Accuracy and diagnostic performance of the Bethesda system for reporting thyroid cytopathology in a tertiary endocrine surgical referral center in Belgium

Accuracy and diagnostic performance of the Bethesda system for reporting thyroid cytopathology in a tertiary endocrine surgical referral center in Belgium.

Kinet Sam, MD[1,2](ORCID: 0000-0001-7220-3489)*; Cornette Hendrik, MD*[2]*; Van Den Heede Klaas, MD*[2](ORCID: 0000-0003-3642-9514)*; Brusselaers Nele, MD, PhD*[3,4,5](ORCID: 0000-0003-0137-447X)*; Van Slycke Sam, MD, PhD*[2,4,6](ORCID: 0000-0002-8292-9475)

1 *Faculty of Medicine, KU Leuven, Herestraat 49, 3000 Leuven, Belgium*

2 *Department of General and Endocrine Surgery, Onze-Lieve-Vrouw (OLV) Hospital Aalst-Asse-Ninove, Moorselbaan 164, 9300 Aalst, Belgium*

3 *Centre for Translational Microbiome Research Department of Microbiology, Tumour and Cell biology, Karolinska Institutet, Karolinska Hospital, Tomtebodavagen 16, 17165 Stockholm, Sweden*

4 *Department of Head and Skin, University Hospital Ghent, Corneel Heymanslaan 10, 9000 Ghent, Belgium*

5 *Global Health Institute, University of Antwerp*, Doornstraat 331, 2610 Wilrijk, Belgium

6 *Department of General Surgery, AZ Damiaan, Gouwelozestraat 100, 8400 Ostend, Belgium*

Corresponding author

Dr. Sam Kinet, MD

Address: Moorselbaan 164, 9300 Aalst, Belgium

Phone number: +32 (0)53 72 41 11

E-mail: sam.kinet@a-kinetics.be

ORCID: 0000-0001-7220-3489

Twitter handle: @SamKinetMD.

Author contributions

Sam Kinet: Acquisition of data - Analysis and interpretation of data - Drafting of revised manuscript.

Klaas Van Den Heede: Acquisition of data - Analysis and interpretation of data - Critical revision of manuscript - Study conception and design.

Hendrik Cornette: Acquisition of data - Analysis and interpretation of data - Drafting of manuscript.

Nele Brusselaers: Analysis and interpretation of data - Critical revision of manuscript - Study conception and design.

Sam Van Slycke: Analysis and interpretation of data - Critical revision of manuscript - Study conception and design.

Funding and Conflicts of Interest

The authors have no conflicts of interest.

No funding was received.

Ethical Requirements

The authors declare that this study complies with the journal's ethical policies.

Short Title

Validation of the Bethesda system in a referral center.

Keywords

Bethesda, Endocrine, Fine-needle aspiration, Surgery, Thyroid

Word Count

Abstract: 249 words.

Manuscript: 2636 words.

1 figure, 3 tables.

A total of 20 pages.

Other Correspondence

An abstract based on this study with results was presented at the 24th Belgian Surgical Week in Ostend, Belgium.

Abstract

Background: The Bethesda System for Reporting Thyroid Cytopathology is a commonly used classification for fine needle aspiration (FNA) cytology of suspicious thyroid nodules. The risk of malignancy (ROM) for each category has recently been analyzed in three international databases. This paper compares the diagnostic performance of the Bethesda classification in a high-volume referral center in Belgium.

Methods: All consecutive thyroid procedures were registered in a prospective database from January 2010 till August 2022. Patient and surgical characteristics, preoperative Bethesda categories, and postoperative pathology results were analyzed.

