Journal of Carcinogenesis
Covering Genesis to Therapeutics
Journal of Carcinogenesis, ISSN: 1477-3163
2025; Vol 24: Issue 10s, https://carcinogenesis.com/

Diagnostic Accuracy of Ultrasound-Guided Fnac for Thyroid Malignancy in ACR Ti-Rads TR4 Nodules: A Prospective Study with Histopathological Correlation

Ahmed Lamey[1]*, Ahmed Mohamed Khalil[2], Hassan E. Younes[3], Ahmed M. Aydarous[4], Ahmed Mohamed Shebl[5], Ahmed M. Kandel[6], Elsayed Fottoh Elgarhey[7], Ragab Elsayed Salah[8], Salah Saber Mabrouk[9], Ahmed Abdou Shaalan[10], Ahmed Abdelwahed Mahmoud[11], Mohamed Hussein Basiony[12], Adel Mohamed Ismail[13], Tamer A. Abouelgreed[14], Mohamed Abdel Aziz Ali Khalil[15]

[1] Department of General Surgery, Faculty of Medicine, Kafr Elsheikh University, Egypt; Burjeel Royal Hospital, Abu Dhabi, United Arab Emirates; Email: dr.ahmedlamey@gmail.com; ORCID: https://orcid.org/0000-0002-4284-5330

[2] Department of General Surgery, Faculty of Medicine, Misr University for Science and Technology, Egypt; Email: drahmedkhalil_80@yahoo.com

[3] Department of General Surgery, Faculty of Medicine, Al-Azhar University, Cairo, Egypt; Email: hassansurg77@yahoo.com

[4] Department of Surgical Oncology, Faculty of Medicine, Al-Azhar University, Cairo, Egypt; Email: ahmedaydarous343@gmail.com; ORCID: https://orcid.org/0009-0008-2931-7227

[5] Department of General Surgery, Faculty of Medicine, Al-Azhar University, Cairo, Egypt; Email: ahmedshibl882@gmail.com

[6] Department of General Surgery, Faculty of Medicine, Al-Azhar University, Damietta, Egypt; Email: kandel2025@gmail.com; ORCID: https://orcid.org/0009-0001-2462-3742

[7] Department of General Surgery, Faculty of Medicine, Al-Azhar University, Cairo, Egypt; Email: drelsayedelgarhey1984@gmail.com

[8] Department of General Surgery, Faculty of Medicine, Al-Azhar University, Cairo, Egypt; Email: ragabsalah078@gmail.com

[9] Department of General Surgery, Faculty of Medicine, Al-Azhar University, Cairo, Egypt.

Email: 01000231689sasa@gmail.com

[10] Department of Radiology, Faculty of Medicine, Al-Azhar University, Cairo, Egypt; Email: dr_ahmedshallan@yahoo.com

[11] Department of Radiology, Faculty of Medicine, Ain Shams University, Cairo, Egypt; Email: ahmed_abdelwahed@yahoo.com; ORCID: https://orcid.org/0000-0003-4224-1622

[12] Department of Surgical Oncology, Ismailia Teaching Oncology Hospital, Ismailia, Egypt; Email: hamada_hassen66@yahoo.com

[13] Department of Surgical Oncology, Ismailia Teaching Oncology Hospital, Ismailia, Egypt; Email: adel.ismail.ioth@gmail.com; ORCID: https://orcid.org/0000-0002-8491-2781

[14] Department of Urology, Faculty of Medicine, Al-Azhar University, Cairo, Egypt; Email: dr_tamer_ali@yahoo.com; ORCID: https://orcid.org/0000-0003-2640-3425

Telephone: +20 100 657 1508

[15] Department of General Surgery, Faculty of Medicine, Arish University, Egypt; Email: mohamed76aziz@gmail.com; ORCID: https://orcid.org/0009-0005-0641-8525

***Corresponding Author: Ahmed Lamey**

Email: dr.ahmedlamey@gmail.com; ORCID: https://orcid.org/0000-0002-4284-5330

© 2025 Journal of Carcinogenesis | Published for Carcinogenesis Press by Wolters Kluwer-Medknow

