The Bethesda System for Reporting Thyroid Cytopathology: Validating at Tribhuvan University Teaching Hospital

Kunjan Acharya[1]   Shreya Shrivastav[2]   Prashant Triipathi[3]   Bigyan Raj Gyawali[4]   Bijaya Kharel[1,2]
Dharma Kanta Baskota[5]   Pallavi Sinha[2]

[1] Tribhuvan University Teaching Hospital, Kathmandu, Nepal
[2] Department of Pathology, Tribhuvan University Teaching Hospital, Kathmandu, Nepal
[3] Tribhuvan University Teaching Hospital, Kathmandu, Nepal
[4] Deparment of ENT, Tribhuvan University Teaching Hospital, Kathmandu, Nepal
[5] Tribhuvan University, Kathmandu, Nepal

Address for correspondence Dr. Kunjan Acharya, Tribhuvan University Teaching Hospital Ringgold standard institution – ENT, Kathmandu, Kathmandu 44600, Nepal (e-mail: kunjan.acharya@iom. edu.np).

Abstract

Introduction   Fine needle aspiration cytopathology (FNAC) is widely used for the stratification of thyroid nodules.

Objective   The objective of the present study is to validate FNAC reporting based on The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) at our institution and to calculate the risk of malignancy in each category.

Methods   This was a descriptive cross-sectional study conducted jointly at the Department of Ear, Nose and Throat and at the Department of Pathology for a period of 1.5 years (May 2018 to November 2018). All cases presenting with thyroid swelling in the outpatient department were investigated with ultrasonography (USG) of the neck, thyroid function test, and FNAC. All FNAC reporting was done according to TBSRTC.

Results   A total of 134 thyroidectomies were performed during the study period. The female to male ratio was 5.3:1. The age ranged from 11 to 74 years old. with a mean age of 51 years old. The FNAC has a specificity of 84.9%, a sensitivity of 89.4%, a positive predictive value of 86.4%, a negative predictive value of 88.2%, and an accuracy of 87.3% in detecting thyroid cancer. The implied risk of malignancy (ROM) in Bethesda II, III, IV, V and VI is 11.7%, 25%,40%,76.6% & 96%, respectively.

Conclusion   The four studied categories had a ROM comparable to other studies, except for the Bethesda III category. Further studies with larger sample sizes and with the use of USG guidance for the aspiration from the thyroid swelling may give better results by reducing the number of false negative and false positive cases.

Keywords
► thyroid nodule
► thyroid neoplasm
► biopsy
► fine needle

Introduction

Thyroid nodules are a very common pathology encountered by otorhinolaryngologists anywhere in the world. They are reported in 5% of the general population in the United States of America (USA) by palpation, and in up to 50% by ultrasound or autopsy[1,2] Fine needle aspiration cytopathology (FNAC) is an accepted modality of investigation in the evaluation of thyroid nodules. It is widely used throughout the world as it is a safe, simple, and cost-effective method for the stratification of thyroid nodules. Fine needle aspiration cytopathology was pioneered by Martin et al. in 1930.[3]

Fine needle aspiration cytopathology reporting by an expert pathologist can help diagnosing thyroid malignancies requiring surgery, thus reducing the burden of unnecessary surgery and its associated morbidities. In view of this, the National Cancer Institute hosted a meeting in in 2007 in Bethesda, Md, USA, framing The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC). It defined the reporting system into 6 different diagnostic criteria, with each having a different risk of malignancy. After 4 years of recommendation, Bongiovanni et al.[4] conducted a meta-analysis using TBSRTC. He compared TBSRTC with the histological outcomes, which showed that TBSRTC is an effective and robust thyroid FNAC classification scheme to guide the clinical management of patients with thyroid nodules.

Although we are following the TBSRTC based on the internationally accepted data, we have not yet validated it in the Nepalese population. Since we are a referral tertiary care center, we planned to provide the validation of the results at our center in order to obtain an overview of the outcome of the system at our center and, therefore, in the country. Only then we can compare our results with the international data and communicate our outcome to international communities. Although there are studies on the accuracy of FNAC in the diagnosis of thyroid lesions, none are done at our center following TBSRTC.

The objective of the present study is to validate FNAC reporting based on TBSRTC at our institution and to report the risk of malignancy in each category.

