ORIGINAL ARTICLE
Turk J Surg

Malignancy rate and independent predictors of malignancy in surgically resected Bethesda IV thyroid nodules: A retrospective cohort study

Saleh Elbalka[1], Ahmed Mahmoud Shoman[2], Islam Elzahaby[1], Farida Shokeir[3], Nada Youssef Mosaad[2], Basma Zaghloul[2], Mohamed Abdelkhalek[1]

[1]Department of Surgical Oncology, Oncology Center Mansoura University, Mansoura, Egypt
[2]Mansoura University Hospitals, Mansoura, Egypt
[3]Department of Pathology, Mansoura University Faculty of Medicine, Mansoura, Egypt

ABSTRACT

Objective: Bethesda IV thyroid nodules present a persistent diagnostic challenge, with reported malignancy rates varying widely and limited consensus on independent preoperative predictors. This study aimed to explore the malignancy rate among surgically resected Bethesda IV nodules and to identify preoperative factors independently associated with malignancy.

Material and Methods: We conducted a retrospective cohort study of consecutive patients with preoperative Bethesda IV cytology who underwent thyroid surgery at a single tertiary referral center. Only sporadic Bethesda IV nodules with available final histopathology were included. The primary outcome was histologically confirmed malignancy. The secondary outcome was clinically significant carcinoma. Preoperative variables included age, sex, focality, suspicious cervical lymph nodes on ultrasound, thyroid imaging reporting and data system (TIRADS) category, nodule size, family history of thyroid cancer, and thyroid functional status. Univariate and multivariable logistic regression analyses were performed.

Results: A total of 154 surgically resected Bethesda IV nodules were included. The overall malignancy rate was 52.6% [81/154; 95% confidence interval (CI) 44.7-60.3]. Papillary thyroid carcinoma was the most frequent malignancy subtype. On univariate analysis, multinodular goiter [odds ratio (OR) 2.31, 95% CI 1.21-4.41; p=0.011], suspicious cervical lymph nodes (OR 3.74, 95% CI 1.50-9.35; p=0.005), and higher TIRADS category (OR 2.13 per category increase, 95% CI 1.25-3.64; p=0.006) were associated with malignancy. In multivariable analysis (n=140), suspicious cervical lymph nodes (adjusted OR 3.52, 95% CI 1.17-10.63; p=0.025) and higher TIRADS category (adjusted OR 2.19 per category increase, 95% CI 1.21-3.97; p=0.010) remained independently associated with malignancy. Age, sex, nodule size, focality after adjustment, and family history were not independently associated with malignancy.

Conclusion: In this tertiary surgical cohort of Bethesda IV nodules, the histopathologic malignancy rate was 52.6%. While this elevated rate reflects inevitable surgical selection bias, higher sonographic risk tier and suspicious cervical lymph nodes were independent predictors of malignancy. These findings highlight the critical role of preoperative ultrasound assessment in refining individual risk stratification within surgically considered Bethesda IV nodules.

Keywords: Bethesda IV, cancer, endocrine surgery, thyroid surgery

INTRODUCTION

Thyroid nodules are highly prevalent in the general population, yet only a minority are malignant, making accurate preoperative risk stratification crucial to avoid both over-treatment and delayed cancer diagnosis. The Bethesda system for reporting thyroid cytopathology standardized cytologic terminology, but indeterminate categories, particularly Bethesda IV (follicular neoplasm/suspicious for follicular neoplasm), remain a persistent diagnostic challenge. Reported malignancy rates for Bethesda IV nodules vary widely across studies and practice settings, ranging from roughly 15-30% in initial Bethesda estimates to 30-60% in more recent surgical series (1-4).

This variability comes from differing referral patterns, surgical thresholds, and evolving histopathology, notably the recognition of non-invasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP) and borderline entities like follicular tumor of uncertain malignant potential (FT-UMP) under the 2022 World Health Organization (WHO) classification. In present practice, diagnostic surgery remains the standard of care because cytology cannot evaluate angiocapsular invasion.

Cite this article as: Elbalka S, Shoman AM, Elzahaby I, Shokeir F, Mosaad NY, Zaghloul B, et al. Malignancy rate and independent predictors of malignancy in surgically resected Bethesda IV thyroid nodules: a retrospective cohort study. *Turk J Surg.* [Epub Ahead of Print]

Corresponding Author
Mohamed Abdelkhalek

E-mail: mabdelkhalek@mans.edu.eg
ORCID ID: orcid.org/0000-0002-2225-500X

Received: 09.03.2026
Accepted: 16.11.2026
Epub: 20.11.2026

DOI: 10.47717/turkjsurg.2026.2026-3-15

Available at **www.turkjsurg.com**

Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of Turkish Surgical Society.
This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

While molecular testing and genomic classifiers can refine risk, they are often cost-prohibitive or unavailable in routine clinical practice. Consequently, structured ultrasound systems, such as thyroid imaging reporting and data system (TIRADS), and nodal assessment serve as the primary non-invasive guides for clinical decision-making. Still, most studies pool Bethesda III-V categories together, leaving the specific predictive utility of ultrasound within isolated Bethesda IV cohorts incompletely defined, particularly with regard to distinguishing histologic malignancy from clinically significant carcinoma (3,5-9).

In this context, we conducted a retrospective cohort study of sporadic Bethesda IV thyroid nodules managed at a tertiary referral center to address two key questions. First, we aimed to quantify both the overall malignancy rate and the proportion of clinically significant carcinoma in surgically resected Bethesda IV nodules. Second, we sought to identify independent preoperative predictors of malignancy.