Results: Out of 2219 consecutive thyroid procedures, 1226 patients underwent preoperative FNA. Papillary thyroid cancer was the most prevalent malignancy (N=119, 70.4%), followed by follicular (N=17, 10.1%), and medullary thyroid cancer (N=15, 8.9%). Micropapillary thyroid cancer was incidentally found in 46 (3.8%) patients. Bethesda categories I, II, III, IV, V, and VI respectively represented 250 (20.4%; ROM 4.4%), 546 (44.5%; ROM 3.8%), 96 (7.8%; ROM 20.8%), 231 (18.8%; ROM 15.2%), 62 (5.1%; ROM 72.6%), and 41 (3.3%; ROM 90.2%) patients. Overall ROM was 13.8%. An NPV of 96.2% was found. Overall specificity was 64.2% with a positive predictive value (PPV) of 31.9%. Diagnostic accuracy was 67.8%. Compared to international databases (CESQIP, EUROCRINE, UKRETS), ROM in this study appeared lower for Bethesda category IV (15.2 vs 26.7%, *p=0.612*).

Conclusion: Despite being validated in numerous studies, ROM based on preoperative FNA cytology classified according to the Bethesda classification may vary amongst surgical centers and countries as this study reveals a higher NPV and lower PPV.

Introduction

Thyroid nodules are highly prevalent as about 5% of the adult population has palpable nodules in the thyroid region and up to 70% of adults show thyroid nodules on neck ultrasound [1]. Most thyroid nodules are asymptomatic and are detected by patients themselves or during a routine check-up. The most common benign cause of thyroid nodules are adenomas, single or as part of a multinodular goiter, however, thyroid cancer is seen in 7-15% of incidentally found nodules [1].

Incidence of thyroid cancer has strongly increased over the last few decades. According to the Global Cancer Observatory (IARC), 586 202 new cases of thyroid cancer were estimated worldwide in 2020, with an age-standardized rate of 10.1/100 000 and 3.1/100 000 in women and men respectively. Recent data from the Belgian Cancer Registry also reveal an increase in thyroid cancer as in 2004 an incidence of 3.1/100 000 in men and 8.9/100 000 in women was seen compared to an incidence of 5.2/100 000 in men and 13.1/100 000 in women in 2017.

Increased detection of asymptomatic thyroid nodularity due to liberal use of thyroid ultrasound has been the most important cause of the elevated incidence of thyroid cancer [2]. A less aggressive diagnosis and treatment of asymptomatic, incidentally found nodules has stabilized the number of surgical procedures in recent years [3]. When referred for a thyroid nodule, an ultrasound of the neck will be carried out [4].

The American Thyroid Association (ATA) guidelines to assess and classify thyroid ultrasound findings are used in many countries. When a nodule is considered suspicious for its size based on several criteria, fine needle aspiration (FNA) is offered to the patient, as it is the most accurate and cost-effective method of evaluating thyroid nodules [5].

The ACR Thyroid Imaging, Reporting and Data System (TI-RADS) classifies ultrasound findings into five categories, each with an ascending suspicion for malignancy [6]. TI-RADS 1 to 5 are respectively considered “Benign”, “Not Suspicious”, “Mildly Suspicious”, “Moderately

Suspicious" and "Highly Suspicious". Nodules classified as TI-RADS 1 and 2 do not warrant additional FNA. FNA is considered appropriate for TI-RADS 3 nodules ≥ 2.5 cm, TI-RADS 4 nodules ≥ 1.5 cm and TI-RADS 5 nodules ≥ 1 cm.

'The Bethesda System for Reporting Thyroid Cytopathology' is the most used classification for FNA cytology. The Bethesda classification divides cytology specimens into six categories: I 'Nondiagnostic' or 'Unsatisfactory', II 'Benign', III 'Atypia of undetermined significance (AUS)' or 'Follicular lesion of undetermined significance' (FLUS), IV 'Follicular neoplasm' or 'Suspicious for a follicular neoplasm', V 'Suspicious for malignancy', and VI 'Malignant'. Each of these categories is linked to a specific risk of malignancy (ROM) and subsequently to an evidence-based clinical guideline for further diagnosis and treatment [7]. Recently, the ROM for each Bethesda category has been assessed within three international databases [8]. Whether these ROMs apply to individual centers remains to be analyzed. Differences in health care organization and quality, as well as environmental factors and treatment policies might influence local results. Results of big datasets should not be blindly applied to local centers [9].

