Endocrine Pathology (2019) 30:8–15
https://doi.org/10.1007/s12022-018-9560-5

Evaluation of *167 Gene Expression Classifier* (*GEC*) and *ThyroSeq v2* Diagnostic Accuracy in the Preoperative Assessment of Indeterminate Thyroid Nodules: Bivariate/HROC Meta-analysis

Martyna Borowczyk[1] · Ewelina Szczepanek-Parulska[1] · Michał Olejarz[1] · Barbara Więckowska[2] · Frederik A. Verburg[3] · Szymon Dębicki[1] · Bartłomiej Budny[1] · Małgorzata Janicka-Jedyńska[4] · Katarzyna Ziemnicka[1] · Marek Ruchała[1]

Published online: 27 December 2018

Abstract

The objective of this meta-analysis was to evaluate the performance of the *Gene Expression Classifier* (*GEC*) and *ThyroSeq v2* (*ThyroSeq*) in the preoperative diagnosis of thyroid nodules with indeterminate fine-needle aspiration biopsy results. We searched literature databases from January 2001 to April 2018. The bivariate model analysis was performed to estimate pooled sensitivity, specificity, positive likelihood ratio (LR+), negative likelihood ratio (LR−), positive predictive value (PPV), and negative predictive value (NPV). Pooled data from 1086 nodules with histopathologic confirmation from 16 *GEC* studies enabled calculation of diagnostic parameters (95% confidence interval): sensitivity 98% (96–99%), specificity 12% (8–20%), PPV 45% (37–53%), and NPV 91% (85–96%). Pooled data from five *ThyroSeq* studies assessing 459 nodules showed sensitivity of 84% (74–91%), specificity 78% (50–92%), PPV 58% (31–81%), and NPV 93% (89–97%). When both tools were compared, *GEC* had a significantly higher sensitivity ($p = 0.003$), while *ThyroSeq* had a significantly higher specificity ($p < 0.001$) and accuracy ($p = 0.015$). Pooled LR+ was higher for *ThyroSeq*: 3.79 (1.40–10.27) vs. 1.12 (1.05–1.20). Pooled LR− was higher for *GEC*, 0.20 (0.10–0.39) vs. 0.13 (0.05–0.31). The bivariate summary estimates of sensitivity and specificity for *GEC* and *ThyroSeq* and their pooled accuracy showed a superiority of the *ThyroSeq* test. The *GEC* with a high sensitivity and NPV may be helpful in ruling out malignancy in cases of indeterminate thyroid nodule cytology. *ThyroSeq* has a significantly higher specificity and accuracy with an acceptable sensitivity so that it has the potential for use as an all-round test of malignancy of thyroid nodules.

Keywords Follicular thyroid cancer · Follicular thyroid adenoma · Next-generation sequencing · Indeterminate cytology · Molecular diagnosis · Thyroid nodules · Genetics

Introduction

Up to 35% of fine-needle biopsy (FNB) procedures on thyroid nodules produce an indeterminate result [1, 2], which include Bethesda classification categories III and IV: atypia of undetermined significance/follicular lesion of undetermined significance (AUS/FLUS) or follicular neoplasm/suspicious for follicular neoplasm (FN/SFN) [2]. Most patients with indeterminate FNB results, especially those with malignant features in the USA [3] are referred for surgery. However, the malignancy rate in indeterminate nodules is reported between 15 and 54% [1, 2]. This means that many patients with indeterminate cytological results are exposed to potentially unnecessary surgery with the associated increased risks of mortality and complications.

The 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer recommends the use of molecular testing to support malignancy risk assessment in thyroid nodules with indeterminate cytology results [3]. The two most common genetic tests used as preoperative molecular markers

✉ Martyna Borowczyk
martyna.borowczyk@gmail.com

[1] Department of Endocrinology, Metabolism and Internal Medicine, Poznań University of Medical Sciences, 49, Przybyszewskiego Street, 60-355 Poznan, Poland

[2] Department of Computer Science and Statistics, Poznań University of Medical Sciences, Poznań, Poland

[3] Department of Nuclear Medicine, University Hospital Marburg, Marburg, Germany

[4] Department of Clinical Pathology, Poznań University of Medical Sciences, Poznań, Poland

in this context are *ThyroSeq*, *version 2* (*ThyroSeq*), a seven-gene panel of genetic mutations and rearrangements [4], and a gene expression classifier (GEC) testing mRNA expression of 167 genes [5]. Several studies have assessed these panels with various degrees of outcome blinding. NPV and PPV appear strictly linked to the pretest prevalence of malignancy in the population being studied [3]. Furthermore, there are considerable discrepancies in FNB categories of nodules assessed by genetic testing between studies. Moreover, only for a part of results, a post-surgical follow-up which allows the assessment of the test reliability by comparing to histopathological results is available. Therefore, a wide inter-institutional variation in performance of a molecular classifier for indeterminate thyroid nodules may be observed depending on an institution's practice and patients' characteristics [6]. As a consequence, the results of single-center prospective trials may not be generalized.

