v7.n11.2026

SCIENTIFIC JOURNAL - RECIMA21 ISSN 2675-6218

A CONTINUOUS MULTIVARIABLE MODEL COMBINING SERUM TSH AND ACR TI-RADS FOR PRE-BIOPSY RISK STRATIFICATION OF THYROID NODULES: DEVELOPMENT AND INTERNAL VALIDATION

UM MODELO MULTIVARIÁVEL CONTÍNUO QUE COMBINA TSH SÉRICO E ACR TI-RADS PARA ESTRATIFICAÇÃO DE RISCO PRÉ-BIÓPSIA DE NÓDULOS TIREOIDIANOS: DESENVOLVIMENTO E VALIDAÇÃO INTERNA

UN MODELO MULTIVARIABLE CONTINUO QUE COMBINA TSH SÉRICA Y ACR TI-RADS PARA LA ESTRATIFICACIÓN DEL RIESGO PREBIOPSIA DE NÓDULOS TIROIDEOS: DESARROLLO Y VALIDACIÓN INTERNA

Diego Batista de Montalvão Cunha[1], Ruan Everton de Souza Silva[2], Josie Haydée Lima Ferreira[3]

e7119356
https://doi.org/10.47820/recima21.v7i11.9356

PUBLISHED: 11/2026

ABSTRACT

Objective: To determine the incremental diagnostic value of serum thyroid-stimulating hormone (TSH) beyond ACR TI-RADS for pre-biopsy risk stratification of thyroid nodules, comparing binary integration strategies with a continuous multivariable risk model. Methods: This retrospective study included 172 adults who underwent thyroidectomy (2021–2024) at two Brazilian hospitals. The optimal TSH cutoff was determined by ROC analysis. A continuous multivariable model (TSH, ACR TI-RADS category, age, sex, and nodule size on ultrasound) was internally validated by bootstrapping and compared, using decision curve analysis, with the ACR TI-RADS FNA criteria alone, TSH alone, and two binary TSH–ACR TI-RADS strategies. Results: Malignancy prevalence was 47.1%. Median TSH was higher in patients with malignant nodules (1.74 vs. 1.31 mIU/L; p<0.01). TSH alone yielded an AUC of 0.652 (95% CI 0.571–0.734; p<0.01), with an optimal cutoff of 1.545 mIU/L. TSH was an independent predictor of malignancy (OR=1.67 per mIU/L; 95% CI 1.13–2.47; p=0.01). The binary strategies increased sensitivity from 71.6% to 84.0–90.1%, at the cost of lower specificity. The continuous model showed the highest optimism-corrected net benefit at threshold probabilities above approximately 10% (corrected AUC 0.829; calibration slope 0.83). Conclusions: In this surgical cohort, TSH provided incremental, low-cost diagnostic value beyond ACR TI-RADS, particularly when incorporated into a continuous multivariable model. External validation is warranted before clinical use.

KEYWORDS: *Thyroid Neoplasms. Thyrotropin. Ultrasonography. Biopsy, Fine-Needle. Sensitivity and Specificity.*

RESUMO

Objetivo: Determinar o valor diagnóstico incremental do hormônio tireoestimulante (TSH) sérico em relação ao ACR TI-RADS na estratificação de risco pré-biópsia de nódulos tireoidianos, comparando estratégias binárias de integração com um modelo de risco multivariável contínuo. Métodos: Este estudo retrospectivo incluiu 172 adultos submetidos a tireoidectomia (2021–2024) em dois hospitais brasileiros. O ponto de corte ideal do TSH foi determinado por análise ROC.

[1] Federal University of Piauí (UFPI), specialist in Internal Medicine and in Endocrinology and Metabolism; MSc student in Sciences and Health.
[2] Federal University of Piauí (UFPI), nurse; MSc student in Sciences and Health.
[3] Federal University of Piauí (UFPI), PhD, Associate Professor in Parasitology and Microbiology.

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

Um modelo multivariável contínuo (TSH, categoria ACR TI-RADS, idade, sexo e tamanho do nódulo à ultrassonografia) foi validado internamente por *bootstrapping* e comparado, por meio de análise de curva de decisão, com os critérios de PAAF do ACR TI-RADS isoladamente, o TSH isoladamente e duas estratégias binárias TSH–ACR TI-RADS. Resultados: A prevalência de malignidade foi de 47,1%. A mediana do TSH foi maior nos pacientes com nódulos malignos (1,74 vs. 1,31 mUI/L; p<0,01). O TSH isoladamente apresentou AUC de 0,652 (IC 95% 0,571–0,734; p<0,01), com ponto de corte ideal de 1,545 mUI/L. O TSH foi preditor independente de malignidade (OR=1,67 por mUI/L; IC 95% 1,13–2,47; p=0,01). As estratégias binárias aumentaram a sensibilidade de 71,6% para 84,0–90,1%, à custa de menor especificidade. O modelo contínuo apresentou o maior benefício líquido corrigido para otimismo em probabilidades-limiar acima de aproximadamente 10% (AUC corrigida de 0,829; inclinação de calibração de 0,83). Conclusões: Nesta coorte cirúrgica, o TSH agregou valor diagnóstico incremental e de baixo custo ao ACR TI-RADS, sobretudo quando incorporado a um modelo multivariável contínuo. Validação externa é necessária antes do uso clínico.

PALAVRAS-CHAVE: Neoplasias da Glândula Tireoide. Tireotropina. Ultrassonografia. Biópsia por Agulha Fina. Sensibilidade e Especificidade.

RESUMEN

Objetivo: Determinar el valor diagnóstico incremental de la hormona estimulante de la tiroides (TSH) sérica respecto al ACR TI-RADS en la estratificación del riesgo prebiopsia de nódulos tiroideos, comparando estrategias binarias de integración con un modelo de riesgo multivariable continuo. Métodos: Estudio retrospectivo de 172 adultos sometidos a tiroidectomía (2021–2024) en dos hospitales brasileños. El punto de corte óptimo de la TSH se determinó mediante análisis ROC. Un modelo multivariable continuo (TSH, categoría ACR TI-RADS, edad, sexo y tamaño del nódulo en la ecografía) se validó internamente mediante bootstrapping y se comparó, por análisis de curvas de decisión, con los criterios de PAAF del ACR TI-RADS solos, la TSH sola y dos estrategias binarias TSH–ACR TI-RADS. Resultados: La prevalencia de malignidad fue del 47,1%. La mediana de la TSH fue mayor en pacientes con nódulos malignos (1,74 vs. 1,31 mUI/L; p<0,01). La TSH sola mostró AUC de 0,652 (IC 95% 0,571–0,734; p<0,01), con punto de corte óptimo de 1,545 mUI/L. La TSH fue predictor independiente de malignidad (OR=1,67 por mUI/L; IC 95% 1,13–2,47; p=0,01). Las estrategias binarias aumentaron la sensibilidad del 71,6% al 84,0–90,1%, a costa de menor especificidad. El modelo continuo mostró el mayor beneficio neto corregido por optimismo en probabilidades umbral superiores al 10% aproximadamente (AUC corregido 0,829; pendiente de calibración 0,83). Conclusiones: En esta cohorte quirúrgica, la TSH aportó valor diagnóstico incremental y de bajo costo al ACR TI-RADS, especialmente al incorporarse a un modelo multivariable continuo. Se requiere validación externa antes de su uso clínico.

