The Frequency of ‘Atypia of Undetermined Significance’ Interpretations for Thyroid Fine-Needle Aspirations Is Negatively Correlated with Histologically Proven Malignant Outcomes

Paul A. VanderLaan Jeffrey F. Krane Edmund S. Cibas

Department of Pathology, Brigham and Women’s Hospital and Harvard Medical School, Boston, Mass., USA

Key Words

Thyroid • Fine-needle aspiration • Cytology • Atypia of undetermined significance

Abstract

Objective: Cytopathologists’ usage patterns for ‘atypia of undetermined significance’ (AUS) in thyroid fine-needle aspiration (FNA) are not well understood. AUS rates over a 5-year period were analyzed to quantify variability and identify correlations with experience and histologic outcomes.

Study Design: A retrospective review of thyroid FNAs from a tertiary-care hospital from 2005 to 2009 was performed. Results were compiled for individual cytopathologists, stratified by year, and correlated with histologic outcomes. **Results:** Thyroid FNAs (5,327) were evaluated by 7 cytopathologists, with an overall AUS rate of 11.2%. The annual AUS rate remained relatively constant over this time period, though notable inter- and intrapathologist variability was seen. The AUS rate was significantly lower for those with cytopathology boards (10.3%) compared to those without (14.0%). There was no correlation between the AUS rate and cytopathologist experience or thyroid FNA volume. The AUS rate and malignant outcome were inversely related: the higher an individual’s AUS rate was, the lower the rate of malignancy for that AUS cohort was. **Conclusions:** Individual cytopa-

thologist AUS rates were variable and often exceeded the recommended target of 7%. The application of recently published defined diagnostic criteria, along with directed cytopathologist feedback, may reduce observer variability and appropriately lower AUS utilization.

Copyright © 2011 S. Karger AG, Basel

Introduction

The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC), born of the 2007 National Cancer Institute State of the Science Conference, established a uniform 6-tiered diagnostic system for the evaluation and reporting of thyroid fine-needle aspirations (FNAs) [1, 2]. In this standardized reporting framework, each category carries an implied risk of malignancy ranging from 0 to 3% for a benign diagnosis to nearly 100% for a malignant diagnosis. Since the introduction of TBSRTC, a number of reports have focused on the category ‘atypia of unde-

This work was presented in part as a platform presentation at the 100th Annual Meeting of the United States and Canadian Academy of Pathology in San Antonio, Tex., on March 1, 2011, by Paul A. VanderLaan MD, PhD.

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2011 S. Karger AG, Basel
0001–5547/11/0556–0512\$38.00/0

Accessible online at:
www.karger.com/acy

Correspondence to: Dr. Edmund S. Cibas
Department of Pathology, Brigham and Women’s Hospital
75 Francis Street
Boston, MA 02115 (USA)
Tel. +1 617 732 6797, E-Mail ecibas@partners.org

terminated significance' (AUS), also called 'follicular lesion of undetermined significance' (FLUS). This category is heterogeneous by definition and reserved for specimens with features not sufficient to be classified as 'suspicious for malignancy' or 'suspicious for a follicular neoplasm' but showing more atypia than can be ascribed to benign changes alone [2]. In TBSRTC, a target utilization rate for the AUS category was set at approximately 7% of thyroid aspirates, a figure based on the observed AUS rates from two large studies published around the time of the 2007 NCI conference [3, 4], although it was acknowledged that this figure might be revisited. To date, the majority of laboratories that have retrospectively examined their experience with AUS in TBSRTC framework report rates higher than the 7% target, with a range of 2.1–18% [5–15]. It is not well understood what factors contribute to this variable usage by cytopathologists. To this end, we set out to identify variables that might influence the frequency with which a cytopathologist uses the AUS category. Additionally, we evaluated the correlation between the frequency of AUS use by a cytopathologist and a histologically proven malignant outcome for the cytopathologist's AUS cohort, the so-called 'malignancy risk'.

