

# Long-term Follow-up of Cytologically Indeterminate Thyroid Nodules Found Benign on Molecular Testing: A Validation Study

Michelle K. White, MD<sup>1</sup> , William B. Thedinger<sup>2</sup>,  
and Jagdish K. Dhingra, MD<sup>1,3</sup>

OTO Open  
 2022, Vol. 6(1) 1–5  
 © The Authors 2022  
 Article reuse guidelines:  
[sagepub.com/journals-permissions](http://sagepub.com/journals-permissions)  
 DOI: 10.1177/2473974X221083542  
<http://oto-open.org>  


## Abstract

**Objective.** Molecular testing has revolutionized management of indeterminate thyroid nodules (Bethesda categories III and IV). Few studies have attempted to validate the negative predictive value of molecular tests. Using long-term observation as a surrogate for surgical resection, we sought to examine the false-negative rate of “benign” indeterminate thyroid nodules on molecular testing.

**Study Design.** Case series with retrospective data collection and chart review.

**Setting.** Large community-based practice with multiple satellite offices.

**Methods.** All patients with thyroid nodules that underwent ultrasound-guided fine-needle aspiration biopsy between 2013 and 2019 were evaluated through retrospective analysis. Cytologically indeterminate nodules reflexively underwent molecular testing to guide clinical management. Observation was recommended for lesions with benign molecular testing, and these nodules were followed clinically and by ultrasound.

**Results.** A total of 2011 nodules underwent fine-needle aspiration, of which 280 (14%) were indeterminate thyroid nodules. Of those 280 nodules, 100 (36%) were benign on molecular testing. Three samples were excluded from analysis due to patient deaths from unrelated causes. Surgical resection was recommended in 16 of the 97 nodules (17%), with the majority due to size and compressive symptoms. Histopathology was available in 14 nodules that underwent surgery, with 1 demonstrating minimally invasive follicular carcinoma.

**Conclusion.** While molecular testing is safe to use in guiding management of indeterminate thyroid nodules, consideration of individualized clinical factors and close long-term follow-up remains paramount.

## Keywords

thyroid, indeterminant thyroid nodules, molecular testing, Bethesda

Received January 18, 2022; accepted February 8, 2022.

Thyroid nodules are a common entity, with an estimated prevalence of 19% to 68% in randomly selected individuals.<sup>1</sup> Appropriate diagnostic workup is necessary when a thyroid nodule is discovered, especially given a 5%–15% risk of malignancy.<sup>1,2</sup> Management is generally directed by cytologic findings from ultrasound-guided fine-needle aspiration (US-FNA), which are classified into I of VI categories by the Bethesda system predicting varying risks of malignancy.<sup>3,4</sup> Cytologic indeterminate results (Bethesda categories III and IV) occur in 15% to 30% of US-FNAs, carrying up to 30% risk of malignancy on eventual surgical pathology.<sup>2,5–7</sup> A clinical decision must be made on whether to manage these cytologic indeterminate thyroid nodules (ITNs) conservatively with long-term observation or through surgical intervention and histopathologic analysis. Thyroidectomy carries a cost of ~\$12,000 at 2010 Medicare rates, and there are additional costs of lost wages and potential surgical complications, including infection, hematoma, vocal cord dysfunction, hypothyroidism, hypocalcemia, airway obstruction, and death.<sup>8</sup> Given the prevalence of thyroid nodules and inherent risks of diagnostic thyroidectomy, testing to elucidate the benign or malignant nature of cytologic ITNs is desired.

Molecular testing has revolutionized management of cytologic ITNs by providing evidence to rule in or rule out

<sup>1</sup>Department of Otolaryngology–Head and Neck Surgery, Tufts Medical Center, Boston, Massachusetts, USA

<sup>2</sup>School of Medicine, Tufts University, Boston, Massachusetts, USA

<sup>3</sup>ENT Specialists, Inc, Brockton, Massachusetts, USA

This article was presented at the AAO-HNSF 2020 Virtual Annual Meeting & OTO Experience; September 13–October 25, 2020.