In order to bring more reliable data on real-world performance of the aforementioned tests, the aim of the present study was to perform a meta-analysis and to this end assess aggregate study data on the performance of the *167 Gene Expression Classifier* (*GEC*) and the *2 ThyroSeq* test for the preoperative assessment of thyroid nodules with indeterminate FNB result.

Materials and Methods

Search Strategy

Our search strategy followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines [7]. We searched PubMed/MEDLINE, Cochrane Library, Scopus, CINAHL, Academic Search Complete, Web of Knowledge, PubMed Central, PubMed Central Canada, and Clinical Key databases from January 2001 up to April 2018 to find all relevant, full-text journal articles written in English. The search strategy included Medical Subject Headings terms and keywords: "thyroid and ("follicular cancer" or "follicular carcinoma" or "follicular neoplasm" or "follicular adenoma" or "follicular nodule") and ("genetic" or "genetics" or "molecular" or "mutation" or "gene" or "genes") or ("ThyroSeq v2") or ("thyroid" and "Gene Expression Classifier")". Reference lists of all the selected articles, previous reviews, and meta-analyses were hand-searched for any additional articles.

Data Extraction

Two authors (MB and MO) independently selected trials which fulfilled the inclusion criteria and extracted data for the outcomes using a standardized data extraction form. Relevant data included articles comparing the number of malignant and benign nodules with the conclusive histopathological result in a group of benign and suspicious changes according to GEC, previously assessed as Bethesda categories III and IV in thyroid FNB or positive changes according to *ThyroSeq* were included in the meta-analysis. We excluded cases without a post-surgical diagnosis, as well as papers assessing only the AUS category (without FLUS).

Another author (ESP) rechecked the extracted data. We used data from the analysis of available papers or the accompanying illustrations.

Assessment of Methodological Quality

The risk of bias in the included studies was independently assessed by two authors (MB and MO), in accordance with the Cochrane risk of bias tool [8]. It assessed factors such as sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, and selective outcome reporting. As recommended for diagnostic accuracy test studies, the revised Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool was also used [9] to cover patient selection, reference standard, index test, and flow of patients through the study and timing of the index test(s), as well as the reference standard ("flow and timing") alike to assess each domain in terms of the risk of bias and the concerns regarding applicability. A third author (ESP) was the adjudicator when no consensus was achieved. All included studies were assessed using the Newcastle-Ottawa Scale [10]. Studies with a result of seven stars or more were included.

Statistical Analysis

p values < 0.05 were considered to indicate statistical significance.

Authors referred to the Cochrane Handbook for Systematic Reviews of Interventions [8] and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. For every study, the number of true positive (TP), true negative (TN), true negative (TN), and false negative (FN) results was noted. Univariate estimation of sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV) with 95% confidence intervals based on the exact binomial Clopper-Pearson method were calculated and presented by forest plots. In addition, we summarized the obtained results by determining the proportions describing pooled accuracy, assuming the lack of asymmetry according to Egger's test.

The meta-analysis focused on two hierarchical logistic regression models currently recommended by the Cochrane Collaboration [11]: the bivariate model and the hierarchical summary ROC (HSROC) model. Bivariate meta-analyses jointly model both sensitivity and specificity.

Based on these, the negative likelihood ratio (LR−), positive likelihood ratio (LR+), and diagnostic odds ratio (DOR) were calculated. Receiver operating characteristic (ROC) curve and estimated summary parameters were presented by HSROC. Heterogeneity, which comes from the difference between the results achieved by comparable diagnostic methods, was investigated by adding methods of preoperative assessment of indeterminate thyroid nodules (*GEC* and *ThyroSeq*) as a covariate to the bivariate meta-regression hierarchical model. Using this model, we compared summary estimates of sensitivity and specificity. Summary estimation of combined NPV and PPV was given on the base on the same bivariate meta-analysis model that was used for sensitivity and specificity [12].

The bivariate/HSROC method was performed using *STATA* version 14.2 (*StataCorp LP*, USA). To add covariates and compare *GEC* and *ThyroSeq* in meta-regression of the diagnostic accuracy model, we used *R CRAN*, version 3.4.2.

Results

After a complete systematic review was performed, 16 studies met the inclusion criteria. The search results and steps of selection are shown in the flowchart (Fig. 1).

The specificity for individual studies based on the *GEC* varied from 0 to 52%, and the sensitivity from 83 to 100%. For nodules screened with the *ThyroSeq* method, specificity varied from 45 to 96%, while sensitivity ranged from 70 to 91% (Tables 1 and 2).

Table 3 presents the pooled estimation of sensitivity, specificity, LR+, LR−, and DOR obtained from the bivariate model for both methods.

