

## The interobserver reproducibility of thyroid cytopathology using Bethesda Reporting System: Analysis of 200 cases

Safina Ahmed,<sup>1</sup> Mumtaz Ahmad,<sup>2</sup> Masood Ahmad Khan,<sup>3</sup> Faiza Kazi,<sup>4</sup> Fozia Noreen,<sup>5</sup> Samia Nawaz,<sup>6</sup> Iram Sohail<sup>7</sup>

### Abstract

**Objective:** To determine interobserver reproducibility of thyroid cytopathology in cases of thyroid fine needle aspirates.

**Methods:** The retrospective, descriptive study, was conducted at the Foundation University Medical College, Islamabad, using cases related to period between 2009 and 2011. A total of 200 cases of fine-needle aspirations were retrieved from the archives. Three histopathologists independently categorised them into 6 groups according to Bethesda reporting system guidelines without looking at previous reports. Kappa statistics were used for analysis of the results on SPSS 17.

**Results:** Of the 200 patients, 194 (97%) were females and 6 (3%) were males. The overall mean age of patients was 46±20 years. Kappa value calculated for observer-1 and observer-2 was 0.735; for observer-1 and observer-3, 0.841; and for observer-2 and observer-3, 0.838, showing substantial interobserver agreement. Histopathological correlation was available, for 39(19.5%). Of these cases, 5(13%) were 'non-diagnostic, 20(51%) 'benign, 2(5%) 'atypia of undetermined significance/follicular lesion of undetermined significance, 6(15%) 'follicular neoplasm, 1(3%) 'suspicious for malignancy, and 5(13%) 'malignant.

**Conclusions:** Good overall interobserver agreement was found, but discordance was seen when certain categories were analysed separately.

**Keywords:** Thyroid, Fine-needle aspiration, Bethesda System, Interobserver reproducibility. (JPMA 63: 1252; 2013)

### Introduction

Thyroid fine needle aspiration (FNA) is one of the most commonly performed diagnostic procedures in out-patient departments.<sup>1,2</sup> It is quick, cost-effective, and minimally invasive technique which reduces the need for an un-necessary surgery.<sup>3</sup> FNA has been recommended as initial diagnostic test in the evaluation and management of thyroid nodules, having sensitivity and specificity of 95% and 90% respectively.<sup>4</sup>

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 didn't show proper clinical relevance and created confusion among pathologists, endocrinologists, surgeons and radiologists.<sup>5,6</sup>

National Cancer Institute, Bethesda, USA, after holding a series of conferences and discussions, introduced the Bethesda Reporting System in 2007,<sup>7</sup> which established a six-tier classification system for reporting thyroid lesions.

It was expected to provide standardised framework for laboratory reports which would reduce confusion among different healthcare providers. It was to enable pathologists to diagnose thyroid lesions as distinct entities according to accepted criteria and provide clear guidance for clinical management.<sup>8</sup> Moreover, it will also help in the prospective research work related to thyroid malignancies.<sup>9</sup>

The purpose of this study was to evaluate Bethesda Classification for reporting thyroid by recording interobserver reproducibility. To the best of our knowledge, it is the first study done in Pakistan in which interobserver agreement for Bethesda System was analysed.

### Patients and Methods

A total of 200 thyroid FNA cases were retrieved from the archives from 2009 to 2011 at the Pathology Department of Foundation University Medical College, Islamabad. All cases were independently reviewed by 3 histopathologists and were placed into 6 categories: Non-diagnostic/Unsatisfactory ND' (category-1), Benign B' (category-2); Atypia of undetermined significance/Follicular Lesion of Undetermined Significance 'AUS/FLUS' (category-3), Follicular Neoplasm/Suspicious for Follicular Neoplasm 'FN/SFN' (category-4); Suspicious for malignancy 'SFM'

<sup>1-4</sup>Department of Pathology, <sup>5-7</sup>FCPS II trainee, Foundation University Medical College, Islamabad.

