ORIGINAL RESEARCH PAPER

Pathology

ACCURACY OF THYROID FINE NEEDLE ASPIRATION AND CYTO-HISTOLOGIC CORRELATION: A 5 YEAR-EXPERIENCE

KEY WORDS: fine needle aspiration; Bethesda system; thyroid cytology

Monique Freire Santana* Departamento de Ensino e Pesquisa, Fundação Centro de Controle de Oncologia do Amazonas *Corresponding Author

Luiz Carlos de Lima Ferreira Departamento de Patologia e Medicina Legal, Hospital Universitário Getulio Vargas

ABSTRACT

The objective of this study was to describe the accuracy of the thyroid FNA and a and cyto-histologic correlation. This is a retrospective review, of diagnostic accuracy, using the reports of FNA samples and histopathology from surgical specimens obtained from thyroidectomies as a gold standard. 3,811 cytological reports from 3,364 patients were analyzed. The median age 51 ± 13.3 years (18-99) and male to female ratio 1:15.5 (female: n=3.581, 94%, median age 51) (male: n=230, 6%, 6%, median age 52.2). In 1,721 satisfactory smears, 1,359 (35.7%) were benign, 30 (0.8%) atypical/follicular lesions of undetermined significance, 182 (4.8%) follicular neoplasm/suspicious for follicular neoplasm, 92 (2.4%) suspicious for malignancy and 58 (1.5%) were malignant, with malignancy rates of 16%, 11%, 25%, 23%, 58% and 88%, respectively. The accuracy FNA was 66.2% with 68.2% sensitivity and 63.9% specificity. This study is the biggest describe analysis of thyroid FNA samples performed in Brazil. Our accuracy was similar to other prospective studies and meta-analysis.

INTRODUCTION

Thyroid nodules are very common in clinical practice. The asymptomatic nodules detected by imaging or in thyroid surgery have an estimated detection rate of up to 6,594/100 ultrasounds (Uppal et al., 2015). In the last 20 years, the thyroid ultrasonographic examination has become the most valued tool in screening these nodules(Alexander & Cooper, 2013). The use of fine needle aspiration (FNA) reduces unnecessary surgery in benign diseases and provides proper management in malignant nodules. It is critical that communication must be succinct and unambiguous between the cytopathologist writing the thyroid report and the clinician or surgeon providing care. "The Bethesda System for Reporting Thyroid Cytopathology" (BSRTC) divides results into 6 categories: "Nondiagnostic" (I), "Benign" (II), "Atypia of Undetermined Significance/Follicular Lesion of Undetermined Significance" (III), "Follicular Neoplasm/ Suspicious for a Follicular Neoplasm" (IV), "Suspicious for Malignancy" (V) and "Malignant" (VI) (Ali & Cibas, 2018). The accuracy of thyroid FNA is well established. In recent meta-analysis, with 25,445 thyroid FNAs, describe sensitivity, specificity and accuracy as 97, 50.7 and 68.8%, respectively; the positive predictive and negative predictive values were 55.9 and 96.3% respectively, and the rates of false negatives and false positives were 3 and 0.5% (Bongiovanni, Spitale, Faquin, Mazzucchelli, & Baloch, 2012). The objective of this study is to describe the accuracy of the thyroid FNA, compare to histopathology as the gold standard.

Materials and Methods: This is diagnostic accuracy study, using the reports of FNA samples and histopathology from surgical specimens obtained from thyroidectomies as a gold standard. The study period extended from January 2011 to December 2015, in three laboratories (one oncologic care center and the other a laboratory of the federal school medicine) and a private one. Approximately twelve physicians from different specialties (head and neck surgeons, endocrinologists, radiologists, and pathology trainees) performed the procedures, guided by ultrasound, and eight pathologists performed the readings from slides and tissue diagnoses. All of the cytological specimens were stained with Papanicolaou, Giemsa or hematoxylin-eosin, and the surgical specimens were stained with hematoxylin-eosin. Excluded reports: illegible, no-BSRTC and with ancillary studies (immunocytochemical or immunohistochemical studies). The database and analyses with descriptive statistics were performed in Epi Info™ and accuracy in Open Epi®. To compare two means was used Wilcoxon test or ANOVA for non-parametrical and parametrical dates, respectively. For accuracy, results of BSRTC I and BSRTC III were excluded, BSRTC II and BSRTC IV with no malignancy histology were considered true negative, BSRTC V and VI cases with malignant histology were considered true positives, and BSRTC V or VI with no malignant histology were considered false positives.