The *GEC* studies are characterized by a significantly higher sensitivity (98%, 95%CI 96–99% vs. 84%, 95%CI 74–91%, $p = 0.003$). In contrast, a significantly higher specificity was found for the *ThyroSeq* method (78%, 95%CI 50–92% vs. 12%, 95%CI 8–20%, $p < 0.001$). Moreover, the combined specificity and sensitivity, as well as the calculated LR+, confirm the advantage of the *ThyroSeq* method. For *ThyroSeq*, the chance that a positive test result will be given to a patient with malignancy is almost four times higher (LR+ = 3.79, 95%CI 1.40–10.27) when we compare it with the chance of a positive test result in healthy people. For the *GEC* method, the same proportion was much lower and amounted to 1.12 (95%CI 1.05–1.20). However, the proportion of patients with malignancy who are correctly identified was higher ($p = 0.003$) for the *GEC* method (sensitivity GE 0.98, 95%CI 0.96–0.99) than for *ThyroSeq* (0.84, 95%CI 0.74–0.91). The value of LR− was similar for both methods and DOR was slightly higher for *ThyroSeq*. Patients with malignancy are almost 19 times more likely to have a positive test with the *ThyroSeq* than disease-free individuals (DOR = 19.86), while the diagnostic odds ratio value observed for *GEC* is lower (DOR = 8.87).

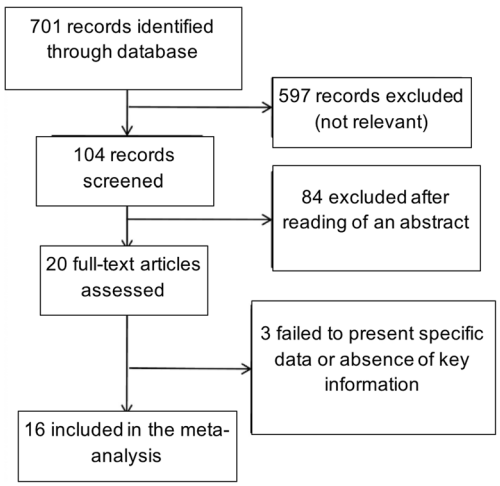


Fig. 1 Methodological flow diagram of the meta-analysis for the *Gene Expression Classifier* test

Figure 2 shows the bivariate summary estimates of sensitivity and specificity for both the *GEC* and the *ThyroSeq* malignancy screening method, together with their corresponding 95% credibility ellipses represented in the ROC space. The ellipses indicate the area that may contain the actual values of the average test sensitivity and specificity for each screening method. We observe a clear difference between the sensitivity and the specificity of the *ThyroSeq* compared with GEC algorithm. The closer the curve is to the upper left corner of the plot (summary value of sensitivity and specificity is both close to 100%), the more accurate is a diagnostic test. In our study, summary point of specificity and sensitivity for *ThyroSeq* is localized much closer to this point of maximal specificity and sensitivity than the point evaluating GEC.

The PPV for individual studies based on the GEC diagnostic varied from 0.33 to 0.86, and the NPV between 0.00 and 1.00. For nodules screened with the *ThyroSeq* method, the PPV varied from 0.27 to 0.89, and the NPV from 0.88 to 0.97 (Table 2). PPV and NPV were compared using bivariate meta-analysis in Table 4 and did not differ significantly, 0.45 vs. 0.58 and 0.91 vs. 0.93, respectively, for *GEC* and *ThyroSeq* (p values > 0.05 in both cases). However, pooled accuracy (Table 5) was significantly higher for *ThyroSeq* ($p = 0.015$). The significance of the latter comparison increased after exclusion of the results influencing asymmetry according to Egger's test ($p < 0.0001$) [31].

Discussion

The present study to the best of our knowledge is the first one to provide (a) an aggregate analysis of performance data of two different genetic analysis–based methods for the assessment of malignancy of thyroid nodules with an indeterminate cytology

Table 1 The number of patients with false negative (FN), false positive (FP), true negative (TN), and true positive (TP) results with forest plots presenting the punctual estimates of sensitivity and specificity and 95% confidence intervals of each study across two diagnostic tests: *GEC* and *ThyroSeq*

GEC (1086 nodules)

Study	Year of publication	TP	FP	FN	TN	Sensitivity (95%CI)	Specificity (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)
Alexander et al.[5]	2012	46	77	5	82	0.90 (0.79-0.97)	0.52 (0.44-0.60)		
Harrell et al.[13]	2013	21	9	2	3	0.91 (0.72-0.99)	0.25 (0.05-0.57)		
Alexander et. al.[14]	2014	47	66	0	3	1.00 (0.92-1.00)	0.04 (0.01-0.12)		
Han et al.[15]	2014	9	11	0	0	1.00 (0.66-1.00)	0.00 (0.00-0.28)		
Lastra et al.[16]	2014	22	26	0	2	1.00 (0.85-1.00)	0.07 (0.01-0.24)		
McIver et al.[17]	2014	5	27	1	3	0.83 (0.36-1.00)	0.10 (0.02-0.27)		
Celik et al.[18]	2015	10	7	0	2	1.00 (0.69-1.00)	0.22 (0.03-0.60)		
Noureldine et al.[19]	2015	72	99	2	10	0.97 (0.91-1.00)	0.09 (0.04-0.16)		
Yang et al.[20]	2015	39	30	0	5	1.00 (0.91-1.00)	0.14 (0.05-0.30)		
Zhu et al.[21]	2015	6	4	0	0	1.00 (0.54-1.00)	0.00 (0.00-0.60)		
Abeykoon et al.[22]	2016	12	2	0	0	1.00 (0.74-1.00)	0.00 (0.00-0.84)		
Chaudhary et al.[23]	2016	28	45	0	8	1.00 (0.88-1.00)	0.15 (0.07-0.28)		
Marti et al.[6]	2016	27	36	0	7	1.00 (0.87-1.00)	0.16 (0.07-0.31)		
Sacks et al.[24]	2016	18	36	0	4	1.00 (0.81-1.00)	0.10 (0.03-0.24)		
Witt et al.[25]	2016	6	9	0	0	1.00 (0.54-1.00)	0.00 (0.00-0.34)		
Wu et al.[26]	2016	50	31	2	12	0.96 (0.87-1.00)	0.28 (0.15-0.44)		
Total		418	515	12	141			0 0.2 0.4 0.6 0.8 1	0 0.2 0.4 0.6 0.8 1