Materials and Methods

Following approval by the institutional review board, a retrospective review of thyroid FNAs at the Brigham and Women's Hospital from January 2005 through December 2009 was performed with the aid of a customized report program written for the laboratory information system. Based on the PowerPath client platform (software version 8.3.26.7; Tamtron Corporation), the report generated a diagnostic category breakdown for thyroid FNA interpretations for the entire laboratory as well as for individual cytopathologists over any specified time period. All thyroid FNAs over this time period were performed under ultrasound guidance by an endocrinologist using a 25-gauge needle (typically 3 or 4 passes without routine on-site evaluation). The specimens were collected immediately in CytoLyt® (Hologic, Inc., Marlborough, Mass., USA), and a single Papanicolaou-stained ThinPrep® slide was prepared from each specimen using ThinPrep 2000® (Hologic). Slides were screened by a cytotechnologist and reviewed by a cytopathologist, who reported the result using a six-tiered category-based diagnostic system identical to TBSRTC except for minor terminological differences. The diagnostic categories (followed by the corresponding Bethesda System designation) were: nondiagnostic specimen (nondiagnostic), no malignant cells identified (benign), atypical cells of undetermined significance (AUS/FLUS), suspicious for a Hürthle cell (oncocyctic) neoplasm/suspicious for a follicular neoplasm (suspicious for a Hürthle cell/follicular neoplasm), suspicious for malignancy (suspicious for malignancy), and positive for malignant cells (malignant).

Results were compiled for the entire laboratory as well as for individual cytopathologists and were stratified by year. To ensure robustness of the data, only those specimens evaluated by cytopathologists reporting a minimum of 450 thyroid FNAs over this 5-year time period were included in the analysis. Seven cytopathologists met the inclusion criterion, with their specimens accounting for over 95% of the in-house thyroid FNA case volume over the study period. All cytopathologists were board certified in anatomic pathology, and five were board certified in cytopathology. When follow-up surgical results were available, the histologic outcomes were correlated with the AUS diagnosis rate. Detailed information on the histologic outcome data characteristics comprising a large part of this cohort of cases has been published previously [14].

Data processing and statistical analysis was performed using Microsoft Excel and GraphPad software. The χ^2 test was used to compare categorical data between groups, and linear regression analysis was utilized for variable correlations. $p < 0.05$ was considered statistically significant.

Results

Over the 5-year period, seven cytopathologists evaluated a total of 5,327 thyroid FNAs. An AUS interpretation was rendered on 595 (11.2%) cases. The distribution of thyroid FNA interpretations rendered by each cytopathologist, as well as the lab average, for the entire 5-year period is shown in figure 1.

To determine how consistent the AUS utilization rate was over time, data for the laboratory and individual cytopathologists were compiled and stratified by year. As depicted in figure 2, the annual average AUS rate for the laboratory remained relatively constant over this time period (9.9–12.4%), but notable variability was observed not only between individual cytopathologists but also for a given cytopathologist over time. The overall AUS utilization rate for individual cytopathologists ranged from 6.1 to 18.7%, with individual cytopathologists demonstrating a standard deviation of 1.4–4.0 percentage points in the personal AUS utilization rate over the 5-year time period (table 1). The intercytopathologist variability in AUS use tended to decrease over time (fig. 2). Linear regression analysis demonstrated a decrease in standard deviation for the laboratory over time, although this trend did not reach statistical significance (data not shown).

A number of variables were analyzed to determine if cytopathologist characteristics were related to the AUS utilization rate. The first was the cytopathologist's level of experience, as measured by the number of years practicing cytopathology from the beginning of practice to the end of the study period (range 3–24 years; table 1). The years of experience were not significantly correlated with the AUS

Fig. 1. Distribution of aggregate thyroid FNA interpretations over the 5-year period of 2005–2009 (inclusive), stratified by cytopathologist. CP = Cytopathologist; Sus Malig = suspicious for malignancy; Sus Hur = suspicious for a Hürthle cell neoplasm; Sus Fol = suspicious for a follicular neoplasm; Lab Ave = laboratory average.

Fig. 2. Annual frequencies of 'AUS' interpretations, stratified by individual cytopathologist. CP = Cytopathologist; Lab Ave = laboratory average.

Fig. 3. Correlation between the 'AUS' rate of a CP and the frequency of a histologically confirmed malignant outcome for the AUS cases (single institution data, current study). Compare with figure 4.

Fig. 4. Correlation between the published 'AUS' rate of a laboratory and the frequency of a histologically confirmed malignant outcome for the AUS cases. Data extracted from a literature review. Numbers attached to data points correspond to citations in References.

utilization rate as determined by linear regression (data not shown). The second cytopathologist characteristic was cytopathology subspecialty board certification. Five of 7 cytopathologists were cytopathology board certified (table 1). The AUS rate was significantly lower for those with cytopathology boards (10.3%, 419/4,066 cases) compared to those without (14.0%, 176/1,261 cases) ($\chi^2 = 12.9$, $p < 0.001$), though it should be noted that the numbers of cytopathologists comprising each group were small. Finally, both the annual and the total number of thyroid FNAs evaluated by each cytopathologist during the study period

were compared to their resultant AUS rate. Linear regression analyses failed to show any significant correlation with thyroid case volume (data not shown).