ABSTRACT

Background: Ultrasound (US)-guided fine-needle aspiration cytology (FNAC) is pivotal in evaluating thyroid nodules, but its performance varies with the Bethesda threshold, sample adequacy, and referral population. We assessed fine-needle aspiration cytology against surgical histopathology in American College of Radiology (ACR) Thyroid Imaging Reporting and Data System (TI-RADS) category 4 (TR4) nodules. Patients and Methods: In this prospective Standards for Reporting of Diagnostic Accuracy (STARD)-compliant study, 120 consecutive adults with a category 4 nodule ≥1.5 cm undergoing thyroidectomy were analysed. Fine-needle aspiration cytology was reported using the 2023 Bethesda System for Reporting Thyroid Cytopathology (TBSRTC); histopathology was the reference standard. The primary analysis (Bethesda V-VI) included diagnostic aspirates only (n=108). Category I (n=12) was reported separately. Alternative thresholds, nondiagnostic handling, noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP)/follicular tumour of uncertain malignant potential (FT-UMP) reclassification, and patient-level analyses were exploratory. Results: Index-nodule malignancy was confirmed in 32/120 patients (26.7%; 27 papillary thyroid carcinomas (PTC), 5 follicular thyroid carcinomas (FTC)); seven incidental papillary thyroid microcarcinomas (PTMCs) lay outside the index nodules. Nondiagnostic aspirates (10.0%) were associated with cystic composition (6/11 [54.5%] versus (vs) 6/109 [5.5%]; p<0.001). Among 108 diagnostic aspirates, the observed risk of malignancy was 3.3% (Bethesda II), 14.3% (III), 18.2% (IV), 66.7% (V), and 100% (VI). At Bethesda V-VI, sensitivity was 84.4% (95% confidence interval (CI) 68.2-93.1%), specificity 96.1% (89.0-98.6%), positive predictive value (PPV) 90.0% (74.4-96.5%), negative predictive value (NPV) 93.6% (85.9-97.2%), accuracy 92.6% (86.1-96.2%), positive likelihood ratio (LR+) 21.37, negative likelihood ratio (LR-) 0.16, and Cohen's kappa (κ) 0.82. Sensitivity rose to 90.6% at the IV-VI threshold (specificity 84.2%). Four of five false negatives were follicular carcinomas. Inter-observer cytological agreement was substantial (κ=0.78). Conclusions: At Bethesda V-VI, ultrasound-guided fine-needle aspiration cytology in category 4 nodules showed high specificity and negative predictive value, with moderate, threshold-dependent sensitivity (84.4%) limited mainly by follicular carcinoma. Observed malignancy risks were broadly consistent with 2023 Bethesda System for Reporting Thyroid Cytopathology benchmarks, although Categories IV and V fell at or below the lower limits, with wide confidence intervals. Nondiagnostic aspirates were driven by fluid-rich composition. These estimates derive from a surgically verified, high-prevalence cohort subject to verification bias and require external validation.

Keywords: *Thyroid Nodule, Fine-Needle Aspiration Cytology, Bethesda System 2023, Papillary Thyroid Carcinoma, Diagnostic Accuracy.*

How to Cite: Ahmed Lamey, Ahmed Mohamed Khalil, Hassan E. Younes, Ahmed M. Aydarous, Ahmed Mohamed Shebl, Ahmed M. Kandel, Elsayed Fottoh Elgarhey, Ragab Elsayed Salah, Salah Saber Mabrouk, Ahmed Abdou Shaalan, Ahmed Abdelwahed Mahmoud, Mohamed Hussein Basiony, Adel Mohamed Ismail, Tamer A. Abouelgreed, Mohamed Abdel Aziz Ali Khalil, (2025) Diagnostic Accuracy of Ultrasound-Guided Fnac for Thyroid Malignancy in ACR Ti-Rads TR4 Nodules: A Prospective Study with Histopathological Correlation, *Journal of Carcinogenesis*, *Vol.24, No.10s*, 810-818

1. INTRODUCTION

Thyroid nodules are detected in up to 50-60% of adults via high-resolution ultrasonography. Although the majority represent benign lesions, approximately 7-15% harbour malignancy. Papillary thyroid carcinoma (PTC) accounts for 80-85% of thyroid malignancies, followed by follicular thyroid carcinoma (FTC) [1,2]. The widespread use of high-frequency ultrasonography has markedly increased the detection of thyroid nodules, and very high nodule prevalence has been documented even in apparently healthy populations [3]. This has coincided with a rising incidence of small papillary carcinomas in many countries, a trend that has been attributed largely to overdiagnosis rather than to a true increase in disease occurrence [4]. The central clinical task is therefore to identify the minority of nodules that require surgery while sparing the large majority of patient's unnecessary biopsy, repeat sampling and operation. Ultrasound risk-stratification systems, notably ACR TI-RADS, standardize sonographic feature assessment to estimate malignancy risk and guide FNAC decisions [5]. TR4 nodules carry an estimated malignancy risk of 5-20% in screening populations, with FNAC recommended for lesions ≥1.5 cm [5,6]. ACR TI-RADS assigns points for five sonographic feature groups (composition, echogenicity, shape, margin and echogenic foci) and maps the total to categories TR1-TR5. TR4 (moderately suspicious) occupies the intermediate zone of this scale. The pooled sensitivity and specificity of ACR TI-RADS in a meta-analysis were 0.89 and 0.70, respectively [7], indicating that ultrasound alone cannot settle the diagnosis in many nodules. Size-based biopsy thresholds also carry a trade-off: in one cohort, 5.7% of TR4 cancers measured less than 1.5 cm [8]. Cytology therefore remains the pivotal test for TR4 nodules, and combining TI-RADS category with Bethesda category has been

