**Medical Policy**

**Title:** Molecular Markers in Fine Needle Aspirates of the Thyroid

### Populations
- **Individuals:** With thyroid nodule(s) and indeterminate findings on fine needle aspirate

### Interventions
- **Interventions of interest are:**
  - Fine needle aspirate sample testing with the Afirma Gene Expression Classifier to predict benignancy

### Comparators
- **Comparators of interest are:**
  - Surgical biopsy

### Outcomes
- Relevant outcomes include:
  - Disease-specific survival
  - Test accuracy
  - Test validity
  - Morbid events
  - Resource utilization

### Populations
- **Individuals:** With thyroid nodule(s) and indeterminate findings on fine needle aspirate

### Interventions
- **Interventions of interest are:**
  - Fine needle aspirate sample testing with molecular markers to predict malignancy

### Comparators
- **Comparators of interest are:**
  - Surgical management based on clinicopathologic risk factors

### Outcomes
- Relevant outcomes include:
  - Disease-specific survival
  - Test accuracy
  - Test validity
  - Morbid events
  - Resource utilization
DESCRIPTION
Cytologic examination of fine needle aspirate (FNA) samples from a thyroid lesion to identify which patients need thyroid resection has diagnostic limitations. Assays using molecular markers have been developed in an attempt to improve the accuracy of thyroid FNA biopsies.

OBJECTIVE
The objective of this evidence review is to evaluate the analytic validity, clinical validity, and clinical utility of molecular markers in fine needle aspiration samples from thyroid lesions.

BACKGROUND
Fine Needle Aspirates of the Thyroid
Thyroid nodules are common, present in 5% to 7% of the U.S. adult population. Most are benign, and most cases of thyroid cancer are curable by surgery when detected early. Fine needle aspirate (FNA) samples of the thyroid is currently the most accurate procedure to distinguish benign thyroid lesions and malignant ones, reducing the rate of unnecessary thyroid surgery for patients with benign nodules and triaging patients with thyroid cancer to appropriate surgery.

About 60% to 70% of thyroid nodules are classified cytologically as benign, and 4% to 10% of nodules are cytologically deemed malignant. However, the remaining 20% to 30% have equivocal findings, usually due to overlapping cytologic features between benign and malignant nodules; these nodules usually require surgery for a final diagnosis. Thyroid FNA cytology is classified by Bethesda System criteria into the following groups: nondiagnostic; benign; follicular lesion of undetermined significance (FLUS) or atypia of undetermined significance (AUS); follicular neoplasm (or suspicious for follicular neoplasm); suspicious for malignancy; and malignant. Lesions with FNA cytology in the AUS or FLUS or follicular neoplasm categories are often considered indeterminate.

There is some individualization of management for patients with FNA-indeterminate nodules, but many patients will require a surgical biopsy, typically thyroid lobectomy, with intraoperative pathology. Consultation would typically be the next step in diagnosis. Approximately 80% of patients with indeterminate cytology undergo surgical resection; postoperative evaluation has revealed a malignancy rate ranging from 6% to 30%, making this a clinical process with very low specificity. Thus, if analysis of FNA samples could reliably identify the risk of malignancy as low, there is potential for patients to avoid surgical biopsy.

Preoperative planning of optimal surgical management in patients with equivocal cytologic results is challenging, because different thyroid malignancies require different surgical procedures (eg, unilateral lobectomy vs total or subtotal thyroidectomy with or without lymph node dissection) depending on several factors, including histologic...
subtype and risk-stratification strategies (tumor size, patient age). If a diagnosis cannot be made intraoperatively, a lobectomy is typically performed, and, if on postoperative histology the lesion is malignant, a second surgical intervention may be necessary for completion thyroidectomy.

**Thyroid Cancer**

Most thyroid cancers originate from thyroid follicular cells and include well-differentiated papillary thyroid carcinoma (PTC; 80% of all thyroid cancers) and follicular carcinoma (15%). Poorly differentiated and anaplastic thyroid carcinomas are uncommon and can arise de novo or from preexisting well-differentiated papillary or follicular carcinomas. Medullary thyroid carcinoma originates from parafollicular or C cells, and accounts for about 3% of all thyroid cancers.

The diagnosis of malignancy in the case of PTC is primarily based on cytologic features. If FNA in a case of PTC is indeterminate, surgical biopsy with intraoperative pathology consultation is most often diagnostic, although its efficacy and therefore its use will vary across institutions, surgeons, and pathologists.

For follicular carcinoma, the presence of invasion of the tumor capsule or of blood vessels is diagnostic and cannot be determined by cytology, because tissue sampling is necessary to observe these histologic characteristics. Intraoperative diagnosis of follicular carcinoma is challenging and often not feasible, because extensive sampling of the tumor and capsule is usually necessary and performed on postoperative permanent sections. New approaches for improving the diagnostic accuracy of thyroid FNA include mutation analysis for somatic genetic alterations, to more accurately classify which patients need to proceed to surgery (and may include the extent of surgery necessary), and a gene expression classifier to identify patients who do not need surgery and can be safely followed.

**Mutations Associated With Thyroid Cancer**

Various mutations have been discovered in thyroid cancer. The most common 4 gene mutations that carry the highest impact on tumor diagnosis and prognosis are BRAF and RAS point mutations and RET/PTC and PAX8/PPARγ rearrangements.

Papillary carcinomas carry point mutations of the BRAF and RAS genes, as well as RET/PTC and TRK rearrangements, all of which are able to activate the mitogen-activated protein kinase pathway. These mutually exclusive mutations are found in more than 70% of papillary carcinomas. BRAF mutations are highly specific for PTC. Follicular carcinomas harbor either RAS mutations or PAX8/PPARγ rearrangement. These mutations identified in 70% to 75% of follicular carcinomas. Genetic alterations involving the PI3K/AKT signaling pathway also occur in thyroid tumors, although they are rare in well-differentiated thyroid cancers and have higher prevalence in less differentiated thyroid carcinomas. Additional mutations known to occur in poorly differentiated and anaplastic
cancerous involve the TP53 and CTNNB1 genes. Medullary carcinomas, which can be familial or sporadic, frequently possess point mutations located in the RET gene.

Studies have evaluated the association between various genes and cancer phenotype in individuals with diagnosed thyroid cancer.4-6

**Molecular Diagnostic Testing**

**Mutation and Rearrangement Testing**

Point mutations in specific genes, including BRAF, RAS, and RET, and evaluation for rearrangements associated with thyroid cancers can be accomplished by with Sanger sequencing or pyrosequencing or with real-time polymerase chain reaction (PCR) of single or multiple genes or by next-generation sequencing (NGS) panels. Panels of tests for mutations associated with thyroid cancer, with varying compositions, are also available. For example, Quest Diagnostics offers a Thyroid Cancer Mutation Panel, which includes BRAF and RAS mutation analysis and testing for RET/PTC and PAX8/PPARγ rearrangements.

The ThyroSeq® v.2 Next Generation Sequencing panel (CBLPath, Ocala, FL) is a NGS sequencing panel of more than 60 genes. According to the CBLPath’s website, the test is indicated when FNA cytology indicates atypia of uncertain significance or follicular lesion of undetermined significance, follicular neoplasm or suspicious for follicular neoplasm, or suspicious for malignancy.7 In particular, it has been evaluated in patients with follicular neoplasm and/or suspicious for follicular neoplasm on FNA as a test to increase both sensitivity and specificity for cancer diagnosis.

The ThyGenX™ Thyroid Oncogene Panel (formerly miRInform® Thyroid; Interpace Diagnostics, Parsippany, NJ; testing done at Asuragen Clinical Laboratory) is another NGS panel designed to assess patients with indeterminate thyroid FNA results. It includes sequencing of 8 genes associated with papillary thyroid carcinoma and follicular carcinomas.

**Gene Expression Profiling**

Genetic alterations associated with thyroid cancer can be assessed using gene expression profiling, which refers to analysis of messenger RNA (mRNA) expression levels of many genes simultaneously. Several gene expression profiling tests are now available to biologically stratify tissue from thyroid nodules.

