Molecular Markers in Fine Needle Aspirates of the Thyroid

Policy
Blue Cross and Blue Shield of Kansas City (Blue KC) will provide coverage for gene expression classifiers when it is determined to be medically necessary because the criteria shown below are met.

When Policy Topic is covered
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 the following characteristics:
- Thyroid nodules without strong clinical or radiologic findings suggestive of malignancy.
- In whom surgical decision making would be affected by test results.

When Policy Topic is not covered
Genetic variant analysis in fine-needle aspirates of the thyroid is investigational.

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

Combined genetic variant analysis and microRNA gene expression classifier in fine needle aspirates of the thyroid is considered investigational.

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

GENETICS NOMENCLATURE UPDATE
Human Genome Variation Society (HGVS) nomenclature is used to report information on variants found in DNA and serves as an international standard in DNA diagnostics. It is being implemented for genetic testing medical evidence review updates starting in 2017 (see Table PG1). HGVS nomenclature is
recommended by HGVS, the Human Variome Project, and the HUman Genome Organization (HUGO).

The American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP) standards and guidelines for interpretation of sequence variants represent expert opinion from ACMG, AMP, and the College of American Pathologists. These recommendations primarily apply to genetic tests used in clinical laboratories, including genotyping, single genes, panels, exomes, and genomes. Table PG2 shows the recommended standard terminology—“pathogenic,” “likely pathogenic,” “uncertain significance,” “likely benign,” and “benign”—to describe variants identified that cause Mendelian disorders.

Table PG1. Nomenclature to Report on Variants Found in DNA

<table>
<thead>
<tr>
<th>Previous</th>
<th>Updated</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutation</td>
<td>Disease-associated variant</td>
<td>Disease-associated change in the DNA sequence</td>
</tr>
<tr>
<td>Variant</td>
<td>Change in the DNA sequence</td>
<td></td>
</tr>
<tr>
<td>Familial variant</td>
<td>Disease-associated variant identified in a proband for use in subsequent targeted genetic testing in first-degree relatives</td>
<td></td>
</tr>
</tbody>
</table>

Table PG2. ACMG-AMP Standards and Guidelines for Variant Classification

<table>
<thead>
<tr>
<th>Variant Classification</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathogenic</td>
<td>Disease-causing change in the DNA sequence</td>
</tr>
<tr>
<td>Likely pathogenic</td>
<td>Likely disease-causing change in the DNA sequence</td>
</tr>
<tr>
<td>Variant of uncertain significance</td>
<td>Change in DNA sequence with uncertain effects on disease</td>
</tr>
<tr>
<td>Likely benign</td>
<td>Likely benign change in the DNA sequence</td>
</tr>
<tr>
<td>Benign</td>
<td>Benign change in the DNA sequence</td>
</tr>
</tbody>
</table>

ACMG: American College of Medical Genetics and Genomics; AMP: Association for Molecular Pathology.

GENETIC COUNSELING

Genetic counseling is primarily aimed at patients who are at risk for inherited disorders, and experts recommend formal genetic counseling in most cases when genetic testing for an inherited condition is considered. The interpretation of the results of genetic tests and the understanding of risk factors can be very difficult and complex. Therefore, genetic counseling will assist individuals in understanding the possible benefits and harms of genetic testing, including the possible impact of the information on the individual’s family. Genetic counseling may alter the utilization of genetic testing substantially and may reduce inappropriate testing. Genetic counseling should be performed by an individual with experience and expertise in genetic medicine and genetic testing methods.

Description of Procedure or Service

<table>
<thead>
<tr>
<th>Populations</th>
<th>Interventions</th>
<th>Comparators</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individuals:</td>
<td>Interventions of interest are:</td>
<td>Comparators of interest are:</td>
<td>Relevant outcomes include:</td>
</tr>
<tr>
<td>• With thyroid nodule(s)</td>
<td>• Fine needle aspirate sample testing with the Afirma Gene Expression Classifier to predict benignancy</td>
<td>• Surgical biopsy</td>
<td>• Disease-specific survival</td>
</tr>
<tr>
<td>and indeterminate</td>
<td></td>
<td></td>
<td>• Test accuracy</td>
</tr>
<tr>
<td>findings on fine needle aspirate</td>
<td></td>
<td></td>
<td>• Test validity</td>
</tr>
<tr>
<td>aspirate</td>
<td></td>
<td></td>
<td>• Morbid events</td>
</tr>
<tr>
<td>Individuals:</td>
<td>Interventions of interest are:</td>
<td>Comparators of interest are:</td>
<td>Relevant outcomes include:</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>• With thyroid nodule(s) and indeterminate findings on fine needle aspirate</td>
<td>• Fine needle aspirate sample testing with molecular markers to predict malignancy</td>
<td>• Surgical management based on clinicopathologic risk factors</td>
<td>• Disease-specific survival</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Test accuracy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Test validity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Morbid events</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Resource utilization</td>
</tr>
</tbody>
</table>

Cytologic examination of fine needle aspiration (FNA) samples from a thyroid lesion to identify which patients need to undergo surgery has diagnostic limitations. Assays using molecular markers have been developed in an attempt to improve the accuracy of thyroid FNA biopsies.

For individuals with thyroid nodule(s) and indeterminate findings on 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 has evidence 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. The evidence is insufficient to determine the effects of the technology on health outcomes.

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.