© 2025 Journal of Carcinogenesis | Published for Carcinogenesis Press by Wolters Kluwer-Medknow

reported to refine malignancy risk stratification in cytologically indeterminate nodules [9]. FNAC remains the first-line diagnostic investigation [6]. The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) established six reporting categories, updated in 2023 [10]. Reported diagnostic performance metrics vary due to threshold selection, nondiagnostic sample handling, and selection bias inherent to surgical cohorts [11]. Performance within the TR4 stratum, where the prior probability of malignancy is intermediate and FNAC decisions are most consequential, is less well characterised using the 2023 terminology, particularly in surgically verified cohorts that report nondiagnostic and borderline entities transparently. The Bethesda System, first published in 2009 [12], standardized thyroid cytology reporting into diagnostic categories, each linked to an implied risk of malignancy (ROM) and a management recommendation. The 2023 edition retained the six-category structure while updating ROM ranges and management guidance [10]. Because ROM and diagnostic accuracy depend on how indeterminate and nondiagnostic categories are treated, published sensitivity and specificity values are difficult to compare across series [11,13]. Regional data add a further layer of heterogeneity: cytology-histology correlation studies from Oman and Saudi Arabia are mostly retrospective, differ in their handling of indeterminate categories, and report widely varying accuracy [14,15]. To our knowledge, prospective evidence from the Gulf region that applies the 2023 terminology specifically to TR4 nodules, with transparent reporting of nondiagnostic and borderline results, is limited. Non-diagnostic aspirates are a related and frequently under-reported source of uncertainty. A predominantly cystic composition and hypoechogenicity are recognized predictors of nondiagnostic results [16], and repeat ultrasound-guided sampling of the solid component is recommended after an initial nondiagnostic aspirate [2,17]. Clarifying how often this occurs in TR4 nodules, and what the resected histology shows, is relevant to both radiologists and cytopathologists. The primary objective of this prospective study was to evaluate the diagnostic accuracy, sensitivity, specificity, predictive values, and likelihood ratios of US-guided FNAC using 2023 TBSRTC criteria across multiple cutoff thresholds in consecutive patients with ACR TI-RADS TR4 index nodules undergoing surgical thyroidectomy.

2. PATIENTS AND METHODS

This prospective single-arm diagnostic accuracy study is reported in compliance with STARD 2015 [18]. The study was conducted at Burjeel Royal Hospitals, United Arab Emirates, between August 2023 and August 2024. Institutional ethical approval was obtained from the Medical Research Ethics Committee of Burjeel Hospitals. Written informed consent was obtained from all participating patients. Sample size was calculated using PASS software (v16.0). Assuming expected FNAC sensitivity of 85% at Bethesda V-VI, α=0.05, permissible marginal error d=0.12, and expected surgical cohort malignancy prevalence of 27%, Buderer's formula [19] yielded a required sample size of approximately 126 patients. Due to institutional volume constraints and consecutive enrollment, 120 patients were enrolled; 108 had diagnostic aspirates and constituted the primary analysis set. Post-hoc precision analysis showed a sensitivity 95% CI width of 24.9% (68.2-93.1%) and a specificity 95% CI width of 9.6% (89.0-98.6%), indicating adequate precision for specificity and overall accuracy but wide sensitivity intervals because of the limited number of malignant cases (n=32). We prospectively enrolled 120 consecutive adult patients (aged ≥18 years) with an ACR TI-RADS TR4 nodule (4-6 points) measuring ≥1.5 cm scheduled for hemithyroidectomy or total thyroidectomy. Surgical indications were recorded for each patient, and more than one indication could apply; the sum of indication categories therefore exceeded the number of patients: suspicious cytology (Bethesda IV-VI, n=41), compressive symptoms (n=48), cosmetic disfigurement (n=22), and retrosternal extension or rapid growth (n=18); nine patients had two indications each (41+48+22+18=129; 129-9=120). Cytology therefore contributed to surgical selection in at least some patients, creating differential verification; this is discussed in Sections 2.4 and 4.1. Exclusion criteria were: (a) prior thyroid surgery or neck irradiation; (b) purely cystic lesions without solid mural components (0 points); (c) concurrent secondary head and neck malignancies; and (d) uncorrected coagulopathy. In patients with multinodular goitre (MNG, n=77), precise sonographic-pathological mapping was performed. If multiple TR4 nodules were present, the index nodule was selected by highest total ACR TI-RADS score; ties were resolved by largest maximum diameter, then by presence of the most suspicious sonographic feature. The index nodule's anatomical quadrant, lobar location, and three-dimensional size were marked for gross specimen correlation. The patient was the unit of analysis, with one index nodule per patient. Seven incidental papillary thyroid microcarcinomas (PTMC ≤10 mm) outside the index TR4 nodule were catalogued separately and evaluated in patient-level sensitivity analyses. Because all enrolled patients were selected from a surgical management stream, verification bias is inherent. Of 450 patients with an ACR TI-RADS TR4 nodule evaluated during the study period, 330 were not resected (310 managed non-operatively, 20 lost to follow-up or declined surgery) and were excluded from surgical correlation. The decision to operate was made with the FNAC result available, so cytology influenced who was verified. This partial and differential verification tends to overestimate sensitivity, underestimate specificity, and raise observed prevalence (and PPV) relative to unselected TR4 populations [20]. A comparison of resected and non-resected patients was not available; this is acknowledged as a limitation. A STARD flow diagram is provided in Figure 1. High-resolution grey-scale and color Doppler neck ultrasonography was performed by two dedicated interventional radiologists using a 7-12 MHz linear-array transducer (GE Logiq E9, GE Healthcare, Milwaukee, WI, USA). ACR TI-RADS points (median = 5, range 4-6) were assigned based on composition, echogenicity, shape, margin, and echogenic foci. Nodule composition was classified as solid (<10% fluid), mixed cystic-solid (10-50% fluid), or predominantly cystic (>50% fluid); purely cystic 0-point lesions

