



Original Research Article

The interobserver variability of thyroid fine needle aspirates using The Bethesda System for Reporting Thyroid Cytopathology

Chhavi Gupta^{1,*}, Subhash Bhardwaj¹, Sindhu Sharma¹

¹Dept. of Pathology, Government Medical College, Jammu, Jammu and Kashmir, India



ARTICLE INFO

Article history:

Received 21-10-2019

Accepted 23-10-2019

Available online 22-02-2020

Keywords:

Bethesda system

Thyroid fine needle aspiration

cytology

Interobserver variability

ABSTRACT

Introduction: The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) is a six category scheme for reporting thyroid Fine Needle Aspirates (FNAs) and represents a major step towards a uniform reporting system for thyroid FNAs, enabling better understanding of thyroid cytopathology report by referring clinicians and providing rational management guidelines. The present study was carried out to analyse cytological features of thyroid FNAs and categorise them by The Bethesda System for Reporting Thyroid Cytopathology; and to assess interobserver variability between two independent reporting pathologists using the Bethesda system.

Materials and Methods: The study was conducted over three years from Nov. 2015 to Oct. 2018 in the Department of Pathology, Government Medical College, Jammu. The thyroid FNAs were categorized according to the Bethesda system by two independent pathologists in a double blinded fashion. Interobserver variability was assessed by calculating percentage of agreement, disagreement and statistically using Cohen's kappa.

Results: A total of 610 cases were categorized as Nondiagnostic-51 cases (8.36%), 50 cases (8.20%); Benign-524 cases (85.90%), 510 cases (83.61%); Atypia of Undetermined Significance-2 cases (0.33%), 12 cases (1.97%); Follicular Neoplasm-15 cases (2.46%), 16 cases (2.62%); Suspicious for Malignancy-1 case (0.16%), 6 cases (0.98%); Malignant-17 cases (2.79%), 16 cases (2.62%) by two pathologists respectively. The two pathologists agreed in 548 cases (89.84%) and disagreed in 62 cases (10.16%) with Cohen's kappa=0.628 (62.8%). Agreement was highest in Benign category-95.68%, followed by Malignant-87.5%, Follicular Neoplasm -75%, Nondiagnostic-68%. No agreement was seen in Atypia of Undetermined Significance and Suspicious for Malignancy categories.

Conclusion: The present study thus encourages the use of The Bethesda System for Reporting Thyroid Cytopathology as a standardized reporting system with relative ease of reproducibility for effective communication among pathologists and clinicians with regard to thyroid Fine Needle Aspiration Cytology reporting and management.

© 2020 Published by Innovative Publication. This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by/4.0/>)

1. Introduction

Thyroid diseases are among the commonest endocrine disorders worldwide. About 42 million people in India suffer from thyroid diseases.¹ Fine needle aspiration cytology (FNAC) is the first-line diagnostic procedure for evaluating thyroid lesions. It is a simple, rapid, cost-effective test that provides valuable information about the

nature of a thyroid nodule and can effectively distinguish between neoplastic and non-neoplastic lesions of the thyroid therefore guiding appropriate management i.e. follow-up or surgery thus reducing unnecessary surgery for patients with benign disease.²⁻⁴

Reporting of thyroid FNA specimens should follow a standard format that is clinically relevant in order to direct appropriate management.⁵ The terminology for reporting thyroid FNACs has varied markedly worldwide with use of many reporting formats varying from two category schemes

* Corresponding author.

E-mail address: guptachhavi919@gmail.com (C. Gupta).

to six or more category schemes with some relying on descriptive phrases instead of categories.^{3,4} This lack of uniformity creates confusion amongst referring clinicians in the interpretation of thyroid cytopathology report, thereby affecting definitive clinical management.²

To address terminology and other issues related to thyroid FNA, the National Cancer Institute (NCI) hosted the "NCI Thyroid Fine Needle Aspiration State of the Science Conference", a multidisciplinary conference that took place in Bethesda, Maryland, United States in October 2007, that led to the formation of The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC), a six-category scheme for reporting thyroid cytopathology that recommends each report should begin with a general diagnostic category. Each of the categories has an implied risk of malignancy that links it to a rational clinical management guideline.⁶

Recently published data support the clinical utility and wide acceptance of TBSRTC by both practicing pathologists and clinicians. In Indian perspective too, many studies have been carried out that have concluded TBSRTC to be an effective reporting system for thyroid FNA. It improves perceptions of diagnostic terminology between cytopathologists and clinicians and also provides clear management guidelines to the clinicians.^{2,4} In our institution, reporting of thyroid lesions on FNA is variable and no standardized system of reporting is being followed. The present study was thus carried out to analyse the cytological features of thyroid fine needle aspiration smears and categorise them by The Bethesda System for Reporting Thyroid Cytopathology; and to assess interobserver variability between two independent reporting pathologists using The Bethesda System for Reporting Thyroid Cytopathology.

