Paper

Diagnostic terminology for reporting thyroid cytology: Inter-observer agreement between the format currently used in Sri Lanka and the Bethesda system

IHDS Prabath , AAH Priyani, MDS Lokuhetty

Abstract

Background: Thyroid cytology reporting has become standardized with the introduction of the Bethesda System of Reporting Thyroid Cytology (BSRTC), linking up diagnostic categories with clinical management guidelines. The current thyroid cytology reporting format in Sri Lanka (TCRSL) is based on the original 'Thy' classification of the British Thyroid Association. The number of diagnostic categories have been increased in the BSRTC than in TCRSL (6 vs. 5 categories), with a possible adverse impact on inter-observer agreement (IOA). IOA of both reporting formats were thus compared to evaluate the appropriateness of looking beyond TCRSL in Sri Lanka, a country facing a significant thyroid cytology burden.

Objective: To compare inter-observer agreement between the thyroid cytology reporting format in Sri Lanka and the Bethesda format.

Methods: 100 thyroid aspirates received by the department from January 2008 to May 2012 were selected by stratified random sampling.

Retrieved smears were reviewed and classified using TCRSL and BSRTC by two pathologists independently. Results were tabulated using SPSS. Linear weighted κ and composite proportion of agreement values were used to evaluate the IOA of responses in both formats.

Results: Linear weighted κ was 0.6138 (95% C.I. 0.4797 - 0.7479) and the composite proportion of agreement was 68% for TCRSL. In BSRTC, linear weighted κ was 0.6795 (95% C.I. 0.563 - 0.796) and the composite proportion of agreement was 70%.

Conclusion: IOA for both reporting formats is in the substantial agreement range, with no significant difference in κ values. BSRTC shows a slightly higher composite proportion of agreement than TCRSL. Thus increased diagnostic categories of BSRTC have no significant adverse impact on the IOA in our setting.

(Key words: Thyroid cytology, Fine-needle aspiration, Bethesda Thyroid classification, BSRTC, Thy classification, Inter-observer agreement)

Department of Pathology, Faculty of Medicine, University of Colombo, Kynsey Road, Colombo 8, Sri Lanka

Correspondence: Dr A.A.H. Priyani, Department of Pathology, Faculty of Medicine, University of Colombo, Kynsey Road, Colombo 8, Sri Lanka.
E mail: priyaniaah@yahoo.com

Introduction

Thyroid enlargement is a frequent clinical presentation of thyroid disease with a reported prevalence of 7 % in the general population (1). Incidence of thyroid cancer is approximately 213,000 new cases per year worldwide, and accounts for 5.6% of all cancers and for 8.1% of cancers in females in Sri Lanka (2,3). It is important to differentiate neoplastic from non neoplastic thyroid enlargement, to provide optimum healthcare to affected patients. Thyroid fine needle aspiration cytology (T-FNAC) is a widely accepted diagnostic tool for pre-operative assessment of solitary thyroid nodules, diffuse and multi-nodular goiters with specificity ranging between 67%-100% and sensitivity ranging between 77.7%-93% (4-8). T-FNAC diagnostic value is enhanced when combined with radiological investigations and thyroid function assays (6). Its simplicity, cost effectiveness, good patient compliance and the minimum risk of complications enhances its value. T-FNAC is a forerunner in investigating thyroid nodules/enlargements in our setting as well, resulting in a heavy thyroid cytology burden. It has possibly reduced the number of patients subjected to thyroidectomies and expensive isotope scans impacting our national healthcare cost.

Sri Lankan pathologists have performed and reported T-FNAC for the past two decades (9). The reporting format for thyroid cytology in Sri Lanka (TCRSL) was compiled in 2007, based on the original 'Thy' classification of the British Thyroid Association and the Royal College of Pathologists of the United Kingdom (10,11). Without doubt TCRSL has made thyroid

cytology reporting more uniform in our setting. However, TCRSL does not include clinical management guidelines, which is now felt as one of its major deficiencies (12).

