Title: Molecular Markers in Fine Needle Aspirates of the Thyroid

**Professional**
Original Effective Date: March 17, 2017
Revision Date(s): March 17, 2017; October 1, 2017; April 15, 2018; September 14, 2018; January 1, 2019; August 28, 2019
Current Effective Date: August 28, 2019

**Institutional**
Original Effective Date: March 17, 2017
Revision Date(s): March 17, 2017; October 1, 2017; April 15, 2018; September 14, 2018; January 1, 2019; August 28, 2019
Current Effective Date: August 28, 2019

State and Federal mandates and health plan member contract language, including specific provisions/exclusions, take precedence over Medical Policy and must be considered first in determining eligibility for coverage. To verify a member’s benefits, contact Blue Cross and Blue Shield of Kansas Customer Service.

The BCBSKS Medical Policies contained herein are for informational purposes and apply only to members who have health insurance through BCBSKS or who are covered by a self-insured group plan administered by BCBSKS. Medical Policy for FEP members is subject to FEP medical policy which may differ from BCBSKS Medical Policy.

The medical policies do not constitute medical advice or medical care. Treating health care providers are independent contractors and are neither employees nor agents of Blue Cross and Blue Shield of Kansas and are solely responsible for diagnosis, treatment and medical advice.

If your patient is covered under a different Blue Cross and Blue Shield plan, please refer to the Medical Policies of that plan.

<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 molecular tests to rule out</td>
<td>• Surgical biopsy</td>
<td>• Disease-specific survival</td>
</tr>
<tr>
<td>and indeterminate findings on fine needle aspirate</td>
<td>malignancy and to avoid surgical biopsy or resection</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>
<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 molecular tests to rule in-</td>
<td>• Surgical management based on</td>
<td>• Disease-specific survival</td>
</tr>
<tr>
<td>and indeterminate findings on fine needle aspirate</td>
<td>malignancy and to guide surgical planning</td>
<td>clinicopathologic risk factors</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>
DESCRIPTION
To determine which patients need thyroid resection, many physicians will perform a cytologic examination of fine needle aspirate (FNA) samples from a thyroid lesion; however, this method has diagnostic limitations. As a result, assays using molecular markers have been developed to improve the accuracy of thyroid FNA biopsies.

OBJECTIVE
The objective of this policy is to evaluate whether testing for molecular markers in fine needle aspirates of the thyroid improves the net health outcome in individuals with thyroid nodule(s) with an indeterminate finding on the fine needle aspirate.

BACKGROUND
Thyroid Nodules
Thyroid nodules are common, present in 5% to 7% of the U.S. adult population. Most are benign, and most cases of thyroid cancer are curable by surgery when detected early.

Diagnosis
Fine needle aspirate (FNA) samples of the thyroid is currently the most accurate procedure to distinguish benign thyroid lesions and malignant ones, reducing the rate of unnecessary thyroid surgery for patients with benign nodules and triaging patients with thyroid cancer to appropriate surgery.

About 60% to 70% of thyroid nodules are classified cytologically as benign, and 4% to 10% of nodules are cytologically deemed malignant. However, the remaining 20% to 30% have equivocal findings, usually due to overlapping cytologic features between benign and malignant nodules; these nodules usually require surgery for a final diagnosis. Thyroid FNA cytology is classified by Bethesda System criteria into the following groups: nondiagnostic; benign; follicular lesion of undetermined significance (FLUS) or atypia of undetermined significance (AUS); follicular neoplasm (or suspicious for follicular neoplasm); suspicious for malignancy; and malignant. Lesions with FNA cytology in the AUS or FLUS or follicular neoplasm categories are often considered indeterminate.
Management
There is some individualization of management for patients with FNA-indeterminate nodules, but many patients will require a surgical biopsy, typically thyroid lobectomy, with intraoperative pathology. Consultation would typically be the next step in diagnosis. Approximately 80% of patients with indeterminate cytology undergo surgical resection; postoperative evaluation has revealed a malignancy rate ranging from 6% to 30%, making this a clinical process with very low specificity. Thus, if analysis of FNA samples could reliably identify the risk of malignancy as low, there is potential for patients to avoid surgical biopsy.

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

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

The diagnosis of malignancy in the case of PTC is primarily based on cytologic features. If FNA in a case of PTC is indeterminate, surgical biopsy with intraoperative pathology consultation is most often diagnostic, although its efficacy and therefore its use will vary across institutions, surgeons, and pathologists. In 2016, reclassification of encapsulated follicular-variant PTC as a noninvasive follicular tumor with papillary-like nuclei was proposed and largely adopted; this classification removes the word carcinoma from the diagnosis to acknowledge the indolent behavior of these tumors.

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

**Genetic Variants Associated with Thyroid Cancer**

Various genetic variants have been discovered in thyroid cancer. The most common 4 gene mutations are **BRAF** and **RAS** single nucleotide variants (SNVs), and **RET/PTC** and **PAX8/PPARγ** rearrangements.

Papillary carcinomas carry SNVs of the **BRAF** and **RAS** genes, as well as **RET/PTC** and **TRK** rearrangements, all of which are able to activate the mitogen-activated protein kinase pathway. These mutually exclusive variants are found in more than 70% of papillary carcinomas. **BRAF** SNVs are highly specific for PTC. Follicular carcinomas harbor either **RAS** SNVs or **PAX8/PPARγ** rearrangements. These variants have been identified in 70% to 75% of follicular carcinomas. Genetic alterations involving the PI3K/AKT signaling pathway also occur in thyroid tumors, although they are rare in well-differentiated thyroid cancers and have a higher prevalence in less differentiated thyroid carcinomas. Additional variants 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 SNVs located in the **RET** gene.

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

Telomerase reverse transcriptase (**TERT**) promoter variants occur with varying frequency in different thyroid cancer subtypes. Overall, **TERT C228T** or **C250T** variants have been reported in approximately 15% of thyroid cancers, with higher rates in the undifferentiated and anaplastic subtypes compared with the well-differentiated subtypes. **TERT** variants are associated with several demographic and histopathologic features such as older age and advanced TNM stage. **TERT** promoter variants have been reported to be independent predictors of disease recurrence and cancer-related mortality in well-differentiated thyroid cancer. Also, the co-occurrence of **BRAF** or **RAS** variants with **TERT** or **TP53** variants may identify a subset of thyroid cancers with unfavorable outcomes.

**Molecular Diagnostic Testing**

**Variant Detection and Rearrangement Testing**

SNVs in specific genes, including **BRAF**, **RAS**, and **RET**, and evaluation for rearrangements associated with thyroid cancers can be accomplished 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 \textit{BRAF} and \textit{RAS} variant analysis and testing for \textit{RET/PTC} and \textit{PAX8/PPARy} rearrangements.

The ThyroSeq v.3 Next-Generation Sequencing panel (CBLPath, Ocala) is an NGS panel of 112 genes. According to the CBLPath’s website, the test is indicated when FNA cytology suggests atypia of uncertain significance or follicular lesion of undetermined significance, follicular neoplasm or suspicious for follicular neoplasm, or suspicious for malignancy.\textsuperscript{15} 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 an NGS 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.

\textbf{Gene Expression Profiling}

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

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. In 2017, Veracyte migrated the Afirma GEC microarray analysis to a next-generation RNA sequencing platform and now markets the Afirma Gene Sequencing Classifier (Afirma GXC) which evaluates 10196 genes with 1115 core genes.

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]\textsuperscript{16}, Zheng et al [2015]\textsuperscript{17}); they are not addressed in this review.

ThyraMIR\textsuperscript{™} is a microRNA expression–based classifier intended for use in thyroid nodules with indeterminate cytology on FNA following a negative result from the ThyGenX\textsuperscript{™} Thyroid Oncogene Panel.

