

Group Consensus Review Minimizes the Diagnosis of “Follicular Lesion of Undetermined Significance” and Improves Cytohistologic Concordance

Xin Jing, M.D.,* Stewart M. Knoepp, M.D., Ph.D., Michael H. Roh, M.D., Ph.D., Kim Hookim, M.D., Jeremiah Placido, M.D., Robertson Davenport, M.D., Rodolfo Rasche, M.D., and Claire W. Michael, M.D.

We conducted a group consensus review of thyroid aspirates that were previously interpreted as “atypia of undetermined significance/follicular lesion of undetermined significance” (AUS/FLUS) and followed by surgical interventions. The study aimed to investigate if consensus review would minimize the diagnosis of AUS/FLUS with an optimal interobserver agreement and also promote a better cytohistologic concordance. A group of reviewers who were blinded to the corresponding histologic findings simultaneously evaluated a total of 50 aspirates at a multi-headed light microscope. Using the Bethesda System for Reporting Thyroid Cytopathology as a guideline, a consensus interpretation was reached upon review of each aspirate. Interobserver agreement was calculated and recorded. The cytohistologic correlation was then performed between the consensus interpretation and the corresponding histologic diagnosis. The consensus review reclassified 26 (52%) aspirates as non-neoplasia/benign, 10 (20%) as follicular neoplasm/suspicious for a follicular neoplasm, 1 (2%) as papillary thyroid carcinoma, and 2 (4%) as nondiagnostic. Eleven (22%) aspirates remained AUS/FLUS. The interobserver agreement across the five diagnostic categories ranged from 71.6% to 100% with an average level of 88.8%. Cytohistologic concordance was achieved in 24 of 26 (92.3%) and 9 of 11 (81.8%) aspirates that were reclassified as non-neoplasia/benign and neoplasia/malignancy, respectively. A diagnostic accuracy of 89.2% (33/37) was obtained in reclassified

cases. In conclusion, the group consensus review minimized AUS/FLUS, offered an optimal level of interobserver agreement, and most importantly, promoted excellent cytohistologic concordance in reclassified cases and, therefore, could play a substantial role in the future in reducing reaspiration and/or unnecessary surgeries. Diagn. Cytopathol. 2012;40:1037–1042. © 2011 Wiley Periodicals, Inc.

Key Words: FNA; thyroid; follicular lesion of undetermined significance; consensus review; cytohistologic concordance

It has been widely accepted that fine needle aspiration (FNA) of thyroid plays an essential and critical role in the management of thyroid nodules. For the majority of thyroid nodules, FNA cytologic evaluation effectively differentiates neoplastic/malignant nodules requiring surgical intervention from non-neoplastic/benign nodules that may be managed conservatively with clinical and radiologic follow-up.^{1–4} However, in a subset of cases, FNA cytology of thyroid nodules exhibits cytomorphologic features that are neither convincingly non-neoplastic/benign nor sufficient for neoplasia/malignancy, and a definitive cytologic interpretation (non-neoplasia/benign versus neoplasia/malignancy) cannot be established. Instead, in these cases, an indeterminate interpretation such as “atypia of undetermined significance/follicular lesion of undetermined significance” (AUS/FLUS) is rendered.^{5–7} Recognizing the considerable variation among both pathologists and institutions in diagnosing indeterminate lesions, the newly published *The Bethesda System for Reporting Thyroid Cytology* (BSRTC) defines the diagnostic criteria for AUS/

Department of Pathology, University of Michigan, Ann Arbor, Michigan

*Correspondence to: Xin Jing, M.D., Department of Pathology, University of Michigan, 1500 E. Medical Center Drive, 2G332 UH, Ann Arbor, MI 48109-0054. E-mail: xinjing@med.umich.edu

Received 11 January 2011; Accepted 25 February 2011

DOI 10.1002/dc.21702

Published online 2 May 2011 in Wiley Online Library (wileyonlinelibrary.com).

