Outcome of Thyroid Gene Expression Classifier Testing in Clinical Practice

Robert L. Witt, MD, FACS

Objectives/Hypothesis: Gene expression classifiers can safely reduce diagnostic thyroid surgery for fine-needle aspiration cytology (FNAC) indeterminate thyroid nodules.

Study Design: Retrospective review, single-institution, single-practice surgeon.

Methods: Three-year retrospective review of indeterminate FNAC that went on to gene expression classifier testing.

Results: A total of 520 patients met American Thyroid Association guideline criteria for surgeon-performed ultrasound-guided FNAC for a thyroid nodule with on-site cytopathology. The indeterminate (Bethesda III or IV) FNAC rate was 9%. Prevalence of malignancy in FNAC indeterminate was 21%. Thirty-two cases went on to gene expression classifier testing. Fourteen were benign, 15 suspicious, and three with no result.

Conclusions: Benign gene expression classifier testing had an estimated negative predictive value of 100% during the study period. These patients have been observed for a mean and median duration of 14 and 7 months, respectively. In this small series, 14 of 29 patients with indeterminate FNAC were spared diagnostic surgery.

Key Words: Thyroid, fine needle aspiration cytology, gene expression testing.

Level of Evidence: 4

INTRODUCTION

The Afirma gene expression classifier (GEC) (Vera-cyte Inc., South San Francisco, CA) represents a “rule out” approach to the cytologically indeterminate thyroid nodule based on its high sensitivity (if sensitivity is high, a negative test is very likely to represent a true lack of disease) and high negative predictive value (NPV) (the fraction of all benign results that are correct). Many of the studies involving Afirma are from academic centers, whereas some are industry sponsored; thus, a call for real-world experience has been recommended.1 Fine-needle aspiration cytology (FNAC) cannot make a definitive diagnosis in about 25% of thyroid nodules.2 Approximately 70% to 80% of thyroid surgeries for indeterminate thyroid nodules performed solely for diagnostic purposes yield benign surgical pathology from the index nodule, and in retrospect, many of these surgeries were unnecessary. Unnecessary surgeries are costly and result in unnecessary anxiety and risk for the patient, including surgical complications, unplanned hospital readmission, and though rare, perioperative death. The indeterminate FNAC categories routinely tested by the Afirma GEC include Bethesda III (atypia of undetermined significance/follicular lesion of undetermined significance [AUS/FLUS]) and Bethesda IV (follicular neoplasms/suspicious for follicular neoplasm [FN/SFN]).3

Bayes theorem (Fig. 1) in the calculation of NPV depends not only on sensitivity and specificity, but also disease prevalence. The NPV heavily depends on disease prevalence rates that are often completely unknown, not well characterized, and are subject to change. Clinicians have been cautioned to consider how the Afirma GEC will perform in their unique patient populations and to know the individual practice and/or institutional prevalence of cancer for indeterminate thyroid nodules.

A personalized medicine approach includes considering not only prevalence of disease in the population but also risk of disease in the unique patient including history, exam, and ultrasound. Individual clinical factors, such as the nodule ultrasound characteristics, help recharacterize the individual patient’s risk. Ultrasound findings that are not high risk are present in the majority of Bethesda III and IV nodules4,5 and are associated with a malignancy risk of 20% or less,6 and therefore predict Afirma GEC NPV of >95%. High-risk ultrasound findings are associated with a risk of malignancy of 70% or more6 and an expected Afirma GEC NPV of <95%. Additionally, in any individual person with a cytologically indeterminate nodule that is truly malignant, the chance of the Afirma GEC falsely yielding a benign result is limited to 1 sensitivity or 10%.7 The sensitivity of the Afirma GEC is 90%.7

This investigation is the first, to the author’s knowledge, to simultaneously use the Afirma GEC outside of an academic institution, utilize local cytopathologists for...
FNAC interpretation, and calculate cancer prevalence in an individual practice. It is the first to report on practice deployment of GEC in the otolaryngology literature for indeterminate FNAC.

The hypothesis of this study is that the Afirma GEC in a non–industry-sponsored, real-world thyroid surgery practice with calculated cancer prevalence for indeterminate thyroid nodules (Bethesda III and IV) could safely reduce diagnostic thyroid surgery.