The Afirma® Gene Expression Classifier (Afirma GEC; Veracyte, South San Francisco, CA) analyzes the expression of 142 different genes to determine patterns associated with benign findings on surgical biopsy. It is designed to evaluate thyroid nodules that have an “indeterminate” classification on FNA as a method to select patients (“rule out”) who are at low risk for cancer.
Veracyte also markets 2 “malignancy classifiers” that use mRNA expression-based classification to evaluate for *BRAF* mutations (Afirma BRAF) or mutations associated with medullary thyroid carcinoma (Afirma MTC). In a description of the Afirma BRAF test, the following have been proposed as benefits of the mRNA-based expression test for *BRAF* mutations: (1) PCR-based methods may have low sensitivity, requiring that a large proportion of the nodule have a relevant mutation; (2) testing for only 1 mutation may not detect patients with low-frequency mutations that result in the same pattern of pathway activation; and (3) PCR-based approaches with high analytic sensitivity may require a large amount of DNA that is difficult to isolate from small FNA samples.\(^8\) Afirma MTC is an option when Afirma GEC is ordered for thyroid nodules with an “intermediate” classification on FNA, and can also be used for thyroid nodules with “malignant” or “suspicious” results on Afirma GEC. Afirma BRAF is designed to be used for nodules with “suspicious” results on Afirma GEC.

ThyraMIR™ (Interpace Diagnostics, Parsippany, NJ) is a micro-RNA expression–based classifier intended for use in thyroid nodules with indeterminate cytology on FNA. Other gene expression profiles have been reported in investigational settings, but have not been widely validated or used commercially (eg, Barros-Filho et al [2015],\(^9\) Zheng et al [2015]\(^10\)); these are not addressed in this review.

**REGULATORY STATUS**

Clinical laboratories may develop and validate tests in-house and market them as a laboratory service; laboratory-developed tests (LDTs) must meet the general regulatory standards of the Clinical Laboratory Improvement Amendments (CLIA). Thyroid mutation testing and gene expression classifiers are available under the auspices of CLIA. Laboratories that offer LDTs must be licensed by CLIA for high-complexity testing. To date, the U.S. Food and Drug Administration (FDA) has chosen not to require any regulatory review of this test.

In 2013, the THxID™-*BRAF* kit (bioMérieux, Marcy l'Etoile, France), an in vitro diagnostic device, was approved by FDA through the premarket approval process to assess specific *BRAF* mutations in melanoma tissue via real-time polymerase chain reaction. However, there are currently no diagnostic tests for thyroid cancer mutation analysis with approval from FDA.
**POLICY**

A. The use of the Afirma Gene Expression Classifier in fine needle aspirates of the thyroid that are cytologically considered to be indeterminate (follicular lesion of undetermined significance or follicular neoplasm) may be considered **medically necessary** in patients who have **ALL** of the following characteristics:

1. Thyroid nodules without strong clinical or radiologic findings suggestive of malignancy, **AND**
2. In whom surgical decision making would be affected by test results.

B. Mutation analysis in fine needle aspirates of the thyroid is **experimental / investigational**.

C. Gene expression classifiers in fine needle aspirates of the thyroid not meeting criteria outlined above are considered **experimental / investigational**.

**POLICY GUIDELINES**

In patients who do not undergo surgical biopsy or thyroidectomy on the basis of gene expression classifier results, regular active surveillance is indicated.

**RATIONALE**

The most recent policy update includes a review of the literature through March 30, 2016 (see Appendix Table 1 for genetic testing categories).

The evaluation of a diagnostic or prognostic typically includes an assessment of the test’s analytic validity (technical performance), clinical validity (diagnostic or prognostic accuracy), and clinical utility (whether the use of the test is associated with improvements in patient outcomes).

**MOLECULAR MARKERS TO PREDICT BENIGNANCY**

**Clinical Context**

Molecular markers to predict benignancy are tests designed to have a high negative predictive value (NPV). The focus of this section is the Afirma Gene Expression Classifier (GEC), which is proposed as a risk stratifying test for patients who have indeterminate findings based on fine needle aspirate (FNA). These patients presently proceed to surgical resection. The purpose of the test is to select patients at low risk of malignancy who could avoid unnecessary surgery.

**Analytic Validity**

Walsh et al verified the analytic performance of the Afirma GEC in the classification of cytologically indeterminate FNAs from thyroid nodules.\(^{11}\) The analytic performance studies were designed to characterize the stability of the RNA in the aspirates during collection and shipment, the analytical sensitivity and specificity, and the assay performance studies including intranodule, intraassay, interassay, and interlaboratory reproducibility. Concordance of the GEC calls was 100% for samples tested under different shipping conditions, 97.2% across different RNA input amounts, 100% under different dilutions with normal tissue, and 96% across different genomic
DNA contamination amounts. The intra-assay, interassay, interlaboratory, and intranodule concordances were 93.9%, 97%, 100%, and 95%, respectively. The authors concluded that the analytic sensitivity and specificity, robustness, and quality control of the GEC were successfully verified.

**Clinical Validity**

Chudova et al described the development and initial clinical validation of a version of the Afirma GEC. The classifier was trained on 178 retrospectively identified surgical thyroid specimens, which represented a variety of malignant and benign disorders, and separately on a set of 137 FNA samples with known surgical pathology. The classifier was developed with the objective of achieving a NPV specificity of 95% and a specificity of 70%. The tissue-trained classifier was tested on an independent sample of 48 FNAs (24 with indeterminate cytopathology, 24 with a mix of malignant and benign cytopathology). The FNA-trained classifier was tested separately on the same sample of 48 FNAs. In the 24 samples with indeterminate cytopathology, sensitivity and specificity were 100% (95% confidence interval [CI], 64% to 100%) and 73.3% (95% CI, 49% to 89%), respectively.

**Prospective Clinical Validation**

Alexander et al reported on a 19-month, prospective, multicenter (49 academic and community) sites, study of the Afirma GEC. A total of 4812 nodules were screened for inclusion with centralized cytopathology. Local pathology reports of the cytopathologic diagnosis were collected for all patients, and reports without a definitive benign or malignant diagnosis at the local site were reviewed by 3 expert cytopathologists, who reclassified them as atypical, follicular neoplasm, or suspicious for a follicular neoplasm, or suspicious for malignancy. Of all nodules screened, 577 (12%) were considered indeterminate after central review, and 413 of those had tissue pathology available.

The GEC used in the Alexander study was retrained on a set of 468 samples, comprised of 220 banked tissue samples, 14 ex vivo operative FNA samples, and 234 prospective clinical FNA samples. The authors noted that 25 of those prospective clinical FNA samples were derived from the 413 samples described above.

After exclusion of the 25 used for test validation and those without a valid GEC result, 265 FNA samples were evaluated with the Afirma GEC. Of the 265 samples, 85 were malignant; the GEC correctly identified 78 of the 85 as suspicious (92% sensitivity; 95% CI, 84% to 97%). Specificity was 52% (95% CI, 44% to 59%). NPV ranged from 85% for “suspicious cytopathologic findings” to 95% for “atypia of undetermined clinical significance.” There were 7 FNAs with false-negative results, 6 of which were thought to be due to hypocellular aspirate specimens.

**Retrospective Clinical Validation**

In 2014, Alexander et al reported results from a multicenter retrospective analysis of 339 thyroid nodules that underwent Afirma GEC testing for indeterminate cytology on FNA (atypia of undetermined significance [AUS] or follicular lesion of undetermined significance [FLUS], follicular neoplasm, or suspicious for malignancy) at 5 academic medical centers. Most nodules sent for GEC testing were AUS or FLUS or follicular neoplasm. The distribution of GEC testing results for each cytopathologic classification is shown in Table 1.
Table 1: GEC Testing Results From Alexander et al (2014)\textsuperscript{13}

<table>
<thead>
<tr>
<th>Cytologic Classification</th>
<th>N</th>
<th>GEC Testing Results, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Benign</td>
</tr>
<tr>
<td>Atypia or follicular lesion of undetermined significance</td>
<td>165</td>
<td>91 (55%)</td>
</tr>
<tr>
<td>Follicular neoplasm</td>
<td>161</td>
<td>79 (49%)</td>
</tr>
<tr>
<td>Suspicious for malignancy</td>
<td>13</td>
<td>4 (31%)</td>
</tr>
<tr>
<td>Total</td>
<td>339</td>
<td>174</td>
</tr>
</tbody>
</table>

A subset of patients whose nodules underwent GEC testing had a subsequent thyroid resection. Among 148 cases with suspicious Afirma GEC findings, surgery (thyroid resection) was recommended for 141 (95%). For the 174 cases with benign Afirma GEC findings, surgery was recommended for 4 (2%; p<0.01). On the assumption that, absent the GEC results, thyroid surgery would be recommended for patients with cytologically indeterminate FNA results, the authors reported that GEC results altered management in 50% of patients. Table 2 shows thyroidectomy biopsy results for the subset of patients in Table 1 who underwent surgery.

