Implications of a Suspicious Afirma Test Result in Thyroid Fine-Needle Aspiration Cytology: An Institutional Experience

Ricardo R. Lastra, MD; Michelle R. Pramick, MD; Cody J. Crammer, BS; Virginia A. LiVolsi, MD; and Zubair W. Baloch, MD, PhD

BACKGROUND: Fine-needle aspiration (FNA) biopsy is the most frequently used method for thyroid nodule evaluation. However, up to 30% of cases are considered indeterminate. Surgery is typically recommended for these cases, but up to two-thirds of indeterminate cases are found to be benign. The Afirma test is used for the preoperative classification of thyroid nodules with indeterminate cytology. This study reviews the authors’ institutional experience with Afirma.

METHODS: A cohort of 132 cases of thyroid FNA with Afirma testing was selected from the study files and relevant information was recorded and analyzed. At the study institution, Afirma is mainly performed on atypia of undetermined significance (AUS)/follicular lesion of undetermined significance (FLUS) cases when diagnosed as such on repeat FNA.

RESULTS: The cohort included 98 female (74%) and 34 male (26%) patients. Cytology diagnosis was AUS/FLUS in 68 cases (51.5%), follicular neoplasm (FN) in 39 cases (29.5%), and FN with oncocytic features (FNOF) in 25 cases (19.0%). Of the FNOF cases with suspicious Afirma findings, 2 (15%) were malignant and 11 (85%) were benign. Of the FN cases with suspicious Afirma findings, 9 (53%) were malignant and 8 (47%) were benign. Of the AUS/FLUS cases with suspicious Afirma findings, 10 (63%) were malignant and 6 (37%) were benign. CONCLUSIONS: The Afirma classifier is a useful tool to aid in the distinction of cytologically indeterminate nodules. Performing Afirma in cases diagnosed as AUS/FLUS on repeat FNA would increase the positive predictive value, thereby minimizing the number of benign cases referred to surgery. Results of the Afirma test could be limited in cases diagnosed as FNOF. Cancer (Cancer Cytopathol) 2014;000:000-000. © 2014 American Cancer Society.

KEY WORDS: thyroid; Afirma; fine-needle aspiration; cytology; molecular.

INTRODUCTION
Fine-needle aspiration (FNA) biopsy is the most widely used method for the screening and diagnosis of thyroid nodules. In most instances, the FNA results are either definitely benign or malignant, allowing for appropriate action. However, between 15% to 30% of cases are considered indeterminate.1 As per the Bethesda System for Reporting Thyroid Cytopathology, indeterminate diagnostic categories include “atypia of undetermined significance (AUS)/follicular lesion of undetermined significance (FLUS)” “follicular neoplasm or suspicious for follicular neoplasm (FN),” and “suspicious for malignancy.”2,3 Because the possibility of malignancy cannot be excluded in indeterminate cases, surgery is typically recommended; however, up to two-thirds of these cases are found to be benign at the time of surgical resection.1,4 This results in unnecessary surgical procedures, complications, and morbidity, as well as medical expenses.
Molecular testing of indeterminate thyroid nodules as an adjunct to thyroid cytology has been proven to be of significant value. Moreover, it is considered a cost-reducing proposition, particularly due to the reduction in 2-stage thyroidectomy (lobectomy followed by completion thyroidectomy). The Afirma gene expression classifier is a proprietary diagnostic test developed by Veracyte Inc. (South San Francisco, Calif) for the preoperative identification of benign thyroid nodules with a cytology diagnosis of AUS/FLUS and FN. The assay analyzes the mRNA expression of 167 genes to identify the signature of a benign thyroid nodule. Based on this information, it classifies nodules as either “benign” (95% negative predictive value [NPV] for aspirates classified as AUS/FLUS and 94% for aspirates classified as FN) or “suspicious” for malignancy (>50% risk for malignancy). Thus, the Afirma gene expression classifier can be used to identify a subpopulation of patients with a low probability of malignancy in a population of patients for whom surgery would be recommended.