Material and Methods

This was a descriptive cross-sectional study conducted jointly at the Department of Ear, Nose and Throat (ENT) and at the Department of Pathology. The present study was conducted over a period of 1.5 year (May 2018 to November 2019). All cases presenting with thyroid swelling in the outpatient department were investigated with ultrasonography (USG) of the neck, thyroid function test, and FNAC. Despite the fact that USG-guided FNAC confers more diagnostic accuracy than palpation-guided FNAC,[5] concerns like economical constraints of the patients from rural parts of the country, and a high chance of delay in the diagnosis resulting from an overdue appointment, made us restrict its use in cases with Bethesda I and III lesions. Furthermore, very small lesions with difficulty in palpation and posteriorly located lesions were also considered for USG-guided FNAC.

For routine FNAC, all cases were assessed by physical examination and localized by palpation. Four to 10 air-dried and alcohol-fixed smears were prepared for Geimsa stain and Papanicolaou stain, respectively. Cell blocks and on-site examination of FNAC material were not available. Smears were examined by two pathologists, and a Bethesda category was assigned by consensus.

All FNAC reporting was done according to TBSRTC. The different categories of TBSRTC, as well as the risk of malignancy, are shown in ►Table 1, as described by Cibas et al.[6] If the reporting of the lesion was Bethesda I, we further repeated the FNAC under USG guidance, as recommended by TBSRTC. If multiple FNAC samples from a same patient yielded two different diagnoses, only the diagnosis with higher malignant potential was considered for further analysis. After analyzing the USG of the neck, the thyroid function test (TFT), and FNAC, we planned cases for total or hemithyroidectomy with or without neck dissection based on the American Thyroid Association (ATA) guidelines.[7] Patients of any age and of both genders whose FNAC reporting was performed according to TBSRTC and had their final histological report available were included in the study. There was a total of 134 cases meeting the inclusion criteria during the study period. The sampling method was by convenience sampling, and all cases during the study period were included in the study. The present study was approved by the institutional review committee (IRC) of the Institute with the reference number no-378.

We used IBM SPSS Statistics for Windows, version 24 (IBM Corp., Armonk, NY, USA) for the analysis of our results. To determine the specificity, sensitivity, positive predictive value and negative predictive value of FNAC, we set the following definitions:

Positive: Bethesda V and VI cases.

True positive: Positive cases in FNAC are bethesda V and VI, which can be true positive or false positive as mentioned in the text.

False positive: Bethesda V and VI cases with benign diagnosis in the HPE.

Negative: Bethesda II cases.

True negative: Bethesda II cases with benign diagnosis in the HPE.

False negative: Bethesda II cases positive for malignancy in the HPE.

Similarly, we also calculated the risk of malignancy (ROM) in each Bethesda category, except for category I. To calculate the ROM, the following formula was used: ROM = number of malignant cases on the final HPE in each category/total number of cases in that particular group * 100. This was calculated separately for each category, from categories II to VI.

Results

A total of 134 thyroidectomies were performed during the study period. A total of 81 were hemithyroidectomies, and 53 were total thyroidectomies. The female population

Table 1 Original framing of the Bethesda system showing different diagnostic criteria, risk of malignancy, and recommendation

Bethesda category	Definition	ROM (%)	Recommendation
I. Nondiagnostic or unsatisfactory (ND)	Cyst fluid only;virtually acellular specimen; other (obscuring blood, clotting artifact, etc.)		Repeat USG guided FNAC
II. Benign (BN)	Consistent with a benign follicular nodule (includes adenomatoid nodule, colloid nodule, etc.); consistent with lymphocytic (Hashimoto) thyroiditis in the proper clinical context; consistent with granulomatous (subacute) thyroiditis;others	0–3%	Clinical follow-up
III. Atypia of undetermined significance or follicular lesion of undetermined significance (AUS)		5–15%	Repeat FNAC under image guidance
IV. Follicular neoplasm (FN) or Suspicious for a follicular neoplasm (SFN)		15–30%	Surgical lobectomy
V. Suspicious for malignancy (SM)	Suspicious for papillary carcinoma; suspicious for medullary carcinoma; suspicious for metastatic carcinoma; suspicious for lymphoma; others	60–75%	Near-total thyroidectomy or surgical lobectomy
VI. Malignant (MGT)	Papillary thyroid carcinoma;poorly differentiated carcinoma; medullary thyroid carcinoma; undifferentiated (anaplastic) carcinoma; squamous cell carcinoma;carcinoma with mixed features; metastatic carcinoma;non-Hodgkin lymphoma;others	97-99%	Near Total Thyroidectomy

Abbreviations: FNAC, fine needle aspiration cytology, USG, ultrasonography.

Table 2 Final histopathology report in each Bethesda category

	Colloid goiter/ nodular goiter	Hyperplastic nodule	Follicular adenoma	Papillary carcinoma	Follicular carcinoma	Total
Bethesda II	39	2	4	6	—	51
Bethesda III	2	—	1	1	—	4
Bethesda IV	12	—	—	3	5	20
Bethesda V	2	1	4	23	—	30
Bethesda VI	0	—	1	28	—	29
Total	54	4	10	61	5	134

outnumbered the male population by a ratio of 5.3:1. The age of the patients ranged from 11 to 74 years old. The mean age was 51 years old.