Table 2: Thyroidectomy Results From Alexander et al (2014)\textsuperscript{13}

<table>
<thead>
<tr>
<th>GEC Results</th>
<th>N</th>
<th>Surgery Recommended, n</th>
<th>Surgery Completed, n</th>
<th>Pathology Malignant, n (% of those with completed surgery)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suspicious</td>
<td>148</td>
<td>141</td>
<td>121</td>
<td>53 (44%)</td>
</tr>
<tr>
<td>Benign</td>
<td>174</td>
<td>4</td>
<td>11</td>
<td>1 (9%)</td>
</tr>
</tbody>
</table>

Seventeen patients who had indeterminate cytology, benign Afirma GEC results, and did not undergo surgery had follow-up beyond 1 year. Of those, 3 patients had surgery to remove the nodule because of compressive symptoms (n=2) or nodule growth (n=1); all nodules were benign on final histology. The remaining 14 patients had ongoing follow-up with ultrasound with no ongoing evidence of malignancy. The study demonstrated site-to-site variation in the proportion of samples that were GEC benign. A benign GEC result did not completely rule out malignant pathology. Long-term follow-up was available for only a small proportion of patients with benign GEC findings who did not undergo surgery.

In 2016, Santhanam et al conducted a meta-analysis of studies reporting on the performance of the Afirma GEC in cytologically indeterminate nodules.\textsuperscript{14} Seven studies met inclusion criteria, which required that studies reported on the use of the Afirma GEC in nodules found indeterminate on FNA (including AUS or FLUS; suspicious for follicular or Hürthle cell neoplasms; suspicious for malignancy), and thyroidectomy was performed as a reference standard in at least the cases where the index test was suspicious. All studies were judged to be at low risk of bias for patient selection and most for GEC test selection, whereas the risk of bias in the final histopathology was low in 3 studies, unclear in 3 studies, and high in 1 study. In the pooled cohort, the prevalence of malignancy was 37.1%. The main results of the analysis are summarized in Table 3.

Table 3: Pooled GEC Performance From Santhanam et al (2016)\textsuperscript{14}

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Point Estimate</th>
<th>95% Confidence Interval</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>95.7%</td>
<td>92.2% to 97.9%</td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td>30.5%</td>
<td>26.0% to 35.3%</td>
<td></td>
</tr>
<tr>
<td>Positive likelihood ratio</td>
<td>1.20</td>
<td>0.996 to 1.44</td>
<td></td>
</tr>
<tr>
<td>Negative likelihood ratio</td>
<td>0.2</td>
<td>0.11 to 0.36</td>
<td></td>
</tr>
<tr>
<td>Diagnostic odds ratio</td>
<td>7.9</td>
<td>4.1 to 15.1</td>
<td></td>
</tr>
</tbody>
</table>
Retrospective single-center studies, including Harrell and Bimston (2014), Lastra et al (2014), McIver et al (2014), Yang et al (2016) have reported the diagnostic accuracy of the Afirma GEC (see Table 4). All studies were subject to ascertainment bias, because a large proportion of individuals with Afirma benign reports did not undergo surgery, which made determining the sensitivity and specificity of the GEC assay impossible. However, the rates of malignancy among patients with Afirma benign results who did undergo surgery were consistently low. One exception is the study by Harrell and Bimston (2014); it may be reflective of a higher-than-usual overall rate of malignancy in patients with indeterminate FNA results. One additional publication (Celik et al, 2015) reported on Afirma GEC testing, but included in its sample population individuals with benign and suspicious cytology on FNA, who are not the targeted population of the test.

Table 4: Single-Center Studies Reporting Afirma GEC Results

<table>
<thead>
<tr>
<th>Study (Year)</th>
<th>Indeterminate FNA Samples, n (%)</th>
<th>Afirma GEC Test Result</th>
<th>N</th>
<th>With Thyroidectomy, n</th>
<th>With Malignancy on Thyroidectomy, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harrell and Bimston (2014)</td>
<td>58 AUS/FLUS or FN</td>
<td>Suspicious Benign</td>
<td>36</td>
<td>30</td>
<td>21</td>
</tr>
<tr>
<td>Lastra et al (2014)</td>
<td>69 (51.5%) AUS/FLUS</td>
<td>Suspicious Benign</td>
<td>62</td>
<td>48</td>
<td>22</td>
</tr>
<tr>
<td>McIver et al (2014)</td>
<td>12 (11.4%) AUS/FLUS</td>
<td>Suspicious Benign</td>
<td>44</td>
<td>32</td>
<td>5</td>
</tr>
<tr>
<td>Yang et al (2016)</td>
<td>93 (88.6%) FN/HCN</td>
<td>Suspicious Benign</td>
<td>16</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Witt et al (2016)</td>
<td>47 AUS/FLUS or SFN/FN (32 with GEC attempted)</td>
<td>Suspicious Benign</td>
<td>15</td>
<td>15</td>
<td>6</td>
</tr>
</tbody>
</table>

AUS: atypia of undetermined significance; FLUS: follicular lesion of undetermined significance; FN: follicular neoplasm; FNA: fine needle aspirates; FNOF: follicular neoplasm with oncocytic features; HCN: Hürthle cell neoplasm; NA: not applicable; SFN: suspicious for follicular neoplasm.

a Two samples inadequate due to low mRNA content.
b GEC results were available for 60 subjects.
c Three samples were inadequate.

There are limited data on the true negative rates of individuals with indeterminate FNA cytology and Afirma GEC benign results. Supportive information on the accuracy Afirma GEC benign results can be obtained from studies that have reported on long-term follow-up of individuals with indeterminate FNA cytology and Afirma GEC benign results. Angell et al retrospectively compared clinical outcomes for individuals with indeterminate FNA cytology and Afirma GEC benign results with individuals to cytologically benign nodules. A total of 95 cytologically indeterminate/Afirma GEC benign nodules in 90 patients were compared with 1224 cytologically benign nodules identified from a single-center, prospectively collected database. Five nodules in the cytologically indeterminate were resected; of the remaining 90 nodules, 58 (64.4%) had follow-up ultrasound available at a median of 13 months postdiagnosis. When nodule growth was defined by a volume increase of 50% or more, 17.2% cytologically indeterminate/Afirma GEC benign were considered to have grown compared with 13.8% of cytologically benign nodules (p=0.44). Surgical resection was more common in cytologically indeterminate/Afirma GEC benign nodules (13.8% vs 0.9%, p<0.001).
Clinical Utility

No evidence directly demonstrating improved outcomes in patients managed with the Afirma GEC was identified. Therefore, a chain of indirect evidence was developed, which addresses 2 key questions:

1. Does use of the Afirma GEC in individuals with cytologically indeterminate thyroid nodules change clinical management (in this case, reduced thyroid resections)?
2. Do those management changes improve outcomes?

Changes in Management

The clinical setting in which the Afirma GEC is meant to be used is well-defined: individuals with AUS or FLUS or follicular neoplasm or who are suspicious for follicular neoplasm on FNA who do not have other indications for thyroid resection (ie, in whom the GEC results would play a role in surgical decision making).

Decision impact studies, most often reporting on clinical management changes but not on outcomes after surgical decisions were made, have suggested that, in at least some cases, surgical decision making changed. These studies are described briefly.

Duick et al reported on the impact of Afirma GEC test results on physician and patient decision making to operate on thyroid nodules with indeterminate cytology and Afirma GEC benign results in a sample of 395 nodules from 368 patients. Surgery was performed in 7.6% of the patients with indeterminate cytology and a benign GEC result, less than the historical rate of thyroid resection (74%) in patients with indeterminate cytology.

The 2014 study by Alexander et al provided evidence on clinical management changes for patients with indeterminate thyroid nodules tested with Afirma GEC. While the treating physicians presumably elected to obtain the GEC testing with the intent of altering management recommendations, the magnitude of the difference in surgical recommendations for patients with GEC suspicious or benign results was large.

