



Clinical decision support analysis of a microRNA-based thyroid molecular classifier: A real-world, prospective and multicentre validation study

Marcos Tadeu Santos,^{a,b,*} Bruna Moretto Rodrigues,^a Satye Shizukuda,^a Andrei Félix Oliveira,^a Miriane Oliveira,^a David Livingstone Alves Figueiredo,^c Giuliano Molina Melo,^{d,e} Rubens Adão Silva,^f Claudio Fainstein,^g Gerson Felisbino dos Reis,^h Rossana Corbo,ⁱ Helton Estrela Ramos,^j Cléber Pinto Camacho,^k Fernanda Vaisman,^{i,l} and Mário Vaisman^l

^aResearch and Development (R&D), Onkos Molecular Diagnostics, Ribeirão Preto, SP, Brazil

^bMolecular Oncology Research Centre, Barretos Cancer Hospital, Barretos, SP, Brazil

^cHead and Neck, Midwestern State University (UNICENTRO), Guarapuava, PR, Brazil

^dOtorhinolaryngology, Head and Neck Surgery, Paulista Medical School/UNIFESP, São Paulo, SP, Brazil

^eHead and Neck Surgery, The Portuguese Beneficence of São Paulo (BP), São Paulo, SP, Brazil

^fSurgical Cancerology, Complexo ISPON, Ponta Grossa, PR, Brazil

^gGeneral Surgery, Fluminense Federal University (UFF), Niterói, RJ, Brazil

^hHead and Neck Surgery, University of São Paulo (USP) HC-FMRP, Ribeirão Preto, SP, Brazil

ⁱEndocrinology, National Cancer Institute (INCA), Rio de Janeiro, RJ, Brazil

^jHealth and Sciences, Federal University of Bahia (UFBA), Salvador, BA, Brazil

^kEndocrinology, Paulista Medical School/UNIFESP, São Paulo, SP, Brazil

^lEndocrinology, Medical School, Rio de Janeiro Federal University (UFRJ), Rio de Janeiro, RJ, Brazil

Summary

Background The diagnosis of cancer in Bethesda III/IV thyroid nodules is challenging as fine-needle aspiration (FNA) has limitations, and these cases usually require diagnostic surgery. As approximately 77% of these nodules are not malignant, a diagnostic test accurately identifying benign thyroid nodules can reduce “potentially unnecessary” surgery rates. We have previously reported the development and validation of a microRNA-based thyroid classifier (mir-THYpe) with high sensitivity and specificity, which could be performed directly from FNA smear slides. We sought to evaluate the performance of this test in real-world clinical routine to support clinical decisions and to reduce surgery rates.

Methods We designed a real-world, prospective, multicentre study. Molecular tests were performed with FNA samples prepared at 128 cytopathology laboratories. Patients were followed-up from March 2018 until surgery or until March 2020 (patients with no indication for surgery). The final diagnosis of thyroid tissue samples was retrieved from postsurgical anatomopathological reports.

Findings A total of 435 patients (440 nodules) classified as Bethesda III/IV were followed-up. The rate of avoided surgeries was 52.5% for all surgeries and 74.6% for “potentially unnecessary” surgeries. The test achieved 89.3% sensitivity, 81.65% specificity, 66.2% positive predictive value, and 95% negative predictive value. The test supported 92.3% of clinical decisions.

Interpretation The reported data demonstrate that the use of the microRNA-based classifier in the real-world can reduce the rate of thyroid surgeries with robust performance and support clinical decision-making.

Funding The São Paulo Research-Foundation (FAPESP) and Onkos.

Copyright © 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Keywords: MicroRNA; Indeterminate thyroid nodule; Thyroid cancer; Molecular diagnostics; Molecular classifier; Real-world evidence

eBioMedicine 2022;82:
104137
Published online xxx
<https://doi.org/10.1016/j.ebiom.2022.104137>

*Corresponding author at: R&D Division - Onkos Molecular Diagnostics 1945, Florêncio de Abreu St, Ribeirão Preto/SP14020-028, Brazil.

E-mail address: marcos@onkos.com.br (M.T. Santos).

Research in context

Evidence before this study

We have previously reported the development and validation of a microRNA-based molecular classifier (mir-THYpe) for thyroid nodules with indeterminate cytology with high sensitivity and specificity, which can be performed directly from readily available cytological smear slides without the need for a new dedicated FNA and at a significantly lower cost. Despite the high performance observed in this retrospective study, a prospective approach assessing the test performance in the real clinical practice was needed to validate its true potential to reduce thyroid surgeries and support clinical decisions. We performed a systematic search for studies (by title and abstract) evaluating the performance or the clinical utility of thyroid molecular diagnostic tests that classify indeterminate nodules and included articles or conference abstracts published from January 1, 2000 to January 31, 2022, in English, Portuguese, or Spanish. We used the PubMed and the Scientific Electronic Library Online (SciELO) for articles and world conferences of the American Thyroid Association (ATA), European Thyroid Association (ETA), Latin American Thyroid Society (LATS), and World Congress on Thyroid Cancer (WCTC) for abstracts/posters. Search terms included (with or without the Boolean AND) “thyroid nodule,” “thyroid cancer,” “indeterminate thyroid nodule,” “molecular classifiers,” “gene-expression,” “miRNA OR microRNA,” “DNA mutation(s),” “expression profile,” “real-world,” “clinical utility,” and “clinical validation.” Next, we filtered the validation studies on molecular classifier tests that are currently available for commercial use in the clinical setting published by the providers. We were not able to identify a real-world, prospective, multicentre study on microRNA-based thyroid molecular classifiers, highlighting the clear need for a study evaluating their real impact on supporting clinical decisions and reducing thyroid surgery rates.

