ORIGINAL ARTICLE ASIAN JOURNAL OF MEDICAL SCIENCES

Interobserver variability in diagnosing thyroid fine-needle aspiration cytology using the 2023 Bethesda system for reporting thyroid cytopathology

Pratima Manohar Pattar[1], Sameera N Patel[2], Girija[3]

[1,3]Assistant Professor, Department of Pathology, S Nijalingappa Medical College, Bagalkot, [2]Assistant Professor, Department of Pathology, ESIC Medical College and Hospital, Kalaburagi, Karnataka, India

Submission: 21-10-2024 Revision: 24-11-2024 Publication: 01-01-2025

ABSTRACT

Background: Thyroid fine-needle aspiration (FNA) surfaced as a vital first-line diagnostic tool to categorize them into benign or malignant, thereby reducing unnecessary surgeries in benign nodules. The Bethesda system for reporting thyroid cytopathology (TBSRTC) has established a standardized, category-based reporting system and since then it has been widely adopted by cytopathologists in worldwide. It has been shown that thyroid pathology contributes significantly to second-opinion discrepancies for cytopathology specimens. **Aims and Objectives:** To establish the reproducibility of TBSRTC in diagnosing thyroid lesions. To determine the interobserver variation in thyroid cytopathology using the 2023 TBSRTC. **Materials and Methods:** This was a 2-year retrospective study conducted in the Department of Pathology, from January 2018 to December 2019. Previous archived slides of thyroid FNA cytology were examined by three trained pathologists in a double-blinded fashion and were reclassified according to the 2023 TBSRTC. The results were evaluated for interobserver variability using Cohen's kappa statistics and Statistical Package for the Social Sciences software version 24.0. **Results:** The present study included a total of 156 patients with a female: male ratio of 3: 0.1 and the mean age of 37.8 ± 2.4 years. The interobserver agreement, calculated using Kappa statistics, for observer 1 and observer 2 was 0.735; for observer 1 and 3 was 0.841; and for observer 2 and 3 was 0.838. These values together showed substantial interobserver agreement. **Conclusion:** The 2023 TBSRTC gives a distinctive, universal, and homogeneous terminology for reporting the thyroid cytology and its execution should be promoted due to its virtual ease of reproducibility.

Key words: Atypia of undetermined significance; Bethesda system; Cohen's kappa; Interobserver variability; Thyroid cytology

Access this article online

Website: https://ajmsjournal.info/index.php/AJMS/index

DOI: 10.71152/ajms.v16i1.4327

E-ISSN: 2091-0576

P-ISSN: 2467-9100

INTRODUCTION

Thyroid diseases are often encountered in India and approximately 42 million people in India suffer from thyroid diseases.[1] The prevalence of a palpable thyroid nodule in India is approximately 12.2%; however, the incidence of thyroid cancer is 8.7 cases/100,000 population/year.[2] As a palpable thyroid nodule may not always be malignant, it is essential for the clinician to differentiate between benign and malignant nodules for appropriate management. Subsequently, thyroid fine-needle aspiration (FNA) surfaced as a vital first-line diagnostic tool to categorize palpable thyroid nodules and thereby reducing unnecessary surgeries in benign thyroid nodules.[3]

In 2007, the Bethesda system for reporting thyroid cytology

Address for Correspondence:
Dr. Pratima Manohar Pattar, Assistant Professor, Department of Pathology, S Nijalingappa Medical College, Bagalkot - 587 103, Karnataka, India. **Mobile:** +91-9742084001. **E-mail:** pratimapattar@gmail.com

(TBSRTC) was proposed at the National Cancer Institute Thyroid FNA State of Science Conference at Bethesda, Maryland. This reporting system includes suggestions regarding the layout of the report, adequacy of the sample, diagnostic category, risk of malignancy (ROM), and proposed clinical management.[4,5] The thyroid Bethesda Reporting System, first published in 2010, established a standardized, category based reporting system. Since its first publication, the reporting system has been widely adopted by cytopathologists in worldwide. The second edition of the Bethesda Thyroid Reporting System was then published in 2017 to adjust the ROM of each diagnostic category and to add explanatory notes. Consequently, the third edition of the Bethesda Thyroid Reporting System published in June 2023 focuses on simplifying diagnostic categories with single names for each category, adopting the new histologic terminologies according to the 2022 World Health Organization Classification on Thyroid Neoplasms.[6]

As per the 2023 TBSRTC, minimum six groups of at least 10 cells which are undistorted and well-maintained on a single slide is considered as adequate for FNA reporting. It is divided into six categories: Non-diagnostic "(ND)" (category-I), Benign "B" (category-II); Atypia of undetermined significance "(AUS)" (category-III), follicular neoplasm "(FN)" (category-IV); suspicious for malignancy "(SFM)" (category-V); and Malignant "M" (category-VI).