PALABRAS CLAVE: *Neoplasias de la Tiroides. Tirotropina. Ultrasonografía. Biopsia con Aguja Fina. Sensibilidad y Especificidad.*

1. INTRODUCTION

Thyroid nodules are among the most common findings in contemporary endocrine practice, occurring in up to 60% of adults worldwide, although only 1–5% are malignant [1,2]. Despite this low malignancy rate, an estimated 2,290 deaths from thyroid cancer occur annually in the United States alone [3], and global mortality is projected to nearly double by 2050 [4].

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

Clinical management of thyroid nodules must therefore balance the need to avoid unnecessary invasive procedures in low-risk lesions against the early detection of clinically significant cancers.

Ultrasound is the imaging modality of choice for initial evaluation of thyroid nodules, and the TI-RADS system [5,6] is currently the most widely used sonographic system guiding decisions about fine-needle aspiration (FNA). However, substantial overlap between benign and malignant sonographic patterns limits its diagnostic performance. A meta-analysis of 18,750 nodules reported a pooled sensitivity of 74% and specificity of 64% for ACR TI-RADS, with a positive predictive value of only 43% [7]. Thus, more than half of the nodules selected for FNA under this system are ultimately benign, motivating the search for markers that could add incremental diagnostic value without requiring an invasive procedure.

Serum TSH is biologically plausible as such a marker, as it is the principal regulator of thyroid follicular cell proliferation and differentiation, exerting a mitogenic effect that has been mechanistically linked to thyroid tumorigenesis [8]. Moreover, serum TSH is an inexpensive, widely available laboratory test and is recommended as part of the initial evaluation of all patients presenting with thyroid nodules [9]. Multiple studies and meta-analyses have reported higher TSH levels among patients with malignant nodules [10,11], yet the optimal cutoff varies substantially across populations [12-14], and studies evaluating TSH together with ACR TI-RADS remain scarce, particularly in the Brazilian population [15].

When investigators have combined TSH with ACR TI-RADS, they have done so almost exclusively by dichotomizing TSH at a single diagnostic cutoff and applying a simple rule on top of the existing category [16]. This approach is intuitive and easy to implement at the bedside, but it discards the graded, dose-dependent information a continuous biomarker carries and is known to lose statistical power relative to modeling the biomarker as a continuous covariate [17]. Other studies have combined TSH, whether dichotomized [18] or continuous [19], with individual sonographic features rather than the ACR TI-RADS category itself. The present study extends this literature by modeling TSH as a continuous covariate within a multivariable model alongside the ACR TI-RADS category, and by directly comparing binary and continuous strategies for integrating TSH with ACR TI-RADS within the same cohort.

This study aimed to develop and internally validate a continuous multivariable model integrating serum TSH with ACR TI-RADS for pre-biopsy risk stratification of thyroid nodules, and to compare its performance with that of the ACR TI-RADS FNA criteria alone, TSH alone, and simpler binary TSH–ACR TI-RADS integration strategies in a Brazilian tertiary-care surgical cohort.

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

2. MATERIALS AND METHODS

Design

This retrospective diagnostic accuracy study included patients who underwent thyroidectomy between 2021 and 2024 at two affiliated tertiary hospitals of the same healthcare group (Hospital Prontomed and Hospital Santa Maria, both of Grupo Medimagem Ltda, Teresina, Piauí, Brazil), which share the same radiology group and a single centralized reader. All ultrasound images were retrospectively re-reviewed by a single experienced radiologist, blinded to histopathological outcome, to confirm ACR TI-RADS 2017 classification [5].

Population

Eligible patients were adults (18–90 years) with a thyroid nodule who underwent thyroidectomy for any indication and had all of the following available: preoperative serum TSH (measured within 6 months before surgery), ultrasound images for ACR TI-RADS re-review, Bethesda cytology, and surgical histopathology. Patients with prior thyroid surgery, known thyroid dysfunction, or use of medications that alter TSH were excluded.

Patients were identified from a list of all thyroidectomies, retrieved by procedure code from the Brazilian Unified Terminology for Supplementary Health (TUSS) and provided by the institutional information technology department (n=1,006). The corresponding medical records were retrieved using the encounter number and reviewed individually by the first author, and 172 patients were ultimately analyzed (Figure 1). Patients were excluded at screening when the record lacked a documented FNA with Bethesda cytology, a documented ultrasound (examinations performed at external services were frequently not filed in the record), or a preoperative TSH measured at the institution, or when thyroid dysfunction was present. Reasons for exclusion were not tabulated individually; the approximate proportions, estimated from the record review, were as follows: no documented FNA (~50%), no documented ultrasound/ACR TI-RADS classification (~30%), thyroid dysfunction (~10%), and no TSH measured at the institution (~10%). Because patients with missing data on any predictor or outcome variable were excluded during cohort assembly, all analyses were complete-case analyses.

The unit of analysis was the patient, represented by one index nodule, defined as the most suspicious nodule on ultrasound. Assuming an expected AUC of 0.55 for TSH, as previously reported [20], a malignancy prevalence of 33.3%, a 95% confidence level, and a

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

precision of 0.10, the minimum required sample size was 138 patients. This calculation was based on TSH as a standalone diagnostic marker.

Variables

TSH (reference range 0.4–4.5 mIU/L) was measured by electrochemiluminescence immunoassay. At the participating hospitals, serum TSH is measured in the initial work-up of every patient with a thyroid nodule, in line with the recommendations of the American Thyroid Association and the European Thyroid Association [6,9], because a subnormal value often redirects the evaluation to thyroid scintigraphy instead of FNA. The TSH value analyzed was the one recorded within 6 months before surgery; its timing relative to FNA was not systematically documented. Because FNA is not expected to alter circulating TSH levels, this value was considered to reflect preoperative thyroid function regardless of its exact timing.

ACR TI-RADS classification followed the 2017 system [5], based on composition, echogenicity, shape, margins, and echogenic foci, with size-based FNA thresholds per category. For each nodule, a binary FNA-indication variable was derived by applying the ACR TI-RADS 2017 size-based thresholds to the assigned category and the ultrasound diameter (TR1–TR2: no FNA regardless of size; TR3: FNA if ≥2.5 cm; TR4: FNA if ≥1.5 cm; TR5: FNA if ≥1.0 cm), independently of the FNA decision actually made in clinical practice. For the multivariable model, ultrasound nodule size was categorized with the same cut points (<1 cm, 1–1.5 cm, 1.5–2.5 cm, or >2.5 cm).

Nodule size was recorded in two ways: size on ultrasound, which is available before biopsy and was used as a predictor in the model, and size on the surgical specimen, which is obtained only after surgery and was used exclusively for descriptive purposes and to assess agreement with ultrasound. Agreement was summarized using Spearman's correlation coefficient, the proportion of nodules in the same size category, and the linearly weighted kappa. Cytology was classified according to the 2023 Bethesda System for Reporting Thyroid Cytopathology [21]. The histopathological result corresponding to the index nodule, matched by laterality and location in the pathology report, was the reference standard for malignancy. Final histopathology was also used for recurrence-risk stratification (low, low-intermediate, high-intermediate, or high) according to the 2025 ATA guidelines [22].

Statistical analysis

Categorical variables were compared between the benign and malignant groups using the chi-square test, and TSH and age were compared using the Mann-Whitney U test and Student's t-test, respectively. The optimal TSH cutoff for predicting malignancy was identified via

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

ROC curve analysis, using the point closest to the top-left corner as the selection criterion. Optimism in the sensitivity and specificity of TSH and of the combined TSH–ACR TI-RADS models was estimated by bootstrap resampling (2,000 replicates, with replacement): the optimal cutoff was re-derived in each replicate, its performance in the resample was compared with its performance in the original sample, and the mean difference (optimism) was subtracted from the apparent estimates. The distribution of the cutoffs obtained across replicates was also recorded. Diagnostic accuracy metrics (sensitivity, specificity, positive/negative predictive values [PPV/NPV], accuracy) were then calculated for TSH, ACR TI-RADS, and Bethesda using the corresponding cutoffs/categories, with the histopathological result as reference standard.

