Clinical Policy Title: Afirma™ gene expression classifier for indeterminate thyroid nodules

Clinical Policy Number: 02.01.11
Effective Date: September 1, 2014
Initial Review Date: March 19, 2014
Most Recent Review Date: March 15, 2017
Next Review Date: March 2018

Related policies:
None.

ABOUT THIS POLICY: AmeriHealth Caritas Pennsylvania has developed clinical policies to assist with making coverage determinations. AmeriHealth Caritas Pennsylvania’s clinical policies are based on guidelines from established industry sources, such as the Centers for Medicare & Medicaid Services (CMS), state regulatory agencies, the American Medical Association (AMA), medical specialty professional societies, and peer-reviewed professional literature. These clinical policies along with other sources, such as plan benefits and state and federal laws and regulatory requirements, including any state- or plan-specific definition of “medically necessary,” and the specific facts of the particular situation are considered by AmeriHealth Caritas Pennsylvania when making coverage determinations. In the event of conflict between this clinical policy and plan benefits and/or state or federal laws and/or regulatory requirements, the plan benefits and/or state and federal laws and/or regulatory requirements shall control. AmeriHealth Caritas Pennsylvania’s clinical policies are for informational purposes only and not intended as medical advice or to direct treatment. Physicians and other health care providers are solely responsible for the treatment decisions for their patients. AmeriHealth Caritas Pennsylvania will update its clinical policies as necessary. AmeriHealth Caritas Pennsylvania’s clinical policies are not guarantees of payment.

Coverage policy

AmeriHealth Caritas Pennsylvania considers the use of the Afirma™ gene expression classifier test to be clinically proven and, therefore, medically necessary when the following criteria are met:

- Presence of thyroid nodules with one or more prior non-diagnostic fine needle aspirates.
- Any one or combination of the following clinical or microscopic indicators for high risk or suspicion of malignancy:
  - Growth over time.
  - Hoarseness, difficulty swallowing or breathing.
  - History of exposure to ionizing radiation.
  - Hardness compared to rest of gland.
  - Cervical adenopathy.

  Histology on pathology report:
  - Micro-follicular.
  - Trabecular.
  - Hürthle cell.

Limitations:

Policy contains:
- Thyroid disease (nodules and cancer).
- Molecular testing thyroid nodules.
- Biopsy procedures (fine needle, core, surgical).
• One test per patient per lifetime.
• In the event of a second unrelated thyroid nodule, medical necessity must meet the same criteria as the initial thyroid nodule.

Alternative covered services:

Open or closed thyroid biopsy.

Background

Diagnostic testing has been the subject of much methodological discussion, summarized by Gray (1997) and Sackett (1991). Understanding diagnostic test analysis through the progression of research is important to interpretation of the literature and to rational clinical selection and use of tests.

Briefly, evaluation of a diagnostic technology entails five levels (Banta 1993):
• Technical evaluation: the technical output gives accurate information concerning the structure of the body part imaged.
• Diagnostic accuracy: this output concerns information that potentially improves the clinician’s ability to diagnose disease and assess the patient’s prognosis.
• Diagnostic impact: the information can alter plans for additional diagnostic tests.
• Therapeutic impact: the information can lead to changes in therapeutic plans for patients.
• Health impact: the end result may be improved patient outcome.

Genetic testing or gene expression testing/classification includes a variety of laboratory tests (analysis of deoxyribonucleic acid [DNA], ribonucleic acid [RNA], genes or gene products) for the purposes of diagnosing disease, assisting in treatment decisions and in the case of thyroid nodules, making operative decisions about surgical resection of the nodule, predicting future disease, and even identifying carriers of disease or facilitating prenatal testing.

Evaluating molecular genetic tests in the absence of a known consistent reference procedure has always been problematic. Marchionni (2008) pointed out that the field of gene expression testing lacks the “gold standard” (outside the tests under evaluation themselves: microarrays and reverse transcriptase polymerase chain reaction [RT-PCR]) traditionally used to evaluate diagnostic test accuracy or operating characteristics (e.g., sensitivity, specificity, positive and negative predictive values). These shortcomings make definitive evaluation of analytic validity and test accuracy difficult to categorize with certainty.

