Title

Do Ultrasound Patterns and Clinical Parameters Inform the Probability of Thyroid Cancer Predicted by Molecular Testing in Nodules with Indeterminate Cytology?

Journal

Thyroid, 31(11)

ISSN

1050-7256

Authors

Figge, James J
Gooding, William E
Steward, David L
et al.

Publication Date

2021-11-01

DOI

10.1089/thy.2021.0119

Peer reviewed

THYROID
Volume 31, Number 11, 2021

DOI: 10.1089/thy.2021.0119

Do Ultrasound Patterns and Clinical Parameters Inform the Probability of Thyroid Cancer Predicted by Molecular Testing in Nodules with Indeterminate Cytology?

James J. Figge,[1] William E. Gooding,[2] David L. Steward,[3] Linwah Yip,[4] Rebecca S. Sippel,[5] Samantha Peiling Yang,[6,7,i] Randall P. Scheri,[8] Jennifer A. Sipos,[9] Susan J. Mandel,[10] Sarah E. Mayson,[11] Kenneth D. Burman,[12] Jessica M. Folek,[13] Bryan R. Haugen,[11] Julie A. Sosa,[14] Rajeev Parameswaran,[15,16] Wee Boon Tan,[15,16] Yuri E. Nikiforov,[17] and Sally E. Carty[4]

Background: Molecular testing (MT) is commonly used to refine cancer probability in thyroid nodules with indeterminate cytology. Whether or not ultrasound (US) patterns and clinical parameters can further inform the risk of thyroid cancer in nodules predicted to be positive or negative by MT remains unknown. The aim of this study was to test if clinical parameters, including patient age, sex, nodule size (by US), Bethesda category (III, IV, V), US pattern (American Thyroid Association [ATA] vs. American College of Radiology Thyroid Image Reporting and Data System [TI-RADS] systems), radiation exposure, or family history of thyroid cancer can modify the probability of thyroid cancer or noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP) predicted by MT.

Methods: We studied 257 thyroid nodules in 232 patients from 10 study centers with indeterminate fine needle aspiration cytology and informative MT results using the ThyroSeq v3 genomic classifier (TSv3). Univariate and multivariate logistic regression was used for data analysis.

Results: The presence of cancer/NIFTP was associated with positive TSv3 results (odds ratio 61.39, $p < 0.0001$). On univariate regression, patient sex, age, and Bethesda category were associated with cancer/NIFTP probability ($p < 0.05$ for each). Although ATA ($p = 0.1211$) and TI-RADS ($p = 0.1359$) US categories demonstrated positive trends, neither was significantly associated with cancer/NIFTP probability. A multivariate regression model incorporating the four most informative non-MT covariates (sex, age, Bethesda category, and ATA US pattern; Model No. 1) yielded a C index of 0.653; $R^2 = 0.108$. When TSv3 was added to Model number 1, the C index increased to 0.888; $R^2 = 0.572$. However, age ($p = 0.341$), Bethesda category ($p = 0.272$), and ATA US pattern ($p = 0.264$) were nonsignificant, and other than TSv3 ($p < 0.0001$), male sex was the only non-MT parameter that potentially contributed to cancer/NIFTP risk ($p = 0.095$). The simplest and most efficient clinical model (No. 3) incorporated TSv3 and sex (C index = 0.889; $R^2 = 0.588$).

[1]Diabetes & Endocrine Care, St. Peter's Health Partners/Trinity Health, Rensselaer, New York, USA.
[2]Biostatistics Facility, UPMC Hillman Cancer Center, Pittsburgh, Pennsylvania, USA.
[3]Department of Otolaryngology, Head and Neck Surgery, University of Cincinnati Medical Center, Cincinnati, Ohio, USA.
[4]Division of Endocrine Surgery, University of Pittsburgh, Pittsburgh, Pennsylvania, USA.
[5]Division of Endocrine Surgery, University of Wisconsin, Madison, Wisconsin, USA.
[6]Endocrinology Division, Department of Medicine, National University Hospital, Singapore, Singapore.
[7]Endocrinology Division, Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore.
[8]Section of Endocrine Surgery, Department of Surgery, Duke University, Durham, North Carolina, USA.
[9]Division of Endocrinology, Diabetes, and Metabolism, Ohio State University School of Medicine, Columbus, Ohio, USA.
[10]Division of Endocrinology, Diabetes and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA.
[11]Division of Endocrinology, Metabolism and Diabetes, University of Colorado School of Medicine, Aurora, Colorado, USA.
[12]Endocrinology Section, Department of Medicine, MedStar Washington Hospital Center, Washington, District of Columbia, USA.
[13]PeaceHealth Medical Group, Springfield, Oregon, USA.
[14]Department of Surgery, University of California San Francisco, San Francisco, California, USA.
[15]Division of Endocrine Surgery, Department of Surgery, National University Hospital, Singapore, Singapore,
[16]Division of Endocrine Surgery, Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore.
[17]Department of Pathology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA.
[i]ORCID ID (https://orcid.org/0000-0003-0838-6721).

Conclusions: In this multicenter study of thyroid nodules with indeterminate cytology and MT, neither the ATA nor TI-RADS US scoring systems further informed the risk of cancer/NIFTP beyond that predicted by TSv3. Although age and Bethesda category were associated with cancer/NIFTP probability on univariate analysis, in sequential nomograms they provided limited incremental value above the high predictive ability of TSv3. Patient sex may contribute to cancer/NIFTP risk in thyroid nodules with indeterminate cytology.

Keywords: indeterminate cytology, logistic regression models, molecular testing, thyroid cancer, thyroid nodules, thyroid ultrasound

Introduction

THYROID NODULES ARE detected by ultrasound (US) in as many as two-thirds of asymptomatic persons (1,2). While most clinically evaluated nodules are benign, 5–16% harbor thyroid cancer (2–4). Clinical, demographic, and ultrasonographic features have proven useful in selecting nodules for fine-needle aspiration (FNA) biopsy (2,5). Several US-based algorithms are commonly used for risk stratification, including the American Thyroid Association (ATA) sonographic risk stratification system for thyroid nodules (6) and the American College of Radiology (ACR) Thyroid Image Reporting and Data System (TI-RADS) (7). Both algorithms account for nodule size and detailed ultrasonographic characteristics, assigning a risk score accordingly.

