://doi.org/10.31557/APJCP.2019.20.4.1283>
- [25] Shen, Y., Liu, M., He, J., Wu, S., Chen, M., Wan, Y., Gao, L., Cai, X., Ding, J. and Fu, X. (2019) Comparison of Different Risk-Stratification Systems for the Diagnosis of Benign and Malignant Thyroid Nodules. *Frontiers in Oncology*, **9**, Article No. 378.
<https://doi.org/10.3389/fonc.2019.00378>
- [26] Gharib, H., Papini, E., Garber, J.R., Duick, D.S., Harrell, R.M., Hegedüs, L., *et al.* (2016) 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. *Endocrine Practice*, **22**, 1-60. <https://doi.org/10.4158/EP161208.GL>
- [27] Haugen, B.R., Alexander, E.K., Bible, K.C., Doherty, G.M., Mandel, S.J., Nikiforov, Y.E., *et al.* (2016) 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*, **26**, 1-133.
<https://doi.org/10.1089/thy.2015.0020>
- [28] Frates, M.C., Benson, C.B., Charboneau, J.W., Cibas, E.S., Clark, O.H., Coleman,

- B.G., et al. (2006) Management of Thyroid Nodules Detected at US: SOCIETY of Radiologists in Ultrasound Consensus Conference Statement. *Ultrasound Quarterly*, **22**, 231-238. <https://doi.org/10.1097/01.ruq.0000226877.19937.a1>
- [29] Ha, S.M., Ahn, H.S., Baek, J.H., Ahn, H.Y., Chung, Y.J., Cho, B.Y. and Park, S.B. (2017) Validation of Three Scoring Risk-Stratification Models for Thyroid Nodules. *Thyroid*, **27**, 1550-1557. <https://doi.org/10.1089/thy.2017.0363>
- [30] Grani, G., Lamartina, L., Ascoli, V., Bosco, D., Biffoni, M., Giacomelli, L., et al. (2019) Reducing the Number of Unnecessary Thyroid Biopsies While Improving Diagnostic Accuracy: Toward the “Right” TIRADS. *Journal of Clinical Endocrinology & Metabolism*, **104**, 95-102. <https://doi.org/10.1210/jc.2018-01674>
- [31] Castellana, M., Castellana, C., Treglia, G., Giorgino, F., Giovannella, L., Russ, G. and Trimboli, P. (2019) Performance of Five Ultrasound Risk Stratification Systems in Selecting Thyroid Nodules for FNA. A Meta-Analysis. *Journal of Clinical Endocrinology & Metabolism*, **105**, 1659-1669. <https://doi.org/10.1210/clinem/dgz170>
- [32] Ewid, N.M., Alamer, A., El Saka, H., Alduraibi, S., AlGoblan, A. and Ahmed, E. (2019) Updated ACR Thyroid Imaging Reporting and Data Systems in Risk Stratification of Thyroid Nodules: 1-Year Experience at a Tertiary Care Hospital in Al-Qassim. *The Egyptian Journal of Internal Medicine*, **31**, 868-873.
- [33] Mistry, R., Hillyar, C., Nibber, A., Sooriyamoorthy, T. and Kumar, N. (2020) Ultrasound Classification of Thyroid Nodules: A Systematic Review. *Cureus*, **12**, Article No. e7239. <https://doi.org/10.7759/cureus.7239>
- [34] Heller, M.T., Gilbert, C., Otori, N.P. and Tublin, M.E. (2013) Correlation of Ultrasound Findings with the Bethesda Cytopathology Classification for Thyroid nodule Fine-Needle Aspiration: A Primer for Radiologists. *American Journal of Roentgenology*, **201**, W487-W494. <https://doi.org/10.2214/AJR.12.9071>
- [35] Nachiappan, A.C., Metwalli, Z.A., Hailey, B.S., Patel, R.A., Ostrowski, M.L. and Wynne, D.M. (2014) The Thyroid: Review of Imaging Features and Biopsy Techniques with Radiologic-Pathologic Correlation. *RadioGraphics*, **34**, 276-293. <https://doi.org/10.1148/rg.342135067>
- [36] Russ, G., Leboulleux, S., Leenhardt, L. and Hegedüs, L. (2014) Thyroid Incidentalomas: Epidemiology, Risk Stratification with Ultrasound and Workup. *European Thyroid Journal*, **3**, 154-163. <https://doi.org/10.1159/000365289>
- [37] Kim, E.K., Cheong, S.P., Woung, Y.C., Ki, K.O., Dong, I.K., Jong, T.L. and Hyung, S.Y. (2002) New Sonographic Criteria for Recommending Fine-Needle Aspiration Biopsy of Nonpalpable Solid Nodules of the Thyroid. *American Journal of Roentgenology*, **178**, 687-691. <https://doi.org/10.2214/ajr.178.3.1780687>
- [38] Theoharis, C.G.A., Schofield, K.M., Hammers, L., Udelsman, R. and Chhieng, D.C. (2009) The Bethesda Thyroid Fine-Needle Aspiration Classification System: Year 1 at an Academic Institution. *Thyroid*, **19**, 1215-1223. <https://doi.org/10.1089/thy.2009.0155>
- [39] Cosgrove, D., Barr, R., Bojunga, J., Cantisani, V., Chammas, M.C., Dighe, M., Vinayak, S., Xu, J.M. and Dietrich, C.F. (2017) WFUMB Guidelines and Recommendations on the Clinical Use of Ultrasound Elastography: Part 4. Thyroid. *Ultrasound in Medicine and Biology*, **43**, 4-26. <https://doi.org/10.1016/j.ultrasmedbio.2016.06.022>
- [40] Remonti, L.R., Kramer, C.K., Leitão, C.B., Pinto, L.C.F. and Gross, J.L. (2015) Thyroid Ultrasound Features and Risk of Carcinoma: A Systematic Review and Meta-Analysis of Observational Studies. *Thyroid*, **25**, 538-550. <https://doi.org/10.1089/thy.2014.0353>