ThyroSeq (459 nodules)

Study	Year of publication	TP	FP	FN	TN	Sensitivity (95%CI)	Specificity (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)
Nikiforov et al. Nikiforov et al.[27+28]	2015	55	7	6	173	0.90 (0.80-0.96)	0.96 (0.92-0.98)		
Shreshta et al.[29]	2016	19	11	3	23	0.86 (0.65-0.97)	0.68 (0.49-0.83)		
Valderrabano et al.[30]	2017	14	19	6	63	0.70 (0.46-0.88)	0.77 (0.66-0.85)		
Taye et al.[31]	2018	10	27	1	22	0.91 (0.59-1.00)	0.45 (0.31-0.60)		
Total		98	64	16	281			0 0.2 0.4 0.6 0.8 1	0 0.2 0.4 0.6 0.8 1

and (b) a direct comparison of key performance measures of these tests. Although both tests have their advantages, the aggregate performance indicators on balance seem to indicate that the *ThyroSeq v2* may be preferable for clinical practice.

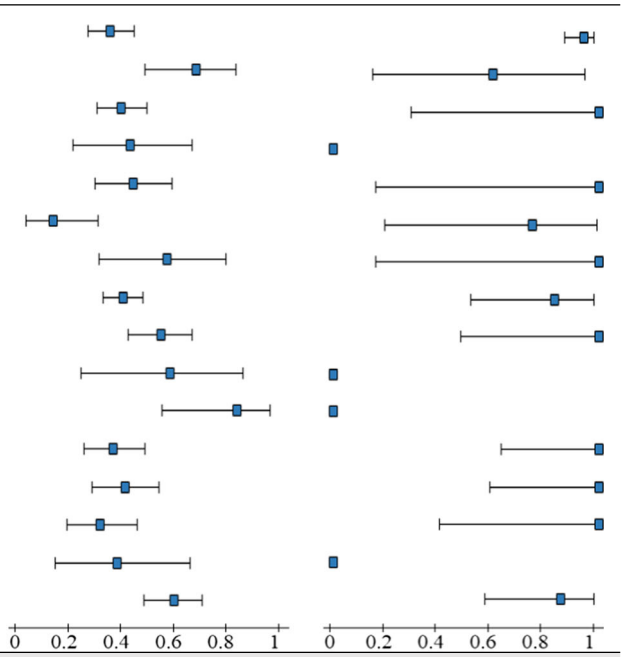
Many molecular markers have been proposed to improve FNB-based diagnosis of thyroid nodules; however, only a small minority demonstrated the potential to modify clinical decision-making and thus impact care [14]. The ATA guidelines' authors suggest that an ideal "rule-in" and "rule-out" test for the assessment of malignancy of cytologically indeterminate nodules should have a PPV for histopathologically proven malignancy similar to a malignant cytological diagnosis (98.6%) and an NPV similar to a benign cytological diagnosis (96.3%) [3] with estimates based on a recent meta-analysis of performance of the Bethesda system [1]. Consequently, the "ideal rule-out test" would have a high sensitivity and high NPV, whereas the "ideal rule-in test" would have a high specificity and PPV [16].

Alexander et al. demonstrated that the *GEC* result of a nodule classified as "benign" had an NPV of 95% for nodules previously categorized cytologically as AUS-FLUS and an NPV of 94% for aspirates that were classified as FN/SFN. In the first clinical study of preoperative use of *GEC*, sensitivity with regard to malignancy for the classification of indeterminate nodules was 90% [5]. The advantages of the test were therefore both a high NPV and a high sensitivity. This result

Table 2 The results of positive predictive value (PPV) and negative predictive value (NPV) and forest plots presenting the punctual estimates with 95% confidence intervals of each study across two diagnostic tests: *GEC* and *ThyroSeq*