Thyroid FNA biopsy accurately classifies most nodules, but a significant proportion (20–38%) lack specific cytologic features for definitive classification (8,9). The Bethesda System for Reporting Thyroid Cytopathology (10,11) includes three categories of indeterminate cytology: Bethesda category III (atypia of undetermined significance or follicular lesion of undetermined significance), which accounts for about 10% of FNA results and has an institution-dependent cancer rate of 6–48% (8); Bethesda category IV (follicular neoplasm/suspicious for follicular or Hurthle cell neoplasm), which accounts for about 10% of FNA results with a variable thyroid cancer rate of 14–34% (8); and Bethesda category V (suspicious for malignancy), which comprises 2–3% of FNA results and has a 53–97% risk of malignancy (8). The uncertainty associated with indeterminate cytopathology greatly complicates patient management, as it often results in unnecessary diagnostic surgery as only 10–40% of patients undergoing surgery for a Bethesda III or IV nodule have cancer on histology (12). Furthermore, diagnostic uncertainty associated with indeterminate cytology can also necessitate two-stage (reoperative completion) thyroidectomy if clinically significant cancer is present on lobectomy (12). In histologic analyses, noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP) is included in the same category as thyroid cancer because it requires surgery for definitive diagnosis (13).

Over the past two decades, great progress has been made in elucidating the molecular landscape of thyroid cancer, and this has set the stage for routine preoperative molecular analysis of indeterminate FNA samples. The ThyroSeq v3 genomic classifier is a 112-gene test that can detect a broad range of thyroid cancer-related point mutations, gene fusions, copy number alterations, insertions, deletions, and gene expression alterations (14). Its clinical validation was recently defined in a prospective blinded multicenter study that demonstrated a high sensitivity (94%) and reasonably high specificity (82%) in classifying Bethesda III or IV indeterminate thyroid nodules as benign versus cancer/NIFTP (15,16). In short-term follow-up, its clinical utility was recently demonstrated to quadruple the surgical yield of cancer, identifying all potentially aggressive malignancies and safely triaging >80% of ThyroSeq v3-negative patients to nonoperative surveillance (17).

Today, several types of molecular testing (MT) are commonly used to refine cancer/NIFTP probability in indeterminate thyroid nodules, although all of the available tests have limitations and do not provide 100% accuracy in classifying such nodules as benign or malignant. Moreover, MT is not available or feasible in all practice settings. In this study, we took advantage of the availability of a large cohort of patients with indeterminate thyroid cytology analyzed by ThyroSeq v3 in a recent multicenter study (15) to examine if readily available clinical, cytologic, and radiologic parameters are independently associated with the probability of same-nodule histologic thyroid cancer/NIFTP. We then systematically compared the relative contributions of the identified covariates in several novel clinical models.

Materials and Methods

This study was a *post hoc* exploratory data analysis with sequential statistical modeling. It utilized the previously described data set from the prospective, double-blind ThyroSeq v3 clinical validity trial (15), which was composed of 257 thyroid nodules from patients aged 18 years or older who were managed at 10 study centers with FNA results yielding indeterminate cytology (154 Bethesda III, 93 Bethesda IV, and 10 Bethesda V), informative MT results using the ThyroSeq v3 genomic classifier, and surgical outcome available for all patients (15). Importantly, the decision to manage these nodules surgically was made based on routine clinical algorithms used at each study center, as the results of MT were not available at the time of patient surgery (15). Furthermore, the study centers did not utilize prior versions of ThyroSeq nor other molecular tests to assist with surgical decision-making.

The study was approved by the University of Pittsburgh Medical Center Institutional Review Board (No. PRO14030694) and by the respective review boards at all participating study centers. For analysis, the previously coded data on patient age, sex, nodule size by US, and Bethesda category (III, IV, or V) were accessed for each participating center, and additional covariate data were retrieved *post hoc*, including personal history of radiation exposure, family history of thyroid cancer, and US pattern. Specifically, the US images were reviewed at each study center to assign the ATA

(6) and TI-RADS (7) risk categories for each nodule as described in Supplementary Table S1. TI-RADS scores were assigned by having the reviewers enter a nodule's sonographic characteristics into a computational tool, and the points associated with each feature were totaled. As in Steward *et al.* (15), ThyroSeq v3 results were coded as positive or negative for cancer/NIFTP, and cases with NIFTP were counted in the same category as cancer.

Statistical analysis

Univariate logistic regression was initially used to study each covariate. Those that were most strongly associated with thyroid malignancy/NIFTP were further assessed using multivariate logistic regression. All analyses assumed that nodules are independent, including multiple nodules from the same individual.

An initial multivariate logistic regression model was then created using all covariates except ThyroSeq v3 that were associated with malignancy or NIFTP in univariate analysis. Those covariates that retained their association with malignancy/NIFTP in a multivariate model were retained as Model number 1. The results of ThyroSeq v3 were introduced into Model 1 to create Model number 2. Model number 3 selected covariates based on explained variation rather than using predetermined criteria and was considered to be a parsimonious model. Multivariate model covariates were selected by Akaike's information criteria to balance fit and complexity. Models were summarized with the C index (equivalent to the area under a receiver operating characteristic curve) and R^2, the proportion of total variation explained by the model. The C index and R^2 were internally cross-validated with 200 bootstrap samples.

In a set of secondary analyses, each of three covariates (ATA US risk category, TI-RADS US risk category, and Bethesda category) was tested for its association with ThyroSeq v3 positivity. For the US patterns, an exact Cochran/Armitage trend was used to assess whether the proportion with positive ThyroSeq v3 determinations differed by risk category and whether an upward trend was significant. Fisher's exact test was used to determine whether there was an association between increasing Bethesda category and the proportion of positive ThyroSeq v3 determinations.

Results

Initial data review revealed that the radiation exposure status for many of the patients was unknown or unclear as to age and latency time from radiation to cancer onset, and so, this parameter was removed from further analysis. Among the 232 study patients with 257 nodules (Table 1), 80% of patients were female, the median age was 53 years (range 18–90), and the median nodule size was 2.1 cm (mean nodule size 2.4 cm; interquartile range 1.5–3.1 cm; maximum nodule size 7 cm; minimum 0.5 cm). Bethesda III was the most frequent cytologic category (60%) followed by Bethesda IV (36%), with the remaining 4% of nodules having Bethesda V cytology. Fifty-nine percent of nodules had negative ThyroSeq v3 results, and after surgery 70% had benign histology. The distribution of the ATA and TI-RADS US risk categories for the study cohort is shown in Table 1.