\textbf{Algorithmic Testing}

Algorithmic testing involves the use of two 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 variants (Afirma BRAF) or variants associated with medullary thyroid carcinoma (Afirma MTC). Table 1 describes the testing algorithms for Afirma MTC and Afirma BRAF.

Table 1. Afirma MTC and Afirma BRAF Testing Algorithms

<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 fine needle aspirate</td>
<td>“Indeterminate”</td>
<td>Afirma MTC</td>
</tr>
<tr>
<td>Afirma GEC</td>
<td>“Malignant” or “suspicious”</td>
<td>Afirma MTC</td>
</tr>
<tr>
<td>Afirma GEC</td>
<td>“Suspicious”</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 variants: (1) PCR-based methods may have low sensitivity, requiring that a large proportion of the nodule have a relevant variant; (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 amount of DNA that is difficult to isolate from small FNA samples.18

The testing strategy for both Afirma MTC and Afirma BRAF is to predict malignancy from an 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- vs 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 an 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 miRInform Thyroid test that assesses for 17 validated gene alterations.

ThyraMIR (Interpace Diagnostics, Parsippany, NJ) is a microRNA 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 first to 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 negative predictive value 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 must meet the general regulatory standards of the Clinical Laboratory Improvement Amendments. Thyroid variant testing and gene expression classifiers are available under the auspices of the Clinical Laboratory Improvement Amendments. Laboratories that offer laboratory-developed tests must be licensed by the Clinical Laboratory Improvement Amendments for high-complexity testing. To date, the U.S. Food and Drug Administration 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 the Food and Drug Administration through the premarket approval process to assess specific BRAF variants in melanoma tissue via real-time PCR. However, there are currently no diagnostic tests for thyroid cancer mutation analysis with approval from the Food and Drug Administration.

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>Predicate</th>
<th>Methodology</th>
<th>Analyte(s)</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afirma® GSC</td>
<td>Afirma®GEC</td>
<td>mRNA gene expression</td>
<td>1,115 genes</td>
<td>Benign/suspicious</td>
</tr>
<tr>
<td>Afirma® BRAF</td>
<td></td>
<td>mRNA gene expression</td>
<td>1 gene</td>
<td>Negative/positive</td>
</tr>
<tr>
<td>Afirma® MTC</td>
<td></td>
<td>mRNA gene expression</td>
<td></td>
<td>Negative/positive</td>
</tr>
<tr>
<td>ThyroSeq v3</td>
<td>ThyroSeq v2</td>
<td>Next-generation sequencing</td>
<td>112 genes</td>
<td>Specific gene variant/translocation</td>
</tr>
<tr>
<td>ThyGeNEXT®</td>
<td>ThyGenX®, miRInform®</td>
<td>Next-generation sequencing</td>
<td>10 genes and 32 gene fusions</td>
<td>Specific gene variant/translocation</td>
</tr>
<tr>
<td>ThyraMIR™</td>
<td></td>
<td>microRNA expression</td>
<td>10 microRNAs</td>
<td>Negative/positive</td>
</tr>
<tr>
<td>RosettaGX™ Reveal</td>
<td></td>
<td>microRNA expression</td>
<td>24 microRNAs</td>
<td>• Benign</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Suspicious for malignancy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• High risk for medullary carcinoma</td>
</tr>
</tbody>
</table>

FNA: fine needle aspirate; NGS: next-generation sequencing; PCR: polymerase chain reaction.

a The miRInform® test is the predicate test to ThyGenX™ and is not commercially available.
b Includes TERT.
c Available literature on TERT testing used PCR.
POLICY

A. The use of either Afirma Gene Expression Classifier or ThyroSeq in fine needle aspirates of thyroid nodules with indeterminate cytologic findings (ie, Bethesda diagnostic category III [atypia / follicular lesion of undetermined significance] or Bethesda diagnostic category IV [follicular neoplasm / suspicion for a follicular neoplasm]) may be considered medically necessary in patients who have ALL of the following characteristics:

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

B. The use of any of the following types of molecular marker testing or gene variant analysis in fine needle aspirates of thyroid nodules with indeterminate findings (Bethesda diagnostic category III [atypia / follicular lesion of undetermined significance] or Bethesda diagnostic category IV [follicular neoplasm / suspicion for a follicular neoplasm]) or suspicious findings (Bethesda diagnostic category V [suspicious for malignancy]) to rule in malignancy to guide surgical planning for initial resection rather than a 2-stage surgical biopsy followed by definitive surgery may be considered medically necessary:

1. ThyroSeq;
2. ThyraMIR microRNA / ThyGenX;
3. Afirma BRAF after Afirma Gene Expression Classifier; or
4. Afirma MTC after Afirma Gene Expression Classifier.

C. Gene expression classifiers, genetic variant analysis, and molecular marker testing in fine needle aspirates of the thyroid not meeting criteria outlined above, including, but not limited to, use of RosettaGX Reveal and single-gene TERT testing, are considered experimental / investigational.

Policy Guidelines

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

2. Use of molecular marker testing based on fine needle aspirate of a thyroid nodule to rule in malignancy prior to surgical biopsy may guide surgical planning,
particularly factors such as choice of surgical facility provider to ensure that the capability is available to conduct a frozen section pathologic reading during surgical biopsy so that surgical approach may be adjusted accordingly in 1 surgery.

3. **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.

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</td>
<td>Disease-associated variant identified in a proband for use in subsequent</td>
</tr>
<tr>
<td></td>
<td></td>
<td>targeted genetic testing in first-degree relatives</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>

**RATIONALE**

The most recent policy update includes a review of the literature through May 14, 2019.

Evidence reviews assess whether a medical test is clinically useful. A useful test provides information to make a clinical management decision that improves the net health outcome. That is, the balance of benefits and harms is better when the test is used to manage the condition than when another test or no test is used to manage the condition.
The first step in assessing a medical test is to formulate the clinical context and purpose of the test. The test must be technically reliable, clinically valid, and clinically useful for that purpose. Evidence reviews assess the evidence on whether a test is clinically valid and clinically useful. Technical reliability is outside the scope of these reviews, and credible information on technical reliability is available from other sources.

**Molecular Tests to Rule Out Malignancy**

**Clinical Context and Test Purpose**

One purpose of molecular testing in individuals with indeterminate findings on fine needle aspirate(s) (FNA) of thyroid nodules is to rule out malignancy and eliminate the need for surgical biopsy or resection.

The relevant question addressed in this evidence review is: Does molecular testing appropriately eliminate the need for surgical biopsy or resection and lead to improved health outcomes?

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

**Patients**

The relevant population of interest are individuals with indeterminate findings on FNAs of thyroid nodules who would be willing to undergo watchful waiting, depending on the results of their molecular testing. Patients with indeterminate findings after FNA of thyroid nodule presently proceed to surgical biopsy or resection.

**Interventions**

The test being considered is molecular testing, which includes either Afirma Gene Sequencing Classifier (GSC) (predicate Afirma Gene Expression Classifier [GEC]) or RosettaGX Reveal.

**Comparators**

The following practice is currently being used: standard surgical management through surgical biopsy or resection for biopsy.

**Outcomes**

The potential beneficial outcome of primary interest would be avoiding an unneeded surgical biopsy or resection (eg, lobectomy or hemithyroidectomy) in a true-negative thyroid nodule that is benign.

Potential harmful outcomes are those resulting from false-negative test results, which may delay diagnosis and surgical resection of thyroid cancer. For small, slow-growing tumors, it is uncertain that a delay in diagnosis would necessarily worsen health outcomes.

The time frame for evaluating the performance of the test is the time from the initial FNA to surgical biopsy or resection measured in weeks to months following an indeterminate result. Papillary thyroid cancer (PTC) is indolent, and a nodule could be observed for many years to ensure no clinical change.
**Afirma GSC**

**Technically Reliable**
Assessment of technical reliability focuses on specific tests and operators and requires a review of unpublished and often proprietary information. Review of specific tests, operators, and unpublished data are outside the scope of this evidence review, and alternative sources exist. This evidence review focuses on the clinical validity and clinical utility.

