Comparison of Results of Bethesda System of Cytology with Final Histopathology in the Evaluation of Thyroid Nodules

Dhanya Ramachandran[1*] **and Madhumita Kumar**[2]

[1]*Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, United States*

[2]*Department of Otorhinolaryngology, Amrita Vishwa Vidyapeetham University, Kochi, Kerala, India*

Abstract

Background: Thyroid nodules are very commonly encountered in clinical practice and there exists a need to effectively diagnose malignancy. Fine Needle Aspiration Cytology (FNAC) plays an important role in the initial evaluation of these patients. However, reports of FNAC samples are limited by subjective error, reporter bias, inter-observer variability, and terminology confusion. The Bethesda System of Reporting of Thyroid Cytology (BSRTC) was developed in 2007 as consensus recommendations to overcome these limitations. Six diagnostic Categories were formed which provided the information regarding the risk of malignancy in each. This study aims to assess the accuracy of the BSRTC in the detection of malignancy at a tertiary referral institute.

Method: A cross-sectional diagnostic accuracy study of 110 patients with thyroid swelling, with diagnostic FNAC and underwent surgical intervention at a tertiary care centre between January 2015 and March 2016. The results were obtained by calculating validity parameters (sensitivity, specificity, positive predictive value, negative predictive value, and accuracy) for cytology with respect to histopathology. Rates of malignancy in each category (II, V, and VI) were also calculated.26 cases of Category III lesions were analyzed separately to calculate the rates of malignancy on histopathology in this category

Results: Validity parameters calculated for cytology with respect to histopathology showed sensitivity of 66%, specificity of 88%, positive predictive value of 87%, negative predictive value of 69%, and accuracy of 76%. 20 of the 64 patients in Category II (31%) were diagnosed to be malignant on histopathology. 25 of 29 in Category V lesions (86%), and 16 of the 17 Category VI lesions (94%) were malignant on histopathology. Of the 26 patients in Category III, 12 (46%) were benign and 14 (54%) were malignant

Conclusion: The risk of malignancy in each of the six diagnostic categories should be independently defined at each institution.

Keywords: Thyroid • Bethesda • Cytology • Validity

Introduction

Thyroid nodules are present in about 60% of the general population and malignancy occurs in about 10% of these nodules. There thus exists a need to effectively diagnose malignancy in order to provide the best treatment for these patients. The strongest data collected in a systematic analysis of nodule size suggest that prevalence of malignancy is higher within larger nodules, whether a cutoff of 3 cm or 4 cm is utilized. Multiple modalities exist for the evaluation of thyroid disease, and these play roles complementary to each other.

Physical examination, thyroid function tests, ultrasonography, and fine-needle aspiration cytology are used in the initial assessment. Each step of evaluation adds up to complete the clinical picture of a patient with thyroid swelling, each with its own benefits and setbacks. It is due to this that no single modality is yet used alone in diagnosis and management of thyroid disease [1]. First developed by Martin and Ellis in 1930, FNAC has been developed to such an extent that it is now the standard for initial assessment of thyroid nodules, owing to its reliability and cost-effectiveness. It may be used as a screening test in the detection of malignancy in thyroid nodules, and accuracy of this increases if the test is done under

*****Address for Correspondence:** Dhanya Ramachandran, Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, United States, Tel: 6093564297; E-mail: dhanya.ramachandran@outlook.com

ultrasonographic guidance. The complications that arise from FNAC are few and not serious, and may include slight pain and skin discoloration at the site of aspiration [2].

Microscopic analysis of samples obtained by FNAC is prone to subjective error and there inherently exists a reporter bias and inter-observer variability. The various terms used to describe a cytology specimen also leads to terminology confusion. The BSRTC was developed in 2007 as consensus recommendations during the National Cancer Institute's thyroid fine needle aspiration state of the science conference which was held at Bethesda, Maryland, USA. These guidelines were formed in order to overcome the various limitations faced during the reporting of thyroid cytology [3]. Based on the definitions, diagnostic criteria, and morphologic criteria which were provided by the BSRTC, six diagnostic Categories were formed, which provided the practitioner with information regarding the risk of malignancy in each Table 1.