Table I. Correlation Between Consensus Cytologic Interpretation and Histologic Diagnosis

Cytologic diagnosis	Interobserver agreement (%)	Histologic Diagnosis				Total (%)
		Benign (non-neoplastic)	Follicular adenoma	Follicular carcinoma	PTC	
Nondiagnostic	100	2	—	—	—	2 (4)
Benign	90.5	24	2	—	—	26 (52)
AUS/FLUS	71.6	5	3	2	1	11 (22)
FN/SFN	81.8	2	5	1	2	10 (20)
PTC	100	—	—	—	1	1 (2)
Total	—	33	10	3	4	50 (100)

FLUS and advises judicious use of this diagnostic terminology.⁸

FNA of thyroid nodules has been an established practice within our institution. Currently, the cytology service assesses more than 800 thyroid FNA specimens per year, of which nearly 6% are classified as AUS/FLUS. Data from a previous study at our institution demonstrated that using proposed standard morphological criteria similar to those of the BSRTC, in conjunction with independent second review, reduces the diagnosis of AUS/FLUS and improves cytohistologic correlation.⁹ The current study was undertaken to investigate if group consensus review would minimize AUS/FLUS diagnoses with optimal interobserver agreement while also promoting a better cytohistologic concordance.

Materials and Methods

A SNOMED search of the electronic pathology database in our institution for the period between January 1998 and June 2010 retrieved a total of 217 aspirates in which the original cytologic interpretation of AUS/FLUS was rendered and that were followed by hemi- or total thyroidectomies. For the retrospective consensus review, 50 aspirates were selected by choosing the first 50 consecutive cases in which a complete set of smears prepared from each aspirate was available. Forty-six of the 50 aspirates consisted of conventional air-dried Diff-Quik and alcohol-fixed Papanicolaou smears along with ThinPrep and/or cell block preparations; the remaining four aspirates contained ThinPrep and cell block preparation only. A group of reviewers, who were blinded to the corresponding histologic findings, simultaneously evaluated slides of each aspirate at a multiheaded light microscope. Reviewers were two cytopathology fellows (KH, JP), three junior (with 2–4 years of experience, XJ, SMK, MHR), and three senior cytopathologists (with more than 15 years of experience, RD, RR, CWM). At least five reviewers representing both junior and senior levels participated in the evaluation of each individual aspirate. Using the BSRTC as a guideline, individual interpretation was presented following the ascending order of the reviewer's levels of experience, and a consensus interpretation was then reached at the end of the group review. Interobserver agreement was calculated by dividing diagnosis of the majority by

the total. For example, interobserver agreement for an individual aspirate was 100% if all participating reviewers agreed on the cytologic interpretation for an individual aspirate; however, interobserver agreement would decrease to 60% if three out of five participating reviewers concurred with the interpretation, while the other two reviewers disagreed with the interpretation. Mean interobserver agreement was calculated for each category at the end of the study. For each aspirate, the cytohistologic correlation (the consensus interpretation versus the corresponding histologic diagnoses) was performed immediately upon obtaining the consensus cytologic interpretation. Cytologic features of each aspirate were discussed in detail throughout the group consensus review. To assess if reviewers gained positive feedback through the approach of immediate cytohistologic correlation, we excluded two aspirates reclassified as nondiagnostic specimen and then compared performance of the consensus review between earlier (review of first 24 diagnostic aspirates) and later phase (review of last 24 diagnostic aspirates).

Results

Using BSRTC⁸ as a guideline, the group consensus review of 50 thyroid aspirates previously interpreted as AUS/FLUS resulted in reclassification of 26 (52%) aspirates as non-neoplasia/benign, including nodular hyperplasia and chronic lymphocytic thyroiditis, 10 (20%) aspirates as follicular neoplasm/suspicious for a follicular neoplasm (FN/SFN), 1 (2%) aspirate as papillary thyroid carcinoma (PTC), and 2 (4%) aspirates as nondiagnostic because of the insufficient number of follicular cells. On the other hand, 11 (22%) aspirates remained AUS/FLUS following the group review. The interobserver agreement across the five diagnostic categories ranged from 71.6% for AUS/FLUS to 100% for PTC and nondiagnostic with an average level of 88.8% (Table I). Table I also shows correlation between the cytologic interpretations rendered by the group consensus review and the corresponding histologic diagnoses. Cytohistologic concordance was achieved in 24 of 26 (92.3%) and 9 of 11 (81.8%) aspirates that were reclassified as non-neoplasia/benign (Fig. 1) and neoplasia/malignancy (Figs. 2 and 3), respectively. The histology-proved neoplasia/malignancy included six follicular adenomas and three