The two binary TSH–ACR TI-RADS integration strategies were compared with the ACR TI-RADS FNA criteria, that is, FNA indication based on the system's size thresholds rather than on the category itself. The Additive model indicated FNA if the nodule met the ACR TI-RADS FNA criteria or if TSH was at or above the cutoff, and the Reclassification model upgraded the nodule by one ACR TI-RADS category when TSH was at or above the cutoff, with the FNA indication recalculated accordingly. A malignancy was considered rescued when FNA was not indicated by the ACR TI-RADS FNA criteria but was indicated by the combined strategy. The proportions of correctly classified nodules across the three strategies were compared using Cochran's Q test, followed by pairwise McNemar tests with Bonferroni correction.

The central analysis of this study was a continuous multivariable model, developed and reported following the TRIPOD guideline [23]. This logistic regression model was fitted by maximum likelihood and included histopathology (malignant vs. benign) as the outcome and age, sex, ultrasound nodule size (categorical), TSH, and ACR TI-RADS category as predictor variables. TSH was included as a continuous covariate, rather than dichotomized at the ROC-derived cutoff, to avoid the loss of statistical information and power associated with categorizing continuous predictors [17]. Linearity of the association between TSH and the log odds of malignancy was assessed with restricted cubic splines (three knots), with no evidence of nonlinearity (likelihood-ratio test, p=0.67). ACR TI-RADS category (TR1–5) entered the model as a single ordinal term (odds ratio per category). A sensitivity analysis treating categories as indicator variables (categories TR1 and TR2 pooled, as only one nodule was category TR1) did not significantly improve fit (likelihood-ratio test, p=0.40). This same continuous risk model also served as the input for the decision curve analysis below. Bethesda cytology was deliberately excluded. It is available only after FNA has been performed and, in a preliminary model, it showed quasi-complete separation (Bethesda V/VI alone was an almost perfect predictor of malignancy),

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

yielding an implausibly high odds ratio, an extremely wide confidence interval, and unstable estimates for the remaining covariates [24,25]. That model was therefore discarded.

In the decision curve analysis, for each threshold probability (pt) for FNA indication, the net benefit of each strategy was calculated [26]. Across pt=1%–90%, five strategies were compared: ACR TI-RADS FNA criteria, TSH ≥1.545 mIU/L, the Additive model, the Reclassification model, and the multivariable model. Each was compared against the "biopsy all" and "biopsy none" reference strategies. The net benefit of the multivariable model was corrected for optimism by bootstrap, refitting the model in each resample, whereas the other strategies, being fixed rules, are shown with apparent values. The discrimination (AUC), calibration slope, and Brier score of the original model, before any shrinkage, were further evaluated using the same bootstrap optimism-correction procedure, following Harrell et al. [27]. The bootstrap-estimated calibration slope (0.83) was then applied as a uniform shrinkage factor to the regression coefficients, modestly reducing their estimated effects, and the intercept was subsequently re-estimated so that the model's average predicted probability matched the observed proportion of malignant nodules in our cohort (47.1%). The final equation, with the shrunken coefficients and the recalibrated intercept, is reported in full to allow external validation, and the shrunken model was also used to generate the illustrative predicted-risk curves. Calibration of the model before shrinkage was additionally assessed graphically with a bootstrap bias-corrected calibration curve (2,000 resamples).

Statistical significance was set at $p<0.05$. Analyses were performed in R (R Core Team), version 4.4.1.

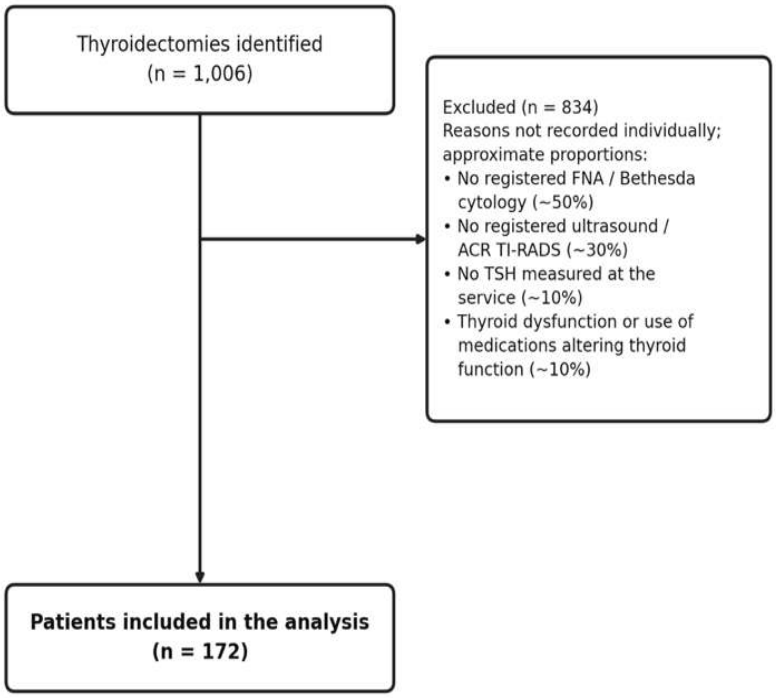
3. RESULTS

Of the 1,006 thyroidectomies identified, 172 patients were included in the analysis (Figure 1). Among the 172 patients (84.3% female; mean age 43.0±12.97 years), histopathology confirmed malignancy in 81 (47.1%) and benign disease in 91 (52.9%) (Table 1). Most nodules were classified as TR4 (45.9%) or TR3 (23.3%), and as Bethesda II (36.0%) or V (27.9%). Among malignant cases, most were low-risk by ATA 2025 stratification (54.3%). Sex and age did not differ between malignant and benign groups (Table 2), while malignant nodules were smaller on ultrasound (<1.5 cm in 46.9% vs. 8.8% of benign nodules, $p<0.01$) and on the surgical specimen, a measurement obtained only after surgery and reported for description only (median 1.20 vs. 2.55 cm, $p<0.01$). Ultrasound and specimen sizes were correlated (Spearman $\rho=0.77$; n=171 with both measurements) but fell in the same size category in 54.4% of nodules, with the ultrasound

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

category higher in 45.0% (weighted kappa 0.53). Median TSH was higher among malignant nodules (1.74 vs. 1.31 mIU/L, p<0.01, Table 2). ROC analysis yielded an AUC of 0.652 (95% CI 0.571–0.734, p<0.01) and identified an optimal cutoff of 1.545 mIU/L (Figure 2).

Figure 1. Flowchart of patient selection



Note: Reasons for exclusion at screening were not recorded individually for each patient; percentages are approximate estimates from the medical record review. FNA: fine-needle aspiration; TSH: thyroid-stimulating hormone.

Source: Authors (2026).

Table 1. Sociodemographic and clinical characterization of the sample (n=172)

Variable	n	%
Sex		
Male	27	15.7
Female	145	84.3
Age (years), mean±SD	43.0±12.97	

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

Variable	n	%
TSH (mIU/L), mean±SD	1.82±1.12	
FNA indication (ACR TI-RADS)		
Yes	123	71.5
No	49	28.5
ACR TI-RADS		
TR1	1	0.6
TR2	17	9.9
TR3	40	23.3
TR4	79	45.9
TR5	35	20.3
Bethesda		
I	15	8.7
II	62	36.0
III	16	9.3
IV	20	11.6
V	48	27.9
VI	11	6.4
Histopathology		
Malignant	81	47.1
Benign	91	52.9
ATA recurrence-risk stratification (malignant cases)		
Low	44	54.3
Low-intermediate	13	16.0
High-intermediate	16	19.8
High	8	9.9

Note: SD: standard deviation.
Source: Authors (2026).