Added value of this study

In this real-world evidence, prospective, multicentre study, 435 patients with 440 Bethesda III/IV nodules were tested and followed-up. Clinical samples (FNA smear slides) used for molecular testing were prepared at 128 cytopathology laboratories. The rate of avoided surgeries was 52.5% for all surgeries and 74.6% for “potentially unnecessary” surgeries. The mir-THYpe test achieved 89.3% sensitivity (CI95% 82–94.3), 81.65% specificity (CI95% 76.6–86), 66.2% positive predictive value (CI95% 60.3–71.7), and 95% negative predictive value (CI95% 91.7–97) at 28.7% (CI95% 24.3–33.5) cancer prevalence. The test supported 92.3% of clinical decisions.

Implications of all the available evidence

The reported data demonstrate that the use of a microRNA-based classifier in the real-world can reduce the rates of thyroid surgeries with robust performance comparable to other well-established thyroid molecular

classifiers and support clinical decision-making. Cost-effectiveness analyses of the test’s financial impact and healthcare system savings are under development, and, together with the present data, will help guide new policies.

Introduction

Thyroid cancer has increased in the last few decades, and it ranks as the ninth most-incident cancer worldwide.¹ Although there is consensus that overdiagnosis is the main reason for increased incidence^{2,3} mainly due to incidental thyroid nodules found on imaging screenings performed for reasons other than thyroid disease evaluation,⁴ there has also been a true increase in the occurrence of thyroid cancer.⁵

Fine-needle aspiration (FNA) cytology is the current gold standard for triaging patients with suspicious thyroid nodules detected clinically or on ultrasound (US),^{6–8} and the six-tier Bethesda System for Reporting Thyroid Cytopathology attempts to standardise cytopathologic analysis.⁹

Approximately 64% of histologically diagnosed thyroid cancer nodules were initially classified on FNA cytology as Bethesda V or VI, with a combined risk of malignancy (RoM) of 91% (only 1.1 surgeries required to detect one case of cancer).¹⁰ Although the proportion of histologically diagnosed thyroid cancer nodules initially classified on FNA cytology as “indeterminate” (Bethesda III and IV) is lower (~29%), the combined RoM for these two classes is about 23% only,¹⁰ implying that around 77% of surgeries are “potentially unnecessary” and could have been reconsidered or avoided (4.4 surgeries required to detect one case of cancer).

Molecular testing has emerged as an option to reduce the need for diagnostic surgery in indeterminate thyroid nodules,⁸ being recommended by the American Thyroid Association (ATA)¹¹ and the National Comprehensive Cancer Network (NCCN)¹² guidelines. We have recently reported the development and validation of a new microRNA-based thyroid molecular classifier test for a precision endocrinology (mir-THYpe) with high sensitivity and specificity, which could be performed directly from readily available cytological smear slides at a significantly lower cost, without the need for a new dedicated FNA.^{13,14}

We present data of a real-world, prospective, and multicentre study to evaluate the support of mir-THYpe results on clinical decisions and on reducing surgery rates in real-world clinical settings and to analyse its performance on Bethesda III and IV nodules.

Methods

Study design and patient data

As we conducted a real-world study to prospect evidence of the utility and performance of the test, patients

eligible for this study were consecutive and had a signed medical prescription for the use of a mir-THYpe test performed in clinical routine with at least one thyroid nodule previously biopsied by FNA and classified as Bethesda III or IV. Ultrasound features and cytological analysis were just considered by the physicians who prescribed the test. We did not interfere with patient selection or exclusion. After molecular test results, patients were assigned to one of the two groups: Group A underwent thyroidectomy (tests performed March/2018–March/2020); Group B did not undergo thyroid surgery (tests performed March/2018–October/2019, follow-up in March/2020). Patients initially assigned to Group B subsequently undergoing thyroidectomy were included in Group A.

Patients who underwent thyroid surgery but for whom we could not obtain original postsurgical anatomopathological (AP) reports were excluded. Samples in which the FNA cytology report and other characteristics, such as size and location, did not correlate with the nodule described in the AP report were also excluded from the performance analysis.

Ethics

Patients were 18 years or older, provided written informed consent, and paid out-of-pocket for the test. The study was approved by the institution's Research Ethics Committee and listed under CAEE number 39107520.0.0000.0106.

