Diagnostic Accuracy of Bethesda Classification of Thyroid Nodules at a Kenyan Hospital: A Retrospective Study

Aurelia Brazeal[1], *Adili Wobenjo*[2], *Chege Macharia*[3], *Morgan Otido*[4], *Rosemary Kamau*[5], *Jana MacLeod*[2]

[1]KEMRI-Wellcome Trust, Clinical Trial Facility, Kilifi, Kenya

[2]Kenyatta University School of Medicine, Department of Surgery, Nairobi, Kenya

[3]AIC Kijabe Hospital, Department of Surgery, Kijabe, Kenya

[4]The Kenya Navy Hospital, Department of Surgery, Mtongwe, Mombasa, Kenya

[5]AIC Kijabe Hospital, Department of Pathology, Kijabe, Kenya

Correspondence to: Aurelia Brazeal; email: aurelia.brazeal@gmail.com

Received: 14 Sep 2023; Revised: 25 Jul 2024; Accepted: 26 Jul 2024; Available online: 5 Aug 2024

Abstract

Background: The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) was developed to standardize the reporting of thyroid nodule fine-needle aspiration cytology (FNAC) results. The adoption of TBSRTC is based on the implied risk of malignancy per category, but this has shown wide variation in different regions worldwide. **Aim:** The aim of this study is to determine the diagnostic accuracy of the Bethesda classification of thyroid nodules in a Kenyan hospital. **Methods:** A retrospective longitudinal study examined FNAC and histopathology data of thyroid gland nodules at a Kenyan hospital from 2010 to 2019; specimens from 347 patients were found. An analysis was performed to determine the diagnostic accuracy of the Bethesda classification in detecting malignancy among these patients. **Results:** The malignancy rate was 16.1%. The risk of malignancy as per the Bethesda category was as follows: non-diagnostic—19.6%, benign—6%, atypia of undetermined significance—20%, follicular neoplasm—16.7%, suspicious for malignancy—80%, and malignant—85.7%. The diagnostic properties of FNAC defining "malignant" and "suspicious for malignancy" categories as malignant were as follows: sensitivity—64.7%, specificity—97.6%, and accuracy—92.7%. **Conclusion:** The Bethesda classification had high diagnostic accuracy, with the risk of malignancy consistent with the Bethesda findings. Classifying malignancy as "malignant" and "suspicious for malignancy" yielded the most favorable diagnostic properties for FNAC.

Keywords: Fine-needle aspiration, Risk of malignancy, The Bethesda System for Reporting Thyroid Cytopathology, Thyroid, Diagnostic accuracy

Ann Afr Surg. 2024; 21(4): 120-126
DOI: http://dx.doi.org/10.4314/aas.v21i4.5

Conflict of interest: None
Funding: None

Introduction

Assessing thyroid nodules in patients is an essential step in clinical decision-making. According to Tamhane and Gharib (1), thyroid nodules have a reported prevalence of 65%, with the majority being benign. Approximately

7–15% of these nodules have thyroid cancer, one of the fastest-growing cancers (2). In 2020, the global incidence of thyroid cancer stood at 6.6 in 100,000, while in eastern Africa the incidence was 2.2 in 100,000 (3). Fine-needle aspiration cytology (FNAC) serves as initial investigations in the evaluation of thyroid nodules.

The fine-needle aspiration (FNA) procedure is performed to obtain the cytology of thyroid nodules. It aids in differentiating between benign and malignant nodules (3). The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) was developed to establish a consensus among cytopathologists regarding the reporting of thyroid nodule FNAC (4). According to the Bethesda system (4-6), the cytological diagnosis of thyroid FNA is divided into six categories: non-diagnostic, benign, atypia of undetermined significance, follicular neoplasm, suspicious for malignancy, and malignant. Grouping the indeterminate categories as positive or negative for malignancy influences the test characteristics of the Bethesda classification (7, 8). Each category has an implied risk of malignancy and recommended clinical management (4-6).