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

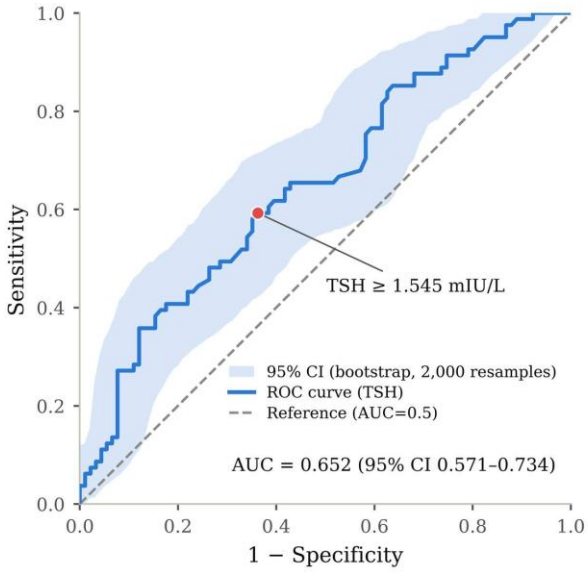
Table 2. Association between sociodemographic, laboratory, and ultrasonographic variables and histopathological outcome

Variable	Malignant (n=81) n (%)	Benign (n=91) n (%)	p-value
Sex			0.90
Male	13 (48.1)	14 (51.9)	
Female	68 (46.9)	77 (53.1)	
Age (years), mean±SD	42.05±12.62	43.91±13.29	0.35
TSH (mIU/L), median (IQR)	1.74 (1.62)	1.31 (1.01)	<0.01
Ultrasound nodule size			<0.01
<1 cm	5 (83.3)	1 (16.7)	
1–1.5 cm	33 (82.5)	7 (17.5)	
1.5–2.5 cm	17 (47.2)	19 (52.8)	
>2.5 cm	26 (28.9)	64 (71.1)	
Nodule size on surgical specimen (cm), median (IQR) – post-surgical, descriptive only	1.20 (0.80)	2.55 (1.90)	<0.01
ACR TI-RADS			<0.01
TR1	0 (0.0)	1 (100)	
TR2	1 (5.9)	16 (94.1)	
TR3	9 (22.5)	31 (77.5)	
TR4	39 (49.4)	40 (50.6)	
TR5	32 (91.4)	3 (8.6)	
FNA indication (ACR TI-RADS)			0.98
Yes	58 (47.2)	65 (52.8)	
No	23 (46.9)	26 (53.1)	

Note: SD: standard deviation, IQR: interquartile range. Ultrasound nodule size (pre-biopsy) is the measurement used as a predictor in the multivariable model. Nodule size on the surgical specimen is the histopathological measurement, obtained only after surgery; it is shown for description only and did not enter any model.
Source: Authors (2026).

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

Figure 2. ROC curve of serum TSH for predicting thyroid malignancy



Note: The shaded band represents the 95% confidence interval obtained from 2,000 bootstrap resamples. The red point indicates the optimal cutoff (TSH ≥1.545 mIU/L, closest-to-top-left criterion). ROC: receiver operating characteristic, AUC: area under the curve; CI: confidence interval.

Source: Authors (2026).

Applied as currently recommended, the ACR TI-RADS FNA criteria showed a sensitivity of 71.6% but a specificity of only 28.6% for FNA indication (Table 3). Combining TSH with ACR TI-RADS improved sensitivity at the cost of specificity (Table 3). The Additive model (ACR TI-RADS-positive or TSH ≥1.545) increased sensitivity from 71.6% to 90.1%, rescuing 15 of 23 malignancies initially missed by the ACR TI-RADS FNA criteria alone, while specificity fell to 20.9% and the proportion of benign nodules with FNA indication increased from 71.4% to 79.1%. The Reclassification model, in which TSH ≥1.545 upgrades the ACR TI-RADS category (to our knowledge not previously described in the literature), reached 84.0% sensitivity (10/23 malignancies rescued), with 76.9% of benign nodules indicated for FNA. Overall correct-classification proportions did not differ significantly across the three strategies (Cochran's Q=4.45, df=2, p=0.11; all pairwise McNemar comparisons p>0.05 after Bonferroni correction). Framed as a trade-off, the Additive and Reclassification models corresponded to approximately one

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

additional benign FNA per 2.1 rescued carcinomas (7 for 15) and per 2.0 rescued carcinomas (5 for 10), respectively.

Table 3. Diagnostic performance of ACR TI-RADS, Bethesda, and TSH-combined models for predicting thyroid malignancy (apparent and bootstrap-corrected)

Test / Model	Sensitivity	Specificity	PPV	NPV	Accuracy
ACR TI-RADS: FNA indication	71.6%	28.6%	47.2%	53.1%	48.8%
Bethesda V or VI	71.6%	98.9%	98.3%	79.6%	86.0%
TSH ≥1.545 mIU/L	59.3% (57.3%)	63.7% (62.1%)	59.3% (57.6%)	63.7% (61.9%)	61.6% (59.8%)
Additive model (ACR TI-RADS + TSH)	90.1% (89.4%)	20.9% (20.4%)	50.3% (50.0%)	70.4% (68.5%)	53.5% (52.9%)
Reclassification model (ACR TI-RADS by TSH)	84.0% (83.0%)	23.1% (22.8%)	49.3% (49.1%)	61.8% (59.9%)	51.7% (51.3%)

Note: PPV: positive predictive value, NPV: negative predictive value. ACR TI-RADS: FNA indication according to the system's size-based criteria. Additive model: FNA indicated if the ACR TI-RADS FNA criteria are met or TSH ≥1.545 mIU/L; Reclassification model: TSH ≥1.545 mIU/L upgrades the nodule by one ACR TI-RADS category, with FNA indication recalculated. Values reflect apparent performance observed in the original sample (n=172); for TSH and the combined models, the bootstrap-corrected value (2,000 resamples, subtracting the estimated mean optimism from the apparent performance) is given in parentheses. The TSH cutoff across resamples had a median of 1.51 mIU/L (95% percentile interval 1.35–2.01); the original 1.545 mIU/L cutoff was therefore estimated with limited precision. Bootstrap correction was not performed for ACR TI-RADS or Bethesda alone, which use fixed, pre-established diagnostic criteria.
Source: Authors (2026).

In the continuous multivariable model, adjustment was made simultaneously for age, sex, ultrasound nodule size, TSH, and ACR TI-RADS category. TSH (continuous variable) remained independently associated with malignancy (OR=1.67 per 1 mIU/L increment; 95% CI 1.13–2.47, p=0.01), as did ACR TI-RADS category (OR=4.37 per category; 95% CI 2.31–8.24, p<0.01) (Table 4). Age was also independently associated with malignancy, though with a small effect size (OR=0.97 per year, 95% CI 0.94–0.998, p=0.04). Sex was not independently associated with malignancy. Ultrasound nodule size, entered as a four-level adjustment covariate, was jointly associated with the outcome (likelihood-ratio test, p=0.003), although individual category estimates were imprecise (Table 4). The model showed good overall discrimination (apparent

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

AUC 0.853), which remained essentially unchanged after bootstrap optimism correction (AUC 0.829; calibration slope 0.83, indicating modest overfitting; Brier score 0.172; all for the model before shrinkage). The bias-corrected calibration curve (Figure 3) showed a tendency to underestimate risk at the lower end and to overestimate it at the higher end, consistent with the slope of 0.83 (mean absolute calibration error of approximately 0.04). The final model after uniform shrinkage is given in Table 4 and by the equation below. To illustrate how predicted risk changes across the range of TSH values observed in our cohort, Figure 4 plots the model's predicted probabilities for a representative patient profile (a 42-year-old woman with a nodule larger than 2.5 cm on ultrasound), across TR3–TR5, the categories in which FNA is potentially indicated. Within TR3, for example, predicted risk rose gradually from 12% to 63% as TSH increased across its observed range.