were excluded. These categories reflect institutional ultrasound reporting definitions rather than formal ACR TI-RADS descriptive sub-criteria. US-guided FNAC was performed by two dedicated interventional radiologists independent of the surgical and pathological teams, using 23-gauge needles without rapid on-site evaluation (ROSE). Each nodule underwent an average of 3 passes (range 2-4). Smears were prepared as air-dried May-Grünwald-Giemsa and alcohol-fixed Papanicolaou stains, supplemented by liquid-based cytology when cellular yield permitted. The median interval between FNAC and surgery was 28 days (IQR 18-42). Cytological smears were independently evaluated by two cytopathologists blinded to clinical findings, ultrasound scores, and final histopathology. Discrepancies were resolved by joint review with a third senior cytopathologist. Per 2023 TBSRTC adequacy guidelines, a sample was diagnostic if it contained at least 6 groups of well-preserved follicular epithelial cells with ≥10 cells per group, except for benign colloid nodules or unequivocal atypia/malignancy. Terminology followed 2023 TBSRTC ("Atypia of Undetermined Significance" [AUS]). Inter-observer agreement before consensus was substantial (κ=0.78; 95% CI 0.66-0.90). No procedure-related adverse events were recorded. Histopathological examination was performed according to the 2022 WHO Classification of Thyroid Neoplasms by two senior surgical pathologists blinded to preoperative cytological Bethesda categories and ultrasound scores [21]. Final diagnoses were reached by consensus. NIFTP (n=3) and FT-UMP (n=2) were categorized as non-malignant in the primary analysis; sensitivity analyses reclassifying them as malignant were performed. Incidental PTMCs ≤10 mm outside the index TR4 nodule (n=7) were recorded as incidental findings and excluded from index-nodule correlation.

3. STATISTICAL ANALYSIS

Data were analyzed using SPSS Statistics version 28.0 and R version 4.2.2. Continuous variables were assessed for normality using Shapiro-Wilk testing and reported as mean ± SD or median (IQR). Continuous variables were compared between benign and malignant groups using the independent-samples t-test (Mann-Whitney U test for non-normal variables). Categorical variables were compared using Fisher's exact test, with the Freeman-Halton extension for multi-category tables. Diagnostic accuracy indices were calculated with Wilson score 95% confidence intervals. LR+ and LR- were calculated using the log-transformation method; when false positives were zero, LR+ was designated "not estimable." Inter-observer agreement was assessed using Cohen's κ with 10,000 bootstrap iterations. Agreement between dichotomized FNAC (V-VI vs II-IV) and dichotomized histology (malignant vs non-malignant) was also summarized using Cohen's κ. The primary pre-specified analysis was diagnostic performance at Bethesda V-VI among the 108 diagnostic aspirates, with Category I reported separately. The V-VI threshold and Category I handling were defined a priori and documented in the PACTR protocol before data analysis. All other analyses—alternative thresholds, alternative nondiagnostic handling, NIFTP/FT-UMP reclassification, patient-level PTMC analysis, and baseline comparisons—were exploratory. All p-values are exploratory; no adjustment for multiple comparisons was applied.

4. RESULTS

A total of 120 consecutive patients met inclusion criteria and completed preoperative FNAC and surgical resection. Mean age was 50.8±11.4 years; 83.3% (n=100) were female. Multinodular goitre was present in 77 patients (64.2%), and 43 (35.8%) had solitary TR4 nodules. Final histopathology confirmed index-nodule malignancy in 32/120 (26.7%; 27 PTC, 5 FTC). Benign lesions comprised 88 cases (73 nodular goitre, 10 follicular adenoma, 3 NIFTP, 2 FT-UMP). Inter-observer cytological agreement before consensus was substantial (κ=0.78, 95% CI 0.66-0.90; 8 discrepancies resolved by consensus; 93.3% exact agreement). Agreement on specimen adequacy was almost perfect (κ=0.93; 95% CI 0.85-0.99).

Table 1. Baseline clinical and ultrasonographic characteristics (n=120)

Characteristic	Total (n=120)	Benign (n=88)	Malignant (n=32)	p-value
Age, years, mean ± SD	50.8 ± 11.4	50.4 ± 11.6	51.9 ± 10.9	0.528
Sex, female/male	100 (83.3%)/20 (16.7%)	74 (84.1%)/14 (15.9%)	26 (81.3%)/6 (18.8%)	0.783
BMI, kg/m², mean ± SD	28.8 ± 4.5	28.6 ± 4.4	29.3 ± 4.8	0.457
Presentation, solitary/MNG	43 (35.8%)/77 (64.2%)	29 (33.0%)/59 (67.0%)	14 (43.8%)/18 (56.3%)	0.290
Index nodule size, cm, mean ± SD	2.71 ± 1.48	2.76 ± 1.52	2.58 ± 1.37	0.551
Composition, solid/mixed/cystic*	81 (67.5%)/28 (23.3%)/11 (9.2%)	69 (78.4%)/11 (12.5%)/8 (9.1%)	12 (37.5%)/17 (53.1%)/3 (9.4%)	<0.001
Microcalcifications present	45 (37.5%)	28 (31.8%)	17 (53.1%)	0.054

*Composition categories reflect institutional definitions (<10% fluid solid; 10-50% mixed; >50% cystic), not formal ACR TI-RADS sub-criteria.