2. Materials and Methods

The present study was conducted in the Department of Pathology, Government Medical College, Jammu over a period of three years from Nov.2015 to Oct.2018. It included all patients presenting with thyroid swelling referred to this department from various clinical departments of this hospital and from other health care centers. Non-cooperative and morbid patients were excluded from the study.

Smears were prepared from the sample obtained by aspiration or non-aspiration method and were stained with May-Grunwald-Giemsa (MGG) and Papanicolau (PAP) stains. Stained smears were examined under light microscopy by two independent pathologists in a double blinded fashion. The cytological features were evaluated and reporting was done according to The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC).

The definitions and cytomorphological criteria as described in the, The Bethesda System for Reporting Thyroid Cytopathology atlas were followed by the two

reporting pathologists (Pathologist I and Pathologist II). After assessing the adequacy as per TBSRTC adequacy criteria, the thyroid fine needle aspirates were categorized into the six categories of The Bethesda System for Reporting Thyroid Cytopathology.⁷

2.1. Nondiagnostic/Unsatisfactory (ND/UNS)

Nondiagnostic is used to convey a sample that does not meet the adequacy criteria as given in the TBSRTC monograph. A thyroid FNA sample is considered adequate for evaluation if it contains a minimum of six groups of well-visualised follicular cells, with atleast ten cells per group, preferably on a single slide. Exceptions to this requirement apply to solid nodules with cytologic atypia, solid nodules with inflammation and colloid nodules where minimum number of follicular cells is not required for adequacy.

Nondiagnostic category included aspirates with fewer than six groups of well-preserved, well-stained follicular cell groups with ten cells each (excluding the exceptional circumstances); aspirates with poorly prepared, poorly stained, or significantly obscured follicular cells; cyst fluid, with or without histiocytes, and fewer than six groups of ten benign follicular cells.

2.2. Benign

The aspiration smears were categorized as benign if they showed cytomorphological features of a benign follicular nodule, lymphocytic thyroiditis, granulomatous thyroiditis or other entities like acute thyroiditis as per TBSRTC.

2.3. Atypia of Undetermined Significance/Follicular Lesion of Undetermined Significance (AUS/FLUS)

This diagnostic category was reserved for aspirates that contained cells with architectural and/or nuclear atypia that was not sufficient to be classified as suspicious for a follicular neoplasm, suspicious for malignancy, or malignant. On the other hand, the atypia was more marked than could be ascribed confidently to benign changes.

2.4. Follicular Neoplasm/ Suspicious for Follicular Neoplasm (FN/SFN)

This category included smears that were moderately or markedly cellular with significant alteration in follicular cell architecture, characterized by cell crowding, microfollicles, and dispersed isolated cells. Colloid was scant to absent.

Aspirates in this category were included as Follicular Neoplasm, Hurthle cell (Oncocytic) type if the aspirate consisted exclusively (or almost exclusively) of Hurthle cells having abundant finely granular cytoplasm, enlarged central or eccentrically located round nucleus, prominent nucleolus. Specimens were moderately to markedly cellular with usually little or no colloid and virtually no lymphocytes

(excluding blood elements) or plasma cells.

2.5. Suspicious for Malignancy (SFM)

Cases showing cytomorphological features that raised a strong suspicion of malignancy (papillary thyroid carcinoma, medullary thyroid carcinoma, lymphoma or metastatic carcinoma) but the findings were not sufficient for a conclusive diagnosis were included in this category.

2.6. Malignant

This Bethesda category was applied whenever the cytomorphologic features were conclusive for malignancy. The malignancies included in this category were Papillary thyroid carcinoma, Medullary thyroid carcinoma, Anaplastic carcinoma, Poorly differentiated carcinoma, and lymphoma.