Samples

Other pathologists did not revise the FNA cytology Bethesda classes and the postsurgical histological classification of resected nodules. The FNA smear slides of the 440 samples (used for the mir-THYpe test – [Figure 1](#), box c) were prepared, and the Bethesda classes were assigned by 128 cytopathology laboratories (Supplementary [Figure S1](#)). All fixation and staining protocols of the FNA cytology smear slides are eligible for the mir-THYpe test. The samples were sent without temperature control (room-temperature). Thyroidectomy AP reports were used as reference standard (gold-standard), when available. The 168 postsurgical AP reports with the histological classification of nodules ([Figure 1](#), boxes h and o) were assigned by 53 pathology laboratories. Although the benign/indolent behaviour of the non-invasive follicular thyroid neoplasms with papillary-like nuclear features (NIFTPs),¹⁵ its diagnosis is established only after applying post-surgery stringent histological criteria.¹⁴ Therefore, we consider NIFTPs as “malignant” tissues and positive tests were considered true positives.

Molecular analysis

The mir-THYpe molecular classifier test was performed according to the previously described method.¹³ A 12th

microRNA (miR-375) was also analysed to screen for medullary thyroid cancer.¹⁶ No mutation or DNA analyses were performed on the samples. Due to its prospective nature, the present study could only be compared with other prospective studies performed by commercial test suppliers.

Statistics

Statistical analyses were performed using the R software, an open-source statistical programming environment. The confidence intervals for sensitivity, specificity, and accuracy were the “exact” Clopper–Pearson confidence intervals, and the confidence intervals for the predictive values were the standard logit confidence intervals.¹⁷ Confidence intervals for the means were calculated using the method described by Walline.¹⁸ The Bayes' theorem used to calculate the negative predictive value (NPV) of non-resected nodules was based on Hall.¹⁹ Consecutive unblinded sampling without randomization was applied in order to meet the inclusion/exclusion criteria described on “Study design and patient data” and the real-world and prospective design aimed for this study.

Role of funders

The funder (Onkos) designed the study with assistance from academic researchers. The funder (Onkos) aided in data collection, analysis, and interpretation, and in report writing. More than one author had full access to all the data from the study and verified the data reported in the manuscript. The corresponding author had final responsibility for the decision to submit for publication and the authors were not precluded from accessing data in the study and accepted responsibility for submitting for publication.

Results

Patient and nodule characteristics

A total of 481 real-world patients with 488 nodules (1.01 nodule/patient) and a medical prescription for the mir-THYpe test provided written informed consent. Of these, 39 nodules (8%) were excluded due to loss to follow-up. Eight nodules (1.6%) classified as Bethesda V and one nodule (0.2%) classified as Bethesda II were also excluded ([Figure 1](#), box b) (only Bethesda III/IV nodules were accepted for this study). The remaining 440 nodules classified as Bethesda III and IV, from 435 patients (1.01 nodules/patient), which fulfilled all initial inclusion criteria described in study design and patient data, were categorised according to the mir-THYpe test results ([Figure 1](#) and [Table 1](#)).

Although the mir-THYpe test was negative for malignancy in 239 samples (Benign Call Rate [BCR] of 54.3%), 17 nodules (7.1%) underwent thyroidectomy