**Clinically Valid**
A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

**Prospective Clinical Validation**
Patel et al (2018) reported a validation study for the Afirma GSC test. The study included 210 thyroid nodules from 183 patients that had indeterminate results (Bethesda III or IV) on FNA, see Table 3. All FNA samples had been previously used in the validation of the Afirma GEC test as reported by Alexander et al (2012) in a 19-month, prospective, multicenter (49 academic and community sites) study. 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 three 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 for a blinded histopathological reference standard. After exclusion of 25 used for test validation and those without a valid GEC result, 265 indeterminate FNA samples were evaluated with the Afirma GEC. Of the 265 samples, 85 nodules were malignant; the GEC correctly identified 78 of the 85 as suspicious (92% sensitivity; 95% confidence interval [CI], 84% to 97%). Specificity was 52% (95% CI, 44% to 59%). The negative predictive value (NPV) ranged from 85% for "suspicious cytologic findings" to 95% for "atypia of undetermined clinical significance." Seven FNAs had false-negative results, six of which were thought to be due to hypocellular aspirate specimens.

The study reported by Patel et al (2018) used the banked samples which were reassayed with next-generation sequencing (NGS) for the Afirma GSC validation study. The previous central, blinded postoperative consensus histopathological diagnosis was used as the reference standard (210 samples) and all personnel were blinded to the other outcomes. The sensitivity of the Afirma GSC study was 91.1% with a specificity of 68.3% and NPV of 96.1% (95% CI 90 to 99) (see Table 4). There were four false negatives in patients with malignant nodules who would have been assigned for active observation. With a sensitivity that was similar to the Afirma GEC test, the specificity was improved with Afirma GSC. There were no notable study limitations.
Table 3. Study Characteristics for Afirma GSC

<table>
<thead>
<tr>
<th>Study</th>
<th>Study Population</th>
<th>Design</th>
<th>Reference Standard</th>
<th>Threshold for Positive Index Test</th>
<th>Timing of Reference and Index Tests</th>
<th>Blinding of Assessors</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patel et al (2018)</td>
<td>183 patients with 210 indeterminate thyroid nodules by FNA</td>
<td>Multicenter, Non-concurrent prospective validation trial</td>
<td>Consensus histopathology diagnosis</td>
<td>Central, blinded histopathological review from Alexander et al (2012)</td>
<td>Assessors were blinded to the pathology</td>
<td>Samples were previously used to validate Afirma GEC</td>
<td></td>
</tr>
</tbody>
</table>

FNA: Fine needle aspirate; Afirma GEC: gene expression classifier; Afirma GSC: gene sequencing classifier.

Table 4. Clinical Validity for Afirma GSC

<table>
<thead>
<tr>
<th>Study</th>
<th>Initial N</th>
<th>Final N</th>
<th>Excluded Samples</th>
<th>Prevalence of Condition</th>
<th>Clinical Validity (95% Confidence Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patel et al (2018)</td>
<td>210 nodules</td>
<td>191 nodules</td>
<td>19 with insufficient residual RNA</td>
<td>Sensitivity: 91.1 (79 to 98) Specificity: 68.3 (60 to 76) PPV: 47.1 (36 to 58) NPV: 96.1 (90 to 99)</td>
<td></td>
</tr>
</tbody>
</table>

Afirma GSC: gene sequencing classifier; NPV: negative predictive value; PPV: positive predictive value.

Retrospective Clinical Validation
Santhanam et al (2016) conducted a meta-analysis of studies reporting on the performance of the predicate Afirma GEC in cytologically indeterminate nodules. Seven studies met inclusion criteria, which required that studies reported on the use of the Afirma GEC in nodules found indeterminate on FNA (including atypia of uncertain significance [AUS] or follicular lesion of undetermined significance [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 three studies, unclear in three studies, and high in one study. Although the authors reported pooled results, these results (particularly specificity) were likely biased given the lack of reference standard diagnosis for most lesions in the included studies (see the following section and Table 3).

Retrospective multicenter and single-center studies, including Harrell and Bimston (2014), Lastra et al (2014), McIver et al (2014), Yang et al (2016), Witt (2016), Baca et al (2017), Harrison et al (2017), Kay-Rivist et al (2017), Hang et al (2017), Samulski et al (2016) and Angell et al (2015) have reported on the diagnostic accuracy of the Afirma GEC and Afirma GSC. 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. Harrell et al (2019) reported a retrospective comparison of Afirma GEC (2011 to July 2017) and Afirma GSC (August 2017 through June 2018) for indeterminate FNA. GSC identified fewer indeterminate
nodules as suspicious (54/139, 38.8%) compared to GEC (281/481, 58.4%) and led to a lower surgery rate, decreasing from 56% in the GEC group to 31% in the GSC group.

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 (2015) retrospectively compared clinical outcomes for individuals who had indeterminate FNA cytology and Afirma GEC benign results with individuals who had cytologically benign nodule results. A total of 95 cytologically indeterminate and 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 and Afirma GEC benign nodules (13.8% vs 0.9%, p<0.001).

Sipos et al (2016) retrospectively analyzed nonacademic medical practices using the Afirma GEC to determine the long-term nonoperative rate of thyroid nodules with benign results. Of the patients with Afirma benign results during 36 months of follow-up, 17.3% underwent surgery. Eighty-eight percent of all surgeries were performed within the first two years after a benign Afirma GEC result.

Deaver et al (2018) retrospectively reviewed the performance of Bethesda category III and IV thyroid nodules and Afirma GEC results from over 5 years. The duration of follow-up of 73 patients with FNA indeterminate/ GEC benign nodules was 46 months (range, 8 to 76 months) for Bethesda III/GEC Benign nodules and 62 months (range, 11 to 74 months) for Bethesda IV/GEC Benign nodules. Five of the cohort (6.8%) underwent surgery over the follow-up period with a single case of incidental microcarcinoma, indicating that GEC Benign nodules remained benign.

Clinically Useful
A test is clinically useful if the use of the results informs management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.

Direct Evidence
Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from randomized controlled trials.

No evidence directly demonstrating improved outcomes in patients managed with the Afirma GEC was identified.
Chain of Evidence
Indirect evidence on clinical utility rests on clinical validity. If the evidence is insufficient to demonstrate test performance, no inferences can be made about clinical utility. Because no direct evidence of utility was identified, a chain of evidence was developed, which addresses two 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 (SFN) 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 resect thyroid nodules with indeterminate cytology and Afirma GEC benign results in a sample of 395 nodules from 368 patients. Surgery was performed in 7.6% of the patients with indeterminate cytology and a benign GEC result, less than the historical rate of thyroid resection (74%) in patients with indeterminate cytology.

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

Abeykoon et al (2016) studied the impact of implementing Afirma GEC at a single center. Surgical recommendations for patients with indeterminate thyroid nodules decreased from 81.5% pre-Afirma GEC to 50% post-Afirma GEC. The rate of malignant surgical pathology diagnosis increased from 20% pre-Afirma GEC to 85.7% post-Afirma. The implementation of Afirma GEC decreased the number of surgical recommendations and increased the rate of malignancy detected for patients who received a surgical biopsy.

Chaudhary et al (2016) studied the impact on surgical outcomes pre- and postimplementation of Afirma GEC. A total of 158 FNAs were sent for Afirma GEC with 73 suspicious and 8 benign Afirma cases going for surgeries. Compared with before implementation of Afirma GEC, the rate for surgical biopsy decreased from 61% to 54% but was not statistically significant. In the SFN, the rate of surgical biopsy significantly decreased from 76% to 52%.