## Corresponding Author:

Jagdish K. Dhingra, MD, ENT Specialists, Inc, 35 Pearl St, Second Floor (Suite 200), Brockton, MA 02301, USA.

Email: [JDhingra@ENTspecialists.com](mailto:JDhingra@ENTspecialists.com)



malignancy and thereby reduce unnecessary diagnostic surgery. An initial molecular test measured expression of 167 gene transcripts implicated in malignant thyroid lesions. The initial validation study reported the test to be successful in ruling out malignancy, with a sensitivity of 92% and a negative predictive value (NPV) of 93%.<sup>3</sup> However, its specificity of 52% and positive predictive value (PPV) of 47% made it a poor predictor of malignancy.<sup>3</sup> A subsequent molecular test was developed to improve the specificity and PPV without sacrificing the high NPV seen with the initial molecular test. In a large single-center study, this new test improved the specificity and PPV to 94.3% and 60%, respectively, in turn reducing the rate of unnecessary surgical intervention for cytologic ITNs.<sup>9</sup>

Several studies have described clinical experiences utilizing molecular testing on cytologic ITNs. However, few studies have sought to validate the false-negative rate of molecular testing results, as this would require resection of all cytologic ITNs irrespective of molecular test results. We hypothesize that close long-term follow-up is the best surrogate for surgical resection, as a stable nodule over a prolonged period may be classified as a true negative. Conversely, long-term follow-up may identify patients for whom surgery is indicated or recommended despite a benign molecular testing result. We analyze our data from a large patient cohort in the same community over a 6-year period. Cytologic ITNs were followed clinically, and in certain cases, surgery was recommended despite benign molecular testing, providing an opportunity to evaluate potential false-negative results.

## Materials and Methods

This is a retrospective analysis of all patients with thyroid nodules that underwent US-FNA between 2013 and 2019 at ENT Specialists, Inc, a busy community thyroid practice in southern Massachusetts. This study was approved by the Institutional Review Board at Brockton Hospital, Signature Healthcare. US-FNA was performed by a single otolaryngologist skilled in performing the procedure. Cytologic ITNs (Bethesda categories III and IV) reflexively underwent Afirma Gene Expression Classifier or Gene Sequencing Classifier testing (Veracyte, Inc) to guide clinical management. Bethesda category III corresponds to “atypia of undetermined significance or follicular lesion of undetermined significance,” and category IV corresponds to “follicular neoplasm or suspicious for a follicular neoplasm.”<sup>4</sup> The determination to proceed with surgery or routine surveillance was made by the patient’s physician based on results of cytologic testing and the patient’s symptoms and preferences. When surgery was performed, specimens were reviewed by hospital pathologists where the surgery was performed to make a final histologic diagnosis.

Retrospective chart review included analysis of all patients with cytologic ITNs with “benign” molecular testing results. Over the period of the study, patients were followed clinically and with surveillance ultrasound. Surgery was recommended in certain patients despite benign molecular testing for a variety of reasons, such as the size of nodule, compressive

<b>Bethesda 3 (232 nodules)</b>	
Benign	36% (83 nodules)
Suspicious	53% (124 nodules)
No result	5% (11 nodules)
Atypical lymphocytes	1% (3 nodules)
No classification	5% (11 nodules)
<b>Bethesda 4 (48 nodules)</b>	
Benign	35.5% (17 nodules)
Suspicious	60.5% (29 nodules)
No classification	4% (2 nodules)