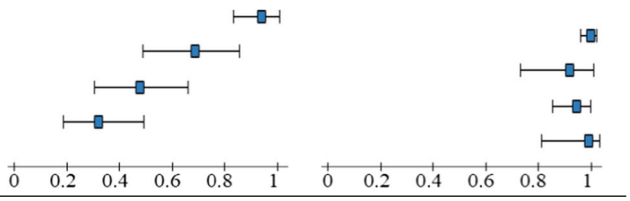
GEC

Study	Year of publication	PPV (95%CI)	NPV (95%CI)	PPV (95%CI)	NPV (95%CI)
Alexander et al.[5]	2012	0.37 (0.29-0.47)	0.94 (0.87-0.98)		
Harrell et al.[13]	2013	0.70 (0.51-0.85)	0.60 (0.15-0.95)		
Alexander et. al.[14]	2014	0.42 (0.32-0.51)	1.00 (0.29-1.00)		
Han et al.[15]	2014	0.45 (0.23-0.68)	0.00 (0.00-1.00)		
Lastra et al.[16]	2014	0.46 (0.31-0.61)	1.00(0.16-1.00)		
McIver et al.[17]	2014	0.16 (0.05-0.33)	0.75 (0.19-0.99)		
Celik et al.[18]	2015	0.59 (0.33-0.82)	1.00 (0.16-1.00)		
Noureldine et al.[19]	2015	0.42 (0.35-0.50)	0.83 (0.52-0.98)		
Yang et al.[20]	2015	0.57 (0.44-0.68)	1.00 (0.48-1.00)		
Zhu et al.[21]	2015	0.60 (0.26-0.88)	0.00 (0.00-1.00)		
Abeykoon et al.[22]	2016	0.86 (0.57-0.98)	0.00 (0.00-1.00)		
Chaudhary et al.[23]	2016	0.38 (0.27-0.50)	1.00 (0.63-1.00)		
Marti et al.[6]	2016	0.43 (0.30-0.56)	1.00 (0.59-1.00)		
Sacks et al.[24]	2016	0.33 (0.21-0.47)	1.00 (0.40-1.00)		
Witt et al.[25]	2016	0.40 (0.16-0.68)	0.00 (0.00-1.00)		
Wu et al.[26]	2016	0.62 (0.50-0.72)	0.86 (0.57-0.98)		



ThyroSeq

Study	Year of publication	PPV (95%CI)	NPV (95%CI)	PPV (95%CI)	NPV (95%CI)
Nikiforov et al. Nikiforov et al.[27+28]	2015	0.89 (0.78-0.95)	0.97 (0.93-0.99)		
Shreshta et al.[29]	2016	0.63 (0.44-0.80)	0.88 (0.70-0.98)		
Valderrabano et al.[30]	2017	0.42 (0.25-0.61)	0.91 (0.82-0.97)		
Taye et al.[31]	2018	0.27 (0.14-0.44)	0.96 (0.78-1.00)		



was also validated in a blinded multicenter prospective trial [14]. The results of our meta-analysis include pooled data from 1086 nodules with histopathologic confirmation from multiple studies and showed that GEC indeed has a high sensitivity of 98% and a high NPV of 91%. However, test limitations were the specificity of 12% and a low PPV of 45%, meaning that on aggregate, a "suspicious" result signifies a 12% risk of malignancy. Considering these performance indicators, the GEC can only be considered valuable as a "rule-out test."

Next-generation sequencing technology allows high-output genomic analysis. A custom next-generation sequencing panel called *ThyroSeq* (*initial version*) was designed to target 12 cancer genes with 284 mutational hotspots for the detection of mutations in thyroid cancer by next-generation

Table 3 The results of bivariate summary estimates of sensitivity, specificity, positive likelihood ratio (LR+), negative likelihood ratio (LR−), and diagnostic odds ratio (DOR) for the *GEC* and *ThyroSeq* tests

Method	Sensitivity (95%CI)	Specificity (95%CI)	LR+ (95%CI)	LR− (95%CI)	DOR (95%CI)
GEC	0.98 (0.96–0.99)	0.12 (0.08–0.20)	1.12 (1.05–1.20)	0.13 (0.05–0.31)	8.87 (3.52–22.44)
ThyroSeq	0.84 (0.74–0.91)	0.78 (0.50–0.92)	3.79 (1.40–10.27)	0.20 (0.10–0.39)	18.86 (3.95–90.00)
Bivariate meta-regression hierarchical model—*p* value	0.003	< 0.001			