**Correspondence:** Safina Ahmed. Email: a\_safina@yahoo.com

(category-5); and Malignant 'M' (category-6). by using the Bethesda System<sup>10</sup> for reporting thyroid cytopathology.

Previous cytology reports were not disclosed to the examining histopathologists. The cytological diagnosis was correlated with histopathology where available. Kappa statistics were applied combining observers 1 and 2, observers 1 and 3 and observers 2 and 3.

Kappa values were interpreted as: 0 to 0.2, slight agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60,

moderate agreement; 0.61 to 0.80, substantial agreement, and 0.8 to 1.00, almost perfect agreement.<sup>11</sup> All statistical analyses were performed using SPSS 17.

### Results

The overall mean age of the 200 cases selected was 46±20 years. Among them, 194(97%) were females and 6(3%) were males. These patients had presented with a single thyroid nodule or a rapidly enlarging dominant nodule in a multinodular goitre.

Table-1: Cross-tabulation between observer 1 and observer 2.

	Observer 2						Total
	Non-diagnostic	Benign	Atypia of undetermined significance	Follicular Neoplasm	Suspicious for malignancy	Malignant	
<b>Observer 1</b>							
Non-diagnostic	43	7	0	0	0	50	
Benign	6	122	0	4	0	0	132
Atypia of undetermined significance	0	0	1	0	0	1	2
Follicular Neoplasm	0	5	0	3	0	1	9
Suspicious for malignancy	0	1	0	0	1	0	2
Malignant	0	0	0	0	1	4	5
Total	49	135	1	7	2	6	200

Table-2: Cross-tabulation between observer 1 and observer 3.

	Observer 3						Total
	Non-diagnostic	Benign	Atypia of undetermined significance	Follicular Neoplasm	Suspicious for malignancy	Malignant	
<b>Observer 1</b>							
Non-diagnostic	44	6	0	0	0	0	50
Benign	3	125	0	4	0	0	132
Atypia of undetermined significance	0	0	1	0	1	0	2
Follicular Neoplasm	0	1	0	7	1	0	9
Suspicious for malignancy	0	0	0	0	2	0	2
Malignant	0	0	0	0	0	5	5
Total	47	132	1	11	4	5	200

Table-3: Cross-tabulation between observer 2 and observer 3.

	Observer 3						Total
	Non-diagnostic	Benign	Atypia of undetermined significance	Follicular Neoplasm	Suspicious for malignancy	Malignant	
<b>Observer 2</b>							
Non-diagnostic	44	5	0	0	0	0	49
Benign	3	127	0	4	1	0	135
Atypia of undetermined significance	0	0	1	0	0	0	1
Follicular Neoplasm	0	0	0	7	0	0	7
Suspicious for malignancy	0	0	0	0	1	1	2
Malignant	0	0	0	0	2	4	6
Total	47	132	1	11	4	5	200

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

The histopathological correlation was available for 39 (19.5%) cases. Among them, 5 (13%) belonged to 'ND' category; 4 (80%) of which were diagnosed as multinodular goitre while 1 (20%) as malignant. Category 'B' had 20 (51%) cases, of which 19 (95%) turned out to be multinodular goitre and 1 (5%) follicular adenoma. The 'AUS/FLUS' category had 2 (5%) cases, of which 1 (50%) was multinodular goitre and 1 (50%) follicular adenoma. Category SFN/FN had 6 (15%) cases, of which 2 (33%) were multinodular goitre 3 (50%) were follicular adenoma and 1 (17%) was malignant. 'SFM' and 'M' categories had 1 (3%) and 5 (13%) cases respectively which were categorized as malignant on histopathology ( $p < 0.01$ ).