Improved Outcomes
A simplified decision model was developed for use with Afirma GEC in individuals with cytologically indeterminate FNA samples. 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 with 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 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 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, the 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. The standard of care for thyroid nodules is based on an intervention that is stratified by FNA cytology results, which are grouped into categories with differing prognosis. Avoiding invasive surgery in situations where patients are at very low likelihood of having an invasive tumor is likely beneficial. Among the low-risk population, the alternative to surgical biopsy is ongoing active surveillance.

*RosettaGX Reveal*

**Technically Reliable**

Assessment of technical reliability focuses on specific tests and operators and requires a review of unpublished and often proprietary information. Review of specific tests, operators, and unpublished data are outside the scope of this evidence review, and alternative sources exist. This evidence review focuses on the clinical validity and clinical utility.

**Clinically Valid**

A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

Lithwick-Yanai et al (2017) described the development and initial clinical validation of the RosettaGX Reveal quantitative real-time polymerase chain reaction assay for 24 microRNA samples in a multicenter, retrospective cohort study using 201 FNA smears. The results of the clinical validation study are reported in Table 5.
## Table 5. Clinical Validity for RosettaGX Reveal

<table>
<thead>
<tr>
<th>Study</th>
<th>Initial N</th>
<th>Final N</th>
<th>Excluded Samples</th>
<th>Prevalence of Condition</th>
<th>Sensitivity (95% Confidence Interval)</th>
<th>Specificity (95% Confidence Interval)</th>
<th>Pos. Pred. Value</th>
<th>Neg. Pred. Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lithwick-Yanai et al (2017)</td>
<td>201 FNA smears</td>
<td>189 passing QC</td>
<td>12</td>
<td></td>
<td>85 (74 to 93)</td>
<td>72 (63 to 79)</td>
<td>NR</td>
<td>91 (84 to 96)</td>
</tr>
<tr>
<td></td>
<td>150 with consensus agreement</td>
<td></td>
<td></td>
<td></td>
<td>98 (87 to 100)</td>
<td>78 (69 to 85)</td>
<td>NR</td>
<td>99 (94 to 100)</td>
</tr>
</tbody>
</table>

FNA: fine needle aspirate; NPV: negative predictive value; NR: not reported; PPV: positive predictive value; QC: quality control.

Wals et al (2018) reported a blinded evaluation of RosettaGX Reveal in 81 archived FNA smears that had Afirma GEC results and histopathology. Afirma GEC had been requested following indeterminate FNA and had classified 74 nodules as suspicious and 7 as benign. The 81 patients underwent surgery based on Afirma GEC results or clinical factors. The final diagnosis from histopathology was 63 benign and 18 malignant thyroid nodules. Reveal classified 14 of the 18 malignant nodules as suspicious for a sensitivity of 77.8% and specificity of 60.3%.

No prospective clinical studies for RosettaGX Reveal were identified.

**Clinically Useful**

**Direct Evidence**

No evidence directly demonstrating improved outcomes in patients managed with the RosettaGX Reveal was identified.

**Chain of Evidence**

Indirect evidence on clinical utility rests on clinical validity. If the evidence is insufficient to demonstrate test performance, no inferences can be made about clinical utility.

### Section Summary: Molecular Tests to Rule Out Malignancy

In a multicenter validation study, Afirma GSC was reported to have a high (NPV 96%; 95% C.I. 90%-99%). These results are consistent with an earlier study on the Afirma GEC in the same study population. In other multicenter and single-center studies, there is suggestive evidence that rates of malignancy are low in Afirma patients who are classified as benign, but the exact NPV is unknown. The available evidence suggests that physician decision making about surgery is altered by Afirma GSC results, however, it should be noted that long-term follow-up of patients with thyroid nodules who avoided surgery based on GEC results is limited. A chain of evidence can be constructed to establish the potential for clinical utility with Afirma GSC testing in cytologically indeterminate lesions, but there is only a single study with the currently marketed test reporting a true NPV. Clinical input, obtained in 2017, supported the use of the predicate Afirma GEC in FNA of thyroid nodules with indeterminate cytologic findings (ie, Bethesda diagnostic category III [atypia/follicular lesion of undetermined significance] or Bethesda diagnostic category IV [follicular neoplasm/suspicion for a follicular neoplasm]) to rule out malignancy and to avoid surgical biopsy.
For the RosettaGX Reveal test, two retrospective clinical validation studies have been reported. No prospective studies for patients managed with the RosettaGX Reveal were identified, so the clinical validity remains uncertain.

**Molecular Tests to Rule in Malignancy**

**Clinical Context and Test Purpose**
The purpose of testing for molecular markers (eg, single nucleotide variants and gene rearrangements) in individuals with indeterminate findings on FNA of thyroid nodules is to rule in malignancy and to guide 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 are individuals with indeterminate findings on FNA(s) of thyroid nodules. Patients with indeterminate findings would presently proceed to surgical biopsy perhaps with intraoperative pathology consultation (ie, intraoperative frozen section) if available.

**Interventions**
The test being considered is testing for molecular markers (eg, single nucleotide variants and gene rearrangements) with Afirma Braf and Afirma MTC to guide surgical planning to ensure the capability for intraoperative pathologic confirmation of malignancy to adjust to definitive surgery for initial resection if appropriate.

**Comparators**
The following practices are currently being used: standard surgical management through surgical resection, including a two-stage surgical biopsy (ie, lobectomy) followed by definitive surgery (ie, hemithyroidectomy or thyroidectomy).

**Outcomes**
The potential beneficial outcome of primary interest is appropriate surgical planning in the preoperative period (eg, hemithyroidectomy or thyroidectomy when malignancy is predicted). This has the potential benefit of reducing the likelihood of having the patient repeating 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 results. However, the use of intraoperative confirmation of malignancy through frozen pathology section in patients with positive molecular marker testing would mitigate any risk of inappropriately performing more extensive thyroidectomy in the absence of malignancy.
Timing
The time frame for evaluating the performance of the test varies from the initial FNA to surgical resection to weeks to months following an indeterminate result.

Setting
The primary setting would be in endocrinology.

Gene Expression Classifiers to Predict Malignancy
Technically Reliable
Assessment of technical reliability focuses on specific tests and operators and requires a review of unpublished and often proprietary information. Review of specific tests, operators, and unpublished data are outside the scope of this evidence review, and alternative sources exist. This evidence review focuses on the clinical validity and clinical utility.

Clinically Valid
A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

Less evidence exists on the validity of gene expression profiling to rule in malignancy (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.

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

Afirma BRAF and Afirma MTC
Diggans et al (2015), described the development and validation of the Afirma BRAF test, for a subset of 213 thyroid nodule FNA samples for which histopathology was available, Afirma BRAF test results were compared with pathologic findings. Afirma BRAF classified all histopathologically benign samples as BRAF V600E-negative (specificity, 100%; 95% CI, 97.4% to 100%). Of the 73 histopathologically malignant samples, the Afirma BRAF test identified 32 as BRAF-positive (sensitivity, 43.8%; 95% CI, 32.2% to 55.9%).

In a study describing the development and validation of the Afirma MTC classifier, Kloos et al (2016) evaluated the MTC classifier in a sample of 10488 thyroid nodule FNA samples referred for GEC testing. In this sample, 43 cases were Afirma MTC-positive, of which 42 were considered to be clinically consistent with MTC on pathology or biochemical testing, for a positive predictive value (PPV) of 97.7% (95% CI, 86.2% to 99.9%).
**Genetic Variants Association with Tumor Behavior**

The presence of *BRAF* or *TERT* variants is strongly associated with malignancy in thyroid nodule FNA samples. *BRAF* or *TERT* variants have also been associated with more aggressive clinicopathologic features in individuals diagnosed with PTC.

Adeniran et al (2011) 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 an extrathyroidal extension or aggressive histologic features were positive for a *BRAF* variant. The authors concluded that patients with an equivocal cytologic diagnosis and a *BRAF* V600E variant could be candidates for total thyroidectomy and central lymph node dissection.