For individuals with thyroid nodule(s) and indeterminate findings on FNA who receive a combined ThyGenX and ThyraMIR testing, the evidence includes 2 retrospective clinical validation studies that utilized a predicate test 17-variant panel (miRInform) to the current ThyGenX and ThyraMIR. Relevant outcomes are disease-specific survival, test accuracy and validity, morbid events, and resource utilization. In 1 retrospective validation study on FNA samples, the 17-variant panel (miRInform) and ThyraMIR had a sensitivity of 89% (95% CI, 73%-97%) and a NPV of 94% (95% CI, 85%-98%). No studies were identified demonstrating the diagnostic characteristics of the marketed ThyGenX. No studies were identified demonstrating evidence of direct outcome improvements. An indirect chain of evidence for the combined ThyGenX and ThyraMIR testing would rely on establishing clinical validity. The evidence is insufficient to determine the effects of the technology on health outcomes.

Clinical input has supported the use of a gene expression classifier with a high NPV for individuals with FNA found to be cytologically indeterminate. Although the current evidence for the clinical validity of the Afirma GEC is not entirely conclusive, given the suggestive evidence with supportive clinical input, this test may be considered medically necessary in the evaluation of FNAs of the thyroid that are cytologically considered to be indeterminate (follicular lesion of undetermined significance or suspicious for follicular neoplasm).

**Background**

**Fine needle aspiration (FNA) of the thyroid**

Thyroid nodules are common, present in 5-7% of the U.S. adult population. The vast majority are benign, and most cases of thyroid cancer are curable by surgery if detected early. Fine needle aspiration (FNA) 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, as different thyroid malignancies may require different surgical procedures (e.g. unilateral lobectomy versus total or sub-total thyroidectomy with or without lymph node dissection) depending on several factors, including histologic subtype and risk-stratification strategies (tumor size, patient age, etc.) 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 ~3% of all thyroid cancers.

The diagnosis of malignancy in the case of PTC is primarily based on cytologic features. If a FNA in a case of PTC is indeterminate, intraoperative consultation is most often diagnostic, although its efficacy and therefore use will vary between 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, as tissue sampling is necessary to observe these histologic characteristics. Intraoperative diagnosis of follicular carcinoma is challenging and often not feasible, as 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, in order 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.

**Genetic Variants associated with thyroid cancer**

Various genetic variants have been discovered in thyroid cancer. The 4 gene variants that are the most common and 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 (MAPK) pathway. (3) These mutually exclusive mutations are found in more than 70% of papillary carcinomas. (3) BRAF mutations are highly specific for PTC. Follicular carcinomas harbor either RAS mutations or PAX8/PPARγ rearrangement. These mutations are also mutually exclusive and identified in 70-75% of follicular carcinomas. (3) Genetic alterations involving the PI3K/AKT signaling pathway also occur in thyroid tumors, although they are rare in well-differentiated thyroid cancer and have higher prevalence in less differentiated thyroid carcinomas. (3) Additional mutations known to occur in poorly differentiated and anaplastic carcinomas 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**

**Variant Detection and Rearrangement Testing**

Single nucleotide variants (SNVs) 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. Panel tests for genes 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 variant 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.
ThyGenX™ is a next-generation sequencing panel that sequences 8 genes and identifies specific gene variants and translocations associated with thyroid cancer. ThyGenX is intended to be used in conjunction with the ThyraMIR microRNA expression test when the initial ThyGenX test is negative.

**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.

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](8) Zheng et al [2015](9)); these are not addressed in this review.

ThyraMIR is a micro-RNA expression-based classifier intended for use in thyroid nodules with indeterminate cytology on FNA following a negative result from The ThyGenX Thyroid Oncogene Panel.

**Algorithmic Testing**
Algorithmic testing involves the use of 2 or more tests in a prespecified sequence, with a subsequent test automatically obtained depending on results of an earlier test.

**Algorithmic Testing Using Afirma GEC with Afirma MTC and Afirma BRAF**
In addition to Afirma GEC, 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). Table 1 describes the testing algorithm for Afirma MTC and Afirma BRAF.

<table>
<thead>
<tr>
<th>Test 1</th>
<th>Test 1 Result</th>
<th>Reflex to Test 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thyroid nodule on FNA</td>
<td>&quot;Intermediate&quot;</td>
<td>Afirma MTC</td>
</tr>
<tr>
<td>Afirma GEC</td>
<td>&quot;Malignant&quot; or &quot;suspicious&quot;</td>
<td>Afirma MTC</td>
</tr>
<tr>
<td>Afirma GEC</td>
<td>&quot;Suspicious&quot;</td>
<td>Afirma BRAF</td>
</tr>
</tbody>
</table>

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 variant may not detect patients with low-frequency variants that result in the same pattern of pathway activation;
and (3) PCR-based approaches with high analytic sensitivity may require a large of amount of DNA that is difficult to isolate from small FNA samples.(10)

The testing strategy for both Afirma MTC and Afirma BRAF is to predict malignancy from a FNA sample with increased pretest probability for malignancy. A positive result with Afirma MTC or Afirma BRAF would inform preoperative planning such as planning for a hemi- versus a total thyroidectomy or performance of a central neck dissection.

**Algorithmic Testing Using ThyGenX and ThyraMIR**

The ThyGenX™ Thyroid Oncogene Panel (Interpace Diagnostics, Parsippany, NJ; testing done at Asuragen Clinical Laboratory) is a 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. ThyGenX has replaced the predicate test miRInform® Thyroid that tested for 17 validated gene alterations.