Cytology-histopathology mapping and determinants of adequacy

Nondiagnostic aspirates (Category I) occurred in 12 patients (10.0%). Nondiagnostic rates differed significantly by composition: 6/11 predominantly cystic nodules (54.5%) vs 6/109 non-cystic nodules (5.5%; 0/28 mixed and 6/81 solid) (Fisher's exact $p<0.001$; Freeman-Halton $p\leq0.001$). All 12 nondiagnostic nodules were benign on final histopathology (10 nodular goitre, 2 follicular adenoma). No repeat FNAC was performed before surgery; all 12 patients proceeded directly to resection because of concurrent surgical indications.

Table 2. Cross-tabulation of 2023 Bethesda categories vs final histopathology (n=120)

Bethesda category	Benign (n=88)	Malignant (n=32)	Total (n=120)	Observed ROM (95% CI)
I. Nondiagnostic	12	0	12	0.0% (0.0-24.2%)
II. Benign	58	2 (FTC)	60	3.3% (0.9-11.4%)
III. AUS	6	1 (PTC)	7	14.3% (2.6-51.3%)
IV. Follicular neoplasm	9	2 (FTC)	11	18.2% (5.1-47.7%)
V. Suspicious for malignancy	3	6 (5 PTC, 1 FTC)	9	66.7% (35.4-87.9%)
VI. Malignant	0	21 (PTC)	21	100% (84.5-100%)
Total	88	32 (27 PTC, 5 FTC)	120	26.7% (19.6-35.2%)

Diagnostic performance across thresholds

At the primary threshold (Bethesda V-VI), sensitivity was 84.4% (68.2-93.1%), specificity 96.1% (89.0-98.6%), PPV 90.0% (74.4-96.5%), NPV 93.6% (85.9-97.2%), accuracy 92.6% (86.1-96.2%), LR+ 21.37 (6.98-65.45), LR- 0.16 (0.07-0.36), and Cohen's $\kappa=0.82$ (0.70-0.94; dichotomized V-VI vs II-IV against malignant vs non-malignant histology).

Table 3. Diagnostic performance of FNAC across Bethesda thresholds (n=108 diagnostic aspirates)

Threshold	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Accuracy, % (95% CI)	LR+	LR-
V-VI (primary)	84.4 (68.2-93.1)	96.1 (89.0-98.6)	90.0 (74.4-96.5)	93.6 (85.9-97.2)	92.6 (86.1-96.2)	21.37 (6.98-65.45)	0.16 (0.07-0.36)
IV-VI (broader)	90.6 (75.8-96.8)	84.2 (74.4-90.7)	70.7 (55.5-82.4)	95.5 (87.6-98.5)	86.1 (78.3-91.4)	5.74 (3.37-9.76)	0.11 (0.04-0.33)
VI only (strict)	65.6 (48.3-79.6)	100 (95.2-100)	100 (84.5-100)	87.4 (78.8-92.8)	89.8 (82.7-94.2)	Not estimable	0.34 (0.21-0.55)

Sensitivity analyses (all exploratory)

All analyses in this section are exploratory and should not be used for confirmatory inference.

Category I handling: Counting all 12 nondiagnostic aspirates as test-positive yielded sensitivity 84.4% (27/32; 95% CI 68.2-93.1%), specificity 83.0% (73/88; 73.8-89.4%), PPV 64.3% (27/42; 49.2-77.0%), and NPV 93.6% (73/78; 85.9-97.2%). Counting all 12 as test-negative yielded sensitivity 84.4% (27/32), specificity 96.6% (85/88; 90.5-98.8%), PPV 90.0% (27/30; 74.4-96.5%), and NPV 94.4% (85/90; 87.6-97.6%). Borderline entities: Reclassifying 3 NIFTP and 2 FT-UMP as malignant (total malignant = 37) reduced sensitivity to 75.7% (28/37; 95% CI 59.9-86.6%) with specificity 97.2% (69/71; 95% CI 90.3-99.2%) at V-VI. False negatives and false positives: The 5 false-negative index nodules comprised 4 FTC (2 in Bethesda II, 2 in Bethesda IV) and 1 PTC (Bethesda III). The 3 false-positive Category V lesions comprised 1 borderline tumour (NIFTP/FT-UMP), 1 follicular adenoma with oxyphilic metaplasia, and 1 nodular goitre with hyperplastic adenomatoid features. The 5 borderline tumours were distributed as 2 in Bethesda II, 1 in III, 1 in IV, and 1 in V. Patient-level PTMC analysis: Seven incidental PTMCs ≤10 mm were identified outside index nodules; all occurred in patients whose index TR4 nodule was histologically benign and whose index FNAC was negative at V-VI. When patient-level malignancy included any malignant focus, total malignant patients were 39/120. Treating Category I aspirates as

negative, the patient-level 2×2 table was TP=27, FP=3, FN=12, TN=78. Patient-level sensitivity was 69.2% (27/39; 95% CI 53.6-81.4%), specificity 96.3% (78/81; 95% CI 89.7-98.7%), PPV 90.0% (27/30; 95% CI 74.4-96.5%), NPV 86.7% (78/90; 95% CI 78.1-92.2%), and accuracy 87.5% (105/120).