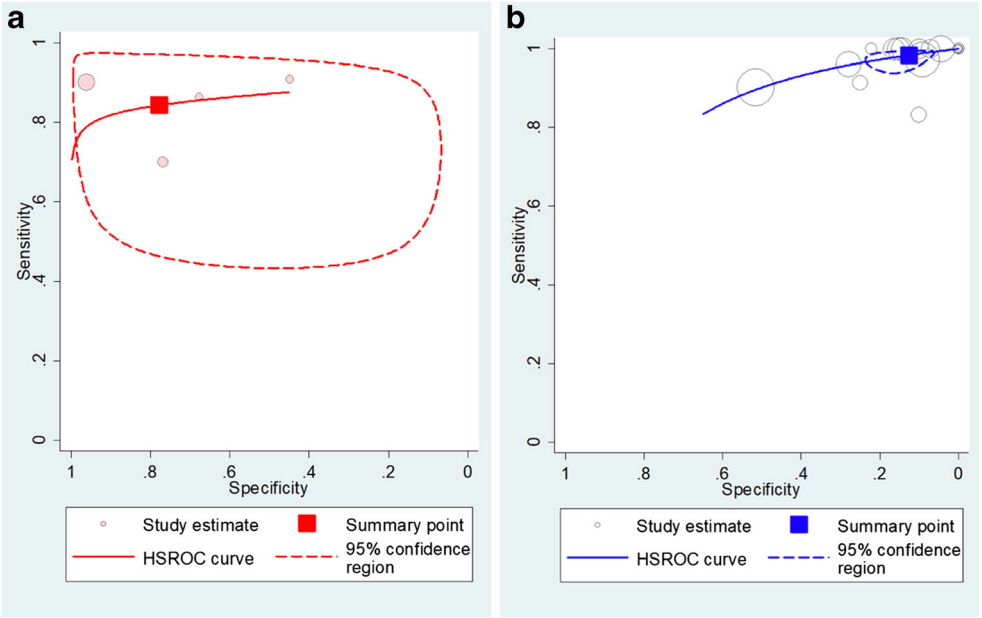


Fig. 2 Bivariate summary estimates of sensitivity and specificity for diagnostic tests (*GEC*, *ThyroSeq*), hierarchical summary receiver operating characteristic, and the corresponding 95% credibility ellipse around the mean values for **a** ThyroSeq and **b** GEC

sequencing technology in fine-needle aspiration and tissue samples [17]. This method was superseded by a second, improved version of the assay (*ThyroSeq v2*) [27]. This version, analyzed in the present study, was designed to detect mutations in > 1000 hotspots of 14 thyroid cancer–related genes (*AKT1*, *BRAF*, *CTNNB1*, *GNAS*, *HRAS*, *KRAS*, *NRAS*, *PIK3CA*, *PTEN*, *RET*, *TP53*, *TSHR*, *TERT*, and *EIF1AX*) and for 42 types of gene fusions or rearrangements known to occur in thyroid cancer (*RET*, *PPARG*, *NTRK1*, *NTRK3*, *BRAF,* and *ALK*) [27]. The results of the studies of the *ThyroSeq v2.1* multigene NGS panel of molecular markers demonstrate that it provides both high sensitivity and high specificity for cancer detection in thyroid nodules with AUS/FLUS cytology, which should allow improved management for these patients [28]. However, the new test to date only has limited real-world experience.

In our analysis, we pooled data from 459 nodules and found both an acceptable sensitivity and specificity, as well as a higher PPV than the GEC and a marginally higher NPV of 93% (89–97%). This is also reflected in the significantly higher overall accuracy found for *ThyroSeq* compared to *GEC*. The combination of all-round acceptable parameters of diagnostic performance suggests *ThyroSeq v2* can be used as both a "rule-in" and "rule-out" test. However, the test still seems to be underdeveloped regarding its interpretation as cutoff points for the recognition of individual mutations, and standard operating procedures still leave room for further study. Therefore, the European Thyroid Association thus far suggests to use this test for scientific purposes, but not yet in the clinical setting [32].

The main limitation of the present study is the scarce source material of four publications for *ThyroSeq v2*, which results in a high heterogeneity of research and publication bias. A more reliable summary may be possible as more such studies will become available. Possibly as a consequence of the limited number of studies compared to the number available for the GEC, a higher heterogeneity with a bigger ellipse representing the confidence region was observed for *ThyroSeq*.

Numerous precautions were taken to minimize potential bias in the study. Bivariate/HSROC meta-analysis is the most rigorous statistical method in the field of diagnostic meta-analysis. This is due to its ability to model the binomial structure as part of the study while taking into account the heterogeneity between the studies [33]. According to guidelines for Meta-analysis of Diagnostic Test Accuracy Studies described by European Network for health technology assessment [34]

Table 4 The results of a bivariate meta-regression hierarchical model with bivariate summary estimates of PPV and NPV and the results of a bivariate meta-regression hierarchical model

Method	PPV (95%CI)	NPV (95%CI)
GEC	0.45 (0.37–0.53)[a]	0.91 (0.85–0.96)[a]
ThyroSeq	0.58 (0.31–0.81)	0.93 (0.89–0.97)
Bivariate meta-regression hierarchical model—*p* value	0.235	0.098

[a] The result after exclusion of the studies, where true negative (TN) results + false negative (FN) results were 0 or 2

Table 5 Calculation of pooled accuracy

Method	Accuracy (95%CI)	*p* value	Accuracy (95%CI)[a]	*p* value
GEC	0.52 (0.45–0.59)	0.015	0.52 (0.45–0.59)[a]	< 0.001
ThyroSeq	0.75 (0.58–0.93)		0.82 (0.67–0.98)[a]	

[a] The result after exclusion of the studies, where true negative (TN) results responsible for the asymmetry according to Egger's test

and recommendations given by the Cochrane Collaboration [11], the bivariate/HSROC method must be used as the standard, together with an analysis of summary ROC curves.

The comparison of the obtained summary results for *GEC* and *ThyroSeq* shows a significant difference between the values of sensitivity and specificity of the studies. The *GEC* studies are characterized by significantly higher sensitivity, while a significantly higher specificity is observed in the *ThyroSeq* group. Therefore, the superiority of the *GEC* test lies in ruling-out of thyroid malignancy in the preoperative assessment of indeterminate thyroid nodules. In contrast, the *ThyroSeq* is characterized with a better performance in the "ruling-in" of malignancy. Pooled sensitivity, specificity, and accuracy as obtained using bivariate/HSROC analysis nonetheless on aggregate showed a clear dominance of the *ThyroSeq* method.