Table 1. The 2023 Bethesda System for Reporting Thyroid Cytopathology: recommended diagnostic categories

Bethesda category	Description
Bethesda I Non-diagnostic	FNAC diagnosed as cyst fluid only, virtually acellular specimen, unsatisfactory smears, hemorrhagic content, old hemorrhagic content, air dry, benign fat, lymphoid cells, and no cells.
Bethesda II Benign	FNACs reported as features consistent with follicular nodular disease (includes adenomatoid nodule, colloid nodule, etc.), consistent with chronic lymphocytic (Hashimoto) thyroiditis in the proper clinical context, consistent with granulomatous (subacute) thyroiditis. Others as nodular goiter, multinodular goiter, adenomatous goiter, benign thyroid tissue.
Bethesda III Atypia of unknown significance	FNACs reported as atypia of undetermined significance.
Bethesda IV Follicular neoplasm	FNACs reported as follicular neoplasm. Specify if oncocytic (formerly Hurthle cell) type.
Bethesda V Suspicious for malignancy	FNACs diagnosed as suspicious for papillary carcinoma, suspicious for medullary carcinoma, suspicious for metastatic carcinoma, and suspicious for lymphoma.
Bethesda VI Malignant	FNACs reported as papillary thyroid carcinoma, high-grade follicular-derived carcinoma, medullary thyroid carcinoma, undifferentiated (anaplastic) carcinoma, squamous cell carcinoma, carcinoma with mixed features (specify), metastatic malignancy, and non-Hodgkin lymphoma.

FNAC, fine-needle aspiration cytology.

TBSRTC has continuously refined the diagnostic categories and the associated risk of malignancy (6). The risk of malignancy differs significantly across regions (7-10), and the application of TBSRTC classification varies across Africa. Therefore, the acceptance of TBSRTC must ensure that the malignancy risk is consistent with local data, as malignancy rates differ among populations and geographic regions. The local evaluation of the malignancy risk is helpful to improve decision-making for appropriate treatment (11). Using data from the pathology registry of a local hospital in Kenya, we aim to investigate the diagnostic accuracy of TBSRTC classification.

Materials and Methods

We conducted a retrospective longitudinal study of patients who underwent a thyroid FNAC after presenting with a thyroid nodule at a semi-rural hospital between January 1, 2010, and December 31, 2019. We retrieved and analyzed data from patients who underwent FNAC and subsequent histopathology from the pathology registry using the following variables:

age, sex, cytology, type of surgery performed, and histopathologic diagnosis.

Thyroid FNAs were performed by surgeons and cytopathologists. All thyroid cytology samples were reviewed, reported, and finalized by general pathologists, each with a postgraduate training in pathology for a minimum duration of 3 years. TBSRTC reporting system was not routinely used in the registry. In cytological reports that did not have the additional Bethesda classification, the pathologists reviewed the cytological descriptions and classified them according to the guidelines of TBSRTC reporting system (4-6). The description and classification of the cytopathology specimen as per TBSRTC reporting system can be seen in Table 1.

In patients whose FNAC was repeated, the repeated cytology result was considered conclusive. The indications for repeat FNAC are non-diagnostic and atypia of unknown significance categories.