Table 4. Patient-level 2×2 analysis including incidental PTMCs (n=120)

Patient-level FNAC at V-VI	Patient-level malignant	Patient-level non-malignant	Total
Positive (V-VI)	TP = 27	FP = 3	30
Negative (II-IV or I)	FN = 12	TN = 78	90
Total	39	81	120

STARD flow diagram

Figure 1 illustrates the participant flow through the study in compliance with STARD 2015 guidelines:

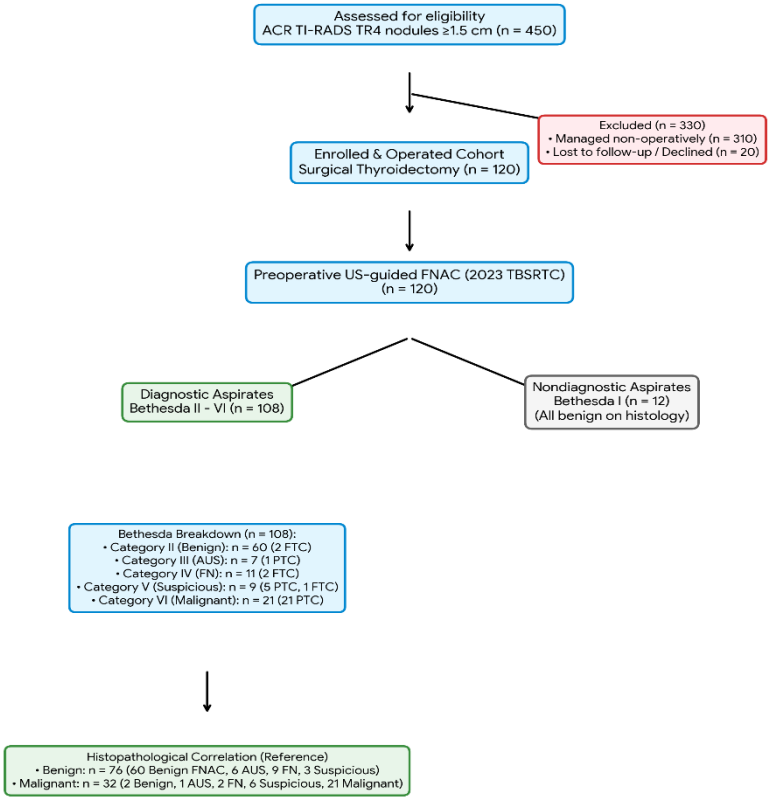

Figure 1: STARD Flow Diagram - Participant selection, FNAC categorization, and histopathological correlation

5. DISCUSSION

In this prospective study of 120 ACR TI-RADS TR4 nodules, US-guided FNAC under 2023 TBSRTC criteria had accuracy 92.6%, sensitivity 84.4%, specificity 96.1%, NPV 93.6%, LR+ 21.37, and LR- 0.16 at the primary Bethesda V-VI threshold. Observed malignancy risks were broadly consistent with 2023 TBSRTC ranges, with Category IV below and Category V marginally below the lower benchmark limits: II 3.3% (expected 2-7%), III 14.3% (13-30%), IV 18.2% (23-34%), V 66.7% (67-83%), and VI 100% (97-100%) [10]. The wide confidence intervals for Categories IV and V include the expected values. Category I ROM was 0% (0/12; 95% CI 0.0-24.2%), below the 5-20% benchmark; however, all 12 nondiagnostic nodules were resected and benign, so this estimate cannot be generalized to nondiagnostic aspirates managed non-operatively. Four of the five false negatives at V-VI were follicular carcinomas. Follicular carcinoma is diagnosed on capsular or vascular invasion in resected tissue, so cytology can at best place such lesions in Category IV [10,22]; a V-VI cut-off therefore limits sensitivity for FTC by design. Accordingly, sensitivity rose to 90.6% at the IV-VI threshold, at the cost of specificity (84.2%) and PPV (70.7%). Two follicular carcinomas were reported as benign (Category II), most plausibly through sampling error; a benign result in a TR4 nodule should therefore be interpreted together with the sonographic features and with continued surveillance rather than as definitive reassurance. Pooled Bethesda estimates depend strongly on threshold and handling of indeterminate categories [11], consistent with the threshold-dependent