ThyraMIR™ (Interpace Diagnostics, Parsippany, NJ) is a micro-RNA expression-based classifier intended for use in thyroid nodules with indeterminate cytology on FNA following a negative result from The ThyGenX Thyroid Oncogene Panel.

The testing strategy for combined ThyGenX and ThyraMIR testing is to first predict malignancy. A positive result on ThyGenX would “rule in” patients for surgical resection. The specific testing results from a ThyGenX positive test would be used to inform preoperative planning when positive. For a ThyGenX negative result, the reflex testing involves the ThyraMIR microRNA expression test to “rule out” for a surgical biopsy procedure given the high NPV of the second test. Patients with a negative result from the ThyraMIR test would be followed with active surveillance and avoid a surgical biopsy.

**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.

Table 2 provides a summary of commercially-available molecular diagnostic tests for indeterminate thyroid pathology.
Table 2. Summary of Molecular Tests for Indeterminate Thyroid Cytopathology FNA Specimens

<table>
<thead>
<tr>
<th>Test</th>
<th>Methodology</th>
<th>Analyte(s)</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afirma GEC</td>
<td>mRNA gene expression</td>
<td>167 genes</td>
<td>Benign/suspicious</td>
</tr>
<tr>
<td>Afirma BRAF</td>
<td>mRNA gene expression</td>
<td>1 gene</td>
<td>Negative/positive</td>
</tr>
<tr>
<td>Afirma MTC</td>
<td>mRNA gene expression</td>
<td></td>
<td>Negative/positive</td>
</tr>
<tr>
<td>ThyroSeq v2</td>
<td>Next-generation sequencing</td>
<td>60+ genes</td>
<td>Specific gene variant/translocation</td>
</tr>
<tr>
<td>ThyGenX*</td>
<td>Next-generation sequencing</td>
<td>8 genes</td>
<td>Specific gene variant/Translocation</td>
</tr>
<tr>
<td>*miRInform (predicate test to ThyGenX and not commercially available)</td>
<td>Multiplex PCR by sequence-specific probes</td>
<td>14 DNA variants, 3 RNA fusions</td>
<td>Specific gene variant/Translocation</td>
</tr>
<tr>
<td>ThyraMIR</td>
<td>microRNA expression</td>
<td>10 microRNAs</td>
<td>Negative/positive</td>
</tr>
</tbody>
</table>

Rationale
This evidence review was originally created in January 2012 and has been updated regularly with searches of the MEDLINE database. The most recent literature review was performed through April 25, 2017 (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 USING AFIRMA GEC

Clinical Context and Test Purpose
The purpose of Afirma Gene Expression Classifier (GEC) in individuals with indeterminate findings on fine needle aspirate(s) of thyroid nodules is to predict benignancy and eliminate or necessitate the need for surgical resection.

The relevant question addressed in this evidence review is: Does Afirma Gene Expression Classifier (GEC) appropriately eliminate or necessitate the need for surgical resection and lead to improved health outcomes?

The following PICOTS were used to select literature to inform this review.

Patients
The relevant population of interest includes individuals with indeterminate findings on fine needle aspirate(s) of thyroid nodules who would be willing to undergo watchful waiting, depending on results of their Afirma GEC test. Patients with indeterminate findings presently proceed to surgical resection.

Interventions
The relevant intervention of interest is the Afirma GEC.
**Comparators**
The relevant comparator of interest is standard surgical management through surgical resection.

**Outcomes**
The potential beneficial outcomes of primary interest are avoiding an unneeded surgical resection (e.g., hemithyroidectomy or thyroidectomy) due to thyroid nodules that are absent of cancer.

Potential harmful outcomes are those resulting from false-positive or false-negative test results. False-positive test results can lead to unnecessary surgical resection and procedure-related complications. False-negative test results can lead to lack of surgical resection for thyroid cancer.

**Timing**
The time frame for evaluating performance of the test varies the time from the initial fine needle aspiration to surgical resection to weeks to months following an indeterminate result to years. Papillary thyroid cancer is an indolent cancer, and a nodule could be observed for many years to ensure no clinical change.

**Setting**
The primary setting would be in endocrinology.

**Analytic Validity**
Walsh et al (2012) 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 (2010) described the development and initial clinical validation of a version of the Afirma GEC.(2) 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 (2012) reported on a 19-month, prospective, multicenter (49 academic and community) sites, study of the Afirma GEC.(12) A total of 4812 nodules were screened for inclusion with centralized cytopathology. Local pathology reports of the cytologic 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 cytologic 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.(13) Most nodules sent for GEC testing were AUS or FLUS or follicular neoplasm. The distribution of GEC testing results for each cytologic classification is shown in Table 3.

**Table 3: GEC Testing Results From Alexander et al (2014)**

<table>
<thead>
<tr>
<th>Cytologic Classification</th>
<th>N</th>
<th>Benign</th>
<th>Suspicious</th>
<th>Non-diagnostic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atypia or follicular lesion of undetermined significance</td>
<td>165</td>
<td>91 (55%)</td>
<td>66 (40%)</td>
<td>8 (5%)</td>
</tr>
<tr>
<td>Follicular neoplasm</td>
<td>161</td>
<td>79 (49%)</td>
<td>73 (45%)</td>
<td>9 (6%)</td>
</tr>
<tr>
<td>Suspicious for malignancy</td>
<td>13</td>
<td>4 (31%)</td>
<td>9 (69%)</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>339</td>
<td>174</td>
<td>148</td>
<td>17</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 4 shows thyroidectomy biopsy results for the subset of patients in Table 3 who underwent surgery.