Xing et al (2009) 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%; p=0.002). During a median follow-up of 3 years (range, 0.6-10 years), PTC persistence or recurrence was seen in 36% of *BRAF* variant-positive patients and 12% of *BRAF* variant-negative patients, with an odds ratio (OR) 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.

Yin et al (2016) reported on a systematic review and meta-analysis evaluating *TERT* promoter variants and aggressive clinical behaviors in PTC. Eight eligible studies (total n=2035 patients; range, 30-507) were included. Compared with wild-type, *TERT* promoter variant status was associated with lymph node metastasis (OR=1.8; 95% CI, 1.3 to 2.5; p=0.001), extrathyroidal extension (OR=2.6; 95% CI, 1.1 to 5.9; p=0.03), distant metastasis (OR=6.1; 95% CI, 3.6 to 10.3; p<0.001), advanced TNM stages III or IV (OR=3.2; 95% CI, 2.3 to 4.5; p<0.001), poor clinical outcome (persistence or recurrence; OR=5.7; 95% CI, 3.6 to 9.3; p<0.001), and mortality (OR=8.3; 95% CI, 3.8 to 18.2; p<0.001).

**Clinically Useful**

A test is clinically useful if the use of the results informs management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.
Direct Evidence
Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from randomized controlled trials.

Testing For Specific Variants Associated with Thyroid Cancer
(eg, BRAF V600E, TERT, and RET variants, RET/PTC and PAX8/PPARY rearrangements) is generally designed to "rule in" cancer in nodules with indeterminate cytology on FNA.48, (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- vs a total thyroidectomy or performance of central neck dissection.

In a retrospective analysis, Yip et al (2014) reported on outcomes after implementation of an algorithm incorporating molecular testing of thyroid FNA samples to guide the extent of initial thyroid resection.49, 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/PPARY rearrangements) was prospectively obtained for all FNAs with indeterminate cytology (FLUS, 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 a molecular diagnosis, a positive molecular diagnostic test was considered an indication for an initial total thyroidectomy. Patients with FLUS and negative molecular diagnostic results were followed with repeat FNA, followed by lobectomy or total thyroidectomy if indeterminate pathology persisted. Patients with a 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/PPARY rearrangements (2/56 [4%]). Compared with those managed without molecular diagnostics (63%), patients managed with molecular diagnostics (69%) were nonsignificantly less likely to undergo total thyroidectomy as an initial procedure (p=0.08). However, they had nonsignificantly higher rates of central compartment lymph node dissection (21% vs 15%, p=0.06). Across both cohorts, 25% (170/671) of patients had clinically significant thyroid cancer, with no difference in thyroid cancer rates based on the type of initial surgery (26% for total thyroidectomy vs 22% for lobectomy, p=0.3). The incidence of clinically significant thyroid cancer after initial lobectomy (ie, requiring a 2-stage surgery) was significantly lower for patients managed with molecular diagnostics (17% vs 43%, p<0.001). An indeterminate FNA result had a sensitivity and specificity for the diagnostic of thyroid cancer of 89% and 27%, respectively, with a PPV of 29% and an NPV of 88%. The addition of molecular diagnostics to FNA results increased the specificity for a cancer diagnosis to 95% and the PPV to 82%.
Chain of Evidence
Indirect evidence on clinical utility rests on clinical validity. If the evidence is insufficient to demonstrate test performance, no inferences can be made about clinical utility.

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

Section Summary: Molecular Tests to Predict Malignancy
The available evidence has suggested that the use of variant testing in thyroid FNA samples is generally associated with 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 a 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. A statement from the American Thyroid Association provides some guidelines for surgeons managing patients with indeterminate nodules. However, adoption of these guidelines in practice and outcomes associated with them is uncertain.

Molecular Tests to Rule Out and Rule in Malignancy
Clinical Context and Test Purpose
The purpose of the ThyroSeq v3 test and the combined ThyGeNEXT Thyroid Oncogene Panel plus ThyraMIR microRNA classifier in individuals with indeterminate findings on FNA(s) of thyroid nodules is to predict malignancy and inform surgical planning decisions with positive results using ThyroSeq v3 or the ThyGeNEXT, and if negative, to predict benignancy using ThyraMIR microRNA classifier to eliminate or necessitate the need for surgical biopsy and guide surgical planning.

The relevant question addressed in this evidence review is: Does the ThyroSeq v3 test or the combined use of ThyGeNEXT and ThyraMIR appropriately eliminate or necessitate the need for surgical resection or biopsy and lead to improved health outcomes?

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

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

Interventions
The tests being considered are either: (a) the ThyroSeq v3 test or (b) the combined ThyGeNEXT Thyroid Oncogene Panel and ThyraMIR microRNA classifier testing.
Comparators
The following practices are currently being used: surgical biopsy and/or standard surgical management through surgical resection.

Outcomes
The potential beneficial outcomes of primary interest are using a true-negative result to avoid an unneeded surgical biopsy or using a true-positive result to guide surgical resection (e.g., hemithyroidectomy or thyroidectomy).

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 and delay in diagnosis.

The time frame for evaluating the performance of the test varies from the initial FNA to surgical resection to weeks to months following an indeterminate result.

ThyroSeq v3 Test
Technically Reliable
Assessment of technical reliability focuses on specific tests and operators and requires a review of unpublished and often proprietary information. Review of specific tests, operators, and unpublished data are outside the scope of this evidence review, and alternative sources exist. This evidence review focuses on the clinical validity and clinical utility.

Clinically Valid
A test must detect the presence or absence of a condition, the risk of developing a condition in the future, or treatment response (beneficial or adverse).

Nikiforova et al (2018) reported on the performance of ThyroSeq v3 with 112 genes. The training sample included 238 surgically removed tissue samples consisting of 205 thyroid tissue samples representing all main types of benign and malignant tumors and nontumoral conditions. The validation sample included an independent set of 175 FNA samples of indeterminate cytology (see Table 6). Using the cutoff identified in the training set, the ThyroSeq v3 sensitivity was 98% (95% CI, 93% to 99%), specificity was 82% (95% CI, 72% to 89%), with accuracy of 91% (95% CI, 86% to 94%) (see Table 7).

Steward et al (2019) conducted a multicenter validation study of ThyroSeq v3 in 256 patients with an indeterminate FNA who had surgery with histopathology (see Table 6). Histopathology was reviewed by a central pathology panel and both cytologists and pathologists were blinded to the molecular results. For a benign result, ThyroSeq v3 had a sensitivity of 93%, a specificity of 81%, PPV of 68%, and NPV of 97% (see Table 7). Out of 152 test-negative samples, 5 (3%) were false-negatives. There were 105 cases with positive results, defined as cancer or noninvasive follicular thyroid neoplasm with papillary-like features. Two nodules had high-risk TERT or TP53 variants (both positive for cancer), 13 had variants in BRAF V600E or NTRK3, or BRAF, or RET fusions (all positive for cancer), and 60 nodules were positive for variants in RAS, BRAF K601E, PTEN, IDH2, or DICER1 or PPARF-THADA fusion (37 [62%] positive for cancer). No major limitations in study design and conduct of this validation study.
study were identified. Because the nodules with low cancer probability genetic alterations were removed for histological analysis, the long-term clinical impact of the genetic alterations could not be determined.