The Bethesda thyroid study group introduced 'The Bethesda System for Reporting Thyroid Cytology (BSRTC)' in the same year (13). BSRTC has standardized reporting and the management by linking diagnostic categories with clinical management guidelines (13). As it is felt to offer a clearer platform of communication among cytopathologists and clinicians, BSRTC is adopted by many countries worldwide. BSRTC is also credited with increasing the inter-observer agreement (IOA) among diagnostic categories, improving the predictive value of diagnoses (14). However, it has more diagnostic categories than TCRSL (6 vs. 5) with a possible impact on the IOA.

Adopting a uniform thyroid cytology reporting format will allow easy comparison of data across borders, achieving the best use of cytology reporting. The Sri Lankan cytopathologists and the clinicians are yet relatively unfamiliar with BSRTC and the diagnostic outcome, compared to TCRSL. Therefore, IOA of BSRTC and TCRSL was compared with the view of providing scientific evidence to evaluate the appropriateness of adopting BSRTC in Sri Lanka.

Materials and Methods

Ethical approval was obtained from the ethics committee of the Faculty of Medicine, University of Colombo. The study sample was selected retrospectively from T-FNAC performed

and reported by the department from January 2008-May 2012. The inclusion criteria included availability of all original smears in good condition and clinical information (age, sex, type of thyroid enlargement, sites of aspiration with clinical/radiological findings, thyroid function status). Stratified random sampling was used to select 100 T-FNAC fulfilling the inclusion criteria. This sampling method was used in order to include all diagnostic categories in the sample, minimizing the effect of prevalence. The serial number and clinical information obtained from clinical records were entered into a data record form as clinical variables. Individual diagnostic categorization was based on TCRSL followed by BSRTC. The evaluation was performed on separate occasions by two pathologists, each functioning independently.

The two pathologists were provided with the case histories and clinical findings, whilst being blinded to the previous diagnosis. The data record form was filled with the diagnosis based on one reporting format at any given time (TCRSL or BSRTC) by a pathologist (observer) after evaluating all the available smears of each case. Data was tabulated in SPSS version 20. The proportions of agreement and the composite agreements were calculated in all categories of both reporting formats. The IOA was analysed by calculating the Cohen kappa co-efficient for both TCRSL and BSRTC. Both unweighted kappa values and linear weighted kappa values were calculated as a measure of discrepancy between the diagnoses in TCRSL and BSRTC. Values of the Cohen kappa were interpreted as follows: 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 0.81 to 1; as almost perfect agreement (15).

Results

The 100 T-FNAC included 9 males and 91 females between 15-76 years of age. Majority had multinodular goiters (43%) while 29% had solitary thyroid nodules. 23% had diffuse gland enlargement, 3% a single lobe enlargement and 2% with minimal/no macroscopically visible thyroid enlargement.

The diagnostic categories of TCRSL and BSRTC and the frequency distribution are depicted in Tables 1, 2 and 3. Both observers were in complete agreement in diagnosing inadequate smears in 7 out of 16 (43%) in TCRSL and 6 out of 12 (50%) non-diagnostic or unsatisfactory smears in BSRTC. Observer 1 had given 16 inadequate/non diagnostic diagnosis in TCRSL and 12 in BSRTC, a higher frequency compared to observer 2. For the diagnosis of benign category/lesion the frequency of agreement was 52/78 (66%) by both observers in TCRSL, while it was 54/76 (71%) in BSRTC (Tables 4 and 5). The (Atypia of undetermined significance/Follicular lesion of undetermined significance) AUS/FLUS category of BSRTC showed the highest variability in the frequency of agreement as the two observers agreed only on 2/19 cases (10%), in AUS/FLUS diagnoses. The frequency of agreement for suspicious for follicular neoplasm/follicular neoplasm category in TCRSL was 4/24 (16%), where as it was 2/10 (20%) for BSRTC. The categories suspicious for malignancy and malignant had frequencies of agreement of 0 and 5 (71%) in TCRSL and 2 (33%) and 4 (57%) in BSRTC respectively.

