Afirma® thyroid FNA analysis for indeterminate thyroid nodules

Clinical Policy ID: CCP.1099
Recent review date: 3/2020
Next review date: 7/2021
Policy contains: Molecular testing; thyroid neoplasm; thyroid nodule.

AmeriHealth Caritas has developed clinical policies to assist with making coverage determinations. AmeriHealth Caritas’ 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 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’ 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’ clinical policies are reflective of evidence-based medicine at the time of review. As medical science evolves, AmeriHealth Caritas will update its clinical policies as necessary. AmeriHealth Caritas’ clinical policies are not guarantees of payment.

Coverage policy

Afirma® Thyroid Analysis (Veracyte, Inc., South San Francisco, California) is clinically proven and, therefore, medically necessary to rule out thyroid neoplasm when all of the following criteria are met (National Comprehensive Cancer Network, 2019):

- One of the following tests is used:
  - Afirma Gene Expression Classifier test.
  - Afirma Genomic Sequencing Classifier test.
- Member is at least 18 years old.
- Thyroid nodule is at least 1 cm on ultrasonography.
- Presence of thyroid nodules with one or more prior non-diagnostic fine needle aspirates, described as either:
  - Atypia (or follicular lesion) of undetermined significance (Bethesda System for Reporting Thyroid Cytopathology Category III; Cibas, 2017).
  - Follicular neoplasm or suspicious for follicular neoplasm, including Hürthle cell (Bethesda System for Reporting Thyroid Cytopathology Category IV).
- Results are expected to influence clinical management (Gharib, 2016).
- Member is not undergoing thyroid surgery for diagnostic confirmation.
The Afirma Xpression Atlas test is investigational and, therefore, not medically necessary.

Limitations

Frequency of testing using the Afirma Gene Expression Classifier test or Afirma Genomic Sequencing Classifier test is limited to once per lifetime per member. In the event of a second unrelated thyroid nodule, additional testing may be medically necessary if the same criteria as the initial thyroid nodule are met.

For testing multiple samples, reimbursement for Afirma testing is limited to once per date of service regardless of the number of nodules submitted for testing.

For Medicare members only

The Afirma Gene Expression Classifier test (Veracyte, Inc., South San Francisco, California) is clinically proven and, therefore, medically necessary to rule out thyroid neoplasm when both criteria are met (Centers for Medicare & Medicaid Services L35396):

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

The medical necessity of the Afirma Gene Expression Classifier test is limited to once per lifetime per member. Additional testing may be medically necessary in the unlikely situation of a second, unrelated thyroid nodule with indeterminate pathology occur, upon appeal with supporting documentation.

Alternative covered services

- Thyroid stimulating hormone levels.
- Conventional imaging (e.g., ultrasonography, elastosonography).
- Functional and molecular imaging (e.g., $^{99m}$Tc-MIBI scintigraphy, positron emission tomography).
- Open or closed thyroid biopsy.

Background

Thyroid cancer is the fastest-growing cancer in the United States, with an estimated 52,070 new cases expected in 2019 (American Cancer Society, 2019). This trend is attributed primarily to the increased use of thyroid ultrasound, which can detect small, indolent thyroid nodules that, otherwise, would have gone undetected. Thyroid cancer affects females more often than males and usually occurs in people ages 25 to 65 years.

There are four main types of thyroid cancer that vary in aggressiveness (National Cancer Institute, 2019). Papillary (the most common) and follicular thyroid carcinomas are well-differentiated tumors, highly treatable, and usually curable. Hürthle cell carcinoma is a variant of follicular carcinoma with a poorer prognosis. Less common anaplastic thyroid cancers are poorly differentiated or undifferentiated and aggressive, metastasize early, and have a poorer prognosis. Medullary thyroid tumors are well-differentiated tumors with an intermediate prognosis.
The significance of Hürthle cells in thyroid pathology is uncertain, as they can present in both benign and malignant neoplastic thyroid diseases (Cannon, 2011). No definitive cytological criteria exist for determining the benign or malignant nature of Hürthle cells before or during thyroidectomy. Clinically, Hürthle cell carcinoma has a distinct genetic profile and more aggressive behavior when compared with papillary and follicular thyroid carcinoma. Malignant Hürthle cells are poorly responsive to chemotherapy, do not concentrate iodine, and are generally not radiosensitive. Therefore, a surgical approach is undertaken for diagnosis and treatment.

