Annals of Diagnostic Pathology 41 (2019) 112–115

Contents lists available at ScienceDirect

Annals of Diagnostic Pathology

journal homepage: www.elsevier.com/locate/anndiagpath

Cytological-Pathologic Correlation

Diagnosing atypia of undetermined significance in thyroid fine needle aspiration samples using nuclear scoring

Aysegul Aksoy Altinboga[a,*], Canan Altunkaya[b], Hilal Ahsen[c], Berrak Gumuskaya[a], Oya Topaloglu[d], Serap Ulusoy[e], İbrahim Kilinc[e], Bekir Cakir[d], Fazli Erdogan[a]

[a] Ankara Yildirim Beyazit University, Department of Pathology, 06800 Ankara, Turkey
[b] Kirikkale University, Department of Pathology, 71451 Kirikkale, Turkey
[c] Ankara Ataturk Training and Research Hospital, Department of Pathology, 06800 Ankara, Turkey
[d] Ankara Yildirim Beyazit University, Department of Endocrinology and Metabolism, 06800 Ankara, Turkey
[e] Ankara Ataturk Training and Research Hospital, Department of General Surgery, 06800 Ankara, Turkey

ARTICLE INFO

Keywords:
Thyroid
Fine needle aspiration cytology
Atypia of undetermined significance/follicular lesion of undetermined significance
Nuclear groove
Nuclear overlapping
Nuclear scoring

ABSTRACT

Background: Atypia of undetermined significance (AUS) is an indeterminate category in the Bethesda system for reporting thyroid cytopathology. Cytological features described as atypia are not always observed in every case, and it is difficult to determine how the small population of cells with enlarged nuclei, a few grooves, and rare elongated nuclei should be classified. Therefore, there is inter-intra observer variability considering these cell types, even though the cytological criteria are well defined. Therefore, this study aimed to establish a nuclear scoring system to help in the differential diagnosis of AUS.

Methods: Fine needle aspiration (FNA) samples that showed AUS and had surgical follow-up were included in this study. The aspirate was scored for the presence of intanuclear cytoplasmic inclusions, nuclear grooves, overlapping, enlargement, and elongation individually. The total nuclear score for each case was calculated. Statistical analysis of the association between each nuclear feature and the presence of papillary thyroid cancer (PTC) in the surgical specimens was performed. Cut-off points from the total score of these nuclear features were also calculated.

Results: Nuclear grooves and overlapping were more common in malignant cases ($p < 0.001$ and $p = 0.048$, respectively). A cut-off point of ≥ 5.5 for the total score was sensitive and specific for defining malignancy.

Conclusion: The risk of PTC was higher in nodules with more prominent nuclear overlapping or nuclear groove in their FNA samples. In order to achieve a more confident AUS diagnosis, our scoring system can be helpful for thyroid FNA samples.

1. Background

The Bethesda system for reporting thyroid cytopathology (BSRTC) was first prepared using the information from a conference held at the National Institute of Health in 2007 [1]. The diagnostic category of atypia of undetermined significance/follicular lesion of undetermined significance (AUS/FLUS) was recommended for defining atypia as ≤10%. The risk of malignancy (ROM) is estimated be 10–30% in this group in the Bethesda book [1], but is as high as 46% in some reports [2]. AUS/FLUS cases that have nuclear atypia are at higher risk of malignancy than other AUS/FLUS cases [3-7].

When distinct features of papillary carcinoma exist in a large population of cells, the appropriate diagnosis is obvious. In contrast, well-known cytological features are not always present in every case [5]. In daily practice, it is not uncommon to encounter a case with cells that have these features quite scantly or that are not well developed [5].

In this study, we tried to determine the most helpful nuclear features of AUS/FLUS cases, which have variable interobserver differences and poor standardization. Moreover, we calculated the total score for the nuclear criteria and determined a cut-off point for AUS/FLUS diagnosis, in order to make the assessment as objective as possible for research purposes.

2. Methods

FNA samples with AUS/FLUS obtained at the Ataturk Training and

* Corresponding author at: Department of Pathology, Ankara Yildirim Beyazit University, Bilkent, Ankara 06800, Turkey.
E-mail address: aysegulaksoy@hotmail.com (A.A. Altinboga).

https://doi.org/10.1016/j.anndiagpath.2019.06.001

Table 1
Distribution and scoring of each nuclear feature.