Table 1 - Categories used for thyroid cytology reporting in TCRSL, BSRTC and Modified “Thy’ classification 2009 by British Thyroid Association

BSRTC	TCRSL	Modified “Thy’ classification 2009
Category 1 Non diagnostic /unsatisfactory	Thy 1 Inadequate	Thy 1 Non-diagnostic for cytological diagnosis Thy 1c Non-diagnostic for cytological diagnosis, Cystic lesion
Category 2 Benign	Thy 2 Benign	Thy 2 Non-neoplastic Thy 2 c Non-neoplastic, cystic lesion
Category 3 Atypia of undetermined significance (AUS) /Follicular lesion of undetermined significance (FLUS)	Thy 3 Follicular proliferation	Thy 3 a Neoplasm Possible atypia/non diagnostic.
Category 4 Suspicious for a follicular neoplasm/Follicular neoplasm		Thy 3 f Neoplasm possible - suggesting follicular neoplasm
Category 5 Suspicious for malignancy	Thy 4 Suspicious for malignancy	Thy 4 Suspicious of malignancy
Category 6 Malignant	Thy 5 Malignant	Thy 5 Malignant

Table 02- Comparison of TCRSL diagnoses between the two observers

O B S E R V E R 1	OBSERVER 2					
		THY 1	THY 2	THY 3	THY 4	THY 5
	THY 1	7	9	0	0	0
	THY 2	0	52	4	1	0
	THY 3	0	12	4	1	0
	THY 4	0	0	3	0	1
	THY 5	0	0	0	1	5

Table 03 - Comparison of BSRTC diagnoses between the two observers

O B S E R V E R 1	OBSERVER 2						
		Category	Category	Category	Category	Category	Category
		01	02	03	04	05	06
	Category 1	6	6	0	0	0	0
	Category 2	0	54	4	0	0	0
	Category 3	0	10	2	0	0	1
	Category 4	0	2	2	2	1	0
	Category 5	0	0	0	2	2	1
Category 6	0	0	0	1	0	4	

Table 04 – Distribution of frequency of agreement among TCRSL diagnoses

TCRSL Category	Maximum Possible	Chance Expected	Frequency of agreement
THY 1	7	1.12	7
THY 2	57	41.61	52
THY 3	11	1.87	4
THY 4	3	0.12	0
THY 5	6	0.36	5
Total	84	45.07	68

Table 05 – Distribution of frequency of agreement among BSRTC diagnoses

BSRTC Category	Maximum Possible	Chance Expected	Frequency of agreement
Category 1	6	0.72	6
Category 2	58	41.76	54
Category 3	8	1.04	2
Category 4	5	0.35	2
Category 5	3	0.15	2
Category 6	5	0.3	4
Total	85	44.32	70

TCRSL had a composite proportion of agreement of 0.68 with 32 discordant cytological diagnoses. BSRTC had a composite proportion of agreement of 0.7 with 30 discordant diagnoses. The calculated unweighted kappa value for

TCRSL was 0.4173 (0.95 C.I. – 0.2508-0.5838). For BSRTC the unweighted kappa value was 0.4612 (0.95C.I.–0.563-0.796). Both unweighted kappa values represented moderate agreement with overlapping confidence intervals. Linear

weighted kappa value for TCRSL was 0.6138 (.95 C.I. - 0.479-0.747) while it was 0.7855 (0.95C.I.- 0.563-0.796) for BSRTC. Both kappa values were in the substantial agreement range with overlapping 0.95 confidence intervals while BSRTC kappa value was higher than that obtained for TCRSL.

Discussion

Both TCRSL and BSRTC had unweighted kappa values in the moderate agreement range. The linear weighted kappa values were in the substantial agreement range with overlapping 0.95 confidence intervals. Even though slightly higher kappa values were observed in BSRTC compared to TCRSL, there was no significant difference indicating that the number of diagnostic categories in BSRTC had no significant impact on the IOA. Though BSRTC is also implicated in increasing the IOA improving the predictive value of diagnoses, this was not significant in this study. The robust, reproducible nature of the 'Thy' classification on which TCRSL was based, could have contributed towards the similar, statistically insignificant IOA obtained for both formats. The familiarity of both observers with the TCRSL format also could also have contributed. It is also worthy to note that there was no significant difference in the agreement between the two formats, though both observers had to familiarize themselves with BSRTC before embarking on the study. The kappa values of BSRTC were also slightly better than what TCRSL obtained. Therefore transition from TCRSL to BRSTC does not appear to be overly difficult in our setting, whilst maintaining the IOA.