Levels of thyroid stimulating hormone, clinical features, and high-resolution ultrasound findings of the thyroid and neck determine risk of malignancy. Cytologic examination of the thyroid with ultrasound-guided, fine-needle aspiration is used to rule out malignancy in suspicious nodules, but it produces indeterminate results in up to 30% of cases (Bose, 2019). Indeterminate aspirates comprise three Bethesda System for Reporting Thyroid Cytopathology categories, each with differing associated risks of malignancy and correspondingly treatment (de Koster, 2018; Krane, 2014):

- Category III — Atypia of undetermined significance/follicular lesion of undetermined significance is considered a low-risk indeterminate category with an estimated 5% to 15% of malignancy. Repeat fine-needle aspiration is often recommended.
- Category IV — Follicular neoplasm/suspicious for a follicular neoplasm has a 15% to 30% risk of malignancy. Lobectomy is recommended in most cases.
- Category V — Suspicious for malignancy has a 60% to 75% risk of malignancy (typically papillary carcinoma with less distinct nuclear features). Thyroid surgery is recommended.

The prevalence of malignancy in indeterminate nodules increases when signs and symptoms are present. In ambiguous cases, thyroid resection is performed for a definitive diagnosis. Molecular diagnostic tests are proposed as a triage tool to either “rule in” or “rule out” malignancy with greater accuracy and defer unnecessary surgeries for benign thyroid nodules. Laboratory-developed testing for genetic mutations associated with thyroid cancer employ reverse transcriptase polymerase chain reaction assay microarrays and next-generation ribonucleic acid sequencing.

Afirma Thyroid FNA Analysis
Afirma Thyroid FNA Analysis comprises three molecular diagnostic tests: Afirma Gene Expression Classifier; Afirma Genomic Sequencing Classifier; and Afirma Xpression Atlas (Afirma, 2019a, 2019b). These tests analyze the genetic content of indeterminate thyroid samples based on cytopathology to identify benign nodules using a proprietary algorithm to report the results. The goal of these tests is to increase the probability that a negative test result is truly benign (i.e., to maximize the test’s negative predictive value), thereby identifying individuals who should have periodic surveillance and avoiding overdiagnosis and unnecessary thyroidectomies.

The Afirma Gene Expression Classifier measures the activity or “expression” levels of 167 genes in a thyroid nodule sample (Afirma, 2019a). The Afirma Genomic Sequencing Classifier, introduced in 2017, and the Afirma Xpression Atlas, introduced in 2018, employ next-generation ribonucleic acid genomic sequencing to measure hundreds more genetic mutations linked to thyroid cancer (Afirma, 2019b). These tests are designed to improve the specificity of the Gene Expression Classifier and identify more individuals who may preserve their thyroid.

Test results are reported as either “benign” or “suspicious for malignancy.” A “benign” result indicates very little chance (< 5%) of cancer. A “suspicious for malignancy” result implies the tests could not rule out with complete certainty that cancer is not present (i.e., the risk of cancer has increased).
Findings

The evidence for this policy comprises a multicenter validation study (Alexander, 2012), a retrospective, multicenter follow-up study (Alexander, 2014), several individual studies (Duick, 2012; Harrell, 2014; Lastra, 2014; McGiver, 2014), one cost effectiveness analysis (Li, 2014), and one evidence-based guideline (Francis, 2015). All studies addressed the Afirma Gene Expression Classifier. The vast majority of cytopathologies subjected to Afirma testing across studies were classified as “atypia (or follicular lesion) of undetermined clinical significance” or “follicular neoplasm or lesion suspicious for follicular neoplasm” in adult populations. The evidence for molecular analysis of indeterminate thyroid nodules in children is absent; therefore, surgical decisions are dictated by fine-needle aspiration results (Francis, 2015).

