

Clinical Research Article

Comparison of Afirma GEC and GSC to Nodules Without Molecular Testing in Cytologically Indeterminate Thyroid Nodules

Preethi Polavarapu,^{1,2} Abbey Fingeret,³ Ana Yuil-Valdes,⁴ Daniel Olson,⁵ Anery Patel,¹ Vijay Shivaswamy,^{1,2} Troy D. Matthias,⁴ and Whitney Goldner¹

¹Department of Internal Medicine, Division of Diabetes, Endocrinology, Metabolism, University of Nebraska Medical Center, Omaha, NE, USA; ²Department of Internal Medicine, Division of Diabetes, Endocrinology, Metabolism, VA Nebraska-Western Iowa Health System, Omaha, NE, USA; ³Department of Surgery, Division of Surgical Oncology, Omaha, NE, USA; ⁴Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE, USA; and ⁵Department of Pathology and Microbiology, VA Nebraska–Western Iowa Health System, Omaha, NE, USA

ORCID numbers: 0000-0002-6625-7779 (P. Polavarapu); 0000-0002-0526-6364 (A. Fingeret); 0000-0002-3301-4965 (V. Shivaswamy); 0000-0002-9724-7987 (W. Goldner).

Received: 30 April 2021; Editorial Decision: 3 September 2021; First Published Online: 7 October 2021; Corrected and Typeset: 25 October 2021.

Abstract

Background: Analysis of cytologically indeterminate thyroid nodules with Afirma Gene Expression Classifier (GEC) and Genomic Sequencing Classifier (GSC) can reduce surgical rate and increase malignancy rate of surgically resected indeterminate nodules.

Methods: Retrospective cohort analysis of all adults with cytologically indeterminate thyroid nodules from January 2013 through December 2019. We compared surgical and malignancy rates of those without molecular testing to those with GEC or GSC, analyzed test performance between GEC and GSC, and identified variables associated with molecular testing.

Results: 468 indeterminate thyroid nodules were included. No molecular testing was performed in 273, 71 had GEC, and 124 had GSC testing. Surgical rate was 68% in the group without molecular testing, 59% in GEC, and 40% in GSC. Malignancy rate was 20% with no molecular testing, 22% in GEC, and 39% in GSC ($P = 0.022$). GEC benign call rate (BCR) was 46%; sensitivity, 100%; specificity, 61%; and positive predictive value (PPV), 28%. GSC BCR was 60%; sensitivity, 94%; specificity, 76%; and PPV, 41%. Those with no molecular testing had larger nodule size, preoperative growth of nodules, and constrictive symptoms and those who underwent surgery in the no molecular testing group had higher body mass index, constrictive symptoms, higher Thyroid Imaging Reporting and Data System and Bethesda classifications. Type of provider was also associated with the decision to undergo surgery.

Conclusion: Implementation of GEC showed no effect on surgical or malignancy rate, but GSC resulted in significantly lower surgical and higher malignancy rates. This study provides insight into the factors that affect the real-world use of these molecular markers preoperatively in indeterminate thyroid nodules.

Key Words: thyroid nodule, Bethesda III, Bethesda IV, indeterminate thyroid FNA cytopathology, Afirma Gene Expression Classifier, GEC, Afirma Genomic Sequencing Classifier, GSC

Thyroid nodules are common, affecting 65% of the population by the age of 60 [1]. Although ultrasound-guided fine needle aspiration (FNA) is the gold standard for evaluation, 15% to 30% of thyroid nodules have indeterminate cytology, which predicts a risk of malignancy of 10% to 40% [2,3]. When evaluating cytologically indeterminate thyroid nodules, current guidelines recommend observation with repeat biopsy, molecular testing, or surgical removal for definitive diagnosis [4,5]. Molecular analysis using the Afirma Gene Expression Classifier (GEC) and Genomic Sequencing Classifier (GSC) are 2 of the available tests used in the evaluation of indeterminate Bethesda System for Reporting Thyroid Cytopathology category III and IV thyroid nodules [4,5]. Afirma GEC became commercially available in 2011 and measures messenger RNA expression of 167 gene patterns [6]. The next-generation GSC replaced GEC in 2017 and also tests for RET, BRAFV600E, and RET/PTC1/3, in addition to the messenger RNA gene expression. It also more accurately identifies parathyroid and Hurthle cell lesions [7].

Implementation of GEC and subsequently GSC has resulted in higher benign call rates (BCR) and reduced overall surgical rates for patients with cytologically indeterminate thyroid nodules [7-15]. In institution-specific analyses, Afirma GEC has reported sensitivities of 83% to 100% [6,8-13,16-21] (Table 1), and GSC has reported sensitivities of 90% to 100% [7,8,10,11,18,19] (Table 2). Positive predictive value (PPV) improved from 16% to 57% [6,8-13,16-25] with GEC to 47% to 85% [7,8,10,11,13,18,19,25] with GSC, and the BCR also improved from 27% to 78% with GEC [6 8-13,15-27] (Table 1) to 54% to 76% with GSC (Table 2) [7,8,10,11,13,15,18,19,25].

Despite these improvements in preoperative evaluation of indeterminate nodules, it is our experience that not all patients undergo molecular testing to aid in decision-making. We evaluated our own institutional experience with cytologically indeterminate thyroid nodules after implementation of GEC and subsequently GSC in 2 Midwest health centers. Test performance for GEC and GSC in this cohort was compared to previously published data. We also evaluated patient and nodule characteristics of those who did not undergo molecular testing to determine the impact of

other variables on the evaluation and management of indeterminate thyroid nodules.