	Score 0, n (%)	Score 1, n (%)	Score 2, n (%)
Nuclear overlapping	9 (6.3)	98 (69)	35 (24.7)
Nuclear enlargement	21 (14.8)	81 (57)	40 (28.2)
Nuclear elongation	19 (13.4)	91 (64)	32 (22.6)
Nuclear grooves	73 (51.4)	66 (46.5)	3 (2.1)
Intranuclear cytoplasmic inclusions	134 (94.4)	8 (5.6)	0 (0)

Research Hospital of Ankara Yildirim Beyazit University between 2010 and 2013 were reviewed by using a system review board. Cases that had surgical specimens during follow-up were included in this study. Clinicians performed all aspirates under ultrasound guidance. All slides were air dried and stained with May-Grünwald Giemza.

A total of 142 FNA samples that were classified as AUS/FLUS according to the BSRTC were reevaluated by a cytopathologist who was blinded to the follow-up information. Each nodule was evaluated for the presence of intanuclear cytoplasmic inclusions, linear chromatin ridges (grooves), and nuclear overlapping, enlargement, and elongation (Table 1). Each nuclear feature was evaluated in atypical follicular epithelial cells observed on the slides. In 121 nodules, each nuclear feature was evaluated in at least 100 cells; 50–80 cells were evaluated in the remaining 20 cases, and in these cases, each nuclear assessment score was proportionally calculated. A two-tier scoring system was used for each nuclear feature: **Score of 0**: none of the cells exhibited the nuclear feature, **Score of 1**: 1–40% of atypical cells exhibited the nuclear feature, and **Score of 2**: 41–100% of atypical cells exhibited the nuclear feature (Table 1). After evaluating scores for each nuclear feature, the total score for each case was calculated.

The slide cellularity and presence of colloid were also evaluated for each case.

Follow-up information was obtained from the medical records. Cases of incidental' papillary carcinomas (papillary carcinomas that were ≤ 1 cm in the largest diameter) were excluded.

For statistical analysis and calculations, we used the IBM SPSS Statistics 21.0 (IBM Corp. Release 2012. IBM SPSS Statistics for Windows, Version 21.0, Armonk, NY: IBM Corp.) program for calculating confidence intervals for sensitivity and specificity, and the MEDCALC free statistical calculator.

Categorical variables such as the total score (0–2), intranuclear cytoplasmic inclusions, and grooves were expressed in numbers (n, %).

The distribution of the evaluated variables (e.g., grooves and intranuclear inclusion) according to the presence of papillary thyroid cancer (PTC) was examined with the chi-squared test. The total score according to PTC presence was compared with the Mann-Whitney *U* test. The discrimination power of the total score between the PTC cases and other cases was assessed via ROC curve analysis; the results of the analysis provided the area under the curve and the standard error of this area. The sensitivity, specificity, and positive-negative predictive values for the cut-point, calculated according to the Youden index, were given as the 95% confidence interval. The statistical significance level was set at $p < 0.05$.

This retrospective study was approved by the Ethical Committee of Ankara Yildirim Beyazit University (Reference number: 26379996/155). The need for additional informed consent was waived owing to the retrospective nature of the analysis.

3. Results

In a total of 8962 aspirates, 1075 nodules were identified as AUS/FLUS between 2010 and 2013. Of these, 161 patients with nodules that had repeated AUS/FLUS diagnosis or showed suspicious clinical and/or ultrasonographic features underwent surgery. Nineteen cases that were diagnosed as 'incidental' papillary carcinomas were excluded, and 142 nodules were finally re-evaluated in this study. These 142 nodules were observed in 139 patients. Among 3 patients, 2 different nodules were sampled from 1 patient, and both nodules were diagnosed as AUS/FLUS. Among these 142 nodules, repeated aspirates were obtained from 42 nodules. On considering the results obtained on the repeated aspirates, 25 were non-diagnostic, 10 showed AUS/FLUS, 6 had a benign diagnosis, and 1 had a malignant diagnosis. Thirty-six cases, which did not have repeat aspirates in our hospital, had AUS/FLUS diagnosis on repeat aspiration at a different medical center.

On using surgical specimens, of the 142 nodules, 55 were diagnosed to be malignant (41 conventional papillary carcinoma, 12 follicular variant of papillary carcinoma, 2 solid variant of papillary carcinoma; i.e., the malignant group) and 87 were nonmalignant (62 nodular goiters, 13 lymphocytic thyroiditis cases, 11 follicular adenomas, and 1 granulomatous thyroiditis case; i.e., the non-malignant group). The median age of patients in the malignant group and non-malignant group was 54 years (range: 24–81 years) and 55 years (range: 27–77 years), respectively; there was no significant difference in the age of patients between the malignant and nonmalignant groups ($p > 0.05$). Four cases had been evaluated for the BRAF V600E mutation using the Entrogen BRAF Mutation Analysis Kit (Tarzana, CA USA) via the PCR method. The BRAF V600E mutation was negative in 3 nodules and positive in 1. The 3 BRAF V600E-negative nodules exhibited nodular goiter on histopathologic examination, while the BRAF V600E-positive nodule exhibited lymphocytic thyroiditis. Moreover, there were no significant differences in the cellularity, presence of colloids, nuclear enlargement, and nuclear elongation between the malignant and nonmalignant groups (according to surgical material; Table 2; $p > 0.05$).