The main aim of this paper is to analyze the diagnostic performance of The Bethesda classification within a high-volume, tertiary referral center in Belgium. Second objective is to analyze possible differences with the international data and evaluate how this could impact future clinical and surgical behavior.

Materials and Methods

All patients who underwent thyroid surgery in a single tertiary referral center (OLV hospital, Aalst, Belgium) were consecutively included in an ongoing, prospectively gathered, endocrine-surgical database from January 2010 onwards. Within this database, a study cohort was retrospectively compiled with patients who received thyroid surgery up to August

2022. Types of surgery included total thyroidectomy, hemithyroidectomy, isthmusectomy, or completion thyroidectomy. All procedures were performed by the same, experienced endocrine surgeon (SVS), with a personal activity of over 200 thyroidectomies per year. Patients were excluded if no preoperative FNA was performed or if the Bethesda category was missing on FNA cytology report (Figure 1). Part of the study cohort has already been described [10, 11]. All patients provided written informed consent prior to the study. Demographics, FNA cytology, data on surgical and associated procedures, and the final histopathology report were collected.

Preoperative Bethesda categories on FNA cytology were compared to postoperative histopathological classifications of the resected specimen to obtain the ROM for each category. The 2017 WHO classification was used to classify thyroid cancer [12]. Sensitivity, specificity, positive (PPV) and negative predictive values (NPV), as well as diagnostic accuracy and risk of malignancy (ROM) of given Bethesda categories were calculated. Incidental thyroid malignancies (i.e., separate from the index nodule that received FNA) consisted of micropapillary thyroid carcinomas and were excluded from calculations of ROM. Bethesda category I is nondiagnostic, therefore sensitivity, specificity, PPV and NPV are meaningless within this category. Bethesda category II on FNA cytology was considered true negative when the final histopathology report was benign, and false negative when the result was malignant. Indications for surgery in patients with Bethesda II thyroid nodules consisted of symptomatic thyroid enlargement, compressive symptoms, Graves' disease, bleeding cysts, branchial cleft cysts, Hashimoto thyroiditis, and other forms of thyroiditis. Bethesda categories III and IV are "indeterminate" as they do not differentiate between benign and malignant. However, because a lobectomy is one of the suggested approaches for these categories, the histology result could be considered "positive" for possible malignancy. To assess the impact of Bethesda categories III and IV on diagnostic accuracy, different calculations were made including and excluding these categories. For analysis, Bethesda categories III, IV, V and VI on FNA cytology were considered true positive when the final

histopathology report was malignant, and false positive when the result was benign. Non-invasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP) was not considered malignant and its effect on ROM was not studied.

Results were compared to recently published data evaluating ROM for each Bethesda category within three international databases (CESQIP, EUROCRINE, and UKRETS) [8].

All quantitative results are presented as median with interquartile ranges (IQR). All statistical analyses were conducted in STATA® (StataCorp, V.16·1/MP).

Results

From January 2010 until August 2022, a total of 2219 consecutive thyroid surgeries were carried out. Of these thyroid surgeries, 1226 patients received preoperative FNA with cytology and a subsequent Bethesda classification. A final histopathology report was obtained for all resected specimens (Figure 1). Median age of the study cohort was 53 years (IQR 43 - 63). Sex ratio (female/male) was 3.94. Out of 1226 included patients, 250 (20.4%) were preoperatively diagnosed as Bethesda category I, 546 (44.5%) as category II, 96 (7.8%) as category III, 231 (18.8%) as category IV, 62 (5.1%) as category V, and 41 (3.3%) as category VI (Table 1).

NIFTP was found in nine patients. A total of 169 malignancies were found (Table 2), the majority of which were papillary carcinomas (N= 119, 70.4%), followed by follicular carcinomas (N=17, 10.1%), and medullary carcinomas (N=15, 8.9%). Other malignant findings consisted of Hürthle cell carcinoma (N=6, 3.6%), anaplastic carcinoma (N=5, 3.0%), metastases of other primary malignancies (N=4, 2.4%), poorly differentiated carcinoma (N=2, 1.2%), and lymphoma (N=1, 0.6%). ROM was 4.4% for Bethesda category I, 3.8% for category II, 20.8% for category III, 15.2% for category IV, 72.6% for category V, and 90.2% for category VI. Overall ROM was 13.8%, excluding incidentally found micropapillary thyroid cancer, which was found in 46 cases.