sensitivity and specificity observed here. The proportion of Category III (AUS) aspirates was 5.8% (7/120; 6.5% among diagnostic aspirates), below the 10% benchmark, although this is a surgically selected cohort. Fluid-rich composition was strongly associated with nondiagnostic aspirates (54.5% cystic vs 5.5% non-cystic, $p<0.001$), supporting sampling of peripheral solid mural components under real-time guidance. The 10.0% nondiagnostic rate was achieved without ROSE.The performance observed here can be placed alongside regional surgical and cytology-histology series. In a retrospective Omani study of 1359 FNAC procedures (137 patients operated), FNAC sensitivity was 80% and specificity 98% after excluding nondiagnostic and indeterminate categories; the 15.7% malignancy rate among benign (Thy2) nodules reflected selection of large (>4 cm) nodules for surgery, a form of verification bias comparable to that affecting our cohort [14]. In a Saudi surgical cohort of 408 aspirates, Bethesda cytology achieved sensitivity 88.9%, specificity 75.6%, PPV 79.7%, NPV 84.4% and accuracy 81.5% [15]; our specificity at the V-VI threshold (96.1%) was higher, which probably reflects differences in cohort composition and in the way indeterminate categories were assigned to the test-positive group [11]. Sharma reported sensitivity 89.5%, specificity 98% and accuracy 97% in 724 cases, and attributed false-negative results mainly to follicular-pattern lesions, cystic papillary carcinoma and papillary microcarcinoma [23]; this pattern parallels the follicular carcinoma false negatives in our series. Large single-institution series with histologic correlation, such as that of Yang et al. (4703 patients), provide a further benchmark for cytology-histology concordance [13]. Our findings are also consistent with studies that combine sonographic risk category with cytology. In a retrospective series of 140 cytologically indeterminate nodules (Bethesda III-V) with surgical histology, the observed malignancy rate was 83.5% when a high-suspicion ultrasound category (TI-RADS 4b-5) coincided with Bethesda IV-V, compared with 8.1% when a low-suspicion category (TI-RADS 2-4a) coincided with Bethesda III [9]. The very high specificity that we observed at the Bethesda V-VI threshold in a TR4-restricted cohort is in keeping with this pattern, in which concordant high-risk cytology carries strong rule-in value, whereas the lower sensitivity mirrors the trade-off of a stricter cytologic threshold. Given that ACR TI-RADS alone has a pooled specificity of about 0.70 [7], and that a small but real fraction of TR4 cancers fall below the size threshold for biopsy [8], these data support using cytology as the decisive step within the TR4 stratum rather than relying on sonographic category or size alone. The nondiagnostic findings are in line with previous reports. In a series of 1493 US-guided aspirations, a cystic portion exceeding 50% was an independent predictor of nondiagnostic cytology (odds ratio 2.82) [16]. In a surgical series of cystic or predominantly cystic nodules, US-guided FNAB gave a nondiagnostic smear rate of 9.2%, and carcinoma was found in 17.6% of the 119 resected nodules, so that a cystic appearance did not exclude malignancy [24]. Our nondiagnostic rate of 10.0% and the strong association with cystic composition (54.5% vs 5.5%) are therefore comparable to these reports, and our finding that all 12 nondiagnostic nodules were benign should not be extrapolated beyond this surgically resected subset. Guidance on repeat ultrasound-guided sampling of initially nondiagnostic nodules [2,17] remains appropriate.

The observed association between mixed cystic-solid composition and malignancy (53.1% in malignant vs 12.5% in benign) reflects referral patterns in this surgical cohort. Because non-standard institutional composition definitions were used, this finding is exploratory and should not be interpreted causally.

Clinical implications and limitations

High specificity (96.1%) and LR+ (21.37) at V-VI indicate that positive FNAC strongly supports definitive surgery. A negative result is less reassuring: 5/32 malignancies were missed, predominantly FTC. PPV (90.0%) and NPV (93.6%) are specific to this surgical cohort with 26.7% prevalence; in lower-prevalence populations, PPV would be lower and NPV higher.

Limitations include: (1) verification bias from surgical selection, with 330 non-resected TR4 patients excluded and FNAC influencing verification; (2) sample size was slightly below the Buderer target, with wide sensitivity CIs and exploratory subgroup analyses; (3) no routine molecular testing.

6. CONCLUSION

US-guided FNAC using 2023 TBSRTC criteria in ACR TI-RADS TR4 nodules provides high specificity (96.1%), good NPV (93.6%), and a high positive likelihood ratio (LR+ = 21.37) at Bethesda V-VI. Sensitivity was moderate (84.4%) and threshold-dependent, with false negatives predominantly FTC. Observed malignancy risks were broadly consistent with 2023 TBSRTC benchmarks, with Category IV below and Category V marginally below the lower limits and wide confidence intervals. Nondiagnostic aspirates were strongly associated with fluid-rich composition, supporting targeted sampling of solid mural components. These estimates derive from a surgically verified, high-prevalence cohort and require external validation before application to unselected thyroid clinic populations.

DECLARATIONS

Institutional Review Board Statement:

The study was conducted in accordance with the Declaration of Helsinki and applicable institutional and regulatory requirements. The study protocol was reviewed and approved by the Institutional Review Board of Burjeel Hospitals, UAE

(approval no. BURJ/RERC/2023/131).

Acknowledgements:
The authors thank the surgical, cytopathological and radiological clinical and nursing staff at Burjeel Royal Hospitals for histopathological, Radiological, administrative and technical support.