The Ethical Committee of the Federal University of Amazonas approved this research following Brazilian laws relating to research with human subjects. The authors do not have disclosures conflicts.

Results: 3,811 cytological reports from 3,364 patients were analyzed. The median age 51 ± 13.3 years (18-99) and male to female ratio 1:15.5 (female: n=3.581, 94%, median age 51) (male: n=230, 6%, 6%, median age 52.2), without statistical difference (p=0,17). The number of slides in reports was in the range of 1–18 (median 3,0); the mean 3.88 and 4.36 in unsatisfactory and satisfactory smears, respectively, with statistical difference ($p < 0,001$). 2,090 (58.8%) FNAs were unsatisfactory, the majority in private laboratory (n=1,477, 70.7%). The principal cause of unsatisfactory smear was low cellularity (n=1,499; 71.7%), hemorrhagic smears (n=354; 16.9%), cyst fluid in 169 (8.1%), and other artifacts (air-drying artifact, for example) in 68 (3.3%) cases (Table 1).

Histopathology correlation was done in 250 cases; in three, it was not possible to define the nature of the lesion; 66 (26.7%) had a malignant diagnosis. The malignancy rates are as follows: BSRTCI 16%, BSRTCII 11%, BSRTCIII 25%, BSRTCIV 23%, BSRTCV 58%, BSRTCVI 88% (Table 1). The accuracy of FNA cytology for 157 cases with comparable cyto-histology diagnosis (Bethesda categories II, IV, V and VI) was 66.2% with 68.2% sensitivity and 63.9% specificity (Table 2).

DISCUSSION

A large number of unsatisfactory smears (54.8%) is unusual but unfortunately not surprising. In literature, the rates vary between 5% to 13.3% (Table 3). The FNAs of this study were performed only with an ultrasonographic examination, without the use of rapid on-site evaluation of fine needle aspiration (ROSE), a method able to increase smears adequability, lowering significantly inadequacy rate for thyroid FNA (6% with ROSE, 17% without ROSE, $p<0.0001$) (Shield, Cosier, Ellerby, Gartrell, & Papadimos, 2014). In the

cases that have he morrhagic smears (16.9%), it is impossible to provide correct diagnosis, despite of technical quality of the cytopathologist (Oertel, 2007). The rate of malignancy in category I of the BRSTC has been described in some studies. Coorough *et al*.'s (2013) identify 12% of thyroid cancer in nondiagnostic FNA (BSRTC I, 259/4.286, 6%), when compare with diagnosis FNA (5%, p<0.001). Park *et al*. (2014) described 35.3% of malignancy in category I. In our study, we found 16% of malignancy, higher than Coorough and lower than Park.

The considerable variability in our study was observed in category II (33.5% in Lab 3 and 44.8% in Lab 1). This fact may be attributed to the different population in the laboratories: in Lab 1 and 3, the samples are derived from the general population; in Lab 2 is from the oncology hospital. The difference of laboratories can also explain the variability observed in category I (Lab 1: 48.3% and Lab 3: 56.5%). In category III, the variability observed in our study reflects the low reproducibility in this category. Like the cases published by Mehrotra and Sams (2013) that evaluated the influence of cytomorphologic criteria in intra-observer variability in category III, in 714 FNAs. This confirms the larger variability in category III (11.4% to 18.8%, much higher than our study) and they propose the review of atypical cases to reach a consensus, to improve reports of thyroid FNAs.

Some factors can explain the high variability to the accuracy in the literature (Table 4). Pre-analytical factors, such as the experience of the professional who performs FNA, the quality ultrasonographic image, use of ROSE or liquid-based preparation, and analytical factors (the experience of the cytopathologist and use of ancillary techniques) undoubtedly improve the quality of FNA in some centers. In our study, the FNAs samples evaluated were obtained from the general population; only one from an oncologic center; this can explain our results.

Another factor that can interfere with the accuracy is the percentage of cases with cytohistological correlation. In some studies, the number of cases with histologic diagnosis is greater than 50%. For example, Deniwar et al. (2015) compare 723 FNA with 375 (51%) cases with histology; Garg et al. (2015) available 100 FNA and 60 (60%) with histology. In others, it does not exceed 40%: Tepeoglu et al. (2014) describe 1021 FNA, and 219 (21%) with histopathological reports. Naz et al. (2014) describes 528 FNA and 61 (11%) with histology. The biggest meta-analysis regarding thyroid FNA review results of 25,445 FNA, with only 6,362 (25%) of which involved surgical excision (Bongiovanni et al., 2012). Possibly the selection criteria or possible selection bias can explain the larger variability.