Although the usefulness of Afirma appears to be evident in the preoperative triage of thyroid nodules with indeterminate cytology, to our knowledge only a limited number of studies have evaluated its use in the clinical setting. In the current study, we reviewed our experience with the Afirma gene expression classifier when applied to thyroid nodules with an indeterminate cytology in an attempt to validate its usefulness at the study institution.

MATERIALS AND METHODS

The institutional pathology files at the Hospital of the University of Pennsylvania were retrospectively searched for all thyroid FNA cases with a diagnosis of AUS/FLUS, FN, or follicular neoplasm with oncocytic features (FNOF) and concurrent Afirma results from February 2011 to January 2014. At this institution, a total of 5 cytopathologists evaluate thyroid FNA specimens. The overall laboratory rates regarding thyroid nodule FNAs is as follows: nondiagnostic, 1%; benign, 60%; AUS/FLUS, 11.5%; FN, 12.2%; FNOF, 5.2%; suspicious for malignancy, 3%; and malignant, 7.1%. Ultrasonographic features considered as “worrisome,” and that triggered biopsy included hypervascularity, calcifications, and increasing size (as compared with prior studies).

The collected data points included patient demographic data (age and sex), as well as the ultrasonographic characteristics of the biopsied nodule, including size, laterality, echogenicity, vascularity, and the presence of calcifications. The prominent cytomorphologic features that resulted in a nodule being diagnosed as AUS/FLUS were recorded and categorized into 4 categories as follows: category 1: the presence of follicular cell groups with crowding and overlapping in a benign-appearing background; category 2: the presence of follicular cell groups with crowding and overlapping in a scant specimen; category 3: the presence of follicular cell groups with crowding and overlapping in a background of lymphocytic thyroiditis; and category 4: other (which included focally present nuclear atypia, nuclear grooves, and elongation, but which were not enough for a diagnosis of “suspicious for malignancy” or “papillary thyroid carcinoma”). When either partial/hemithyroidectomy or total thyroidectomy were performed subsequent to the FNA, the final surgical pathology diagnosis of the biopsied nodule was also collected.

In each case, FNA was performed under ultrasound guidance by a radiologist and/or endocrinologist, most often with a 27-gauge needle. The initial 2 passes were evaluated on-site by a cytopathologist, who confirmed the presence of adequate material for diagnosis and rendered a preliminary diagnostic impression. Up to 2 additional passes were obtained in any given case, if it was believed that the original attempts were unsuccessful.

Based on on-site interpretation of the cytopathologists and the clinical and radiologic impression of the performing clinician, additional pass(es) was performed exclusively for Afirma testing. As a general rule in the study institution, on-site diagnosis of AUS/FLUS in a repeat FNA of a nodule initially diagnosed as AUS/FLUS on first FNA triggers an additional pass for Afirma testing. No set criteria exist for the Afirma testing of nodules with a preliminary diagnosis FN or FNOF. However, an on-site or final diagnosis of FN or FNOF in cases of small and radiologically benign-appearing nodules usually also leads to Afirma testing.

Statistical Analysis

Statistical analyses were performed using the chi-square test. A P value of <.05 was considered to be statistically significant.

RESULTS

The study cohort comprised 132 patients, including 98 females (74%) and 34 males (26%) (male:female ratio of
The age ranged from 19 to 81 years, with a mean age at the time of FNA of 54.4 years. The nodule size ranged from 0.8 cm to 7.0 cm, with an average nodule size of 2.12 cm; 68 (52%) nodules were located in the right lobe, 53 (40%) in the left lobe, and 11 (8%) in the isthmus.

The cytology diagnosis was AUS/FLUS in 68 cases (51.5%), FN in 39 cases (29.5%), and FNOF in 25 cases (19.0%). As mentioned earlier, based on the cytomorphologic features, the AUS/FLUS cases were divided as follows: 46 (68%) cases in category 1, 5 (7%) cases in category 2, 10 (15%) cases in category 3, and 3 (4%) cases in category 4. No description of cytomorphologic features was present in 4 (6%) of the AUS/FLUS cases.

Afirma results were benign in 70 cases (53%) (Fig. 1) and suspicious in 62 cases (47%) (48 of which had surgical follow-up). Of the 62 cases with suspicious Afirma results, the cytology diagnoses were AUS/FLUS in 23 cases (37%), 18 of which had surgical follow-up; FN in 22 cases (35.5%), 17 of which had surgical follow-up; and FNOF in 17 cases (27.5%), 13 of which had surgical follow-up.