MATERIALS AND METHODS

Christiana Care Institutional Review Board approval was obtained for this study. American Thyroid Association (ATA) guidelines§ were used for patient selection for surgeon-performed thyroid ultrasound-guided FNAC. All FNACs from February 2012 to February 2015 were reviewed. Excluded from the series were patients with a family or personal history of thyroid cancer, prior radiation therapy to the neck, compressive symptoms, nodules >4 cm, vocal fold immobility, or suspicious cervical lymphadenopathy. Patient’s status of surgery or clinical observation was determined through April 1, 2015.

A highly experienced cytopathologist was present for on-site staining (Diff Quick; Dade Diagnostics, Aguada, PR), adequacy evaluation, and molecular testing triage. This was followed by Pap stain, cell block, and ThinPrep test. A transverse ultrasound-guided technique using a 27-gauge needle was used in thyroid FNAC. The surgeon performed the FNA while both the surgeon and cytopathologist confirmed that the needle tip was within the selected nodule by ultrasound guidance in all cases. An immediate cytopathological interpretation for satisfactory FNAC yield was made. Cytology that was determined unsatisfactory by the cytopathologist on first pass was immediately repeated with additional passes until adequacy for diagnosis was obtained. If the FNAC was indeterminate, Bethesda Classification III or IV,³ two additional passes were obtained with ultrasound guidance and placed in the FNAProtect solution for Afirma GEC testing.

Patients with benign GEC are followed every 6 months with surgeon-performed ultrasound and surveyed for change of size of the nodules. FNAC cancer prevalence for Bethesda III and IV is calculated on the basis of the authors practice-specific known final histology (surgical truth) for indeterminate FNAC that underwent diagnostic thyroid surgery.

RESULTS

Five hundred twenty patients met ATA guidelines for surgeon-performed FNAC of a thyroid nodule. A total of 47/520 (9%) patients had indeterminate (Bethesda III or IV) FNAC, and 32/47 (68%) agreed to Afirma GEC testing (Fig. 2). Three of the 32 patients (10%), all Bethesda III, had no result (inadequate RNA content for evaluation). Two of the three patients with no result went on to diagnostic thyroid surgery, and both had final benign pathology. One of the three patients elected observation and has had no increase in size of their thyroid nodule in over 2 years. The three no-result patients and their third party payers did not receive a request of payment from Afirma. Twenty-nine had a result of Afirma GEC of benign or suspicious. There were 14/29 Afirma GEC benign patients (48%) (seven patients Bethesda III and seven patients Bethesda IV), and all were followed with clinical and ultrasound at 6-month intervals. The mean size of the nodules followed with benign GEC was 18 mm (range, 1.0–4.0 cm). To date, none of the Afirma benign nodules have undergone surgery and have been under close clinical observation with surgeon (author)-performed ultrasound every 6 months for a mean and median duration of 14 and 7 months, respectively, since Afirma GEC testing (range, 55–1,126 days). None of the Afirma benign nodules have increased in size, and none have required repeat FNAC biopsy. Assuming all unoperated Afirma GEC benign nodules are truly benign and remain benign, the estimated NPV is 100% for the study period.

Fifteen of 29 Bethesda III/IV samples (52%) were Afirma GEC suspicious and went on to diagnostic surgery. The mean size of nodules with suspicious GEC was 19 mm (range, 1.0–4.0 cm). Six of these 15 (40%) were malignant upon surgical pathology. The prevalence of cancer in GEC suspicious Bethesda III and IV nodules was 3/10 (30%) and 3/5 (60%), respectively. The overall calculated prevalence of cancer for indeterminate FNAC in the first author’s individual practice is 6/29 (21%) for the study period.

DISCUSSION

The Afirma GEC, clinically validated in the landmark study by Alexander et al.,¹ uses expression of 167 genes (mRNA transcripts), 142 genes in the main classifier (benign or suspicious) and 25 genes that filter out rare neoplasms. In this second prospective, double-blinded, multicenter validation study, 49 clinical sites and 4,812 fine-needle aspiration aspirates from thyroid nodules 1 cm or larger investigated 265 cytologically indeterminate nodules; the NPVs for FLUS, follicular neoplasm, or suspicious cytologic features were 95%, 94%, and 85%, respectively.¹ This was based on a cancer prevalence in

\[
NPV = \frac{\text{Specificity}(1-\text{Prevalence})}{(\text{Specificity}(1-\text{Prevalence}) + (1-\text{Sensitivity})\text{Prevalence})}
\]

Fig. 1. Bayes theorem in the calculation of negative predictive value.