Understandably, the risk of malignancy increases with each Category. Understanding the accuracy of the risk of malignancy estimated through thyroid nodule FNAC is essential when making management decisions. FNAC results often guide the decision to observe or pursue surgical intervention, given the diagnostic test's typically favorable sensitivity and specificity [4]. Only Categories II, V, and VI provide information on the presence or absence of malignancy in the examined nodule. Category IV includes follicular and oncocytic (formerly called Hurthle cell) neoplasms. Benign and malignant follicular neoplasms and oncocytic adenomas and carcinomas cannot be distinguished on the basis of cytology alone, as tissue architecture is required to make the diagnosis of malignancy through observation of capsular or angiolymphatic invasion [5]. Also, in larger nodules, due to the higher underlying prevalence of malignancy, a malignant FNA result is more likely to be truly positive, whereas a benign FNA result is less likely to be truly negative.

This study aims to assess the accuracy of the BSRTC in the detection of malignancy at a tertiary referral institute by calculating validity parameters for cytology with respect to histopathology using standard statistical definitions as set aside by the AACE/AME Task Force on Thyroid Nodules [6].

Materials and Methods

This is a cross-sectional diagnostic accuracy study of 110 patients with thyroid swelling who underwent FNAC (reported by cytologists at the Department of Clinical Pathology) and surgical intervention in the Departments of Otorhinolaryngology, Head and Neck Surgery and Oncology, and General Surgery at the Amrita Institute of Medical Sciences and Research Centre, Cochin between January 2015 and March 2016, after the approval of the Ethics Committee at this institute [7].

Inclusion criteria

Patients with diagnostic FNAC (Categories II, V, and VI) who underwent surgical intervention at this institute (total thyroidectomy or hemi- thyroidectomy).

Exclusion criteria

Patients with non-diagnostic FNAC (Categories I, III, and IV).

Calculation of sample size

Based on the results of previous studies, and with 20% allowable error and 95% confidence, the minimum sample size comes to 60. In this study, 110 patients were included, which gives 15% error and 95% confidence [8].

Primary objective: To evaluate the diagnostic agreement of Bethesda System of Reporting Thyroid Cytopathology (BSRTC) with final histopathology in patients with thyroid disease who undergo surgical intervention.

Secondary objective:

- Evaluation of outcomes of Category III lesions.
- Assessment of patient profiles.

Statistics

The electronic medical records of patients with thyroid nodules who underwent surgical management were reviewed, and their demographic details, pre-operative cytology reports and final histopathology reports were collected. Patients with pre-operative cytology of Categories II, V and VI of the BSRTC were included in the study group as they provide information of the lesion being benign or malignant [9]. For the purpose of statistical analysis, Categories V and VI were analyzed together as "malignant on FNAC". Patients with pre-operative cytology of Categories I, III and IV were excluded since it does not provide information with regards to presence or absence of malignancy in the smear. Follicular lesions (Category IV) require the detection of angioinvasion on histopathology for the diagnosis of malignancy. The results of the study were obtained by calculating validity parameters using standard definitions (for sensitivity, specificity, positive predictive value, negative predictive value, and accuracy) for cytology with respect to histopathology.

Statistical definitions used:

- Sensitivity=True positives/True positives+False negatives
- Specificity=True negatives/False positives+True negatives
- Positive predictive value=True positives/True positives +False positives
- Negative predictive value=True negatives/True negatives +False negatives
- Accuracy=True positives+True negatives/Total

Rates of malignancy in each category (II, V, and VI) were also calculated. 26 cases of Category III lesions were analyzed separately to calculate the rates of malignancy in this category.

Results

During the study 110 patients with thyroid nodules who underwent FNAC and thyroidectomy were included in the study, after considering the inclusion and exclusion criteria (Table-1).