Table 1. Characteristics of the Study Cohort

	N *(%)*
Patient characteristics ($N=232$)	
Gender	
Male	46 (20)
Female	186 (80)
Age	
Median	53
IQR	42:62
Range	18–90
Family history	14 (6)
Nodule characteristics ($N=257$)	
Nodule size	
Median	2.1
IQR	1.5–3.1
Bethesda category	
III	154 (60)
IV	93 (36)
V	10 (4)
Consensus diagnosis group	
Benign	181 (70)
NIFTP	11 (4)
Malignant	65 (25)
ThyroSeq v3 results	
Positive	105 (41)
Negative	152 (59)
ATA US pattern	
Very low suspicion	21 (8)
Low suspicion	125 (49)
Intermediate suspicion	87 (34)
High suspicion	22 (9)
TI-RADS score	
TR1—benign	12 (5)
TR2—not suspicious	20 (8)
TR3—mildly suspicious	104 (40)
TR4—moderately suspicious	100 (39)
TR5—highly suspicious	21 (8)

ATA, American Thyroid Association; IQR, interquartile range; NIFTP, noninvasive follicular thyroid neoplasm with papillary-like nuclear features; TI-RADS, Thyroid Image Reporting and Data System; US, ultrasound.

Univariate logistic regression

Clinical parameters were evaluated to determine their individual association with thyroid cancer/NIFTP using univariate logistic regression (Table 2). In this cohort of thyroid nodules with indeterminate FNA cytology that were managed surgically, only patient sex, age, Bethesda category, and ThyroSeq v3 were significantly associated with thyroid cancer/NIFTP probability ($p<0.05$ for each). Although the ATA ($p=0.1211$) and TI-RADS ($p=0.1359$) US risk groups demonstrated modest positive trends, neither was significantly associated with cancer/NIFTP probability. However, the ATA US risk groups demonstrated a more continuous progressively increasing trend (Fig. 1), and so, ATA was more appropriate than TI-RADS for inclusion in multivariate models.

The ThyroSeq v3 result dominated the performance of the other parameters, with an odds ratio (OR) of 61.39 ($p<0.0001$).

Male sex (OR 2.48; $p=0.0053$) and Bethesda category V (OR 5.72; $p=0.0003$) were significantly associated with an increased thyroid cancer/NIFTP risk, and age demonstrated a

Table 2. Univariate Effect of Covariates on the Prediction of Malignancy/Noninvasive Follicular Thyroid Neoplasm with Papillary-like Nuclear Features

Covariate	*Reference*	*Odds ratio [CI]*	p
Gender	Male:female	2.48 [1.31–4.68]	0.0053
Age	62:42	0.78 [0.53–1.16]	0.0314
Family history	Yes:no	1.62 [0.55–4.72]	0.3787
Nodule size	3.1:1.5	1.19 [0.83–1.70]	0.3389
Bethesda category	V:III	5.72 [2.23–14.65]	0.0003
ATA US patterns	High suspicion:low suspicion	1.74 [0.86–3.55]	0.1211
ACR TI-RADS	Highly suspicious:mildly suspicious	1.59 [0.86–2.92]	0.1359
ThyroSeq v3	Positive:negative	61.39 [23.02–163.67]	<0.0001

The reference column refers to the range of reference values over which the odds ratio is calculated. A:B signifies that the reported odds ratio is for A relative to B. For example: the odds of malignancy of Bethesda category V compared with BC III are 5.72. The odds of malignancy for ATA high suspicion relative to low suspicion are 1.74.

ACR, American College of Radiology; CI, 95% confidence interval.

nonlinear risk profile (OR = 0.78; $p = 0.0314$), in that age-dependent risk showed a decline until about age 55 years and then leveled off thereafter.

Multivariate logistic regression model number 1

This multivariate logistic regression model incorporated the four parameters found to have the strongest association with cancer/NIFTP risk in univariate analysis other than ThyroSeq v3, which were patient sex, age, Bethesda category, and ATA US patterns (Model No. 1) and yielded a cross-validated C index of 0.653; R^2 was 0.108. The contribution of each parameter to the overall performance of Model number 1 is shown in Table 3 and Figure 2. In this model, patient sex, age, and Bethesda category were significantly associated with thyroid cancer/NIFTP probability ($p < 0.05$ for each). The ATA US patterns demonstrated a positive trend ($p = 0.098$) that was modest in extent.

Male sex (OR 2.57; $p = 0.008$) and Bethesda category V (OR = 4.36; $p = 0.004$) were significantly associated with increased risk of thyroid cancer/NIFTP, and age demonstrated a nonlinear risk profile (OR = 0.66; $p = 0.013$), with cancer/NIFTP risk declining until age 55 years and leveling off thereafter.

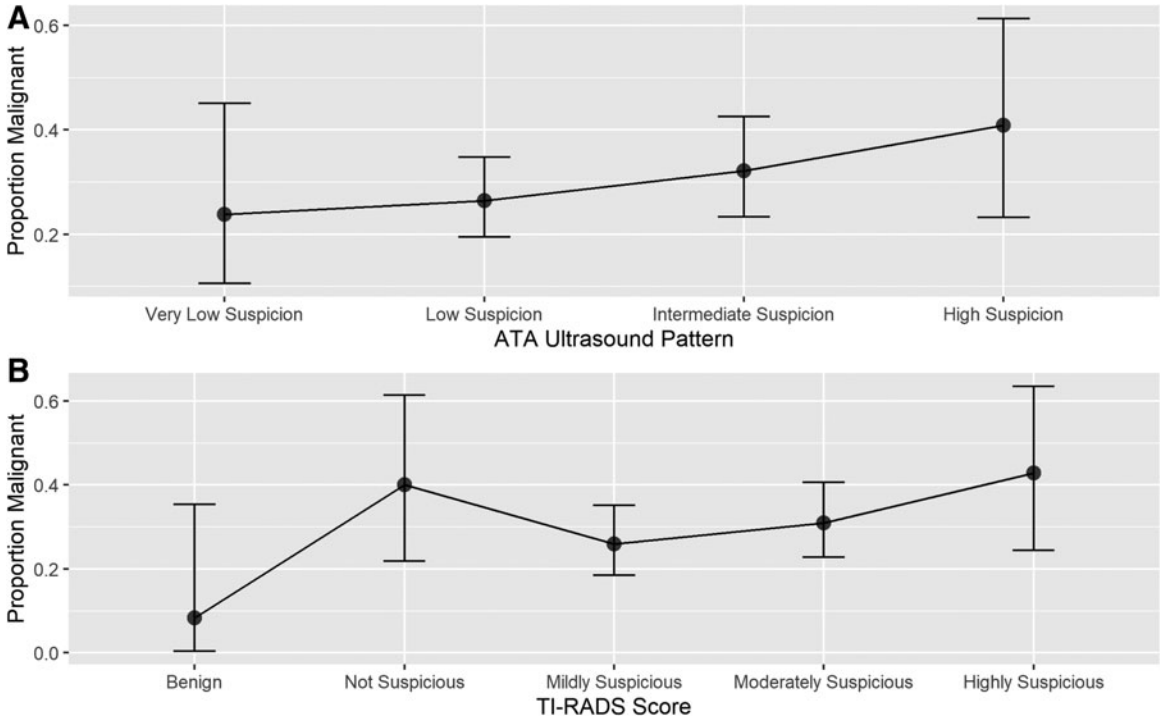

FIG. 1. Proportion of observed malignant nodules or NIFTP with 95% confidence intervals by ATA ultrasound pattern (**A**) and TI-RADS score (**B**). Although the ATA ($p = 0.1211$) and TI-RADS ($p = 0.1359$) ultrasound risk groups demonstrated modest positive trends, neither was significantly associated with cancer/NIFTP probability. However, the ATA ultrasound risk groups demonstrated a more continuous progressively increasing trend, rendering ATA more appropriate than TI-RADS for inclusion in multivariate models. ATA, American Thyroid Association; NIFTP, noninvasive follicular thyroid neoplasm with papillary-like nuclear features; TI-RADS, Thyroid Image Reporting and Data System.