A prospective, multicenter validation study involving 49 clinical sites, 3,789 patients, and 4,812 fine-needle aspirates from thyroid nodules 1 cm or larger found that 85 of the 265 indeterminate nodules were malignant (Alexander, 2012). Afirma correctly identified 78 of the 85 nodules as suspicious, yielding an overall sensitivity of 92% (95% confidence interval 84 to 97) and negative predictive values of 95%, 94%, and 85% for “atypia (or follicular lesion) of undetermined clinical significance,” “follicular neoplasm or lesion suspicious for follicular neoplasm,” and “suspicious cytologic findings,” respectively. Average specificity was 52%.

The retrospective follow-up study analyzed the clinical utility of Afirma testing in 339 patients at five academic medical centers between 2010 and 2013 (Alexander, 2014). Afirma classified 51% of nodules as benign, 44% as suspicious, and 5% as nondiagnostic. The malignancy rate for indeterminate/Afirma suspicious nodules was 44%, and 87% of malignant lesions were papillary thyroid carcinoma. A low malignancy rate (false-negative rate = 1.3%) for the 174 participants with Afirma benign results was confirmed in 75 cases based on surgery (n = 4) or clinical follow up (n = 71; median 8.5 months, range 1 to 24 months). The malignancy rate for the remaining 99 participants was not reported. The Afirma results modified care recommendations in 171 of 339 (50%) participants with the greatest impact occurring in those classified as Afirma benign, as only 2% were recommended for surgery (P < .01).

Institutional experiences with Afirma and a cost-effectiveness analysis demonstrate that the test performance achieved in a validation trial cannot be generalized to every day practice, as several factors can influence test performance and cost-effectiveness (Alexander, 2014; Brauner, 2015; Duick, 2012; Harrell, 2014; Lastra, 2014; Li, 2011; McGiver, 2014). The underlying variation in malignancy rate across institutions and the interobserver variability in cytologic, genetic, and histologic diagnosis of thyroid lesions are major contributors to the variability. Studies inconsistently reported the proportion of Hürthle cell-rich lesions from other types of indeterminate samples and failed to stratify test performance based on as Bethesda III, IV, and V classification of nodules, all of which would be essential to establishing local test performance.

In a significant portion of participants, confirmation of Afirma test findings by histopathology or clinical follow-up were missing, and the true positive and true negative rates could not be determined. These shortcomings make definitive evaluation of analytic validity, test accuracy, and cost effectiveness of Afirma difficult to categorize with certainty and its clinical role less definitive.

The vast majority of thyroid cancer cases are locally or regionally confined with estimated 5-year survival rates exceeding 98% among persons with early stage, low-aggressive cancer types (American Cancer Society, 2019). Increased detection of small, indolent thyroid nodules through use of thyroid ultrasound has not conferred a clear mortality benefit in a population with an already low disease-associated mortality (Krane, 2014). The evidence suggests the clinical benefit of Afirma is in avoiding diagnostic thyroidectomies and associated morbidity in adults with cytologically-indeterminate nodules at least 1 cm in diameter classified as “atypia (or follicular lesion) of
“undetermined significance” (Bethesda III) or “follicular neoplasm or suspicious for follicular neoplasm, including Hürthle cell” (Bethesda IV) (Alexander, 2014; Duick, 2012; Lastra, 2014).