The indication for the thyroid surgery in our study was mainly for Bethesda II lesions (38%), followed by Bethesda categories V (22.3%) and VI (21.6%). There were 14.9% cases in category IV. On the final HPE, papillary carcinoma (61/134) was the most common diagnosis, followed by colloid goiter (54/134) (►Tables 2 and 3).

The ROM was calculated for each Bethesda category by correlating it with the final HPE results. The ROM increased for each category, from category II to VI. Category VI had all but 1 malignant lesion on the final report, with a ROM of 96.6%. For category V, the ROM was 76.6%, with 7 benign lesions out of 30 cases (►Table 3).

Out 51 (38%) Bethesda II cases, 45 were benign (true negative), and 6 were malignant (false negative) in the HPE. Similarly, out of 59 (44%) Bethesda V and VI cases collectively, 51 were malignant (true positive), and 8 were benign (false positive) in the HPE. Considering these values above, the specificity was 84.9%, the sensitivity was 89.4%, the positive predictive value (PPV) was 86.4%, the negative predictive value (NPV) was 88.2%, and the accuracy was 87.3% (►Table 4)

Discussion

The TBSRTC is also endorsed by the 8th edition of the ATA guidelines. It has improved the clarity of communication between clinicians. It has helped clinicians and researchers

Table 3 Implied risk of malignancy in different Bethesda categories

Bethesda category	Number of cases (percentage)	Benign in HPE	Malignant in HPE	Risk of alignancy
II	51 (38%)	45	6	11.7%
III	4(2.9%)	3	1	25%
IV	20(14.9%)	12	8	40%
V	30(22.3%)	7	23	76.6%
VI	29(21.6%)	1	28	96%

Abbreviation: HPE, histopathological examination.

Table 4 Sensitivity, specificity, positive predictive value, negative predictive value and accuracy of fine needle aspiration sytology in diagnosing thyroid malignancy, excluding Bethesda III and IV

Positive cases		Negative cases		Specificity	Sensitivity	PPV	NPV	Accuracy
True	False	True	False					
51	8	45	6	84.9%	89.4%	86.4%	88.2%	87.3%

Abbreviations: NPV, negative predictive value; PPV, positive predictive value.

to share clinical data for their own research and collaborations. The present study provides the data of a unique cohort of the Nepalese population and can be an addition to the literature comparing similar data from different geographical regions of the world.

The overall sensitivity, specificity, PPV, NPV, and accuracy in the present study were 89.9%, 84.4%, 86.4%, 88.2% and 87.9%, respectively, which is slightly lower than the data quoted by Zarif et al.[8] Our study showed slightly better sensitivity, but with lower specificity. This variation could be due to the different sample size. The other factors could be the experience of the cytopathologists, as well as the use of USG guidance during the aspiration for cytology. Nevertheless, the accuracy of ∼ 88% in our study is quite acceptable for the clinical practice. We have not considered Bethesda III and IV categories while calculating this data, keeping them in the grey zone. The new recommendation for this category is discussed later in the present article.

Bethesda category I was not included in the study and was not used for ROM calculation, as it is recommended for reevaluation as in others studies.[6,9] The FNAC is reported as

nondiagnostic based on the criteria proposed by Cibas et al.,[6] that is, A) < 6 groups of well-preserved, well-stained follicular cell groups 10 ten cells each; B) poorly prepared, poorly stained, or obscured follicular cells; C) cyst fluid, with or without histocytes, and < 6 groups of 10 benign follicular cells.

We have also compared the ROM of our study with the ROM from the original study by Cibas et al.[6] and from other 3 studies, as shown in ►**Table 5**.

The ROM in different TBSRTC categories obtained in our study is comparable to other published studies on the similar aspect.[6,9–11] The inhomogeneous number of cases in different categories shows some variation in the results in our study.

In our study, the ROM of category II was 11.7%, which is higher when compared with that of the study by Cibas et al.,[6] Mufti et al.,[10] and Jo et al.,[11] but it was lower than that of the study by Her-Juing et al.[9] In our study, the higher number of malignant cases in Bethesda category II could be due to the lack of USG guidance during FNAC, which can miss the suspicious site in the nodule. Ultrasonography guidance was only advised when conventional blind FNAC reporting yielded category I or III.