In the malignant group, follicular cells exhibited nuclear grooves more commonly (67.3%) compared to the non-malignant group (Table 2; 36.8%; $p < 0.001$). The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for nuclear grooves were 67%, 63%, 54%, and 75%, respectively.

Nuclear overlapping was common in 34.5% ($n = 19$) of patients in the malignant group and in 18.4% ($n = 16$) of patients in the non-malignant group (Table 2); the difference between these two ratios was significant ($p = 0.048$). The sensitivity, specificity, PPV, and NPV for nuclear overlapping were 35%, 82%, 54%, and 66%, respectively.

Table 2
Cytomorphological parameters and their association with the presence of PTC.

	Presence of PTC in surgical material		Test statistic	*P*-value
	Absent, n (%)	Positive, n (%)		
Cellularity			1.443	0.230
Hypocellular	59 (67.8)	31 (56.4)		
Normocellular	28 (32.2)	24 (43.6)		
Colloid			0.102	0.749
Scant or absent	78 (89.7)	51 (92.7)		
Normal	9 (10.3)	4 (7.3)		
Nuclear overlapping			3.905	**0.048**
Absent or rare	71 (81.6)	36 (65.5)		
Common	16 (18.4)	19 (34.5)		
Nuclear enlargement			1.326	0.249
Absent or rare	66 (75.9)	36 (65.5)		
Common	21 (24.1)	19 (34.5)		
Nuclear elongation			0.754	0.385
Absent or rare	70 (80.5)	40 (72.7)		
Common	17 (19.5)	15 (27.3)		
Nuclear grooves			12.541	< **0.001**
Absent	55 (63.2)	18 (32.7)		
Present	32 (36.8)	37 (67.3)		

PTC, papillary thyroid cancer.
Bold designates the P value of statistically significant cytomorphological parameters.

A minority of the cases (1 in the nonmalignant group and 7 in the malignant group) exhibited intranuclear pseudoinclusions. The incidence of intanuclear inclusions was significantly higher in the malignant group compared to the nonmalignant group ($p = 0.006$).

The median total score was calculated as 3 (range: 2–7) in the nonmalignant group and 4 (range: 1–8) in the malignant group, with the total score of the malignant group being higher than that in the nonmalignant group (Z = 2.548 and 0.011, respectively).The median total score was calculated as 5 (range: 1–8)in the conventional papillary carcinoma and 4 (range: 2–7) in the follicular variant of papillary carcinoma. No significant difference was found between conventional papillary carcinoma and follicular variant of papillary carcinoma (Z = 0.830 and 0.407, respectively).

On ROC curve analyses, performed to distinguish between the malignant and non-malignant groups, the area under the curve was calculated as 0.625 ± 0.051 ($p = 0.012$). The sensitivity of the cut-off point ≥5.5, which was calculated by using the Youden index, was 36.4% (95% GA: 23.8%–50.4%); the specificity was 88.5% (95% GA: 79.9%–94.3%), PPV was 66.7% (95% GA: 50.3%–79.8%), and NPV was 68.8% (95% GA: 64.0%–73.2%).

4. Discussion

In this study, we determined that the presence of grooves and nuclear overlapping were the most helpful nuclear feature of AUS/FLUS cases, in which the interobserver differences are variable and standardization is difficult. In addition, the cut-off point of ≥5.5 from the total score of these nuclear features was calculated, using the Youden index, to achieve a more confident AUS/FLUS diagnosis, as it had a high specificity for PTC presence. To the best of our knowledge, this is the first study to evaluate nuclear scoring in thyroid FNA cases.

Although there is standard terminology for reporting the results of thyroid FNA, it is hard to calculate the ROM of gray zone lesions. AUS/FLUS is a type of gray zone lesions. However, in the literature, there is heterogeneity in the ROM of AUS/FLUS, and some studies demonstrated that this risk is underestimated [2,8-11]. The risk of malignancy varied, reported to be 46% and 39% in two studies [2,12]. The AUS/FLUS category comprises a heterogeneous group of lesions defined differently in the BSRTC, which could be one of the reasons for this heterogeneity.