Lesions preoperatively classified as benign (Bethesda category II) were malignant or false negative in 3.8% of cases, which leads to an NPV of 96.2%.

Lesions preoperatively classified as malignant or highly suspicious (Bethesda categories V and VI) were benign or false positive in 20.4% of cases, leading to a specificity of 96.2%. For Bethesda categories III and IV, 83.2% false positives were seen. Together, this resulted in a total of 68.1% false positives for categories III, IV, V, and VI. Given the low ROM in Bethesda categories III and IV, overall specificity reduced to 64.2%.

If only “conclusive” Bethesda categories II, V, and VI were included, a sensitivity of 79.6% was found.

Individual PPVs for Bethesda categories III, IV, V, and VI were 20.8%, 15.2%, 72.6% and 90.2% respectively. An overall PPV of 31.9% was found. Diagnostic accuracy of the Bethesda classification was 67.8%.

In both the studied cohort (44.5%) and international cohorts (32.4%) Bethesda category II is most represented (Table 3). Respective percentages of Bethesda categories I, III, IV, V and VI are 20.4% in the studied cohort versus 6.7% in international data, 7.8% versus 14.5%, 18.8% versus 21.6%, 5.1% versus 7.7% and 3.3% versus 17.2%. ROM of Bethesda category V in the studied cohort (72.6%) was comparable to international data (73.7%). ROM of Bethesda categories I (4.4%), II (3.8%), III (20.8%), IV (15.2%) and VI (90.2%) in the studied cohort was lower than the reported ROM of international data (respectively 13.6%; 7.8%; 24.5%; 26.7% and 95.4%).

Discussion

With 1226 included cases over a time span of over 12 years, the diagnostic accuracy of the Bethesda classification at the OLV Aalst was 67.8%. With the exception of Bethesda

category IV, which showed a lower ROM than Bethesda category III in the studied cohort, a comparable ROM to international data was seen, with Bethesda category I having a higher ROM than Bethesda category II in both the studied cohort and international data.

Thyroid cytology following FNA is a very important part of the diagnostic work-up of thyroid nodules. A clinician needs a performant, non- to little invasive, low-cost pre-operative technique to distinguish benign from malignant thyroid nodules. Cytopathology was standardized by implementation of the Bethesda classification in 2010, with an update in 2018 withholding a higher ROM for the lower Bethesda categories. By using the Bethesda classification clinicians can approximate a ROM for each patient, which in turn has an impact on disease management and decision for surgery. Differences in ROM of Bethesda category I can be explained by the uncertainty related to this category. When Bethesda category I is concluded on FNA, either FNA is repeated, conservative treatment is offered, or surgery is carried out in case of suspicious clinical or ultrasound findings. A different approach to this uncertainty is reflected in the amount of Bethesda category I diagnoses included in the studied cohort (20.4%) compared to the amount in international data (6.7%). A lower ROM of Bethesda category II in the studied cohort (3.8%) might reflect a higher accuracy of benign detection and is associated with a lower rate of false negatives (3.8% versus 7.8%) and a higher NPV (96.2% versus 92.2%) compared to international data. For the “indeterminate” Bethesda categories III and especially IV, the lower ROM in the studied cohort (respectively 20.8% and 15.2% versus 24.5% and 26.7%) might confirm the uncertainty of these diagnoses. This uncertainty also explains the higher false positive rates when Bethesda categories III and IV are included in the false positive calculations.