Methods

We received Institutional Review Board approval from both the University of Nebraska Medical Center and the Nebraska–Western Iowa Health System Veteran’s Hospital. We performed a retrospective cohort analysis of consecutive adult patients with cytologically indeterminate thyroid nodules at University of Nebraska Medical Center and the Nebraska–Western Iowa Health System Veteran’s Hospital from January 2013 through December 2019. GEC was first available at our institutions in 2013 and replaced by GSC in October 2017. The decision to obtain molecular testing for cytologically indeterminate thyroid nodules was based on joint decision-making between the provider and the patient. Molecular testing was not reflexive or mandatory for each patient with indeterminate thyroid nodule cytopathology and was an individual decision. These data include all providers involved in caring for patients with indeterminate thyroid nodules so is representative of multiple groups of providers in the 2 institutions. We included all patients with nodules that were biopsied from both institutions, with indeterminate cytology (Bethesda III and Bethesda IV), with and without molecular testing, who also had follow-up data during the study period. Clinical and demographic variables were assessed including age, sex, race or ethnicity, location of residence, body mass index (BMI), thyroid nodule characteristics, imaging characteristics, generation of molecular testing, Bethesda cytologic category, extent of surgery, time to surgery, histopathologic diagnosis, and length of follow-up. Noninvasive follicular thyroid neoplasm with papillary-like nuclear features was included with malignant tumors in the analysis since they require surgical resection for definitive diagnosis. Malignancy was defined for this cohort as carcinoma present in the index indeterminate nodule. Incidental thyroid cancers were excluded from overall malignancy calculations.

Univariate analysis was performed using nonparametric tests with Chi-squared, Fisher’s exact, and Kruskal-Wallis tests for categorical variables and Mann-Whitney U and 1-way analysis of variance

Table 1. Performance of GEC: comparison of institutional experiences

GEC	Sensitivity, %	Specificity, %	PPV, %	NPV, %	BCR, %	Nodule, n	Surgical rate, %	Malignancy rate, %
Alexander et al [6] ^a	92	52	47	93	38	265	100	32
Angell et al [13] ^a	NR	NR	34	NR	47.9	486	51	31
Chaudhary et al [17]	100	15	38	100	40	158	54.4	33.7
Celik et al [26]	NR	NR	NR	NR	33.3	66	57.6	50
Endo et al [8] ^a	94	19	33	89	48.1	343	52.4	42
Endo et al [8] ^b	94	61	33	98	48.1	343	52.4	42
Gortakowski et al [18]	85.7	60.4	22.2	97	60	92	36.9	19.3
Geng et al [19]	91	28	51	79	49	167	42.5	30
Harrell et al [16] (2014)	94.4	23.5	38-57 ^c	80-90 ^c	34	58	63	33-51.4
Harrell et al [10]	88	32	57	NR	42	509	56	51.4
Jug et al [22] #	NR	NR	30.1	NR	51	207	46.3	21
Livhits et al [20]**	100	15.8	38.5	100	42.9	70	43	34.4
Lastra et al [27]	NR	NR	NR	NR	53	132	37.8	44
Mclver et al [12]**	83	10	16	75	27	60	60	17
San Martin et al [11]**	97	60	40	98.6	41	178	47.8	21.6
Sacks et al [23]	NR	NR	33.3	NR	37.1	140	45.1	36
Roychoudhury et al [30] ^d	NR	NR	NR	NR	NR	69	87	18
Kay-Rivest et al [24] ^e	NR	NR	51.51	NR	46	63	52.3	52
Newfoundland								
Kay-Rivest et al [24] ^e	NR	NR	45.71	NR	55	109	40.3	48
Montreal								
Wu et al [9] ^b	95.2	60.1	NR	93.3-100	46.2	245	52	49
Wei et al [15]	NR	NR	36.7	NR	45.4	194	74.5	37
Yang et al [21]	100	15.4	50.7	100	42	217	33.6	46.5
Yang et al [25]	NR	NR	47	88	53	49	NR	NR

Abbreviations: BCR, benign call rate; GEC, Gene Expression Classifier; GSC, Genomic Sequencing Classifier; NPV, negative predictive value; NR, not reported; PPV, positive predictive value.

^aAll nodules with surgical confirmation.

^bAll nodules with surgical confirmation + benign GEC/GSC nodules.

^cRange accounts for cancer prevalence ranging from 33% to 51.4%.

^dAfirma suspicious nodule only.

^eNondiagnostic not included.

Table 2. Performance of GSC: comparison of institutional experiences

GSC	Sensitivity, %	Specificity, %	PPV, %	NPV, %	BCR, %	Nodules, n	Surgical rate, %	Malignancy rate, %
Angell et al [13] ^a	NR	68.3 ^b	5 ^b	NR	65.8	114	32	46
Endo et al [8] ^a	100	17 ^a	60 ^a	100	76.2	164	17.6	52
Endo et al [8] ^b	100	94 ^b	60 ^b	100	76.2	164	17.6	52
Harrell et al [10]	97	44	76	NR	61.2	146	31	64
Gortakowski et al [18]	100	73.7	61.5	97	78	73	20.5	61.5
Geng et al [19]	100	42	61	100	61	133	34.5	36
Kepal et al [7]	91.1	68.3	47.1 ^a	96.1	54	191	100	24
San Martin et al [11] ^b	90.6	94	85.3 ^b	96.3	67.8	121	34.7	27.6
Wei et al [15] ^a	NR	NR	57.1 ^a	NR	66.7	78	53.8	57
Yang et al [25]	NR	NR	64	100	63	51	NR	NR

Abbreviations: BCR, benign call rate GSC, Genomic Sequencing Classifier; NPV, negative predictive value; NR, not reported; PPV, positive predictive value.

^aAll nodules with surgical confirmation.

^bAll nodules with surgical confirmation + benign GEC/GSC nodules.

tests for continuous variables. A Cochran Armitage nonparametric trend test was used to assess trends over time for surgery and molecular testing for the timeline of no reflex molecular testing available, reflex GEC

available, and reflex GSC available. Statistical analysis was performed using STATA version 15 (StataCorp LLC, College Station, TX, USA). A *P*-value of <0.05 was considered statistically significant.