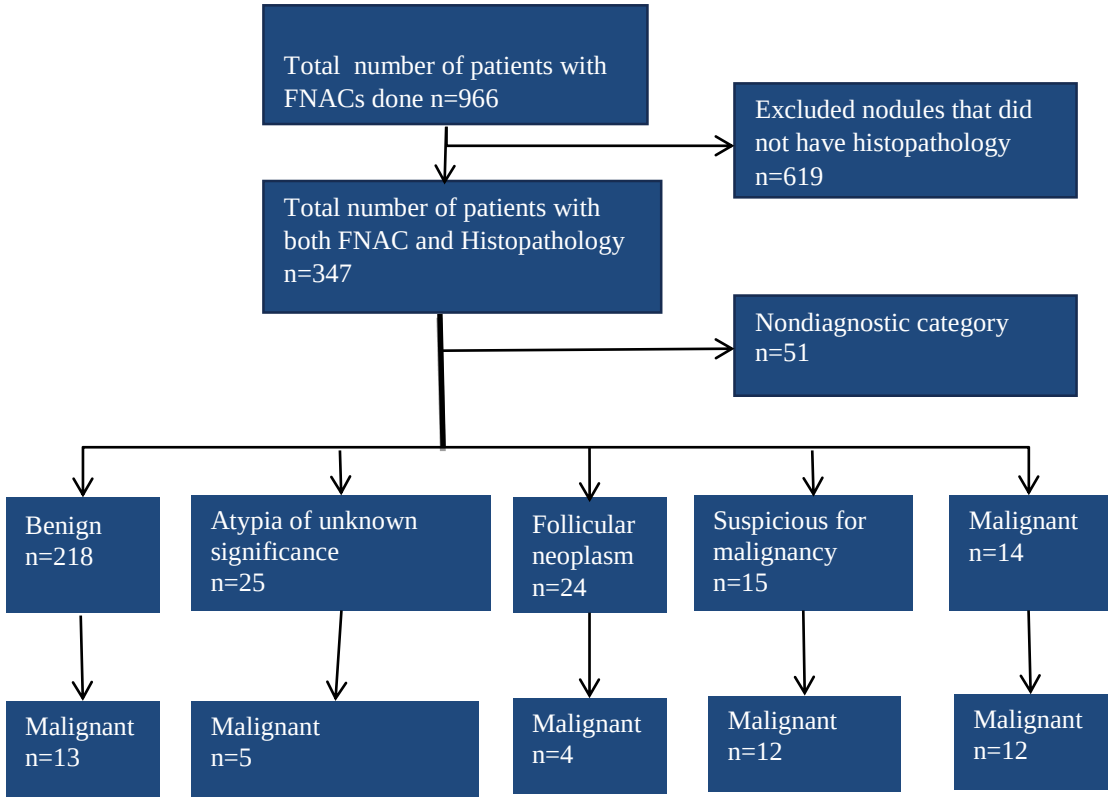

Figure 1. Flow Diagram of Thyroid Nodules and procedures done leading to Malignant histopathology results.

Data cleaning and statistical analysis

The data collected from the pathology registry were compiled into a Microsoft Excel database which we then cleaned and analyzed. For statistical analysis, we used Jamovi version 2.3.18 (https://www.jamovi.org), a statistical software built on the R statistical language.

We considered the histopathology results of thyroid nodules as the gold standard and assessed the correlation between TBSRTC and histopathology results by calculating the risk of malignancy. We divided the number of malignant cases on histopathology by the total number in each category of the Bethesda classification.

For TBSRTC, we calculated the test characteristics: sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, negative likelihood ratio, and accuracy. In all analyses, we excluded the "non-diagnostic" category and considered the "benign" category as true negative. We considered the following definitions of positive for

malignancy in different scenarios and performed the analyses: (a) "malignant" category; (b) "malignant" and "suspicious for malignancy" categories; (c) "malignant", "suspicious for malignancy" and "follicular neoplasm" categories; and (d) "malignant" "suspicious for malignancy" "follicular neoplasm" and "atypia of undetermined significance" categories.

We described categorical data (Bethesda categories, gender, and histopathology) using count and frequency percentages and used cross-tabulation and chi-square test to determine the association between these variables. We described continuous variables as means with standard deviation. In our analyses, we used a *p* value <0.05 to indicate statistical significance.

Ethical consideration

The study received ethical approval (Ref: KH/ISERC/02718/0040/2023) from the Ethics Committee, a nationally accredited ethics board.

Results

Demographic characteristics

Over the 10-year period, samples from 347 patients were submitted for both FNAC and histopathology (Figure 1). The cohort had a mean age of 44 years (SD=13.4). The median was 42 years (15–83) and 45 years (20–73) for females and males, respectively.