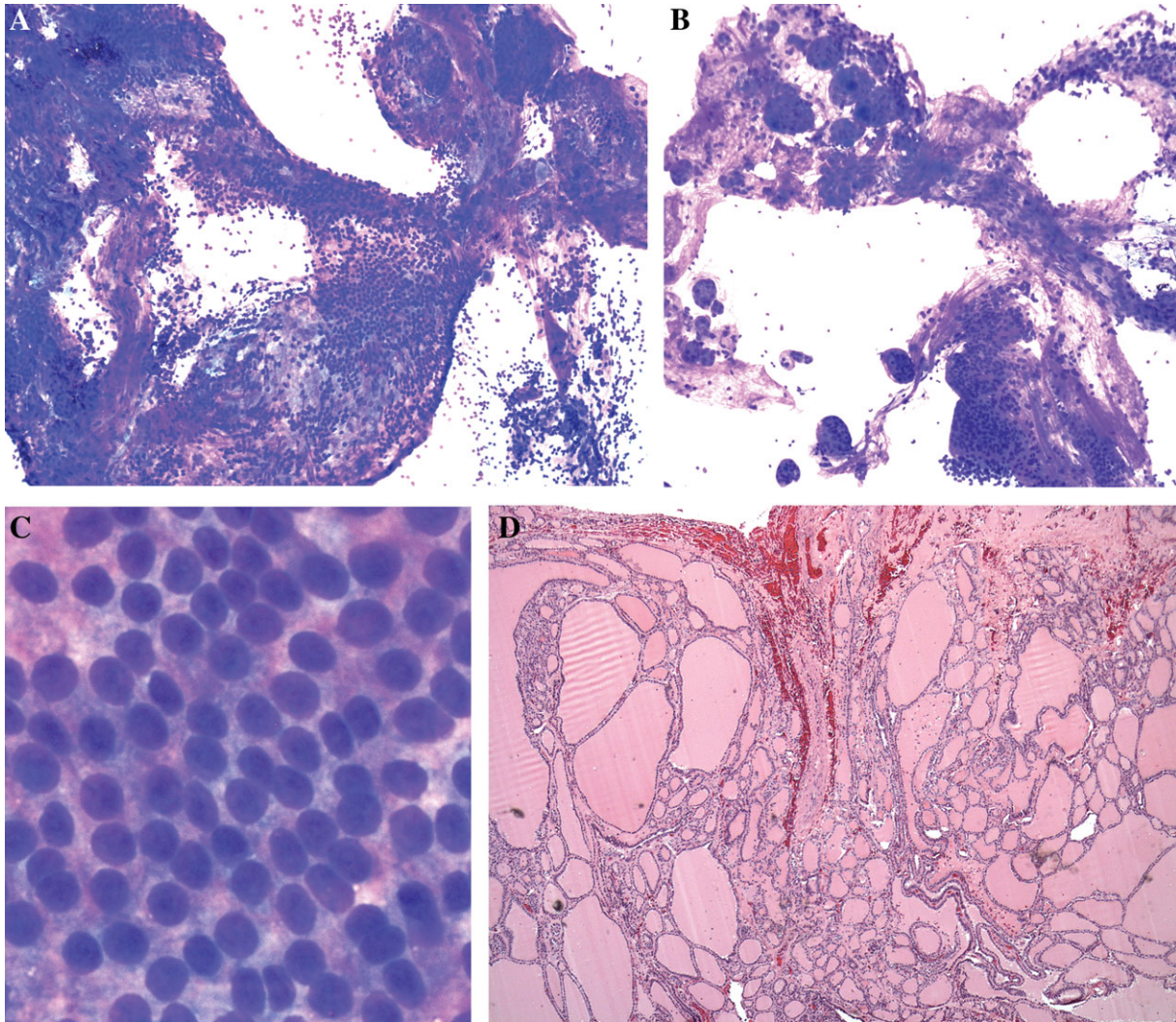


Fig. 1. A nodule reclassified as nodular hyperplasia. (A) and (B): Hypercellular aspirate composed of follicular cells that are arranged in large tissue fragments and intact follicles (Diff-Quik, $\times 100$). (C): Higher power view showing honeycomb pattern with evenly distributed, uniform nuclei (Diff-Quik, $\times 600$). (D): Histology showing nodular hyperplasia (H&E, $\times 100$). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