**Table 4: Thyroidectomy Results From Alexander et al (2014)**

<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.(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 neoplasm; 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 5.

**Table 5: Pooled GEC Performance From Santhanam et al (2016)**

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Point Estimate</th>
<th>95% Confidence Interval</th>
<th>I^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>95.7%</td>
<td>92.2% to 97.9%</td>
<td>45.4%</td>
</tr>
<tr>
<td>Specificity</td>
<td>30.5%</td>
<td>26.0% to 35.3%</td>
<td>92.1%</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>

reported the diagnostic accuracy of the Afirma GEC (see Table 6). 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.(19)

Table 6: 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>36a</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Lastra et al (2014)</td>
<td>69 (51.5%) AUS/FLUS 39 (29.5%) FN 25 (19%) FNOF</td>
<td>Suspicious Benign</td>
<td>62</td>
<td>70</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>McIver et al (2014)</td>
<td>12 (11.4%) AUS/FLUS 93 (88.6%) FN/HCN</td>
<td>Suspicious Benign</td>
<td>44b</td>
<td>16</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Yang et al (2016)</td>
<td>165 (76%) AUS/FLUS 24 (11%) SFN/FN</td>
<td>Suspicious Benign</td>
<td>80</td>
<td>94</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Witt et al (2016)</td>
<td>47 AUS/FLUS or SFN/FN (32 with GEC attemptedc)</td>
<td>Suspicious Benign</td>
<td>15</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</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.(21) 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 (2012) 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.(22) 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.(13) 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] (23) Noureldine et al [2015] (24)) 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 CYTOLGICALLY 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 and Test Purpose
The purpose of testing for molecular markers (eg, single nucleotide variants and gene rearrangement) in individuals with indeterminate findings on fine needle aspirate(s) of thyroid nodules is to predict malignancy and change surgical approach or management.

The relevant question addressed in this evidence review is: Does testing for molecular markers predict malignancy and alter surgical approach or management and lead to improved health outcomes?

The following PICOTS were used to select literature to inform this review.

Patients
The relevant population of interest includes individuals with indeterminate findings on fine needle aspirate(s) of thyroid nodules. Patients with indeterminate findings presently proceed to surgical biopsy with intraoperative pathology consultation.

Interventions
The relevant intervention of interest is testing for molecular markers (eg, single nucleotide variants and gene rearrangements to predict malignancy).

Comparators
The relevant comparator of interest is standard surgical management through surgical resection.

Outcomes
The potential beneficial outcomes of primary interest are to allow for appropriate surgical planning in the preoperative period (eg, hemithyroidectomy or thyroidectomy when malignancy is predicted or lobectomy if malignancy is less likely). This has the potential benefit of reducing the likelihood of having the patient have to have repeat surgery if a diagnosis is not made on frozen pathology section during the initial surgery if lobectomy is done as a first procedure.

Potential harmful outcomes are those resulting from false-positive or false-negative test results. False-positive test results can lead to unnecessary surgical resection and procedure-related complications. False-negative test results can lead to lack of surgical resection for thyroid cancer.
**Timing**
The time frame for evaluating performance of the test varies the time from the initial fine needle aspiration to a surgical resection to weeks to months following an indeterminate result.

**Setting**
The primary setting would be in endocrinology.

**Variant Detection and Rearrangement Testing**

**Analytic Validity**
Single nucleotide variants (SNVs) in specific genes associated with thyroid cancer (eg, the *BRAF* V600E gene) and the detection of genetic rearrangements associated with thyroid cancer (eg, the *RET/PTC* rearrangement) are typically detected with Sanger sequencing or next-generation sequencing (NGS) methods. In the case of testing for gene variants associated with thyroid cancer malignancy, analytic validity refers to a test’s technical accuracy in detecting a variant that is present or in excluding a variant 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* variant detection by qualitative PCR in thyroid FNA samples with high within- and between-run reproducibility.(26)

NGS is expected to have high accuracy for detecting a variant that is present. However, with increasing numbers of tested variants, 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.(27) They developed a custom library of gene sequence variants based on variants 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 variants. 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 variants 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 variants.

**Clinical Validity**
A number of studies have evaluated whether testing for SNVs or gene fusions (either SNVs or panels) can be used to improve the sensitivity and specificity for diagnosing indeterminate FNA of the thyroid, with the goal of identifying variants that predict malignancy in FNA samples.
Variants Association With Malignancy

In 2015, Fnais et al conducted a systematic review and meta-analysis of studies reporting on the test accuracy of \textit{BRAF} variant testing in the diagnosis of PTC.(28) The review included 47 studies with 9924 FNA samples. For all cytologically indeterminate nodules, the pooled sensitivity estimate for \textit{BRAF} variant testing was 31\% (95\% CI, 6\% to 56\%). Among nodules suspicious for malignancy on FNA, the pooled sensitivity estimate for \textit{BRAF} variant testing was 52\% (95\% CI, 39\% to 64\%; \textit{I}^2=77\%).