MATERIAL and METHODS

Study Design and Setting

This is a retrospective cohort study conducted at a single tertiary referral center with a dedicated thyroid surgery unit. All consecutive thyroid nodules with preoperative fine-needle aspiration cytology (FNAC) reported as Bethesda IV that subsequently underwent surgery during the study period from January 2013 until January 2026 were eligible for inclusion.

This study was performed in accordance with the ethical principles of the Declaration of Helsinki. Ethical approval was granted by the Mansoura University Faculty of Medicine Institutional Review Board (MFM-IRB code: R.26.02.3561, date: 24.02.2026). Because this study was a retrospective observational archive review involving existing, de-identified clinical records without direct patient contact or prospective intervention, the requirement for research-specific informed consent was formally waived by the MFM-IRB.

Patient Selection

We identified all patients whose index FNAC for the studied nodule was categorized as Bethesda IV. Patients were included if they underwent thyroid surgery with available final histopathology of the Bethesda IV nodule; two Bethesda IV patients who did not proceed to surgery were excluded, as was one patient with MEN2B and postoperative medullary thyroid carcinoma, because the presence of germline RET mutation confers a distinct biological pathway and preoperative probability of malignancy, which could bias risk estimation for sporadic Bethesda IV nodules. No patients were documented to have been exposed to head or neck irradiation (Figure 1).

Throughout the 13-year study period, surgical resection remained the gold standard diagnostic approach to differentiate benign adenomas from carcinomas because preoperative molecular testing was not routinely available. As a tertiary oncology center, referred patients had definitive operative indications, resulting in near-universal surgical management (only two patients were managed non-operatively due to refusal or severe comorbidity). Over time, our institutional practice evolved in the direction of favoring diagnostic hemithyroidectomy over upfront total thyroidectomy for unilateral, sonographically low-risk lesions.

Radiology Assessment and TIRADS Harmonization

Thyroid ultrasound was performed and reported by dedicated sonographers at our center. Extracted variables included nodule size (maximum diameter in centimeters), focality [solitary nodule vs. multinodular goiter (MNG)], sonographic risk tier, and the presence of suspicious cervical lymph nodes (central and/or lateral compartments).

Over the 13-year study timeframe (2013-2026), reporting evolved alongside international consensus updates. Following institutional adoption of the European Thyroid Association guidelines in late 2017, EU-TIRADS categories (2 to 5) were prospectively assigned in simultaneous routine clinical reports (2018-2026, n=143, 92.9%). For earlier cases evaluated prior to formal institutional EU-TIRADS implementation (2013-2017, n=11, 7.1%), archived reports consistently documented discrete sonographic features (composition, echogenicity, margin characteristics, taller-than-wide orientation, and microcalcifications); these documented features were harmonized into corresponding EU-TIRADS categories based on published European Thyroid Association definitions without subjective image re-interpretation.

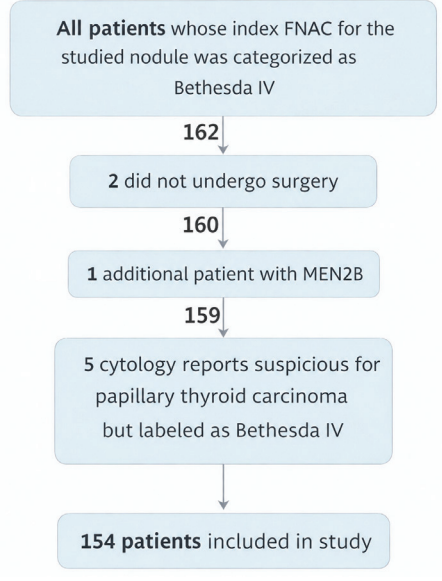


Figure 1. Study flowchart for patient selection and inclusion criteria. FNAC: Fine-needle aspiration cytology.

The criteria for labeling cervical lymph nodes as sonographically suspicious remained consistent across the study interval: presence of microcalcifications, cystic necrosis, abnormal round shape, loss of fatty hilum, or chaotic/peripheral vascularity. Cervical lymph node status was analyzed as a composite binary variable (presence vs. absence of suspicious central and/or lateral nodes). Fine-needle aspiration or thyroglobulin washout of suspicious nodes was selectively performed based on clinical discretion, and definitive nodal status (benign vs. metastatic) was confirmed on postoperative histopathologic examination of resected specimens.

Cytology and Bethesda IV Definition

FNAC was performed under ultrasound guidance using standard technique. Bethesda IV was defined according to the Bethesda System for Reporting Thyroid Cytopathology as "follicular neoplasm/suspicious for follicular neoplasm". However, for this study, Bethesda IV included reports phrased as "Hürthle cell neoplasm," reflecting terminology used in historical cytology reports over the study period and was retained solely for data extraction purposes, while cases in which the cytology was suspicious for papillary thyroid carcinoma but were labeled as Bethesda IV were excluded from the study.

FNAC was not routinely performed in all toxic nodules. Biopsy in these cases was selectively undertaken based on clinical judgment, typically in the presence of suspicious ultrasound features, discordant imaging findings, or atypical clinical presentation.

Surgery and Histopathology

All resected specimens were examined by experienced pathologists. Data were extracted from original institutional records without centralized slide re-review. In accordance with updated WHO criteria, borderline lesions (NIFTP and FT-UMP) were reported prospectively, while historical non-invasive encapsulated follicular variant papillary carcinomas lacking invasion were recorded as NIFTP for analysis. Tumor histology, invasive features (capsular, vascular, lymphatic, and perineural), extrathyroidal extension, TNM stage, and concurrent thyroiditis were systematically recorded.

The analysis was restricted to the index nodule that underwent preoperative FNAC and subsequent resection. Lesion-level concordance between cytology and final histopathology was confirmed in 100% of included cases (154/154) through tripartite correlation of preoperative ultrasound, operative surgical descriptions, and macroscopic pathology gross examination. Any case with indefinite lesion tracking was excluded during initial cohort identification.