A common occurrence in adults (3 percent to 7 percent on physical exam), nodular thyroid disease is heralded by disordered growth of thyroid cells, often combined with fibrosis. Nodules may be solitary or multiple and functional or non-functional. Benign nodules are classified as follicular epithelial cell adenomas, macrofollicular (colloid), normo-follicular (simple), micro-follicular, trabecular or Hürthle cell variant. The risk of malignancy varies according to histologic appearance: risk is very low for macro-
follicular and normo-follicular adenomas, while micro-follicular, trabecular and Hürthle cell variants raise greater concern.

Risk factors for malignancy in a thyroid nodule include history of head and neck irradiation, age less than 20 or more than 45 years, bilateral disease, size greater than 4 cm, new or enlarging neck mass, male gender, family history of thyroid cancer or multiple endocrine neoplasia (MEN) type 2, vocal cord paralysis and hoarseness, nodule fixation to adjacent structures, and extra-thyroidal and lymph node involvement. Thyroid tumors are classified according to histologic features of the follicular epithelium from which they often originate. Differentiated tumors such as papillary thyroid cancer (PTC) or follicular thyroid cancer (FTC) are often curable and the prognosis is good for patients diagnosed with early-stage disease. Anaplastic thyroid cancer (ATC) is aggressive, responds poorly to treatment and is associated with a bleak prognosis. Incidence of thyroid cancer (approximately 9/100,000 people) increases with age and reaches a plateau after about 50. It is twice as common in females, with male gender associated with worse prognosis. All cell types within the thyroid can produce neoplasms. These include follicular cells, calcitonin-producing C cells, lymphocytes, stromal and vascular elements, and tumors metastatic from other sites.

About one-third of palpable nodules are thyroid cysts, recognized by ultrasound appearance or fine needle aspiration (FNA) of large amounts of pink or straw-colored fluid (colloid). Mixed cystic/solid lesions are common and require aspiration of cellular components. Twenty to 35 percent of aspirates will be indeterminate. These require surgical resection of the nodule or entire thyroid gland, after which 80 percent will prove to be benign.

The high incidence of benignity in thyroid biopsy has fostered the development of laboratory tests to resolve ambiguity and ameliorate the plethora of nonmalignant biopsies yielded by traditional clinical practice. Among these is the Afirma™ Thyroid FNA analysis (Veracyte Inc.), which examines genes known to be involved in cell growth to classify indeterminate nodules as either clearly benign or suspicious.

**Searches**

AmeriHealth Caritas Pennsylvania searched PubMed and the databases of:
- UK National Health Services Centre for Reviews and Dissemination.
- Agency for Healthcare Research and Quality’s National Guideline Clearinghouse and other evidence-based practice centers.
- The Centers for Medicare & Medicaid Services (CMS).

We conducted searches on January 30, 2017. Search terms were: “thyroid nodules,” “Afirma™,” and “genetic tests.”

We included:
- **Systematic reviews**, which pool results from multiple studies to achieve larger sample sizes and greater precision of effect estimation than in smaller primary studies. Systematic reviews use predetermined transparent methods to minimize bias, effectively treating the review as a scientific endeavor, and are thus rated highest in evidence-grading hierarchies.
• **Guidelines based on systematic reviews.**

• **Economic analyses**, such as cost-effectiveness, and benefit or utility studies (but not simple cost studies), reporting both costs and outcomes — sometimes referred to as efficiency studies — which also rank near the top of evidence hierarchies.

**Findings**

Brito (2013) evaluated the accuracy of thyroid nodule ultrasound in predicting cancer over a systematic review of 31 cohort studies from 1985 – 2012 inclusive of a total of 18,288 nodules (average size 15mm). The authors found several characteristic ultrasound features of malignancy (i.e., taller than wider, and spongiform appearance). However, they noted only low to moderate quality evidence that ultrasound features are accurate predictors.