**Figure 1.** Molecular testing results for cytologic indeterminate thyroid nodules (Bethesda III and IV).

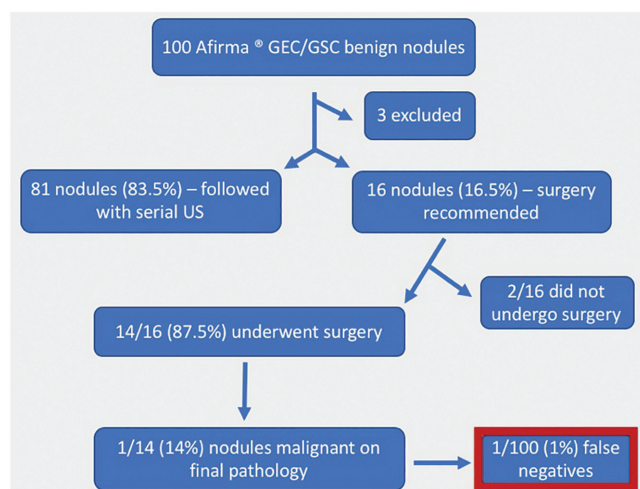
symptoms, growth on serial ultrasound, and the presence of additional suspicious nodules. Of these patients who underwent surgery, surgical pathology was reviewed to determine whether the nodule represented a false-negative molecular test result.

## Results

Between 2013 and 2019, a total of 2011 nodules underwent US-FNA in a dedicated ultrasound clinic, and 280 nodules (14%) were reported back as cytologic ITNs (Bethesda categories III and IV). Reflexive molecular testing provided further diagnostic classification: benign, suspicious, no result, atypical lymphocytes, and no classification (**Figure 1**). Within our practice, the benign call rate for molecular testing is 30.3% and 45.7% for the initial and newer molecular tests, respectively.

Of the 280 ITNs, 100 (35.7%) were benign on molecular testing. All samples between October 2013 and July 2017 were tested with the initial molecular test (51 nodules, 51%), and all samples between August 2017 and December 2019 were tested with a newer molecular test (49 nodules, 49%). Three samples were excluded from analysis due to patient deaths from unrelated causes. The remaining 97 “benign” nodules on molecular testing were followed with serial ultrasound at 12-month intervals for all patients. Of those, 16 (16.5%) were recommended for surgical excision. The majority of these surgical recommendations were for large or increasing nodule size on follow-up ultrasound. One patient with progressive disease received a repeat fine-needle aspiration, which revealed suspicious cytopathology. This patient went for surgical excision, but final pathology was benign. Patients who were not recommended for surgical intervention all had stable disease on serial ultrasound.

Of the 16 patients recommended for surgery, 14 underwent surgery with ENT Specialists, Inc, while 2 were lost to



**Figure 2.** Flowchart representing management of thyroid nodules with benign molecular testing. GEC, Gene Expression Classifier; GSC, Gene Sequencing Classifier; US, ultrasound.

follow-up. On follow-up after the study period, 1 patient had serial ultrasound performed at an outside hospital, which revealed a stable nodule size at 4.3 cm. The patient chose conservative management. We were unable to reach the other patient lost to follow-up. Thus, surgery was performed in 14.7% (14 of 95 patients) of the benign molecular testing group. The follow-up for those who received serial ultrasound and underwent surgery ranged from 0.25 to 5 years, with a mean of 1.3 years (interquartile range, 0.5-1.25). One of these patients with a benign nodule from the newer molecular test, subjected to surgery for increasing size over 3 years, had a final diagnosis of follicular variant of papillary thyroid carcinoma based on final surgical pathology. This represented a 1% false-negative rate (**Figure 2**). The NPVs of the initial and newer molecular tests were 100% and 98.0%, respectively.

## Discussion

In the initial prospective multicenter study for the original molecular test,<sup>3</sup> 100% of cytologic ITNs were surgically removed without prior knowledge of molecular test results. From these data, histopathologic diagnosis was used to validate whether molecular testing results accurately predicted the presence or absence of malignancy.<sup>3</sup> Alexander et al<sup>3</sup> found a risk of malignancy for an ITN with a benign molecular test result to be 4% to 5% ( $1 - \text{NPV}$ ), which is comparable to the ~3% risk of malignancy in benign Bethesda II cytologic diagnosis.<sup>3,5,10</sup> This suggested that a cytologic ITN with a benign molecular test result can be managed similarly to a cytologically benign nodule with routine follow-up and surveillance ultrasound.<sup>8,11-14</sup> As a result, professional organizations revised guidelines recommending observation and ultrasound follow-up in favor of thyroid surgery for benign molecular testing results.<sup>1,4,15</sup>