BSRTC category	Interpretation	Risk of malignancy (%)
I	Non diagnostic/ unsatisfactory smears	-
II	Benign	0-3
III	Atypia of undetermined significance/follicular lesion of undetermined significance (AUS/FLUS)	5-15
IV	Follicular neoplasm/ suspicious for follicular neoplasm; Hurthle cell neoplasm/suspicious for Hurthle cell neoplasm	15-30
V	Suspicious for malignancy	60-75
VI	Malignant	97-99

Table 1. The Bethesda system of reporting of thyroid cytology.

Results of validity parameters

The pre-operative diagnosis of benign or malignant based on cytology was compared to the final histopathological diagnosis of the thyroidectomy specimen (Table 2). Validity parameters calculated for cytology with respect to histopathology showed sensitivity of 66%, specificity of 88%, positive predictive value of 87%, negative predictive value of 69%, and accuracy of 76%.

	Malignant on HPE	Benign on HPE
Malignant on FNAC	40	6
(Cat. V, VI)	87%	13%
N=46	(True positives)	(False positives)
Benign on FNAC	20	44
(Cat. II)	31%	69%
N=64	(False negatives)	(True negatives)

Table 2. Results of the study.

Results of rates of malignancy in each category

20 of the 64 patients in Category II (31%) were diagnosed to be malignant on histopathology. Categories V and VI which were analyzed together for the purpose of statistical calculations, were also analyzed separately (Table 3). 25 of 29 in Category V lesions (86%), and 16 of the 17 Category VI lesions (94%) were malignant on histopathology. This shows that Category VI was more accurate in determining the presence of malignancy than Category V (Figure 1) [10].

Study	Rate of malignancy in category III (%)
Bongiovanni	14
Mondal	20
Mufti	50
Muratli	21
Naz	33
Jo	17
This study	51

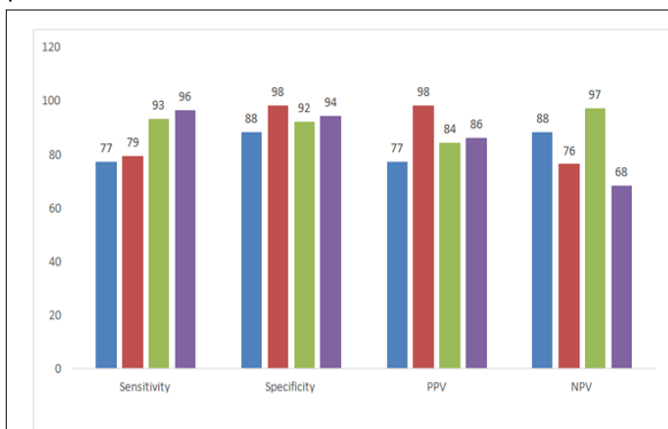**Table 3.** Comparison of outcomes of category III with previous studies.

Figure 1. Comparison of Validity Parameters of this study with previous studies. Note: ■ Mehra, ■ Kesler, ■ Haberal, ■ This study

Results of assessment of patient profiles

Of the 110 patients studied, 80 were females (73%) and 30 were males (27%). Among females, malignancy was seen as commonly as benign lesions. However, males were more likely to have malignant lesions (63%) than benign (37%) (Figure 1).

Results of rate of malignancy in category III

26 cases of Category III lesions which were operated on were analyzed separately to assess the rates of malignancy in this category. Of the 26 patients in Category III, 12 (46%) were benign on histopathology and 14 (54%) were malignant (Figure 2).
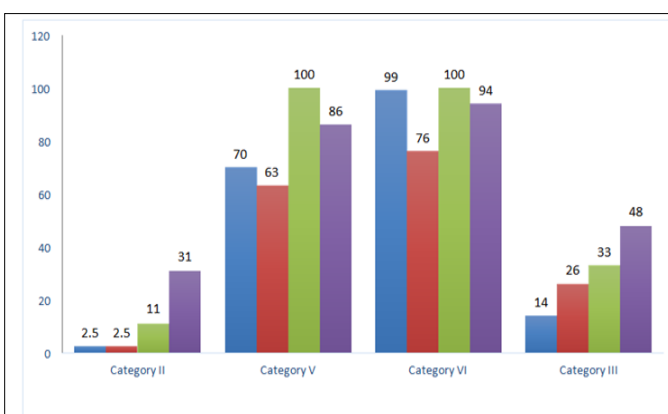