Histologic outcome data were available for 33% (199/595) of the cases interpreted as AUS, with a malignant outcome observed in 43% (85/199) of cases, as reported previously [14]. An inverse relationship was found between the frequency of AUS diagnosis and the rate of malignancy (fig. 3). Stated differently, an AUS interpretation from a cytopathologist who uses the category more frequently has a lower positive predictive value for malig-

Table 1. The frequency of ‘AUS’ interpretations of individual CPs: relationship to experience, board certification, and risk of histologically proven malignant outcome for the AUS cohort

	Experience years	Cytopathology boards	Mean AUS rate (annual range)	Malignancy rate (n)
CP1	22	Yes	9.6 ± 3.6% (6.5–15.0%)	36.1% (13/36)
CP2	9	Yes	8.1 ± 3.5% (4.2–12.6%)	33.3% (8/24)
CP3	24	No	18.7 ± 4.0% (12.7–22.4%)	19.6% (11/56)
CP4	3	No	10.5 ± 3.0% (7.2–13.4%)	12.5% (5/40)
CP5	13	Yes	10.0 ± 1.4% (8.7–11.7%)	36.4% (20/55)
CP6	6	Yes	14.1 ± 2.0% (11.5–16.2%)	23.0% (20/87)
CP7	7	Yes	6.1 ± 1.7% (3.4–7.5%)	61.5% (8/13)
Laboratory average	10	5: yes, 2: no	11.2 ± 0.9% (9.9–12.4%)	27.3% (85/311)

nancy than an AUS interpretation from a cytopathologist who uses AUS less frequently.

To determine if this inverse relationship holds true across different laboratories, similar data were extracted from published studies on thyroid FNAs reported in a TBSRTC-like system [6–14]. Figure 4 depicts the AUS rate of a laboratory plotted against the histologically proven risk of malignancy for the AUS cases. Although it did not reach statistical significance, a similar inverse correlation was observed, supporting the conclusion that higher AUS rates are associated with a lower risk of malignancy.

Discussion

This study confirms that, historically, AUS use by cytopathologists has been variable, both among individual cytopathologists and for a given cytopathologist over time. We were not able to attribute the presumed overuse of AUS to cytopathologist inexperience or the overall volume of thyroid FNAs evaluated, though it might be related to training since we found that cytopathology board certification was correlated with lower AUS utilization. Very similar results have been reported by others [6]. We found that individual cytopathologist rates of AUS diagnosis were negatively correlated with the risk of malignancy. A similar trend was apparent in data extracted from the published literature.

What can one infer from the inverse relationship between AUS usage by cytopathologists and the associated risk of malignancy? Since a high AUS rate is linked to a lower malignancy rate, it is likely that overuse of the AUS category is due to overcalling ‘benign’ FNAs (with a malignancy rate of 0–3%) AUS rather than undercalling specimens more aptly classified as ‘suspicious for malig-

nancy’ (malignancy rate 60–75%) or ‘suspicious for a follicular/Hürthle cell neoplasm’ (malignancy rate 15–45%) [2]. This hypothesis is supported by the fact that most thyroid nodules are benign, and therefore it is statistically more probable that an aspirate interpreted as AUS is an overdiagnosis of a benign nodule rather than an undercall of a malignant nodule. In a recent set of studies involving a rereview of AUS cases incorporating either a consensus review or application of defined diagnostic morphologic criteria, over half of AUS cases were reclassified as benign/nonneoplastic, with excellent concordance achieved on subsequent cytologic-histologic correlation [16, 17]. In these studies, only a minority of cases retained an AUS diagnosis (22–26%) after stringent application of TBSRTC criteria. Therefore, careful consideration of specific cytologic features must be differentiated from those anticipated in the spectrum of benign changes for a diagnosis of AUS to be warranted.

In practice, determining the threshold between ‘AUS’ and ‘benign’ is often a matter of the individual cytopathologist’s judgment. Experience shapes judgment, and it seems logical that those with more years practicing cytology may be less apt to use the AUS category and more willing (and able) to use the more definitive ‘benign’ or ‘suspicious’ categories. Conversely, additional experience could enable the cytopathologist to recognize subtle cytologic features that might otherwise be overlooked, resulting in higher AUS rates. In this regard, we have confirmed what others have previously shown, namely that the number of years practicing cytology does not correlate with AUS use but that cytopathology board certification does correlate with a lower AUS usage rate [6]. Additional training and/or certification may help cytopathologists to better distinguish truly benign changes from those that are worrisome enough to warrant an

atypical diagnosis. In the data presented here, the combined AUS and benign categories account for the majority of thyroid FNA diagnoses (73.4%, individual cytopathologist range 69.3–77.2%). As illustrated in figure 1, the sum of the benign and AUS categories appears to be relatively constant across individual cytopathologists. Therefore, a cytopathologist with a low threshold for calling a thyroid FNA AUS likely does so at the expense of a benign diagnosis rather than undercalling a suspicious for malignancy/follicular neoplasm/Hürthle cell neoplasm.