Two studies (Aragon Han et al [2014], Noureldine et al [2015]) evaluated the potential for the Afirma GEC test to change surgical decision making by comparing actual surgical decision making when Afirma GEC was used to predict surgical decision making based on a management algorithm. In both, surgical decision making was estimated to change in at least some proportion of patients (10%-15%).

Improved Outcomes

A simplified decision model was developed for use with Afirma GEC in individuals with cytologically indeterminate FNA samples. It is shown in Appendix 2. It is assumed that when Afirma GEC is not used, patients with cytologically indeterminate FNA results undergo thyroid resection. When Afirma GEC is used, those with Afirma suspicious lesions undergo resection, while those who have Afirma benign lesions do not. In this case, compared to the standard care plan, some patients without cancer will have avoided a biopsy, which is weighed against the small increase in missed cancers in patients who had cancer but tested as Afirma benign.

Assuming that the rate of cancer in cytologically indeterminate thyroid nodules is approximately 20%, in the standard care plan, 80% of patients with cytologically indeterminate FNA samples will undergo an unnecessary biopsy. Applying the test characteristic values from Alexander et al
(2012), it is estimated that approximately 1.6% of individuals with a true cancer would be missed, but approximately 38%, instead of 80%, would undergo unneeded surgery.

Whether the tradeoff between avoiding unneeded surgeries and the potential for missed cancer is worthwhile depends, in part, on patient and physician preferences. However, some general statements may be made by considering the consequences of a missed malignancy and the consequences of unnecessary surgery. Most missed malignancies will be papillary thyroid carcinomas (PTCs), which have an indolent course. Thyroid nodules are amenable to ongoing surveillance (clinical, ultrasound, and with repeat FNAs), with minimal morbidity.

Thyroid resection is a relatively low risk surgery. However, consequences of surgery can be profound. Patients who undergo a hemi- or subtotal thyroidectomy have a risk of recurrent laryngeal nerve damage and parathyroid gland loss.

At present, the existing standard of care for thyroid nodules is based on intervention that is stratified by FNA cytology results, which are grouped into categories with differing prognosis. Avoiding an invasive surgery in situations where patients are at very low likelihood of having an invasive tumor is likely beneficial, given the small but potentially significant adverse effects associated with thyroidectomy or hemithyroidectomy. The alternative to surgical biopsy in the low-risk population is ongoing active surveillance.

Section Summary: Molecular Markers to Predict Benignancy

In 1 multicenter validation study, the Afirma GEC test has been reported to have a high NPV (range, 90%-95%). These results are supported by an earlier development and clinical validation study (Chudova et al), but the classifiers used in the 2 studies do not appear to be identical. In an additional multicenter and multiple single-center studies, there is suggestive evidence that rates of malignancy are low in Afirma benign patients, but the exact NPV is unknown. The available evidence has suggested that physician decision making about surgery is altered by GEC results, although long-term follow-up of patients with thyroid nodules who avoided surgery based on GEC results is limited. An indirect chain of evidence can be constructed to establish the potential for clinical utility with GEC testing in cytologically indeterminate lesions, but with only 1 study with the marketed test reporting a true NPV, the clinical validity is uncertain.

Molecular Markers to Predict Malignancy

Clinical Context

Molecular markers associated with malignancy in thyroid nodules are generally used as “rule-in” tests to identify cancer or tumors with more aggressive behavior. The purpose of the test is in patients with cytologically indeterminate FNA results is to determine the presence of a certain mutation or to have a high enough pretest probability of cancer would change the surgical approach or some other aspect of management.

For thyroid nodules that have indeterminate findings on FNA cytology, a surgical biopsy with intraoperative pathology consultation would typically be the next step. Following a diagnosis of a thyroid malignancy, preoperative surgical planning with regard to the extent of thyroid resection and lymph node dissection is an important consideration. Conventional factors determining biopsy strategy and surgical resection strategy include histologic subtype and risk stratification based on factors such as tumor size and patient age.
For analytic and clinical validity, point mutation and rearrangement testing to predict malignancy is discussed separately from gene expression classifiers because the development and validation of the classifiers is unique to each specific marketed test. However, the proposed clinical utility is similar.

**Point Mutation and Rearrangement Testing**

**Analytic Validity**

Point mutations in specific genes associated with thyroid cancer (e.g., the *BRAF* V600E gene) and the detection of genetic rearrangements associated with thyroid cancer (e.g., the *RET/PTC* rearrangement) are typically detected with Sanger sequencing or next-generation sequencing (NGS) methods. In the case of testing for gene mutations associated with thyroid cancer malignancy, analytic validity refers to a test’s technical accuracy in detecting a mutation that is present or in excluding a mutation that is absent. The real-time polymerase chain reaction (PCR)–based methods are generally considered to have high accuracy. For example, Smith et al reported on the technical performance characteristics for *BRAF* mutation detection by qualitative PCR in thyroid FNA samples with high within- and between-run reproducibility.

NGS is expected to have high accuracy for detecting a mutation that is present. However, with increasing numbers of tested mutations, there is increased risk of detection of variants of uncertain significance (VUS). The VUS rate for currently available NGS panels for thyroid cancer is not well-characterized. Nikiforova et al described the development and validation of a multigene NGS panel for thyroid cancer, the ThyroSeq panel. They developed a custom library of gene sequence variants based on mutations previously reported in the literature. The assay demonstrated 100% accuracy in evaluating samples of 15 thyroid tumors and 3 cell lines with known genetic alterations and 15 DNA samples with no mutations. In analysis of 229 DNA samples from frozen tissues (n=105), formalin-fixed, paraffin-embedded (FFPE) tissues (n=72), and FNAs (n=52), the panel identified mutations in 19 (70%) of 27 of classic PTCs, 25 (83%) of 30 follicular variant PTCs, 14 (78%) of 18 conventional, and 7 (39%) of 18 Hürthle cell carcinomas, 3 (30%) of 10 poorly differentiated carcinomas, 20 (74%) of 27 anaplastic thyroid carcinomas, and 11 (73%) of 15 medullary thyroid carcinomas. Of 83 benign nodules, 5 (6%) were positive for mutations.

**Clinical Validity**

A number of studies have evaluated whether testing for point mutations or gene fusions (either single mutation or panels of mutations) can be used to improve the sensitivity and specificity for diagnosing indeterminate FNA of the thyroid, with the goal of identifying mutations that predict malignancy in FNA samples.

**Mutations Association With Malignancy**

In 2015, Fnais et al conducted a systematic review and meta-analysis of studies reporting on the test accuracy of *BRAF* mutation testing in the diagnosis of PTC. The review included 47 studies with 9924 FNA samples. For all cytologically indeterminate nodules, the pooled sensitivity estimate for *BRAF* mutation testing was 31% (95% CI, 6% to 56%). Among nodules suspicious for malignancy on FNA, the pooled sensitivity estimate for *BRAF* mutation testing was 52% (95% CI, 39% to 64%; I²=77%).

Ferraz et al evaluated 20 publications that reported on the type and number of mutations in cases of FNA of the thyroid diagnosed as indeterminate and compared the results with final
histology after surgical resection. Sixteen studies analyzed 1 mutation (eg, \textit{BRAF} mutation or \textit{RET/PTC} rearrangement) and 4 studies analyzed a panel of several mutations (\textit{BRAF} and \textit{RAS} mutations, \textit{RET/PTC} and \textit{PAX8/PPAR\gamma} rearrangements). The detection of a mutation in a histologically (surgically resected) benign thyroid lesion was categorized as a false-positive case, detecting no mutation in an FNA sample from a histologically benign surgical sample was considered a true negative, and finding no mutation in a histologically malignant lesion was categorized as a false negative. Based on 4 studies that examined a panel of mutations, there was a broad sensitivity range (38%-85.7%; mean, 63.7%), a mean specificity of 98% (range, 95%-100%), mean false-positive rate of 1.25% (range, 0%-4%), and mean false-negative rate of 9% (range, 1%-21%). Based on 2 studies that examined \textit{RET/PTC} rearrangements, mean sensitivity was 55% (range, 50%-60%), specificity 100%, a false-positive rate of 0% and mean false-negative rate 3.5% (91%-6%). Based on 3 studies that examined \textit{BRAF} mutations, mean sensitivity was 13% (range, 0%-37.5%), mean specificity was 92.3% (range, 75%-100%), mean false-positive rate was 0.5% (0%-1%), and mean false-negative rate was 6% (range, 3%-12%). Authors concluded that testing for a panel of mutations leads to an improvement in the sensitivity and specificity for indeterminate FNA of the thyroid but that further standardizations and further molecular markers are needed before broad application of molecular FNA cytology for the diagnosis of thyroid nodules.