Figure 2. Comparison of rates of malignancy in this study with previous studies. Note: ■ Bongiovanni, ■ Muratli, ■ Naz, ■ This study

Discussion

An analysis of the validity parameters calculated by various others studies was performed. The validity parameters calculated in this study are not as high as compared to other studies (Figure 1). However, there exists much variability in the statistical calculations among those studies, and this can affect the results [11]. FNA has multiple levels of results, and thus holds inherent potential variability in sensitivity and specificity depending on how intermediate/indeterminate results are categorized. If suspicious lesions are considered positive (as in this study), the sensitivity increases while the specificity decreases. If suspicious lesions are excluded, then the sensitivity decreases and the false negative rates increase. Also, this study only evaluates those patients who underwent surgical intervention, and this inherently injects bias into the analysis. False negative results mean the missed malignancies [12]. The most likely cause of these is improper sampling, especially if the nodule is less than 1 cm or more than 4 cm. In this study, it has been demonstrated that a significant percentage (31%) of lesions diagnosed to be benign on FNAC was in fact malignant as seen on histopathology. This is far in excess of the predicted rate of malignancy in Category II according to the BSRTC (0%-3%), and also higher than reported in previous studies [13].

The rates of malignancy in Categories V and VI according to this study are comparable to previous studies as well as the BSRTC (Figure 2) [14]. The rate of detection of malignancy in the Category V is higher than that estimated by the BSRTC, and is lower for Category VI. This shows that it may in fact be more accurate to classify some of the Category V lesions as Category VI at this institute. Category III lesions also have a high rate of malignancy, and this is compared to previous studies (Table 3) [15]. It should be noted that any estimate of the risk of malignancy in indeterminate thyroid nodules is limited by the fact that not all nodules undergo surgery. A large proportion of patients diagnosed in our institute as Category III do not undergo surgical intervention. It is thus likely that the rate of malignancy in Category III lesions is falsely elevated in this study since the true number of Category III diagnoses is unknown. The follicular variant of Papillary Thyroid Carcinoma (fvPTC) is not as aggressive as classical papillary carcinoma, and has a lower risk of regional and distant metastasis [16,17]. A study by Kleiman et al shows that Category III or IV nodules are most likely fvPTC, and Category V or VI nodules are most likely classic PTC. This may also explain the high rate of malignancy in Category III lesions.

Limitations

One limitation of this study is that cases of microcarcinomas detected in specimens which are predominantly benign, and cases of fvPTC were included in the malignant group and this may also have altered the results. The strength of this study would also have improved if every FNAC was performed under ultrasonographic guidance, and if the ultrasonographic features were taken into account as well.

Recommendations

Care should also be taken when interpreting a reported false-negative rate in the literature. It is possible that the reason for the variability in the validity parameters may be that standard definitions were often not utilized in many of the primary reports. Thus, following the standard definitions is required to improve future studies. Similar to standardizing the reporting of intermediate results, it would allow direct comparison of individual studies' reported results.

Conclusion

In conclusion, this study illustrates that the risk of malignancy in each of the six diagnostic categories should be independently defined at each cytology centre or institution. Further studies are required which utilize strict inclusion criteria, and also analyze the role of nodule size on the occurrence of malignancy. The role of ultrasonography in the evaluation of thyroid nodules also needs to be considered. The BSRTC represents a significant step forward to enable standardization, reproducibility, and predictive value in the evaluation of thyroid nodules, and institutions need to be encouraged to utilize the BSRTC to facilitate easy sharing of data across the world.

References

1. Bukhari, M H, Khan A A, Niazi S, and Arshad M, et al. "Better Thyroid Cytopathology Reporting System May Increase the Clinical Management and Patients Outcome." *J Cytol Histol* 3 (2012): 158.

2. Shin, Jennifer J, Caragacianu Diana, and Randolph Gregory W. "Impact of Thyroid Nodule Size on Prevalence and Post-Test Probability of Malignancy: A Systematic Review." *Laryngoscope* 125 (2015): 263-272.