Najafzadeh (2012) studied the cost-effectiveness of molecular diagnostic testing to improve the preoperative diagnosis in thyroid cancer and facilitate treatment planning. The authors found that, when used with FNA, molecular diagnostic testing resulted in a gain of 0.046 quality-adjusted life years (QALYs) and savings of $1,087 over a 10-year horizon.

Samir (2012) found that molecular diagnostic testing in combination with ultrasound-guided percutaneous thyroid nodule core (CB) biopsy in 82 patients with one or more prior non-diagnostic fine FNA had significantly higher diagnostic yield than either component alone.

Li (2011) conducted an economic evaluation for molecular test for cytologically indeterminate thyroid nodules with a sensitivity of 91 percent and specificity of 75 percent using 142-gene RNA expression by proprietary algorithm (Afirma™). One notable criticism of the study was its sponsorship: it was funded by the Veracyte company which makes the Afirma™ test.

**Policy updates:**

A retrospective study (n=63) of noninvasive follicular variant papillary thyroid carcinoma (NFVPTC) a distinct subset of FTC with an exceedingly indolent clinical course, examined the clinical efficacy of the Afirma™ gene-expression classifier in the management of thyroid nodules with indeterminate FNA results (Wong 2016).

Surgical pathology specimens demonstrated 16 (25 percent) FTCs, five (8 percent) follicular thyroid carcinomas, one (2 percent) classical type PTC, and 41 (65 percent) benign tumors/nodules. Of the 16 FTCs, 14 (88 percent) were NFVPTCs and accounted for 64 percent of the carcinomas. The authors concluded that the Afirma™ test detects NFVPTCs and that many of the carcinomas detected by Afirma™ are NFVPTCs.

Santhanam (2016) demonstrated that the gene expression classifier (GEC) could identify a benign gene expression signature in those thyroid nodules of indeterminate cytology with a negative predictive value of ≥95 percent. A total of 58 studies and a meta-analysis revealed a high pooled sensitivity and a low
specificity for the Afirma™ GEC test — two key indicators that the test is an excellent tool to rule out malignancy.

Sipos (2016) examined the operative rate of patients with thyroid nodules and a benign Afirma™ GEC result. The secondary endpoint of this study was the treating physician's opinion regarding the safety of GEC use compared to the hypothetical situation of providing thyroid nodule management without the GEC. During 36 months (+/- 3 months) of follow up, 17 of 98 patients (17.3 percent) with benign GEC results underwent surgery. The treating physicians reported that patient safety was improved by using the GEC compared to not using the GEC in 78 of 91 (86 percent) cases. A non-operative approach to follow-up was felt to be a safe alternative to diagnostic surgery by the majority of responsible physicians in the study.

Angeli (2015) found that the Afirma™ GEC test has a high negative predictive value for ruling out malignancy in thyroid nodules with indeterminate cytology. Many patients with a cytologically indeterminate and GEC benign (Cyto-I/GEC-B) nodule safely undergo monitoring instead of diagnostic surgery.

Thyroid FNA was used to assess appropriate management of nodular thyroid lesions safely in a retrospective study by Celik (2015). Medical records of patients with thyroid FNA and GEC results were obtained from archived material. Results were compared to thyroidectomy histologic diagnoses. Among 66 patients with FNA results (47 women and 19 men aged 26-89 years [mean, 59.4 years]), surgical reports were available for 38. Afirma™ GEC results were nondiagnostic for 10 of 66 (15.2 percent), benign for 22 (33.3 percent), and suspicious for 34 (51.5 percent). Surgical diagnosis was available for 38 of 66 patients (57.6 percent); GEC results for 6 (15.8 percent) of these were nondiagnostic, 27 (71.0 percent) were suspicious, and 5 (13.2 percent) were benign. One of 6 (16.7 percent) samples with nondiagnostic results, 1 of 5 (20 percent) with benign results, and 15 of 27 (55.6 percent) with suspicious results were malignant on histology. Papillary carcinoma was the most common tumor type (15 of 38; 39.5 percent). The authors concluded that Afirma™ GEC results minimized the number of unnecessary operations.