TABLE 3. MULTIVARIATE LOGISTIC REGRESSION MODEL NUMBER 1 (WITHOUT THYROSEQ V3) FOR PREDICTION OF THYROID MALIGNANCY/NONINVASIVE FOLLICULAR THYROID NEOPLASM WITH PAPILLARY-LIKE NUCLEAR FEATURES

Covariate	*Reference*	*Odds ratio [CI]*	p
Gender	Male:female	2.57 [1.28–5.19]	0.0083
Age	62:42	0.66 [0.43–1.01]	0.0127
Bethesda category	V:III	4.36 [1.62–11.73]	0.0036
ATA US patterns	High suspicion:low suspicion	1.92 [0.89–4.14]	0.0981

Model number 1 yielded a cross-validated C index of 0.653; $R^2 = 0.108$.

A potential clinical prediction nomogram based on Model number 1 (without MT) is illustrated in Figure 3. This nomogram incorporates information about patient sex, age, Bethesda category III–V, and the ATA US pattern. Points are accumulated based on the value of each clinical parameter, and the probability of thyroid malignancy/NIFTP is determined by comparing the total points against the probability scale at the bottom of the diagram.

Multivariate logistic regression model number 2

When ThyroSeq v3 was added to Model number 1, the association with thyroid cancer/NIFTP significantly improved as the cross-validated C index increased to 0.888; $R^2 = 0.572$ (Table 4). However, patient age (OR = 0.75; $p = 0.341$), Bethesda category (OR = 2.12; $p = 0.272$), and the ATA US patterns (OR = 1.86; $p = 0.264$) became redundant and had limited incremental information content in comparison with ThyroSeq v3 (OR = 51.94; $p < 0.0001$), which dominated the overall performance of the model. Patient sex was the only parameter showing tendency for significance beyond that of MT (OR = 2.31; $p = 0.095$).

A proposed clinical prediction nomogram based on Model number 2 (with MT) is illustrated in Figure 4. This nomogram incorporates information about patient sex, age, Bethesda category III–V, the ATA US pattern, and the ThyroSeq v3 results (positive/negative).

Multivariate logistic regression model number 3

In this model, covariates were selected from all available candidates (clinical, demographic, molecular), with the goal being to create a model with few parameters but high predictive potential. The two most informative covariates from Model number 2, ThyroSeq v3 and sex, were retained to comprise Model number 3 (Table 5). Model 3 showed better discrimination potential with a cross-validated C index of 0.889 and $R^2 = 0.588$, achieving a higher C index with fewer variables. Sex showed a weak contribution (OR = 2.30; $p = 0.078$), whereas ThyroSeq v3 again dominated the performance of the model (OR = 60.56; $p < 0.0001$). This model provided a highly accurate association with thyroid cancer/NIFTP in nodules with indeterminate cytology using the smallest variable input.

A proposed clinical prediction nomogram based on Model number 3 is illustrated in Figure 5. This nomogram incorporates information about sex and the ThyroSeq v3 results (positive/negative).

Secondary analyses

Despite suggestive upward trends, neither the ATA US risk categories ($p = 0.234$) nor the TI-RADS summation categories ($p = 0.672$) differed in the proportion of nodules with ThyroSeq v3 positivity (Supplementary Fig. S1). In contrast, the Bethesda categories demonstrated

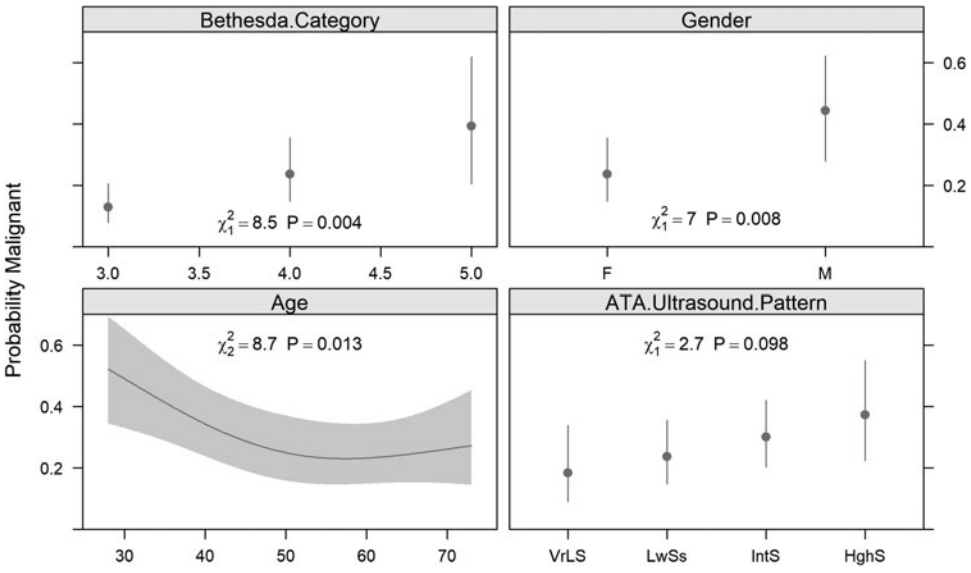

FIG. 2. Effects of model 1 covariates expressed as probability of malignancy/NIFTP. The relative contribution of each of the 4 covariates comprising model 1 (gender, age, Bethesda category, and ATA ultrasound pattern) is expressed as probability of malignancy/NIFTP. For ATA ultrasound pattern (bottom right); HghS, high suspicion; IntS, intermediate suspicion; LwS, low suspicion; VeLS, very low suspicion.