Subsequent studies that sought to verify these findings are limited, as true validation of molecular testing would require histopathology data on surgical resection of all benign

nodules to determine the rate of false negativity. Designing such a study would be impractical. Valderrabano et al<sup>16</sup> performed a systematic review of 19 studies encompassing 2568 nodules. They found significant differences in the sensitivity and specificity of the molecular test using the correlation between benign call rate and PPV from the pooled independent studies as compared with the PPV of the initial study by Alexander et al.<sup>3</sup> Other studies attempted to validate the NPV of molecular testing through eventual resection of benign ITNs and histopathologic diagnosis.<sup>17-21</sup> Many of these studies yielded less robust NPVs than the initial validation study: 83.3% in Noureldine et al<sup>20</sup> vs 94% in Alexander et al. This variation called for further investigation into molecular testing and its role in influencing the clinical management of ITNs.

Ultimately, most nodules with benign molecular testing are not removed, and true histopathologic diagnoses are never known. Postmarketing evaluation of false-negative rates and NPV of molecular testing is challenging due to low resection rates for benign nodules. Long-term follow-up may be the best surrogate for histopathology, under the assumption that nodules that have reassuring characteristics on surveillance ultrasound over time may be considered “true negative” nodules. We hypothesized that resection and histopathologic analysis of nodules with previous benign molecular testing results that developed suspicious features on subsequent surveillance ultrasound could be used to validate the performance of molecular testing.

After retrospective analysis of 95 benign nodules, 1 patient had a malignancy that was misclassified by molecular testing. This represents a false-negative rate of ~1% for molecular testing in our patient cohort. Of the 95 nodules that were followed with serial ultrasound, 14 underwent surgery. In a similarly designed study from a busy community endocrine surgical referral center,<sup>22</sup> 193 nodules with benign molecular testing results were examined, which is a larger group than in the validation study by Alexander et al.<sup>3</sup> Surgery was performed in 42 patients, and 14 were diagnosed with malignancies or noninvasive follicular thyroid tumor with papillary-like nuclear features. The authors found a false-negative rate of 7.3%, which they appropriately considered a “false-negative percentage for an incomplete surgical group,” given that histopathologic confirmation was not available for all patients with benign molecular test results.<sup>22</sup> For this reason, these findings represent the minimum possible false-negative percentage for the population. The same conclusion can be drawn from this study: our false-negative rate of 1% is a modified percentage that does not encompass all potentially missed malignancies by molecular testing, given that not all benign nodules are surgically removed for histopathologic diagnosis.

Despite these findings, the chance of missing clinically significant malignancy in a cytologic ITN with a benign molecular test result is still reasonably low. The modified false-negative rate of 7.3% in Harrell et al<sup>22</sup> seems to fall within this realm of risk. With close follow-up and routine surveillance, it seems unlikely that potential benign molecular

testing false negatives would lead to significant patient harm, given that changes in patient symptoms or nodule characteristics over time are likely to prompt timely intervention. Clinical decision making should be individualized according to a patient's presenting symptoms, medical history, and findings on ultrasound and cytology.

One main limitation of our study is the small sample size. Only 14 (14.7%) nodules with benign molecular testing results underwent surgical resection and histopathologic analysis, as compared with 42 (21.8%) nodules in a 2018 study by Harrell et al.<sup>22</sup> This difference in sample size could explain the difference in false-negative rates between these studies: 1% vs 7.3%. Furthermore, limited sample sizes and the resulting low number of false-negative benign nodules can greatly alter the NPV and evaluation of molecular testing. Our NPV for the initial molecular test (100%) was higher than in most validation studies, including the initial validation study by Alexander et al (93%).<sup>3</sup> This is likely due to our limited sample size.