Measurement of test performance for both GEC and GSC was calculated by 2 different methods, using a priori designations for each category using the following assumptions: (1) patients with benign molecular testing were evaluated as true negative only if surgical pathology was available to confirm, and (2) patients with benign molecular testing results were assumed to be true negatives if they did not undergo surgery for definitive diagnosis. Test performance was assessed with calculation of sensitivity, specificity, PPV, and negative predictive value (NPV) for each Bethesda category by generation of molecular test.

Results

A total of 468 Bethesda III and IV thyroid nodules met inclusion criteria for analysis (Fig. 1). Of the 468 nodules analyzed, 273 did not undergo molecular testing, 71 underwent GEC, and 124 underwent GSC testing. There were no differences between the groups with regard to age, sex, BMI, race or ethnicity, and location of residence (local vs out of town). The presence of preexisting hypothyroidism and number of thyroid nodules was similar between groups, but there was a significant difference in dominant nodule size between groups with size being the smallest in the GSC group. Nodule size in greatest dimension was significantly larger at 2.8 ± 1.4 cm in those without molecular testing and 2.8 ± 1.2 cm in the GEC group, compared to 2.3 ± 1.0 cm in the GSC group ($P = 0.0001$). In addition, the presence of constrictive symptoms was significantly higher in those who did not undergo molecular testing (19.6%) vs those who underwent GEC (8.2%) or GSC (6.8%) ($P = 0.0018$). Imaging characteristics using both Thyroid Imaging Reporting and Data System (TIRADS) and American Thyroid Association (ATA) thyroid ultrasound

imaging criteria were not significantly different between groups (TIRADS: $P = 0.0767$; ATA: $P = 0.465$). Surgical rates were significantly different between groups, with lowest rates occurring in the GSC group (39.5%) vs GEC (59.2%) and no molecular testing (67.8%) ($P = 0.0001$). (Table 3) The median time to surgery was longest for the GEC group at 90 days [interquartile range (IQR) 56.5-269 days] compared to 58 days (IQR 44-86 days) for GSC and 44 days for the group without molecular testing (IQR 30-75 days; $P = 0.0001$). The proportion of patients who underwent surgery after 180 days was also highest in the GEC group at 34.4%, compared to 15.2% for the GSC group and 10.6% for the group without molecular testing ($P = 0.0022$) (Table 3).

In a Cochrane Armitage nonparametric time trend analysis, there was no difference in the rate of surgery over time comparing the 3 timeframes: the timeframe without reflex molecular testing sample collection, the timeframe with reflex GEC, and the timeframe with reflex GSC sample collection at the time of FNA ($P = 0.0723$). There was, however, a significant increase in the rate of molecular testing during these time periods from 20.7% to 23.5% to 74.6% respectively ($P < 0.0001$).

The distribution of Bethesda III atypia of undetermined significance (AUS)/follicular neoplasm of undetermined significance (FLUS) and Bethesda IV follicular neoplasm (FN) and Hurthle cell neoplasm (HCN) nodules was significantly different between the 3 groups ($P = 0.0073$). Sixty-seven percent of the GSC were from the Bethesda III AUS/FLUS group, compared to 17% from the GSC Bethesda IV FN and 16% in the GSC Bethesda IV HCN groups. This compares with the GEC group who had 60% in the Bethesda III AUS/FLUS, 30% Bethesda IV FN, and 10% Bethesda IV HCN groups.

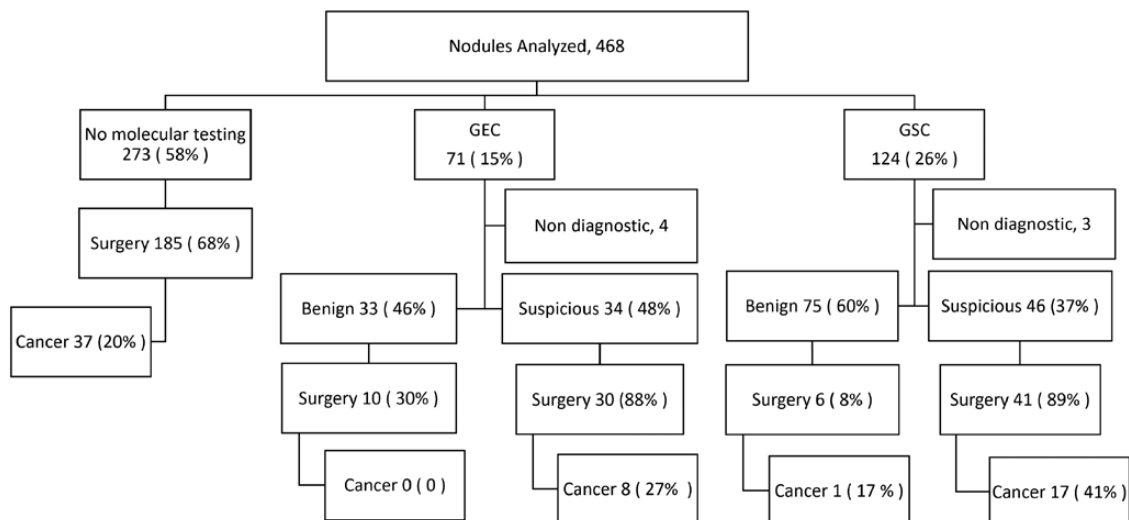


Figure 1. Flow chart demonstrating distribution of the indeterminate thyroid nodules, surgical and malignancy rates.