The results of both the pathologists were evaluated for interobserver variability by calculating the percentage of agreement, disagreement; and interobserver variability was statistically assessed using Cohen's kappa as a measure of concordance between the two observers beyond chance.

3. Results

A total of 610 patients were included in the study. The thyroid fine needle aspirates were categorized into the six categories of The Bethesda System for Reporting Thyroid Cytopathology by two independent pathologists.

Pathologist I reported 51 cases (8.36%) as Nondiagnostic, 524 cases (85.90%) as Benign, 2 cases (0.33%) as Atypia of Undetermined Significance, 15 cases (2.46%) as Follicular Neoplasm, 1 case (0.16%) as Suspicious for Malignancy, and 17 cases (2.79%) as Malignant while Pathologist II reported 50 cases (8.20%) as Nondiagnostic, 510 cases (83.61%) as Benign, 12 cases (1.97%) as Atypia of Undetermined Significance, 16 cases (2.62%) as Follicular Neoplasm, 6 cases (0.98%) as Suspicious for Malignancy, and 16 cases (2.62%) as Malignant (Table 1) (Figures 1, 2, 3, 4, 5 and 6).

Benign category was the largest category followed by Nondiagnostic category. Benign follicular nodule was the predominant subcategory followed by chronic lymphocytic thyroiditis as reported by both pathologists. Papillary Thyroid Carcinoma was the most common malignancy reported by both the pathologists in our study (Table 1).

The two pathologists were in agreement in 548 cases (89.84%) and disagreed in 62 cases (10.16%). Diagnostic agreement was highest in the Benign category (95.68%), followed by Malignant category (87.5%), Follicular Neoplasm (75%), Nondiagnostic (68%). No agreement was seen in AUS category and Suspicious for Malignancy category (0%, 0% respectively) between the two reporting pathologists (Table 2).

Discordant cases of Nondiagnostic category by one pathologist were categorized in Benign category by other

pathologist. Discordant cases in Benign category were put into Non diagnostic, AUS, Follicular neoplasm and Suspicious for Malignancy categories by other pathologist. Of the two cases in AUS category by one pathologist, there was disagreement in both the cases with one case reported in Benign category and other case in Suspicious for Malignancy category by other pathologist. Of the 15 cases in Follicular neoplasm category by one pathologist, there was disagreement in 3 cases with 2 cases reported in Benign category and 1 case in Malignant category by other pathologist. Of one case in Suspicious for Malignancy category by one pathologist, there was complete disagreement with the case being reported in Malignant category by other pathologist. Of the 17 cases in Malignant category by one pathologist, there were 3 discordant cases, 2 cases in Benign category and 1 case in Suspicious for Malignancy category by other pathologist (Table 2).

The interobserver variability was statistically assessed using Cohen's kappa coefficient with observed value of 0.628; 62.8% (95% Confidence Interval 54.0-71.6) i.e. actual agreement between the two pathologists beyond chance was 62.8% (moderate) in our study.

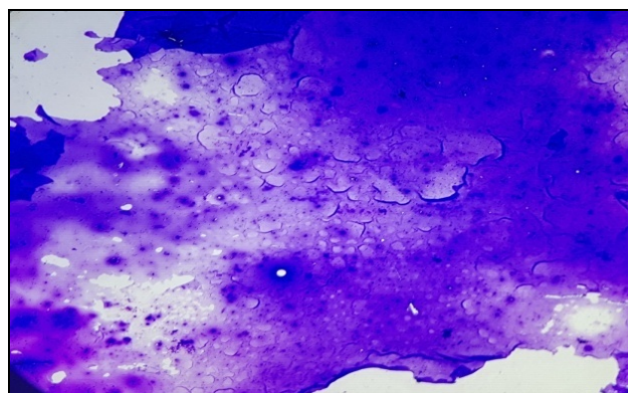


Fig. 1: Photomicrograph from a case of Colloid Nodule showing abundant colloid with few follicular cells (MGG 40 X).

4. Discussion

TBSRTC has been widely adopted in the United States and in many places worldwide and has been endorsed by the American Thyroid Association.^{7,8} The distribution of cases into various TBSRTC categories by the two pathologists in our study were compared with other studies as shown in Table 3.