Differences in ROM of Bethesda category VI can be seen as a statistical pitfall, as only 41 patients (3.3%) received a Bethesda category VI diagnosis in the relatively small, studied cohort of 1226 patients. Overall ROM in the studied cohort (13.8%) was much lower than international data (33.7%) and more comparable to data reported in Bethesda guidelines [7]. An interindividual difference in ROM of Bethesda categories I to IV between the three major

databases featured in the review by Inabnet *et al* is also seen, which reflects variation in the application of Bethesda guidelines in each center [8]. Lower ROM in Bethesda categories III, IV, and VI might reflect a more cautious interpretation of guidelines when examining FNA specimens. The results suggest overtreatment of Bethesda III and IV nodules compared to the literature, with similar treatment of Bethesda V nodules.

Other single-center studies of surgical patients who received preoperative FNA, report distributions of Bethesda I between 3.0 - 4.2% [13, 14, 15]; of Bethesda II between 20.6 – 36.0% [13, 14, 15]; of Bethesda III between 11.5 – 28.0% [13, 14, 15, 16]; of Bethesda IV between 7.5 – 24.9% [13, 14, 15, 16]; of Bethesda V between 3.2 – 8.9% [13, 14, 15, 16]; and of Bethesda VI between 18.0 – 25.8% [13, 14, 15]. In comparison, findings in this study show a higher proportion of Bethesda I and II, a lower proportion of Bethesda III and VI, and comparable proportions of Bethesda IV and V. ROM in these single-center studies varies between 0.0 – 29.0% for Bethesda I [13, 14, 15, 17]; between 2.8 – 11.0% for Bethesda II [13, 14, 15, 17]; between 7.3 – 51.0% for Bethesda III [13, 14, 15, 16, 17, 18]; between 15.5 – 57.9% for Bethesda IV [13, 14, 15, 16, 17, 18]; between 65.0 – 100.0% for Bethesda V [13, 14, 15, 16, 17]; and between 96.5 – 98.8% for Bethesda VI [13, 14, 15, 17]; with total ROM varying between 30.0 – 61.0% [13, 14, 15]. ROMs in this study are similar, except for a slightly lower ROM for Bethesda IV, a lower ROM for Bethesda VI, and lower overall ROM. These percentages may implicate differences in FNA quality between centers. Some of these discrepancies and the wide ranges of distribution and ROM can also be explained using molecular testing before surgery in some studies, and the exclusion of malignancies not related to the index nodules (i.e., the nodule in which FNA was performed). For indeterminate nodules (Bethesda III, IV and V on FNAC), consideration of molecular testing is recommended for further diagnosis [19]. Molecular testing has been shown to predict aggressiveness of thyroid malignancies [20], to decrease the surgical rate in patients with indeterminate nodules [21] and to guide optimal management of Bethesda VI nodules [22]. Molecular testing or gene expression profiling was not offered to patients given the higher

cost, lack of reimbursement, and the absence of specific recommendations regarding this topic in Belgium.

Distribution of histological subtypes of malignancies in the studied cohort was grossly similar to international cohorts. Less frequent subtypes such as anaplastic carcinoma, metastases of other primary malignancies, poorly differentiated carcinoma, lymphoma, and C-cell hyperplasia (which was not observed in the studied cohort) were not fully proportional to international data, presumably due to the smaller population in this study.

Limitations of this study include the relatively small study population with higher risk of statistical pitfalls. Another limitation is the risk of certain types of bias. Since only patients who underwent thyroid surgery were included, patients who received conservative treatment after FNA remain unnoticed, which may lead to selection bias. The studied population might differ from other databases given a different approach to Bethesda I (i.e., lower threshold for surgery) and higher portion of Bethesda II with more symptomatic benign goiters. The database used in this study was prospectively gathered by different assessors, which might cause information or diagnostic bias. Differentiation between index nodule or malignancy separate from index nodule was not possible on gathered information. It is possible that some malignancies were not index nodules, therefore affecting FNA accuracy.

A strength of this study is the single center approach with standardized work-up and all surgeries being performed by a single surgeon, thus limiting interobserver bias. Furthermore, the strength of diagnostic pathway execution in the study center (OLV hospital, Aalst, Belgium) is reflected by the study cohort since all FNAs were followed by cytopathological examination and allocation to a Bethesda category.