Table 3. Demographics and clinicopathologic features of those with and without molecular testing

	No molecular testing (n = 273)	Molecular testing GEC (n = 71)	Molecular testing GSC (n = 124)	P-value
Age in years	5.9 ± 14.9	55.4 ± 16.7	56.17 ± 15.7	0.434
Female	202 (74)	45 (63)	85 (69)	0.502
BMI, kg/m ²	30.3 ± 6.9	30.5 ± 6.3	31.1 ± 6.9	0.566
Race				0.3109
White	229/273 (83.9)	62/71 (87.3)	111/124 (89.5)	
Black	29/273 (10.6)	4/71 (5.6)	9/124 (7.3)	
Hispanic	3/273 (1.1)	0/71 (0)	1/124 (0.8)	
Asian	5/273 (1.8)	2/71 (2.8)	2/124 (1.6)	
Other	7/273 (2.6)	3/72 (4.2)	1/124 (0.8)	
Location of residence, local	146/273 (54.5)	37/71 (52.1)	71/124 (57.3)	0.7238
Clinical characteristics				
Preexisting hypothyroidism		12/71 (16.9)	16/124 (12.9)	0.6667
Nodules	2.2 ± 1.2	2.1 ± 1.5	2.1 ± 1.0	0.178
Nodule size, cm	2.8 ± 1.4	2.8 ± 1.2	2.3 ± 1.0	0.0001
Increased growth prior to biopsy	78/221 (35.3)	14/54 (25.9)	24/113 (21.2)	0.0235
Constrictive symptoms	46/235 (19.6)	5/61 (8.2)	8/118 (6.8)	0.0018
Imaging characteristics				
TIRADS				0.0767
1	1/238 (0.4)	0/56 (0)	0/113 (0)	
2	8/238 (3.4)	3/56 (5.4)	5/113 (4.4)	
3	76/238 (31.9)	24/56 (42.9)	28/113 (24.8)	
4	124/238 (52.1)	23/56 (41.1)	62/113 (54.9)	
5	29/238 (12.2)	6/56 (10.7)	18/113 (15.9)	
ATA				0.465
Very low	3/238(1.3)	0/56 (0)	4/113 (3.5)	
Low	86/238 (36.1)	26/56 (46.4)	36/113 (31.9)	
Intermediate	114/238 (47.9)	23/56 (41.1)	54/114 (47.8)	
High	35/238 (14.7)	7/56 (12.5)	19/113 (16.8)	
Hypoechoic	151/237 (63.7)	30/56 (53.8)	69/111 (62.2)	0.3724
Calcifications	67/241 (27.8)	12/56 (21.4)	36/113 (31.9)	0.3622
Cytopathology characteristics				0.0073
Bethesda				
III, AUS/FLUS	126 (46)	43 (60)	83 (67)	
IV, FN	115 (42.1)	21 (30)	21 (17)	
IV, HCN	32 (12)	7 (10)	20 (16)	
Underwent surgery	185/273 (67.8)	42/71 (59.2)	49/124 (39.5)	0.0001
Time to surgery, days	44 (30-75)	90 (56.5-269)	58 (44-86)	0.0001
Time to surgery >180 days	18 (10.6)	11 (34.4)	7 (15.2)	0.0022

Data are given as mean ± SD, n or n/N (%), or median (interquartile range). Bolded P-value indicates significance ≤0.05.

Abbreviations: ATA, American Thyroid Association; AUS, atypia of undetermined significance; BMI, body mass index; FLUS, follicular neoplasm of undetermined significance; FN, follicular neoplasm; GEC, Gene Expression Classifier; GSC, Genomic Sequencing Classifier; HCN, Hurthle cell neoplasm; TIRADS, Thyroid Imaging Reporting and Data System.

When evaluating only the group that did not undergo molecular testing, those who underwent surgery had a significantly higher BMI ($P = 0.019$) and the presence of constrictive symptoms ($P = 0.022$). They also had nodules with higher TIRADS scores: 54.9% of the surgery group had a nodule with TIRADS 4 and 13.6% had a TIRADS 5 nodule, compared to 46.1% TIRADS 4 and 9.2% TIRADS 5 nodules in the no-surgery group ($P = 0.0353$). The type of provider evaluating the nodule was also significantly

different ($P = 0.027$). In the surgery group, 61.8% were seen by a surgeon, compared to 37.4% of the no-surgery group (Table 4).

Reasons for not undergoing molecular testing in our cohort of 273 people were also evaluated. Fifty-six percent (153/273) were recommended surgery and not offered molecular testing by the treating physician, 32% (87/273) were offered testing but declined, and 12% (33/273) had no data. Of the 87 who were offered molecular testing, 54%

(47/87) preferred surgery for other reasons including concurrent Graves' disease, concurrent hyperparathyroidism, growth of nodule, constrictive symptoms, and anxiety/worry about undiagnosed cancers. Seven percent (6/87) had concurrent, nonthyroid cancers, 8% (7/87) had HCNs on FNA and opted not to undergo testing (during the GEC testing when there was a known high false-positive rate), 11% (10/87) required a second biopsy to do testing and declined, 3% (3/87) felt the cost of molecular testing was prohibitive, and 16% (14/87) opted for second FNA rather than molecular testing and repeat FNA was benign, hence eliminating the need for molecular testing.

Overall, BCR was not significantly different between groups with GEC at 46% and GSC 60% ($P = 0.7855$) (Table 5). Surgical rates overall were significantly different between groups, with GSC having the lowest rates at 40% ($P < 0.001$). However, for Bethesda III AUS/FLUS nodules specifically, surgical rates were not different ($P = 0.2959$). There was a significant reduction in surgical rate in the Bethesda IV FN group with 79% for those without molecular testing, 62% for GEC, and 33% for GSC ($P < 0.0001$). For nodules with Bethesda IV HCN cytology, surgical rates were 81% in those without molecular testing, 86% with GEC, and 30% with GSC ($P < 0.0005$) (Table 6).