**Afirma Results Versus Surgical Follow-Up**

Of the 48 cases with suspicious Afirma results and available surgical follow-up, 22 cases (46%) were diagnosed as malignant and 26 cases (54%) were diagnosed as benign at the time of surgical resection. The malignant diagnoses were 16 cases of follicular variant of papillary thyroid carcinoma (FV-PTC), 3 cases of classic PTC (cPTC), and 3 cases of follicular carcinoma (FC). The benign diagnoses were 9 cases of follicular adenoma (FA), 14 cases of follicular adenoma with oncocyctic features/Hurthle cell adenoma (HA), and 3 cases of adenomatoid nodule (AN) (Table 1).

Surgical follow-up was available in only 2 cases with benign Afirma results. Both of these cases had been diagnosed as FNOF on cytology and proved to be benign at the time of surgical resection. The final surgical pathology diagnosis in these 2 cases was FA and AN.

**Afirma Results Versus Cytology Diagnosis**

Of the 18 cases of AUS/FLUS with suspicious Afirma results and available surgical follow-up, 11 (61%) were diagnosed as malignant at the time of surgical resection (8 FV-PTC cases, 2 cPTC cases, and 1 FC case) (Figs. 2A and 2B) and 7 (39%) were diagnosed as benign (5 FA cases, 1 HA case, and 1 AN case). Of the 17 cases of FN with suspicious Afirma results and available surgical follow-up, 9 (53%) were diagnosed as malignant (all FV-PTC cases) and 8 (47%) were diagnosed as benign (4 FA cases, 3 HA cases, and 1 AN case). Of the 13 cases diagnosed as FNOF with suspicious Afirma results and available surgical follow-up, 2 (15%) were diagnosed as malignant (both FC cases) and 11 (85%) were diagnosed as benign (10 HA cases and 1 AN case) (Figs. 3A-3C) (Table 2). The difference in the rate of malignancy between suspicious Afirma cases diagnosed as AUS/FLUS on repeat cytology and those diagnosed as FN was not statistically significant ($P=.48$). However, this difference was statistically significant when comparing cases diagnosed as FNOF with cases diagnosed as AUS/FLUS ($P=.007$) and FN ($P=.034$).

Combining cases diagnosed as FN and FNOF into a single category (“follicular neoplasm/suspicious for...
“follicular neoplasm” as per the Bethesda classification scheme, without subclassification based on the presence of oncocytic features) yielded 30 cases with suspicious Afirma results and available surgical follow-up. The malignancy rate in this combined category was found to be 37% (11 cases) (9 FV-PTC cases and 2 FC cases).

In relation to the subclassification of AUS/FLUS cases based on the cytomorphologic features, the majority of cases were included in the first category (46 cases; 68%). The relatively low number of cases in categories 2, 3, and 4 precludes meaningful evaluation or comparison among these groups. Available follow-up information in relation to this subgrouping is included in Table 3.

Figure 2. (A) Smear and liquid-based preparations (Inset) of a case with atypia of undetermined significance/follicular lesion of undetermined significance cytology (category 1) and a suspicious Afirma result are shown. Follow-up was follicular variant of papillary thyroid carcinoma (FV-PTC) (Papanicolaou stain, × 63; Inset: Papanicolaou stain, × 40). (B) Surgical resection specimen of the case in panel A is shown demonstrating macrofollicular FV-PTC (H & E, × 40).

Figure 3. (A) Cytology diagnosis of follicular neoplasm with oncocytic features (FNOF) is shown. The subsequent Afirma diagnosis was suspicious (Diff-Quik, × 10). (B) Cytology diagnosis of FNOF is shown. The subsequent Afirma diagnosis was suspicious (Papanicolaou stain, × 20). (C) Resection of an FNOF case with a suspicious Afirma diagnosis is shown. The final surgical pathology diagnosis was Hurthle cell adenoma. Note the thick capsule surrounding the nodule (H & E, × 10).
The ultrasound examination revealed “nodular calcifications” in 21 of the 132 cases (16%). Of these, the Afirma result was suspicious in 13 cases and benign in 8 cases (corresponding to 21% and 11% of the total cases with suspicious and benign Afirma results, respectively). Surgical follow-up was available in 10 of the 13 nodules with calcifications and a suspicious Afirma result.