3. Gharib, Hossein, Papini Enrico, Paschke Ralf, and Duick Daniel S, et al. "American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and European Thyroid Association Medical Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules: Executive Summary of the Recommendations." *J Endocrinol Investigat* 33 (2010): 287-291.

4. Tuluc, M, Solomides C. "Thyroid Cytology." *Otolaryn Clin North Am* 47 (2014): 475-489.

5. Camargo, Rosalinda YA, Tomimori Eduardo K, Knobel Meyer, and Medeiros-Neto Geraldo. "Preoperative Assessment of Thyroid Nodules: Role of Ultrasonography and Fine Needle Aspiration Biopsy Followed by Cytology." *Clinics* 62 (2007): 411-418.

6. Mehra, Payal, and Verma Anand Kumar. "Thyroid Cytopathology Reporting by the Bethesda System: A Two-Year Prospective Study in an Academic Institution." *Pathol Res Int* (2015).

7. Cibas, Edmund S, and Ali Syed Z. "The Bethesda System for Reporting Thyroid Cytopathology." *Am J Clin Pathol* 132 (2009): 658-665.

8. Kessler, Alexander, Gavriel Haim, Zahav Sergei, and Vaiman Michael, et al. "Accuracy and Consistency of Fine-Needle Aspiration Biopsy in the Diagnosis and Management of Solitary Thyroid Nodules." *Isr Med Assoc J* 7 (2005): 371-373.

9. Haberal, A N, Toru S, Ozen O, and Arat Z, et al. "Diagnostic Pitfalls in the Evaluation of Fine Needle Aspiration Cytology of the Thyroid: Correlation with Histopathology in 260 Cases." *Cytopathology* 20 (2009): 103-108.

10. Petersen, S, Greisen O. "Fine-Needle Aspiration Biopsy of the Thyroid Gland." *Head Neck Surg Off J* 9 (1984): 1-24.

11. Haugen, Bryan R, Alexander Erik K, Bible Keith C, and Doherty Gerard M, 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* 26 (2016): 1-133.

12. Muratli, Asli, Erdogan Nilsen, Sevim Sezgin, and Unal Isik, et al. "Diagnostic Efficacy and Importance of Fine-Needle Aspiration Cytology of Thyroid Nodules." *J Cytology/Indian Acad Cytol* 31 (2014): 73.

13. Naz, Samreen, Ali Hashmi Atif, Faridi Naveen, and Muzzammil Edhi Muhammad, et al. "Diagnostic Accuracy of Bethesda System for Reporting Thyroid Cytopathology: An Institutional Perspective." *Int Archiv Med* 7 (2014): 1-5.

14. Mondal, Santosh Kumar, Sinha Simanti, Basak Bijan, and Nag Roy Dipanwita, et al. "The Bethesda System for Reporting Thyroid Fine Needle Aspirates: A Cytologic Study with Histologic Follow-Up." *J Cytol/Indian Acad Cytolog* 30 (2013): 94.

15. Mufti, Shagufta Tahir, and Molah Rihab. "The Bethesda System for Reporting Thyroid Cytopathology: A Five-Year Retrospective Review of One Center Experience ." *Int J Health Sci* 6 (2012): 159.

16. Jo, Vickie Y, Stelow Edward B, Dustin Simone M, and Hanley Krisztina Z. "Malignancy Risk for Fine-Needle Aspiration of Thyroid Lesions According to the Bethesda System for Reporting Thyroid Cytopathology ." *Am J Clin Pathol* 134 (2010): 450-456.

17. Kleiman, David A, Beninato Toni, Soni Ashwin, and Shou Yiyan, et al. "Does Bethesda Category Predict Aggressive Features in Malignant Thyroid Nodules?." *Ann Surg Oncol* 20 (2013): 3484-3490.

How to cite this article: Ramachandran, Dhanya and Madhumita Kumar. "Comparison of Results of Bethesda System of Cytology with Final Histopathology in the Evaluation of Thyroid Nodules." *J Cytol Histol* 13 (2022) : 651