Table 2. Distribution of Bethesda categories and histopathology findings

Bethesda category	Distribution of Bethesda categories	Benign on histopathology	Malignant on histopathology	Risk of malignancy
Non-diagnostic	14.7%	41	10	19.6%
Benign	62.8%	205	13	6.0%
Atypia of unknown significance	7.2%	20	5	20.0%
Follicular neoplasm	6.9%	20	4	16.7%
Suspicious for malignancy	4.3%	3	12	80.0%
Malignant	4%	2	12	85.7%
Total	**100%**	**291**	**56**	**16.1%**

Table 3. Histopathologic diagnoses

Distribution of histopathology diagnoses (n=347)		
Benign (n=291, 83.9%)		
Nodular hyperplasia	104	30.0%
Multinodular	71	20.5%
Follicular adenoma	42	12.1%
Colloid goiter	24	6.9%
Hashimoto thyroiditis	13	3.7%
Others (adenoma, cyst, thyroiditis, etc.)	37	10.7%
Malignant (n=56, 16.1%)		
Papillary carcinoma	37	10.7%
Follicular carcinoma	9	2.6 %
Others	10	2.9%

The age at presentation with thyroid nodules did not differ between genders (p = 0.073). Most patients were females (87%). However, the proportion of male patients with thyroid malignancies (18.2%) was higher than that of females (9.4%) with a relative risk of 1.11 (p = 0.001). This suggests a potential gender difference in malignancy risk within the population.

Correlation of FNAC and histopathology findings

The distribution of FNACs according to TBSRTC and the histopathological diagnoses is shown in Tables 2 and 3, respectively. In this cohort, 10 FNACs were repeated and reclassified. The correlation between the Bethesda classification and the histopathology revealed a linear association between high Bethesda categories and the risk of malignancy (Table 2).

Several interpretations of positive results for malignancies are presented in Table 4. Considering the "malignant" category as positive for malignancy

resulted in a false-negative and false-positive rate of 6% and 14.3%, respectively. Histopathologically, 8 of 13 false-negative FNAC samples were papillary carcinoma, and the 2 false-positive FNAC samples were follicular adenomas.

The test sensitivity improved with the addition of other Bethesda categories as positive with an associated reduction in specificity and diagnostic accuracy (Table 4). The diagnostic accuracy was excellent in the "malignant" and "suspicious for malignancy" categories, with the area under the receiver operating characteristic (AUROC) curve being the largest (0.812) (Table 4).

Table 4. Test characteristics of the Bethesda classification system

Positive for malignancy	Sensitivity	Specificity	Accuracy	PPV	NPV	LR+	LR−	AUROC
Malignant	48.0%	99.0%	93.5%	85.7%	94.0%	49.7	0.525	0.735
Malignant and Suspicious for malignancy	64.9%	97.6%	92.7%	82.8%	94.0%	27.2	0.36	0.812
Malignant, Suspicious for malignancy, and Follicular neoplasm	68.3%	89.1%	86.0%	52.8%	94.0%	6.28	0.356	0.787
Malignant, Suspicious for malignancy, Follicular neoplasm, and Atypia of unknown significance	68.3%	89.1%	86.0%	52.8%	94.0%	6.28	0.356	0.769

AUROC, area under the receiver operating characteristic curve; LR+, positive likelihood ratio; LR−, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value.

Discussion

The 347 patients with thyroid nodules had a mean age of 44 years (SD=13.4), a finding comparable to other studies (12, 13). The majority of nodules were diagnosed as benign by FNAC, consistent with the published literature (7, 8, 12-14). However, the proportion of thyroid nodules classified as non-diagnostic was high (14.7%). Ali (15) proposed that each laboratory should limit the proportion of non-diagnostic categories to <10%. Limiting this proportion requires the assessment of thyroid FNA for sample adequacy (5, 6). According to TBSRTC guidelines, thyroid nodules classified as non-diagnostic require repeat FNA under ultrasound guidance. In our study, a small proportion of thyroid FNAs were repeated in this category. The use of ultrasound was not documented. This highlights an important area for improvement of clinical practice to improve the diagnostic yield of thyroid FNA.

Papillary carcinoma was the most common thyroid malignancy in our study, similar to the regional published literature (12-14). There is an increasing trend towards papillary thyroid carcinoma. (12) The iodization of salt for goiter prevention is ascribed to the increasing frequency of papillary thyroid carcinoma in iodine-deficient areas (12, 16). However, there is conflicting evidence on this (17).

The malignancy rate was 16.1%. Masereka et al. (12) reported a malignancy incidence rate of 13.1% in Uganda, while Bongiovanni et al. (7) reported an average incidence of 33.4% in a meta-analysis. The lower malignancy rate is likely due to the lower incidence of thyroid cancer in East Africa (3). This regional variation in the malignancy rate highlights the importance of considering geographical factors when interpreting findings and the need for a tailored approach in the management of patients in different populations.