Of these, 5 (50%) were diagnosed as malignant (3 cases as FV-PTC, 1 case as cPTC, and 1 case as FC) and 5 (50%) as benign (3 cases as FA, 1 case as HA, and 1 case as AN). In contrast, 17 of 38 cases with suspicious Afirma results but without reported calcifications by ultrasound were found to be malignant at the time of surgical resection (45%). The difference in the rate of malignancy in nodules with a suspicious Afirma result, with and without calcifications, was not statistically significant ($P = 0.76$) (Table 4).

Hypervascularity was reported in 18 of the 132 cases (14%). Of these, 6 cases were suspicious and 12 were benign by Afirma (corresponding to 10% and 17% of the total cases with suspicious and benign Afirma results, respectively). Surgical follow-up was available in 3 of the 6 cases with hypervascularity and suspicious Afirma results. Malignancy was reported in 1 case (33%) (FC) and 2 cases (67%) were benign (both HA).

### Table 2. Diagnosis on Surgical Pathology Follow-Up of Cases With Suspicious Afirma Results According to Cytology Diagnosis (n=48)

<table>
<thead>
<tr>
<th>Category</th>
<th>Malignant</th>
<th>Benign</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUS/FLUS (n=18)</td>
<td>11 (61%)</td>
<td>7 (39%)</td>
</tr>
<tr>
<td>FN (n=17)</td>
<td>9 (53%)</td>
<td>8 (47%)</td>
</tr>
<tr>
<td>FNOF (n=13)</td>
<td>2 (15%)</td>
<td>11 (85%)</td>
</tr>
</tbody>
</table>

**Abbreviations:** AUS/FLUS, atypia of undetermined significance/follicular lesion of undetermined significance; FN, follicular neoplasm; FNOF, follicular neoplasm with oncocytic features; FV-PTC, follicular variant of papillary thyroid carcinoma; PTC, papillary thyroid carcinoma.

### Table 3. Distribution of AUS/FLUS Cases According to Cytomorphologic Subcategorization (n=68)

<table>
<thead>
<tr>
<th>Category</th>
<th>No. of Suspicious Cases (% of Cases Within Category)</th>
<th>No. of Suspicious Cases With Follow-Up</th>
<th>Diagnosis of Suspicious Cases With Follow-Up</th>
</tr>
</thead>
</table>
| 1 (46 cases; 68%) | 16 (35%) | 12 | Malignant (8 cases; 67%)
| 2 (5 cases; 7%) | 2 (40%) | 2 | Malignant (1 case; 50%)
| 3 (10 cases; 15%) | 4 (40%) | 3 | Malignant (1 case; 33%)
| 4 (3 cases; 4%) | 1 (33%) | 1 | Malignant (1 case; 100%)

**Abbreviations:** AUS/FLUS, atypia of undetermined significance/follicular lesion of undetermined significance; cPTC, classic papillary thyroid carcinoma; FC, follicular carcinoma; FV-PTC, follicular variant of papillary thyroid carcinoma.

### Table 4. Incidence of Malignancy in Nodules With Suspicious Afirma Results Based on Ultrasonographic Findings

<table>
<thead>
<tr>
<th>Ultrasonographic Characteristics</th>
<th>Malignant</th>
<th>Benign</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calcifications Present (n=10)</td>
<td>5 (50%)</td>
<td>5 (50%)</td>
<td>.76</td>
</tr>
<tr>
<td>Absent (n=38)</td>
<td>17 (45%)</td>
<td>21 (55%)</td>
<td></td>
</tr>
<tr>
<td>Echogenicity Hyperechoic (n=9)</td>
<td>6 (67%)</td>
<td>3 (33%)</td>
<td>.18</td>
</tr>
<tr>
<td>Hypoechoic (n=11)</td>
<td>4 (36%)</td>
<td>7 (64%)</td>
<td></td>
</tr>
</tbody>
</table>