Table 4. Development of the multivariable prediction model for factors associated with thyroid malignancy

Variable	β (SE)	β after shrinkage	OR	95% CI	p-value
Intercept	−4.572 (2.075)	−3.817	0.01	0.00–0.60	0.03
Age	−0.033 (0.016)	−0.027	0.97	0.94–0.998	0.04
Female sex	−0.436 (0.530)	−0.362	0.65	0.23–1.83	0.41
Ultrasound size 1–1.5 cm	1.132 (1.262)	0.939	3.10	0.26–36.77	0.37
Ultrasound size 1.5–2.5 cm	−0.818 (1.215)	−0.679	0.44	0.04–4.78	0.50
Ultrasound size >2.5 cm	−0.665 (1.220)	−0.552	0.51	0.05–5.62	0.59
TSH (per mIU/L)	0.511 (0.200)	0.424	1.67	1.13–2.47	0.01
ACR TI-RADS (per category)	1.474 (0.324)	1.223	4.37	2.31–8.24	<0.01

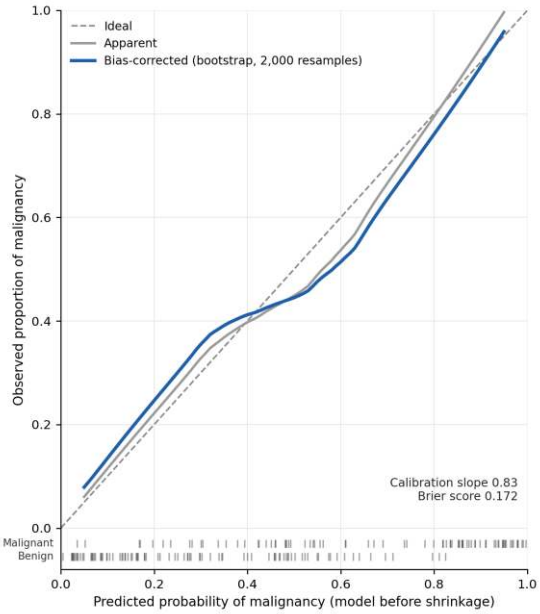
Note: β: regression coefficient, SE: standard error, OR: odds ratio, CI: confidence interval. Reference category for ultrasound nodule size: <1 cm. TSH was entered as a continuous variable (per 1 mIU/L) and ACR TI-RADS as a single ordinal term (OR per category). The joint effect of the three ultrasound nodule-size terms was assessed by likelihood-ratio test (p=0.003). β (SE), OR, 95% CI, and p-values refer to the original, unshrunken model, whose bootstrap-corrected AUC, calibration slope, and Brier score were 0.829, 0.83, and 0.172, respectively. β after

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

shrinkage: coefficients multiplied by the calibration slope (0.83), with the intercept re-estimated so that the mean predicted risk equals the observed prevalence (47.1%). These coefficients are used in the final prediction equation.
Source: Authors (2026).

Final model equation (after shrinkage). The predicted probability of malignancy is p = 1 / (1 + e^(−LP)), where LP = −3.817 − 0.027 × age (years) − 0.362 × [female sex] + 0.939 × [ultrasound size 1–1.5 cm] − 0.679 × [ultrasound size 1.5–2.5 cm] − 0.552 × [ultrasound size >2.5 cm] + 0.424 × TSH (mIU/L) + 1.223 × ACR TI-RADS category (1–5). Bracketed terms equal 1 when the condition is present and 0 otherwise (reference ultrasound size: <1 cm). For example, for a 42-year-old woman with a nodule larger than 2.5 cm on ultrasound, a TSH of 2.0 mIU/L, and ACR TI-RADS category 4, LP = −3.817 − 1.134 − 0.362 − 0.552 + 0.848 + 4.892 = −0.125, corresponding to a predicted probability of malignancy of 47%. The intercept was recalibrated to the prevalence of this surgical cohort (47.1%) and should be re-estimated before application in other populations; the equation is provided to allow external validation and reproducibility and is not intended for clinical use.

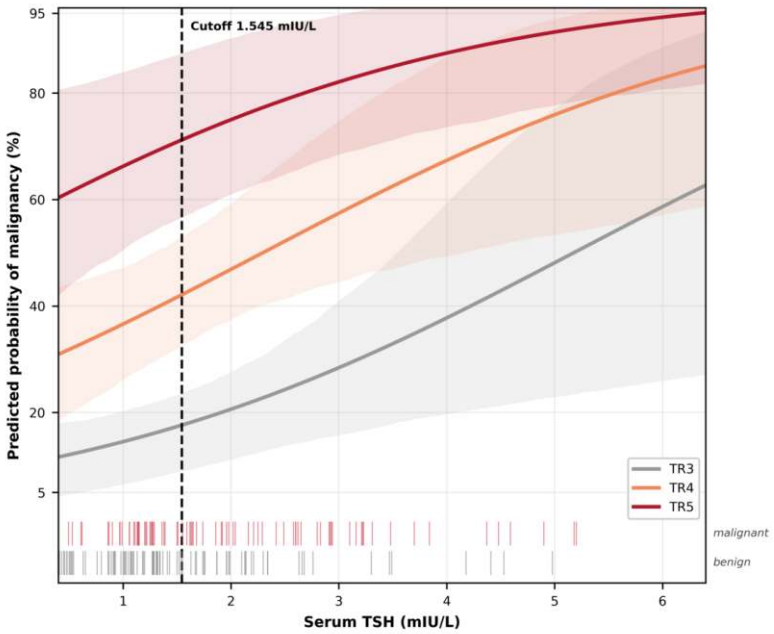
Figure 3. Calibration plot of the multivariable model for thyroid malignancy



Note: Apparent and bootstrap bias-corrected (2,000 resamples) calibration curves of the model before shrinkage; the dashed diagonal indicates perfect calibration. Tick marks along the bottom show the predicted probability of each patient, by histopathological outcome. Calibration slope 0.83; Brier score 0.172.
Source: Authors (2026).

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

Figure 4. Predicted probability of thyroid malignancy as a function of continuous serum TSH, stratified by ACR TI-RADS category (TR3–TR5)



Note: Derived from the multivariable logistic model after shrinkage (coefficients multiplied by the bootstrap-estimated calibration slope, 0.83; intercept re-estimated so that mean predicted risk matched the observed prevalence, 47.1%). Curves reflect a reference profile of 42 years, female sex, and ultrasound nodule size >2.5 cm. Shaded areas indicate the 2.5th–97.5th percentiles of bootstrap refits using the same procedure. Tick marks along the bottom (rug plot) show each patient's observed TSH value, stratified by histopathological outcome. The dashed line marks the cohort's TSH cutoff (1.545 mIU/L).
Source: Authors (2026).