The frequency of Nondiagnostic interpretations varies notably from laboratory to laboratory (range, 3-34%).⁷ The findings of our study (8.36%, 8.20%) are consistent with this range and are comparable with other studies.^{4,9} As most thyroid nodules are benign, a benign result is the most common FNA interpretation (approximately 60-70% of all

Table 1: Distribution of cases into various categories and subcategories as per TBSRTC by Pathologists I and II

Bethesda Category	Sub-categories	Pathologist I Number of cases	Percentage (%)	Pathologist II Number of cases	Percentage (%)
I. Non diagnostic		51	8.36	50	8.20
	Cyst fluid only	0	0.00	2	0.33
	Virtually acellular specimen	42	6.89	42	6.89
	Others (obscuring blood, clotting artifact, etc.)	9	1.47	6	0.98
II. Benign		524	85.90	510	83.61
	Consistent with a benign follicular nodule (colloid nodule, adenomatoid nodule)	351	57.54	377	61.80
	Consistent with chronic lymphocytic (Hashimoto) thyroiditis	162	26.56	118	19.34
	Consistent with granulomatous (subacute) thyroiditis	3	0.49	6	0.98
	Other	8	1.31	9	1.48
III. Atypia of Undetermined Significance		2	0.33	12	1.97
IV. Follicular Neoplasm		15	2.46	16	2.62
	Follicular Neoplasm	12	1.97	11	1.80
	Follicular Neoplasm, Hurthle Cell (Oncocytic) type	3	0.49	5	0.82
V. Suspicious for Malignancy		1	0.16	6	0.98
	Suspicious for Papillary Thyroid Carcinoma	1	0.16	6	0.98
VI. Malignant		17	2.79	16	2.62
	Papillary Thyroid Carcinoma	8	1.31	8	1.31
	Poorly differentiated carcinoma	1	0.16	2	0.33
	Medullary thyroid carcinoma	3	0.49	3	0.49
	Undifferentiated (Anaplastic) carcinoma	4	0.66	3	0.49
	Non-Hodgkin Lymphoma	1	0.16	0	0.00
Total		610	100.00	610	100.00

cases).⁷ The cases in benign category in our study are higher than this range (85.90%, 83.61%) and also from study by Jo et al.¹⁰ but are comparable with other studies.^{2,4,9,11} Being the only tertiary care center of our province, it caters to a large number of patients on both direct and referral basis, so a large population representative of general population is encountered that could be a reason for higher number of cases in benign category.

AUS/FLUS usage varies from 1% to 22% of thyroid FNAs with practical upper limit of 10%.⁷ AUS cases

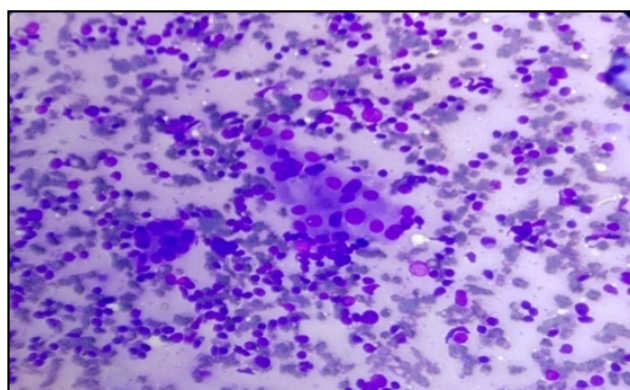
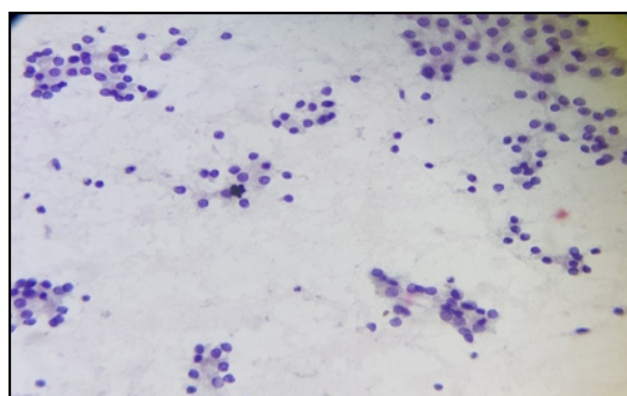
and Follicular Neoplasm cases reported in our study are comparable with studies in Table 3. Suspicious for Malignancy (SFM) diagnoses account for approximately 3% (range 1.0-6.3%) of all thyroid FNAs. As with any indeterminate diagnosis, this category should be used judiciously so that patients are managed as appropriately as possible.⁷ The SFM cases in our study as reported by Pathologist I are lesser than the lower limit of the given range (0.16%) but are comparable with study by Laishram et al.,¹¹ while are 0.98% i.e. ~1% as reported by Pathologist