**Table 6.** Study Characteristics of Clinical Validity ThyroSeq v3

<table>
<thead>
<tr>
<th>Study</th>
<th>Study Population</th>
<th>Design</th>
<th>Reference Standard</th>
<th>Threshold for Positive Index Test</th>
<th>Timing of Reference and Index Tests</th>
<th>Blinding of Assessors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nikiforov et al (2018)⁵¹</td>
<td>175 samples with indeterminate cytology and known surgical follow-up</td>
<td>Retrospective</td>
<td>Histopathologic diagnosis</td>
<td>Cutoffs determined in the training sample</td>
<td>Samples were tested after surgical outcome was known</td>
<td>Unclear</td>
</tr>
<tr>
<td>Steward et al (2019) ⁵²</td>
<td>256 patients (286 nodules) with an indeterminate FNA (Bethesda III, IV, or V) and underwent thyroid surgery</td>
<td>Multicenter (10 sites) prospective validation study</td>
<td>Central pathology review</td>
<td>Classified as malignant or NIFPT or benign</td>
<td>Cross-sectional</td>
<td>Yes</td>
</tr>
</tbody>
</table>

FNA: fine needle aspirate; NIFPT: noninvasive follicular thyroid neoplasm with papillary-like features

**Table 7.** Clinical Validity of ThyroSeq v3

<table>
<thead>
<tr>
<th>Study</th>
<th>Initial N</th>
<th>Final N</th>
<th>Excluded Samples</th>
<th>Prevalence of Condition</th>
<th>Sensitivity (95% Confidence Interval)</th>
<th>Specificity (95% Confidence Interval)</th>
<th>PPV (95% Confidence Interval)</th>
<th>NPV (95% Confidence Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nikiforov et al (2018)⁵¹</td>
<td>286</td>
<td>57</td>
<td>29 (10%)</td>
<td>30%</td>
<td>98 (93 to 100)</td>
<td>81 (72 to 89)</td>
<td>68 (58 to 76)</td>
<td>97 (93 to 99)</td>
</tr>
<tr>
<td>Steward et al (2019) ⁵²</td>
<td>175</td>
<td>175</td>
<td></td>
<td></td>
<td>98 (93 to 100)</td>
<td>81 (72 to 89)</td>
<td>68 (58 to 76)</td>
<td>97 (93 to 99)</td>
</tr>
</tbody>
</table>

NPV: negative predictive value; PPV: positive predictive value.

Additional studies describing the clinical validity of the ThyroSeq v2 panel in external settings (outside of the institution where it was developed) have reported on the diagnostic performance to predict malignancy in thyroid nodules that are indeterminate on FNA have been reported (see Table 8). These studies differed from the previous studies in that noninvasive follicular thyroid neoplasm with papillary-like nuclear features was classified as not malignant for calculation of performance characteristics.
Table 8. Additional Clinical Validity Studies of ThyroSeq to Predict Malignancy in Indeterminate Thyroid FNA Samples

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Genes and Rearrangements Tested</th>
<th>Insufficient or Inadequate for Analysis</th>
<th>Measures of Agreement (95% CI), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valderrabano et al (2017)53.</td>
<td>190 indeterminate thyroid nodules</td>
<td>ThyroSeq v2 (60+ genes)</td>
<td>2</td>
<td>Sen 70 (46 to 88) Spec 77 (66 to 85) PPV 42 (25 to 61) NPV 91 (82 to 97)</td>
</tr>
<tr>
<td>Taye et al (2018)54.</td>
<td>156 indeterminate thyroid nodules</td>
<td>ThyroSeq v2 (60+ genes)</td>
<td>3</td>
<td>Sen 89 (52 to 100) Spec 43 (29 to 58) PPV 22 (10 to 38) NPV 96 (78 to 99)</td>
</tr>
</tbody>
</table>

CI: confidence interval; FNA: fine needle aspiration; NPV: negative predictive value; PPV: positive predictive value; Sen: sensitivity; Spec: specificity.

Additional studies describing the clinical validity of the genes that comprise the ThyroSeq panel or other individual variants and combinations of variants to predict malignancy in thyroid nodules that are indeterminate on FNA have been reported. The results that pertain to the use of gene testing in indeterminate thyroid nodules are summarized in Table 9.

Table 9. Clinical Validity of Molecular Markers to Predict Malignancy in Indeterminate Thyroid FNA Samples

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Genes and Rearrangements Tested</th>
<th>Insufficient or Inadequate for Analysis</th>
<th>Measures of Agreement, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moses et al (2010)55.</td>
<td>110 indeterminate thyroid nodules</td>
<td>BRAF, KRAS, NRAS, RET/PTC1, RET/PTC3, NTRK1</td>
<td>2</td>
<td>Sen 38 Spec 95 PPV 67 NPV 79 Acc 77</td>
</tr>
<tr>
<td>Ohori et al (2010)56.</td>
<td>100 patients with 117 atypia or follicular lesions of uncertain significance</td>
<td>BRAF, NRAS, HRAS, KRAS, RET/PTC1, RET/PTC3, PAX8/PPARγ</td>
<td>NR</td>
<td>Sen 60 Spec 100 PPV 100 NPV 92 Acc 93</td>
</tr>
<tr>
<td>Beaudenon-Huibregtse et al (2014)57.</td>
<td>53 nodules with indeterminate or nondiagnostic FNA</td>
<td>BRAF, HRAS, KRAS, NRAS, PAX8-PPARγ, RET-PTC1, RET-PTC3</td>
<td>48</td>
<td>Sen 48 Spec 89 PPV 81 NPV 64</td>
</tr>
</tbody>
</table>

Acc: accuracy; 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.
ThyGenX Thyroid Oncogene Panel and ThyraMIR microRNA Classifier

Technically Reliable
Assessment of technical reliability focuses on specific tests and operators and requires a review of unpublished and often proprietary information. Review of specific tests, operators, and unpublished data are outside the scope of this evidence review, and alternative sources exist. This evidence review focuses on the clinical validity and clinical utility.

Clinically Valid
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 follicular neoplasm or SFN across 12 endocrinology centers. A summary of the sensitivity and specificity of the combined test is listed in Table 10.

Table 10. Summary of Clinical Validity for 17-Variant Panel and ThyraMIR on FNA Samples

<table>
<thead>
<tr>
<th>Groups</th>
<th>No. of Cases</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>PPV (95% CI)</th>
<th>NPV (95% CI)</th>
<th>Odds Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort</td>
<td>109</td>
<td>89 (73 to 97)</td>
<td>85 (75 to 92)</td>
<td>74 (58 to 86)</td>
<td>94 (85 to 98)</td>
<td>44 (13 to 151)</td>
</tr>
<tr>
<td>AUS/FLUS</td>
<td>58</td>
<td>94 (73 to 100)</td>
<td>80 (64 to 91)</td>
<td>68 (46 to 85)</td>
<td>97 (84 to 100)</td>
<td>68 (8 to 590)</td>
</tr>
<tr>
<td>FN/SFN</td>
<td>51</td>
<td>82 (57 to 96)</td>
<td>91 (76 to 98)</td>
<td>82 (57 to 96)</td>
<td>91 (76 to 98)</td>
<td>48 (9 to 269)</td>
</tr>
</tbody>
</table>

Adapted from Labourier et al (2015). AUS: atypia of undetermined significance; CI: confidence interval; FLUS: follicular lesion of undetermined significance; FN: follicular neoplasm; FNA: fine needle aspiration; NPV: negative predictive value; PPV: positive predictive value; SFN: suspicious for a follicular neoplasm.

Clinically Useful
A test is clinically useful if the use of the results informs management decisions that improve the net health outcome of care. The net health outcome can be improved if patients receive correct therapy, or more effective therapy, or avoid unnecessary therapy, or avoid unnecessary testing.

Direct Evidence
Direct evidence of clinical utility is provided by studies that have compared health outcomes for patients managed with and without the test. Because these are intervention studies, the preferred evidence would be from randomized controlled trials.

Direct evidence for the clinical utility for the ThyroSeq v2 test and the combined ThyGenX and ThyraMIR diagnostic testing algorithm is lacking.

Chain of Evidence
Indirect evidence on clinical utility rests on clinical validity. If the evidence is insufficient to demonstrate test performance, no inferences can be made about clinical utility.