Table 4. Demographics and clinicopathologic features of those without molecular testing who did and did not undergo surgery

	No surgery (n = 88)	Surgery (n = 185)	P-value
Age in years	57.7 ± 16.0	53.6 ± 14.2	0.199
Race			0.086
White	69/88 (78.4)	160/185 (86.5)	
Black	12/88 (13.6)	17/185 (9.2)	
Hispanic	1/88 (1.1)	2/185 (1.1)	
Asian	3/88 (3.4)	2/185 (1.1)	
Other	3/88 (3.4)	4/185 (2.2)	
Female	65/88 (73.9)	137/185 (74.1)	0.9734
BMI, kg/m ²	28.3 ± 5.8	31.2 ± 7.2	0.019
Location of residence, local	54/88 (61.4)	92/185 (49.7)	0.072
Preexisting hypothyroidism	11/82 (13.4)	23/181 (12.7)	0.874
Nodules	2.4 ± 1.3	2.1 ± 1.2	0.298
Max size of nodule cm	2.46 ± 1.3	2.96 ± 1.5	0.103
Increased growth prior to biopsy	22/53 (41.5)	56/168 (33.3)	0.277
Cytology result			0.0001
AUS/FLUS	58/88 (65.9)	68/185 (36.8)	
FN	24/88 (27.3)	91/185 (49.2)	
HCN	6/88 (6.8)	26/185 (14.1)	
Constrictive symptoms	7/68 (10.3)	39/167 (23.4)	0.022
TIRADS			0.0353
1	0/76 (0)	1/162 (0.6)	
2	5/76 (6.6)	3/162 (1.9)	
3	29/76 (38.2)	47/162 (29.0)	
4	35/76 (46.1)	89/162 (54.9)	
5	7/76 (9.2)	22/162 (13.6)	
ATA			0.0912
Very low	2/76 (2.6)	1/162 (0.6)	
Low	32/76 (42.1)	54/162 (33.3)	
Intermediate	33/76 (43.4)	81/162 (50)	
High	9/76 (11.8)	26/162 (16.1)	
Hypoechoogenicity	47/76 (61.8)	104/161 (64.6)	0.681
Calcifications	20/78 (25.6)	47/163 (28.8)	0.605
Type of provider			0.027
Endocrine	34/83 (40.96)	60/170 (35.3)	
Surgeon	31/83 (37.4)	105/170 (61.8)	
Other	18/83 (21.7)	5/170 (2.9)	

Data are given as mean ± SD or n/N (%). Bolded P-value indicates significance ≤0.05.

Abbreviations: ATA, American Thyroid Association; AUS, atypia of undetermined significance; BMI, body mass index; FLUS, follicular neoplasm of undetermined significance; FN, follicular neoplasm; HCN, Hurthle cell neoplasm; TIRADS, Thyroid Imaging Reporting and Data System.

When analyzing only those with benign molecular testing, surgical rates were lower in the GSC group at 8%, as compared to GEC at 30% ($P < 0.001$). Conversely, when evaluating nodules with suspicious molecular testing, surgical rates were 88% and 89%, respectively, for GEC and GSC ($P = 0.853$) (Fig. 1).

Overall malignancy rates were highest in the GSC group at 39%, compared to 20% and 22% in the no-molecular-testing and GEC groups, respectively ($P = 0.0222$) (Table 7). When evaluating malignancy rates by individual Bethesda categories, Bethesda III AUS/FLUS was the only category with significant differences: 15% malignancy rate in the no-molecular-testing group, GEC 26%, and GSC 39% ($P = 0.0217$). Malignancy rates were not significantly different within the Bethesda IV groups (Table 7) In the nodules with suspicious molecular testing that underwent

surgery, 8 (27%) of GEC and 17 (42%) of GSC had cancer in the index nodule ($P = 0.0222$) (Fig. 1). There were only 4 cases of noninvasive follicular thyroid neoplasm with papillary-like nuclear features in this cohort; 3 did not have molecular testing, and the fourth had GSC with a suspicious result. Incidental thyroid cancer found on histopathology outside of the index indeterminate nodule was noted in 9%, although these did not contribute to malignancy rate calculations for analysis of test performance.

For Bethesda IV HCN nodules specifically, the BCR was significantly different: BCR was 1/7 (14.3%) for GEC and 13/20 (65%) for GSC ($P = 0.0254$) (Table 5) Surgical rates were highest for the no-molecular-testing group at 81% as compared to GEC at 86% and GSC at 30% ($P = 0.0005$) (Table 6). Malignancy rates were 12% for those without molecular testing, compared to 0% with GEC and 33%

Table 5. Benign call rate

	GEC	GSC	P-value
Overall	33/71 (46)	75/124 (60)	0.7855
Bethesda			
III, AUS/FLUS	21/43 (49)	50/83 (60)	0.544
IV, FN	11/21 (52)	12/21 (57)	0.451
IV, HCN	1/7 (14)	13/20 (65)	0.0254

Data are given as n (%). Bolded P-value indicates significance ≤ 0.05 .

Abbreviations: AUS, atypia of undetermined significance; FLUS, follicular neoplasm of undetermined significance; FN, follicular neoplasm; GEC, Gene Expression Classifier; GSC, Genomic Sequencing Classifier; HCN, Hurthle cell neoplasm.

Table 6. Surgical rates

	Molecular testing not performed	GEC	GSC	P-value
Overall	185/273 (68)	42/71 (59)	49/124(40)	0.0001
Bethesda				
III, AUS/FLUS	68/126 (54)	23/43 (53)	36/83 (43)	0.2959
IV, FN	91/115 (79)	13/21 (62)	7/21 (33)	0.0001
IV, HCN	26/32 (81)	6/7 (86)	6/20 (30)	0.0005

Data are given as n (%). Bolded P-value indicates significance ≤ 0.05 .

Abbreviations: AUS, atypia of undetermined significance; FLUS, follicular neoplasm of undetermined significance; FN, follicular neoplasm; GEC, Gene Expression Classifier; GSC, Genomic Sequencing Classifier; HCN, Hurthle cell neoplasm.

Table 7. Malignancy rates

	Molecular testing not performed	GEC	GSC	P-value
Overall	37/185 (20)	9/41 (22)	19/49 (39)	0.0222
Bethesda				
III, AUS/FLUS	10/68 (15)	6/23 (26)	14/36 (39)	0.0217
IV, FN	24/91 (26)	3/12 (25)	3/7 (43)	0.6322
IV, HCN	3/26 (12)	0/6 (0)	2/6 (33)	0.2204

Data are given as n (%). Bolded P-value indicates significance ≤ 0.05 .