Recently, it has become clear that even the "gold standard" for diagnosing follicular-patterned thyroid lesions, histopathology, shows considerable interobserver variability.[7] Furthermore, it has been shown that thyroid pathology contributes significantly to second-opinion discrepancies for histopathology and cytopathology specimens.[8] Due to the "gray zone" that exists in the interpretation of cytologic samples of thyroid lesions composed predominantly of colloid and follicular groups, we expected considerable interobserver variability might occur for the interpretation of these cases.[9]

The main purpose of TBSRTC is to provide standardization in the nomenclature of thyroid nodules and recommended surgical and clinical management for such. There has been a growing concern that there is variability with the reporting of TBSRTC at different institutions.[10] The purpose of this study is aimed at assessing the diagnostic utility of the 2023 TBSRTC at our institution with consideration of interobserver variation.

Aims and objectives

To establish the reproducibility of TBSRTC in diagnosing thyroid lesions and to determine the interobserver variation in thyroid cytopathology by using the 2023 TBSRTC.

MATERIALS AND METHODS

This was a 2-year retrospective study conducted in the Department of Pathology, ESIC Medical College and Hospital, Kalaburagi, from January 2018 to December 2019. After taking Institutional Ethics Committee permission (No: ESICMC/GLB/IEC/16/2021), medical records data pertaining to demographic details, clinical history, and radiological details were retrieved. Previous archived slides of thyroid FNA cytology (FNAC) ranging from a spectrum of benign to malignant lesions, belonging to different categories which can be classified according to the 2023 Bethesda system were examined by three trained pathologists who have experience of >3 years in cytopathology in a double-blinded fashion. Previous cytology reports were not disclosed to the examining pathologists. Hemorrhagic aspirate samples were excluded from the study.

Interobserver variability can be due to variation in FNA sample quality, preparation, and staining. As we have taken all the samples from single institute, these variations were assumed to be in accountable or minimal. The other factors influencing variability can arise from differences in experience or interpretation criteria among the cytopathologists. Therefore, the 2023 TBSRTC has a role in standardizing the diagnostic criteria and also in reducing the interobserver variability.

After examination, the FNAC smears were reclassified according to the six diagnostic categories of 2023 TBSRTC. The results of all three pathologists were evaluated for interobserver variability by calculating the percentage of agreement. All the data were entered in Microsoft Excel and the interobserver variability was estimated using Cohen's kappa statistics and Statistical Package for the Social Sciences software version 24.0.

RESULTS

A total of 156 patients were included in the present study, out of which 151 (96.8%) were female and 05 (3.2%) were male, with a female: male ratio of 3:0.1. The age of the patients ranged from 21 to 60 years, with a mean of 37.8±2.4 years. Majority of the cases (24.3%) belonged to the age group of 31–40 years.

The cases were re-categorized (Table 1) as 6–7% of cases in category I (ND), 56–61% of cases in category II (Benign), 14–18% of cases in category III (AUS), 9–10% of cases in category IV (FN), 3–4% of cases in category V (SFM), and 2–3% of cases in category VI (M). The highest number of cases were in the category II (Benign), followed by category III (AUS) and then category IV (FN). The category III

(AUS) showed more disparity in the diagnosis among three cytopathologists than the category II (Benign). The categories I, IV, V, and VI also showed minimum disparity in the diagnosis.

The interobserver agreement, calculated using Cohen's Kappa statistics, for observer 1 and observer 2 was 0.735 (Table 2); for observer 1 and 3, 0.841 (Table 3); and for observer 2 and 3, 0.838 (Table 4). These values together showed substantial interobserver agreement.