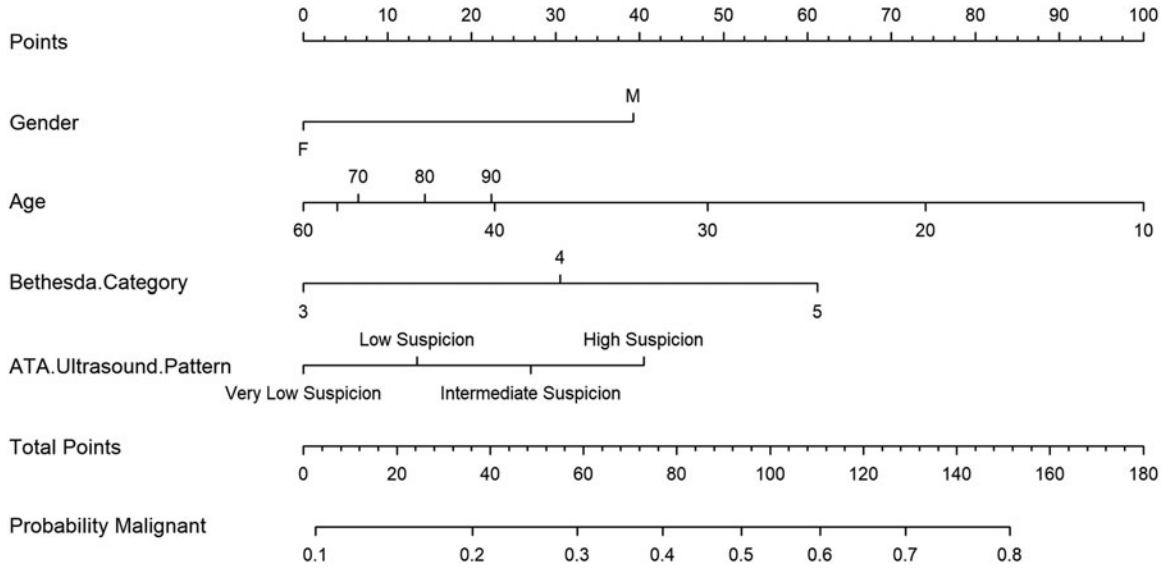

FIG. 3. Clinical prediction nomogram based on model 1. Model 1 uses 4 covariates: gender, age, Bethesda category, and ATA ultrasound pattern, to predict the probability of a malignant nodule or NIFTP. Points are accumulated based on the value of each clinical parameter; the probability of malignancy/NIFTP is determined by comparing the total points against the probability scale at the bottom of the diagram. The nomogram is applicable to thyroid nodules 0.5–7 cm in diameter with indeterminate cytology. The predictions for patient age <18 years are extrapolated; hence, the nomogram is most appropriate for patients age 18 years and older.

a significant positive association with the proportion of positive ThyroSeq v3 determinations ($p = 0.0006$, Supplementary Table S2).

Discussion

Under the clinical care paradigm(s) that are standard today (12), US is typically used to select thyroid nodules for FNA, and when cytology is reported as indeterminate, MT is often used to triage patients to either surgery or surveillance (12,17–24). However, whether clinical characteristics can further refine cancer probability predicted by MT in nodules with indeterminate cytology remains unknown. Furthermore, since MT is not available in certain clinical settings, it is important to determine whether commonly available clinical, cytologic, and/or US findings can augment or substitute for MT in predicting the presence of cancer/NIFTP in cytologically indeterminate thyroid nodules. The results of this study demonstrate that among nodules selected for FNA that yield indeterminate cytology, MT is much more informative than all other examined parameters, which do not contribute to improved cancer/NIFTP risk stratification in such nodules. The only possible exception to this is male sex, which showed a trend toward association with thyroid cancer/NIFTP risk in addition to MT results. Furthermore, the findings of this study suggest that while US patterns are highly informative in selecting nodules for FNA, they are less helpful for further predicting cancer/NIFTP after the nodule was diagnosed as indeterminate on FNA cytology.

Although thyroid cancer occurs more frequently in females, male sex is a known risk factor for thyroid cancer in a given nodule (2,5), but its prognostic significance for thyroid cancer aggressiveness has been questioned (25). In this study, we observed that male sex was strongly associated with cancer/NIFTP on univariate analysis, and on multivariate analysis that included ThyroSeq v3, male sex was the only clinical factor with a tendency to provide additive value. Using Model 3 as an example (nomogram in Fig. 5), the probability of malignancy for a ThyroSeq v3-positive female is about 0.65 and for a ThyroSeq v3-positive male is about 0.8; for a ThyroSeq v3-negative patient, the corresponding probabilities are 0 and 0.1. In short, the observed findings suggest that male sex is associated with a higher risk of cancer/NIFTP irrespective of MT results. Clinically this could be interpreted to prompt closer surveillance for men who do not undergo surgery, and it certainly bears further investigation.

Table 4. Multivariate Logistic Regression Model Number 2 (with ThyroSeq v3) for Prediction of Thyroid Malignancy/Noninvasive Follicular Thyroid Neoplasm with Papillary-like Nuclear Features

Covariate	*Reference*	*Odds ratio [CI]*	p
Gender	Male:female	2.31 [0.86–6.16]	0.0948
Age	62:42	0.75 [0.43–1.32]	0.3407
Bethesda category	V:III	2.12 [0.56–8.05]	0.2720
ATA US patterns	High suspicion:low suspicion	1.86 [0.62–5.52]	0.2641
ThyroSeq v3	Positive:negative	51.94 [19.23–140.3]	<0.0001

Model number 2 yielded a cross-validated C index of 0.888; $R^2 = 0.572$.