Abbreviations: AUS, atypia of undetermined significance; FLUS, follicular neoplasm of undetermined significance; FN, follicular neoplasm; GEC, Gene Expression Classifier; GSC, Genomic Sequencing Classifier; HCN, Hurthle cell neoplasm.

with GSC ($P = 0.2204$) (Table 7). None of the resected HCN nodules were malignant in the GEC group.

Measurements of test performance were calculated for both GEC and GSC. We calculated performance using 2 different methods. First, we included only the surgically resected nodules. Using this definition, GEC sensitivity was 100%; specificity, 32%; PPV, 28%; and NPV, 100%. This compares to GSC with a sensitivity of 94%; specificity, 17%; PPV, 41%; and NPV, 83% (Table 8). The second method included both surgically resected nodules and unresected GEC or GSC benign nodules as true negatives. Using this definition, GEC sensitivity was 100%; specificity, 61%; PPV, 28%; and NPV, 100%. GSC sensitivity was 94%; specificity, 76%; PPV, 41%; and NPV, 97% (Table 9).

Test performance for GEC and GSC was also measured for each Bethesda category (Tables 8 and 9). Due to the small number of Hurthle cell lesions, performance measures were not calculated.

Discussion

We report Afirma GEC and GSC use in cytologically indeterminate thyroid nodules in 2 Midwest academic institutions. We evaluated BCR and surgical and malignancy rates, as well as sensitivity, specificity, PPV, and NPV and found our experience to be similar to multiple previously published institutional experience studies [6-13,15,17-22,24-27] (Tables 1 and 2).

Over the last decade, there have been progressive improvements to commercially available molecular tests for cytologically indeterminate thyroid nodules. Surgical rates were reported as 34% to 87% for GEC and reduced to 18% to 54% with GSC [7-13,15,16,17-19,21-24,26,27]. In our cohort, surgical rate for indeterminate nodules without molecular testing was 68%, and implementation of GEC did not significantly reduce this. Only after implementation of GSC was the rate of surgery significantly reduced to 40%. This is similar to the results published by Sacks et al, where overall surgical rate was similar for those

Table 8. Performance of GEC and GSC: all nodules with surgical confirmation

	Sensitivity	Specificity	PPV	NPV	BCR
GEC					
All nodules	100	32	28	100	46
Bethesda nodules					
III	100	29	33	100	49
IV, FN	100	44	29	100	52
GSC					
All nodules	94	17	41	83	60
Bethesda nodules					
III	92	19	43	80	60
IV, FN	100	0	43	0	57

Data given as %.

Abbreviations: BCR, benign call rate; FN, follicular neoplasm; GEC, Gene Expression Classifier; GSC, Genomic Sequencing Classifier; NPV, negative predictive value; PPV, positive predictive value.

Table 9. Performance of GEC and GSC: all nodules with surgical confirmation + benign GEC/GSC nodules

	Sensitivity	Specificity	PPV	NPV	BCR
GEC					
All nodules	100	61	28	100	46
Bethesda nodules					
III	100	64	33	100	49
IV, FN	100	69	29	100	52
GSC					
All nodules	94	76	41	97	60
Bethesda nodules					
III	93	74	43	98	60
IV, FN	100	75	43	100	57

Data given as %.

Abbreviations: BCR, benign call rate; FN, follicular neoplasm; GEC, Gene Expression Classifier; GSC, Genomic Sequencing Classifier; NPV, negative predictive value; PPV, positive predictive value.

without molecular testing at 43.5%, compared to 46.5% for those with GEC [23]. We captured follow-up data and included all patients who underwent surgery at a later date with records available at our institution. We had a mean follow-up of 20.6 months (SD = 12), minimizing the possibility of missed surgeries within 2 years of biopsy so we do not believe this was a falsely low estimate of total number of surgeries. Additionally, time to surgery was the highest with GEC and the lowest for no molecular testing. When GEC testing was offered at our institutions, it initially required a second biopsy, which could explain the time to surgery. Once we implemented sample collection at the time of initial FNA with reflex testing for indeterminate cytology, time to surgery was reduced. We also captured those who had surgery during follow up >180 days after biopsy. GEC had the highest proportion of patients (34.4%) who had surgery >180 days after biopsy compared to 15.2% and 10.6% of the GSC and no-molecular-testing groups, respectively. Data for time to surgery for benign GEC result are available for 8 of the 10 patients with a median of 312 days (IQR 158.5-778 days) and a mean of 488 days (SD = 478). Nodule growth characteristics were available for 6 of these patients with median growth of 0.55 cm (IQR 0.3-0.6 cm) and mean of 0.62 cm (SD = 0.46).

This difference between GEC and GSC could be attributed to decreased time for overall follow-up, or it could be due to improved performance of the test.

When assessing the surgical rate by Bethesda category in our cohort, Bethesda III surgical rates were not significantly different between those without molecular testing and those with GEC or GSC. For patients with Bethesda IV nodules, however, surgical rate significantly declined progressively from those without molecular testing to GEC and further to GSC, indicating GSC does have a significant impact on surgical rate reduction. This was also true for Bethesda IV HCN groups, consistent with previous studies reporting improved BCR in Hurthle cell lesions.[7,10,11,13,16].