The largest body of literature on mutation testing for prediction of malignancy in indeterminate thyroid nodules is related to the development a NGS panel (ThyroSeq) that includes \textit{BRAF}, \textit{RAS}, \textit{RET/PTC}, or \textit{PAX8/PPAR\gamma}. Studies that address these panels are described in more detail; studies that include subsets of these mutations or additional mutations are summarized in the following section.

Nikiforov et al prospectively tested a panel of mutations (\textit{BRAF}, \textit{RAS}, \textit{RET/PTC}, \textit{PAX8/PPAR\gamma}) in 470 FNA samples of thyroid nodules from 328 consecutive patients. Mutational status correlated with cytology and either surgical pathology diagnosis or follow-up (mean, 34 months). Forty patients were excluded for poor quality specimens or loss to follow-up. Sixty-nine patients (with 86 thyroid FNA samples) underwent surgery soon after completing the cytologic evaluation; preoperative cytologic diagnosis was: positive for malignancy in 22 samples, indeterminate (including atypical and suspicious for malignancy) in 52 samples, and negative for malignancy in 12 samples. By FNA, 32 mutations were found (18 \textit{BRAF}, 8 \textit{RAS}, 5 \textit{RET/PTC}, 1 \textit{PAX8/PPAR\gamma}); after surgery, 31 (97%) mutation-positive nodules were diagnosed as malignant on pathologic examination and 1 (3%) as a benign tumor. Thirteen of the 32 mutation-positive FNA samples had a definitive cytologic diagnosis of malignancy, whereas the rest were either indeterminate or negative for malignancy.

Of the remaining 219 patients, 147 (229 FNAs) who did not undergo surgery were followed using serial ultrasound with no change in the nodule status (124 patients) or using repeated FNA with cytology negative for malignancy (23 patients) and no mutation found in the FNA material. These nodules were considered negative for malignancy. The remaining 72 patients who were initially in the follow-up group underwent subsequent surgery. Combining all 3 groups, the specificity for malignancy was high (99.7%), but the sensitivity of the molecular test alone was not (62%).

Ohori et al performed mutation screening in 117 FNA samples classified as AUS or FLUS. \textit{BRAF}, \textit{RAS}, \textit{RET/PTC}, or \textit{PAX8/PPAR\gamma} mutations were detected in 10% of this category. The screening demonstrated that the probability of having a malignancy in this cytology category together with
a detection of 1 of the somatic mutations investigated was 100%, whereas the probability of having a thyroid malignancy without a mutation detected was 7.6%.

In 2011, Nikiforov et al reported results of a prospective study that assessed the clinical validity of a panel of mutations to predict the likelihood of malignancy in thyroid nodules found indeterminate on FNA. The authors included 1056 consecutive samples with indeterminate cytology on FNA that underwent mutation testing, with 967 of those adequate for molecular analysis (653 AUS or FLUS; 247 follicular or Hürthle cell neoplasms or suspicious for follicular neoplasm; 67 suspicious for malignant cells). (One hundred seventeen of the samples were included in the Ohori et al study described above and summarized in Table 5). Eighty-seven BRAF, RAS, RET/PTC, or PAX8/PPARγ mutations were detected. At analysis, 479 patients had undergone thyroidectomy for further evaluation, providing a histopathologic diagnosis for 513 samples. The presence of a mutation had low sensitivity for predicting malignant histology (63%, 57%, 68% for samples with AUS or FLUS, follicular or Hürthle cell neoplasms or suspicious for follicular neoplasm, and suspicious for malignant cells on cytology, respectively), but high specificity (99%, 97%, 96%, respectively). The NPV for the mutation analysis results was 94%, 86%, and 72% for samples with AUS or FLUS, follicular or Hürthle cell neoplasms or suspicious for follicular neoplasm, and suspicious for malignant cells on cytology, respectively. The authors concluded that mutation analysis might be useful in surgical planning, such as determining whether patients should undergo a thyroid lobectomy or a complete thyroidectomy as a first surgery.

In a subsequent study, Nikiforov et al evaluated the accuracy of an NGS panel that included tests for point mutations in 13 genes and for 42 types of gene fusions (ThyroSeq v2 NGS panel) in a series of 143 consecutive thyroid FNA samples with a cytologic diagnosis of follicular or Hürthle cell neoplasm /suspicious for follicular or Hürthle cell neoplasm. Molecular testing was retrospectively performed for 91 samples and prospectively performed for the remaining 52. The prevalence of cancer on histology was 27.5% and 26.9% in the retrospective and prospective cohorts, respectively. In the retrospective cohort, of the 25 malignant nodules, 22 were PTCs and 3 were follicular thyroid carcinomas (FTCs). In the prospective cohort, of the 14 malignant nodules, 11 were PTCs and 3 were FTCs. The performance of the ThyroSeq in both cohorts is shown in Table 5.

<p>| Table 5: Performance of ThyroSeq Panel in Nikiforov et al (2014) and (2015) |</p>
<table>
<thead>
<tr>
<th>-------------------------------------------------</th>
<th>-------------------------------------------------</th>
<th>-------------------------------------------------</th>
<th>-------------------------------------------------</th>
<th>-------------------------------------------------</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>Retrospective (n=91)</td>
<td>Prospective (n=52)</td>
<td>Overall (N=143)</td>
<td>Patients With Known Outcome (n=98)</td>
</tr>
<tr>
<td>Negative</td>
<td>64 (2 cancer; 62 benign)</td>
<td>37 (2 cancer; 35 benign)</td>
<td>73 (2 cancer; 71 benign)</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>27 (23 cancer; 4 benign)</td>
<td>15 (12 cancer; 3 benign)</td>
<td>26 (20 cancer; 6 benign)</td>
<td></td>
</tr>
<tr>
<td>Sensitivity (95% CI)</td>
<td>92%</td>
<td>86%</td>
<td>90% (80% to 99%)</td>
<td>90.9% (78.8% to 100%)</td>
</tr>
<tr>
<td>Specificity (95% CI)</td>
<td>94%</td>
<td>92%</td>
<td>93% (88% to 98%)</td>
<td>92.1% (86.0% to 98.2%)</td>
</tr>
<tr>
<td>PPV (95% CI)</td>
<td>85%</td>
<td>80%</td>
<td>83% (72% to 95%)</td>
<td>76.9% (60.7% to 93.1%)</td>
</tr>
<tr>
<td>NPV (95% CI)</td>
<td>97%</td>
<td>95%</td>
<td>96% (92% to 95%)</td>
<td>97.2% (78.8% to 100%)</td>
</tr>
</tbody>
</table>

CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value.
The authors noted that, compared with the mutations panel used in their 2011 study, the NGS panel was associated with marked increase in NPV, with a similar positive predictive value (PPV). In this case, they proposed that the panel could be used to both “rule in” and “rule out” invasive cancers.

The same group (Nikiforov et al) reported on the performance of a subsequent generation ThyroSeq panel (ThyroSeq v2.1) with an expanded gene panel in a series of 465 thyroid FNA samples with a diagnosis of AUS or FLUS. Molecular analysis was performed prospectively in all patients. Ninety patients (96 nodules) underwent thyroid surgery, based on either patient preference, the presence of another nodule with a diagnosis of suspicious for malignancy or malignant on FNA, or positive molecular testing. Two other patients were considered to have a definitive nonsurgical diagnosis of primary hyperparathyroidism based on biochemical testing.

In addition to studies that describe the clinical validity of the mutations that comprise the ThyroSeq panel, studies have reported on the diagnostic performance of individual mutations and combinations of mutations to predict malignancy in thyroid nodules that are indeterminate on FNA. The results that pertain to the use of mutation testing in indeterminate thyroid nodules are summarized in Table 6. (In some cases, measures of agreement were calculated from data provided in the published article.)