Incidental thyroid carcinomas (specifically, occult papillary microcarcinomas ≤10 mm) were identified in 14 of the 154 patients (9.1%), discovered within the contralateral lobe (n=8) or ipsilateral background parenchyma (n=6) separate from the index nodule. These incidental foci were systematically documented but were not counted toward the primary malignancy outcome of the index lesion. Specifically, in nine cases where the index Bethesda IV nodule proved to be benign (follicular adenoma or benign nodular hyperplasia), the case was definitively categorized as non-malignant regardless of the presence of concurrent background microcarcinomas.

Preoperative Bethesda IV cytology encouraged diagnostic resection in all cases, with surgical extent determined through multidisciplinary evaluation of clinical and sonographic features. Diagnostic lobectomy was preferred for unilateral solitary nodules without suspicious lymphadenopathy, contralateral disease, or radiation history. Total thyroidectomy was performed for bilateral nodularity, compressive symptoms, high-risk sonographic features (EU-TIRADS 5), or suspicious lymph nodes. Therapeutic neck dissection was performed for suspicious or confirmed nodal metastasis, and completion thyroidectomy was reserved for high-risk invasive carcinomas identified on post-lobectomy pathology.

Outcome Definitions

The primary outcome was the malignancy rate among Bethesda IV nodules, defined as the proportion of nodules with final histology showing any invasive thyroid carcinoma, including micro-papillary thyroid carcinoma. NIFTP and FT-UMP were categorized as non-malignant lesions in line with the contemporary 2022 WHO criteria. Adenomas, benign nodular hyperplasia, and isolated thyroiditis were also considered non-malignant.

The secondary outcome was clinically significant carcinoma; an operational composite endpoint defined according to American Thyroid Association principles and the 2022 WHO classification to distinguish consequential malignancies from indolent or borderline lesions. It was defined as:

1. Any invasive thyroid carcinoma displaying gross or microscopic extrathyroidal extension;
2. Documented regional lymph node or distant metastasis;
3. Histopathologically aggressive subtypes (e.g., tall cell, columnar cell, or poorly differentiated carcinoma);
4. Invasive carcinomas measuring ≤1.0 cm in maximum diameter; or
5. Follicular thyroid carcinomas exhibiting unequivocal capsular and/or vascular invasion.

On the contrary, benign lesions (follicular adenomas, nodular hyperplasia, thyroiditis), borderline/low-risk neoplasms (NIFTP and FTUMP), and incidental papillary thyroid microcarcinomas

(<1.0 cm) lacking invasive or metastatic features were categorized as non-significant lesions. Analyses of this operational secondary endpoint were predefined as exploratory.

Predictor Variables

Candidate preoperative predictors of malignancy were selected based on clinical plausibility and prior literature: age, sex, focality [solitary thyroid nodule (STN) vs. MNG], presence of suspicious cervical lymph nodes on ultrasound (central and/or lateral), TIRADS category (modeled as ordinal), maximum nodule diameter in centimeters, family history of thyroid cancer, and thyroid functional status (euthyroid, toxic, or hypothyroid).

Statistical Analysis

The malignancy rate was calculated with 95% confidence intervals (CIs) for proportions. Baseline preoperative characteristics of the entire cohort were summarized using means and standard deviations for normally distributed data or medians and ranges for non-normally distributed continuous variables. Categorical variables were presented as counts and percentages. Comparisons between malignant and non-malignant arms were applied using appropriate univariate tests (t-test or Mann-Whitney U test for continuous variables, as well as chi-square or Fisher's exact test for categorical variables).

Because TIRADS is an ordinal variable, it was summarized descriptively as median [interquartile range (IQR)] and compared between groups using the Mann-Whitney U test but was entered into logistic regression as an ordinal numeric predictor to estimate the odds associated with each one-category increase.

To identify independent predictors of malignancy, a multivariable logistic regression model was planned, with malignancy (yes/no) as the dependent variable and (age, sex, focality, radiologic lymph-node suspicion, TIRADS, nodule size, and family history of thyroid cancer) as predictors. Given the low proportion of missing data (<10% for all variables), a complete case analysis was deemed unlikely to introduce substantial bias. Missing data were restricted exclusively to TIRADS category, which was not recorded in 14 of 154 patients (9.1%; primarily due to incomplete external referral imaging in earlier years); all other candidate predictors had complete data (0% missingness across all 154 patients). Patients with missing TIRADS data were compared with the complete-case cohort to assess potential selection discrepancies.

For the secondary exploratory endpoint (clinically significant carcinoma vs. benign/indolent lesions), unadjusted univariate comparisons were conducted using the chi-square test, Fisher's exact test, or Mann-Whitney U test as appropriate. Because this composite outcome was exploratory and highly collinear with the primary malignancy endpoint (accounting for 91.4% of all malignant lesions), multivariable modeling was restricted to the primary endpoint, and findings for the secondary outcome were interpreted as unadjusted associations.

RESULTS

A total of 154 surgically resected thyroid nodules with preoperative Bethesda IV cytology and available final-histopathology were included in the analysis. The overall malignancy rate was 52.6% (81/154; 95% CI 44.7-60.3).