This study shows that even though specific guidelines exist, the diagnostic accuracy of FNA and the Bethesda classification differs between centers. The studied population included a greater proportion of patients with benign FNA results who received surgery, which alters ROM compared to other databases. Reported ROMs for each Bethesda category by

international guidelines can serve as indicators, but values should not be blindly copied and applied to individual centers. Centers should evaluate their own results and preoperative ROM based on specific center data should be discussed with patients instead of, or in combination with data in literature.

References

1. Wong R, Farrell SG, Grossmann M (2018) Thyroid nodules: diagnosis and management. Med J Aust 209(2):92-98

2. Van Den Heede K, Tolley NS, Di Marco AN, et al (2021) Differentiated Thyroid Cancer: A Health Economic Review. Cancers (Basel) 13(9): 2253

3. Roman BR, Morris LG, Davies L (2017) The thyroid cancer epidemic, 2017 perspective. Curr Opin Endocrinol Diabetes Obes 24(5):332-336

4. Detweiler K, Elfenbein DM, Mayers D (2019) Evaluation of Thyroid Nodules. Surg Clin North Am 99(4):571-586

5. Haugen BR, Alexander EK, Bible KC, et al (2016) 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid 26(1):1-133

6. Tessler FN, Middleton WD, Grant EG, et al (2017) ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. J Am Coll Radiol 14(5):587-595

7. Cibas ES, Ali SZ (2017) The 2017 Bethesda System for Reporting Thyroid Cytopathology. Thyroid 27(11):1341-1346

8. Inabnet WB 3rd, Palazzo F, Sosa JA et al (2020) Correlating the Bethesda System for Reporting Thyroid Cytopathology with Histology and Extent of Surgery: A Review of 21,746 Patients from Four Endocrine Surgery Registries Across Two Continents. World J Surg 44(2):426-435

9. Househ M, Aldosari B (2017) The Hazards of Data Mining in Healthcare. Stud Health Technol Inform 238:80-83

10. Van Slycke S, Van Den Heede K, Bruggeman N, et al (2021) Risk factors for postoperative morbidity after thyroid surgery in a PROSPECTIVE cohort of 1500 patients. Int J Surg 88:105922

11. Van Slycke S, Simons AS, Van Den Heede K, et al (2021) Combined cervicosternotomy and cervicotomy for true retrosternal goiters: a surgical cohort study. Updates Surg 73(4):1-10

12. Bai Y, Kakudo K, Jung CK (2020) Updates in the Pathologic Classification of Thyroid Neoplasms: A Review of the World Health Organization Classification. Endocrinol Metab (Seoul) 35(4):696-715

13. Kiernan CM, Broome JT, Solórzano CC (2014) The Bethesda system for reporting thyroid cytopathology: a single-center experience over 5 years. Ann Surg Oncol 21(11):3522-3527

14. Magister MJ, Chaikhoutdinov I, Schaefer E, et al (2015) Association of Thyroid Nodule Size and Bethesda Class With Rate of Malignant Disease. JAMA Otolaryngol Head Neck Surg 141(12):1089-1095

15. Linhares SM, Handelsman R, Picado O, et al (2021) Fine needle aspiration and the Bethesda system: Correlation with histopathology in 1,228 surgical patients. Surgery 170(5):1364-1368

16. Baraf L, Avidor Y, Bahat Dinur A, et al (2023) Rates of Malignancy in Cytology Indeterminate Thyroid Nodules: A Single Center Surgical Series. Isr Med Assoc J 25(2):147-151

17. Acar Y, Doğan L, Güven HE, et al (2017) Bethesda Made It Clearer: A Review of 542 Patients in a Single Institution. Oncol Res Treat 40(5):277-280

18. Yaprak Bayrak B, Eruyar AT (2020) Malignancy rates for Bethesda III and IV thyroid nodules: a retrospective study of the correlation between fine-needle aspiration cytology and histopathology. BMC Endocr Disord 20(1):48