### Table 6. Studies of Clinical Validity of Molecular Markers to Predict Malignancy in Indeterminate Thyroid FNA Samples

<table>
<thead>
<tr>
<th>Study (Year)</th>
<th>Population</th>
<th>Genes Tested</th>
<th>Insufficient or Inadequate for Analysis</th>
<th>Measures of Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ohori et al (2010)</td>
<td>100 patients with 117 atypia or follicular lesions of undetermined significance</td>
<td>BRAF, NRAS, HRAS, KRAS, RET/PTC1, RET/PTC3, PAX8-PPARY</td>
<td>NR</td>
<td>Sen: 60, Spec: 100, PPV: 100, NPV: 92, Acc: 93</td>
</tr>
<tr>
<td>Cantara et al (2010)</td>
<td>41 indeterminate and 54 suspicious thyroid nodules</td>
<td>BRAF, H-K-NRAS, RET/PTC, TRK, PAX8-PPARY</td>
<td>53</td>
<td>Sen: 86(^a) 80(^d), Spec: 97(^a) 100(^d), PPV: 86(^a) 100(^d), NPV: 97(^a) 47(^b), Acc: 95(^a) 83(^b)</td>
</tr>
<tr>
<td>Rossi et al (2015)</td>
<td>140 indeterminate or suspicious for malignancy or malignant nodules</td>
<td>BRAF</td>
<td>NR</td>
<td>Sen: 90(^c) 50(^d), Spec: 100(^c) 100(^d), PPV: 90(^c) 69(^d), NPV: 40(^c) 14(^d), Acc: 77(^d)</td>
</tr>
</tbody>
</table>

Acc: accuracy; CI: confidence interval; FNA: fine needle aspiration; NPV: negative predictive value; NR: not reported; PPV: positive predictive value; PTC: papillary thyroid carcinoma; Sen: sensitivity; Spec: specificity.

\(^a\) FNA indeterminate nodules.
\(^b\) FNA suspicious nodules.
Atypia of indeterminate significance.
Follicular neoplasm or suspicious for follicular neoplasm.
Suspicious for malignancy.

Additional studies have reported on differences in mutation frequency in malignant versus benign tumors, and report on the sensitivity and specificity of mutation testing in unselected populations (i.e., all patients with nodules, rather than just those with indeterminate cytology). These studies are summarized next.

Mathur et al collected thyroid FNA samples, thyroid tissue, clinical and histopathology data, and tumor genotyping for BRAF V600E, NRAS, and KRAS mutations, and RET/PTC1, RET/PTC3, and NTRK1 rearrangements for 341 patients with 423 dominant thyroid nodules. A cytologic examination of the samples showed that 51% were benign (25% were surgically resected), 21% were malignant, 11% were atypical lesions, 12% were follicular or Hürthle cell neoplasms, and 4% were suspicious for malignancy. On final analysis, 165 nodules were benign and 123 malignant. In the 423 FNA samples, 24 BRAF V600E, 7 KRAS, and 21 NRAS mutations, and 4 PAX8-PPARγ, 3 RET/PTC1, and 2 RET/PTC3 rearrangements were detected. In all, 17 (10.3%) of 165 benign thyroid nodules had a mutation compared with 26% (32/123) malignant tumors (p<0.05).

Eszlinger et al retrospectively analyzed a panel of mutations (BRAF and RAS point mutations, PAX8-PPARγ and RET/PTC rearrangements) in a sample of 310 thyroid air-dried FNA specimens with available corresponding FFPE thyroid biopsy samples (164 indeterminate, 57 malignant, 89 benign on FNA). Forty-seven mutations were detected on FNA: 22 BRAF, 13 NRAS, 3 HRAS mutations, and 8 PAX8-PPARγ and 1 RET/PTC rearrangements. The addition of mutation analysis to cytology results was associated with a sensitivity of 75.3% and a specificity of 90.4% for the detection of malignancy, with a PPV of 77.2% and NPV of 89.4%. The presence of a BRAF mutation or a RET/PTC rearrangement was associated with cancer in 100% of samples.

The association between BRAF mutations and PTC is supported in a report by Park et al (2015) on 294 patients with thyroid nodules whose FNA samples were evaluated for BRAF mutations using 2 methods, real-time PCR with TaqMan minor groove-binding probes and allele-specific PCR using dual-priming oligonucleotides. The detection rate of PTC by BRAF mutation testing by real-time PCR and allele-specific PCR were 80.2% (95% CI, 71.9% to 86.9%) and 76.9% (95% CI, 68.3% to 84.0%), respectively.

Mutations Association With Tumor Behavior
As already noted, the presence of BRAF mutations is strongly associated with malignancy in thyroid nodule FNA samples. BRAF mutations have also been associated with more aggressive clinicopathologic features in individuals diagnosed with PTC.

Adeniran et al assessed 157 cases with equivocal thyroid FNA readings (indeterminate and suspicious for PTC) or with a positive diagnosis for PTC and concomitant BRAF mutation analysis. The results of histopathologic follow-up correlated with the cytologic interpretations and BRAF status. Based on the follow-up diagnosis after surgical resection, the sensitivity for diagnosing PTC was 63.3% with cytology alone and 80.0% with the combination of cytology and BRAF testing. No false positives were noted with either cytology or BRAF mutation analysis. All PTCs with extrathyroidal extension or aggressive histologic features were positive for BRAF.
mutation. The authors concluded that patients with an equivocal cytologic diagnosis and \textit{BRAF} V600E mutation could be candidates for total thyroidectomy and central lymph node dissection. Xing et al investigated the utility of \textit{BRAF} mutation testing of thyroid FNA specimens for preoperative risk stratification of PTC in 190 patients. A \textit{BRAF} mutation in preoperative FNA specimens was associated with poorer clinicopathologic outcomes for PTC. Compared with the wild-type allele, a \textit{BRAF} mutation strongly predicted extrathyroidal extension (23\% vs 11\%; \textit{p}=0.039), thyroid capsular invasion (29\% vs 16\%; \textit{p}=0.045), and lymph node metastasis (38\% vs 18\%; \textit{p}=0.002). During a median follow-up of 3 years (range, 0.6-10 years), PTC persistence or recurrence was seen in 36\% of \textit{BRAF} mutation–positive patients versus 12\% of \textit{BRAF} mutation–negative patients, with an odds ratio of 4.16 (95\% CI, 1.70 to 10.17; \textit{p}=0.002). The PPV and NPV for preoperative FNA-detected \textit{BRAF} mutation to predict PTC persistence or recurrence were 36\% and 88\%, respectively, for all histologic subtypes of PTC. The authors concluded that preoperative \textit{BRAF} mutation testing of FNA specimens might provide a novel tool to preoperatively identify PTC patients at higher risk for extensive disease (extrathyroidal extension and lymph node metastases) and those more likely to manifest disease persistence or recurrence.

### Gene Expression Classifiers to Predict Malignancy

**Analytic Validity**

In 2015, Diggans et al described the development and validation Afirma BRAF malignancy classifier.\(^8\) The study included FNA biopsies from 716 thyroid nodules. Biopsies were evaluated with quantitative PCR for the \textit{BRAF} V600E gene, with 181 used as a training sample and 535 used as a validation sample. The Afirma BRAF malignancy classifier was generated using robust multichip average-normalized gene expression summaries, and the classifiers were evaluated for positive percent agreement (PPA) and negative percent agreement (NPA) with the PCR-derived gene classification. The highest scoring classification method and gene set were then used in a final round of model building. The maximum PPA and NPA for all cytology categories were observed when the threshold for \textit{BRAF}-positive status was 5\% or more \textit{BRAF} mutations. At 5\% analytic sensitivity, Afirma BRAF demonstrated a PPA with PCR results of 90.4\% (95\% CI, 83.5\% to 95.1\%) and an NPA of 99.0\% (95\% CI, 97.6\% to 99.7\%). Two samples in the training set and 4 samples in the validation set were Afirma BRAF–positive but negative (0\% mutation) on PCR, which the authors attributed to technical variability in either assay or to mutations other than the \textit{BRAF} V600E mutation that cause similar gene expression changes.