Table 1: Observers diagnoses of thyroid cytology cases according to the 2023 Bethesda system for reporting thyroid cytopathology

Diagnostic categories	Observer 1	Observer 2	Observer 3
ND	11	11	10
Benign	96	90	88
AUS	23	28	29
FN	15	16	17
Suspicious malignancy	07	06	07
Malignant	04	05	05
Total	156	156	156

DISCUSSION

Thyroid FNA has been in use for many years and is now the mainstay of preoperative diagnosis of thyroid lesions.[11] The thyroid gland is the organ most commonly sampled by FNA as it can make an apparent difference in patient management. FNA is rapid, cost-effective, and minimally invasive technique which reduces the need of unnecessary surgery.[12]

Previously, pathologists have been using variable terminologies for cytological reporting of thyroid lesions which were all ambiguous and inconsistent. Different reporting criteria were used in different laboratories. The results did not show proper clinical relevance and created confusion among pathologists, endocrinologists, surgeons, and radiologists.[13,14] This led to the introduction of TBSRTC in October 2007. The TBSRTC system is a universal standardized reporting system which helps cytologists and physicians in understanding and foreseeing the scenario by estimating the malignant potential of individual categories.[15]

Table 2: Cross tabulation between observer 1 and observer 2

Observer 1	Observer 2					
	ND	Benign	AUS	FN	Suspicious	Malignant
ND	11	00	00	00	00	00
Benign	00	90	05	01	00	00
AUS	00	00	23	00	00	00
FN	00	00	00	15	00	00
Suspicious	00	00	00	00	06	01
Malignant	00	00	00	00	0	04

ND: Non-diagnostic, AUS: Atypia of undetermined significance, FN: Follicular neoplasm

Table 3: Cross tabulation between observer 2 and observer 3

Observer 2	Observer 3					
	ND	Benign	AUS	FN	Suspicious	Malignant
ND	10	01	00	00	00	00
Benign	00	87	01	01	01	00
AUS	00	00	28	00	00	00
FN	00	00	00	16	00	00
Suspicious	00	00	00	00	06	00
Malignant	00	00	00	00	00	05

ND: Non-diagnostic, AUS: Atypia of undetermined significance, FN: Follicular neoplasm

Table 4: Cross tabulation between observer 3 and observer 1

Observer 3	Observer 1					
	ND	Benign	AUS	FN	Suspicious	Malignant
ND	10	00	00	00	00	00
Benign	00	88	00	00	00	00
AUS	00	06	23	00	00	00
FN	00	02	00	15	00	00
Suspicious	00	00	00	00	07	00
Malignant	01	00	00	00	00	04

ND: Non-diagnostic, AUS: Atypia of undetermined significance, FN: Follicular neoplasm

In our study, we observed that the interobserver variation in category I of 2023 TBSRTC was very less. Most of the observers agreed on the ND category due to acellularity or obscuring blood elements. The category II has mentioned the clear criterias for benign lesions of thyroid. Therefore, the interobserver variation was mild in this category also. Category III of AUS showed more interobserver variation. In spite of the criteria defined for this category, some observers interpreted these lesions as FN or adenomatoid nodule.

TBSRTC defines the diagnostic criteria for AUS, which is mainly reserved for aspirates with architectural and/or cytologic atypia that is not sufficient to be categorized as suspicious for neoplasia or malignancy, yet on the other hand, it is more marked than can be ascribed confidently to benign changes. Therefore, the category III (AUS) is the most heterogeneous and challenging category. It is not uncommon that thyroid nodules categorized as AUS are subsequently proven to be benign on histologic examination.[16-18]

Category IV of FN poses problem for the observers if the cellularity is less; therefore, moderate variation is observed. Category V, of SFM also showed moderate variation. When the features are not prominent to label the lesion as malignant one; the observers showed variation in the diagnosis. While in category VI, that is, in malignant the variation was very less, as the features were clear to label the lesion as malignant and if it was not clear then they were categorized into SFM.