The malignancy rates and test performance in our cohort are similar to previously published data. Our malignancy rate is 20% for those without molecular testing, 22% for those with GEC, and 39% for those with GSC. These are comparable to San Martin et al's malignancy prevalence of 22% in GEC and 28% in GSC [11]. When assessing test performance, our findings of high sensitivity and NPV are also comparable to previously published data [7-13,15,16]. The GSC has been reported as higher specificity and higher PPV compared to GEC.[7,8,10,11,13]. In our study, this finding was replicated with improvement of specificity from GEC to GSC from 61% to 76% and PPV from 28% to 41% when using all nodules. For those with surgical confirmation, the performance was not as robust, with GEC vs GSC specificity 32% vs 17% and PPV 28%

vs 41%. Other studies have shown PPV of 16% to 57% for GEC and 47% to 85% for GSC [6-8,10-13,15-25]. Our GEC and GSC PPV is comparable to other previously published institutional analyses. (Tables 1 and 2) In our study, NPV was 100% for GEC, but 83% for GSC when only evaluating those with surgical confirmation. There were only 7 people in the GSC group, and 1 had cancer. The low numbers in this group likely contribute to the low NPV in this study, which is lower than previously published studies. NPV is higher at 99% in our cohort when including benign GSC nodules without surgical confirmation.

In addition to evaluating cytologically indeterminate thyroid nodules with Afirma GEC or GSC testing, we also looked at those with indeterminate thyroid nodules who did not undergo molecular testing. Not surprisingly, patients who did not undergo molecular testing had significantly larger nodule size, growth of the nodule prior to biopsy, and constrictive symptoms. All these factors likely influenced the joint patient/provider decision to forego pre-operative molecular testing and proceed straight to surgery. Those without molecular testing had significantly higher rates of surgery. Ultrasound characteristics evaluated by both TIRADS and ATA sonographic risk of malignancy criteria were not different between groups overall, and specifically echogenicity and presence of calcifications was also not significantly different. However, it should be noted that many of the years included in this retrospective study were prior to the current ATA thyroid nodule guidelines that first recommended consideration of molecular testing in indeterminate thyroid nodules [4]. We also did not find significant differences between Bethesda categories in those who did and did not undergo molecular testing. This is in contrast to Lee et al [28] who explored patient preferences for molecular testing of indeterminate thyroid nodules and reported a higher number of patients with Bethesda IV cytology or high-risk ultrasound features opted for molecular testing, which was not seen in our study. Time trend analysis showed there was no difference in the rate of surgery over the 3 different time frames in our analysis: prereflex molecular testing, reflex GEC, and reflex GSC testing. However, there was a significant increase in the rate of molecular testing overall during these time frames, indicating molecular testing became more acceptable and commonplace over time.

To better understand the decision-making regarding molecular testing and surgery, we also evaluated the differences between those without molecular testing who did and did not undergo surgery. Unlike the comparison between groups for or against molecular testing, among those who did not undergo molecular testing, there was a difference between those who underwent surgery and those who did not regarding ultrasound TIRADS score and cytology.

Those who underwent surgery had higher TIRADS scores and were more likely to have Bethesda IV (FN or HCN) cytologic diagnosis. Type of provider was also a significant predictor of surgery. If a person with an indeterminate thyroid nodule was seen by an endocrinologist, 41% did not undergo surgery and 35% did undergo surgery. If seen by a surgeon, 62% underwent surgery compared to 37% who did not, and if seen by providers other than an endocrinologist or surgeon, they were more likely to avoid surgery. Depending on the location, patients have variable access to different types of providers. Provider familiarity with guidelines, available testing, and interpretation of testing can vary widely among providers and likely have a much larger impact on decision-making regarding molecular testing in indeterminate thyroid nodules than has been previously evaluated.

One of the limitations of this study is the retrospective methodology. We reviewed all ultrasounds and sonographic characteristics as well as clinical notes to determine reasons for or against molecular testing, but some had missing data. When evaluating these factors, there was no apparent difference between those who did not have molecular testing and those who had GEC or GSC in regards to pre biopsy sonographic characteristics, Bethesda cytology classification or demographic characteristics including age, sex, BMI, race, and distance from the treating facility. This makes it less likely that selection of lower risk nodules for molecular testing reduced the PPV and specificity in our study.

Another limitation of this study is the inability to formally characterize patients with benign molecular testing who did not undergo surgery for confirmation of benignity. Up to a 6% false-negative rate has been reported in those with benign molecular testing, discovered by changes on serial ultrasound assessment longitudinally [29]. However, this is a limitation of all retrospective indeterminate thyroid FNA molecular studies, given surgery is avoided in all those with benign molecular test results.

Conclusions

Surgical rates in patients with cytologically indeterminate thyroid nodules without molecular testing and those with Afirma GEC were not significantly different in our cohort. Only those with Afirma GSC testing had a significant reduction in surgical rates and increase in malignancy rates after implementation of testing. Sensitivity and NPV were high for both GEC and GSC. Those who do not undergo molecular testing of thyroid nodule more commonly have larger nodule size, growth of thyroid nodule, and constrictive symptoms. In patients who did not undergo molecular testing, those that underwent surgery had overall

higher BMI, constrictive symptoms, higher TIRADS score on ultrasound imaging, and higher Bethesda classification. In those without molecular testing, if patients had seen a surgeon, they were more likely to undergo surgery, as compared to patients seeing an endocrinologist or other provider. Further studies are needed to understand the practical application of these molecular markers preoperatively in cytologically indeterminate thyroid nodules as not all patients opt for the use molecular testing for further evaluation in this clinical scenario.

Additional Information

Correspondence: Preethi Polavarapu, MBBS, 984120 Nebraska Medical Center, Omaha, NE 68198-4120, USA. E-mail: preethi.polavarapu@unmc.edu.

Disclosures: V.S.: Site principal investigator for multicenter trials for NovoNordisk, Eli Lilly, and Kowa Pharmaceuticals. W.G.: Site principal investigator for Roche and Siemens. All other authors have nothing to disclose.