**Summary of clinical evidence:**

<table>
<thead>
<tr>
<th>Citation</th>
<th>Content, Methods, Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wong (2016)</td>
<td>Key points:</td>
</tr>
<tr>
<td>Noninvasive Follicular</td>
<td>- A retrospective study (n=63) of NFVPTC.</td>
</tr>
<tr>
<td>Variant of Papillary</td>
<td>- Surgical pathology specimens demonstrated 16 (25%) FTCs, five (8%) follicular thyroid carcinomas, one (2%) classical type PTC, and 41 (65%) benign tumors/nodules.</td>
</tr>
<tr>
<td>Thyroid Carcinoma and</td>
<td>- Of the 16 FTCs, 14 (88%) were NFVPTCs and accounted for 64 percent of the carcinomas.</td>
</tr>
<tr>
<td>the Afirma™ GEC-</td>
<td></td>
</tr>
<tr>
<td>Expression Classifier.</td>
<td></td>
</tr>
<tr>
<td>Santhanam (2016)</td>
<td>Key points:</td>
</tr>
<tr>
<td>Gene expression classifier for the diagnosis of indeterminate thyroid</td>
<td>- Verified that the GEC test methodology could identify a benign gene expression signature in those nodules with indeterminate cytology with a negative predictive value of greater than 95%.</td>
</tr>
<tr>
<td>Citation</td>
<td>Content, Methods, Recommendations</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------------</td>
</tr>
</tbody>
</table>
| nodules  | • Specifically looked at the performance of the Afirma™ gene expression classifier in predicting benign and malignant nodules in patients with cytologically indeterminate nodules.  
• The reference standard for determination of benign or malignant nodules was the histopathology of the thyroidectomy specimen.  
• The pooled sensitivity of the GEC was 95.7% (95% CI 92.2-97.9, I (2) value 45.4%, p = 0.09), and the pooled specificity was 30.5% (95% CI 26.0-35.3, I (2) value 92.1%, p < 0.01).  
• The meta-analysis revealed a high pooled sensitivity and a low specificity for the Afirma™ GEC test for indeterminate thyroid nodules.  
• The authors concluded that this makes it an excellent tool to rule out malignancy. |
| Sipos (2016) Long-Term Non-Operative Rate Of Thyroid Nodules | **Key points:**  
• Examined the operative rate in patients with a benign result from the Afirma™ GEC during long-term follow-up at nonacademic medical facilities.  
• The secondary endpoint of this study was the treating physician’s opinion regarding the safety of GEC use compared to the hypothetical situation of providing thyroid nodule management without the GEC.  
• Treating physicians reported that patient safety was improved by using the GEC compared to not using the GEC in 78 of 91 (86%) cases.  
• The authors concluded that a benign result on the GEC is associated with a reduction in the rate of thyroid surgeries compared to published data when patients are followed for 36 months after testing. |
| Angeli (2015) Afirma™ Benign Thyroid Nodules | **Key points:**  
• Retrospective study of the sonographic changes and clinical outcomes for patients with Cyto-I/GEC-B nodules compared with patients with cytologically benign (Cyto-B) nodules.  
• Ninety-five Cyto-I/GEC-B nodules in 90 patients were identified. Five patients underwent primary surgical resection. Of the remaining 90 nodules, 58 (64.4%) had sonographic follow-up available at a median of 13 months (range 4-40 mo).  
• Cyto-I/GEC-B nodules showed similar growth compared with 1224 Cyto-B nodules using either of the following criteria: 20% or greater in two dimensions (8.6% vs 8.3%, P = .80) or 50% or greater in volume (17.2% vs 13.8%, P = .44).  
• Thyroidectomies were more frequent in the Cyto-I/GEC-B group (13.8% vs 0.9%, P < .0001), but cancer was found in only one patient, with no evidence of persistent disease after initial treatment.  
• Although Cyto-I/GEC-B nodules were more frequently resected, only one malignancy was found. |
| Celik (2015) Afirma™ GEC and thyroid lesions | **Key points:**  
• Retrospective study of 66 patients with non-diagnostic but suggestive (for cancer) thyroid FNA results (47 women and 19 men aged 26-89 years [mean, 59.4 years]).  
• Afirma™ GEC results were nondiagnostic for 10 of 66 (15.2%), benign for 22 (33.3%), and suspicious for 34 (51.5%).  
• Surgical reports were available for 38 patients. |
GEC results for 6 (15.8%) of these were nondiagnostic, 27 (71.0%) were suspicious, and 5 (13.2%) were benign.