Table 2: Comparison of cases reported according to TBSRTC by two pathologists as agreement versus disagreement

Bethesda Category		Pathologist I Non Diagnostic	Benign	Atypia of Undetermined Significance	Follicular Neoplasm	Suspicious for Malignancy	Malignant	Total
Pathologist II	I. Non diagnostic	34	16	0	0	0	0	50
	II. Benign	17	488	1	2	0	2	510
	III. Atypia Of Undetermined Significance	0	12	0	0	0	0	12
	IV.Follicular Neoplasm	0	4	0	12	0	0	16
	V. Suspicious for Malignancy	0	4	1	0	0	1	6
	VI. Malignant	0	0	0	1	1	14	16
	Total	51	524	2	15	1	17	610

Table 3: Comparison of percentages of distribution of cases of present study with other studies

Bethesda category	Mondal et al. ²	Mehra et al. ⁴	Bhat et al. ⁹	Jo et al. ¹⁰	Laishram et al. ¹¹	Present Study Pathologist I	Pathologist II
I Non diagnostic	1.2	7.2	6.6	18.6	5.2	8.36	8.20
II Benign	87.5	80	82	59	89.9	85.90	83.61
III AUS	1	4.9	2	3.4	0	0.33	1.97
IV Follicular Neoplasm	4.2	2.2	2.5	9.7	2.2	2.46	2.62
V Suspicious for malignancy	1.4	3.6	1.6	2.3	0.3	0.16	0.98
VI Malignant	4.7	2.2	5.1	7.0	2.2	2.79	2.62

**Fig. 2:** Photomicrograph from a case of Lymphocytic Thyroiditis—showing sheet of follicular cells infiltrated by mature lymphocytes along with numerous lymphoid cells in the background (MGG 400 X)**Fig. 3:** Photomicrograph from a case of AUS showing few clusters as well as singly scattered follicular cells with mild anisonucleosis and focal microfollicle formation (PAP 400 X).

II which are comparable with other studies.^{2,4} A malignant thyroid FNA diagnosis accounts for approximately 5% (range, 2-16%) of all thyroid FNAs.⁷ Malignant cases (Category VI) in our study are within this range (2.79%, 2.62%) and comparable with studies in Table 3.

The two pathologists were in agreement in 548 cases (89.84%) and disagreed in 62 cases (10.16%) which is consistent with other studies.^{12–15} Cohen's kappa score of 0.628 is comparable to kappa score in a study by Awasthi et al.¹² In a study by Salillas et al.,¹⁴ the strength of agreement

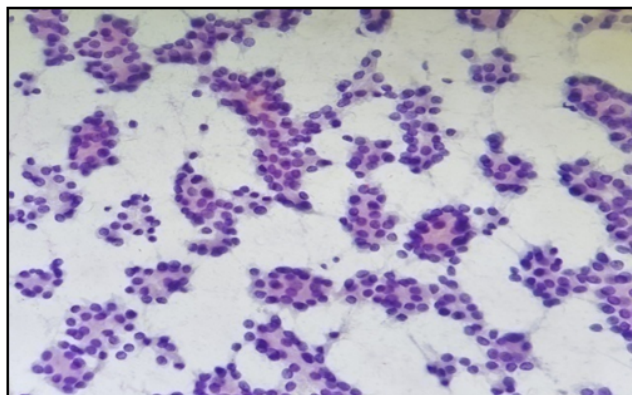


Fig. 4: Photomicrograph from a case of Follicular Neoplasm showing follicular cells arranged in crowded clusters and microfollicles (PAP 400 X).