Several studies were planned to standardize this category, which has interobserver variability. Mathur et al. divided the AUS in four different subgroups as presence of nuclear atypia; AUS-N, focal microfollicular proliferation; AUS-F, focal Hürthle cell proliferation AUS-HC, and other; AUS-O and reported the malignancy risks for these subgroups as 54%, 39%, 19%, and 26%, respectively [12]. In their study, only AUS-H and AUS-O had rates that were similar to those given in the BSRTC. In another study, even when the features of papillary carcinoma were focal, their presence was strongly associated with the follicular variant of papillary carcinoma [13]. In contrast, in another study, although minimal findings of papillary carcinoma were strongly associated with papillary carcinoma at resection (in > 50% of the cases), there was no strong correlation with the follicular variant of papillary carcinoma [14].

In the current study, among the nuclear criteria, nuclear overlapping and nuclear groove were significantly more common in malignant cases. When the presence of grooves is considered, the sensitivity and specificity were intermediately high (67% and 63%, respectively). In contrast, the sensitivity of nuclear overlapping was low (35%) while the specificity was high (82%). These findings indicate that if AUS/FLUS is observed in a thyroid FNA sample, the presence of grooves and/or the presence of nuclear overlapping can be more reliable criteria for atypia. Therefore, a malignancy may be suspected when nuclear grooves and overlapping are prominent in a given nodule with otherwise diagnostic criteria for AUS, which also has suspicious clinical and ultrasound features. Hence, surgery instead of follow-up may be an option for such cases.

There is also interpreter variability when reporting thyroid cytology, with conspicuous differences in the rates of cytopathologic diagnoses for AUS. Interobserver disagreement occurs in 11–34% of cases [4,14-20]. In the present study, all the AUS cases were evaluated and scored by only one cytopathologist. Therefore, interobserver variability cannot be calculated, and this is one of the limitations of this study.

The utility of BRAF V600E-mutation testing in thyroid FNA samples is controversial. In some studies, it was proved an effective diagnostic tool for thyroid FNA [21], while it had limited value in others [22]. False-negative and false-positive BRAF V600E mutation test results were also reported in thyroid FNA samples [23]. In AUS/FLUS cases, the sensitivity and specificity of BRAF V600E analyses was found to be 40.1% and 99.5%, respectively [24]. As BRAF V600E mutation analyses were performed in a limited number of cases in the current study, statistical evaluation is not possible. However, even among 4 cases of AUS/FLUS, 1 was found to be false positive.

In the present study, the specificity of the calculated cut-off point (≥5.5) was considerably high (88.5%). In daily practice, the use of a scoring system for each case is time consuming and cumbersome. However, such scoring systems can be used for cases in which the differential diagnosis between AUS and suspected malignancy is difficult. In these cases with intra- and inter-observer variability, using a scoring system can justify the diagnosis and improve reproducibility. However, interobserver variability could not be calculated in this study as the evaluation and scoring of the cases were performed by only one cytopathologist, which is one of the limitations of this study.

5. Conclusions

According to the results of our study, the presence of prominent nuclear grooves and nuclear overlapping indicate a higher rate of PTC. Therefore, when the differential diagnosis is AUS or suspected malignancy when FNA samples show these nuclear features, a suspected malignancy may be more likely. In addition, the total calculated score of ≥5.5 had a high specificity for PTC. Thus, in cases with difficulty in establishing the differential diagnoses between AUS and suspected malignancy, observing the prominent nuclear grooves and overlapping, and even using a nuclear scoring system, may have additional benefits.

Funding support

None.

Declaration of Competing Interest

The authors declare that they have no conflict of interests.

Acknowledgments

We would like to thank Editage (www.editage.com) for English language editing.

References

[1] Krane JF, Nayar R, Renshaw AA. Atypia of undetermined significance/follicular lesion of undetermined significance. In: Ali SZ, Cibas E, editors. The Bethesda system for reporting thyroid cytopathology: Definitions, criteria, and explanatory notes. 2nd ed.New York: Springer; 2018. p. 49–70.
[2] VanderLaan PA, Marqusee E, Krane JF. Clinical outcome for atypia of undetermined significance in thyroid fine-needle aspirations: should repeated FNA be the preferred initial approach? Am J Clin Pathol 2011;135:770–5.
[3] Ohori NP, Schoedel KE. Variability in the atypia of undetermined significance/follicular lesion of undetermined significance diagnosis in the Bethesda system for reporting thyroid cytopathology: sources and recommendations. Acta Cytol 2011;55:492–8.
[4] Renshaw AA. Focal features of papillary carcinoma of the thyroid in fine needle aspiration material are strongly associated with papillary carcinoma at resection. Am J Clin Pathol 2002;118:208–10.