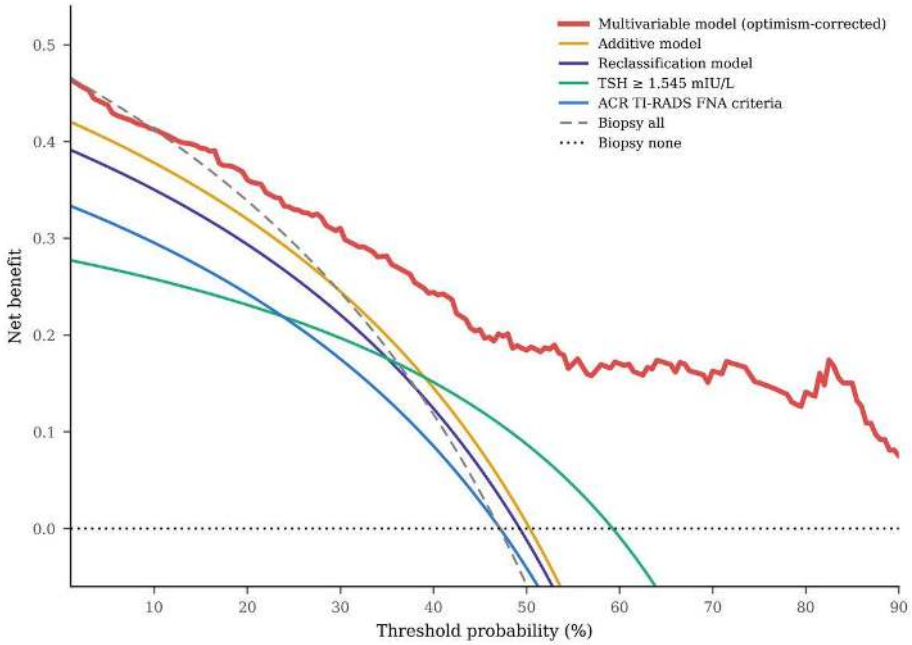
Bootstrap resampling (2,000 replicates), repeating the entire TSH cutoff-selection process in each resample, showed only modest optimism in the performance estimates: TSH's apparent sensitivity/specificity of 59.3%/63.7% were corrected to 57.3%/62.1% (Table 3). This optimism propagated, at similarly modest magnitude, to the combined models: the Additive model's sensitivity was corrected from 90.1% to 89.4%, and the Reclassification model's from 84.0% to 83.0% (Table 3). The resampled cutoff, however, had a median of 1.51 mIU/L (95% percentile interval 1.35–2.01), indicating that the original 1.545 mIU/L cutoff was estimated with limited precision. The correction therefore quantifies the optimism in the performance estimates but does not, by itself, demonstrate that the cutoff is stable.

In the decision curve analysis, across threshold probabilities from 1% to 90%, the optimism-corrected net benefit of the continuous multivariable model (age, sex, ultrasound nodule size, TSH, and ACR TI-RADS category) exceeded that of all other strategies tested, including the

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

"biopsy all" and "biopsy none" references, at threshold probabilities above approximately 10%; at lower thresholds, it did not exceed "biopsy all" (Figure 5). The fixed-rule strategies, by contrast, outperformed both reference strategies only over bounded ranges: the Additive model between pt≈30% and 50%, the Reclassification model between pt≈38% and 49%, and TSH ≥1.545 mIU/L between pt≈36% and 59%, while the ACR TI-RADS FNA criteria alone offered no appreciable advantage over "biopsy all" at any threshold.

Figure 5. Decision curve analysis for FNA indication in thyroid nodules



Note: Net benefit of five FNA-indication strategies across threshold probabilities from 1% to 90%, referenced against "biopsy all" and "biopsy none". The multivariable model curve (Table 4) is corrected for optimism by bootstrap (model refitted in each resample); the other strategies are fixed rules, shown with apparent values and using the TSH cutoff derived in this sample. The corrected multivariable model exceeded all other strategies at threshold probabilities above approximately 10%, but not "biopsy all" below it. TSH ≥1.545 mIU/L and the Additive and Reclassification models were advantageous only within limited ranges of threshold probability, and the ACR TI-RADS FNA criteria at none. FNA: fine-needle aspiration.
Source: Authors (2026).

4. DISCUSSION

The main finding of this study is that, in decision curve analysis, a multivariable model with continuous serum TSH showed higher estimated net benefit, a balance of cancers detected

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

against additional benign biopsies, than the ACR TI-RADS FNA criteria alone or either binary TSH–ACR TI-RADS strategy across almost the entire range of threshold probabilities evaluated, even after correction for optimism. Adding TSH to ACR TI-RADS by either binary strategy also increased sensitivity, rescuing malignancies missed by the ACR TI-RADS FNA criteria (15 and 10 of 23, compared with 5 of 8 in Trimboli et al. [16]), at a cost of about one additional benign FNA per 2 cancers rescued.

Because serum TSH is inexpensive and already recommended as part of the initial evaluation of thyroid nodules, incorporating it into risk stratification would not require an additional test in most patients. Rather than functioning as an isolated adjunct test, TSH behaved as a continuous covariate whose contribution was measurable and clinically interpretable within a combined risk-stratification framework.

Three converging lines of evidence support this incremental value. First, TSH levels were consistently higher in malignant than in benign nodules. Second, ROC-based discrimination, although modest when TSH was considered in isolation (AUC 0.652), was statistically significant (95% CI 0.571–0.734). Third, the direction of the association agreed with that reported by previous cohorts [10,28]. The optimal cutoff (1.545 mIU/L), however, was imprecisely estimated (bootstrap 95% percentile interval 1.35–2.01 mIU/L) and should not be taken as a validated threshold.

Incorporating serum TSH values with ACR TI-RADS data, whether through an Additive or a Reclassification model, increased sensitivity to a clinically relevant degree relative to ACR TI-RADS alone, rescuing a substantial proportion of malignancies that would otherwise go unnoticed, in the same range as the 5/8 rescue rate reported by Trimboli et al. [16]. However, these sensitivity gains came at a quantifiable cost, with approximately one additional benign FNA per 2.1 (Additive model; 7 additional benign FNAs for 15 rescued carcinomas) or 2 (Reclassification model; 5 for 10) rescued carcinomas. From a resource-management perspective, these increases of 7.7 and 5.5 percentage points in the proportion of benign nodules with FNA indication, respectively, could be interpreted as an economic loss. From the standpoint of patient safety and medical ethics, however, we consider this trade-off potentially justifiable given the diagnostic and prognostic cost of an undetected carcinoma, although a formal cost-effectiveness evaluation was not performed.

Although the proposed binary strategies are simple and readily applicable in routine clinical practice, dichotomizing a continuous biomarker inevitably results in the loss of information [17]. This is consistent with our decision curve analysis, in which the binary strategies were advantageous only within narrow windows of threshold probability, whereas the continuous formulation had the highest optimism-corrected net benefit at thresholds above approximately

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

10%. We therefore additionally developed a multivariable prediction model using TSH as a continuous variable, reasoning that this additional statistical complexity is only justified if it can be shown to translate into a tangible gain in clinical utility, rather than remaining a theoretical refinement.

In this approach, TSH remained an independent predictor of malignancy, adding incremental diagnostic value even after adjustment for the morphological features already captured by the ACR TI-RADS system. By combining variables such as age, sex, ultrasound nodule dimensions, TSH levels, and the TI-RADS category, the model showed good discriminative ability, with an AUC of 0.853, remaining robust even after optimism correction (AUC 0.829). Calibration, however, was less favorable: the corrected slope of 0.83 (against an ideal value of 1.00) indicated a tendency to overestimate risk in higher-risk patients and underestimate it in lower-risk patients when the model is applied without adjustment to new data, an expected consequence of a relatively modest number of events per predictor (81 malignant outcomes for 7 estimated parameters).