The lesser number of non-diagnostic diagnoses obtained in BSRTC (16%) compared to TCRSL (12%) depicts the less stringent BSRTC adequacy criteria. Therefore using BSRTC may prevent the additional health care cost of a repeat FNAC on the health care system and the burden of undergoing a repeat FNAC for a patient, having only a low risk of malignancy of 1- 4% (13).. The findings of Vickie et al reiterate that a repeat non-diagnostic thyroid aspirate does not confer a lower malignant risk (16). The least proportion of agreement among all diagnostic categories were seen in the AUS/FLUS (10%) category of BSRTC. This category has always been a diagnostic grey zone with high variability among cytopathologists.

Many studies have shown similar results, with higher interobserver variability in AUS/FLUS category compared with other categories (17,18). The 'Thy3' category of TCRSL represents AUS/FLUS (category 3) and suspicious for a follicular neoplasm/follicular neoplasm (category 4) categories of BSRTC (7) (Table 1). Therefore, the significant risk of malignancy of Thy 3 category of TCRSL may not be fully appreciated during the clinical decision making process (3,7). BSRTC is placed a step ahead of TCRSL in this context because of the clinical management guidelines AUS/FLUS and suspicious for a follicular neoplasm/follicular neoplasm categories are linked with.

The higher percentage of agreement seen in the diagnosis of Thy 5/malignant category in TCRSL (71%), in comparison to BSRTC category 6 (57%) was an unexpected finding. This could be a reflection of the two observers'

familiarity with the TCRSL. It could also be due to a possible increased in the specificity of category 6 in BSRTC in view of the additional diagnostic options of AUS/FLUS and suspicious for follicular neoplasm/follicular neoplasm categories. However, further comment is not possible without correlating the cytological findings of Thy 5/category 6 with the gold standard of histological diagnosis.

Since the original 'Thy' diagnostic format was published in 2002 by the British Thyroid Association of the United Kingdom and adopted by Sri Lanka in 2007, it has been modified further in 2009 (11) (Table 1). This modified 'Thy' classification addresses the need for uniformity. The 'Thy3' category is now refined as 'Thy3a' and 'Thy 3f'. We in Sri Lanka however have not incorporated these modifications to our national guidelines. Even the current unmodified Thy classification based on TCRSL did not show a significant difference among the kappa values compared to BSRTC. However, it should be kept in mind that using kappa has its own limitations.

T-FNAC, in addition to having a high negative predictive value is documented to have a sensitivity and specificity ranging from 77.7%-93% and 67%-100% respectively (4-8). In order to maintain a high level of sensitivity and specificity while optimizing clinical care, it is necessary to have a robust and reproducible terminology for reporting thyroid cytology. BSRTC is widely adopted in the U.S.A and in many other developed countries due to its high sensitivity, high negative predictive value and also for its superiority in guiding the clinical management of patients (19). Thus, BSRTC with all its merits is an option. This is especially in

view of its linkage with clinical management guidelines; interobserver agreement and the easy transition from TCRSL to BSRTC as shown in the present study.

Adopting the 2009 modified "Thy" classification by the British Thyroid Association and Royal College of Pathologists (11), and recommending the clinicians to link up Thy category with the management guidelines documented by the British Thyroid Association and the Royal College of Physicians (20) is another option. This option is further supported by the robust and reproducible nature of TCRSL based on the original 'Thy' classification as demonstrated by this study. Further study correlating the cytological findings of TCRSL and BSRTC with the gold standard of histological diagnosis may provide additional valuable information impacting this decision.

The utility of evidence based medicine tools are not commonly seen in pathology while it is the governing trend in clinical medicine. While systematic reviews, meta-analysis and randomised control trials are accepted as the best available high quality evidence in clinical medicine, most attempts to develop pathology guidelines and recommendations have been based on observational studies or subjective views (21). The availability of primary studies representing different settings is valuable in getting high quality evidence in the field of pathology as well. We hope the findings of this study would initiate and promote further study supporting evidence based development of guidelines and recommendations for thyroid cytology reporting in Sri Lanka.