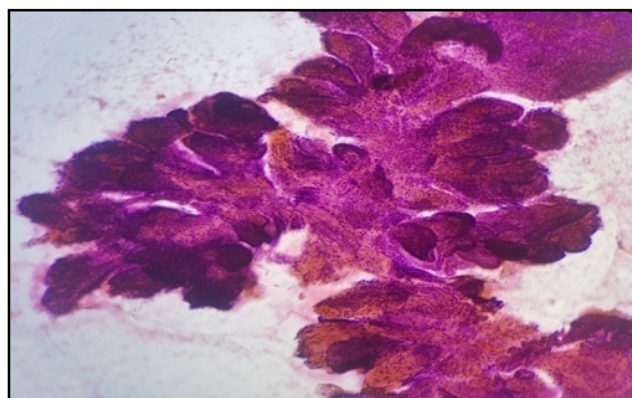


Fig. 5: Photomicrograph from a case of Papillary Thyroid Carcinoma (Bethesda Malignant category) showing multiple complex papillae with peripheral palisading of the tumor cells (PAP 100 X).

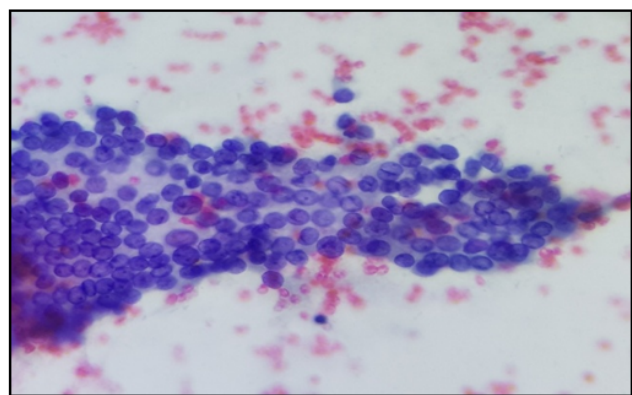


Fig. 6: Photomicrograph from a case of Papillary Thyroid Carcinoma showing optically clear nuclei with intranuclear grooves (PAP 400 X).

was very good with a kappa statistic of 0.90. In a study by Pathak et al.,¹⁵ unweighted Cohen's kappa score for consultant and SR was 0.7517 (strong), between consultant and JR was 0.5907 (moderate).

Diagnostic agreement was highest in the Benign category (95.68%), followed by Malignant category (87.5%) in our study. In a study by Bhasin et al.,¹³ maximum degree of agreements were noted in nondiagnostic (100%), malignant (100%) and benign (93.87%) categories. In a study by Awasthi et al.,¹² there was absolute agreement on the cases in ND/US and AUS/FLUS categories followed by 94.8% concordance in benign category. Lesser number of cases reported under ND/US category in above two studies could be reason for absolute agreement in ND category in their studies. In a study by Salillas et al.,¹⁴ the category under which maximum degree of agreement was noted was malignant (100%, $k=0.61$) followed by benign ($k=0.60$).

In our study, no agreement was seen in AUS (0%) category and SFM (0%) category between the two reporting pathologists. In a study by Bhasin et al.,¹³ maximum disagreement was noted in AUS/FLUS category that is consistent with our study. Of 7 discordant cases in study by Salillas et al.,¹⁴ 3 were SFM and 2 AUS which is consistent with our study. In a study by Pathak et al.,¹⁵ poor interobserver agreement level ($K=0.1301$) was observed in the AUS/FLUS category which is consistent with our study. In a study by Kocjan et al.,¹⁶ there was poor agreement for Thy3a ($\kappa = 0.11$) and Thy4 ($\kappa = 0.17$) categories which are similar to AUS and SFM categories respectively of Bethesda system and comparable to our study.

Low reproducibility for AUS/FLUS has been reported, variability in criteria used for AUS/FLUS is responsible for significant interobserver and inter-institutional variation in making diagnoses.¹⁵ The reproducibility of AUS/FLUS category is at best only fair.⁷ Significant interobserver variability of thyroid FNA though well established has been reported to be smaller (although still significant) for "benign" and "malignant" categories. AUS/FLUS category is one with the most interobserver variability among the cytopathologists. Differing threshold levels in applying the diagnostic criteria as noted in one of the studies could be a reason for this variability in our study too.¹⁷

5. Conclusion

The findings of our study are consistent with other published studies in literature. The systematic reporting according to TBSRTC has led to clear interpretation of the thyroid FNAC report. The use of this uniform terminology by two reporting pathologists revealed a moderate interobserver agreement that favours its use because of its relative ease of reproducibility. It also provides malignancy risk and management guideline for each category. Thus, the present study encourages the use of TBSRTC as a standardized reporting system in our institution and elsewhere for

effective communication among pathologists and clinicians with regard to thyroid FNAC reporting and management.