Data Availability: Some or all data sets generated during and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

References

- Guth S, Theune U, Aberle J, Galach A, Bamberger CM. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. *Eur J Clin Invest*. 2009;39(8):699-706.
- Cibas ES, Ali SZ. The 2017 Bethesda System for Reporting Thyroid Cytopathology. *Thyroid*. 2017;27(11):1341-1346.
- Baloch ZW, Cooper DS, Gharib H, Alexander EK. Overview of diagnostic terminology and reporting. In: Ali SZ, Cibas ES, eds. *The Bethesda System for Reporting Thyroid Cytopathology: Definitions, Criteria, and Explanatory Notes*. Springer International Publishing; 2018:1-6.
- 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.
- National Comprehensive Cancer Network. Thyroid carcinoma (version 2.2020). Accessed August 24, 2020. https://www.nccn.org/professionals/physician_gls/pdf/thyroid.pdf
- 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.
- Patel KN, Angell TE, Babiarz J, et al. Performance of a genomic sequencing classifier for the preoperative diagnosis of cytologically indeterminate thyroid nodules. *JAMA Surg*. 2018;153(9):817-824.
- Endo M, Nabhan F, Porter K, et al. Afirma Gene Sequencing Classifier compared with Gene Expression Classifier in indeterminate thyroid nodules. *Thyroid*. 2019;29(8):1115-1124.

9. Wu JX, Young S, Hung ML, et al. Clinical factors influencing the performance of Gene Expression Classifier testing in indeterminate thyroid nodules. *Thyroid*. 2016;**26**(7):916-922.
10. Harrell RM, Eyerly-Webb SA, Golding AC, Edwards CM, Bimston DN. Statistical comparison of Afirma GSC and Afirma GEC outcomes in a community endocrine surgical practice: early findings. *Endocr Pract*. 2019;**25**(2):161-164.
11. San Martin VT, Lawrence L, Bena J, et al. Real-world Comparison of Afirma GEC and GSC for the assessment of cytologically indeterminate thyroid nodules. *J Clin Endocrinol Metab*. 2020;**105**(3):dgz099.
12. 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.
13. Angell TE, Heller HT, Cibas ES, et al. Independent Comparison of the Afirma Genomic Sequencing Classifier and Gene Expression Classifier for Cytologically Indeterminate Thyroid Nodules. *Thyroid*. 2019;**29**(5):650-656.
14. Liu Y, Pan B, Xu L, Fang D, Ma X, Lu H. The diagnostic performance of Afirma Gene Expression Classifier for the indeterminate thyroid nodules: a meta-analysis. *Biomed Res Int*. 2019;**2019**:7150527.
15. Wei S, Veloski C, Sharda P, Ehya H. Performance of the Afirma Genomic Sequencing Classifier versus Gene Expression Classifier: an institutional experience. *Cancer Cytopathol*. 2019;**127**(11):720-724.
16. 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.
17. Chaudhary S, Hou Y, Shen R, Hooda S, Li Z. Impact of the Afirma Gene Expression Classifier result on the surgical management of thyroid nodules with category III/IV cytology and its correlation with surgical outcome. *Acta Cytol*. 2016;**60**(3):205-210.
18. Gortakowski M, Feghali K, Osakwe I. Single institution experience with Afirma and Thyroseq testing in indeterminate thyroid nodules. *Thyroid*. 2021;**31**(9):1376-1382.
19. Geng Y, Aguilar-Jakthong JS, Moatamed NA. Comparison of Afirma Gene Expression Classifier with Gene Sequencing Classifier in indeterminate thyroid nodules: a single-institutional experience. *Cytopathology*. 2021;**32**(2):187-191.
20. Livhits MJ, Kuo EJ, Leung AM, et al. Gene Expression Classifier vs targeted next-generation sequencing in the management of indeterminate thyroid nodules. *J Clin Endocrinol Metab*. 2018;**103**(6):2261-2268.
21. Yang SE, Sullivan PS, Zhang J, et al. Has Afirma Gene Expression Classifier testing refined the indeterminate thyroid category in cytology? *Cancer Cytopathol*. 2016;**124**(2):100-109.
22. Jug RC, Datto MB, Jiang XS. Molecular testing for indeterminate thyroid nodules: performance of the Afirma Gene Expression Classifier and ThyroSeq panel. *Cancer Cytopathol*. 2018;**126**(7):471-480.
23. Sacks WL, Bose S, Zumsteg ZS, et al. Impact of Afirma Gene Expression Classifier on cytopathology diagnosis and rate of thyroidectomy. *Cancer Cytopathol*. 2016;**124**(10):722-728.
24. Kay-Rivest E, Tibbo J, Bouhabel S, et al. The first Canadian experience with the Afirma® Gene Expression Classifier test. *J Otolaryngol Head Neck Surg*. 2017;**46**(1):25.
25. Yang Z, Zhang T, Layfield L, Esebua M. Performance of Afirma Gene Sequencing Classifier versus Gene Expression Classifier in thyroid nodules with indeterminate cytology. *J Am Soc Cytopathol*. 2021;**S2213-2945**(21)00071-5.
26. Celik B, Whetsell CR, Nassar A. Afirma GEC and thyroid lesions: an institutional experience. *Diagn Cytopathol*. 2015;**43**(12):966-970.
27. Lastra RR, Pramick MR, Crammer CJ, LiVolsi VA, Baloch ZW. Implications of a suspicious Afirma test result in thyroid fine-needle aspiration cytology: an institutional experience. *Cancer Cytopathol*. 2014;**122**(10):737-744.
28. Lee DJ, Xu JJ, Brown DH, et al. Determining patient preferences for indeterminate thyroid nodules: observation, surgery or molecular tests. *World J Surg*. 2017;**41**(6):1513-1520.
29. Zhu CY, Donangelo I, Gupta D, et al. Outcomes of indeterminate thyroid nodules managed nonoperatively after molecular testing. *J Clin Endocrinol Metab*. 2021;**106**(3):e1240-e1247.
30. Roychoudhury S, Klein M, Souza F, et al. How “suspicious” is that nodule? Review of “suspicious” Afirma Gene Expression Classifier in high risk thyroid nodules. *Diagn Cytopathol*. 2017;**45**(4):308-311.