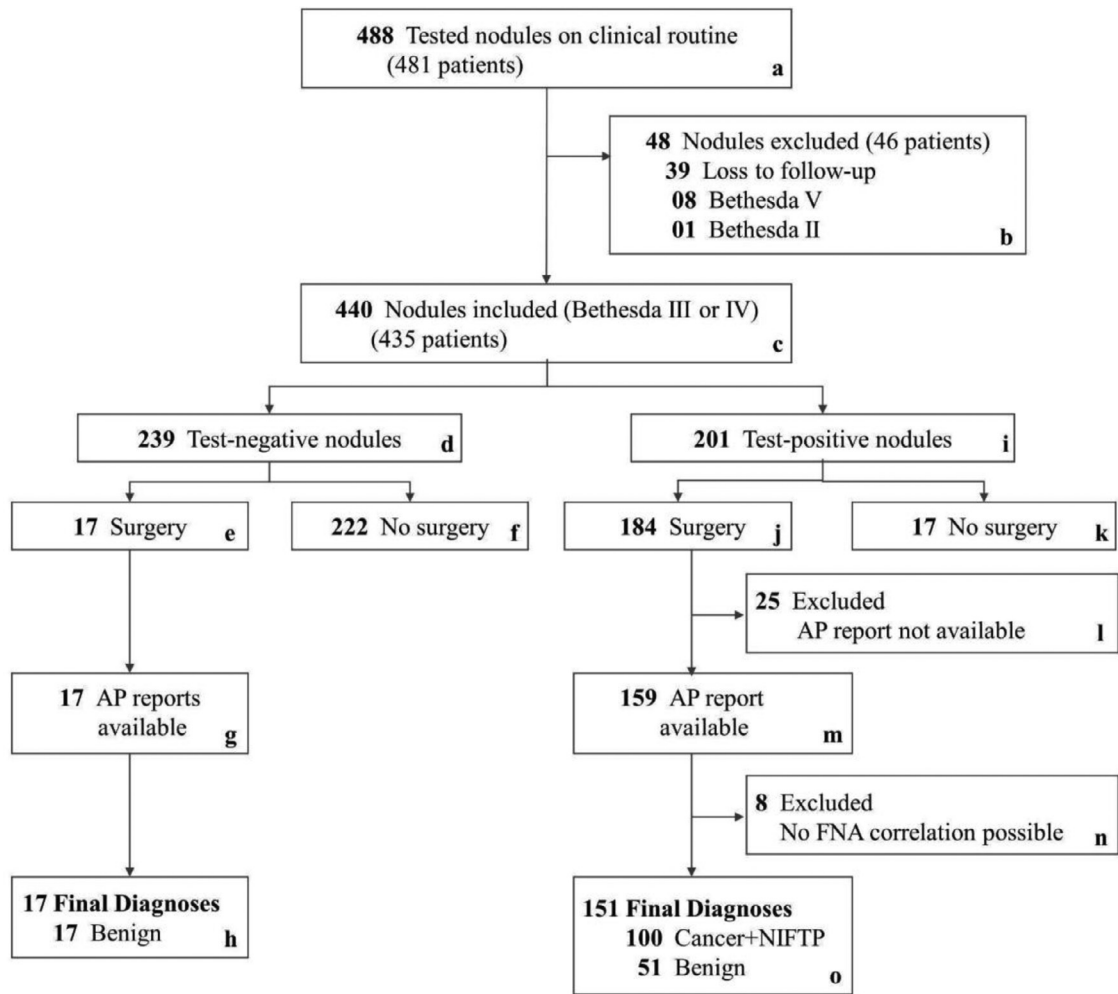


Figure 1. Patients and Nodules Exclusion Flow Chart. AP, anatomopathological; FNA, fine-needle aspiration; NIFTP, non-invasive follicular thyroid neoplasm with papillary-like nuclear features.

(Group A), (nine previously classified as Bethesda III [52.9%], and eight as Bethesda IV [47.1%]). The mean age of these 17 patients was 49 years, the mean nodule size was 1.8 cm, and the mean test-to-surgery time was 170 days; for details, see Supplementary Table S1. The mean age of the other 222 patients (92.9%) (Group B) was 52 years. The mean follow-up time was 412.9 days. Of these, 117 nodules (52.7%) were classified as Bethesda III and 105 (47.3%) as Bethesda IV; Supplementary Table S2 presents the 95% confidence intervals and ranges. Follow-up times are described in Supplementary Table S3.

The mir-THYpe test was positive for malignancy in 201 samples (Positive Call Rate [PCR] of 45.7%), and 184 nodules (91.6%) were surgically treated (Group A). However, 25 nodules (13.6%) were excluded due to lack of access to the original postsurgical AP report, and eight samples (4.3%) were excluded because the punctured nodule described in the FNA cytology did not

correlate with the nodule(s) described in the postsurgical AP report (all AP reports described at least one benign and one malignant nodule and limited information on nodule side, position and size were provided, making the exact correlation impossible). Of the remaining 151 resected nodules (82.1%), 83 were previously classified as Bethesda III (55%) and 68 as Bethesda IV (45%). The mean age of the patients was 49 years, the mean nodule size was 1.13 cm, and the mean test-to-surgery time was 79 days. The mean age of the other 17 patients (8.4%) (Group B) was 54 years, and the mean follow-up time was 359.6 days. Among them, 13 nodules (76.5%) were classified as Bethesda III and four (23.5%) as Bethesda IV; Supplementary Table S2 presents the 95% confidence intervals and ranges. Test-to-surgery and follow-up times are described with periods in Supplementary Table S3.

Overall, the “typical” profile of a real-world patient referred for the mir-THYpe test was women (82%),

Variable	Total	mir-THYpe test result	
		Negative	Positive
Cohort, No (%)			
Nodules Bethesda III/IV	440	239 (54.3)	201 (45.7)
Unique Patients ^a	435	-	-
Sex, No (%)			
Male	79 (18)	50 (63.3)	29 (36.7)
Female	361 (82)	189 (52.4)	172 (47.6)
Age, No (%), y			
20-54	252 (57.3)	124 (49.2)	128 (50.8)
≥ 55-90	188 (42.7)	115 (61.2)	73 (38.8)
Bethesda Classes^b, No (%)			
AUS/FLUS (III)	247 (56.1)	126 (51)	121 (49)
FN/SFN (IV)	193 (43.9)	113 (58.5)	80 (41.5)
Nodule size^c, No (%), cm			
< 2.00	145 (86.3)	12 (70.6)	133 (88.1)
2.00-4.00	21 (12.5)	4 (23.5)	17 (11.3)
> 4.00	2 (1.2)	1 (5.9)	1 (0.7)
Nodule Lobe Position, No (%)			
Left	175 (39.8)	83 (47.4)	92 (52.6)
Right	236 (53.6)	139 (58.9)	97 (41.1)
Isthmus	29 (6.6)	17 (58.6)	12 (41.4)

Table 1: Demographic and clinical characteristics of the study cohort.