PTC among which two were follicular variant of PTC. A diagnostic accuracy of 89.2% (33/37) was obtained in the reclassified aspirates. Cytologic misinterpretation during the group review contributed to the false-negative and false-positive interpretations. Dividing the study period into earlier (review of first 24 diagnostic aspirates) and later phase (review of last 24 diagnostic aspirates), 20 and 13 aspirates with cytohistologic concordance were reviewed during later and earlier phase of the study, respectively; all four aspirates with cytohistologic discordance were reassessed during the early phase of the group review. Additionally, out of 11 aspirates remained as AUS/FLUS, 7 and 4 aspirates were results of earlier and later group review, respectively (Table II).

Histologic examination of the 11 aspirates remaining as AUS/FLUS revealed five benign diagnoses, three follicu-

lar adenomas, two follicular carcinomas, and one conventional PTC. Thus, malignancy was detected in corresponding surgical specimens in 27% (3/11) of the remaining indeterminate aspirates.

Discussion

Profound variations among both pathologists (i.e., 2.5–28.6%) and institutions (i.e., 3.3–14.9%) have been previously described with regard to applying the indeterminate diagnostic terminology for FNA cytology evaluation of thyroid nodules.¹⁰ One of the dominant contributing factors leading to the variation is the lack of precisely defined, objective morphological criteria.^{10–12} In addition, threshold level and stringency in applying the diagnostic criteria may differ among reviewers and institutions. According to the newly published BSRTC,⁸ AUS/FLUS