A chain of evidence may be constructed to infer the potential clinical utility of the combined diagnostic testing algorithm. No studies using ThyGenX NGS panel in FNA samples were identified. However, available evidence has suggested that the use of variant testing using NGS in thyroid FNA samples is generally associated with high specificity and PPV for clinically significant thyroid cancer. There is the 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, the variant analysis does not achieve an NPV sufficiently high enough 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 using a 17-variant panel with ThyraMIR showed an NPV of 94%. Therefore, the high NPV of ThyraMIR has the potential to accurately predict benignancy and triage patients to active surveillance.

Section Summary: Molecular Markers to Rule Out and Rule in Malignancy
Evidence for the clinical validity of the ThyroSeq v3 NGS panel comes from a prospective clinical validity study, the performance characteristics were sensitivity, 93%; specificity, 81%; PPV, 68%; NPV, 97%. In 2 independent validation studies with a predicate test (ThyroSeq v2) in which noninvasive follicular thyroid neoplasm with papillary-like nuclear features was categorized as not malignant, performance characteristics were lower and variable (sensitivity, 70%-89%; specificity, 43%-77%; PPV, 22%-42%; NPV, 91%-96%).

Evidence for the clinical validity of combined testing for miRNA gene expression using ThyraMIR and a targeted 17-variant panel comes from 2 retrospective studies using 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%, a 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.

Direct evidence for the clinical utility for the ThyroSeq v2 test and the combined ThyGenX and ThyraMIR reflex testing is lacking. However, available evidence has suggested that testing for gene variants and rearrangements can predict malignancy and inform surgical planning decisions when the test is positive. Pooled retrospective and prospective clinical validation studies of ThyroSeq v2 have reported a combined NPV of 96% (95% CI, 92% to 95%) and PPV of 83% (95% CI, 72% to 95%) and might potentially assist in selecting patient to avoid surgical biopsy in negative and guide surgical planning if positive. 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 testing may provide incremental information in identifying patients appropriately for active surveillance, but improvements in health outcomes are still uncertain.

Clinical input, obtained in 2017, considered ThyroSeq v2 to provide a clinically meaningful improvement for patients with indeterminate cytologic findings to rule out malignancy and avoid surgical biopsy and in patients with cytologic findings suspicious for malignancy to guide surgical planning for the initial resection.

SUMMARY OF EVIDENCE
To determine which patients need thyroid resection, many physicians will perform a cytologic examination of FNA samples from a thyroid lesion; however, this method has diagnostic
limitations. As a result, assays using molecular markers have been developed 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 molecular tests to rule out malignancy and to avoid surgical biopsy or resection, the evidence includes a prospective clinical validity study with the Afirma GSC and a chain of evidence to support clinical utility. The relevant outcomes are disease-specific survival, test accuracy and validity, morbid events, and resource utilization. In a multicenter validation study, the Afirma GSC was reported to have a high (NPV 96%; 95% CI, 90%-99%). These results are consistent with an earlier study on the Afirma GEC in the same study population. In other multicenter and single-center studies, there is suggestive evidence that rates of malignancy are low in Afirma patients who are classified as benign, but the exact NPV is unknown. The available evidence suggests that the decisions a physician makes regarding surgery are altered by Afirma GEC/GSC results; however, it should be noted that long-term follow-up of patients with thyroid nodules who avoided surgery based on GEC results is limited. A chain of evidence can be constructed to establish the potential for clinical utility with GEC testing in cytologically indeterminate lesions, but there is only a single study of the marketed test reporting a true NPV. Clinical input, obtained in 2017, supported the use of the previous version of the Afirma test in FNA of thyroid nodules with indeterminate cytologic findings to rule out malignancy and avoid surgical biopsy with an acceptably low trade-off in missed malignancy. The evidence is sufficient to determine that the technology improves the net health outcome.

For individuals with thyroid nodule(s) and indeterminate findings on FNA who receive FNA sample testing with molecular tests to rule in malignancy and to guide surgical planning, the evidence includes prospective and retrospective studies of clinical validity. The 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 for the initial resection. Prospective studies in additional populations are needed to validate these results. 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. Clinical input, obtained in 2017, considered ThyraMIR microRNA/ThyGenX, Afirma BRAF after Afirma GEC, and Afirma MTC after Afirma GEC to provide a clinically meaningful improvement for patients with cytologic findings suspicious for malignancy to guide surgical planning for the initial resection. The evidence is sufficient to determine that the technology results in a meaningful improvement in the net health outcome.

For individuals with thyroid nodule(s) and indeterminate findings on FNA who receive FNA sample testing with molecular tests to rule out malignancy and avoid surgical biopsy or to rule in malignancy for surgical planning, the evidence includes multiple retrospective and prospective clinical validation studies for the ThyroSeq test and 2 retrospective clinical validation studies that used a predicate test 17-variant panel (miRInform) test to the current ThyGenX and ThyraMIR. The relevant outcomes are disease-specific survival, test accuracy and validity, morbid events, and resource utilization. In a retrospective validation study on FNA samples, the
17-variant panel (miRInform) test and ThyraMIR had a sensitivity of 89%, and an NPV of 94%. A prospective clinical validation study of ThyroSeq v3 reported an NPV of 97% and PPV of 68%. No studies were identified demonstrating the diagnostic characteristics of the marketed ThyGenX. No studies were identified demonstrating evidence of direct outcome improvements. A chain of evidence for the ThyroSeq v3 test and combined ThyGenX and ThyraMIR testing would rely on establishing clinical validity. Clinical input, obtained in 2017, considered ThyroSeq v2 to provide a clinically meaningful improvement for patients with indeterminate cytologic findings to rule out malignancy and avoid surgical biopsy and in patients with cytologic findings suspicious for malignancy to guide surgical planning for the initial resection. The evidence is sufficient to determine that the technology results in a meaningful improvement in the net health outcome.

### 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.

#### 2017 Input

In response to requests, clinical input on 7 tests for molecular markers was received from 9 respondents, including 1 specialty society-level response, 1 physician from academic center, and 7 physicians from 2 health systems while this policy was under review in 2017. Based on the evidence and independent clinical input, the clinical input supports that the following indications provide a clinically meaningful improvement in the net health outcome and are consistent with generally accepted medical practice:

- Use of the following types of molecular marker testing in fine needle aspirate (FNA) of thyroid nodules with indeterminate cytologic findings (i.e., Bethesda diagnostic category III [atypia/follicular lesion of undetermined significance] or Bethesda diagnostic category IV [follicular neoplasm/suspicion for a follicular neoplasm]) to rule out malignancy and to avoid surgical biopsy:
  - Afirma Gene Expression Classifier; or
  - ThyroSeq v2

- Use of the following type of molecular marker testing in FNA of thyroid nodules with indeterminate cytologic findings or Bethesda diagnostic category V (suspicious for malignancy) to rule in the presence of malignancy to guide surgical planning for the initial resection rather than a 2-stage surgical biopsy followed by definitive surgery:
  - ThyroSeq v2;
  - ThyraMIR microRNA/ThyGenX;
  - Afirma BRAF after Afirma Gene Expression Classifier; or
  - Afirma MTC after Afirma Gene Expression Classifier.

Based on the evidence and independent clinical input, the clinical input does not support whether the following indication provides a clinically meaningful improvement in the net health outcome or is consistent with generally accepted medical practice:

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

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

PRACTICE GUIDELINES AND POSITION STATEMENTS
American Association of Clinical Endocrinologists et al
In 2016, the American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi updated their joint guidelines and made the following statements on molecular testing for cytologically indeterminate thyroid nodules:

- “Cytopathology expertise, patient characteristics, and prevalence of malignancy within the population being tested impact the negative predictive values (NPVs) and positive predictive values (PPVs) for molecular testing.”
- “Consider the detection of BRAF and RET/PTC and, possibly, PAX8/PPARG and RAS mutations if such detection is available.”
- “TERT mutational analysis on FNA, when available, may improve the diagnostic sensitivity of molecular testing on cytologic samples.”
- “Because of the insufficient evidence and the limited follow-up, we do not recommend either in favor of or against the use of gene expression classifiers (GECs) for cytologically indeterminate nodules.”