^a Five patients with two nodules. One patient with both nodules with negative result. One patient with both nodules with positive result. Three patients with one nodule with negative and the other with a positive result.

^b According to the origin pathology lab (real-world).

^c Nodule sizes were retrieved only by the postsurgical AP reports.

AUS/FLUS, atypia of undetermined significance/follicular lesion of undetermined significance; FN/SFN, follicular or oncocytic (Hürthle cell) neoplasm/suspicious for a follicular or oncocytic (Hürthle cell) neoplasm; No, number of.

younger than 55 years (57.3%), with a Bethesda III (56.1%) thyroid nodule <2 cm (86.3%). [Table 1](#) displays the demographic and clinical characteristics of the study cohort.

Test performance

[Table 2](#) summarises the test performance across all 168 resected nodules according to the histopathological type found post-surgery.

All 17 samples with negative results treated with surgery, regardless of Bethesda class, were confirmed as benign lesions (true negatives) in the postsurgical AP reports, achieving an overall NPV of 100% (one-side CI_{95%} 80.5–100). This result must be interpreted with caution because the sample size was too small (patients with a negative test usually do not receive a surgery recommendation). This result included four Hürthle cell adenomas (23.5% [CI_{95%} 6.8–49.9]).

Of the 151 resected nodules with positive results included in the final analysis, 100 nodules were confirmed as cancer/NIFTP lesions (true positives) in the postsurgical AP report, including nine NIFTPs (9% [CI_{95%} 4.2–16.4]) and one medullary carcinoma (1% [CI_{95%} 0.02–5.5]), achieving an overall positive

predictive value (PPV) of 66.2% (CI_{95%} 60.3–71.7). Splitting the results by cytological class when the mir-THYpe test was positive, 52 of the 83 Bethesda III (62.7%), and 48 of the 68 Bethesda IV nodules (70.6%) were accurately classified. Fifty-one nodules showed false-positive results due to benign tissue identified in the postsurgical AP reports (33.8% [CI_{95%} 26.3–41.9]).

Overall, the test accuracy was 70% (CI_{95%} 62–76.5) as it accurately classified 117 of the 168 resected samples.

Although expected, the small number of resected test-negative nodules impaired the realistic calculation of the sensitivity and NPV (both at 100%). To minimise this effect, we performed a theoretical calculation based on the Bayes' theorem,¹⁹ applying the sensitivity observed during the validation study¹³ (94.6%) to the 222 non-resected nodules with test-negative results ([Figure 1](#), box f), resulting in 210 true negatives and 12 false negatives. In this scenario (base case), the mir-THYpe test achieved 89.3% sensitivity (CI_{95%} 82–94.3), 81.65% specificity (CI_{95%} 76.6–86), 66.2% PPV (CI_{95%} 60.3–71.7), and 95% NPV (CI_{95%} 91.7–97) at a 28.7% (CI_{95%} 24.3–33.5) disease prevalence ([Table 3](#)). Sensitivity analysis in other two scenarios, excluding the 17 test-negative resected-nodules (scenario #1) and analysing all the 239 test-negative nodules (regardless of

mir-THYpe result / Postsurgical tissue	FNA smear slides Bethesda classes		
	AUS/FLUS (III)	FN/SFN (IV)	Total
Test-Negative nodules	9	8	17
Malignant - False Negatives, No (%)	0 (0%)	0 (0%)	0 (0%)
Benign - True Negatives, No (%)	9 (100%)	8 (100%)	17 (100%)
Hürthle cell adenoma	2	2	4
Follicular Adenoma	1	3	4
Adenomatous goitre / follicular hyperplasia	-	2	2
Colloid goitre	5	1	6
Hashimoto's thyroiditis	1	-	1
Test-Positive nodules	83	68	151
Malignant - True Positives, No (%)	52 (62.7)	48 (70.6)	100 (66.2%)
Papillary carcinoma, usual type	23	7	30
Papillary carcinoma, follicular variant	16	30	46
Papillary carcinoma, Hürtle cell variant	2	1	3
NIFTP ^a	6	3	9
Follicular carcinoma, microinvasive	4	5	9
Follicular carcinoma, widely invasive	-	1	1
Follicular carcinoma, Hürtle cell variant	-	1	1
Medullary carcinoma	1	-	1
Benign - False Positives, No (%)	31 (37.3)	20 (29.4)	51 (33.8%)
Hürthle cell adenoma	4	3	7
Follicular Adenoma	10	11	21
Adenomatous goitre / follicular hyperplasia	9	5	14
Colloid goitre	6	1	7
Hashimoto's thyroiditis	2	-	2
Total	92 (54.8%)	76 (45.2%)	168 (100%)

Table 2: Performance of microRNA-based molecular classifier according to histopathological subtype.

^a Considering positive test result for NIFTP as correct classification.

NIFTP, Non-invasive follicular thyroid neoplasm with papillary-like nuclear features; AUS/FLUS, atypia of undetermined significance/follicular lesion of undetermined significance; FN/SFN, follicular or oncocytic (Hürthle cell) neoplasm/suspicious for a follicular or oncocytic (Hürthle cell) neoplasm; No, number of.