19. Patel KN, Yip L, Lubitz CC, et al (2020) The American Association of Endocrine Surgeons Guidelines for the Definitive Surgical Management of Thyroid Disease in Adults. Ann Surg 271(3):e21-e93

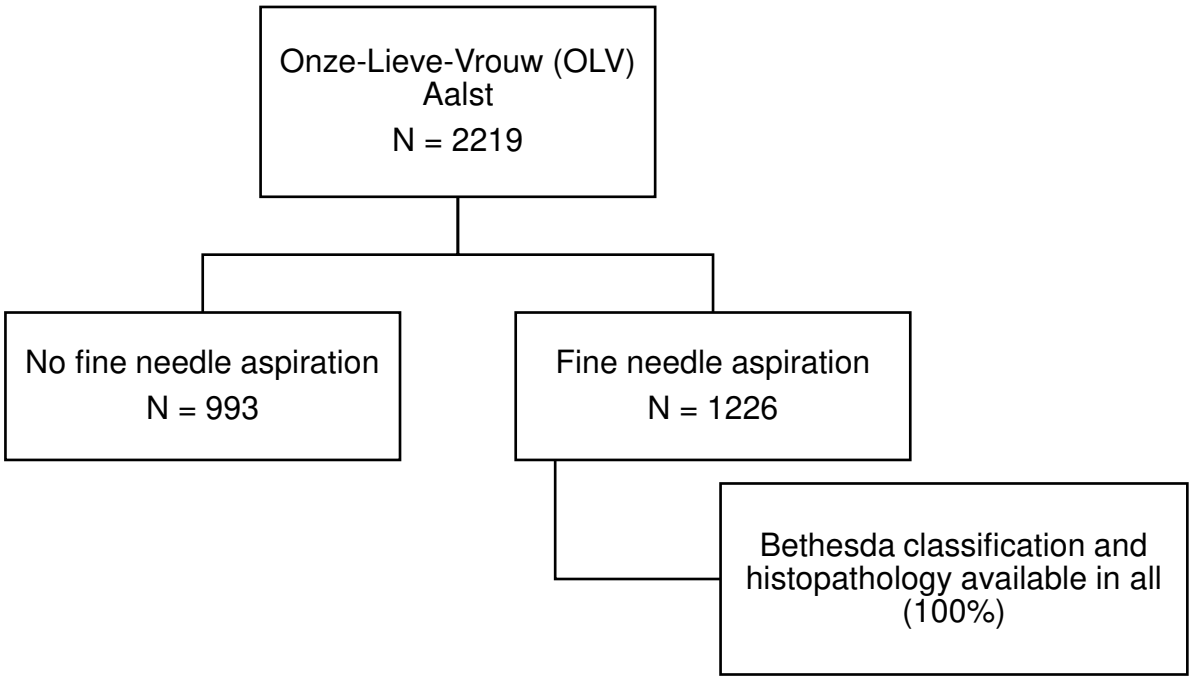
20. Krasner JR, Alyouha N, Pusztaszeri M, et al (2019) Molecular mutations as a possible factor for determining extent of thyroid surgery. J Otolaryngol Head Neck Surg 48(1):51

21. Kandil E, Metz TA, Issa PP, et al (2023) Diagnostic Performance of Afirma and Interpace Diagnostics Genetic Testing in Indeterminate Thyroid Nodules: A Single Center Study. Cancers (Basel) 15(7):2098

22. Hier J, Avior G, Pusztaszeri M, et al (2021) Molecular testing for cytologically suspicious and malignant (Bethesda V and VI) thyroid nodules to optimize the extent of surgical intervention: a retrospective chart review. J Otolaryngol Head Neck Surg 50(1):29

Figures

Figure 1 Flow Chart of Study Cohort



Tables

Table 1 Distribution, malignancies, ROM, and test statistics.