Intra- and interrun reproducibility of the classifier were evaluated using 9 FNA biopsies and 3 tissue controls selected from training samples with high (\textit{BRAF}-positive) or low (\textit{BRAF}-negative) classifier scores and scores near the classifier decision boundary. Each FNA biopsies and tissue was processed from total RNA in triplicate in each of 3 different runs across days, operators, and reagent lots. The intraassay standard deviation (SD) of Afirma BRAF scores was 0.171 (95\% CI, 0.146 to 0.204). Of the 106 Afirma BRAF calls produced (2 arrays failed quality control requirements), 106 resulted in concordant calls across all 3 runs (100\% concordance). The interassay SD of scores was 0.204 (95\% CI, 0.178 to 0.237) for scores measured on a 6-point scale. These results suggest low intra- and interrun variability.

In 2016, Kloos et al described the development of the Afirma MTC classifier in a study that also described the clinical validity of the MTC classifier.\(^44\)
Clinical Validity

Less evidence exists on the validity of gene expression profiling (specifically, the Afirma BRAF and Afirma MTC tests). Mutations can be used to improve the sensitivity and specificity for diagnosing indeterminate FNA of the thyroid, with the goal of identifying mutations that predict malignancy in FNA samples.

In the Diggans study, describing the development and validation of the Afirma BRAF test (previously described), for a subset of 213 thyroid nodule FNA samples for which histopathology was available, Afirma BRAF test results were compared with pathologic findings. Afirma BRAF classified all histopathologically benign samples as BRAF V600E-negative (specificity, 100%; 95% CI, 97.4% to 100%). Of the 73 histopathologically malignant samples, the Afirma BRAF test identified 32 as BRAF-positive (sensitivity, 43.8%; 95% CI, 32.2% to 55.9%).

In the Kloos study describing the development and validation of the Afirma MTC classifier, the MTC classifier was evaluated in a sample of 10,488 thyroid nodule FNA samples referred for GEC testing (the Afirma GEC described below). In this sample, 43 cases were Afirma MTC-positive, of which 42 were considered to be clinically consistent with medullary thyroid carcinoma on pathology or biochemical testing, for a PPV of 97.7% (95% CI, 86.2% to 99.9%).

Labourier et al reported on the sensitivity and specificity of a test algorithm combining micro-RNA measurements from 17 genes (miR/inform, Asuragen Laboratory, Austin, TX) with a 10-gene GEC in 109 FNA samples with AUS or FLUS or follicular neoplasms or suspicious for follicular neoplasm on cytology evaluated at the Asuragen Laboratory with known final pathology. Seventy-four nodules were diagnosed as benign and 35 as malignant. Performance of the combined test (micro-RNA measurements and the 10-gene GEC) is summarized in Table 7.


<table>
<thead>
<tr>
<th>Outcomes</th>
<th>All Patients</th>
<th>Atypia or Follicular Lesions of Undetermined Significance</th>
<th>Follicular Neoplasms or Suspicious for Follicular Neoplasms</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>109</td>
<td>58</td>
<td>51</td>
</tr>
<tr>
<td>Sensitivity (95% CI)</td>
<td>89% (73% to 97%)</td>
<td>94% (73% to 100%)</td>
<td>82% (57% to 96%)</td>
</tr>
<tr>
<td>Specificity (95% CI)</td>
<td>85% (75% to 92%)</td>
<td>80% (64% to 91%)</td>
<td>91% (76% to 98%)</td>
</tr>
<tr>
<td>PPV (95% CI)</td>
<td>74% (58% to 86%)</td>
<td>68% (46% to 85%)</td>
<td>82% (57% to 96%)</td>
</tr>
<tr>
<td>NPV (95% CI)</td>
<td>94% (85% to 98%)</td>
<td>97% (84% to 100%)</td>
<td>91% (76% to 98%)</td>
</tr>
</tbody>
</table>

CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value.

Clinical Utility

Testing for specific mutations associated with thyroid cancer (eg, BRAF V600E and RET mutations, RET/PTC and PAX8/PPARγ rearrangements) is generally designed to “rule in” cancer in nodules that have indeterminate cytology on FNA. (Of note, some mutation panels, such as the ThyroSeq panel, may have a high enough NPV that their clinical use could also be considered as a molecular marker to predict benignity; see next section.) A potential area for clinical utility for this type of mutation testing would be in informing preoperative planning for thyroid surgery following initial thyroid FNA, such as planning for a hemi- versus a total thyroidectomy or performance of a central neck dissection.

In a retrospective analysis, Yip et al reported outcomes after implementation of an algorithm incorporating molecular testing of thyroid FNA samples to guide the extent initial thyroid
The study included a cohort of patients treated at a single academic center at which molecular testing (BRAF V600E, BRAF K601E, NRAS codon 61, HRAS codon 61, and KRAS codon 12 and 13 point mutations; RET/PTC1, RET/PTC3, and PAX8/PPARγ rearrangements) was prospectively obtained for all FNAs with indeterminate cytology (follicular lesion of undetermined significance, follicular neoplasm, suspicious for malignancy), and for selective FNAs at the request of the managing physician for selected nodules with benign or nondiagnostic cytology. The study also included a second cohort of patients who did not have molecular testing results available. For patients treated with molecular diagnosis, a positive molecular diagnostic test was considered an indication for an initial total thyroidectomy. Patients with follicular lesion of undetermined significance and negative molecular diagnostic results were followed with repeat FNA, followed by a lobectomy or total thyroidectomy if indeterminate pathology persisted. Patients with follicular neoplasm or suspicious for malignancy results on cytology and a negative molecular diagnostic result were managed with lobectomy or total thyroidectomy.

The sample included 671 patients, 322 managed with and 349 without molecular diagnostics. Positive molecular testing results were obtained in 56 (17% of those managed with molecular diagnostics) patients, most commonly RAS mutations (42/56 [75%]), followed by BRAF V600E (10/56 [18%]) and BRAF K601E (2/56 [4%]) mutations, and PAX8/PPARγ rearrangements (2/56 [4%]). Compared with those managed without molecular diagnostics (63%), patients managed with molecular diagnostics (69%) were nonsignificantly less likely to undergo total thyroidectomy as an initial procedure (p=0.08). However, they had nonsignificantly higher rates of central compartment lymph node dissection (21% vs 15%, p=0.06). Across both cohorts, 25% (170/671) of patients had clinically significant thyroid cancer, with no difference in thyroid cancer rates based on the type of initial surgery (26% for total thyroidectomy vs 22% for lobectomy, p=0.3). The incidence of clinically significant thyroid cancer after initial lobectomy (ie, requiring a 2-stage surgery) was significantly lower for patients managed with molecular diagnostics (17% vs 43%, p<0.001). An indeterminate FNA result had a sensitivity and specificity for the diagnostic of thyroid cancer of 89% and 27%, respectively, with a PPV and NPV of 29% and 88%, respectively. The addition of molecular diagnostics to FNA results increased the specificity for a cancer diagnosis to 95% and the PPV to 82%.

In 2015, a task force from the American Thyroid Association (ATA) published a review with recommendations for the surgical management of FNA-indeterminate nodules with various molecular genetic tests. This review reported on the estimated likelihood of malignancy in an FNA-indeterminate nodule depending on results of the Afirma GEC test (described above) and other panels designed to rule in malignancy. Depending on the estimated prebiopsy likelihood of malignancy, recommendations for surgery included observation, active surveillance, repeat FNA, diagnostic lobectomy, or oncologic thyroidectomy.

Section Summary: Molecular Markers to Predict Malignancy

The available evidence has suggested that use of mutation testing in thyroid FNA samples is generally associated with a high specificity and PPV for clinically significant thyroid cancer. The most direct evidence related to the clinical utility of mutation testing for genes associated with malignancy in thyroid cancer comes from a single-center retrospective study that reported surgical decisions and pathology findings in patients managed with and without molecular diagnostics. There is potential clinical utility for identifying malignancy with higher certainty on FNA if such testing permits better preoperative planning at the time of thyroid biopsy, potentially avoiding the need for a separate surgery. An ATA statement provides some guidelines for
surgeons managing patients with indeterminate nodules. However, adoption of these guidelines in practice and outcomes associated with them are uncertain.