Ferraz et al evaluated 20 publications that reported on the type and number of variants in cases of FNA of the thyroid diagnosed as indeterminate and compared the results with final histology after surgical resection.(29) Sixteen studies analyzed 1 variant (eg, \textit{BRAF} variant or \textit{RET/PTC} rearrangement) and 4 studies analyzed a panel of several variants (\textit{BRAF} and \textit{RAS} variants, \textit{RET/PTC} and \textit{PAX8/PPAR\gamma} rearrangements). The detection of a variant in a histologically (surgically resected) benign thyroid lesion was categorized as a false-positive case, detecting no variant in an FNA sample from a histologically benign surgical sample was considered a true negative, and finding no variant in a histologically malignant lesion was categorized as a false negative. Based on 4 studies that examined a panel of variants, 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} variants, 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 variants 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 variant 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 variants or additional variants are summarized in the following section.

Nikiforov et al prospectively tested a panel of variants (\textit{BRAF}, \textit{RAS}, \textit{RET/PTC}, \textit{PAX8/PPAR\gamma}) in 470 FNA samples of thyroid nodules from 328 consecutive patients.(30) Variant 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 variants were found (18
BRAF, 8 RAS, 5 RET/PTC, 1 PAX8/PPARγ); after surgery, 31 (97%) variant-positive nodules were diagnosed as malignant on pathologic examination and 1 (3%) as a benign tumor. Thirteen of the 32 variant-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 variant 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 variant screening in 117 FNA samples classified as AUS or FLUS. (31) BRAF, RAS, RET/PTC, or PAX8/PPARγ variants 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 variants investigated was 100%, whereas the probability of having a thyroid malignancy without a variant detected was 7.6%.

In 2011, Nikiforov et al reported results of a prospective study that assessed the clinical validity of a panel of variants to predict the likelihood of malignancy in thyroid nodules found indeterminate on FNA. (32) The authors included 1056 consecutive samples with indeterminate cytology on FNA that underwent variant 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 variants were detected. At analysis, 479 patients had undergone thyroidectomy for further evaluation, providing a histopathologic diagnosis for 513 samples. The presence of a variant 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 variant 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 variant 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 single nucleotide variants 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. (33) 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.

Table 5: Performance of ThyroSeq Panel in Nikiforov et al (2014)\textsuperscript{33} and (2015)\textsuperscript{34}

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Retrospective (n=91)</td>
<td>Prospective (n=52)</td>
</tr>
<tr>
<td>Negative</td>
<td>64 (2 cancer; 62 benign)</td>
<td>37 (2 cancer; 35 benign)</td>
</tr>
<tr>
<td>Positive</td>
<td>27 (23 cancer; 4 benign)</td>
<td>15 (12 cancer; 3 benign)</td>
</tr>
<tr>
<td>Sensitivity (95% CI)</td>
<td>92%</td>
<td>86%</td>
</tr>
<tr>
<td>Specificity (95% CI)</td>
<td>94%</td>
<td>92%</td>
</tr>
<tr>
<td>PPV (95% CI)</td>
<td>85%</td>
<td>80%</td>
</tr>
<tr>
<td>NPV (95% CI)</td>
<td>97%</td>
<td>95%</td>
</tr>
</tbody>
</table>

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

The authors noted that, compared with the gene 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.\textsuperscript{34} 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 genes that comprise the ThyroSeq panel, studies have reported on the diagnostic performance of individual variants and combinations of variants to predict malignancy in thyroid nodules that are indeterminate on FNA. The results that pertain to the use of gene 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>
<thead>
<tr>
<th>Study (Year)</th>
<th>Population</th>
<th>Genes Tested</th>
<th>Insufficiency or Inadequate for Analysis</th>
<th>Measures of Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moses et al (2010)</td>
<td>35 indeterminate thyroid nodules</td>
<td>BRAF, KRAS, NRAS, RET/PTC1, RET/PTC3, NTRK1</td>
<td>2</td>
<td>38 95 67 79 77</td>
</tr>
<tr>
<td>Ohori et al (2010)</td>
<td>100 patients with atypia or follicular lesions of undetermined significance</td>
<td>BRAF, NRAS, HRAS, KRAS, RET/PTC1, RET/PTC3, PAX8/PPARY</td>
<td>NR</td>
<td>60 100 100 92 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>86a 80b 97a 100b 6a 80b 97a 47b 95a 83b</td>
</tr>
<tr>
<td>Xing et al (2009)</td>
<td>25 indeterminate, dominant nodules</td>
<td>BRAF</td>
<td>NR</td>
<td>14 100 100 48 52</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>90c 50d 100c 100d 100e 100c 100d 100e 93c 69d 14e 96c 77d 46e</td>
</tr>
<tr>
<td>Beaudenon-Huibregts e et al (2014)</td>
<td>53 nodules with indeterminate/ nondiagnostic FNA</td>
<td>BRAF, HRAS, KRAS, NRAS, PAX8-PPARY, RET-PTC1, RET-PTC3</td>
<td>48</td>
<td>89 81 64</td>
</tr>
<tr>
<td>Valderrobano et al (2017)</td>
<td>190 indeterminate thyroid nodules</td>
<td>ThyroSeq v2 (60+ genes)</td>
<td>2</td>
<td>70 77 42 91</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. c Atypia of indeterminate significance. d Follicular neoplasm or suspicious for follicular neoplasm. e Suspicious for malignancy.