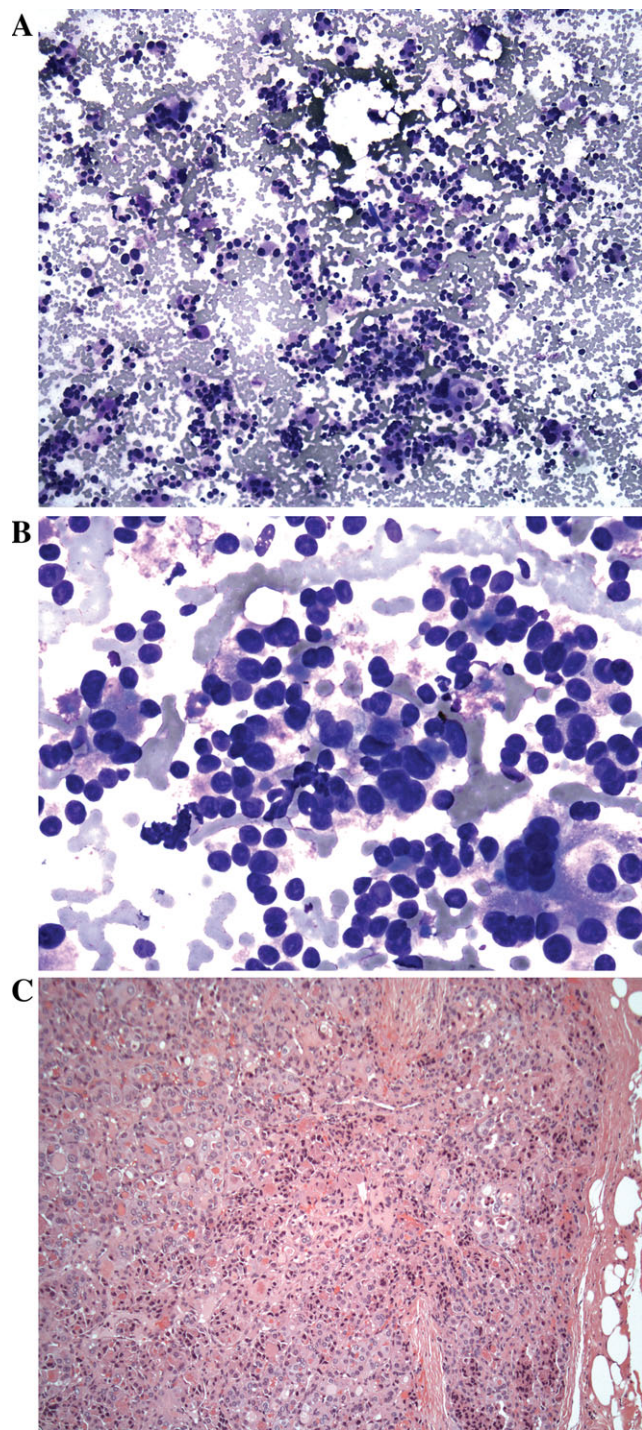


Fig. 2. A nodule reclassified as follicular neoplasm. (A): Hypercellular aspirate consisted of numerous microfollicles with nuclear crowding/overlapping (Diff-Quik, $\times 100$). (B): The follicular cells demonstrate marked cytologic atypia (Diff-Quik, $\times 400$). (C): Histology revealing follicular carcinoma (H&E, $\times 100$). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

is mainly reserved for aspirates with architectural and/or cytologic atypia that is not sufficient to be categorized as suspicious for neoplasia/malignancy, yet on the other

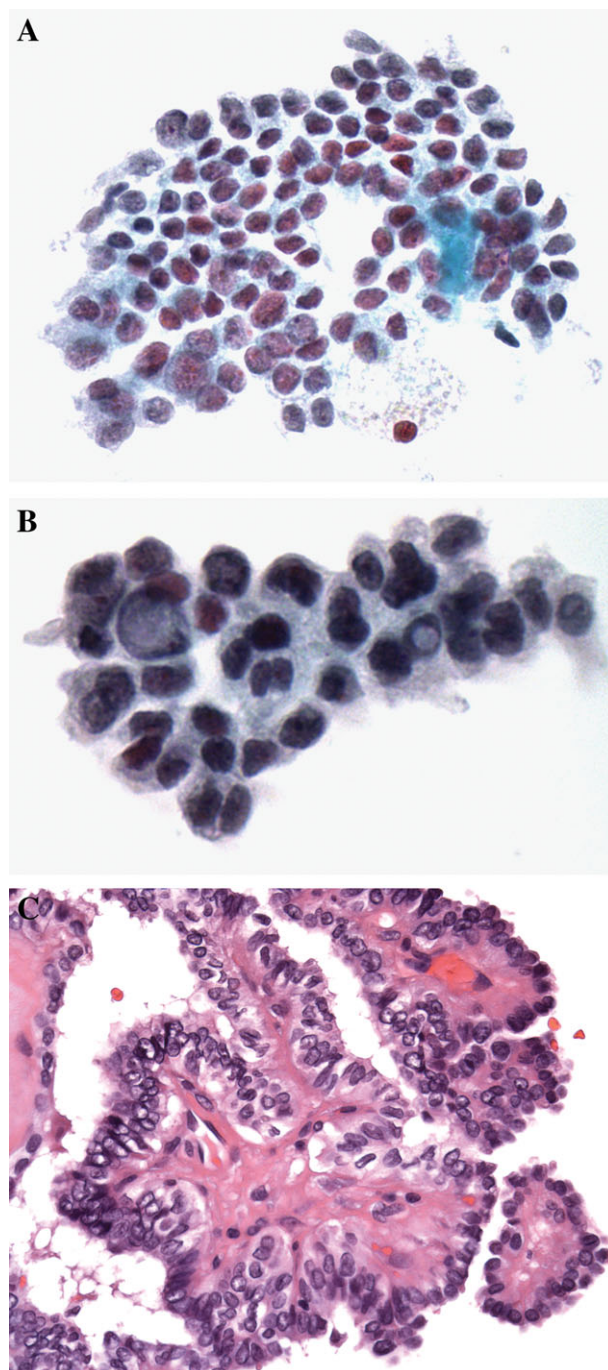


Fig. 3. A nodule reclassified as papillary thyroid carcinoma. (A) and (B): ThinPrep preparation showing follicular cells that are arranged in syncytial sheets. The nuclei exhibit features of papillary thyroid carcinoma (Papanicolaou, $\times 600$). (C): Histology revealing papillary thyroid carcinoma (H&E, $\times 400$). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

hand, is more marked than can be ascribed confidently to benign changes. BSRTC recommends the indeterminate interpretation be used in $<7\%$ of all thyroid FNA interpretation and suggests clinical correlation and repeat aspiration after an appropriate interval following initial