**SUMMARY OF EVIDENCE**

For individuals with thyroid nodule(s) and indeterminate findings on fine needle aspirate (FNA) who receive FNA sample testing with the Afirma Gene Expression Classifier (GEC) to predict benignancy, the evidence includes 1 prospective clinical validity study with the marketed test, and an indirect chain of evidence to support clinical utility. Relevant outcomes are disease-specific survival, test accuracy and validity, morbid events, and resource utilization. In 1 multicenter validation study, the Afirma GEC was reported to have a high negative predictive value (NPV; range, 90%-95%). These results are supported by an earlier development and clinical validation study (Chudova et al), but the classifiers used in the 2 studies do not appear to be identical. In an additional multicenter and multiple single-center studies, there is suggestive evidence that rates of malignancy are low in Afirma benign patients, but the exact NPV is unknown. The available evidence has suggested that physician decision making about surgery is altered by GEC results, although long-term follow-up of patients with thyroid nodules who avoided surgery based on GEC results is limited. An indirect chain of evidence can be constructed to establish the potential for clinical utility with GEC testing in cytologically indeterminate lesions, but with only 1 study of the marketed test reporting a true NPV, the clinical validity is uncertain.

For individuals with thyroid nodule(s) and indeterminate findings on FNA who receive FNA sample testing with molecular markers to predict malignancy, the evidence includes prospective and retrospective studies of clinical validity. Relevant outcomes are disease-specific survival, test accuracy and validity, morbid events, and resource utilization. Mutation analysis has the potential to improve the accuracy of an equivocal FNA of the thyroid and may play a role in preoperative risk stratification and surgical planning. Single-center studies have suggested that testing for a panel of mutations associated with thyroid cancer may allow for the appropriate selection of patients for surgical management with an initial complete thyroidectomy. Prospective studies in additional populations are needed to validate these results. Mutation analysis does not achieve a high enough NPV to identify which patients can undergo active surveillance over thyroid surgery. Although the presence of certain mutations may predict more aggressive malignancies, the management changes that would occur as a result of identifying higher risk tumors are not well-established. The evidence is insufficient to determine the effects of the technology on health outcomes.

**CLINICAL INPUT FROM PHYSICIAN SPECIALTY SOCIETIES AND ACADEMIC MEDICAL CENTERS**

While the various physician specialty societies and academic medical centers may collaborate with and make recommendations during this process, through the provision of appropriate reviewers, input received does not represent an endorsement or position statement by the physician specialty societies or academic medical centers, unless otherwise noted.

2016 Input

In response to requests, input was received from 2 physician specialty societies (1 of which provided 3 responses) and 1 academic medical center while this policy was under review in 2016. Input focused on the use of gene expression classifiers designed to with a high negative predictive value (NPV) in nodules indeterminate on fine needle aspirate (FNA). Although individual uses of a gene expression classifier with NPV in these situations varied, there was
general agreement that the tests are considered standard in the evaluation of some indeterminate cases of FNA.

2013 Input
In response to requests, input was received from 1 physician specialty society (4 reviewers) and 6 academic medical centers, for a total of 10 reviewers, while this policy was under review in 2013. There was general agreement with the policy statements that mutation analysis and use of the gene expression classifier is investigational. Input was mixed as to whether either test changes patient management and whether prospective randomized trials are necessary to establish the clinical utility of these tests.

PRACTICE GUIDELINES AND POSITION STATEMENTS
American Thyroid Association
In 2016, the American Thyroid Association (ATA) updated its guidelines on the management of thyroid nodules and differentiated thyroid cancer in adults.49 These guidelines made the following statements on molecular diagnostics in thyroid nodules that are atypia of undetermined significance (AUS) or follicular lesion of undetermined significance (FLUS) on cytology:

“For nodules with AUS/FLUS cytology, after consideration of worrisome clinical and sonographic features, investigations such as repeat FNA [fine needle aspirate] or molecular testing may be used to supplement malignancy risk assessment in lieu of proceeding directly with a strategy of either surveillance or diagnostic surgery. Informed patient preference and feasibility should be considered in clinical decision-making.” (Weak recommendation, Moderate-quality evidence)

“If repeat FNA cytology, molecular testing, or both are not performed or inconclusive, either surveillance or diagnostic surgical excision may be performed for an AUS/FLUS thyroid nodule, depending on clinical risk factors, sonographic pattern, and patient preference.” (Strong recommendation, Low-quality evidence)

The guidelines made the following statements on molecular diagnostics in thyroid nodules that are follicular neoplasm (FN) or suspicious for follicular neoplasm (SFN) on cytology:

“Diagnostic surgical excision is the long-established standard of care for the management of FN/SFN cytology nodules. However, after consideration of clinical and sonographic features, molecular testing may be used to supplement malignancy risk assessment data in lieu of proceeding directly with surgery. Informed patient preference and feasibility should be considered in clinical decision-making.” (Weak recommendation, Moderate-quality evidence)

The guidelines also stated: “there is currently no single optimal molecular test that can definitively rule in or rule out malignancy in all cases of indeterminate cytology, and long-term outcome data proving clinical utility are needed.”

National Comprehensive Cancer Network
National Comprehensive Cancer Network (NCCN) guidelines on the treatment of thyroid cancer (v1.2016) make the following comments on the use of molecular diagnostics in thyroid cancer50:

For thyroid nodules evaluated with FNA, molecular diagnostics may be employed when lesions are suspicious for (category 2B recommendation):

- Follicular or Hürthle cell neoplasms.
- Atypia of undetermined significance or follicular lesion of undetermined significance.
The guidelines also state: “Molecular testing (both the Gene Expression Classifier and individual mutation analysis) was available in the majority of NCCN Member Institutions (>75%). About 70% of the panelists would recommend using a gene expression classifier in the evaluation of follicular lesions.”

**U.S. PREVENTIVE SERVICES TASK FORCE RECOMMENDATIONS**
Not applicable.

**ONGOING AND UNPUBLISHED CLINICAL TRIALS**
A search of ClinicalTrials.gov in November 2016 did not identify any ongoing or unpublished trials that would likely influence this review.

**CODING**
The following codes for treatment and procedures applicable to this policy are included below for informational purposes. Inclusion or exclusion of a procedure, diagnosis or device code(s) does not constitute or imply member coverage or provider reimbursement. Please refer to the member’s contract benefits in effect at the time of service to determine coverage or non-coverage of these services as it applies to an individual member.

<table>
<thead>
<tr>
<th>CPT/HCPCS</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>81545</td>
<td>Oncology (thyroid), gene expression analysis of 142 genes, utilizing fine needle aspirate, algorithm reported as a categorical result (eg, benign or suspicious)</td>
</tr>
</tbody>
</table>

**ICD-10 Diagnoses**
- C73: Malignant neoplasm of thyroid gland
- D44.0: Neoplasm of uncertain behavior of thyroid gland

**REVISIONS**

**REFERENCES**


**APPENDIX**

**Appendix Table 1. Categories of Genetic Testing Addressed**

<table>
<thead>
<tr>
<th>Category</th>
<th>Addressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Testing of an affected individual’s germline to benefit the individual</td>
<td></td>
</tr>
<tr>
<td>1a. Diagnostic</td>
<td></td>
</tr>
<tr>
<td>1b. Prognostic</td>
<td></td>
</tr>
<tr>
<td>1c. Therapeutic</td>
<td></td>
</tr>
<tr>
<td>2. Testing cancer cells from an affected individual to benefit the individual</td>
<td>X</td>
</tr>
<tr>
<td>2a. Diagnostic</td>
<td></td>
</tr>
<tr>
<td>2b. Prognostic</td>
<td>X</td>
</tr>
<tr>
<td>2c. Therapeutic</td>
<td>X</td>
</tr>
<tr>
<td>3. Testing an asymptomatic individual to determine future risk of disease</td>
<td></td>
</tr>
<tr>
<td>4. Testing of an affected individual’s germline to benefit family members</td>
<td></td>
</tr>
<tr>
<td>5. Reproductive testing</td>
<td></td>
</tr>
<tr>
<td>5a. Carrier testing: preconception</td>
<td></td>
</tr>
<tr>
<td>5b. Carrier testing: prenatal</td>
<td></td>
</tr>
<tr>
<td>5c. In utero testing: aneuploidy</td>
<td></td>
</tr>
<tr>
<td>5d. In utero testing: mutations</td>
<td></td>
</tr>
<tr>
<td>5e. In utero testing: other</td>
<td></td>
</tr>
<tr>
<td>5f. Preimplantation testing with in vitro fertilization</td>
<td></td>
</tr>
</tbody>
</table>