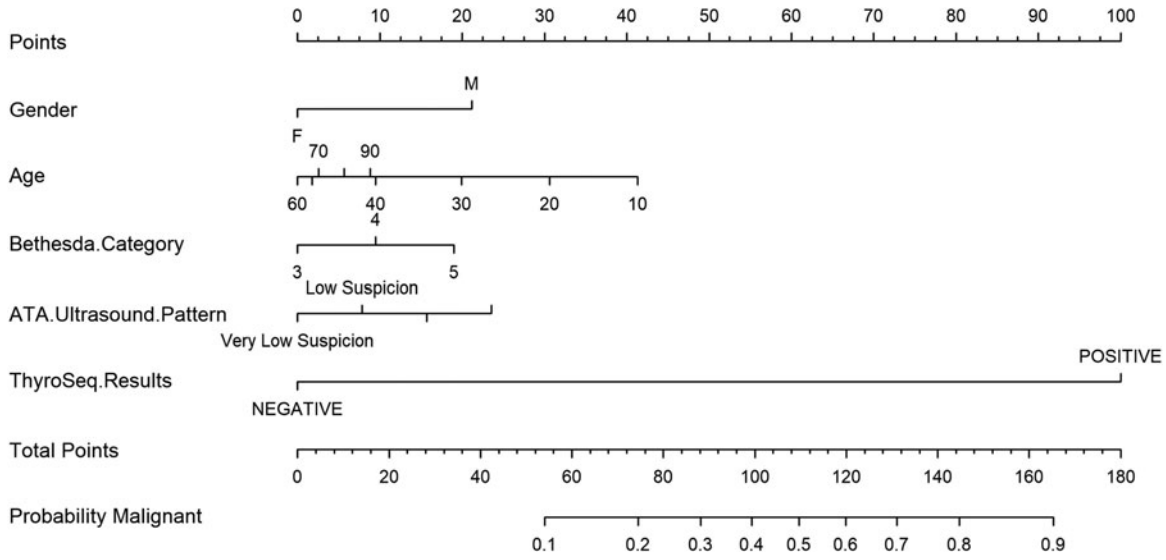

FIG. 4. Clinical prediction nomogram based on model 2. Model 2 reuses the four covariates of model 1: gender, age, Bethesda category, and ATA ultrasound pattern, but adds ThyroSeq v3 results to predict the probability of a malignant nodule or NIFTP. Points are accumulated based on the value of each clinical parameter; the probability of malignancy/NIFTP is determined by comparing the total points against the probability scale at the bottom of the diagram. Note that a positive ThyroSeq v3 result alone contributes 100 points. The nomogram is applicable to thyroid nodules 0.5–7 cm in diameter with indeterminate cytology. The predictions for patient age <18 years are extrapolated; hence, the nomogram is most appropriate for patients age 18 years and older.

In the absence of molecular data, some experts have conjectured that clinical data and US patterns could suffice to determine whether indeterminate thyroid nodules can be monitored or require surgery (26–32). However, analysis (Model No. 1) demonstrates that the four most informative parameters other than MT (patient sex, age, Bethesda category, and ATA US patterns), when combined, are insufficiently accurate to predict malignancy in routine clinical use. The cross-validated C index for Model 1 was only 0.653. It is generally accepted that a model should have a C index of at least 0.700 for clinical use, and a C index of 0.800 or more is preferable. Hence, we cannot recommend Model 1 for clinical use. In contrast, when MT was added to Model 1, the cross-validated C index increased to 0.888 (Model 2).

A recent meta-analysis of US risk stratification systems found that the ACR TI-RADS had a higher performance than other systems in selecting nodules for FNA (33), with a diagnostic OR of 4.9. In comparison, the diagnostic OR of the ATA US patterns was 3.1. A recent retrospective study of 463 Bethesda III and IV nodules with complete data reported significant correlations of ATA US patterns with thyroid cancer/NIFTP risk ($p < 0.001$), with an OR of 5.18 when ATA low- and intermediate-risk groups were pooled and compared with ATA high-risk groups (26). Here, however, we observed that in indeterminate-cytology nodules, the ATA/TI-RADS US patterns, although demonstrating modest positive trends (Fig. 1), were not further associated with malignancy/NIFTP independently of MT. Whether these trends would become statistically significant with a larger sample size would be an important question for further study. Although the ATA system performed slightly better than the ACR TI-RADS in our group of indeterminate cytology nodules, neither factor was observed to be a significant determinant. Similarly, Capezzone *et al.* (34), in a surgical series of 73 indeterminate thyroid nodules, found a higher rate of malignancy in nodules with EU-TIRADS 4–5 US patterns versus nodules with EU-TIRADS 2–3 US scores; however, the difference was not statistically significant ($p = 0.10$). The authors conjectured that the small sample size might have limited the statistical significance of the findings.

Table 5. Multivariate Logistic Regression Model Number 3 (with ThyroSeq v3) for Prediction of Thyroid Malignancy/Noninvasive Follicular Thyroid Neoplasm with Papillary-like Nuclear Features

Covariate	*Reference*	*Odds ratio [(CI]*	p
Gender	Male:female	2.30 [0.91–5.83]	0.0781
ThyroSeq v3	Positive:negative	60.56 [22.56–162.59]	<0.0001

Model number 3 yielded a cross-validated C index of 0.889; $R^2 = 0.588$.

Limitations

This study has several limitations, including its *post hoc* nature; however, the research question was carefully framed before data analysis was undertaken. In addition, although the US images were read by experts at each study center, no central review of the images was performed; hence, the results could be impacted by interobserver variability. We attempted to mitigate classification errors by having the study centers enter the characteristics of each nodule into a systematic data table that accurately assigned a TI-RADS risk score based on the nodule's radiologic characteristics. In addition, as noted above, the lack of

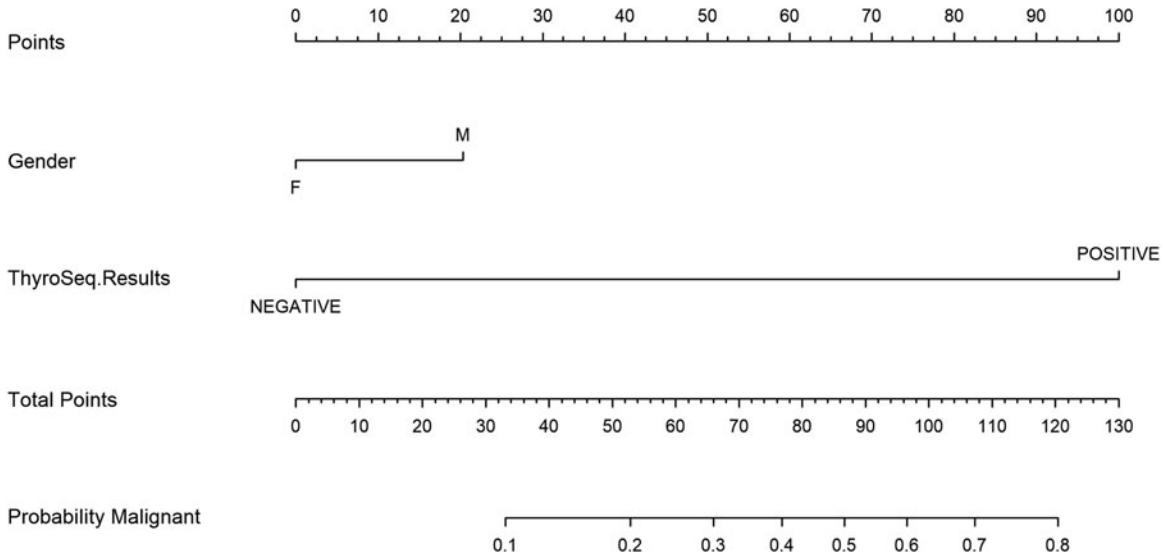

FIG. 5. Clinical prediction nomogram based on model 3. Model 3 uses only 2 covariates, gender and ThyroSeq v3 score, to predict the probability of a malignant nodule or NIFTP. Points are accumulated based on the value of each clinical parameter; the probability of malignancy/NIFTP is determined by comparing the total points against the probability scale at the bottom of the diagram. A positive ThyroSeq v3 score contributes 100 points, male gender about 20. The nomogram is applicable to thyroid nodules 0.5–7 cm in diameter with indeterminate cytology. The nomogram is most appropriate for patients age 18 years and older.