Baseline Preoperative and Surgical Characteristics

Our analysis found no statistical difference in age (p=0.09) or sex distribution between malignant and benign groups (female: 87.7% vs. 86.3%, p=0.82). Nodular focality differed significantly between groups: solitary nodules were more frequent in the benign cohort, whereas MNG was more common among malignant cases (STN: 58.9% vs. 38.3%; MNG: 61.7% vs. 41.1%; p=0.017). Suspicious cervical lymph nodes on ultrasound were significantly more common in malignant nodules (28.4% vs. 9.6%, p=0.004). Radiological nodule size did not differ significantly between groups (3.03±1.38 cm vs. 3.17±1.31 cm, p=0.35). Family history of thyroid cancer was not associated with malignancy (3.7% vs. 11.0%, p=0.12). TIRADS category was significantly higher among malignant nodules. Although the median TIRADS was 4 in both groups, the distribution was shifted toward higher categories in malignant nodules [median (IQR) 4 (3.5-4) vs. 4 (3-4); p=0.005]. Thyroid functional status was not associated with malignancy (p=0.23) (Table 1).

Surgical Procedures and Extent of Resection

Among the 154 included patients, diagnostic thyroid lobectomy/hemithyroidectomy was performed in 85 patients (55.2%), while 69 patients (44.8%) underwent initial total thyroidectomy. Completion thyroidectomy was subsequently performed in 26 patients (16.9% of the total cohort; 30.6% of initial lobectomies) following an initial lobectomy revealing invasive carcinoma, leaving 59 patients (38.3%) who underwent hemithyroidectomy alone. Therapeutic cervical nodal dissection was carried out in 30 patients (19.5%), consisting of central compartment neck dissection alone in 14 patients (9.1% of the cohort; 46.7% of dissections) and combined central and lateral neck dissection in 16 patients (10.4% of the cohort; 53.3% of dissections). Initial total thyroidectomy was significantly less frequent among patients with malignant histopathology than those with benign disease [34.6% (28/81) vs. 56.2% (41/73), p=0.009], whereas lymph node dissections were significantly more frequent in the malignant cohort [34.6% (28/81) vs. 2.7% (2/73), p<0.001].

Final Histopathology

Among malignant cases, papillary thyroid carcinoma was the predominant subtype [51/81, 63.0%; comprising 44 conventional PTCs (54.3%) and 7 papillary thyroid microcarcinomas (8.6%)] followed by follicular thyroid carcinoma representing (19/81,

23.5%). Among non-malignant indolent cases, benign nodular hyperplasia accounted for 41/73 (56.2%), and follicular adenoma for 17/73 (23.3%) (Table 2).

Detailed histopathologic variants of malignant tumors, including papillary, follicular, medullary, and other carcinoma subtypes, are provided in Supplementary Table S1.

Regression Analysis

The 14 patients with unrecorded EU-TIRADS categories did not differ significantly from the 140 complete cases in malignancy rate [71.4% (10/14) vs. 50.7% (71/140), p=0.168], mean age (37.2±14.6 vs. 34.4±13.0 years, p=0.439), or sex distribution (92.9% vs. 86.4% female, p=0.696). On univariate analysis, MNG was associated with higher odds of malignancy compared with STNs [odds ratio (OR) 2.31, 95% CI 1.21-4.41; p=0.011], and suspicious cervical lymph nodes on ultrasound were strongly associated with malignancy (OR 3.74, 95% CI 1.50-9.35; p=0.005). Higher TIRADS category was also associated with malignancy (per-category OR 2.13, 95% CI 1.25-3.64; p=0.006). Sex, radiologic nodule size, family history of thyroid cancer, and thyroid functional status were not significantly associated with malignancy in univariate analyses (Table 3).

The multivariable logistic regression model was fitted using complete cases (n=140). In this model, each one-unit increase in TIRADS category was associated with approximately a 2.2-fold increase in the odds of malignancy (adjusted OR 2.19, 95% CI 1.21-3.97; p=0.01). Suspicious lymph nodes on the US were

Table 1. Baseline preoperative characteristics of surgically resected Bethesda IV nodules (n=154)

Characteristic	Total (n=154)	Malignant (n=81)	Benign/indolent (n=73)	p-value
Age (years)	34.6±13.1	33.0±12.8	36.4±13.3	0.09
Female (%)	134 (87.0)	71 (87.7)	63 (86.3)	0.82
STN (%)	74 (48.1)	31 (38.3)	43 (58.9)	
MNG (%)	80 (51.9)	50 (61.7)	30 (41.1)	**0.017**
Suspicious lateral and/or central LN (%)	30 (19.5)	23 (28.4)	7 (9.6)	**0.004**
Maximum nodule diameter (cm)	3.10±1.34	3.03±1.38	3.17±1.31	0.35
TIRADS category median (IQR)	4	4 (3.5-4)	4 (3-4)	**0.005**
Family history of thyroid cancer (%)	11 (7.1)	3 (3.7)	8 (11.0)	0.12
Thyroid function: euthyroid (%)*	131 (85.1)	67 (82.7)	64 (87.7)	
Thyroid function: hypothyroid (%)*	12 (7.8)	5 (6.2)	7 (9.6)	
Thyroid function: toxic (%)*	6 (3.9)	5 (6.2)	1 (1.4)	
Thyroid function (euthyroid/hypo/toxic)*	149 (100)	67/5/5	64/7/1	0.23

*: Thyroid function data were missing in a small number of cases (5/154 overall); IQR: Interquartile range; MNG: Multinodular goiter; TIRADS: Thyroid imaging reporting and data system; STN: Solitary thyroid nodule; LN: Lymph node.

Table 2. Final histopathologic diagnoses

Histologic diagnosis	n	%
Malignant	**81**	
Papillary thyroid carcinoma	44	54%
Follicular thyroid carcinoma	19	23%
Medullary thyroid carcinoma	4	5%
Oncocytic thyroid carcinoma	6	7%
Primary thyroid lymphoma	1	1%
Micro-papillary thyroid carcinoma	7	9%
Non-malignant	**73**	
NIFTP	4	5%
Follicular adenoma	17	23%
Benign nodular hyperplasia	41	56%
Hürthle cell adenoma	10	14%
FT-UMP	1	1%

FT-UMP: Follicular tumor of uncertain malignant potential; NIFTP: Non-invasive follicular thyroid neoplasm with papillary-like nuclear features.