**Afirma Results Versus Ultrasonographic Characteristics of the Biopsied Nodule**

The ultrasound examination revealed “nodular calcifications” in 21 of the 132 cases (16%). Of these, the Afirma result was suspicious in 13 cases and benign in 8 cases (corresponding to 21% and 11% of the total cases with suspicious and benign Afirma results, respectively). Surgical follow-up was available in 10 of the 13 nodules with calcifications and a suspicious Afirma result.
Of the 132 cases, 20 (15%) were described as hyper-echoic and 35 (24%) as hypoechoic. Among the hyper-echoic nodules, 10 were diagnosed as suspicious and 10 as benign by Afirma (corresponding to 16% and 14% of the total cases with suspicious and benign Afirma results, respectively). Surgical follow-up was available in 9 of the 10 hyperechoic nodules with suspicious Afirma results. Of these, 6 (67%) were diagnosed as malignant (5 cases of FV-PTC and 1 case of FC) and 3 cases (33%) were diagnosed as benign (2 cases of FA and 1 case of HA). In comparison, 15 of the 35 hypoechoic nodules were diagnosed as suspicious and 20 as benign by Afirma (corresponding to 24% and 29% of the total cases with suspicious and benign Afirma results, respectively). Surgical follow-up was available in 11 of the 15 hypoechoic nodules with suspicious Afirma results. Of these, 4 (36%) were diagnosed as malignant (3 cases of FV-PTC and 1 case of FC) and 7 (64%) as benign (1 case of FA, 5 cases of HA, and 1 case of AN). The difference in the rate of malignancy in hyper-echoic and hypoechoic nodules with a suspicious Afirma result was not statistically significant (P=.18) (Table 4).

DISCUSSION

Well-defined cytomorphologic criteria have been established for the interpretation of thyroid FNA specimens by the Bethesda System for Reporting Thyroid Cytopathology. According to this tiered classification scheme, thyroid nodules can be broadly categorized into benign, suspicious for malignancy/malignant, and indeterminate. The risk of malignancy in thyroid nodules diagnosed as “benign” is reported to be between 0% to 3%, and most will require only clinical and radiologic follow-up; conversely, nodules diagnosed as “suspicious for malignancy” and “malignant” carry a risk of malignancy of 60% to 75% and 97% to 99%, respectively, and the widely accepted management is total or near-total thyroidectomy. The indeterminate category includes “AUS/FLUS” and “FN/suspicious for follicular neoplasm,” with an associated risk of malignancy of 5% to 15% and 15% to 30%, respectively. In cases classified as AUS/FLUS, repeat FNA after 3 months is recommended; however, if the diagnosis remains as AUS/FLUS after repeat FNA, surgical lobectomy is generally recommended. Surgical lobectomy is recommended for cases diagnosed as FN. Evidently, performing surgery in these persistent AUS/FLUS and FN cases results in partial thyroidectomies for benign reasons in a majority of cases, due to the low risk of malignancy.

In recent years, a significant portion of the literature related to thyroid cytopathology has been focused on attempting to further characterize these cytologically indeterminate nodules by means of adjunct molecular analysis. Up to two-thirds of follicular cell-derived thyroid carcinomas are reported to harbor somatic genetic mutations. Hence, testing for a panel of mutations that include BRAF V600E, point mutations in NRAS codon 61, HRAS codon 61, and KRAF codons 12/13, as well as rearrangements in RET/PTC1, RET/PTC3, and PAX8/PPARγ has been described as a useful and cost effective tool to aid in this distinction. It is important to note that although most reports agree with the potential usefulness of molecular testing in these indeterminate cases, some authors still suggest that molecular testing of thyroid nodules does not significantly alter the surgical management of these patients.

Attempts to develop gene expression classifiers to aid in the preoperative characterization of thyroid nodules have been made. These tests need to have a high sensitivity and a high NPV to be effective in reducing unnecessary surgery; to date, limited sensitivity has restricted their use in clinical practice. The Afirma gene expression classifier uses an algorithm based on the expression of 167 genes to classify thyroid FNA specimens as either benign or suspicious. Of these 167 genes, 142 are involved in the main classifier (ie, benign vs malignant) and the remaining 25 genes filter out rare neoplasms. The NPV for the gene expression classifier for FNA aspirates classified as AUS/FLUS and FN is 95% and 94%, respectively, with a sensitivity of 90% for both AUS/FLUS and FN cases. Overall, 68% of cases with indeterminate cytology and suspicious Afirma results are found to be benign at the time of surgical resection.