Additional studies have reported on differences in variant frequency in malignant versus benign tumors, and report on the sensitivity and specificity of gene testing in unselected populations (ie, 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 variants, and RET/PTC1, RET/PTC3, and NTRK1 rearrangements for 341 patients with 423 dominant thyroid nodules. (42) 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 variants, 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 variant compared with 26% (32/123) malignant tumors (p<0.05).

Eszlinger et al retrospectively analyzed a panel of variants (BRAF and RAS single nucleotide variants, 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 variants were detected on FNA: 22 BRAF, 13 NRAS, 3 HRAS variants, and 8 PAX8/PPARγ and 1 RET/PTC rearrangements. The addition of variant 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 variant or a RET/PTC rearrangement was associated with cancer in 100% of samples.

The association between BRAF variants 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 variants 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 variant 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.

**Genetic Variants Association With Tumor Behavior**

As already noted, the presence of BRAF variants is strongly associated with malignancy in thyroid nodule FNA samples. BRAF variants 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 variant 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 variant analysis. All PTCs with extrathyroidal extension or aggressive histologic features were positive for BRAF variant. The authors concluded that patients with an equivocal cytologic diagnosis and BRAF V600E variant could be candidates for total thyroidectomy and central lymph node dissection.

Xing et al investigated the utility of BRAF variant testing of thyroid FNA specimens for preoperative risk stratification of PTC in 190 patients. A BRAF variant in preoperative FNA specimens was associated with poorer clinicopathologic outcomes for PTC. Compared with the wild-type allele, a BRAF variant strongly predicted extrathyroidal extension (23% vs 11%; p=0.039), thyroid capsular invasion (29% vs 16%; p=0.045), and lymph node metastasis (38% vs 18%;
During a median follow-up of 3 years (range, 0.6–10 years), PTC persistence or recurrence was seen in 36% of BRAF variant–positive patients versus 12% of BRAF variant–negative patients, with an odds ratio of 4.16 (95% CI, 1.70 to 10.17; p=0.002). The PPV and NPV for preoperative FNA-detected BRAF variant to predict PTC persistence or recurrence were 36% and 88%, respectively, for all histologic subtypes of PTC. The authors concluded that preoperative BRAF variant 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.(10) The study included FNA biopsies from 716 thyroid nodules. Biopsies were evaluated with quantitative PCR for the 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 BRAF-positive status was 5% or more BRAF variants. 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 variant) on PCR, which the authors attributed to technical variability in either assay or to variants other than the BRAF V600E variant 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 (BRAF-positive) or low (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.(45)
Pankratz et al (2016) studied the analytic performance of Afirma MTC classifier from fresh-frozen tissue specimens with a confirmed medullary thyroid carcinoma (MTC) diagnosis. (46) 27 MTC tissue specimens were compared with 20 de-identified FNA samples from normal donors. The reported clinical sensitivity of the Afirma MTC classifier was 96.3% [95% CI, 81.0%-99.9%].

**Clinical Validity**

Less evidence exists on the validity of gene expression profiling (specifically, the Afirma BRAF and Afirma MTC tests). Genetic variants can be used to improve the sensitivity and specificity for diagnosing indeterminate FNA of the thyroid, with the goal of identifying variants 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. (10) 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). (45) 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%).

**Clinical Utility**

Testing for specific variants associated with thyroid cancer (eg, *BRAF* V600E and *RET* variants, *RET/PTC* and *PAX8/PPARγ* rearrangements) is generally designed to “rule in” cancer in nodules that have indeterminate cytology on FNA. (47) (Of note, some gene 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 benignancy; see next section.) A potential area for clinical utility for this type of variant 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 resection. (48) 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 single nucleotide variants; *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 variants (42/56 [75%]), followed by BRAF V600E (10/56 [18%]) and BRAF K601E (2/56 [4%]) variants, 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 diagnosis 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.(49) 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 variant 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 variant 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.

**COMBINED THYGENX VARIANT DETECTION AND THYRMIR MICRO RNA TESTING**

**Clinical Context and Test Purpose**
The purpose of ThyGenX™ Thyroid Oncogene Panel and ThyraMIR micro RNA classifier in individuals with indeterminate findings on fine needle aspirate(s) of thyroid nodules is to predict malignancy and inform surgical planning decisions with positive results using ThyGenX, and if negative, predict benignancy using ThyrmIR micro RNA classifier to eliminate or necessitate the need for surgical biopsy.

The relevant question addressed in this evidence review is: Does the combined use of ThyGenX and ThyraMIR appropriately eliminate or necessitate the need for surgical resection or biopsy and lead to improved health outcomes?

The following PICOTS were used to select literature to inform this review.

**Patients**
The relevant population of interest includes individuals with indeterminate findings on fine needle aspirate(s) of thyroid nodules. Patients with indeterminate findings presently proceed to surgical resection.

**Interventions**
The relevant intervention of interest is combined ThyGenX™ Thyroid Oncogene Panel and ThyraMIR micro RNA classifier testing.

**Comparators**
The relevant comparator of interest is surgical biopsy standard surgical management through surgical resection.

**Outcomes**
The potential beneficial outcomes of primary interest are avoiding an unneeded surgical biopsy or surgical resection (eg, hemithyroidectomy or thyroidectomy) due to thyroid nodules that are absent of cancer.

Potential harmful outcomes are those resulting from false-positive or false-negative test results. False-positive test results can lead to unnecessary surgical biopsy or resection and procedure-related complications. False-negative test results can lead to lack of surgical biopsy or resection for thyroid cancer.