Result	Postsurgical tissue class, no		Test performance		
	Cancer+NIFTP	Benign	Analytical parameter	%	CI95%
Positive	100	51	Sensitivity	89.3	82-94.3
			Specificity	81.6	76.6-86
Negative	12^a	210^a +17 (227)^a	NPV	95.0	91.7-97
			PPV	66.2	60.3-71.7

Table 3: Calculated performance of microRNA-based molecular classifier in cytologically indeterminate thyroid nodules.

^a Theoretical values, considering the published sensitivity (13) of 94.6% for the 222 test-negative not surgically resected nodules.

CI, Confidence Interval; NPV/PPV, Negative/Positive Predictive Value; NIFTP, Non-invasive follicular thyroid neoplasm with papillary-like nuclear features; No, number of.

the 17 test-negative resected-nodules surgery results) (scenario #2) demonstrated that the test NPV range from 81.6 to 99.2% at a 95% confidence interval, (See Supplementary Table S5).

Clinical decisions support and surgery reduction rates

The mir-THYpe test supported 93% of clinical decisions when the result was negative for malignancy (222

nodules with no surgery, out of 239 test-negatives). Positive test results supported 91.5% of clinical decisions (184 nodules treated surgically, out of 201 test-positives). Overall, the mir-THYpe test supported 92.3% of clinical decisions (406/440 cases).

To calculate the rates of total avoided surgeries among the 440 eligible patients, we assumed that 423 would have undergone thyroidectomy if no molecular test were available (we subtracted 17 patients not treated

surgically, despite a positive result – Figure 1, box k). Of these, 222 patients avoided surgery (Figure 1, box f), resulting in 52.5% (CI95% 47.6–57.3) of all surgeries.

Following Figure 1, we considered that 280 patients (17 [box e] + 222 [box f] + 51 [box o]) had a benign nodule and would undergo a “potentially unnecessary” surgery if no molecular test were available. As we could not assume that all 222 samples from box f were truly benign, based on the Bayes’ theorem,¹⁹ we applied the sensitivity observed in the validation study¹³ (94.6%) and estimated that around 210 samples were true-negative nodules in this cohort, resulting in a total of 268 patients who would have undergone a “potentially unnecessary” surgery if no molecular test were available. Considering that only 68 patients (17 [box e] + 51 [box o]) were surgically treated, the mir-THYpe test avoided 74.6% (CI95% 69–79.2) of “potentially unnecessary” surgeries.

Discussion

The present study aimed to prospectively evaluate the real-world performance, the support on clinical decisions, and the true potential of a previously described microRNA-based thyroid molecular classifier for avoiding surgeries in a multicentre cohort.¹³

Although randomised controlled trials (RCTs) are considered the gold standard for analysing the efficacy of therapies, in order to associate cause and effect due to control groups, they can be limited to a subset of patients who could not be fully representative of the unselected real-world patients.^{22–24} Real-world evidence studies might better represent routine practice, providing a valuable reflection on the range and distribution of patients observed in clinical practice and including the bias of mis-classification due to physician judgement during clinical care.^{23–25} A real-world evidence study design can tell physicians which results they might experience in practice, providing a clear view of which level of support in clinical decisions the molecular tests can provide.

The analysed specimens belonged to patients for whom a physician considered the molecular testing a clinical indication in real-world, thus avoiding the bias of including retrospective samples based on the Bethesda classification only. Another advantage was that the FNA smear slides were sent at room-temperature. The FNA smear slides were prepared at 128 cytology laboratories, ensuring a highly heterogeneous sample cohort. As summarised in Table 1, patient demographics and nodule characteristics corroborated the expected real-world proportions, as illustrated in the “typical” patient profile described above. Lastly, patient and physician received the molecular test results before deciding the next step (unlike blinded studies in which all nodules are resected, and test support in clinical decisions is unmeasurable). Therefore, we could follow the

patients and observe the real impact of their more-informed clinical decisions and test utility on surgery avoidance. The support of the mir-THYpe test on real-world clinical decision-making (92.3% overall) re-emphasises that the Bethesda categories III and IV present a challenge for physicians to decide the best option for their patients, with 74.5% of “potentially unnecessary” surgeries being avoided, suggesting that the molecular test could minimise this problem and potentially improve the efficiency of the healthcare system.