Table 3. Univariate logistic regression for malignancy			
Predictor [N]	**OR**	**95% CI**	**p-value**
Age (per year) [154]	0.98	0.96-1.01	0.119
Female sex [154]	1.13	0.44-2.88	0.803
Multinodular goiter [154]	**2.31**	1.21-4.41	**0.011**
Suspicious LN on US [154]	**3.74**	1.50-9.35	**0.005**
Nodule size (per cm) [154]	0.93	0.73-1.17	0.529
TIRADS (per category) [140]	**2.13**	1.25-3.64	**0.006**
Family history thyroid cancer [154]	0.31	0.08-1.23	0.095
Dysthyroid (vs. euthyroid) [149]	1.19	0.44-3.22	0.726
TIRADS: Thyroid imaging reporting and data system; LN: Lymph node; US: Ultrasonography; CI: Confidence interval; OR: Odds ratio.			

associated with a more than three-fold increase in odds of malignancy (adjusted OR 3.52, 95% CI 1.17-10.63; p=0.025). No independent associations were observed for age, sex, focality after adjustment, nodule size, or family history of thyroid cancer (Table 4). In a sensitivity analysis using a reduced multivariable model excluding TIRADS (n=154), suspicious cervical lymph nodes remained independently associated with malignancy (adjusted OR 3.53, 95% CI 1.27-9.79; p=0.015), consistent with the primary analysis. Age and focality remained borderline, while family history of thyroid cancer showed an inverse association in this reduced model, likely reflecting sparse data and model instability (Supplementary Table S2).

Secondary Outcome (Exploratory Analysis)

In exploratory unadjusted analyses of the predefined composite endpoint, 74 nodules (48.1% of the entire cohort; 91.4% of all malignancies) were classified as clinically significant carcinomas. On unadjusted univariate analysis, sonographically suspicious cervical lymph nodes were significantly more prevalent in clinically significant cases compared with the indolent/benign group (31.1% vs. 8.8%, p=0.001). Likewise, the distribution of EU-TIRADS categories was shifted toward higher tiers among clinically significant tumors [median (IQR) 4 (3-4) vs. 4 (3-4), p=0.008). A borderline unadjusted trend was observed for MNG (60.8% vs. 43.8%, p=0.050). Age, sex, nodule diameter, family history, and thyroid functional status showed no statistically significant unadjusted associations with this endpoint (Table 5).

DISCUSSION

This retrospective cohort study of 154 surgically resected Bethesda IV thyroid nodules found an overall malignancy rate of 52.6%, placing our series at the higher end of the risk spectrum reported for this cytologic category. In multivariable analysis, higher TIRADS category and the presence of suspicious cervical lymph nodes on ultrasound were independent predictors of malignancy, whereas age, sex, nodule size, focality, and family history of thyroid cancer were not. Most of the malignant nodules 74/81 (91.4%) met criteria for clinically significant carcinoma, underscoring that, in this surgically selected cohort, malignant disease was predominantly clinically relevant rather than indolent.

Malignancy Rate in Context

Original Bethesda System estimates suggested a malignancy risk of approximately 15-30% for category IV nodules (10), but multiple subsequent series have reported higher rates, particularly in surgical cohorts. In a recent single-center study of Bethesda III-IV nodules, the malignancy rate for Bethesda IV was 51.6% almost identical to our rate (11). Other cohorts have reported Bethesda IV malignancy rates of 47-56%, like our findings and indicating that, in many real-world surgical series,

Table 4. Multivariable logistic regression for independent predictors of malignancy (complete cases, n=140)			
Variable	**Adjusted OR**	**95% CI**	**p-value**
Age (per 1-year increase)	0.97	0.94-1.00	0.072
Nodule size (per 1 cm increase)	0.95	0.72-1.25	0.7
TIRADS (per 1-unit increase)	2.19	1.21-3.97	**0.01**
Male sex (vs. female)	0.7	0.23-2.10	0.53
Solitary nodule (vs. MNG)	0.75	0.35-1.60	0.45
Suspicious LN on US (yes vs. no)	**3.52**	1.17-10.63	**0.025**
Family history of thyroid cancer (yes vs. no)	0.27	0.06-1.26	0.095
TIRADS: Thyroid imaging reporting and data system; LN: Lymph node; US: Ultrasonography; CI: Confidence interval; OR: Odds ratio; MNG: Multinodular goiter.			

Table 5. Baseline preoperative characteristics by clinically significant carcinoma (n=154)				
Characteristic	**Total (n=154)**	**Clinically significant malignant (n=74)**	**Benign/indolent (n=80)**	**p-value**
Age (years)	33.05±13.34	36.06±12.80	36.4±13.3	0.15
Female (%)	134 (87.0)	64 (86.5)	70 (87.5)	1
STN (%)	74 (48.1)	29 (39.2)	45 (56.2)	
MNG (%)	80 (51.9)	45 (60.8)	35 (43.8)	**0.05**
Suspicious lateral and/or central LN (%)	30 (19.5)	23 (31.1)	7 (8.8)	**0.001**
Maximum nodule diameter (cm)	3.10±1.34	3.04±1.41	3.15±1.29	0.61
TIRADS category median (IQR)	4	4 (3-4)	4 (3-4)	**0.008**
Family history of thyroid cancer (%)	11 (7.1)	3 (4.1)	8 (10)	0.26
Thyroid function: euthyroid (%)*	131 (85.1)	62 (83.8)	69 (86.3)	
Thyroid function: hypothyroid (%)*	12 (7.8)	6 (8.1)	6 (7.5)	
Thyroid function: toxic (%)*	6 (3.9)	6 (8.1)	0	
Thyroid function (euthyroid/hypo/toxic)*	149 (100)	62/6/6	69/6/0	0.06