Performance measures in cytology are useful for ensuring that an individual's practice patterns are in line with published benchmarks and departmental averages. Currently, cytopathologists in our laboratory receive annual confidential feedback on certain metrics, including turnaround time, and, for gynecologic specimens, their ASCUS-to-SIL ratio and HPV positivity rate for ASCUS cases [18, 19]. Over the 5-year period studied here, the AUS utilization rate was not routinely measured or reported to cytopathologists, and TBSRTC was formally published only at the conclusion of the study period. Therefore, this study reflects historic cytopathologist practice patterns. Linear regression analysis demonstrated a trend towards a decrease in the standard deviation range for AUS interpretations over time. This would suggest that, as the laboratory has gained experience with the AUS diagnostic category, there has been a tendency for increased intralaboratory reproducibility likely achieved by the sharing of difficult cases and participation in regularly scheduled consensus conferences. Looking forward, it remains to be seen if the recently introduced (for calendar year 2010) confidential feedback on thyroid AUS rates provided to the cytopathologists in our laboratory will result in a lower AUS rate over time.

TBSRTC recommends a target rate of 7% for the AUS category. Many laboratories currently exceed this benchmark, which has prompted consideration of ways to reduce unnecessary usage. Although overuse of the AUS category is likely due to the overcalling of benign cytomorphologic changes or preparation artifacts, which lowers specificity, underuse of AUS may result in abnormally low AUS rates, which would lower sensitivity. Hence, there may be a limit to a 'lower is better' approach. When appropriately used, the proportion of cases allocated to the AUS category should have a malignancy risk that would justify conservative management, like a repeat FNA in 3–6 months as currently recommended by TBSRTC. Extrapolating the 7% target from TBSRTC to the linear regressions from this study (fig. 3) and the meta-analysis of published studies (fig. 4) gives an estimated

malignancy rate of 42 and 32%, respectively. These malignancy rates are higher than the estimated overall malignancy rate for the AUS category in the Bethesda System of 5–15% because they are calculated from only a subset of nodules interpreted as AUS, namely those that were selected for surgery (only 33% of all AUS nodules in this series). Indeed, there may be a significant selection bias for these AUS-diagnosed nodules based on clinical and sonographic features. In order to assess the true malignancy rate for thyroid nodules carrying an AUS diagnosis under current TBSRTC criteria, one would need a prospective cohort with 100% surgical follow-up, but such a study design would be overly aggressive in the management of these nodules.

Despite the challenges associated with the AUS diagnostic category, it is important to recognize that, when properly utilized, it serves an important function in thyroid FNA interpretation. Clearly, as outlined in TBSRTC, this category is useful for dealing with specimens that have technical artifacts precluding definitive characterization (such as clotting or air drying artifacts or a sparsely cellular specimen). Indeed, it has been demonstrated that eliminating this diagnostic category results in a decrease in the sensitivity and an increase in the false-positive and false-negative rates of thyroid FNAs [9]. Although some degree of ambiguity surrounds an AUS diagnosis, TBSRTC currently offers a clear recommendation for the subsequent management of these low-risk nodules: a repeat aspirate in 3–6 months, with surgery for those nodules diagnosed as AUS or worse.

In the management of patients with a thyroid nodule, the FNA is one component, albeit an important one, in determining the best course of action. Clearly, the clinician interprets the thyroid aspirate in the context of all of the information available, including radiographic and clinical data [20]. As such, it is important for the cytologist to take steps to reduce the 'atypical' diagnoses whenever possible, either in a broad system-wide fashion [21] or through more focused, specimen-specific measures, such as group consensus reviews [16] and/or the confidential feedback recently instituted in our laboratory. Looking ahead, familiarity with and application of TBSRTC-defined diagnostic criteria for AUS, along with periodic feedback, may help achieve AUS rates approximating the recommended lower target for this category.

Disclosure Statement

None.