## Discussion

Thyroid nodule is a common clinical presentation in a surgical out-patient department (OPD). About 5-15% of these nodules turn out to be malignant, and among them papillary and follicular carcinomas are the most common, comprising almost 90% of thyroid malignancies.<sup>12</sup> Although thyroid FNA has been widely used as a first-line intervention to assess thyroid nodules, but no standard terminology was available for its reporting. Mostly reports were made on the descriptive format.<sup>13</sup> Introduction of Bethesda System was an attempt to bring uniformity in reporting.

In this study, thyroid nodules were more common in females compared to males which is consistent with previously published data.<sup>14</sup> Majority of our cases were classified as 'benign'. Sushel et al<sup>15</sup> and other similar studies done on histopathological pattern of diagnosis in thyroid patients show that benign lesions are more common.

Good interobserver agreement was found for inadequate, benign and malignant categories among all the 3 observers, which was anticipated because any good reporting system should have robust criteria to differentiate benign from malignant lesions. This is in accordance with a study done in the UK by Kocjan et al<sup>16</sup> on 200 thyroid FNA cases. It used Royal College of Pathologists Classification System and showed good interobserver agreement ( $k=0.72$ ) for combined categories implying surgical management (Thy3f, Thy4, Thy5) as well as for combined categories implying conservative management (Thy1, Thy2, Thy3a).

However, in our study, discordance was found in AUS/FLUS, SFN/FN and SFM categories. Kocjan et al<sup>16</sup> also reported poor agreement for categories Thy3a ( $k=0.11$ ) and Thy4 ( $k=0.17$ ) when they were assessed separately.

The main difficulty is always posed by borderline lesions. Category AUS/FLUS has features that raise the possibility of neoplasia, but are insufficient to categorise it as malignant. This category carries 5-10% risk of malignancy. In our study, the percentage of cases in AUS/FLUS was 5% (among those cases for which histopathology diagnosis was available). It was in keeping with Bethesda System recommendation which says that diagnostic rate for atypical category should be less than 7% and its over-diagnosis must be avoided.<sup>17</sup>

The interobserver agreement for this category was poor in our series probably because of lower number of cases and essentially due to the fact that many criteria of atypia are based on subjective morphological features. Similarly, a study done by Paul A et al,<sup>18</sup> in which they analysed AUS rates over a five-year period, revealed notable intra and interobserver variability. The overall AUS rate in this study was 11.2%. Although the criteria for atypical category are well defined, but different pathologists have variable thresholds in applying them.

We also observed discordance among our pathologists for SFN/FN category. et al<sup>19</sup> studied thyroid lesions showing predominantly colloid and follicular groups, and revealed poor interobserver agreement ( $k=0.35$ ) for them. The level of agreement improved markedly when the categories of follicular lesion and follicular neoplasm were combined together. Similarly, Gerhard et al<sup>20</sup> also reported major intra and interobserver agreement on follicular thyroid lesions. We speculate that it may be because of the lack of stringent criteria on cellularity, proportion of follicular cells forming microfollicles and amount of colloid present in the background. On the contrary, Clary et al<sup>21</sup> analysed interobserver agreement of follicular lesions among four observers. The range of agreement was fair to substantial ( $k=0.199-0.617$ ).

Although the use of standard Bethesda Reporting System provides a good distinction between benign and malignant lesions, but borderline categories still impart diagnostic problem, showing poor interobserver agreement. A strict application of the defined diagnostic criteria and help of ancillary techniques is required to overcome these issues.

## Conclusion

Substantial interobserver agreement was found for thyroid cytological lesions using Bethesda Reporting

Criteria. Introduction of the system is a good step towards standardisation of cytology reports. It provides fine distinction between benign and malignant cytological lesions. However, the AUS/FLUS category is heterogeneous and its efficacy as separate cytologic group is still controversial.

### Acknowledgement

We are grateful to all the laboratory technologists who were very helpful.