Policy updates
In 2016, we added five retrospective studies (Angeli, 2015; Celik, 2015; Sacks, 2016; Sipos, 2016; Wong, 2016) and a meta-analysis (Santhanam, 2016) to the policy. Their results are consistent with previous findings of the discriminative power of the Afirma Gene Expression Classifier as a tool to rule out malignancy and its ability to safely reduce unnecessary thyroidectomies. However, Afirma may have unintended consequences. Sacks (2016) observed a decrease in the incidence of benign diagnoses and an increase in the number of indeterminate cytopathological diagnoses, with a shift of fine-needle aspiration interpretation toward Bethesda III/IV in which molecular testing is used, while not affecting the institutional thyroidectomy rate or malignancy yield.

In 2018, we added a systematic review (Duh, 2017) to the policy describing the methodologic quality of studies examining the diagnostic accuracy of the Afirma Gene Expression Classifier. Duh (2017) confirmed prior policy findings that studies: 1) failed to consistently apply a reference standard to benign indeterminate thyroid nodules that were not excised for histopathologic confirmation; and 2) restricted participants to those with indeterminate thyroid nodules who had already been selected for referral for thyroidectomy or lobectomy. These biases can result in unreliable estimates of specificity and negative predictive value. The authors recommended that studies define and assign a “true negative” label to Afirma benign nodules that do not develop malignant signs or symptoms during a pre-specified period of follow-up and including these nodules in calculations of diagnostic accuracy.

In 2019, we added three evidence-based guidelines (Gharib, 2016; Haugen, 2017; National Comprehensive Cancer Network, 2018), one cost-effectiveness analysis (Balentine, 2018), and six retrospective studies (Brauner, 2015; Hang, 2017; Harrell, 2018a, 2018b; Jug, 2018; Livhits, 2018) to the policy. Three of the studies examined the Afirma Gene Expression Classifier (Brauner, 2015; Hang, 2017; Harrell, 2018a). Three studies compared the Afirma Gene Expression Classifier to two next generation sequencing panels—the Afirma Genomic Sequencing Classifier (Harrell, 2018b) and the ThyroSeq V2 panel (CBLPath and University of Pittsburgh Medical Center) (Jug, 2018; Livhits, 2018).

The results confirm prior policy findings for the Afirma Gene Expression Classifier and present new findings comparing it to next-generation sequencing. Preliminary results suggest higher specificity with next-generation sequencing, particularly the ability to discern benign from malignant Hürthle cell nodules (Harrell, 2018b; Livhits, 2018; National Comprehensive Cancer Network, 2018). A retrospective study confirmed Afirma Gene Expression Classifier’s low test performance for identifying Hürthle cell cancer without an appreciable reduction in unnecessary thyroidectomies (Brauner, 2015).

Guideline recommendations for the Afirma Gene Expression Classifier are equivocal, reflecting uncertainty in its clinical role. The American College of Clinical Endocrinologists recommends neither for nor against the use of this test for cytologically indeterminate nodules, stressing a complementary role to cytopathology when the results are expected to influence clinical management (Gharib, 2016). The National Comprehensive Cancer Network (2018) suggests molecular diagnostic testing for evaluating Bethesda III or IV cytopathology results, taking into consideration the clinical, radiographic, and cytologic features of each patient. The American Thyroid Association strongly recommended (based on weak evidence) counseling patients regarding the potential benefits and limitations of molecular testing and uncertainties in the therapeutic and long-term clinical implications of results (Haugen, 2017).
Thyroid tumors diagnosed as “encapsulated follicular variant of papillary thyroid carcinoma without capsular or vascular invasion” represent a distinct class of thyroid tumors with very low risk of adverse outcome. The American Thyroid Association endorsed renaming these tumors to “noninvasive follicular thyroid neoplasms with papillary-like nuclear features” with a set of reproducible diagnostic criteria to reflect a highly indolent form (Haugen, 2017).