References

- 1 Cibas ES, Ali SZ, NCI Thyroid FNA State of the Science Conference: The Bethesda System for Reporting Thyroid Cytopathology. *Am J Clin Pathol* 2009;132:658–665.
- 2 Ali SZ, Cibas ES: The Bethesda System for Reporting Thyroid Cytopathology, ed 1. New York, Springer, 2010.
- 3 Yassa L, Cibas ES, Benson CB, Frates MC, Doubilet PM, Gawande AA, Moore FD, Kim BW, Nosé V, Marqusee E, Larsen PR, Alexander EK: Long-term assessment of a multidisciplinary approach to thyroid nodule diagnostic evaluation. *Cancer Cytopathol* 2007;111:508–516.
- 4 Yang J, Schnadig V, Logrono R, Wasserman PG: Fine-needle aspiration of thyroid nodules: a study of 4,703 patients with histologic and clinical correlations. *Cancer* 2007;111:306–315.
- 5 Krane JF, Vanderlaan PA, Faquin WC, Renshaw AA: The atypia of undetermined significance/follicular lesion of undetermined significance: malignant ratio: a proposed performance measure for reporting in the Bethesda System for Thyroid Cytopathology. *Cancer Cytopathol* 2011, DOI: [10.1002/cncy.20192](https://doi.org/10.1002/cncy.20192).
- 6 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.
- 7 Nayar R, Ivanovic M: The indeterminate thyroid fine-needle aspiration: experience from an academic center using terminology similar to that proposed in the 2007 National Cancer Institute Thyroid Fine Needle Aspiration State of the Science Conference. *Cancer Cytopathol* 2009;117:195–202.
- 8 Theoharis CG, Schofield KM, Hammers L, Udelsman R, Chhieng DC: The Bethesda thyroid fine-needle aspiration classification system: year 1 at an academic institution. *Thyroid* 2009;19:1215–1223.
- 9 Shi Y, Ding X, Klein M, Sugrue C, Matano S, Edelman M, Wasserman P: Thyroid fine-needle aspiration with atypia of undetermined significance: a necessary or optional category? *Cancer Cytopathol* 2009;117:298–304.
- 10 Renshaw AA: Should 'atypical follicular cells' in thyroid fine-needle aspirates be subclassified? *Cancer Cytopathol* 2010;118:186–189.
- 11 Jo VY, Stelow EB, Dustin SM, Hanley KZ: Malignancy risk for fine-needle aspiration of thyroid lesions according to the Bethesda System for Reporting Thyroid Cytopathology. *Am J Clin Pathol* 2010;134:450–456.
- 12 Somma J, Schlecht NF, Fink D, Khader SN, Smith RV, Cajigas A: Thyroid fine needle aspiration cytology: follicular lesions and the gray zone. *Acta Cytol* 2010;54:123–131.
- 13 Marchevsky AM, Walts AE, Bose S, Gupta R, Fan X, Frishberg D, Scharre K, Zhai J: Evidence-based evaluation of the risks of malignancy predicted by thyroid fine-needle aspiration biopsies. *Diagn Cytopathol* 2010;38:252–259.
- 14 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;135:770–775.
- 15 Faquin WC, Baloch ZW: Fine-needle aspiration of follicular patterned lesions of the thyroid: diagnosis, management, and follow-up according to National Cancer Institute (NCI) recommendations. *Diagn Cytopathol* 2010;38:731–739.
- 16 Jing X, Knoepp SM, Roh MH, Hookim K, Placido J, Davenport R, Rasche R, Michael CW: Group consensus review minimizes the diagnosis of 'follicular lesion of undetermined significance' and improves cytohistologic concordance. *Diagn Cytopathol* 2011, DOI: [10.1002/dc.21702](https://doi.org/10.1002/dc.21702).
- 17 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* 2010;39:737–742.
- 18 Juskevicius R, Zou KH, Cibas ES: An analysis of factors that influence the ASCUS/SIL ratio of pathologists. *Am J Clin Pathol* 2001;116:331–335.
- 19 Nascimento AF, Cibas ES: The ASC/SIL ratio for cytopathologists as a quality control measure: a follow-up study. *Am J Clin Pathol* 2007;128:653–656.
- 20 Abele JS, Levine RA: Diagnostic criteria and risk-adapted approach to indeterminate thyroid cytodiagnosis. *Cancer Cytopathol* 2010;118:415–422.
- 21 Koen TM, Mody DR, Scheiber-Pacht M, Fairley T, Thrall MJ: Limiting the use of atypical/inconclusive as a category in nongynecologic cytology specimens. *Arch Pathol Lab Med* 2010;134:1016–1019.