An additional complexity of interpreting molecular testing relates to the varying prevalence of thyroid malignancy among institutions and geographic regions. When compared with meta-analyses validating molecular testing, studies such as ours and Harrell et al<sup>22</sup> have a specific advantage of long-term follow-up data within a specific, localized population cohort. Meta-analyses have an inherent disadvantage, as data are pooled from different population groups that may have different incidences of thyroid cancer. Marti et al<sup>23</sup> found significant variation in benign call rate, PPV, and NPV of molecular testing between 2 New York City institutions: 1 cancer referral center and 1 comprehensive health care system. Similarly, we found a lower benign call rate for the initial and newer molecular tests (30.3% and 45.7%, respectively) as compared with other validation studies. This difference could be due to sample size or varying rates in thyroid cancer across regions and practice settings.

Marti et al<sup>23</sup> hypothesized that the performance of the original molecular test best approximated the initial validation study of Alexander et al<sup>3</sup> when prevalence of thyroid malignancy was between 12% and 25%. In settings with a prevalence >25%, a benign molecular test result cannot rule out malignancy with sufficient NPV.<sup>23</sup> This phenomenon can explain the variation in rates of false negativity when comparing studies such as ours, Harrell et al,<sup>22</sup> and the initial validation study of Alexander et al. The higher false negative found by Harrell et al (7.3%) vs Alexander et al (1.6%) could be explained by a higher prevalence of thyroid cancer in the former study (33%) vs the latter (24%).

Molecular testing can offer reassurance for patients with cytologic ITNs and their providers and has been utilized in avoiding thousands of unnecessary thyroid operations since molecular testing became available.<sup>11</sup> Long-term stability of cytologically benign ITNs over a 6-year period in this study provides further validation of molecular testing. However, as molecular testing is still in its infancy, diligent long-term follow-up and careful interpretation are advised while deciding on management of cytologic ITNs.

## Author Contributions

**Michelle K. White**, contributed to the formulation of the research idea, data collection, statistical analysis, and manuscript writing and editing and approved the final submission; **William B. Thedinger**, contributed to manuscript writing and editing and approved the final submission; **Jagdish K. Dhingra**, contributed to the formulation of the research idea, data collection, statistical analysis, and manuscript writing and editing and approved the final submission.


## Disclosures

**Competing interests:** None.

**Sponsorships:** None.

**Funding source:** None.

## ORCID iD

Michelle K. White  <https://orcid.org/0000-0002-4921-6174>

## References

1. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*. 2016;26(1):1-133. doi:10.1089/thy.2015.0020
2. Nishino M. Molecular cytopathology for thyroid nodules: a review of methodology and test performance—molecular tests for thyroid FNA. *Cancer Cytopathol*. 2016;124(1):14-27. doi:10.1002/cncy.21612
3. Alexander EK, Kennedy GC, Baloch ZW, et al. Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. *N Engl J Med*. 2012;367(8):705-715. doi:10.1056/NEJMoa1203208
4. Cibas ES, Ali SZ. The 2017 Bethesda System for Reporting Thyroid Cytopathology. *Thyroid*. 2017;27(11):1341-1346. doi:10.1089/thy.2017.0500
5. Cibas ES, Ali SZ. The Bethesda System for Reporting Thyroid Cytopathology. *Thyroid*. 2009;19(11):1159-1166. doi:10.1089/thy.2009.0274
6. Baloch Z, Cibas E, Clark D, et al. The National Cancer Institute Thyroid Fine Needle Aspiration State of the Science Conference: a summation. *Cytojournal*. 2008;5(1):6.
7. Cooper DS, Doherty GM, Haugen BR, et al. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2009;19(11):1167-1215. doi:10.1089/thy.2009.0110
8. Duick DS, Klopper JP, Diggans JC, et al. The impact of benign gene expression classifier test results on the endocrinologist-patient decision to operate on patients with thyroid nodules with indeterminate fine-needle aspiration cytopathology. *Thyroid*. 2012;22(10):996-1001. doi:10.1089/thy.2012.0180
9. Endo M, Nabhan F, Porter K, et al. A firm gene sequencing classifier compared with gene expression classifier in indeterminate thyroid nodules. *Thyroid*. 2019;29(8):1115-1124. doi:10.1089/thy.2018.0733

10. Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda System for Reporting Thyroid Cytopathology: a meta-analysis. *Acta Cytol.* 2012;56(4):333-339. doi:10.1159/000339959
11. Kloos RT. Molecular profiling of thyroid nodules: current role for the Afirma gene expression classifier on clinical decision making. *Mol Imaging Radionucl Ther.* 2017;26(suppl 1):36-49. doi:10.4274/2017.26.suppl.05
12. Angell TE, Frates MC, Medici M, et al. Afirma benign thyroid nodules show similar growth to cytologically benign nodules during follow-up. *J Clin Endocrinol Metab.* 2015;100(11):E1477-E1483. doi:10.1210/jc.2015-2658
13. Alexander EK, Schorr M, Klopper J, et al. Multicenter clinical experience with the Afirma gene expression classifier. *J Clin Endocrinol Metab.* 2014;99(1):119-125. doi:10.1210/jc.2013-2482
14. Sipos JA, Blevins TC, Shea HC, et al. Long-term nonoperative rate of thyroid nodules with benign results on the Afirma gene expression classifier. *Endocr Pract.* 2016;22(6):666-672. doi:10.4158/EP151006.OR
15. National Comprehensive Cancer Network. Thyroid carcinoma. Version 2.2017. Published May 17, 2017. Accessed March 31, 2021. [https://oncolife.com.ua/doc/nccn/Thyroid\\_Carcinoma.pdf](https://oncolife.com.ua/doc/nccn/Thyroid_Carcinoma.pdf)
16. Valderrabano P, Hallanger-Johnson JE, Thapa R, Wang X, McIver B. Comparison of postmarketing findings vs the initial clinical validation findings of a thyroid nodule gene expression classifier: a systematic review and meta-analysis. *JAMA Otolaryngol Neck Surg.* 2019;145(9):783. doi:10.1001/jamaoto.2019.1449
17. Al-Qurayshi Z, Deniwar A, Thethi T, et al. Association of malignancy prevalence with test properties and performance of the gene expression classifier in indeterminate thyroid nodules. *JAMA Otolaryngol Neck Surg.* 2017;143(4):403. doi:10.1001/jamaoto.2016.3526
18. McIver B, Castro MR, Morris JC, et al. An independent study of a gene expression classifier (Afirma) in the evaluation of cytologically indeterminate thyroid nodules. *J Clin Endocrinol Metab.* 2014;99(11):4069-4077. doi:10.1210/jc.2013-3584
19. Harrell RM, Bimston DN. Surgical utility of Afirma: effects of high cancer prevalence and oncocytic cell types in patients with indeterminate thyroid cytology. *Endocr Pract.* 2014;20(4):364-369. doi:10.4158/EP13330.OR
20. Noureldine SI, Olson MT, Agrawal N, Prescott JD, Zeiger MA, Tufano RP. Effect of gene expression classifier molecular testing on the surgical decision-making process for patients with thyroid nodules. *JAMA Otolaryngol Neck Surg.* 2015;141(12):1082. doi:10.1001/jamaoto.2015.2708
21. Vora A, Holt S, Haque W, Lingvay I. Long-term outcomes of thyroid nodule AFIRMA GEC testing and literature review: an institutional experience. *Otolaryngol Head Neck Surg.* 2020;162(5):634-640. doi:10.1177/0194599820911718
22. Harrell RM, Eyerly-Webb SA, Pinnar NE, Golding AC, Edwards CM, Bimston DN. Community endocrine surgical experience with false-negative Afirma Gec results: 2011-2017. *Endocr Pract.* 2018;24(7):622-627. doi:10.4158/EP-2017-0263
23. Marti JL, Avadhani V, Donatelli LA, et al. Wide inter-institutional variation in performance of a molecular classifier for indeterminate thyroid nodules. *Ann Surg Oncol.* 2015;22(12):3996-4001. doi:10.1245/s10434-015-4486-3