Table 5 Comparison of the risk of malignancy of the present study with those of other studies

Diagnostic TBSRTC	Current study (n = 134)	Cibas et al[6] 2009	ST Mufti (n = 250)[10] 2012	HerJuing Wu H et al[9] (n = 221)[10] 2011	Jo VY et al (n = 892)[11] 2010
I		–	20%	–	8.9
II	11.7%	0-3%	3.1%	14%	1.1
III	25%	5-15%	50%	44%	17
IV	40%	15-30%	20%	67%	25.4
V	76%	60-75%	80%	77%	70
VI	96%	97-99%	100%	100%	98.1

Abbreviation: TBSRTC, The Bethesda System for Reporting Thyroid Cytopathology.

This discrepancy is also seen in the study done by Zarif et al.,[8] who further analyzed data in this category, in which, initially, 20 out of 128 cases were reported as malignant in the final HPE but, on a detailed analysis, 12 of those 20 cases reported as malignant were later found to have occult papillary carcinoma, which decreased the ROM from 15.6 to 6%. Similarly, we also reviewed all 6 cases in this category, out which 4 were papillary carcinoma measuring between 1 and 1.5 cm, 1 was a follicular variant of papillary carcinoma, and 1 was multifocal papillary carcinoma. These findings support that there was inadequate or inappropriate sampling from the thyroid nodules, leading to false negative reporting on FNAC, rather than an inadequacy of the reporting system itself. This would lead to the argument that thyroid FNAC should be done under image guidance, especially with suspicious features on imaging.

Category III is generally considered in the grey zone. There has been a lot of discussion and changes between 2007 and 2017 regarding its terminology, from atypia of undetermined significance (AUS) to follicular lesion of undetermined significance (FLUS). The ROM in this category is generally higher than that documented by the original TBSRTC, as the minority of the AUS cases will be resected only if there are worrisome clinical or sonographic features. This can be a selection bias, since more suspicious nodules were resected leaving behind the benign nodule for routine follow-up. The malignancy rate varies from 44 to 79% in various studies.[9,12,13] However, it was very low (between 5 and 15%) in a study by Cibas et al. from 2017.[14] Our study presented a ROM of 25%, but we had very few cases from this category that were operated. The varying results in various studies could be due to the consensus that the majority of cases in TBSRTC III category are not operated.

. In the present study, 1 out 4 Bethesda III category cases were reported as papillary carcinoma measuring 3.7 cm in the final HPE. The FNAC sample in this category is very small ($n = 4$) in order to observe a significant difference in the results; therefore, we need larger sample sizes in this particular group to validate our data. It is also worth remembering here that the management of Bethesda III has been changed recently in 2017 after a study by Cibas et al.,[14] which advocates molecular testing rather than performing multiple FNACs in this category. Similarly, noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP) has replaced the noninvasive follicular variant of papillary carcinoma. This change has redistributed the ROM in the original TBSRTC into two category, that is, NIFTP is considered as cancer (ROM = 10–30%), and NIFTP not considered as cancer (ROM = 6–18%) in AUS.[14] The ROM diminishes from between 10 and 30% to between 6 and 18%, which suggests that NIFTP constitutes a substantial proportion of the "malignancies" hidden in this category, which is not considered as a cancer by many of us nowadays.[15] In our setup, we also have not taken NIFTP in any consideration, and ROM was 25% in category III, which lies in the group where NIFTP is considered as cancer (ROM = 10–30%), as show in the study by Cibas et al.[14]

In our study, 14.9% (20/134) of the aspirates were reported as Bethesda IV (FN/SFN); out of these, 8 were reported as malignant lesions in the final HPE, and the ROM was 40%. The reported ROM in this category by the majority of studies is lower, ranging from 15 to 26.1%.[4,6,10,11] Among 8 malignancy cases reported, 5 were follicular carcinoma, and 3 were papillary carcinoma. Similarly, out of 12 benign cases, 7 were follicular adenoma, and 5 were hyperplastic nodules. Suspicious for follicular neoplasm (SFN) is preferred to follicular neoplasm (FN) by many cytopathologists because a significant proportion of cases (up to 35%) proves to not be neoplasms, but rather hyperplastic proliferations of follicular cells. Similar results were also observed in our study; 5 out of 20 cases (25%) were of hyperplastic nodule. For ages, it has been challenging for a cytopathologist to differentiate FN/SFN between hyperplastic nodule or follicular variant of papillary carcinoma (FVPC). Obviously, there is less harm to the patients if FN is reported as hyperplastic nodule or follicular adenoma in the final HPE, because in both scenarios, the surgeon can perform hemithyroidectomy. However, widely invasive FVPC mandates total thyroidectomy. The keys to separate FN from FVPCT are the nuclear features, especially powdery chromatin and oval-shaped nuclei.[16]

There were 22.3 and 21.6% of aspirates in Bethesda V and VI. The ROM was also 76 and 96% in these two categories, respectively. These results obtained in our study were comparable to the results of published results, including that of a meta-analysis.[4,6,10,11] Therefore, there were few cases of false positive reports by the TBSRTC, leading to fewer cases of unintended surgeries. This was very true for category VI, since only 1 of the total 29 cases was benign.