One of 6 (16.7%) samples with nondiagnostic results, 1 of 5 (20%) with benign results, and 15 of 27 (55.6%) with suspicious results were malignant on histology.

Papillary carcinoma was the most common tumor type (15 of 38; 39.5%).

The authors concluded that Afirma™ GEC results minimize the number of unnecessary operations.

Key points:
- 31 cohort studies (total 18,288 nodules, average size 15mm) from 1985 – 2012 evaluated the accuracy of thyroid nodule ultrasound in predicting cancer.
- Ultrasound features of malignancy were consistently found to be taller than wider, with a benign, spongiform appearance.
- There was low-/moderate-quality evidence that ultrasound features other than those noted above are not accurate predictors.

Key points:
- Economic analysis of the cost-effectiveness of molecular diagnostic testing to improve preoperative diagnosis in thyroid cancer.
- Assuming 95% sensitivity and specificity when used with FNA, molecular diagnostic testing resulted in a gain of 0.046 QALYs and savings of $1,087 over 10-year period.

Key points:
- Cross-sectional diagnostic accuracy study of 82 subjects from 2006 – 2008 examined ultrasound-guided percutaneous thyroid nodule core biopsy in patients with one or more prior non-diagnostic fine needle (FNA) aspirate.
- Reference standard was office follow-up over 4 – 37 mo (mean 180) for development of clinically evident malignancy.
- Combination had significantly higher diagnostic yield than either component alone.

Key points:
- Economic evaluation of molecular testing for cytologically indeterminate thyroid nodules
- Found selectivity of 91% and specificity of 75% using 142-gene RNA expression by proprietary algorithm (Afirma™ by Veracyte).
- Authors concluded that Afirma™ test was cost-effective vs. FNA alone.
- Study funded by Veracyte.