Fig. 2. Flow diagram of the outcome of thyroid gene expression classifier testing. FNAC = fine-needle aspiration cytology; GEC = gene expression classifier; US = ultrasound.
TABLE I.
Non–Industry-Sponsored Published Clinical Experiences and Pooled Analysis of the Afirma Gene Expression Classifier

<table>
<thead>
<tr>
<th>Study</th>
<th>Afirma GEC Suspicious, n (%)</th>
<th>Afirma GEC Benign, n (%)</th>
<th>Afirma GEC Benign Operative Rate, n (%)</th>
<th>Practice Setting</th>
<th>Cytology Interpretation Setting</th>
<th>NPV*</th>
</tr>
</thead>
<tbody>
<tr>
<td>McIver et al. 201411</td>
<td>36 (65%)</td>
<td>19 (35%)</td>
<td>2 (6.3%)</td>
<td>Community hospital</td>
<td>TCP</td>
<td>94.7%</td>
</tr>
<tr>
<td>Alexander et al. 201414</td>
<td>148 (46%)</td>
<td>174 (54%)</td>
<td>11 (6.3%)</td>
<td>Multicenter, academic center</td>
<td>Local</td>
<td>99.4%</td>
</tr>
<tr>
<td>McIver et al. 2014†12</td>
<td>44 (73%)</td>
<td>16 (27%)</td>
<td>4 (25%)</td>
<td>Academic center</td>
<td>Local</td>
<td>93.8%</td>
</tr>
<tr>
<td>Lastra et al. 201415</td>
<td>62 (47%)</td>
<td>70 (53%)</td>
<td>2 (2.9%)</td>
<td>Academic center</td>
<td>Local</td>
<td>100%</td>
</tr>
<tr>
<td>Sullivan et al. 201416</td>
<td>7 (54%)</td>
<td>6 (46%)</td>
<td>1 (16.7%)</td>
<td>Academic center</td>
<td>Local</td>
<td>100%</td>
</tr>
<tr>
<td>Current study</td>
<td>15 (52%)</td>
<td>14 (48%)</td>
<td>0 (0%)</td>
<td>Community hospital</td>
<td>Local</td>
<td>100%</td>
</tr>
<tr>
<td>Pooled analysis, n = 611</td>
<td>312 (51%)</td>
<td>289 (49%)</td>
<td>22 (7.4%)</td>
<td>Community hospital</td>
<td>Local</td>
<td>99.0%</td>
</tr>
</tbody>
</table>

*NPV calculated assuming unoperated Afirma GEC benign nodules are truly benign.
†Excluding <1 cm false-negative nodule as described in this text.
‡Includes 13 Bethesda V nodules.

GEC = gene expression classifier; NPV = negative predictive value; TCP = Thyroid Cytopathology Partners.

The prevalence of cancer for indeterminate FNAC in the author’s individual practice was 6/29 (21%). In this single surgeon Veracyte-enabled practice, where FNAC was determined locally, the estimated NPV was 100% for a benign GEC for the duration of clinical and ultrasound follow-up. Surgical truth was not established for indeterminate Afirma GEC benign nodules (nor is surgical truth generally established for FNAC benign nodules that carry a 5% false negative rate9). The 21% prevalence of malignancy in this study suggests that finding many malignancies among the Afirma GEC benign unoperated nodules is unlikely. These data for the study period are consistent with the reported high sensitivity and NPV of the test. Fourteen of 29 (48%) patients with indeterminate FNAC had a benign Afirma GEC result. Thus, almost half of the tested patients avoided diagnostic thyroid surgery and have been followed carefully with clinical exams and ultrasound. Some patients have been in follow-up for longer than 2 to 3 years. These findings support that GEc is safe and durable for an FNAC indeterminate benign Afirma GEC result.