observed significance for male sex may relate to underpowering. An attempt was made to compensate for the small sample size by including parameters that showed promising trends, such as the ATA US patterns—for which the small sample size could have impacted significance, especially for the more extreme US categories for both TI-RADS (TR1, TR2, and TR5) and ATA (very low and high suspicion) classifications. In future studies, it will be important to study the interaction of specific types of molecular alterations with clinical parameters in predicting malignancy/NIFTP; with larger case numbers, the specific types of molecular alterations, such as *RAS* point mutations, could be independently analyzed. Finally, the proposed clinical prediction nomograms presented in this report (Figs. 3–5) should be prospectively validated against independent data in future clinical studies.

Conclusions

In summary, neither of the scoring systems, ATA US or the TI-RADS, was informative for association with thyroid cancer/NIFTP in thyroid nodules after the FNA cytology was read as indeterminate. Although patient age and Bethesda category were associated with thyroid cancer/NIFTP probability on univariate analysis, they had no independent incremental value above the high potential predictive ability of ThyroSeq v3. Generation of several clinical models demonstrated that sex was the only parameter with potential contribution to prediction of thyroid cancer/NIFTP in addition to MT results.

Acknowledgment

The authors are indebted to Michelle S. Keller, MHA, who served as the honest broker for this study.

Authors' Contributions

Study concept and design: J.J.F., W.E.G., Y.E.N., and S.E.C. Statistical analysis: W.E.G. Data interpretation and article writing and editing: all 18 authors.

Author Disclosure Statement

J.A.S.: member of the Data Monitoring Committee of the Medullary Thyroid Cancer Consortium Registry supported by GlaxoSmithKline, Novo Nordisk, Astra Zeneca, and Eli Lilly. Institutional Research funding is supported by Exelixis and Eli Lilly. B.R.H.: Eisai advisory board. Y.E.N.: owns intellectual property and receives royalties related to ThyroSeq from the University of Pittsburgh and serves as consultant for Sonic Healthcare USA. All other authors have no competing financial interests.

Funding Information

This study was supported, in part, by the Head and Neck Cancer SPORE grant P50CA097190 to the University of Pittsburgh. S.E.C. gratefully acknowledges generous support from the William and Susan Johnson Fund for Endocrine Surgery Research.

Supplementary Material

Supplementary Figure S1
Supplementary Table S1
Supplementary Table S2

References

1. Ezzat S, Sarti DA, Cain DR, Braunstein GD 1994 Thyroid incidentalomas: prevalence by palpation and ultrasonography. Arch Intern Med **154:**1838–1840.