Therefore, we can apply this data in our practice.

Inhomogeneity in the use of USG-guided FNAC was one of the major limitations of the present study. However, as we did not consider Bethesda I and III lesions to determine the sensitivity, specificity, PPV and NPV, bias due to the selective inclusion for USG-guided FNAC can be excluded.

Conclusion

Thyroid FNAC remains a well-accepted screening modality for thyroid nodules in a multidisciplinary setup of surgeons and pathologists, and its results can guide clinical management. The FNAC reporting according to TBSRTC at our hospital has specificity of 84.9%, sensitivity of 89.4%, PPV of 86.4%, NPV of 88.2%, and accuracy of 87.3% in detecting thyroid cancer. The six TBSRTC diagnostic categories are very useful for triaging patients for clinical management in our practice. Although there was discrepancy between our ROM results for Bethesda category III and those from other studies, other four studied categories had a ROM comparable to other studies. Further studies with larger sample sizes and with use of USG guidance for aspiration from thyroid swelling may provide better results by reducing the number of false negative and false positive cases.

References

1 Hegedüs L. Clinical practice. The thyroid nodule. N Engl J Med 2004;351(17):1764–1771

2 Tyler DS, Shaha AR, Udelsman RA, et al. Thyroid cancer: 1999 update. Ann Surg Oncol 2000;7(05):376–398

3 Martin HE, Ellis EB. Biopsy by needle puncture and aspiration. Ann Surg 1930;92(02):169–181

4 Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda system for reporting thyroid cytopathology: a meta-analysis. Acta Cytol 2012;56(04):333–339

5 Kumari KA, Jadhav PD, Prasad C, Smitha NV, Jojo A, Manjula VD. Diagnostic efficacy of ultrasound-guided fine needle aspiration combined with the bethesda system of reporting. J Cytol 2019;36 (02):101–105

6 Cibas ES, Ali SZ. The Bethesda system for reporting thyroid cytopathology. Thyroid 2009;19(11):1159–1165

7 Haugen BR. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: what is new and what has changed? Cancer 2017;123(03):372–381

8 Zarif HA, Ghandurah SE, Al-Garni MA, Binmahfooz SK, Alsaywid BS, Satti MB. Thyroid nodules cytopathology applying the Bethesda system with histopathological correlation. Saudi J Med Med Sci 2018;6(03):143–148

9 Her-Juing HWu, Rose C, Elsheikh TM. The Bethesda system for reporting thyroid cytopathology: an experience of 1,328 cases in a community practice setting with the implication for risk of neoplasm and risk of malignancy. Diagnostic Cytopathology 2012May40(05):399–403

10 Mufti ST, Molah R. The bethesda system for reporting thyroid cytopathology: a five-year retrospective review of one center experience. Int J Health Sci (Qassim) 2012;6(02): 159–173

11 Jo VY, Stelow EB, Dustin SM, Hanley KZ. Malignancy risk for fine-needle aspiration of thyroid lesions according to the Bethesda System for Reporting Thyroid Cytopathology. Am J Clin Pathol 2010;134(03):450–456

12 Park JH, Yoon SO, Son EJ, Kim HM, Nahm JH, Hong S. Incidence and malignancy rates of diagnoses in the bethesda system for reporting thyroid aspiration cytology: an institutional experience. Korean J Pathol 2014;48(02):133–139

13 Lee K, Jung CK, Lee KY, Bae JS, Lim DJ, Jung SL. Application of Bethesda System for Reporting Thyroid Aspiration Cytology. Korean J Pathol 2010;44(05):521–527

14 Cibas ES, Ali SZ. The 2017 Bethesda system for reporting thyroid cytopathology. Thyroid 2017;27(11):1341–1346

15 Strickland KC, Howitt BE, Marqusee E, et al. The impact of noninvasive follicular variant of papillary thyroid carcinoma on rates of malignancy for fine-needle aspiration diagnostic categories. Thyroid 2015;25(09):987–992

16 Renshaw AA. Focal features of papillary carcinoma of the thyroid in fine-needle aspiration material are strongly associated with papillary carcinoma at resection. Am J Clin Pathol 2002;118(02): 208–210