**Timing**
The time frame for evaluating performance of the test varies the time from the initial fine needle aspiration to surgical resection to weeks to months following an indeterminate result.
Setting
The primary setting would be in outpatient endocrinology.

Test Development and Analytic Validity
Hadd et al reported on targeted next-generation sequencing (NGS) of cancer genes in 38 formalin-fixed paraffin-embedded and 10 fine needle aspiration tumor specimens.(50) The results show an accuracy of 96.1% (95% CI, 96.1% to 99.3%) compared with Sanger sequencing; Sanger sequencing has an analytic sensitivity of approximately 15% to 20%. When NGS was compared to a multiplex detection system with a 1% variant detection rate, the accuracy was reported to be 99.6% (95% CI, 97.9% to 99.9%).

Wylie et al (2016) reported on the development of the ThyraMIR miRNA classifier, along with a 17-variant oncogene panel including BRAF, RAS, RET or PAX.(51) An miRNA classifier was originally developed using rtPCR methodology in a sample of 257 surgical specimens, and validated in an independent set of 42 nodules with indeterminate cytology. A 17-variant panel covering validated oncogenic gene alterations for BRAF, RAS, RET or PAX8 genes was tested on preoperative FNA and surgical specimens. Optimization of miRNA classifiers A and B resulted in the commercial ThyraMIR Classifier. ThyraMIR was used on a subset of thyroid tissues negative by the targeted 17-variant panel and resulted in a sensitivity of 85% and specificity of 95%.

Section Summary: Analytic Validity
The analytic validity of targeted next-generation sequencing (NGS) of cancer genes is expected to be high. Concordance rates between Sanger sequencing and NGS are high but limited lower analytic sensitivity of Sanger sequencing. Concordance rates increased when NGS is compared to an orthogonal technology with a 1% variant detection rate. One study describing the development of a miRNA classifier was identified; a description of the analytic validity of the corresponding commercially-available NGS version of the oncogene pane has not been identified but is expected to be high.

Clinical Validity
Labourier et al (2015) evaluated the diagnostic algorithm combining a 17-variant panel with ThyraMIR on a cross-sectional cohort of thyroid nodules comprised of 109 FNA samples with AUS/FLUS or FN/SFN across 12 endocrinology centers.(52) A summary of the sensitivity and specificity of the combined test is listed in Table 9.

Table 9. Summary of Clinical Validity for 17-variant Panel and ThyraMIR on FNA Samples

<table>
<thead>
<tr>
<th>No. of Cases</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort, %</td>
<td>109 (95% CI)</td>
<td>89 (73–97)</td>
<td>85 (75–92)</td>
<td>74 (58–86)</td>
<td>94 (85–98)</td>
</tr>
<tr>
<td>AUS/FLUS, % (95% CI)</td>
<td>58</td>
<td>94 (73–100)</td>
<td>80 (64–85)</td>
<td>97 (84–100)</td>
<td>68 (8–590)</td>
</tr>
<tr>
<td>----------------------</td>
<td>----</td>
<td>-------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>FN/SFN, % (95% CI)</td>
<td>51</td>
<td>82 (57–96)</td>
<td>91 (76–98)</td>
<td>82 (57–96)</td>
<td>48 (9–269)</td>
</tr>
</tbody>
</table>

FNA, fine-needle aspiration; AUS, atypia of undetermined significance; FLUS, follicular lesion of undetermined significance; FN; follicular neoplasm; SFN, suspicious for a follicular neoplasm; PPV, positive predictive value; NPV, negative predictive value.

**Section Summary: Clinical Validity**

Evidence for clinical validity of combined testing for miRNA gene expression using ThyraMIR and a targeted 17-variant panel comes from a two retrospective studies utilizing archived surgical specimens and FNA samples. One study combined a 17-variant panel with ThyraMIR testing on archived surgical specimens and resulted in a sensitivity of 85% and specificity of 95%. The second study combined a 17-variant panel (miRInform) with ThyraMIR testing on FNA samples and resulted in a sensitivity of 89%, specificity of 85%, PPV of 74% and NPV of 94%. No studies were identified that demonstrated the clinical validity of a combined ThyGenX and ThyraMIR test on FNA samples.

**Clinical Utility**

Direct evidence for the clinical utility for the combined ThyGenX and ThyraMIR diagnostic testing algorithm are lacking. In the absence of direct evidence for the clinical utility of the combined testing, an indirect chain of evidence may be constructed to infer potential clinical utility of the combined diagnostic testing algorithm. No studies utilizing ThyGenX next-generation sequencing panel in FNA samples was identified. However, available evidence has suggested that use of variant testing using NGS in thyroid FNA samples is generally associated with a high specificity and PPV for clinically significant thyroid cancer. 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. However, variant analysis does not achieve a high enough NPV to identify which patients can undergo active surveillance over thyroid surgery. In the diagnostic algorithm that reflexes to the ThyraMIR after a negative ThyGenX result, patients receiving reflex testing could identify who may undergo active surveillance over thyroid surgery. A single study utilizing a 17-variant panel with ThyraMIR showed a NPV of 94%. Therefore, the high NPV of ThyraMIR has the potential to accurately predict benignancy and triage patients to active surveillance.