References

- Gaitan E, Nelson NC, Poole GV: Endemic goiter and endemic thyroid disorders: *World Journal of Surgery* 1991;15: 205-206.
- Stewart BW, Kleihues P. *World Cancer Report*. Lyon: 2003. 257- 260.
- Cancer Incidence Data: Sri Lanka Year 2001-05. Cancer Registry. 7th Publication. Sri Lanka: National Cancer Control Programme; 2009.
- Priyani AAH, Illeperuma A, Fernando R: Validity of fine needle aspiration cytology in diagnosing thyroid gland neoplasms: *Journal of Diagnostic Pathology* 2011(6);1: 34-41.
- Beneragama DH, Jayasuriya W, Samarawickrama R I U, Rupasinghe RATK, Piyasena MGY, Dayarathna UJP, et al. Validity of fine needle aspiration cytology in the diagnosis of thyroid diseases. *Journal of Diagnostic Pathology*: 2006/2007;5: 19-29.
- Kaur K, Sonkhya N, Bapna AS, Mital P. A comparative study of fine need aspiration cytology, ultrasonography and radionuclide scan on the management of solitary thyroid nodule: a prospective analysis of fifty cases. *Indian Journal of Otolaryngology and Head and Neck Surgery* April – June 2002; 54 (2): 96-101.
- Tabaqchali MA, Hanson JM, Johnson SJ, Wadehra V, Lennard TWJ, Proud G. Thyroid aspiration cytology in Newcastle: a six year cytology/histology correlation study. *Annals of Royal College of Surgeons of England* 2000; 82: 149-155.
- Bakhos R, Selvaggi S M, DeJongS, Gordon D L, Pitale S U, Herrmann M, et al Fine Needle Aspiration of the Thyroid : Rate and Causes of Cytolathologic Discordance. *Diagnostic Cytopathology* 2000; 23: 233-237.
- Kumarasinghe P. Morphologic assessment of thyroid disease by cytology and pattern of thyroid enlargement in Sri Lanka. *Journal of the Ceylon College of Physicians* 1998; 31(1&2):31-40.
- Lokuhetty D. Guidelines for reporting cytology, Introduction, Cytology of Breast, Thyroid and Body fluids. In: CPSL National Guidelines. Colombo: Ministry of Healthcare and Nutrition. Sri Lanka 2007:63-68.
- British Thyroid Association, Royal College of Pathologists. Guidelines on the reporting of thyroid cytology specimens. London. RCP 2009. Available at: <http://www.rcpath.org/Resources/RCPPath/Migrated%20Resources/Documents/G/g089guidance on the reporting of thyroid cytology final.pdf>. (Accessed 18.06.2013).
- Lokuhetty MDS: Beyond the horizon of current thyroid cytology reporting in Sri Lanka. *Journal of Diagnostic Pathology* 2011(6); 1: 6-11.
- Cibas ES, Ali SZ. The Bethesda System for Reporting Thyroid Cytopathology. *American Journal of Clinical Pathology* 2009; 132: 658-665.
- Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda System for Reporting Thyroid Cytopathology: A Meta-Analysis. *Acta Cytologica* 2012;56: 333-339.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159-174.
- Jo YV, Vanderlaan PA, Marqusee E, Krane FJ: Repeatedly Non-diagnostic Thyroid Fine-Needle Aspirations Do Not Modify Malignency Risk. *Acta Cytologica* 2011; 55: 539-543.

17. 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 Cytologica* 2011; 55: 512-517.
18. Osion MT, Clark DP, Erozan YS, Ali SZ: Spectrum of Risk of Malignancy in Subcategories of 'Atypia of Undetermined Significance'. *Acta Cytologica* 2011; 55: 518-525.
19. Cibas ES, Bibbo M: Thyroid FNA: Challenges and Opportunities. *Acta Cytologica* 2011; 55:489-491.
20. British Thyroid Association, Royal College of Physicians. Guidelines for the management of thyroid cancer. 2nd edn. Report of the Thyroid Cancer Guidelines Update Group. London: RCP 2007.
Available at: www.rcplondon.ac.uk/pubs/contents/5e0be7d7-0726-48af-a510-5cc5b9ec574d.pdf (accessed 24.05.09)
21. Gaffar BA. Systematic Review, Meta-Analysis and Randomized Control Trials in Cytopathology. *Acta Cytologica* 2012; 56: 221-227.