Potential bias in our study includes the great variability in the number of inter-observer (between cytopathologists), variability in FNAs (experience of professional responsible for needle aspiration or ultrasonography, quality of ultrasonography machine), pre-analytical factors, such as stain quality and quality of slides in different laboratories, and pos-analytical factors (report transcription errors, for example). All this bias occurs in the daily clinical practice, which makes the study even closer to the local reality, with their inherent limitations.

CONCLUSIONS

Despite of the FNA is a simple procedure, cost-effective, as an initial screening test that provides diagnosis with good accuracy, our results are not as comparable to those of other studies. In our results, we found a greater number of unsatisfactory FNAs, that may be a reality in other centers. Factors pre-analytical that interfere with adequability of smears, selection bias and variability in nomenclature of cytopathological reports can explain the different values in accuracy and the considerable number of unsatisfactory smears.

Acknowledgments: To Luciana Fujimoto, for scientific contributions for this study. To managers of three laboratories where data collection was performed, Jeconias Câmara, J. A. S. Loureiro, Rosilene Viana, and their support team.

Table 1. Cytologic diagnosis and rate of malignancy in Bethesda category

Cytological diagnosis	Lab 1		Lab 2		Lab 3		Total		Rate of malignancy
	n	%	N	%	n	%	n	%	
I	280	48.3	333	54.1	1,477	56.5	2,090	54.8	16
II	260	44.8	222	36	877	33.5	1,359	35.7	11
III	4	0.7	17	2.8	9	0.3	30	0.8	25
IV	11	1.9	9	1.5	162	6.2	182	4.8	23
V	17	2.9	28	4.5	47	1.8	92	2.4	58
VI	8	1.4	7	1.1	43	1.6	58	1.5	88
Total	580	100	616	100	2,615	100	3,811	100	-

Table 2. Accuracy of fine needle aspiration cytology according to Bethesda System of Reporting Thyroid Cytopathology

	Number of cases	Percentage	95%IC
True negative	46	29.3	
False negative	27	17.2	
False positive	26	16.6	
True positive	58	36.9	
Sensitivity		68.2	57.73 – 77.16
Specificity		63.9	52.35 – 74.02
Positive predictive value (PPV)		69	58.51 – 77.92
Negative predictive value (NPV)		63	51.55 – 73.18
Accuracy		66.2	58.54 – 73.17

Table 3. Comparison of the distribution of Bethesda diagnostic categories of the present study with published studies.

	FNA	I		II		III		IV		V		VI	
	n	n	%	n	%	n	%	n	%	n	%	n	%
Santana; Ferreira	3,811	2,090	54.8	1,359	35.7	30	0.8	182	4.8	92	2.4	58	1.5
(Arul & Masilamani, 2015)	483	24	5	215	44.5	14	2.9	104	21.5	74	15.3	52	10.8
(Arul, Akshatha, & Masilamani, 2016)	603	16	2.7	393	65.2	60	10	64	10.6	32	5.3	38	6.3
(Naz et al., 2014)	528	25	4.7	403	76.3	67	12.7	11	2.1	18	3.4	4	0.8
(Park et al., 2014)	1,730	230	13.3	702	40.6	157	9.1	7	0.4	335	19.3	299	17.3
(Bongiovanni et al., 2012)	25,445	3,271	12.9	15,104	59.4	2,441	9.6	2,571	10.1	680	2.7	1,378	5.4

Table 4. Comparison of the accuracy of the present study with published studies

	n	Sensitivity	Specificity	PPV	NPV	Accuracy
Santana, Ferreira	157	68.2	63.9	69	63	66.2
Bongiovanni et al., 2012	6,362	97	50.7	55.9	96.3	68.8
Tseng et al., 2008	1,064	81	98.7	94.4	95	94.9
Arul & Masilamani, 2015	209	94.4	97.6	98.1	93.2	95.8
Arul et al., 2016	392	64.3	85.1	50.3	88.9	80.3