However, our study had limitations. Due to the real-world study design, we were not able to collect important pre-test nodule risk characteristics, such as ultrasonographic characteristics of the nodules, and, in Bethesda III cases, to identify whether the tested sample was from the first FNA or from the repeat FNA. The lack of these pre-test data limited our conclusions for what nodule risk characteristics might have been preferentially selected by physicians during clinical practice. The multicentre design made it difficult to closely monitor patients as they were not centralised to a few institutions, and there was limited access to other clinical parameters. Ideally, non-operated patients should be followed with US at least to identify possible nodule progression suggestive of management changes. Although the physicians routinely monitored the patients, we had restricted access. Therefore, we had to limit the follow-up monitoring according to whether the patient underwent thyroid surgery or not and we could not access the reasons why the 17 test-negative patients underwent surgery and the 17 test-positive patients were not surgically treated (clinical decision or patients’ choice). We were also unable to collect nodule size of the non-operated patients, limiting the understanding of the profile of patients undergoing the test in real-world. Moreover, we did not review the patients’ FNA or AP slides in order to guarantee the real-world approach, even though we assumed the risk of potential misclassifications at the pathology laboratories. Finally, as most patients who tested negative did not undergo thyroid surgery (222/239, 93%), we did not have the statistical power to evaluate the performance of negative results (mainly sensitivity and NPV). Although all 17 resected nodules with negative results were confirmed benign, the sample size impaired the confidence on 100% sensitivity and NPV. However, the statistical adjustment based on Bayes’ theorem applied (described above on Results) for the 222 test-negative samples to minimise this effect and calculate the test performance was valid as the sensitivity of a diagnostic test is a parameter not influenced by disease prevalence. Thus, in this specific and supposed scenario, the present study achieved an NPV of 95%, which seems realistic (similar to the NPV of 95.9% observed in the previous validation study).¹³ The sensitivity analysis of the data in other two supposed scenarios (See Supplementary Table S5) also suggests a high NPV of the mir-THYpe test in real-world, but not

at 100%, as observed in the small cohort of 17 test-negative resected nodules, all with benign histology.

The statistical parameters used to analyse the performance of a test are crucial in determining practical applications and clinical adoption. To evaluate the real-world performance of the mir-THYpe test, we contextualised it with the results of prospective studies on the latest version of two leading commercially available thyroid molecular classifier tests, the ThyroSeq v3²⁰ and Afirma GSC.²¹ We also compared the molecular tests results with the performance of the FNA alone using data from a meta-analysis.¹⁰

ThyroSeq v3 and Afirma GSC are the two main thyroid molecular classifiers available, widely adopted worldwide with proven clinical impact, but restricted in poor countries due to price and logistics. Hence, we compared the mir-THYpe test performance observed in this study with prospective studies describing the performance of the latest version of these two tests.^{20,21} Based on the true positives for each test, the mir-THYpe achieved similar performance to the ThyroSeq v3 test (62.7% vs. 64%, 70.6% vs. 68.1%, and 66.2% vs. 66% for Bethesda III, IV, and overall, respectively) and better performance than the Afirma GSC in all scenarios (62.7% vs. 44.7%, 70.6% vs. 47.4%, and 66.2% vs. 45.8% for Bethesda III, IV, and overall). In practical and quantitative terms, when either mir-THYpe or ThyroSeq v3 is used alongside FNA indeterminate cytology, around 1.5 surgeries are required to identify one cancer patient. If Afirma GSC is chosen, 2.18 surgeries are required to identify one cancer patient (1.45 more than when using mir-THYpe) and 4.43 surgeries are required with FNA alone (2.94 more than adding mir-THYpe) (See Supplementary Table S4). Based on the NPV of each test, the mir-THYpe achieved similar performance (95%) to ThyroSeq v3 (97%) and Afirma GSC (96%). Moreover, the sample type requested to perform the test is also relevant, because if no material was stored from the first FNA (frozen), ThyroSeq v3 and Afirma GSC need a new and dedicated FNA. Furthermore, delicate logistics operations are needed as samples must be refrigerated and delivered on time. Although ThyroSeq v3 is under validation for FNA smear slides, the available results are limited to only 31 samples, with different performance depending on slide staining (Papanicolaou, 100%; Diff-Quick, 65%; and overall, 81%).²⁶ Furthermore, FNA smear slides were not used for the prospective study validation.²⁰

Conclusion

The reported data demonstrate over a heterogeneous and large cohort of patients that the real-world use of the mir-THYpe test can reduce surgery rates for Bethesda III and IV indeterminate cytology nodules and support clinical decisions. Moreover, the real-world test performance is robust and comparable to other

well-established thyroid molecular classifiers. Furthermore, it can be performed using already available cytology smear slides at a significantly lower cost. Cost-effectiveness analyses of the test's financial impact and healthcare system savings are under development, and together with the present data, will help guide new policies.

Contributors

MTS designed and supervised the study. CPC, DLAF, GMM, RAS, CF, GFR, RC, HER, FV and MV contributed with data and medical expertise. MTS, BMR, SS, AFO and MO contributed with data collection, verification, integration, and patients' follow-up. MTS drafted the manuscript. MTS and BMR contributed as molecular biologists. CPC contributed with statistical analysis. All authors read and approved the final version.

Data sharing statement

Due to the nature of this research, to protect the participants' privacy and confidentiality, raw data are not shared publicly. Relevant data supporting the findings and conclusions of this study are available within the article and/or supplementary materials. Upon a justifiable request, the share of de-identified data should be approved by the board of an investigational ethics committee.

Declaration of interests

MTS holds equity at Onkos Molecular Diagnostics. BMR, SS, AFO and MO are formal employees at Onkos Molecular Diagnostics. All other authors declare no competing interests.