According to 2023, Bethesda system good interobserver correlation was found in ND, benign, and malignant categories among all the three observers. Category III, AUS has borderline findings which makes difficult to categorize them in malignant category and carries 5–10% ROM. The foremost interobserver disagreement in the present study was found to be in the cases with the presence of abundant colloid in the background which prompted for benign category diagnosis. Some cases in the benign category were categorized as AUS by one of the observer as it showed focal follicular atypia. These focal changes were attributed to repair by the other observer to make it a benign category diagnosis.

In the present study, a concurrence of diagnosis was observed in 92.5% of the cases, and substantial agreement was found by Cohen's kappa (0.841). This was in corroboration with findings of other studies in the literature.[12,15] In the study by Ahmed et al.,[12] interobserver variability was calculated by kappa value of 0.735, Awasthi et al.,[15] found the kappa value of 0.613, and Stelow et al.,[19] found it to be 0.65.

However, along with its many advantages, a few pitfalls were noted in the present study. The AUS category appears to be a heterogenous category, and the categorization of a lesion in this category is very subjective.[19,20]

Limitations of the study

The limitation of the present study was a lack of histopathological correlation for the AUS and FN cases. Therefore, similar studies with large sample size and histopathological correlation can be attempted to refine the categories in the TBSRTC and the efficacy of AUS as a separate category.

CONCLUSION

TBSRTC is a distinctive, universal, and homogeneous terminology for reporting the thyroid cytology and its execution should be promoted due to its virtual ease of reproducibility. The practicing cytopathologists as well as clinicians should be adequately trained for the diagnostic criteria, terminology, and guidelines of TBSRTC for a more efficient usage of this system.

Substantial interobserver agreement was found for thyroid cytological lesions using Bethesda Reporting Criteria. It provides fine distinction between benign and malignant cytological lesions. However, the AUS category is heterogeneous and its efficacy as separate cytologic group is still controversial.

ACKNOWLEDGMENT

The authors take this opportunity to thank Medical Records Department and all the technical staff, Department of Pathology ESIC Medical College for their whole hearted support for this study.

REFERENCES

1. Unnikrishnan AG and Menon UV. Thyroid disorders in India: An epidemiological perspective. Indian J Endocrinol Metab. 2011;15(Suppl 2):S78-S81.
   https://doi.org/10.4103/2230-8210.83329
2. Unnikrishnan AG, Kalra S, Baruah M, Nair G, Nair V, Bantwal G, et al. Endocrine society of India management guidelines for patients with thyroid nodules: A position statement. Indian J Endocrinol Metab. 2011;15(1):2-8.
   https://doi.org/10.4103/2230-8210.77566
3. Somma J, Schlecht NF, Fink D, Khader SN, Smith RV and Cajigas A. Thyroid fine needle aspiration cytology: Follicular lesions and the gray zone. Acta Cytol. 2010;54(2):123-131.
   https://doi.org/10.1159/000324996
4. Bhasin TS, Mannan R, Manjari M, Mehra M, Sekhon AK, Chandey M, et al. Reproducibility of 'the bethesda system for reporting thyroid cytopathology': A multicenter study with review

of the literature. J Clin Diagn Res. 2013;7:1051-1054.
https://doi.org/10.7860/jcdr/2013/5754.3087