Table II. Distribution of the Results Between Earlier and Later Phase of the Group Review

Results	Earlier phase (<i>n</i> = 24)	Later phase (<i>n</i> = 24)
Cytohistic concordance	13	20
Cytohistic discordance	4	0
AUS/FLUS	7	4

indeterminate interpretation.⁸ In our institution, nearly 6% of all thyroid aspirates fall into the indeterminate category during the last 12 years. On the other hand, virtually all thyroid nodules categorized as AUS/FLUS at our institution were managed with surgical interventions prior to the implementation of the newly published BSRTC recommendation.

Our data demonstrates that using BSRTC as a guideline, group consensus review of the aspirates previously categorized as AUS/FLUS: (1) minimizes AUS/FLUS without compromising diagnostic accuracy; (2) achieves an optimal level of interobserver agreement; and (3) promotes a better cytohistic concordance and diagnostic accuracy. Overall, the group consensus review could have played a substantial role in reducing unnecessary surgical intervention, and more than 90% of non-neoplastic/benign thyroid nodules could have been managed with clinical/radiological follow-up instead of surgical intervention.

In comparison to results of our previously published data generated from the individualized second review of 123 indeterminate thyroid aspirates,⁹ the current data illustrates an identical distribution in the diagnostic categories of both non-neoplasia/benign (52%) and neoplasia/malignancy (22%). The percentage of remaining AUS/FLUS in the current study (22%) appears to be slightly lower than that of prior study (26%). We thus believe that previously categorized indeterminate lesions in our institution were well-represented by the 50 aspirates in the study. However, it is noteworthy that a more accurate cytohistic concordance is achieved in the current study for both non-neoplasia/benign (92.3% versus 73.4%) and neoplasia/malignancy (81.8% versus 55.6%). Thus, the current data demonstrates a higher level of diagnostic accuracy for reclassified aspirates (89.2% versus 68.1%). Considering that both studies employed similar cytomorphologic criteria as a guideline, we suggest that in addition to the approach of consensus review itself, positive feedback obtained through cytohistic correlation performed immediately upon completion of the group review of each aspirate played a significant role in real-time affirmation of the diagnostic cytologic features. In this regard, it is interesting to see that later phase of the study generated more cytohistic concordant cases (*n* = 20) compared with earlier phase (*n* = 13), none of the cytohistic discordant cases occurred during the later phase and furthermore, later phase of the study generated less AUS/FLUS (*n* = 4) compared with earlier

phase (*n* = 7). We believe that the group consensus review along with a consistent and effective feedback contributed to marked improvement in diagnostic accuracy. In fact, diagnostic accuracy obtained in the current study is comparable to the previously reported data, which were generated from a study of general cohort of thyroid aspirates.¹³ The findings support our postulate that the refined consensual approach combined with a consistent and effective feedback brings enhanced accuracy in the categorization of the thyroid lesions that pose diagnostic challenges.

With regard to assessing influence of reviewer's experience on classifying aspirates as AUS/FLUS, we were fully aware that different conclusions were drawn from the studies that used different approaches. For example, a review of original cytology reports from multi-institutions did not show a relationship between the level of reviewer's experience and usage of the indeterminate terminology,¹⁰ while a cytologic re-review of previously diagnosed indeterminate lesions demonstrated less usage of the indeterminate terminology by the reviewers with more experience.¹¹ We believed that it is unlikely to generate additional information by applying the similar approaches to our existing cohort of AUS/FLUS. Thus, the current study focused on the role of group consensus review in the reassessment of previously categorized indeterminate aspirates. Before reaching a consensus interpretation for each aspirate, individual interpretation was presented following the ascending order of the reviewer's levels of experience so that the junior reviewers made their decision without being influenced by the senior reviewers. Noting that the reassessment was performed on the thyroid lesions with diagnostic challenges, the interobserver agreement across the diagnostic categories descended in the following order: malignant PTC (100%) > non-neoplasia/benign (90.5%) > FN/SFN (81.8%) > AUS/FLUS (71.6%). Not surprisingly, the least interobserver agreement was obtained for the category of AUS/FLUS.