The cohort's risk of malignancy in the Bethesda diagnostic categories was consistent with TBSRTC (6), except for the "malignant" category (Table 5). Misclassification contributes to false-positive and false-negative rates. The explanations for the misclassification in the FNAC results are divided into sampling problems and interpretation errors (18).

Sampling problems include sampling error, suboptimal samples, and inadequate slide preparation.

Table 5. Comparison of cohort malignancy in our study with the Bethesda classification (Ali et al)

Bethesda category	Risk of malignancy (current study)	Risk of malignancy TBSRTC (Ali et al., 2023)
Non-diagnostic	19.6%	13% (5–20%)
Benign	6.0%	4% (2–7%)
Atypia of unknown significance	20.0%	22% (13–30%)
Follicular neoplasm	16.7%	30% (23–34%)
Suspicious for malignancy	80.0%	74% (67–83%)
Malignant	85.7%	97% (97–100%)

TBSRTC, The Bethesda System for Reporting Thyroid Cytopathology.

The interpretation errors occur when cytopathologists under-diagnose cases to avoid false-positive results (19). In our study, FNA sampling was performed by surgeons and pathologists, without image guidance, which may contribute to misclassification. The use of the Thyroid Imaging, Reporting and Data System (TI-RADS) improves the detection of thyroid nodules requiring FNAC (20).

The Bethesda classification demonstrated low sensitivity, high specificity, and high false-positive rates in our study, comparable to the findings by Masereka et al. (12). In contrast, Bongiovanni et al. (7), Sheffield et al. (8), and Abdullahi et al. (14) reported high sensitivity, specificity, and accuracy. These studies reported a high incidence of thyroid malignancies, the utilization of ultrasound to detect nodules at high risk of malignancy, and image-guided FNA. The false-positive rate (14.3%) was high in our study. According to Malheiros et al. (21), reported false-positive rates range from 2% to 10%. Sidawy et al. (22) proposed that FNACs have limited accuracy in identifying follicular pattern lesions, papillary microcarcinoma, and cystic papillary thyroid carcinoma. Contributing to this are the diagnostic features of papillary cancer that occur in benign conditions such as nodular goiter, adenomatous goiter, follicular neoplasm, and thyroiditis (3). This could explain the false-positive results in our study. Despite the sensitivity and specificity observed in our study, the diagnostic accuracy of the Bethesda classification was generally very good to excellent (86 – 93.5%) and was consistent with the results of the meta-analysis by Bongiovanni et al. (7). The AUROC value, which was highest for the "malignant" and "suspicious for malignancy" categories, suggested that this definition of malignancy was the most accurate in correctly classifying patients with thyroid malignancies.

The limitations of our study include its retrospective nature and therefore challenges of missing data and ambiguity in some cytopathology results. Additionally, the study cohort may not be representative of the general population. We recommend prospective research on the malignancy risk using TI-RADS and FNAC in pre-operative decision-making as this will lead to a better understanding of their diagnostic role in a goiter endemic population. This will also offer insights into optimizing patient management strategies.

Conclusion

The risk of malignancy in our cohort was comparable to TBSRTC findings. In our study, defining positive for malignancy as malignant and suspected malignancy provided the best diagnostic properties of the Bethesda classification. This definition improves the prediction of malignancy risk in the local population and facilitates more informed decision-making regarding patient management. Overall, the Bethesda classification should be used routinely in the evaluation of patients with thyroid nodules.

Author contributions

AB led in conceptualization and data curation, AW led in formal analysis, methodology and in review & editing of the original draft, CM led in resources and JM led in conceptualization and supervision. All authors equally contributed to project administration, validation, visualization and in writing of the original draft.