REFERENCES

1. Alexander, E. K., & Cooper, D. (2013). The Importance , and Important Limitations , of Ultrasound Imaging for Evaluating Thyroid Nodules. JAMA Internal Medicine, 31(2), 1–2. https://doi.org/10.1001/jamainternmed.2013.9245.2
2. Ali, S. Z., & Cibas, E. S. (Eds.). (2018). The Bethesda System for Reporting Thyroid Cytopathology. https://doi.org/10.1007/978-3-319-60570-8
3. Arul, P., Akshatha, C., & Masilamani, S. (2016). A study of malignancy rates in different diagnostic categories of the Bethesda system for reporting thyroid cytopathology: An institutional experience. Biomedical Journal, 38(6), 517–522. https://doi.org/10.1016/j.bj.2015.08.001
4. Arul, P., & Masilamani, S. (2015). A correlative study of solitary thyroid nodules using the bethesda system for reporting thyroid cytopathology. Journal of Cancer Research and Therapeutics, 11(3), 617. https://doi.org/10.4103/0973-1482.157302
5. Bongiovanni, M., Spitale, A., Faquin, W. C., Mazzucchelli, L., & Baloch, Z. W. (2012). The Bethesda system for reporting thyroid cytopathology: A meta-analysis. Acta Cytologica, 56(4), 333–339. https://doi.org/10.1159/000339959
6. Coorough, N., Hudak, K., Jaume, J. C., Buehler, D., Selvaggi, S., Rivas, J., … Chen, H. (2013). Nondiagnostic fine-needle aspirations of the thyroid: Is the risk of malignancy higher? Journal of Surgical Research, 184(2), 746–750. https://doi.org/10.1016/j.jss.2013.02.018
7. Deniwar, A., Hambleton, C., Thethi, T., Moroz, K., & Kandil, E. (2015). Examining the bethesda criteria risk stratification of thyroid nodules. Pathology Research and Practice, 211(5), 345–348. https://doi.org/10.1016/j.prp.2015.02.005
8. Garg, S., Desai, N. J., Mehta, D., & Vaishnav, M. (2015). To establish bethesda system for diagnosis of thyroid nodules on the basis of FNAC with histopathological correlation. Journal of Clinical and Diagnostic Research, 9(12), EC17–EC21. https://doi.org/10.7860/JCDR/2015/14823.6897
9. Mehrotra, S., & Sams, S. (2013). The Role of Cytomorphologic Criteria on Intraobserver Variability Rate of Indeterminate Thyroid Fine Needle Aspiration Cytology. Journal of the American Society of Cytopathology, 2(1), S76–S77. https://doi.org/10.1016/j.jasc.2013.08.206
10. Naz, S., Hashmi, A. A., Khurshid, A., Faridi, N., Edhi, M. M., Kamal, A., & Khan, M. (2014). Diagnostic accuracy of Bethesda system for reporting thyroid cytopathology: an institutional perspective. International Archives of Medicine, 7(1), 46. https://doi.org/10.1186/1755-7682-7-46
11. Oertel, Y. C. (2007). Fine-Needle Aspiration of the Thyroid: Technique and Terminology. Endocrinology and Metabolism Clinics of North America, 36(3), 737–751. https://doi.org/10.1016/j.ecl.2007.05.001
12. Park, J. H., Yoon, S. O., Son, E. J., Kim, H. M., Nahm, J. H., & Hong, S. W. (2014). Incidence and malignancy rates of diagnoses in the bethesda system for reporting thyroid aspiration cytology: An institutional experience. Korean Journal of Pathology, 48(2), 133–139. https://doi.org/10.4132/KoreanJPathol.2014.48.2.133
13. Shield, P. W., Cosier, J., Ellerby, G., Gartrell, M., & Papadimos, D. (2014). Rapid on-site evaluation of fine needle aspiration specimens by cytology scientists: A review of 3032 specimens. Cytopathology, 25(5), 322–329. https:// doi.org/10.1111/cyt.12157
14. Tepeoğlu, M., Bilezikçi, B., & Bayraktar, S. G. (2014). A histological assessment of the Bethesda system for reporting thyroid cytopathology (2010) abnormal categories: A series of 219 consecutive cases. Cytopathology, 25(1), 39–44. https://doi.org/10.1111/cyt.12051
15. Tseng, C.-E., Wei, C.-K., Kuo, C.-S., Yan, S.-T., Chen, P.-F., Lien, W.-C., … Tseng, Y.-H. (2008). Fine Needle Aspiration Cytology of Thyroid Nodules: Evaluation of Diagnostic Accuracy. Tzu Chi Medical Journal, 20(4), 296–303. https://doi.org/10.1016/S1016-3190(08)60054-3
16. Uppal, A., White, M. G., Nagar, S., Aschebrook-Kilfoy, B., Chang, P. J., Peter, A., … Grogan, R. H. (2015). Benign and Malignant Thyroid Incidentalomas are Rare in Routine Clinical Practice: A Review of 97,908 Imaging Studies. Cancer Epidemiology, Biomarkers & Prevention, 24(9), 1327–1331. https://doi.org/10.1158/1055-9965.EPI-15-0292