Decision curve analysis reinforces this interpretation. While the binary strategies showed an advantage only within limited threshold ranges (the Additive model was clinically advantageous between threshold probabilities of 30% and 50%, and the Reclassification model between 38% and 49%; Figure 5), outside these ranges, none of the binary strategies consistently outperformed the "biopsy all" or "biopsy none" reference strategies, despite the substantial number of carcinomas rescued. The continuous multivariable model, in turn, had the highest optimism-corrected net benefit at thresholds above approximately 10%, suggesting that the practical value of TSH lies less in a single binary cutoff and more in its integration into a continuous risk score, alongside conventional clinical and ultrasonographic variables.

Figure 4 further illustrates that the contribution of serum TSH to malignancy prediction is gradual rather than threshold-dependent. Instead of supporting a single diagnostic cutoff, the model demonstrates how increasing TSH progressively shifts the predicted probability of malignancy within each ACR TI-RADS category, reinforcing the rationale for modeling TSH as a continuous predictor in multivariable risk estimation. This is where the incremental value of TSH becomes most tangible. Because most patients with a thyroid nodule already have a TSH result on record, the test may widen the range of clinical scenarios in which risk-based stratification outperforms a one-size-fits-all biopsy policy, without requiring molecular testing or additional imaging, although this remains to be shown outside surgical cohorts. Taken together, these results indicate that the multivariable model maintains satisfactory, though imperfect,

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

performance after optimism correction, and that external, prospective validation remains a necessary step before clinical adoption.

Beyond the present evaluation of the proposed cutoff and integration strategies, two extensions of this work merit further investigation. First, the continuous multivariable model could be formalized into a structured scoring system, analogous to how ACR TI-RADS itself assigns points to individual ultrasonographic features, in which serum TSH contributes a weighted, continuous adjustment to the composite risk score rather than functioning as a separate binary rule. Such a tool would preserve the interpretability clinicians expect from ACR TI-RADS while retaining the statistical advantages of a continuous covariate. Second, serum TSH could be evaluated in combination with molecular markers of thyroid carcinogenesis (e.g., BRAF V600E, RAS mutations, or RET/PTC rearrangements) in indeterminate cytology categories, where diagnostic uncertainty is greatest and additional low-cost, pre-surgical information would be most clinically valuable [29]. Prospective studies designed to test either extension would help clarify whether the incremental value of TSH observed here translates into a clinically actionable tool.

Strengths and limitations

Strengths of this study include the use of histopathology as an objective reference standard and the evaluation of a TSH-based reclassification strategy that has, to our knowledge, not been previously described. The study also incorporated internal validation, calibration analysis, and decision curve analysis, providing a more comprehensive assessment of model performance and potential clinical utility. Modeling TSH as a continuous covariate, rather than as a fixed cutoff, may also make this model easier to adapt to other populations in future work: unlike a dichotomized rule, which typically requires re-deriving an entirely new cutoff whenever disease prevalence or the biomarker's distribution differs, a continuous model can often be updated through simple recalibration of its intercept and slope [17], a possibility that still requires empirical confirmation through external validation.

Several limitations should also be acknowledged. First, this was a retrospective study conducted at two affiliated hospitals of the same healthcare group, with a single radiology group and a single centralized reader, and restricted to surgically treated patients. Such cohorts are inherently enriched for malignancy because surgical referral is guided by preoperative clinical, ultrasonographic, and cytological findings. As a result, estimates of diagnostic performance and clinical utility may differ from those observed in an unselected population of patients undergoing initial thyroid nodule evaluation, owing to spectrum and verification bias [30]. This design likely contributed to the high prevalence of malignancy (47.1%) in our cohort and may limit the transportability of the proposed strategies and their predictive values to routine clinical practice.

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

Because decision curve analysis is sensitive to prevalence, the net benefit observed here should not be interpreted as evidence that the multivariable model improves FNA decisions in an unselected population. Additionally, because our cohort was drawn exclusively from patients who underwent thyroidectomy, it likely overrepresents large, benign nodules resected for compressive symptoms rather than for cancer suspicion. Such nodules disproportionately triggered the size-based FNA-indication threshold of ACR TI-RADS despite low-risk sonographic features, contributing to the low specificity of ACR TI-RADS observed in this sample. The inverse association between larger ultrasound nodule size and malignancy risk estimated by the multivariable model may partly reflect this same referral pattern rather than a generalizable biological relationship, reinforcing the need for external validation in a less selected population before clinical adoption. Moreover, ultrasound measurements systematically exceeded specimen measurements (higher size category in 45.0% of nodules, same category in 54.4%; weighted kappa 0.53), so the two are not interchangeable; only ultrasound size, available before biopsy, entered the model, whereas specimen size, obtained after surgery, was used only for description. The retrospective ACR TI-RADS re-review by a single experienced radiologist, blinded to the histopathological result, reduces intra-reader variability but precludes the assessment of inter-observer variability. In routine practice, ACR TI-RADS is applied by different professionals with varying experience, so the discrimination of the system and the incremental value of TSH observed here may differ in other settings. Further limitations include the complete-case selection of 172 of 1,006 thyroidectomies owing to incomplete records, which may introduce additional selection bias. Reasons for exclusion were not recorded individually for the 834 patients excluded at screening, and the proportions reported are approximate; no comparison between included and excluded patients was therefore possible, and selection bias cannot be excluded. Serum TSH was obtained within 6 months before surgery and its timing relative to FNA was not recorded for each patient, although it is routinely measured before FNA at the participating hospitals. The pre-biopsy framing therefore refers to the intended use of the model, which relies on variables that do not depend on the cytological result. The sample was also modest for the development of a multivariable model (81 events for 7 estimated parameters; approximately 11.6 events per parameter). Applying the criteria of Riley et al. [31] to our sample, 172 patients were borderline-adequate to limit overfitting (minimum of 134–170 patients for an expected shrinkage of 0.90) but insufficient to estimate the overall outcome risk precisely (minimum of 383 patients for a margin of ±0.05). The bootstrap-corrected calibration slope of 0.83 is consistent with some degree of overfitting, only partly mitigated by uniform shrinkage; the model's coefficients and absolute predicted risks must therefore be interpreted with caution and require external validation. Finally,

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

internal validation by bootstrap corrects for optimism but not for transportability: the model has not been validated in an independent cohort, and its coefficients, particularly those for ultrasound nodule size, which were imprecise, should not be applied to other settings.

5. FINAL CONSIDERATIONS

Among patients who underwent thyroidectomy, serum TSH provided measurable incremental value beyond ACR TI-RADS-based ultrasonographic risk stratification, improving pre-biopsy selection for fine-needle aspiration by increasing carcinoma detection, albeit at the cost of additional biopsies in benign nodules. Beyond the simplified binary integration strategies, the continuous multivariable model captured this incremental value most fully, showing the highest estimated net benefit in this cohort at threshold probabilities above approximately 10%, with calibration not yet established beyond this population. Given its low cost and wide availability, serum TSH may represent a practical source of incremental diagnostic value within ultrasound-based risk stratification, rather than a standalone diagnostic marker. These findings are restricted to this surgical cohort from two affiliated hospitals of a single healthcare group and require external validation in unselected, prospectively evaluated patients before any clinical application.

STATEMENTS AND DECLARATIONS

Competing Interests: The authors declare no financial or non-financial interests directly or indirectly related to the work submitted for publication.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics approval: This study was approved by the Research Ethics Committee of the Federal University of Piauí (UFPI) (CAAE 91941625.9.0000.5214; opinion 7.965.122; approved on 10 November 2025), in accordance with the 1964 Declaration of Helsinki and its later amendments.