	Bethesda I	Bethesda II	Bethesda III	Bethesda IV	Bethesda V	Bethesda VI	Total
Number of cases, N (%)	250 (20.4)	546 (44.5)	96 (7.8)	231 (18.8)	62 (5.1)	41 (3.3)	1226 (100)
Malignancies, N	11	21	20	35	45	37	169
ROM, %	4.4	3.8	20.8	15.2	72.6	90.2	13.8
False negatives, %	NA	3.8	NA	NA	NA	NA	NA
False positives, %	NA	NA	83.2		20.4		NA
			68.1				
Sensitivity, %							86.7
Specificity, %							64.2
NPV, %							96.2
PPV, %							31.9
Diagnostic accuracy, %							67.8

N: Number of cases; ROM: Risk of malignancy; NA: Not applicable; NPV: Negative predictive value; PPV: Positive predictive value.

Table 2 Histology by Bethesda category.

Histology, % of malignancies (N)	Bethesda I	Bethesda II	Bethesda III	Bethesda IV	Bethesda V	Bethesda VI	Total
Papillary	72.7	66.7	70.0	57.1	82.2	70.3	70.4 (119)
Follicular	9.1	23.7	15.0	17.1	4.4	0.0	10.1 (17)
Medullary	0.0	4.8	5.0	5.7	6.7	21.6	8.9 (15)
Hürthle cell	0.0	0.0	0.0	17.1	0.0	0.0	3.6 (6)
Anaplastic	9.1	0.0	0.0	0.0	2.2	8.1	3.0 (5)
Metastatic	9.1	4.8	5.0	0.0	2.2	0.0	2.4 (4)
Poorly differentiated	0.0	0.0	5.0	2.9	0.0	0.0	1.2 (2)
Lymphoma	0.0	0.0	0.0	0.0	2.2	0.0	0.6 (1)

N: Number of cases.

Table 3 Comparison with international data [8].

	Bethesda I	Bethesda II	Bethesda III	Bethesda IV	Bethesda V	Bethesda VI	Total
Number of cases, N (%)							
CESQIP [8]	269 (3.2)	2394 (28.5)	1714 (20.4)	1320 (15.7)	705 (8.4)	1994 (23.7)	8396 (100)
UKRETS [8]	747 (11.1)	1736 (25.8)	713 (10.6)	2232 (33.2)	424 (6.3)	879 (13.1)	6731 (100)
Eurocrine [8]	436 (6.6)	2911 (44.0)	722 (10.9)	1138 (17.2)	552 (8.3)	860 (13.0)	6619 (100)
Pooled international data [8]	1452 (6.7)	7041 (32.4)	3149 (14.5)	4690 (21.6)	1681 (7.7)	3733 (17.2)	21746 (100)
Studied cohort	250 (20.4)	546 (44.5)	96 (7.8)	231 (18.8)	62 (5.1)	41 (3.3)	1226 (100)
ROM, %							
Bethesda guidelines [7]	5 – 10	0 – 3	6 – 18	10 – 40	50 – 75	97 – 99	NA
CESQIP [8]	16.2	9.9	27.4	34.4	74.5	95.5	42.4
UKRETS [8]	11.6	6.2	22.3	22.7	72.7	97.6	29.0
Eurocrine [8]	15.6	7.1	20.2	26.2	73.6	92.7	27.9
Pooled international data [8]	13.6	7.8	24.5	26.7	73.7	95.4	33.7
Studied cohort	4.4	3.8	20.8	15.2	72.6	90.2	13.8

SD: Standard deviation; N: Number of cases; ROM: Risk of malignancy; NA: Not applicable; CESQIP: The Collaborative Endocrine Surgery Quality Improvement Program; UKRETS: UK Registry of Endocrine and Thyroid Surgery.

[7] Cibas ES, Ali SZ (2017) The 2017 Bethesda System for Reporting Thyroid Cytopathology. Thyroid 27(11):1341-1346

[8] Inabnet WB 3rd, Palazzo F, Sosa JA et al (2020) Correlating the Bethesda System for Reporting Thyroid Cytopathology with Histology and Extent of Surgery: A Review of 21,746 Patients from Four Endocrine Surgery Registries Across Two Continents. World J Surg 44(2):426-435