The Afirma GEC proprietary algorithm is based on a gene expression “signature” and classifies a sample as either (molecularly) benign or suspicious. Benign results for FLUS and FN, incorporating clinical judgment, typically undergo clinical and ultrasound follow-up, and Afirma-suspicious results (those with not clearly benign molecular signatures) typically undergo diagnostic thyroid lobectomy, with possible total thyroidectomy.8 The Afirma GEC is positioned in clinical care as a rule-out test given its high NPV, and not as a rule-in test given its modest positive predictive value (PPV). Clinically, this is implemented by applying the Afirma GEC to cytologically indeterminate nodules that are headed to the operating room only due to the perceived risk of malignancy, but for which the treating clinician would be willing to observe the nodule without surgery if that risk could be substantially reduced by a negative result from a rule-out test (e.g., high NPV). Conversely, the Afirma GEC has a lower specificity (52%), meaning that about half of the truly benign nodules are expected to be called GEC suspicious (and be falsely positive). Given this lower specificity, this test is not a good rule-in test, which relies on PPV. Although false-positive results among patients otherwise headed to surgery are not a major consequence, a nodule that is not otherwise headed to the operating room should typically not have that fate changed simply by an Afirma GEC suspicious result.

The clinical utility of the Afirma GEC is the avoidance of diagnostic surgery in selected cases. The first study to measure this impact was an industry-sponsored series of 368 patients treated by 51 physicians with benign Afirma GEC results, which demonstrated a 10-fold reduction in surgery rates for cytologically indeterminate nodules, from a previous historical rate of 74% to just 7.6% for Afirma GEC benign results. Of 2,040 Afirma GEC results, 52.3% were benign.10 Access to the Afirma GEC is via one of two models. In the first model, physicians outside of academia or integrated healthcare systems are generally required to submit the FNAC for cytologic evaluation by Veracyte’s independent industry partner, Thyroid Cytopathology Partners (TCP) (Austin, TX). Although this group reports high-volume cytological interpretation experience and a rate of indeterminate readings consistent with the medical literature, the prevalence of malignancy among their indeterminate cases is largely unknown. For this reason, it has been questioned1 whether the NPV of a TCP indeterminate GEC benign result will be similar to the 95% and 94% rates published in the validation trial by Alexander et al.8 for AUS/FLUS and FN/SFN, respectively.

McIver et al.11 an independent academic Veracyte-enabled group that performed their own cytopathology, report an NPV of 94%, with a prevalence of malignancy of 17%. Only 27% of their patients were reported with a benign GEC result. They express concern that...
widespread implementation of molecular tests could result in increased costs, and that only one patient in every four would avoid surgery.

The only community hospital–published experience to date that had cytology interpreted by TCP was published by Harrell and Bimston.12 The authors report their community clinical practice experience in 645 FNAs performed over 27 months and yielding 58 (9%) specimens deemed cytologically indeterminate by TCP. Based on the prevalence of cancer in this specific series, the authors found, in operated Afirma GEC suspicious and benign cases, an estimated cancer prevalence of 33% among their cytologically indeterminate cases. The authors calculated their practice sensitivity and specificity (true negatives divided by the sum of true negatives + false positives) based on surgical (truth) results alone, but because most Afirma GEC benign cases were not operated and thus not included in their specificity calculation, true-negative cases were largely undercounted, and their calculated specificity was predictably very low, as was their NPV. Using these values and practice cancer prevalence estimate, the authors calculated a best case NPV of 89.6%. Among their 19 Afirma GEC benign cases >1 cm in size, one was found to be malignant among the four operated cases. Alternatively, should the remaining 15 nonoperated GEC benign cases be truly benign, then their practice NPV would be 18/19 (94.7%), and very similar to the experience described in the two Veracyte-sponsored validation studies,9,13 this report, and others (Table I).11,14–16

To date, several non–industry-sponsored reports from academic groups who read their own cytology have described their clinical experience with the Afirma GEC.11,14–16 This is the first description from a community-hospital setting where cytopathology readings were performed locally. High-risk history, and physical and ultrasound findings, were assessed for appropriate selection of patients with indeterminate FNAC for GEC. This report supports that in this practice setting, and with cytopathology read locally, nearly 50% of Afirma GEC results are benign, and these patients are clinically followed in lieu of diagnostic surgery. This report is also in support of the National Comprehensive Cancer Network Thyroid Carcinoma Guideline, which states that cytologically indeterminate thyroid nodules determined to have a malignancy risk of 5% or less with a molecular test can be clinically observed.17

CONCLUSION
Benign Afirma GEC testing in this investigation had an estimated NPV of 100% during the study period. These patients have been observed for a mean and median duration of 14 and 7 months, respectively. In this small series, 14 of 29 patients tested with indeterminate FNAC were spared unnecessary surgery for indeterminate thyroid nodules.

BIBLIOGRAPHY