For the role of molecular testing for deciding extent of surgery the following recommendations were made:

- “Currently, with the exception of mutations such as BRAFV600E that have a PPV approaching 100% for papillary thyroid carcinoma (PTC), evidence is insufficient to recommend in favor of or against the use of mutation testing as a guide to determine the extent of surgery.”

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. 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 and follicular neoplasm (FN) or suspicious for follicular neoplasm (SFN) on cytology (see Table 11):
Table 11. Molecular Diagnostics in Thyroid Nodules That Are AUS or FLUS or FN or SFN on Cytology

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>SOR</th>
<th>QOE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AUS or FLUS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;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.”</td>
<td>Weak</td>
<td>Moderate</td>
</tr>
<tr>
<td>&quot;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.”</td>
<td>Strong</td>
<td>Low</td>
</tr>
<tr>
<td><strong>FN or SFN</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;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.”</td>
<td>Weak</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

AUS: atypia of undetermined significance; FLUS: follicular lesion of undetermined significance; FN: follicular neoplasm; FNA: fine needle aspirate; QOE: quality of evidence; SFN: suspicious for follicular neoplasm; SOR: strength of 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 (v.1.2019) guidelines on the treatment of thyroid cancer comment on the use of molecular diagnostics in thyroid cancer.\(^{61}\) For thyroid nodules evaluated with FNA, molecular diagnostics may be employed when lesions are suspicious for:

- Follicular or Hürthle cell neoplasms.
- Atypia of undetermined significance or follicular lesion of undetermined significance.

The guidelines state that molecular diagnostics may not perform well for Hurthle cell neoplasms.

**U.S. PREVENTIVE SERVICES TASK FORCE RECOMMENDATIONS**

Not applicable.

**ONGOING AND UNPUBLISHED CLINICAL TRIALS**

Some currently unpublished trials that might influence this review are listed in Table 12.

Table 12. Summary of Key Trials

<table>
<thead>
<tr>
<th>NCT No.</th>
<th>Trial Name</th>
<th>Planned Enrollment</th>
<th>Completion Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ongoing</td>
<td>Genomic Profiling of Nodular Thyroid Disease and Thyroid Cancer</td>
<td>200</td>
<td>Jan 2020</td>
</tr>
</tbody>
</table>

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

CPT/HCPCS

81345  TERT (telomerase reverse transcriptase) (eg, thyroid carcinoma, glioblastoma multiforme) gene analysis, targeted sequence analysis (eg, promoter region)
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, PDGFRα, PDGFRβ, 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)
0018U  Oncology (thyroid), microRNA profiling by RT-PCR of 10 microRNA sequences, utilizing fine needle aspirate, algorithm reported as a positive or negative result for moderate to high risk of malignancy
0026U  Oncology (thyroid), DNA and mRNA of 112 genes, next-generation sequencing, fine needle aspirate of thyroid nodule, algorithmic analysis reported as a categorical result ("Positive, high probability of malignancy" or "Negative, low probability of malignancy")

ICD-10 Diagnoses

C73   Malignant neoplasm of thyroid gland
D44.0 Neoplasm of uncertain behavior of thyroid gland

REVISIONS

<table>
<thead>
<tr>
<th>Date</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-01-2017</td>
<td>In Coding section:</td>
</tr>
<tr>
<td></td>
<td>- Added CPT code: 0018U.</td>
</tr>
<tr>
<td>04-15-2018</td>
<td>Updated Description section.</td>
</tr>
<tr>
<td></td>
<td>In Policy section:</td>
</tr>
<tr>
<td></td>
<td>- In Item A, added &quot;either&quot;, &quot;or ThyroSeq v2&quot;, &quot;nodules with&quot;, &quot;cytologic findings (ie, Bethesda diagnostic category III [atypia], &quot;or Bethesda diagnostic category IV&quot;, &quot;suspicion for a follicular neoplasm&quot;, and removed &quot;that are cytologically considered to be&quot; to read, &quot;The use of either the Afiama Gene Expression Classifier or ThyroSeq v2 in fine needle aspirates of thyroid nodules with indeterminate cytologic findings (ie, Bethesda diagnostic category III [atypia/follicular lesion of undetermined significance] or Bethesda diagnostic category IV [follicular neoplasm/suspicion for a follicular neoplasm]) may be considered medically necessary in patients who have ALL of the following characteristics.&quot;</td>
</tr>
<tr>
<td>Date</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Updated Description section.</td>
</tr>
<tr>
<td></td>
<td>In Policy section:</td>
</tr>
<tr>
<td></td>
<td>- In Item C, added &quot;and single-gene TERT testing&quot; to read, &quot;Gene expression</td>
</tr>
<tr>
<td></td>
<td>classifiers, genetic variant analysis, and molecular marker testing in</td>
</tr>
<tr>
<td></td>
<td>fine needle aspirates of the thyroid not meeting criteria outlined above,</td>
</tr>
<tr>
<td></td>
<td>including, but not limited to, use of RosettaGX Reveal and single-gene</td>
</tr>
<tr>
<td></td>
<td>TERT testing, are considered experimental / investigational.&quot;</td>
</tr>
<tr>
<td></td>
<td>Updated Rationale section.</td>
</tr>
<tr>
<td></td>
<td>In Coding section:</td>
</tr>
<tr>
<td></td>
<td>- Added CPT code: 0026U.</td>
</tr>
<tr>
<td></td>
<td>Updated References section.</td>
</tr>
<tr>
<td>01-01-2019</td>
<td>In Coding section:</td>
</tr>
<tr>
<td></td>
<td>- Added new CPT code: 81345.</td>
</tr>
<tr>
<td>08-28-2019</td>
<td>Updated Description section.</td>
</tr>
<tr>
<td></td>
<td>In Policy section:</td>
</tr>
<tr>
<td></td>
<td>- In Item A, removed &quot;v2&quot; to read, “The use of either Afirma Gene Expression</td>
</tr>
<tr>
<td></td>
<td>Classifier or ThyroSeq in fine needle aspirates of thyroid nodules with</td>
</tr>
<tr>
<td></td>
<td>indeterminate cytologic findings (ie, Bethesda diagnostic category III</td>
</tr>
<tr>
<td></td>
<td>[atypia / follicular lesion of undetermined significance] or Bethesda</td>
</tr>
<tr>
<td></td>
<td>diagnostic category IV [follicular neoplasm / suspicion for a follicular</td>
</tr>
<tr>
<td></td>
<td>neoplasm]) may be considered medically necessary in patients who have ALL</td>
</tr>
<tr>
<td></td>
<td>of the following characteristics”.</td>
</tr>
<tr>
<td></td>
<td>- In Item B 1, removed &quot;v2&quot; to read, “ThyroSeq”.</td>
</tr>
<tr>
<td></td>
<td>Updated Rationale section.</td>
</tr>
<tr>
<td></td>
<td>Updated References section.</td>
</tr>
<tr>
<td></td>
<td>Removed Appendix section.</td>
</tr>
</tbody>
</table>
REFERENCES
13. Song YS, Lim JA, Choi H, et al. Prognostic effects of TERT promoter mutations are enhanced by coexistence with BRAF or RAS mutations and strengthen the risk prediction by the ATA or TNM staging system in differentiated thyroid cancer patients. Cancer. May 1 2016;122(9):1370-1379. PMID 26969876.


**Other References**

1. Blue Cross and Blue Shield of Kansas Pathology Liaison Committee, May 2018.