Informed Consent Statement:
Written informed consent was obtained from all participants before enrolment and randomization, after the study objectives, procedures, potential risks, and the use of their clinical data and photographs for research and publication had been explained. The study was approved by the Institutional Review Board and was conducted in accordance with the Declaration of Helsinki.

Data Availability Statement:
The data supporting the findings of this study are available from the corresponding author upon reasonable request, subject to applicable institutional, ethical, and patient-confidentiality requirements.

Funding:
This research received no external funding.

Conflicts of Interest:
The authors declare no conflicts of interest.

Author Contributions:
AL, AM, HE, AM, AM: manuscript preparation, protocol, data collection and management, manuscript editing. AM, EF, RE, SS, AA: Data acquisition, data analysis and management, manuscript editing. AA, MH, AM, TA, MA: Manuscript editing, project development, data analysis, and project development. All the authors have read and approved the manuscript.

REFERENCES

[1] Alexander EK, Cibas ES. Diagnosis of thyroid nodules. Lancet Diabetes Endocrinol. 2022; 10(7):533-539.
[2] Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer. Thyroid. 2016; 26(1):1-133.
[3] 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.
[4] Vaccarella S, Franceschi S, Bray F, Wild CP, Plummer M, Dal Maso L. Worldwide thyroid-cancer epidemic? The increasing impact of overdiagnosis. N Engl J Med. 2016; 375(7):614-617.
[5] Tessler FN, Middleton WD, Grant EG, et al. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): white paper of the ACR TI-RADS Committee. J Am Coll Radiol. 2017; 14(5):587-595.
[6] Gharib H, Papini E, Garber JR, et al. AACE/ACE/AME medical guidelines for clinical practice for the diagnosis and management of thyroid nodules, 2016 update. Endocr Pract. 2016; 22(5):622-639.
[7] Li W, Wang Y, Wen J, Zhang L, Sun Y. Diagnostic performance of American College of Radiology TI-RADS: a systematic review and meta-analysis. AJR Am J Roentgenol. 2021; 216(1):38-47.
[8] Huang EYF, Kao NH, Lin SY, et al. Concordance of the ACR TI-RADS classification with Bethesda scoring and histopathology risk stratification of thyroid nodules. JAMA Netw Open. 2023; 6(9):e2331612.
[9] Barbosa TLM, Mesa Junior CO, Graf H, Cavalvanti T, Trippia MA, da Silveira Ugino RT, de Oliveira GL, Granella VH, de Carvalho GA. ACR TI-RADS and ATA US scores are helpful for the management of thyroid nodules with indeterminate cytology. BMC Endocr Disord. 2019; 19(1):112.
[10] 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.
[11] Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda System for Reporting Thyroid Cytopathology: a meta-analysis. Acta Cytol. 2012; 56(4):333-339.
[12] Cibas ES, Ali SZ. The Bethesda System for Reporting Thyroid Cytopathology. Thyroid. 2009; 19(11):1159-1165.
[13] Yang J, Schnadig V, Logrono R, Wasserman PG. Fine-needle aspiration of thyroid nodules: a study of 4703 patients with histologic and clinical correlations. Cancer. 2007; 111(5):306-315.
[14] Alhassan R, Al Busaidi N, Al Rawahi AH, et al. Features and diagnostic accuracy of fine needle aspiration cytology of thyroid nodules: retrospective study from Oman. Ann Saudi Med. 2022; 42(4):246-251. doi:10.5144/0256-4947.2022.246.
[15] Zarif HA, Ghandurah SE, Al-Garni MA, Binmahfooz SK, Alsaywid BS, Satti MB. Thyroid nodules cytopathology applying the Bethesda system with histopathological correlation. Saudi J Med Med Sci. 2018; 6(3):143-148.
[16] Moon HJ, Kwak JY, Kim EK, Kim MJ. Ultrasonographic characteristics predictive of nondiagnostic results for

fine-needle aspiration biopsies of thyroid nodules. Ultrasound Med Biol. 2011; 37(4):549-555.
[17] Alexander EK, Heering JP, Benson CB, et al. Assessment of nondiagnostic ultrasound-guided fine needle aspirations of thyroid nodules. J Clin Endocrinol Metab. 2002; 87(11):4924-4927.
[18] Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527.
[19] Buderer NM. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. Acad Emerg Med. 1996; 3(9):895-900.
[20] Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics. 1983; 39(1):207-215.
[21] Baloch ZW, Asa SL, Barletta JA, et al. Overview of the 2022 WHO classification of thyroid neoplasms. Endocr Pathol. 2022; 33(1):27-63.
[22] Cibas ES, Ali SZ. The 2017 Bethesda System for Reporting Thyroid Cytopathology. Thyroid. 2017;27(11):1341-1346.
[23] Sharma C. Diagnostic accuracy of fine needle aspiration cytology of thyroid and evaluation of discordant cases. J Egypt Natl Canc Inst. 2015;27(3):147-153.
[24] Bellantone R, Lombardi CP, Raffaelli M, et al. Management of cystic or predominantly cystic thyroid nodules: the role of ultrasound-guided fine-needle aspiration biopsy. Thyroid. 2004;14(1):43-47.

© 2025 Journal of Carcinogenesis | Published for Carcinogenesis Press by Wolters Kluwer-Medknow