There are 2 significant factors to consider when evaluating the results of the current study: the effect of the presence of oncocytic cells on the Afirma results and the performance of Afirma testing on cases diagnosed as AUS/FLUS on first FNA compared with those that persist as such on repeat FNA.

The positive predictive value (PPV) for malignancy of Afirma in cases diagnosed as “suspicious for follicular neoplasm,” without subclassifying into FN or FNOF, is 37% in the current study; this is similar to that noted in the published literature. It is interesting to note that
when the same group is divided into FN and FNOF, the PPV is 53% and 15%, respectively, which is a statistically significant difference. It has been shown that microRNA profiles of conventional and oncocyotic follicular carcinomas are substantially different. Whether this is related to the presence of abundant mitochondrial DNA in oncocyotic cells or simply to selective expression of certain genes in oncocyotic tumors is unclear. Either way, this difference appears to cause an increased false-positive “suspicious rate” by Afirma in cases of FNOF.

The reported literature indicates a PPV for malignancy of 38% to 48% in cases classified as AUS/FLUS on first FNA and suspicious Afirma results. However, the incidence of malignancy in these AUS/FLUS cases without Afirma testing is reported to be between 10% and 30%, corresponding to that described by the Bethesda System for Reporting Thyroid Cytopathology. As mentioned earlier, Afirma analysis in the study institution is mainly performed on repeat AUS/FLUS cases. The incidence of malignancy at the time of thyroidectomy in cases diagnosed as AUS/FLUS on repeat FNA, but on which Afirma testing has not been performed, is 39.4%. From this data, it is evident that cases diagnosed as AUS/FLUS on repeat FNA have a significantly higher incidence of malignancy than those diagnosed as AUS/FLUS on first FNA. Hence, we have performed Afirma testing on a selected cohort of AUS/FLUS cases with a higher malignancy rate (ie, selection bias), and it is therefore not surprising that our PPV for malignancy in AUS/FLUS cases with a suspicious Afirma result (61%) is significantly higher than that reported in the literature (38%-48%).

As mentioned previously, the high NPV of Afirma testing leads to the clinical follow-up of indeterminate nodules with a benign result, rather than surgery. Hence, we do not have significant surgical pathology follow-up data in these cases, and a direct comparison between the “gold standard” surgical pathology diagnosis in benign and suspicious Afirma cases cannot be made. However, at the study institution, thyroid nodules diagnosed as FN and resected without prior Afirma testing have a malignancy rate of <30%, which is consistent with the reported malignancy rate of FN by the Bethesda System. Although the lack of follow-up in the benign Afirma cases represents a limitation of the current study, the comparison of the malignancy rate in FN cases with a suspicious Afirma result and those without Afirma testing (53% as shown in the current study vs <30%) demonstrates the usefulness of this test in the evaluation of these nodules.

With regard to ultrasonographic findings, we found no significant differences in the rate of malignancy when comparing suspicious Afirma results in hyperechoic and hypoechoic nodules, the presence or absence of calcifications, or hypervascularity.

The results of the current study demonstrate that the Afirma gene expression classifier is a useful tool to aid in the distinction of cytologically indeterminate nodules. Limiting Afirma testing only to those cases diagnosed as AUS/FLUS on repeat FNA would increase the PPV of the test by selecting a cohort of cases with a higher incidence of malignancy. This may avoid performing a costly molecular test in a considerable number of cases that are diagnosed as AUS/FLUS on first FNA. Clinicians and pathologists should understand the limitations of Afirma testing in nodules diagnosed as FNOF; however, we do not have enough data to fully support not performing Afirma testing in these cases.

FUNDING SUPPORT
No specific funding was disclosed.

CONFLICT OF INTEREST DISCLOSURES
Dr. LiVolsi previously worked as a histology consultant for Veracyte Inc.

REFERENCES