[TIRADS displayed scaled; clinically corresponds to 4 (3-4) vs. 4 (3-4); Mann-Whitney p=0.008]. *: Thyroid function data were missing in a small number of cases (5/154 overall). TIRADS: Thyroid imaging reporting and data system; LN: Lymph node; US: Ultrasonography; IQR: Interquartile range; STN: Solitary thyroid nodule; MNG: Multinodular goiter.

category IV nodules carry a substantially higher risk than the original consensus suggested. These discrepancies likely reflect both referral and surgical selection biases, patients with higher-risk sonographic or clinical features are more likely to undergo surgery, as well as regional differences in iodine status, practice patterns, and the application of updated histopathologic criteria (2,3,7). Importantly, our observed 52.6% malignancy rate reflects histopathologic cancer prevalence within a surgically resected tertiary cohort, rather than the true baseline risk across all Bethesda IV nodules. In settings where preoperative molecular testing is not routinely utilized, diagnostic surgery is considered standard, and tertiary referral centers naturally accumulate nodules characterized by more concerning clinical or sonographic features. Therefore, while our data confirm that malignancy prevalence can exceed 50% among surgically selected Bethesda IV patients in specialized referral centers, this figure cannot be generalized as an unselected baseline risk for all cytologically indeterminate nodules.

Role of Ultrasound and TIRADS

The most clinically actionable finding in our study is the independent predictive value of TIRADS category and suspicious lymph nodes on ultrasound. Higher sonographic risk categories were associated with noticeably increased odds of malignancy (adjusted OR 2.19, 95% CI 1.21-3.97; p=0.010), and nodules harboring sonographically suspicious lymph nodes demonstrated more than a threefold increase in risk, even after adjustment for other preoperative clinical variables. These results are consistent with recent work showing that higher TIRADS scores are strongly associated with malignancy in indeterminate (Bethesda III/IV/V) nodules and can meaningfully refine risk estimates beyond cytology alone (8,9). Similarly, recent investigations comparing preoperative sonographic risk stratification with postoperative histopathology in cytologically indeterminate nodules (such as AUS/FLUS) have underscored that high-risk ultrasound characteristics provide crucial discriminative value for identifying true malignancy and guiding timely surgical intervention versus conservative follow-up (12).

Our findings add to this body of evidence by focusing specifically on Bethesda IV nodules and demonstrating that, within this already high-risk cytologic category, EU-TIRADS still provides meaningful additional stratification. This aligns with a large JAMA Network Open cohort that examined ACR TI-RADS in relation to cytology and histopathology and demonstrated that TI-RADS 4 and 5 nodules, particularly when larger, have substantially higher positive predictive values for malignancy and should be prioritized for intervention (13). Clinically, these findings highlight the utility of structured ultrasound in refining preoperative risk stratification for Bethesda IV nodules. Because our study evaluated diagnostic predictors rather than comparative surgical outcomes or recurrence rates, these data support using sonographic stratification to guide thorough preoperative staging, targeted nodal mapping, and individualized multidisciplinary planning, rather than dictating the definitive extent of parenchymal resection.

Interpreting suspicious cervical lymph nodes in indeterminate nodules warrants caution. While cytologically proven nodal involvement establishes definitive metastatic disease, routine sampling is often impractical, leaving surgical planning dependent on sonographic suspicion alone. In our series, 40% (12/30) of preoperatively suspicious nodes were ultimately

benign reactive hyperplasia on surgical pathology. Despite this noteworthy false-positive rate, sonographically suspicious nodes remained independently associated with a more than threefold increase in the odds of malignancy (adjusted OR 3.52). Therefore, while radiologic nodal suspicion operates as a strong preoperative risk modifier that justifies meticulous regional evaluation and planning, it should not be confused with pathologically proven disease.

Focality, Size and Other Clinical Variables

We observed that MNG was associated with malignancy in univariate analysis, with MNG conferring more than a two-fold increase in the odds of cancer compared with solitary nodules; however, this association did not persist after adjustment for TIRADS and lymph-node status. Data on focality in Bethesda IV nodules are mixed, with some series suggesting that solitary nodules confer higher risk and others find no independent effect once ultrasound features are considered. Our results support the view that focality per se is less important than the detailed sonographic phenotype of the index nodule and its regional lymph node basin. However, our findings do not support the notion that MNG confers lower malignant potential than solitary nodules once detailed ultrasound characteristics are considered (2,3,14).

Nodule size was not associated with malignancy in either univariate or multivariate analyses in our cohort. This is in line with prior studies of indeterminate nodules showing that, beyond very small (<1 cm) or very large lesions, size alone is a poor predictor of malignancy. Similarly, sex and thyroid functional status were not associated with malignancy in univariable analysis in our cohort, which is consistent with the broader thyroid cancer literature where male sex and hyperthyroidism can modify risk but are generally weaker predictors than ultrasound and cytology for individual nodules (2,8,9,13).

Family history of thyroid cancer showed a non-significant negative point estimate in the complete-case model and reached nominal significance in the reduced model without TIRADS. We caution against interpreting this as a true biological protective effect; this finding is attributable to sparse data bias and statistical instability resulting from the very small number of patients with a positive family history in our cohort (11/154, 7.1%)

Histopathology, NIFTP and Clinically Significant Carcinoma

As in other Bethesda IV series, papillary thyroid carcinoma and follicular thyroid carcinoma constituted most malignant diagnoses in our cohort, with additional cases of medullary carcinoma, other thyroid carcinoma subtypes, and primary thyroid lymphoma. On the benign/indolent side, benign nodular hyperplasia and follicular adenoma predominated, with smaller numbers of Hürthle cell adenoma, in addition to low-risk tumors including NIFTP, and FT-UMP (1,7).