Acknowledgements

We are grateful to the SUPERA Technology and Innovation Park team for institutional support. We also thank The São Paulo Research-Foundation (FAPESP - Project 2017/16417-9 and 2021/06579-7). Finally, and most importantly, the authors would like to thank all the patients who made this study possible.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.ebiom.2022.104137](https://doi.org/10.1016/j.ebiom.2022.104137).

References

- 1 Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68:394-424.

- 2 Welch HG, Doherty GM. Saving thyroids - overtreatment of small papillary cancers. *N Engl J Med.* 2018;379:310–312.
- 3 Vaccarella S, Franceschi S, Bray F, et al. Worldwide thyroid-cancer epidemic? The increasing impact of overdiagnosis. *N Engl J Med.* 2016;375:614–617.
- 4 Fisher SB, Perrier ND. The incidental thyroid nodule. *CA Cancer J Clin.* 2018;68:97–105.
- 5 Lim H, Devesa SS, Sosa JA, et al. Trends in thyroid cancer Incidence and mortality in the United States, 1974–2013. *JAMA.* 2017;317:1338–1348.
- 6 Chen T, Gilfix BM, Rivera J, et al. The role of the ThyroSeq v3 molecular test in the surgical management of thyroid nodules in the Canadian public health care setting. *Thyroid.* 2020;30:1280–1287.
- 7 Lyu YJ, Shen F, Yan Y, et al. Ultrasound-guided fine-needle aspiration biopsy of thyroid nodules <10 mm in the maximum diameter: does size matter? *Cancer Manag Res.* 2019;11:1231–1236.
- 8 Grani G, Sponziello M, Pecce V, et al. Contemporary thyroid nodule evaluation and management. *J Clin Endocrinol Metab.* 2020;105:2869–2883.
- 9 Cibas ES, Ali SZ. The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid.* 2017;27:1341–1346.
- 10 Bongiovanni M, Spitale A, Faquin WC, et al. The Bethesda system for reporting thyroid cytopathology: a meta-analysis. *Acta Cytol.* 2012;56:333–339.
- 11 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;26:1–133.
- 12 National Comprehensive Cancer Network. Thyroid Cancer (Version 3.2021). https://www.nccn.org/professionals/physician_gls/pdf/thyroid.pdf. Accessed 4 February 2022.
- 13 Santos MT, Buzolin AL, Gama RR, et al. Molecular classification of thyroid nodules with indeterminate cytology: development and validation of a highly sensitive and specific new miRNA-based classifier test using fine-needle aspiration smear slides. *Thyroid.* 2018;28:1618–1626.
- 14 Fish S. New miRNA-based classifier test uses FNA cytology to evaluate cytologically indeterminate thyroid nodules. *Clin Thyroidol.* 2019;31:23–26.
- 15 Canini V, Leni D, Pincelli AI, et al. Clinical-pathological issues in thyroid pathology: study on the routine application of NIFTP diagnostic criteria. *Sci Rep.* 2019;9:13179.
- 16 Santos MT, Oliveira AF, Rodrigues BM, et al. Preoperative early identification and exclusion of medullary thyroid carcinoma: development and validation of a highly sensitive and specific micro-RNA-based molecular classifier test using fine-needle aspiration smear slides. *Arch Endocrinol Metab.* 2021;65(suppl 3):S74. (Abstract 104531).
- 17 Mercaldo ND, Lau KF, Zhou XH. Confidence intervals for predictive values with an emphasis to case-control studies. *Stat Med.* 2007;26:2170–2183.
- 18 Hulley SB, Cummings SR, Browner WS, et al. *Designing Clinical Research: An Epidemiologic Approach.* 4th Ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2013.
- 19 Hall GH. The clinical application of the Bayes' theorem. *Lancet.* 1967;290:555–557.
- 20 Steward DL, Carty SE, Sippel RS, et al. Performance of a multigene genomic classifier in thyroid nodules with indeterminate cytology: a prospective blinded multicenter study. *JAMA Oncol.* 2019;5:204–212.
- 21 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;153:817–824.
- 22 Revicki DA, Frank L. Pharmacoeconomic evaluation in the real world. Effectiveness versus efficacy studies. *Pharmacoeconomics.* 1999;15:423–434.
- 23 Garrison LP, Neumann PJ, Erickson P, et al. Using real-world data for coverage and payment decisions: the ISPOR real-world data task force report. *Value Health.* 2007;10:326–335.
- 24 Association of the British Pharmaceutical Industry (ABPI). *The Vision for Real World Data – Harnessing the Opportunities in the UK. Demonstrating Value With Real World Data.* London, UK: Association of the British Pharmaceutical Industry; 2011.
- 25 Association of the British Pharmaceutical Industry (ABPI). *Demonstrating Value with Real World Data: A Practical Guide.* London, UK: Association of the British Pharmaceutical Industry; 2011.
- 26 Nikiforova MN, Lepe M, Tolino LA, et al. Thyroid cytology smear slides: an untapped resource for ThyroSeq testing. *Cancer Cytopathol.* 2021;129:33–42.