[5] Renshaw AA. How closely do thyroid fine needle aspirates need to be screened? Diagn Cytopathol 2002;27:259–60.
[6] Renshaw AA. Accuracy of thyroid fine-needle aspiration using receiver operator characteristic curves. Am J Clin Pathol 2001;116:477–82.
[7] Wu HH, Jones JN, Grzybicki DM, Elsheikh TM. Sensitive cytologic criteria for the identification of follicular variant of papillary thyroid carcinoma in fine-needle aspiration biopsy. Diagn Cytopathol 2003;29:262–6.
[8] Ho AS, Sarti EE, Jain KS, et al. Malignancy rate in thyroid nodules classified as Bethesda category III (AUS/FLUS). Thyroid 2014;24:832–9.
[9] Çuhaci N, Arpaci D, Üçler R, et al. Malignancy rate of thyroid nodules defined as follicular lesion of undetermined significance and atypia of undetermined significance in thyroid cytopathology and its relation with ultrasonographic features. Endocr Pathol 2014;25:248–56.
[10] Onder S, Firat P, Ates D. The Bethesda system for reporting thyroid cytopathology: an institutional experience of the outcome of indeterminate categories. Cytopathology 2014;25:177–84.
[11] Choi YJ, Baek JH, Ha EJ, et al. Differences in risk of malignancy and management recommendations in subcategories of thyroid nodules with atypia of undetermined significance or follicular lesion of undetermined significance: the role of ultrasound-guided core-needle biopsy. Thyroid 2014;24:494–501.
[12] Mathur A, Najafian A, Schneider EB, Zeiger MA, Olson MT. Malignancy risk and reproducibility associated with atypia of undetermined significance on thyroid cytology. Surgery 2014;156:1471–6.
[13] Logani S, Gupta P, LiVolsi VA, Mandel S, Baloch ZW. Thyroid nodules with FNA cytology suspicious for follicular variant of papillary thyroid carcinoma: follow-up and management. Diagn Cytopathol 2000;23:380–5.
[14] Bajaj J, Morgenstern N, Sugrue C, Wasserman J, Wasserman P. Clinical impact of second opinion in thyroid fine needle aspiration cytology (FNAC): a study of 922 interinstitutional consultations. Diagn Cytopathol 2012;40:422–9.
[15] Olson MT, Boonyaarunnate T, Aragon Han P, Umbricht CB, Ali SZ, Zeiger MA. A tertiary center's experience with second review of 3885 thyroid cytopathology specimens. J Clin Endocrinol Metab 2013;98:1450–7.
[16] 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.
[17] Walts AE, Bose S, Fan X, et al. A simplified Bethesda system for reporting thyroid cytopathology using only four categories improves intra- and inter-observer diagnostic agreement and provides non-overlapping estimates of malignancy risks. Diagn Cytopathol 2012;40:E62–8.
[18] Davidov T, Trooskin SZ, Shanker BA, et al. Routine second-opinion cytopathology review of thyroid fine needle aspiration biopsies reduces diagnostic thyroidectomy. Surgery 2010;148:1294–9.
[19] Kocjan G, Chandra A, Cross PA, 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.
[20] Bongiovanni M, Crippa S, Baloch Z, et al. Comparison of 5-tired and 6-tired diagnostic systems for the reporting of thyroid cytopathology: a multi-institutional study. Cancer Cytopathol 201225;120:117–125.
[21] Jia Y, Yu Y, Li X, et al. Diagnostic value of B-RAF (V600E) in difficult-to-diagnose thyroid nodules using fine-needle aspiration: systematic review and meta-analysis. Diagn Cytopathol 2014;42:94–101.
[22] Trimboli P, Treglia G, Condorelli E, et al. BRAF-mutated carcinomas among thyroid nodules with prior indeterminate FNA report: a systematic review and meta-analysis. Clin Endocrinol (Oxf) 2016;84:315–20.
[23] Kim DS, Kim DW, Heo YJ, et al. Utility of including BRAF mutation analysis with ultrasonographic and cytological diagnoses in ultranography-guided fine-needle aspiration of thyroid nodules. PLoS One 2018;13:1–10.
[24] Su X, Jiang X, Xu X, et al. Diagnostic value of BRAF (V600E)-mutation analysis in fine-needle aspiration of thyroid nodules: a meta-analysis. Onco Targets Ther 2016;9:2495–509.