**Section Summary: Clinical Utility**

Direct evidence for the clinical utility of combined ThyGenX and ThyraMIR reflex testing is lacking. However, available evidence suggests that testing for gene variants and rearrangements can predict malignancy and inform surgical planning decisions when the test is positive, but the NPV of the ThyGenX to identify patients who should undergo active surveillance over thyroid surgery is unknown. In a reflex testing setting, the high NPV for a microRNA gene expression test used on the subset of patients with a negative result from a variant and gene rearrangement test has the potential clinical utility in identifying patients appropriately for active surveillance.
SUMMARY OF EVIDENCE
For individuals with thyroid nodule(s) and indeterminate findings on 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 has evidence 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. The evidence is insufficient to determine the effects of the technology on health outcomes.

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. Variant 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 genetic variants 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. Variant analysis does not achieve a high enough NPV to identify which patients can undergo active surveillance over thyroid surgery. Although the presence of certain variants 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.

For individuals with thyroid nodule(s) and indeterminate findings on FNA who receive a combined ThyGenX and ThyraMIR testing, the evidence includes 2 retrospective clinical validation studies that utilized a predicate test 17-variant panel (miRInform) to the current ThyGenX and ThyraMIR. Relevant outcomes are disease-specific survival, test accuracy and validity, morbid events, and resource utilization. In 1 retrospective validation study on FNA samples, the 17-variant panel (miRInform) and ThyraMIR had a sensitivity of 89% (95% CI, 73%-97%) and a NPV of 94% (95% CI, 85%-98%). No studies were identified showing the clinical validity or clinical utility of the currently marketed combined ThyGenX and
ThyraMIR tests. An indirect chain of evidence for the combined ThyGenX and ThyraMIR testing relies on evidence from other variant detection and gene arrangement tests and microRNA gene expression performed in conjunction with a target 17-variant panel. The evidence is insufficient to determine the effects of the technology on health outcomes.

SUPPLEMENTAL INFORMATION

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 variant 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.(53) 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 (v2.2017) make the following comments on the use of molecular diagnostics in thyroid cancer(54):

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 variant 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.
MEDICARE NATIONAL COVERAGE
There is no national coverage determination (NCD). In the absence of an NCD, coverage decisions are left to the discretion of local Medicare carriers.

Palmetto GBA determines coverage and reimbursement for laboratories that perform molecular diagnostic testing and submit claims to Medicare in Medicare Jurisdiction E (California, Nevada, and Hawaii). Palmetto GBA’s decisions apply for all molecular diagnostic tests for Medicare.

Palmetto GBA completed an assessment of the Afirma GEC and determined that the test meets criteria for analytic and clinical validity and clinical utility as a reasonable and necessary Medicare benefit. Effective January 2012, Palmetto GBA began to reimburse Afirma GEC services for patients with the following conditions:

- “Patients with 1 or more thyroid nodules with a history or characteristics suggesting malignancy such as:
  - Nodule growth over time
  - Family history of thyroid cancer
  - Hoarseness, difficulty swallowing or breathing
  - History of exposure to ionizing radiation
  - Hard nodule compared with rest of gland consistency
  - Presence of cervical adenopathy
- Have an indeterminate follicular pathology on fine needle aspiration”

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.

References:

**Billing Coding/Physician Documentation Information**

**81445** Targeted genomic sequence analysis panel, solid organ neoplasm, DNA analysis, and RNA analysis when performed, 5-50 genes (eg, ALK, BRAF, CDKN2A, EGFR, ERBB2, KIT, KRAS, NRAS, MET, PDGFRA, PDGFRB, PGR, PIK3CA, PTEN, RET), interrogation for sequence variants and copy number variants or rearrangements, if performed

**81479** Unlisted molecular pathology procedure

**81545** Oncology (thyroid), gene expression analysis of 142 genes, utilizing fine needle aspirate, algorithm reported as a categorical result (eg, benign or suspicious) (new code 1/1/2016)

**ICD-10 Codes**

**C73** Malignant neoplasm of thyroid gland.

**D44.0** Neoplasm of uncertain behavior of thyroid gland

There is a specific CPT code for the Afirma® Gene Expression Classifier test:

**81545** Oncology (thyroid), gene expression analysis of 142 genes, utilizing fine needle aspirate, algorithm reported as a categorical result (eg, benign or suspicious).
Additional Policy Key Words
RosettaGX Reveal™
Rosetta GX Reveal™

Policy Implementation/Update Information
3/1/13  New policy; considered investigational.
3/1/14  No policy statement changes.
3/1/15  Added CPT information in Billing Coding Information. No policy statement changes.
3/1/16  No policy statement changes. Added new CPT code.
1/1/17  A medically necessary statement was added for the use of Afirma Gene Expression Classifier in patients with indeterminate thyroid fine needle aspirates based on results of clinical input.
3/1/17  No policy statement changes.
8/1/17  Policy statements revised to add investigational statement for combined genetic variant analysis and microRNA gene expression classifier (ie, ThyGenX/ThyraMir). Updated main policy statement: Blue Cross and Blue Shield of Kansas City (Blue KC) will provide coverage for gene expression classifiers when it is determined to be medically necessary because the criteria shown below are met.

State and Federal mandates and health plan contract language, including specific provisions/exclusions, take precedence over Medical Policy and must be considered first in determining eligibility for coverage. The medical policies contained herein are for informational purposes. The medical policies do not constitute medical advice or medical care. Treating health care providers are independent contractors and are neither employees nor agents Blue KC and are solely responsible for diagnosis, treatment and medical advice. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, photocopying, or otherwise, without permission from Blue KC.