Citation | Content, Methods, Recommendations
---|---
7 | • GEC results for 6 (15.8%) of these were nondiagnostic, 27 (71.0%) were suspicious, and 5 (13.2%) were benign.
\[\text{One of 6 (16.7%) samples with nondiagnostic results, 1 of 5 (20%) with benign results, and 15 of 27 (55.6%) with suspicious results were malignant on histology.}\]
\[\text{Papillary carcinoma was the most common tumor type (15 of 38; 39.5%).}\]
\[\text{The authors concluded that Afirma™ GEC results minimize the number of unnecessary operations.}\]
Brito (2013) | The accuracy of thyroid nodule ultrasound to predict thyroid cancer
\[\text{Key points:}\]
\[\text{• 31 cohort studies (total 18,288 nodules, average size 15mm) from 1985 – 2012 evaluated the accuracy of thyroid nodule ultrasound in predicting cancer.}\]
\[\text{• Ultrasound features of malignancy were consistently found to be taller than wider, with a benign, spongiform appearance.}\]
\[\text{• There was low-/moderate-quality evidence that ultrasound features other than those noted above are not accurate predictors.}\]
\[\text{Key points:}\]
\[\text{• Economic analysis of the cost-effectiveness of molecular diagnostic testing to improve preoperative diagnosis in thyroid cancer.}\]
\[\text{• Assuming 95% sensitivity and specificity when used with FNA, molecular diagnostic testing resulted in a gain of 0.046 QALYs and savings of $1,087 over 10-year period.}\]
Samir (2012) | Ultrasound-guided percutaneous thyroid nodule core biopsy
\[\text{Key points:}\]
\[\text{• Cross-sectional diagnostic accuracy study of 82 subjects from 2006 – 2008 examined ultrasound-guided percutaneous thyroid nodule core biopsy in patients with one or more prior non-diagnostic fine needle (FNA) aspirate.}\]
\[\text{• Reference standard was office follow-up over 4 – 37 mo (mean 180) for development of clinically evident malignancy.}\]
\[\text{• Combination had significantly higher diagnostic yield than either component alone.}\]
Li (2011) | Cost-effectiveness of a novel molecular test for cytologically indeterminate thyroid nodules
\[\text{Key points:}\]
\[\text{• Economic evaluation of molecular testing for cytologically indeterminate thyroid nodules}\]
\[\text{• Found selectivity of 91% and specificity of 75% using 142-gene RNA expression by proprietary algorithm (Afirma™ by Veracyte).}\]
\[\text{• Authors concluded that Afirma™ test was cost-effective vs. FNA alone.}\]
\[\text{• Study funded by Veracyte.}\]

References

Professional society guidelines/other:


**Peer-Reviewed References:**


**CMS National Coverage Determination (NCDs):**

No NCDs identified as of the writing of this policy.

**Local coverage determination (LCDs):**

No LCDs identified as of the writing of this policy.

**Commonly submitted codes**

Below are the most commonly submitted codes for the service(s)/item(s) subject to this policy. This is not an exhaustive list of codes. Providers are expected to consult the appropriate coding manuals and bill accordingly.

<table>
<thead>
<tr>
<th>CPT Codes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>81479</td>
<td>Unlisted molecular pathology procedure</td>
</tr>
<tr>
<td>81545</td>
<td>Oncology (thyroid), gene expression analysis of 142 genes, utilizing fine needle aspirate, algorithm reported as a categorized result (benign or suspicious).</td>
</tr>
<tr>
<td>ICD-10 Code</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>C73</td>
<td>Malignant neoplasm of thyroid gland</td>
</tr>
<tr>
<td>D44.0</td>
<td>Neoplasm of uncertain behavior of thyroid gland</td>
</tr>
<tr>
<td>D44.2</td>
<td>Neoplasm of uncertain behavior of parathyroid gland</td>
</tr>
<tr>
<td>D44.9</td>
<td>Neoplasm of uncertain behavior of unspecified endocrine gland</td>
</tr>
<tr>
<td>E01.1</td>
<td>Iodine-deficiency related multinodular (endemic) goiter</td>
</tr>
<tr>
<td>E04.1</td>
<td>Nontoxic single thyroid nodule</td>
</tr>
<tr>
<td>E04.2</td>
<td>Nontoxic multinodular goiter</td>
</tr>
<tr>
<td>E04.8</td>
<td>Other specified nontoxic goiter</td>
</tr>
<tr>
<td>E04.9</td>
<td>Nontoxic goiter, unspecified</td>
</tr>
<tr>
<td>R13.0</td>
<td>Aphagia</td>
</tr>
<tr>
<td>R13.10</td>
<td>Dysphagia, unspecified</td>
</tr>
<tr>
<td>R49.0</td>
<td>Dysphonia</td>
</tr>
<tr>
<td>R59.0</td>
<td>Localized enlarged lymph nodes</td>
</tr>
<tr>
<td>R59.1</td>
<td>Generalized enlarged lymph nodes</td>
</tr>
<tr>
<td>R59.9</td>
<td>Enlarged lymph nodes, unspecified</td>
</tr>
<tr>
<td>Z92.3</td>
<td>Personal history of irradiation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HCPCS Level II Codes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>