The handling of NIFTP in malignancy-risk estimates has been an evolving issue since the 2016 proposal to reclassify non-invasive encapsulated follicular variant PTC (ENIFVPTC) as NIFTP, a borderline or premalignant neoplasm with extremely low recurrence risk. The current WHO classification treats NIFTP and related borderline lesions as non-malignant, and multiple studies have shown that when strict histologic criteria are applied, NIFTP behaves indolently with negligible risk of nodal or distant spread. FT-UMP as well clinically behaves in a benign fashion and warrants a more conservative approach as per WHO guidelines. We followed this contemporary approach by classifying NIFTP and FT-UMP as non-malignant (indolent) and additionally introduced a secondary endpoint of "clinically significant carcinoma", which required invasive and/or aggressive features (5,6,15,16).

Notably, 91.4% (74/81) of malignant nodules met our criteria for clinically significant carcinoma. Although unadjusted associations with higher EU-TIRADS tiers and suspicious nodes reflected our primary findings, this endpoint remains exploratory. Its primary clinical utility is descriptive: confirming that most malignancies in this surgical cohort were conventional invasive carcinomas requiring active oncologic treatment rather than occult microcarcinomas or borderline lesions. Amid growing concerns over the overdiagnosis of indolent thyroid neoplasms, separating baseline histologic malignancy from clinically consequential disease is critical for both patient counseling and health-policy perspectives.

Strengths

This study has several strengths. First, it focuses exclusively on Bethesda IV nodules with clear cytologic definitions and excludes cases reported as Bethesda IV with cytology suspicious for classic papillary carcinoma, eliminating any reporting bias and yielding a relatively homogeneous cohort that aligns closely with the intended Bethesda IV category. Second, ultrasound features including TIRADS and lymph-node status were prospectively reported by licensed sonographers in routine practice and only later extracted from the reports, avoiding the bias of retrospective image re-reading. Third, we applied WHO criteria to classify NIFTP and FT-UMP as non-malignant and introduced a clinically significant endpoint, which better reflects current thinking about overdiagnosis and overtreatment of indolent thyroid neoplasms.

Study Limitations

Chief among our limitations is the surgical selection bias being a retrospective single-center study. Because the denominator was restricted to resected specimens in a tertiary referral cancer center, these findings cannot establish the population-

based malignancy risk across all Bethesda IV nodules, which includes indolent lesions managed conservatively in primary or secondary care settings. Accordingly, our 52.6% malignancy rate reflects histologic prevalence within an operative cohort rather than an unselected baseline cytologic risk probability.

Second, the retrospective evaluation of ultrasound features over a 13-year timeframe introduces potential variability. Although EU-TIRADS was assigned prospectively following guideline adoption in late 2017, earlier cases (2013-2017) required category harmonization based on standardized descriptors in archived reports, introducing potential interobserver or temporal inconsistency. Additionally, modeling TIRADS as an ordinal predictor assumes comparable risk increases between successive tiers; while higher categories clearly reflect greater risk, the exact per-category OR should be viewed as an indicative risk gradient rather than an absolute mathematical increase.

Third, with over half of the cohort presenting with MNG, incidental background carcinomas could theoretically confuse outcome attribution. We mitigated this by establishing strict lesion-level concordance across ultrasound, operative descriptions, and gross pathology in all 154 cases, while explicitly excluding incidental non-index microcarcinomas from the primary endpoint. The consistency of our primary and secondary findings confirms that incidental microfoci did not drive our conclusions.

Fourth, histopathologic data were extracted from original institutional records without centralized glass-slide re-review. Although historical non-invasive encapsulated follicular variant papillary carcinomas lacking invasion were recorded as NIFTP based on detailed microscopic descriptions, evolving diagnostic nomenclature over the 13-year study period carries potential for subtle diagnostic drift and inter-pathologist variability over time.

Fifth, although missing data was minimal and confined exclusively to EU-TIRADS (9.1%), comparisons between the complete-case model (n=140) and the alternative reduced model (n=154) demonstrated parameter instability for variables with sparse event counts, particularly family history. Because model reduction evaluates an alternative specification rather than directly assessing missing-data bias, effect estimates for sparse variables must be interpreted cautiously.

Finally, our secondary endpoint of "clinically significant carcinoma" represents an operational combined measure rather than an externally validated international benchmark. Because multivariable modeling was not performed for this composite outcome, reported associations remain unadjusted and exploratory, and conclusions regarding independent predictors apply strictly to the primary histopathologic malignancy model.

Despite these limitations, our findings reinforce the central role of high-quality ultrasound in the evaluation of Bethesda IV nodules and suggest that TIRADS category and nodal assessment should be explicitly integrated into preoperative risk stratification and surgical planning. In clinical practice, recognizing that malignancy risk varies widely according to sonographic appearance can enhance preoperative counseling and shared decision-making. Bethesda IV nodules presenting low-to-intermediate sonographic risk and normal cervical nodes represent a group where the probability of malignancy is relatively lower, strengthening the role of conservative diagnostic lobectomy in eligible patients. Conversely, identifying high-risk sonographic phenotypes or suspected nodal involvement needs dedicated preoperative neck mapping and multidisciplinary consensus before establishing the operative plan. Prospective studies assessing long-term disease-free survival, surgical complications, and completion thyroidectomy rates remain necessary to determine whether risk-adapted surgical algorithms improve patient outcomes compared with standard practice. Future research should aim to validate our findings in larger, multicenter cohorts, ideally incorporating standardized TIRADS, detailed nodal mapping, and molecular markers to build robust, externally validated risk models tailored to Bethesda IV nodules (11).