Consent to participate: A waiver of informed consent was granted by the Research Ethics Committee for the retrospective analysis of de-identified medical records, in accordance with Resolution No. 466/2012 of the Brazilian National Health Council.

Consent to publish: Not applicable. This manuscript does not contain any individual person's identifiable data (images, videos, or other personal or clinical details) in any form.

Data availability: The de-identified dataset analyzed in the current study is not publicly available due to institutional and ethical restrictions on patient-level clinical data, but is available from the

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

corresponding author upon reasonable request and subject to approval by the UFPI Research Ethics Committee.

Use of artificial intelligence: The authors used an AI assistant (Claude, Anthropic) as a support tool for generating figures and language editing. All data, analyses, interpretation, and conclusions were reviewed by the authors, who take full responsibility for the content.

REFERENCES

1. Guth S, Theune U, Aberle J, Galach A, Bamberger CM. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. Eur J Clin Invest. 2009;39(8):699-706. doi:10.1111/j.1365-2362.2009.02162.x

2. Grani G, Sponziello M, Pecce V, Ramundo V, Durante C. Contemporary thyroid nodule evaluation and management. J Clin Endocrinol Metab. 2020;105(9):2869-2883. doi:10.1210/clinem/dgaa322

3. Siegel RL, Kratzer TB, Giaquinto AN, Sung H, Jemal A. Cancer statistics, 2025. CA Cancer J Clin. 2025;75(1):10-45. doi:10.3322/caac.21871

4. Lyu Z, Zhang Y, Sheng C, Huang Y, Zhang Q, Chen K. Global burden of thyroid cancer in 2022: incidence and mortality estimates from GLOBOCAN. Chin Med J (Engl). 2024;137(21):2567-2576. doi:10.1097/CM9.0000000000003284

5. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): white paper of the ACR TI-RADS committee. J Am Coll Radiol. 2017;14(5):587-595. doi:10.1016/j.jacr.2017.01.046

6. Durante C, Hegedüs L, Czarniecka A, Paschke R, Russ G, Schmitt F, et al. 2023 European Thyroid Association clinical practice guidelines for thyroid nodule management. Eur Thyroid J. 2023;12(5):e230067. doi:10.1530/ETJ-23-0067

7. Castellana M, Castellana C, Treglia G, Giorgino F, Giovanella L, Russ G, et al. Performance of five ultrasound risk stratification systems in selecting thyroid nodules for FNA. J Clin Endocrinol Metab. 2020;105(5):1659-1669. doi:10.1210/clinem/dgz170

8. Zhou C, Liu W, Zheng J, Wu Q, Ai Z. Molecular mechanisms of thyroid hormone signaling in thyroid cancer: oncogenesis, progression, and therapeutic implications. Biomedicines. 2025;13(10):2552. doi:10.3390/biomedicines13102552

9. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on thyroid nodules and differentiated thyroid cancer. Thyroid. 2016;26(1):1-133. doi:10.1089/thy.2015.0020

10. Wang Z, Lin Y, Jiang Y, Fu R, Wang Y, Zhang Q. The associations between thyroid-related hormones and the risk of thyroid cancer: an overall and dose-response meta-analysis. Front Endocrinol (Lausanne). 2022;13:992566. doi:10.3389/fendo.2022.992566

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

11. Fan XY, You W, Chen Y, Nie CC, Wang XL, Lei C, et al. A meta-analysis of the value of serum TSH concentration in the diagnosis of differentiated thyroid cancer in patients with thyroid nodules. Heliyon. 2024;10(2):e24391. doi:10.1016/j.heliyon.2024.e24391

12. Boelaert K. The association between serum TSH concentration and thyroid cancer. Endocr Relat Cancer. 2009;16(4):1065-1072. doi:10.1677/ERC-09-0150

13. Golbert L, de Cristo AP, Faccin CS, Farenzena M, Folgierini H, Graudenz MS, et al. Serum TSH levels as a predictor of malignancy in thyroid nodules: a prospective study. PLoS One. 2017;12(11):e0188123. doi:10.1371/journal.pone.0188123

14. McLeod DSA, Watters KF, Carpenter AD, Ladenson PW, Cooper DS, Ding EL. Thyrotropin and thyroid cancer diagnosis: a systematic review and dose-response meta-analysis. J Clin Endocrinol Metab. 2012;97(8):2682-2692. doi:10.1210/jc.2012-1083

15. Duccini K, de Souza MD, Delfim R, Aguiar A, Teixeira P, Vaisman M. High serum thyrotropin concentrations within the reference range: a predictor of malignancy in nodular thyroid disease. Med Princ Pract. 2018;27(3):272-277. doi:10.1159/000488196

16. Trimboli P, Curti M, Colombo A, Scappaticcio L, Leoncini A. Combining TSH measurement with TIRADS assessment to further improve the detection of thyroid cancers. Minerva Endocrinol (Torino). 2024;49(2):125-131. doi:10.23736/S2724-6507.24.04207-6

17. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med. 2006;25(1):127-141. doi:10.1002/sim.2331

18. Tang ZW, Li XX, Luo J. Development and validation of the nomogram based on ultrasound, thyroid stimulating hormone, and inflammatory marker in papillary thyroid carcinoma: a case-control study. Transl Cancer Res. 2023;12(3):490-501. doi:10.21037/tcr-22-2478

19. Li H, Lou P, Huang Y, Ma J, Zhang T, Ma L. Nomogram for predicting papillary thyroid carcinoma risk combined with clinical, biochemical and ultrasonographic indicators. BMC Cancer. 2026;26(1):352. doi:10.1186/s12885-026-15686-z

20. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. J Biomed Inform. 2014;48:193-204. doi:10.1016/j.jbi.2014.02.013

21. Ali SZ, Baloch ZW, Cochand-Priollet B, Schmitt FC, Vielh P, VanderLaan PA. The 2023 Bethesda System for Reporting Thyroid Cytopathology. Thyroid. 2023;33(9):1039-1044. doi:10.1089/thy.2023.0141

22. Ringel MD, Sosa JA, Baloch Z, Bischoff L, Bloom G, Brent GA, et al. 2025 American Thyroid Association management guidelines for adult patients with differentiated thyroid cancer. Thyroid. 2025;35(8):841-985. doi:10.1177/10507256251363120

23. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). Ann Intern Med. 2015;162(1):55-63. doi:10.7326/M14-0697

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.

24. Albert A, Anderson JA. On the existence of maximum likelihood estimates in logistic regression models. Biometrika. 1984;71(1):1-10. doi:10.1093/biomet/71.1.1

25. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Stat Med. 2002;21(26):2409-2419. doi:10.1002/sim.1047

26. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565-574. doi:10.1177/0272989X06295361

27. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361-387. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4

28. Su A, Zhao W, Wu W, Wei T, Ruan M, Li Z, et al. The association of preoperative thyroid-stimulating hormone level and the risk of differentiated thyroid cancer in patients with thyroid nodules: a systematic review and meta-analysis. Am J Surg. 2020;220(3):634-641. doi:10.1016/j.amjsurg.2020.01.009

29. Muzza M, Colombo C, Pogliaghi G, Karapanou O, Fugazzola L. Molecular markers for the classification of cytologically indeterminate thyroid nodules. J Endocrinol Invest. 2020;43(6):703-716. doi:10.1007/s40618-019-01164-w

30. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics. 1983;39(1):207-215. doi:10.2307/2530820

31. Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ. 2020;368:m441. doi:10.1136/bmj.m441

This article is published in open access under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.