Although molecular testing of cytology specimens from thyroid nodules has the potential to play a major role in the evaluation of indeterminate thyroid lesions, to date, the results of its performance are still not fully satisfactory. The ideal test combining both—high sensitivity and NPV with a high sensitivity and PPV, allowing a test to be both a "ruling-in" and "ruling-out" test—still needs to be developed. Furthermore, such an ideal test also would have the ability to stratify prognostic risk based on the mutation also providing a direction for therapy.

Recently, a new version of the *ThyroSeq* panel—*ThyroSeq v3*—has been proposed [35]. This DNA- and RNA-based next-generation sequencing assay analyzes 112 genes for a *variety* of genetic alterations, including point mutations, insertions/deletions, gene fusions, copy number alterations, and abnormal gene expression, and it uses a genomic classifier (GC) to separate malignant lesions from benign lesions [35]. It is conceivable that this further development of the *ThyroSeq* test, now being rolled out in clinical practice, has the potential to improve on the already clinically interesting diagnostic performance of the *ThyroSeq v2*. However, real-life study data are currently still lacking on this test, precluding its inclusion in the present analysis.

Conclusion

The 167 gene *GEC* with a high sensitivity and NPV may be helpful in ruling out malignancy in cases of indeterminate thyroid nodule cytology. However, *ThyroSeq v2* has a significantly higher specificity and accuracy with acceptable sensitivity, so that it has the potential for use as an all-round test of malignancy of thyroid nodules. However, long-term outcome data are necessary before a clear recommendation can be made.

Funding Information This study was supported by PRELUDIUM Grant of Polish National Center for Science number 2015/19/N/NZ5/02257.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda System for Reporting Thyroid Cytopathology: a meta-analysis. *Acta Cytol.* 2012;56(4):333–339.
2. Ho AS, Sarti EE, Jain KS, Wang H, Nixon IJ, Shaha AR, Shah JP, Kraus DH, Ghossein R, Fish SA, Wong RJ, Lin O, Morris LGT Malignancy rate in thyroid nodules classified as Bethesda category III (AUS/FLUS). *Thyroid.* 2014;24(5):832–839.
3. 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.
4. Nikiforov YE, Ohori NP, Hodak SP, Carty SE, LeBeau SO, Ferris RL, Yip L, Seethala RR, Tublin ME, Stang MT, Coyne C, Johnson JT, Stewart AF, Nikiforova MN Impact of mutational testing on the diagnosis and management of patients with cytologically indeterminate thyroid nodules: a prospective analysis of 1056 FNA samples. *J Clin Endocrinol Metab.* 2011;96(11):3390–3397.
5. Alexander EK, Kennedy GC, Baloch ZW, Cibas ES, Chudova D, Diggans J, Friedman L, Kloos RT, LiVolsi VA, Mandel SJ, Raab SS, Rosai J, Steward DL, Walsh PS, Wilde JI, Zeiger MA, Lanman RB, Haugen BR Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. *N Engl J Med.* 2012;367(8):705–715.
6. Marti JL, Avadhani V, Donatelli LA, Niyogi S, Wang B, Wong RJ, Shaha AR, Ghossein RA, Lin O, Morris LGT, Ho AS Wide Interinstitutional Variation in Performance of a Molecular Classifier for Indeterminate Thyroid Nodules. *Ann Surg Oncol.* 2015;22(12):3996–4001.
7. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6(7):e1000097.
8. Higgins JPTG, S., eds. *Cochrane hand book for systematic reviews of interventions version 5.1.0.* 2011.
9. Zeng X, Zhang Y, Kwong JS, et al. The methodological quality assessment tools for preclinical and clinical studies, systematic