Different approaches are explored for further stratification of thyroid lesions categorized as AUS/FLUS. In this regard, potential benefit of applying molecular analysis as an adjunct of FNA cytology has been investigated.^{14,15} Regardless, the cytomorphologic assessment is an essential and critical, as well as a cost-effective, method in the management of thyroid nodules.

In conclusion, the group consensus review minimized AUS/FLUS, offered an optimal level of interobserver agreement, and most importantly, promoted excellent cytohistic concordance in reclassified cases and, therefore, could play a substantial role in the future in reducing reaspiration and/or unnecessary surgeries.

References

- Gharib H. Fine-needle aspiration biopsy of thyroid nodules: Advantages, limitations, and effect. *Mayo Clin Proc* 1994;69:44-49.

2. Holleman F, Hoekstra JB, Ruitenbergh HM. Evaluation of fine needle aspiration (FNA) cytology in the diagnosis of thyroid nodules. *Cytopathology* 1995;6:168–175.
3. Mandreker SR, Nadkarni NS, Pinto RG, Menezes S. Role of fine needle aspiration cytology as the initial modality in the investigation of thyroid lesions. *Acta Cytol* 1995;39:898–904.
4. Zagorianakou P, Malamou-Mitsi V, Zagorianakou N, Stefanou D, Tsatsoulis A, Agnantis NJ. The role of fine-needle aspiration biopsy in the management of patients with thyroid nodules. *In Vivo* 2005;19:605–609.
5. Busseniers AE, Oertel YC. “Cellular adenomatoid nodules” of the thyroid: Review of 219 fine-needle aspirates. *Diagn Cytopathol* 1993;9:581–589.
6. Poller DN, Ibrahim AK, Cummings MH, Mikel JJ, Boote D, Perry M. Fine-needle aspiration of the thyroid. *Cancer* 2000;90:239–244.
7. Greaves TS, Olvera M, Florentine BD, et al. Follicular lesions of thyroid: A 5-year fine-needle aspiration experience. *Cancer* 2000;90:335–341.
8. Krane JF, Nayar R, Renshaw AA. Atypia of undetermined significance/follicular lesion of undetermined significance. In: Ali SZ, Cibas ES, editors. *The Bethesda System for Reporting Thyroid Cytopathology*. New York: Springer; 2010. p 37.
9. Jing X, Roh MH, Knoepp SM, Zhao L, Michael CW. Minimizing the diagnosis of “follicular lesion of undetermined significance” and identifying predictive features for neoplasia. *Diagn Cytopathol*. DOI: 10.1002/dc.21459 [Epub ahead of print].
10. Layfield LJ, Morton MJ, Cramer HM, Hirschowitz S. Implications of the proposed thyroid fine-needle aspiration category of “follicular lesion of undetermined significance”: A five-year multi-institutional analysis. *Diagn Cytopathol* 2009;37:710–714.
11. 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;49:378–382.
12. Shi Y, Ding X, Klein M, et al. Thyroid fine-needle aspiration with atypia of undetermined significance: A necessary or optional category?. *Cancer Cytopathol* 2009;117:298–304.
13. Baloch ZW, Sack MJ, Yu GH, Livolsi VA, Gupta PK. Fine-needle aspiration of thyroid: An institutional experience. *Thyroid* 1998;8:565–569.
14. Nikiforov YE, Steward DL, Robinson-Smith TM, et al. Molecular testing for mutations in improving the fine-needle aspiration diagnosis of thyroid nodules. *J Clin Endocrinol Metab* 2009;94:2092–2098.
15. Nikiforova MN, Nikiforov YE. Molecular diagnostics and predictors in thyroid cancer. *Thyroid* 2009;19:1351–1361.