As a result of this change, a significant percentage of indeterminate nodules classified as Afirma “suspicious” would be reclassified as nonmalignant, resulting in more false positives and a lower positive predictive value on both microarray and next-generation sequencing (Hang, 2017; Jug, 2018). Clinicians who rely on an Afirma “suspicious” result to justify thyroidectomy may, in fact, be overtreating these nodules. Early validation studies of the Gene Expression Classifier did not account for this category. Prospective studies are needed to validate the observed patient outcomes, test performance in predicting thyroid cancer outcomes, and implications on patients' psychosocial health and economics.

A cost-effectiveness analysis compared thyroid lobectomy to an Afirma Gene Expression Classifier strategy using a base case of a 40-year-old woman with a 1-cm indeterminate nodule (Bethesda III and IV) (Balentine, 2018). When accounting for the costs of long-term surveillance for indeterminate nodules classified as benign, diagnostic lobectomy was both more effective and less costly than the Afirma strategy for ruling out malignancy of indeterminate thyroid nodules. The results were highly sensitive to estimates of utilities after lobectomy and patient values and preferences affecting treatment decisions.

The policy was modified to add: 1) medical necessity criteria in accordance with National Comprehensive Cancer Network (2018) criteria for molecular testing using the Afirma Gene Expression Classifier; 2) a coverage statement of investigational for the Afirma Genomic Sequencing Classifier and Afirma Xpression Atlas due to insufficient evidence of efficacy; and 3) a separate policy section for Medicare-specific criteria. The policy ID was changed from CP# 02.01.11 to CCP.1099.

In 2020, we updated one guideline (National Comprehensive Cancer Network, 2019) and added two analytic validation studies of the Afirma Xpression Atlas test (Angell, 2019b) and the Afirma Gene Sequencing Classifier test (Hao, 2019), one cost effectiveness analysis (Nicholson, 2019), and five independent clinical studies to the policy. The National Comprehensive Cancer Network (2019) noted that the Afirma Gene Sequencing Classifier improved the sensitivity and specificity of detecting Hürthle cell neoplasms over that of the Afirma Gene Expression Classifier in indeterminate thyroid nodules. Results from single-site, retrospective studies confirmed these findings in populations representative of current clinical practice in the United States (Angell, 2019a; Endo, 2019; Harrell, 2019; San Martin, 2019; Wei, 2019). These studies reported statistically significant increases in benign call rates using the Afirma Gene Sequencing Classifier compared to the Afirma Gene Expression Classifier, indicating fewer false positive results, and corresponding reductions in unnecessary diagnostic surgeries.

Nicholson (2019) compared the cost effectiveness of diagnostic lobectomy, the Afirma Gene Sequencing Classifier, and ThyroSeq version 3. In hypothetical modeling using a base case of a 40-year-old euthyroid woman with a solitary 2 cm Bethesda III or IV thyroid nodule, either molecular test was considerably more cost-effective than diagnostic lobectomy. The cost per correct diagnosis was $14,277 for ThyroSeq version 3, $17,873 for Afirma Gene Sequencing Classifier, and $38,408 for diagnostic lobectomy. The results were robust to variations in cost, cancer prevalence, and length of surveillance. The policy was changed to include the Afirma Gene Sequencing Classifier test as medically necessary to improve diagnosis and reduce unnecessary surgeries in patients with indeterminate thyroid nodules.
On December 19, 2019, we searched PubMed and the databases of the Cochrane Library, the U.K. National Health Services Centre for Reviews and Dissemination, the Agency for Healthcare Research and Quality, and the Centers for Medicare & Medicaid Services. Search terms were “Thyroid Nodule” (MeSH), “Gene Expression Profiling” (MeSH), “thyroid nodules,” and “Afirma.” We included the best available evidence according to established evidence hierarchies (typically systematic reviews, meta-analyses, and full economic analyses, where available) and professional guidelines based on such evidence and clinical expertise.


CCP.1099


**Policy updates**

3/2014: initial review date and clinical policy effective date: 9/2014


3/2020: Policy references updated and coverage expanded to include Afirma Gene Sequencing Classifier test.