5. Crowe A, Linder A, Hameed O, Salih C, Roberson J, Gidley J, et al. The impact of implementation of the Bethesda System for reporting thyroid cytopathology on the quality of reporting, "Risk" of malignancy, surgical rate, and rate of frozen sections requested for thyroid lesions. Cancer Cytopathol. 2011;119(5):315-321.
https://doi.org/10.1002/cncy.20174
6. Han M and Fan F. Bethesda system for reporting thyroid cytopathology-an updated review. J Clin Transl Pathol. 2023;3(2):84-98.
https://doi.org/10.14218/jctp.2023.00005
7. Franc B, De la Salmoniere P, Lange F, Hoang C, Louvel A, De Roquancourt A, et al. Interobserver and intraobserver reproducibility in the histopathology of follicular thyroid carcinoma. Hum Pathol. 2003;34(11):1092-1100.
https://doi.org/10.1016/s0046-8177(03)00403-9
8. Hamady ZZ, Mather N, Lansdown MR, Davidson L and Maclennan KA. Surgical pathological second opinion in thyroid malignancy: Impact on patients' management and prognosis. Eur J Surg Oncol. 2005;31(1):74-77.
https://doi.org/10.1016/j.ejso.2004.08.010
9. Baloch ZW, Fleisher S, Li Volsi VA and Gupta PK. Diagnosis of "follicular neoplasm": A gray zone in thyroid fine-needle aspiration cytology. Diagn Cytopathol. 2002;26(1):41-44.
https://doi.org/10.1002/dc.10043
10. Jin X and Jing X. Cytologic assessment of thyroid nodules-updates in 2023 Bethesda reporting system, diagnostic challenges and pitfalls. Hum Pathol Rep. 2024;36(1):300743.
https://doi.org/10.1016/j.hpr.2024.300743
11. Cramer H. Fine-needle aspiration cytology of the thyroid: An appraisal. Cancer. 2000;90(6):325-329.
https://doi.org/10.1002/1097-0142(20001225)90:6<325:aid-cncr1>3.0.co;2-u
12. Ahmed S, Ahmad M, Khan MA, Kazi F, Noreen F, Nawaz S, et al. The interobserver reproducibility of thyroid cytopathology using Bethesda reporting system: Analysis of 200 cases. J Pak Med Assoc. 2013;63(10):1252-1255.
13. Cibas ES. Fine-needle aspiration in the work-up of thyroid nodules. Otolaryngol Clin North Am. 2010;43(2):257-271, vii-viii.
https://doi.org/10.1016/j.otc.2010.01.003
14. Park JH, Kim HK, Kang SW, Jeong JJ, Nam KH, Chung WY, et al. Second opinion in thyroid fine-needle aspiration biopsy by the Bethesda System. Endocr J. 2012;59(3):205-212.
https://doi.org/10.1507/endocrj.ej11-0274
15. Awasthi P, Goel G, Khurana U, Joshi D, Majumdar K and Kapoor N. Reproducibility of "the Bethesda system for reporting thyroid cytopathology:" A retrospective analysis of 107 Patients. J Cytol. 2018;35(1):33-36.
https://doi.org/10.4103/joc.joc_215_16
16. Cibas ES and Ali SZ. The 2017 Bethesda system for reporting thyroid cytopathology. Thyroid. 2017;27(11):1341-1346.
https://doi.org/10.1089/thy.2017.0500
17. Lee B, Smola B, Roh MH, Hughes DT, Miller BS and Jing X. The impact of using the Bethesda system for reporting thyroid cytology diagnostic criteria on the follicular lesion of undetermined significance category. J Am Soc Cytopathol. 2014;3(3):131-136.
https://doi.org/10.1016/j.jasc.2014.01.009
18. Kantola S, Virani N, Haus C, Hipp J, Zhao L and Jing X. Prospective evaluation of impact of using the Bethesda system for reporting thyroid cytopathology: An institutional experience. J Am Soc Cytopathol. 2015;4(1):5-29.
https://doi.org/10.1016/j.jasc.2014.08.003
19. Stelow EB, Bardales RH, Crary GS, Gulbahce HE, Stanley MW, Savik K, et al. Interobserver variability in thyroid fine-needle aspiration interpretation of lesions showing predominantly colloid and follicular groups. Am J Clin Pathol. 2005;124(2):239-244.
https://doi.org/10.1309/p5mfma7umful2kwt
20. Bansal I, Sale S, Mane V, Swami L and Bhatt N. Interobserver and intraobserver variation in thyroid cytopathology by conventional method and the Bethesda system. Indian J Pathol Oncol. 2020;7(3):425-427.
https://doi.org/10.18231/j.ijpo.2020.084

Authors Contribution:
PMP- Concept, design, study protocol, data collection, data analysis, manuscript preparation and submission of article; **SNP-** Design of study, literature survey, statistical analysis and interpretation, review manuscript; **G-** Preparation of tables, coordination and manuscript revision.

Work attributed to:
ESIC Medical College and Hospital, Kalaburagi, Karnataka, India.

Orcid ID:
Pratima Manohar Pattar - https://orcid.org/0000-0002-4183-3311
Sameera N Patel - https://orcid.org/0009-0002-7096-6049
Girija - https://orcid.org/0000-0001-8987-0317

Source of Support: Nil, **Conflicts of Interest:** None declared.