References

1. Tamhane S, Gharib H. Thyroid nodule update on diagnosis and management. Clin Diabetes Endocrinol. 2016; 2: 17.

2. Grussendorf M, Ruschenburg I, Brabant G. Malignancy rates in thyroid nodules: a long-term cohort study of 17,592 patients. Eur Thyroid J. 2022; 11(4): e220027.
3. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid. 2016; 26(1): 1-133.
4. Cibas ES, Ali SZ. The Bethesda System for Reporting Thyroid Cytopathology. Am J Clin Pathol. 2009; 132(5): 658-65.
5. Cibas ES, Ali SZ. The 2017 Bethesda System for Reporting Thyroid Cytopathology. Thyroid. 2017; 27(11): 1341-6.
6. Ali SZ, Baloch ZW, Cochand-Priollet B, et al. The 2023 Bethesda System for Reporting Thyroid Cytopathology. Thyroid. 2023; 33(9): 1039-44.
7. Bongiovanni M, Spitale A, Faquin WC, et al. The Bethesda System for Reporting Thyroid Cytopathology: a meta-analysis. Acta Cytol. 2012; 56(4): 333-9.
8. Sheffield BS, Masoudi H, Walker B, et al. Preoperative diagnosis of thyroid nodules using the Bethesda System for Reporting Thyroid Cytopathology: a comprehensive review and meta-analysis. Expert Rev Endocrinol Metab. 2014; 9(2): 97-110.
9. Straccia P, Rossi ED, Bizzarro T, et al. A meta-analytic review of the Bethesda System for Reporting Thyroid Cytopathology: Has the rate of malignancy in indeterminate lesions been underestimated? TBSRTC for indeterminate lesions. Cancer Cytopathol. 2015; 123(12): 713-22.
10. Zahid A, Shafiq W, Nasir KS, et al. Malignancy rates in thyroid nodules classified as Bethesda categories III and IV; a subcontinent perspective. J Clin Transl Endocrinol. 2021; 23: 100250.
11. Hathi K, Rahmeh T, Munro V, et al. Rate of malignancy for thyroid nodules with AUS/FLUS cytopathology in a tertiary care center – a retrospective cohort study. J Otolaryngol Head Neck Surg. 2021; 50(1): 58.
12. Masereka R, Okeny PK, Fualal JO, et al. Diagnostic accuracy of fine needle aspiration cytology in patients undergoing thyroidectomy in Uganda: tertiary hospital experience. Afr Health Sci. 2017; 16(4): 1143.
13. ElShallaly GEH, Ibrahim BAB, M. E. Salih MME, et al. Histopathologic Pattern of Thyroid Disease in 1351 Thyroidectomy Patients. Sudan J Med Sci. 2022; 17(4): 577-88.
14. Abdullahi IM, Yasin NA, Dirken ES, et al. Comparative study of fine needle aspiration cytology and histopathology in thyroid nodules at a tertiary care hospital: First report from Somalia. Asian J Surg. 2023; 46(10): 4202-7.
15. Ali SZ. Thyroid cytopathology: Bethesda and beyond. Acta Cytol. 2011; 55(1): 4-12.
16. Zeng Z, Li K, Kang W, et al. Changing patterns of thyroid cancer in different stages of Universal Salt Iodization in Peking Union Medical College Hospital, 1986–2018. Gland Surg. 2020; 9(5): 1338-45.
17. Zhang X, Zhang F, Li Q, et al. Iodine nutrition and papillary thyroid cancer. Front Nutr. 2022; 9: 1022650.
18. Zhu Y, Song Y, Xu G, et al. Causes of misdiagnoses by thyroid fine-needle aspiration cytology (FNAC): our experience and a systematic review. Diagn Pathol. 2020; 15(1): 1.
19. Park JH, Yoon SO, Son EJ, et al. Incidence and Malignancy Rates of Diagnoses in the Bethesda System for Reporting Thyroid Aspiration Cytology: An Institutional Experience. Korean J Pathol. 2014; 48(2): 133.
20. 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): 587-95.
21. Malheiros DC, Canberk S, Poller DN, et al. Thyroid FNAC: Causes of false-positive results. Cytopathology. 2018; 29(5): 407-17.
22. Sidawy MK, Del Vecchio DM, Knoll SM. Fine-needle aspiration of thyroid nodules: correlation between cytology and histology and evaluation of discrepant cases. Cancer. 1997; 81(4): 253-9.