review and meta-analysis, and clinical practice guideline: a systematic review. *J Evid Based Med.* 2015;8(1):2–10.
10. Wells GAS, B.; O'Connell, D. et al. *The Newcastle-Ottawa scale (nos) for assessing the quality of nonrandomized studies in meta-analyses. The 3rd symposium on systematic reviews: Beyond the basics.* Oxford, England2000.
11. Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. *Chapter 10: Analysing and Presenting Results.* The Cochrane Collaboration; 2010.
12. Leeflang MM, Deeks JJ, Rutjes AW, Reitsma JB, Bossuyt PM. Bivariate meta-analysis of predictive values of diagnostic tests can be an alternative to bivariate meta-analysis of sensitivity and specificity. *J Clin Epidemiol.* 2012;65(10):1088–1097.
13. 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.
14. Alexander EK, Schorr M, Klopper J, Kim C, Sipos J, Nabhan F, Parker C, Steward DL, Mandel SJ, Haugen BR Multicenter clinical experience with the Afirma gene expression classifier. *J Clin Endocrinol Metab.* 2014;99(1):119–125.
15. Han LO, Li XY, Cao MM, Cao Y, Zhou LH. Development and validation of an individualized diagnostic signature in thyroid cancer. *Cancer Med.* 2018;7(4):1135–1140.
16. 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.
17. McIver B, Castro MR, Morris JC, Bernet V, Smallridge R, Henry M, Kosok L, Reddi H 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.
18. Celik B, Whetsell CR, Nassar A. Afirma GEC and thyroid lesions: An institutional experience. *Diagn Cytopathol.* 2015;43(12):966–970.
19. Noureldine SI, Olson MT, Agrawal N, Prescott JD, Zeiger MA, Tufano RP. Effect of Gene Expression Classifier Molecular Testing on the Surgical Decision-Making Process for Patients With Thyroid Nodules. *JAMA Otolaryngol Head Neck Surg.* 2015;141(12):1082–1088.
20. Yang SE, Sullivan PS, Zhang J, Govind R, Levin MR, Rao JY, Moatamed NA Has Afirma gene expression classifier testing refined the indeterminate thyroid category in cytology? *Cancer Cytopathol.* 2016;124(2):100–109.
21. Zhu QL, Faquin WC, Samir AE. Relationship Between Sonographic Characteristics and Afirma Gene Expression Classifier Results in Thyroid Nodules With Indeterminate Fine-Needle Aspiration Cytopathology. *AJR Am J Roentgenol.* 2015;205(4):861–865.
22. Abeykoon JP, Mueller L, Dong F, Chintakuntlawar AV, Paludo J, Mortada R. The Effect of Implementing Gene Expression Classifier on Outcomes of Thyroid Nodules with Indeterminate Cytology. *Horm Cancer.* 2016;7(4):272–278.
23. 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.
24. Sacks WL, Bose S, Zumsteg ZS, et al. Impact of Afirma gene expression classifier on cytopathology diagnosis and rate of thyroidectomy. *Cancer.* 2016;124(10):722–728.
25. Witt RL. Outcome of thyroid gene expression classifier testing in clinical practice. *Laryngoscope.* 2016;126(2):524–527.
26. Wu JX, Lam R, Levin M, Rao J, Sullivan PS, Yeh MW. Effect of malignancy rates on cost-effectiveness of routine gene expression classifier testing for indeterminate thyroid nodules. *Surgery.* 2016;159(1):118–126.
27. Nikiforov YE, Carty SE, Chiosea SI, Coyne C, Duvvuri U, Ferris RL, Gooding WE, Hodak SP, LeBeau SO, Ohori NP, Seethala RR, Tublin ME, Yip L, Nikiforova MN Highly accurate diagnosis of cancer in thyroid nodules with follicular neoplasm/suspicious for a follicular neoplasm cytology by ThyroSeq v2 next-generation sequencing assay. *Cancer.* 2014;120(23):3627–3634.
28. Nikiforov YE, Carty SE, Chiosea SI, Coyne C, Duvvuri U, Ferris RL, Gooding WE, LeBeau S, Ohori NP, Seethala RR, Tublin ME, Yip L, Nikiforova MN Impact of the Multi-Gene ThyroSeq Next-Generation Sequencing Assay on Cancer Diagnosis in Thyroid Nodules with Atypia of Undetermined Significance/Follicular Lesion of Undetermined Significance Cytology. *Thyroid.* 2015;25(11):1217–1223.
29. Shrestha RT, Evasovich MR, Amin K, Radulescu A, Sanghvi TS, Nelson AC, Shahi M, Burmeister LA Correlation Between Histological Diagnosis and Mutational Panel Testing of Thyroid Nodules: A Two-Year Institutional Experience. *Thyroid.* 2016;26(8):1068–1076.
30. Valderrabano P, Khazai L, Leon ME, Thompson ZJ, Ma Z, Chung CH, Hallanger-Johnson JE, Otto KJ, Rogers KD, Centeno BA, McIver B Evaluation of ThyroSeq v2 performance in thyroid nodules with indeterminate cytology. *Endocr Relat Cancer.* 2017;24(3):127–136.
31. Taye A, Gurciullo D, Miles BA, Gupta A, Owen RP, Inabnet III WB, Beyda JN, Marti JL Clinical performance of a next-generation sequencing assay (ThyroSeq v2) in the evaluation of indeterminate thyroid nodules. *Surgery.* 2018;163(1):97–103.
32. Paschke R, Cantara S, Crescenzi A, Jarzab B, Musholt TJ, Sobrinho Simoes M. European Thyroid Association Guidelines regarding Thyroid Nodule Molecular Fine-Needle Aspiration Cytology Diagnostics. *Eur Thyroid J.* 2017;6(3):115–129.
33. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol.* 2005;58(10):982–990.
34. *Guideline EJ. Meta-analysis of diagnostic test accuracy studies.* 2014.
35. Nikiforova MN, Mercurio S, Wald AI, Barbi de Moura M, Callenberg K, Santana-Santos L, Gooding WE, Yip L, Ferris RL, Nikiforov YE Analytical performance of the ThyroSeq v3 genomic classifier for cancer diagnosis in thyroid nodules. *Cancer.* 2018;124(8):1682–1690.