6. Conflict of interest

The authors have no conflicts of interest to declare.

7. Source of funding

The study was not funded by any source.

References

- Unnikrishnan AG, Menon UV. Thyroid disorders in India: An epidemiological perspective. *Indian J Endocr Metab.* 2011;15:78–81.
- Mondal SK, Sinha S, Basak B, Roy DN, Sinha SK. The Bethesda system for reporting thyroid fine needle aspirates: A cytologic study with histologic follow-up. *J Cytol.* 2013;30:94–99.
- Cibas ES, Sanchez MA. The National Cancer Institute thyroid fine-needle aspiration state-of-the-science conference: Inspiration for a uniform terminology linked to management guidelines. *Cancer Cytopathol.* 2008;114:71–73.
- Mehra P, Verma AK. Thyroid cytopathology reporting by the Bethesda system: A two-year prospective study in an academic institution. *Patholog Res Int.* 2015;p. 11.
- Orell SR, Sterrett GF. Orell & Sterrett's Fine Needle Aspiration Cytology. 5th ed. RELX India Private Limited ; 2012,.
- Cibas ES, Ali SZ. The Bethesda System for Reporting Thyroid Cytopathology. *Am J Clin Pathol.* 2009;132:658–665.
- Ali SZ, Cibas ES. The Bethesda System for Reporting Thyroid Cytopathology. Definitions, Criteria and Explanatory Notes. 2nd ed. Springer ; 2018,.
- Haugen BR, Alexander EK, Bible KC. 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.* 2015;26(1):1–133.
- Bhat S, Bhat N, Bashir H. The Bethesda System for Reporting Thyroid Cytopathology: a two year institutional audit. *Int J Cur Res Rev.* 2016;8(6):5–11.
- 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;134:450–456.
- Laishram RS, Zothanmawii T, Joute Z, Yasung P, Debnath K. The Bethesda system of reporting thyroid fine needle aspirates: A 2-year cytologic study in a tertiary care institute. *J Med Soc.* 2017;31(1):3–7.
- Awasthi P, Goel G, Khurana U, Joshi D, Majumdar K, Kapoor N. Reproducibility of “The Bethesda system for reporting thyroid cytopathology:” A retrospective analysis of 107 patients. *J Cytol.* 2018;35(1):33–36.
- Bhasin TS, Mannan R, Manjari M. Reproducibility of “The Bethesda System for reporting Thyroid Cytopathology”: A Multicenter study with review of the literature. *J Clin Diagn Res.* 2013;7(6):1051–1054.
- Salillas AL, Sun F, Almocera EG. Review of the Bethesda System for Reporting Thyroid Cytopathology: A Local Study in Bohol Island. *Acta Cytologica.* 2015;59:77–82.
- Pathak P, Srivastava R, Singh N, Arora VK, Bhatia A. Implementation of the Bethesda system for reporting thyroid cytopathology: Inter-observer concordance and reclassification of previously inconclusive aspirates. *Diagn Cytopathol.* 2014;42(11):944–949.
- Kocjan G, Chandra A, Cross PA. The interobserver reproducibility of thyroid fine needle aspiration using the UK Royal College of Pathologists' Classification System. *Am J Clin Pathol.* 2011;135:852–859.
- Padmanabhan V, Marshall CB, Barkan GA. Reproducibility of atypia of undetermined significance/follicular lesion of undetermined significance category using the Bethesda system for reporting thyroid cytology when reviewing slides from different institutions: A study of interobserver variability among cytopathologists. *Diagn Cytopathol.* 2017;00:1–7.

Author biography

Chhavi Gupta Resident

Subhash Bhardwaj Professor

Sindhu Sharma Associate Professor

Cite this article: Gupta C, Bhardwaj S, Sharma S. The interobserver variability of thyroid fine needle aspirates using The Bethesda System for Reporting Thyroid Cytopathology. *Indian J Pathol Oncol* 2020;7(1):76–82.