CONCLUSION

In this single-center cohort of 154 surgically resected Bethesda IV thyroid nodules, the malignancy rate exceeded half (52.6%), and more than 90% of malignant lesions met predefined criteria for clinically significant carcinoma. Within this already high-risk cytologic category, higher TIRADS category and the presence of suspicious cervical lymph nodes on ultrasound were the only consistent independent predictors of malignancy. These findings suggest that structured ultrasound assessment, specifically TIRADS scoring combined with systematic nodal evaluation, should be central to preoperative decision-making in this cytologic category.

Ethics

Ethics Committee Approval: Ethical approval was granted by the Mansoura University Faculty of Medicine Institutional Review Board (MFM-IRB code: R.26.02.3561, date: 24.02.2026).

Informed Consent: Because this study was a retrospective observational archive review involving existing, de-identified clinical records without direct patient contact or prospective intervention, the requirement for research-specific informed consent was formally waived by the MFM-IRB.

Footnotes

Author Contributions

Concept - S.E., M.A.; Design - S.E., M.A.; Data Collection or Processing - I.E., F.S., B.Z.; Analysis or Interpretation - A.M.S., I.E., F.S.; Literature Search - A.M.S., I.E., B.Z.; Writing - S.E., M.A.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

Supplementary Tables: https://d2v96fxpocvxx.cloudfront.net/688d2d00-d207-464d-89b6-73f393f4f50c/content-images/f1a1a0c7-3a02-432f-8b20-2b15ae3e7c6f.pdf

REFERENCES

1. *Azzam EZ, Salah MA, Aboelwafa WA, Essam RM, Bondok ME. Rates and predictors of malignancy in bethesda III and IV thyroid nodules: a prospective study. Cureus. 2024;16:e76615.*
2. *Zahid A, Shafiq W, Nasir KS, Loya A, Abbas Raza S, Sohail S, et al. Malignancy rates in thyroid nodules classified as Bethesda categories III and IV; a subcontinent perspective. J Clin Transl Endocrinol. 2021;23:100250.*
3. *Loderer T, Bonati E, Donato V, Viani L, Cozzani F, Del Rio P. Malignancy risk in Bethesda class IV thyroid nodules in an iodine deficient region. Gland Surg. 2023;12:884-893.*
4. *Godoi Cavalheiro B, Kober Nogueira Leite A, Luongo de Matos L, Palermo Miazaki A, Marcel Ientile J, Aurelio V Kulcsar M, et al. Malignancy rates in thyroid nodules classified as Bethesda categories III and IV: retrospective data from a tertiary center. Int J Endocrinol Metab. 2018;16:e12871.*
5. *Paniza AC de J, Mendes TB, Viana MDB, Thomaz DMD, Chiappini PBO, Colozza-Gama GA, et al. Revised criteria for diagnosis of NIFTP reveals a better correlation with tumor biological behavior. Endocr Connect. 2019;8:1529-1538.*
6. *Bai Y, Kakudo K, Jung CK. Updates in the pathologic classification of thyroid neoplasms: a review of the world health organization classification. Endocrinol Metab (Seoul). 2020;35:696-715.*
7. *Alyusuf EY, Alhmayin L, Albasri E, Enani J, Altuwaijri H, Alsomali N, et al. Ultrasonographic predictors of thyroid cancer in Bethesda III and IV thyroid nodules. Front Endocrinol (Lausanne). 2024;15:1326134.*
8. *George NA, Suresh S, Jiji V, Renu S, Thomas S, Janardhan D, et al. Correlation of TIRADS and Bethesda scoring systems with final histopathology of thyroid nodules - an institutional experience. Indian J Otolaryngol Head Neck Surg. 2022;74:5753-5758.*
9. *He YP, Xu HX, Zhao CK, Sun LP, Li XL, Yue WW, et al. Cytologically indeterminate thyroid nodules: increased diagnostic performance with combination of US TI-RADS and a new scoring system. Sci Rep. 2017;7:6906.*
10. *Cibas ES, Ali SZ. The Bethesda system for reporting thyroid cytopathology. Thyroid. 2009;19:1159-1165.*
11. *Khan AA, Jasim NKE, Al-Mannai NEAH, Ata F, Goyal R, Jaber T. Clinical outcomes of patients with Bethesda III or IV cytology on fine needle aspiration of thyroid nodules-a retrospective study. Endocrinol Diab Metab. 2025;8:e70076.*
12. *Demir F, Hançerlioğulları O, Uçaner B, Buldanlı MZ, Özkara M. Comparison of preoperative imaging and FNAB results with postoperative pathology results in patients undergoing AUS/FLUS. Gazi Med J. 2024;35:291-296.*
13. *Huang EYF, Kao NH, Lin SY, Jang IJH, Kiong KL, See A, et al. Concordance of the ACR TI-RADS classification with Bethesda scoring and histopathology risk stratification of thyroid nodules. JAMA Netw Open. 2023;6:e2331612.*
14. *Elbalka SS, Metwally IH, Shetiwy M, Awny S, Hamdy O, Kotb SZ, et al. Prevalence and predictors of thyroid cancer among thyroid nodules: a retrospective cohort study of 1,000 patients. Ann R Coll Surg Engl. 2021;103:683-689.*
15. *Juhlin CC, Mete O, Baloch ZW. The 2022 WHO classification of thyroid tumors: novel concepts in nomenclature and grading. Endocr Relat Cancer. 2023;30:e220293.*
16. *Yuan Y, Shu H, Li L, Wu L, Yu F. A new scoring system for risk stratification of thyroid tumors. BMC Med Imaging. 2025;25:114.*