2. Burman KD, Wartofsky L 2015 Thyroid nodules. N Engl J Med **373:**2347–2356.
3. Grani G, Sponziello M, Pecce V, Ramundo V, Durante C 2020 Contemporary thyroid nodule evaluation and management. J Clin Endocrinol Metab **105:**2869–2883.
4. Durante C, Grani G, Lamartina L, Filetti S, Mandel SJ, Cooper DS 2018 The diagnosis and management of thyroid nodules: a review. JAMA **319:**914–924.
5. Valderrabano P, McIver P 2017 Evaluation and management of indeterminate thyroid nodules: the revolution of risk stratification beyond cytological diagnosis. Cancer Control **24:**1–14.
6. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, Pacini F, Randolph GW, Sawka AM, Schlumberger M, Schuff KG, Sherman SI, Sosa JA, Steward DL, Tuttle RM, Wartofsky L 2016 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer. Thyroid **26:**1–133.
7. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, Cronan JJ, Beland MD, Desser TS, Frates MC, Hammers LW, Hamper UM, Langer JE, Reading CC, Scoutt LM, Stavros AT 2017 ACR thyroid imaging, reporting and data systems (TI-RADS): white paper of the ACR TI-RADS committee. J Am Coll Radiol **14:**587–595.
8. Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW 2012 The Bethesda system for reporting thyroid cytopathology: a meta-analysis. Acta Cytol **56:**333–339.
9. Manning AM, Yang H, Falciglia M, Mark JR, Steward DL 2017 Thyroid ultrasound-guided fine-needle aspiration cytology: observed increase in indeterminate rate over the past decade. Otolaryngol Head Neck Surg **156:**611–615.
10. Baloch ZW, LiVolsi VA, Asa SL, Rosai J, Merino MJ, Randolph G, Vielh P, DeMay RM, Sidawy MK, Frable WJ 2008 Diagnostic terminology and morphologic criteria for cytologic diagnosis of thyroid lesions: a synopsis of the National Cancer Institute Thyroid Fine-Needle Aspiration State of the Science Conference. Diagn Cytopathol **36:**425–437.
11. Cibas ES, Ali SZ 2017 The 2017 Bethesda system for reporting thyroid cytopathology. Thyroid **27:**1341–1346.
12. Patel KN, Yip L, Lubitz CC, Grubbs EG, Miller BS, Shen W, Angelos P, Chen H, Doherty GM, Fahey TJ, 3rd, Kebebew E, Livolsi VA, Perrier ND, Sipos JA, Sosa JA, Steward D, Tufano RP, McHenry CR, Carty SE 2020 The American Association of Endocrine Surgeons guidelines for the definitive surgical management of thyroid disease in adults. Ann Surg **271:**e21–e93.
13. Nikiforov YE, Seethala RR, Tallini G, Baloch ZW, Basolo F, Thompson LDR, Barletta JA, Wenig BM, Ghuzlan AA, Kakudo K, Giordano TJ, Alves VA, Khanafshar E, Asa SL, El-Naggar AK, Gooding WE, Hodak SP, Lloyd RV, Maytal G, Mete O, Nikiforova MN, Nosé V, Papotti M, Poller DN, Sadow PM, Tischler AS, Tuttle RM, Wall KB, LiVolsi VA, Randolph GW, Ghossein RA 2016 Nomenclature revision for encapsulated follicular variant of papillary thyroid carcinoma: a paradigm shift to reduce overtreatment of indolent tumors. JAMA Oncol **2:**1023–1029.
14. Nikiforova MN, Mercurio S, Wald AI, Barbi de Moura M, Callenberg K, Santana-Santos L, Gooding WE, Yip L, Ferris RL, Nikiforov YE 2018 Analytical performance of the ThyroSeq v3 genomic classifier for cancer diagnosis in thyroid nodules. Cancer **124:**1682–1690.
15. Steward DL, Carty SE, Sippel RS, Yang SP, Sosa JA, Sipos JA, Figge JJ, Mandel S, Haugen BR, Burman KD, Baloch ZW, Lloyd RV, Seethala RR, Gooding WE, Chiosea SI, Gomes-Lima C, Ferris RL, Folek JM, Khawaja RA, Kundra P, Loh KS, Marshall CB, Mayson S, McCoy KL, Nga ME, Ngiam KY, Nikiforova MN, Poehls JL, Ringel MD, Yang H, Yip L, Nikiforov YE 2019 Performance of a multigene genomic classifier in thyroid nodules with indeterminate cytology: a prospective blinded multicenter study. JAMA Oncol **5:**204–212.
16. Desai D, Lepe M, Baloch ZW, Mandel SJ 2021 ThyroSeq v3 for Bethesda III and IV: an institutional experience. Cancer Cytopathol **129:**164–170.
17. Carty SE, Ohori NP, Hilko DA, McCoy KL, French EK, Manroa P, Morariu E, Sridharan S, Seethala RR, Yip L 2020 The clinical utility of molecular testing in the management of thyroid follicular neoplasms (Bethesda IV nodules). Ann Surg **272:**621–627.
18. Livhits MJ, Kuo EJ, Leung AM, Rao J, Levin M, Douek ML, Beckett KR, Zanocco KA, Cheung DS, Gofnung YA, Smooke-Praw S, Yeh MW 2018 Gene expression classifier vs targeted next-generation sequencing in the management of indeterminate thyroid nodules. J Clin Endocrinol Metab **103:**2261–2268.
19. Nicholson KJ, Roberts MS, McCoy KL, Carty SE, Yip L 2019 Molecular testing versus diagnostic lobectomy in Bethesda III/IV thyroid nodules: a cost-effectiveness analysis. Thyroid **29:**1237–1243.
20. Vora A, Holt S, Haque W, Lingvay I 2020 Long-term outcomes of thyroid nodule AFIRMA GEC testing and literature review: an institutional experience. Otolaryngol Head Neck Surg **162:**634–640.
21. Zhu CY, Donangelo I, Gupta D, Nguyen DT, Ochoa JE, Yeh MW, Livhits MJ 2020 Outcomes of indeterminate thyroid nodules managed nonoperatively after molecular testing. J Clin Endocrinol Metab **106:**e1240–e1247.
22. Jug R, Parajuli S, Ahmadi S, Jiang XS 2019 Negative results on thyroid molecular testing decrease rates of surgery for indeterminate thyroid nodules. Endocr Pathol **30:**134–137.
23. Chen T, Gilfix BM, Rivera J, Sadeghi N, Richardson K, Hier MP, Forest VI, Fishman D, Caglar D, Pusztaszeri M, Mitmaker EJ, Payne RJ 2020 The role of the ThyroSeq v3 molecular test in the surgical management of thyroid nodules in the Canadian public health care setting. Thyroid **30:**1280–1287.
24. Ohori NP, Landau MS, Carty SE, Yip L, LeBeau SO, Manroa P, Seethala RR, Schoedel KE, Nikiforova MN, Nikiforov YE 2019 Benign call rate and molecular test result distribution of ThyroSeq v3. Cancer Cytopathol **127:**161–168.
25. Glikson E, Alon E, Bedrin L, Talmi YP 2017 Prognostic factors in differentiated thyroid cancers revisited. Isr Med Assoc J **19:**114–118.
26. Valderrabano P, McGettigan MJ, Lam CA, Khazai L, Thompson ZJ, Chung CH, Centeno BA, McIver B 2018 Thyroid nodules with indeterminate cytology: utility of the American Thyroid Association sonographic patterns for cancer risk stratification. Thyroid **28:**1004–1112.
27. Kim DW, Lee EJ, Lee JH 2011 Role of ultrasound diagnosis in assessing and managing thyroid nodules with inadequate cytology. AJR Am J Roentgenol **197:**1213–1219.

28. Mendez W, Rodgers SE, Lew JI, Montano R, Solorzano CC 2008 Role of surgeon-performed ultrasound in predicting malignancy in patients with indeterminate thyroid nodules. Ann Surg Oncol **15:**2487–2492.
29. Gweon HM, Son EJ, Youk JH, Kim JA 2013 Thyroid nodules with Bethesda system III cytololgy: can ultrasonography guide the next step? Ann Surg Oncol **20:**3083–3088.
30. Yoo WS, Choi HS, Cho SW, Moon JH, Kim KW, Park HJ, Park SY, Choi SI, Choi SH, Lim S, Yi KH, Park DJ, Jang HC, Park YJ 2014 The role of ultrasound findings in the management of thyroid nodules with atypia or follicular lesions of undetermined significance. Clin Endocrinol (Oxf) **80:**735–742.
31. Ahmadi S, Herbst R, Oyekunle T, Jiang XS, Strickland K, Roman S, Sosa JA 2019 Using the ATA and ACR TI-RADS sonographic classifications as adjunctive predictors of malignancy for indeterminate thyroid nodules. Endocr Pract **25:**908–917.
32. Hong HS, Lee JY 2019 Diagnostic performance of ultrasound patterns by K-TIRADS and 2015 ATA guidelines in the risk stratification of thyroid nodules and follicular lesions of unknown significance. AJR Am J Roentgenol **213:**444–450.
33. Castellana M, Castellana C, Treglia G, Giorgino F, Giovanella L, Russ G, Trimboli P 2020 Performance of five ultrasound risk stratification systems in selecting thyroid nodules for FNA. J Clin Endocrinol Metab **105:**1659–1669.
34. Capezzone M, Cantara S, Di Santo A, Sagnella A, Pilli T, Brilli L, Ciuoli C, Maino F, Forleo R, Cartocci A, Castagna MG 2021 The combination of sonographic features and the seven-gene panel may be useful in the management of thyroid nodules with indeterminate cytology. Front Endocrinol (Lausanne) **12:**613727.

Address correspondence to:
James J. Figge, MD, MBA
Diabetes & Endocrine Care
St. Peter's Health Partners/Trinity Health
279 Troy Road
Rensselaer, NY 12144
USA

E-mail: jfiggemd@gmail.com

Sally E. Carty, MD
Division of Endocrine Surgery
University of Pittsburgh
101 Kauffmann, 3471 Fifth Avenue
Pittsburgh, PA 15213
USA

E-mail: cartyse@upmc.edu