### References

1. Khan A, Khan MM, Shah S. Effectiveness of fine needle aspiration cytology in diagnosis of cold thyroid nodules. *J Med Sci* 2005; 13: 148-50.
2. Zhang YX, Zhang B, Zhang ZH, Guo HQ, Wang Y, Xu ZG et-al. Fine-needle aspiration cytology of thyroid nodules: a clinical evaluation. *Zhonghua Er Bi Yan Hou Tou Jing Wai Ke Za Zhi* 2011; 46: 892-6.
3. Baqqa PK, Mahajan NC. Fine needle aspiration cytology of thyroid swelling: how useful and accurate is it? *Indian J Cancer* 2010; 47: 437-42.
4. Smadi AA, Ajarmeh K, Wreikat F. Fine-needle aspiration of thyroid nodules has high sensitivity and specificity. *Rawal Med J* 2008; 33: 221-4.
5. Cibas ES. Fine-needle aspiration in the workup of thyroid nodules. *Otolaryngol Clin North Am* 2010; 43: 257-71.
6. 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: 205-12.
7. Schinstine M. A brief description of the Bethesda system for reporting thyroid fine needle aspirates. *Hawaii Med J* 2010; 69: 176-8.
8. Chung YS, Yoo C, Jung JH, Choi HJ, Suh YJ. Review of atypical cytology of thyroid nodule according to the Bethesda system and its beneficial effect in the surgical treatment of papillary carcinoma. *J Korean Surg Soc* 2011; 81: 75-84.
9. Crippa S, Mazzucchelli L, Cibas ES, Ali SZ. The Bethesda system for reporting thyroid fine-needle aspiration specimens. *Am J Clin Pathol* 2010; 134: 343-4.
10. Cibas ES, Ali SZ. NCI Thyroid FNA State of the Science Conference. The Bethesda system for reporting thyroid cytopathology. *Am J Clin Pathol* 2009; 132: 658-65.
11. Kundel HL, Polansky M. Measurement of observer agreement. *Radiology* 2003; 228: 303-8.
12. Touzopoulos P, Karanikas M, Zarogoulidis P, Mitrakas A, Porpodis K, Katsikogiannis N, et al. Current surgical status of thyroid diseases. *J Multidiscip Healthc* 2011; 4: 441-9.
13. Redman R, Yoder BJ, Massoll NA. Perceptions of diagnostic terminology and cytopathologic reporting of fine-needle aspiration biopsies of thyroid nodules: a survey of clinicians and pathologists. *Thyroid* 2006; 16: 1003-8.
14. Taddesse A, Yaqub A. Clinical, sonographic and cytological evaluation of small versus large thyroid nodules. *J Pak Med Assoc* 2011; 61: 466-9.
15. Sushel C, Khanzada TW, Zulfikar I, Samad A. Histopathological pattern of diagnosis in patients undergoing thyroid operations. *Rawal Med J* 2009; 34:14-6.
16. Kocjan G, Chandra A, Cross PA, Giles T, Johnson SJ, Stephenson TJ, et-al. The interobserver reproducibility of thyroid fine-needle aspiration using the UK Royal College of Pathologists' classification system. *Am J Clin Pathol* 2011; 135: 852-9.
17. Shi Y, Ding X, Klein M, Sugrue C, Matano S, Edelman M, et-al. Thyroid fine-needle aspiration with atypia of undetermined significance: a necessary or optional category? *Cancer* 2009; 117: 298-304.
18. Vanderlaan PA, Krane JF, Cibas ES. The frequency of 'atypia of undetermined significance' interpretations for thyroid fine-needle aspirations is negatively correlated with histologically proven malignant outcomes. *Acta Cytol* 2011; 55: 512-7.
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: 239-44.
20. Gerhard R, da Cunha Santos G. Inter- and intraobserver reproducibility of thyroid fine needle aspiration cytology: an analysis of discrepant cases. *Cytopathology* 2007; 18: 105-11.
21. Clary KM, Condel JL, Liu Y, Johnson DR, Grzybicki DM, Raab SS. Interobserver variability in the fine needle aspiration biopsy diagnosis of follicular lesions of the thyroid gland. *Acta Cytol* 2005; 49: 378-82.