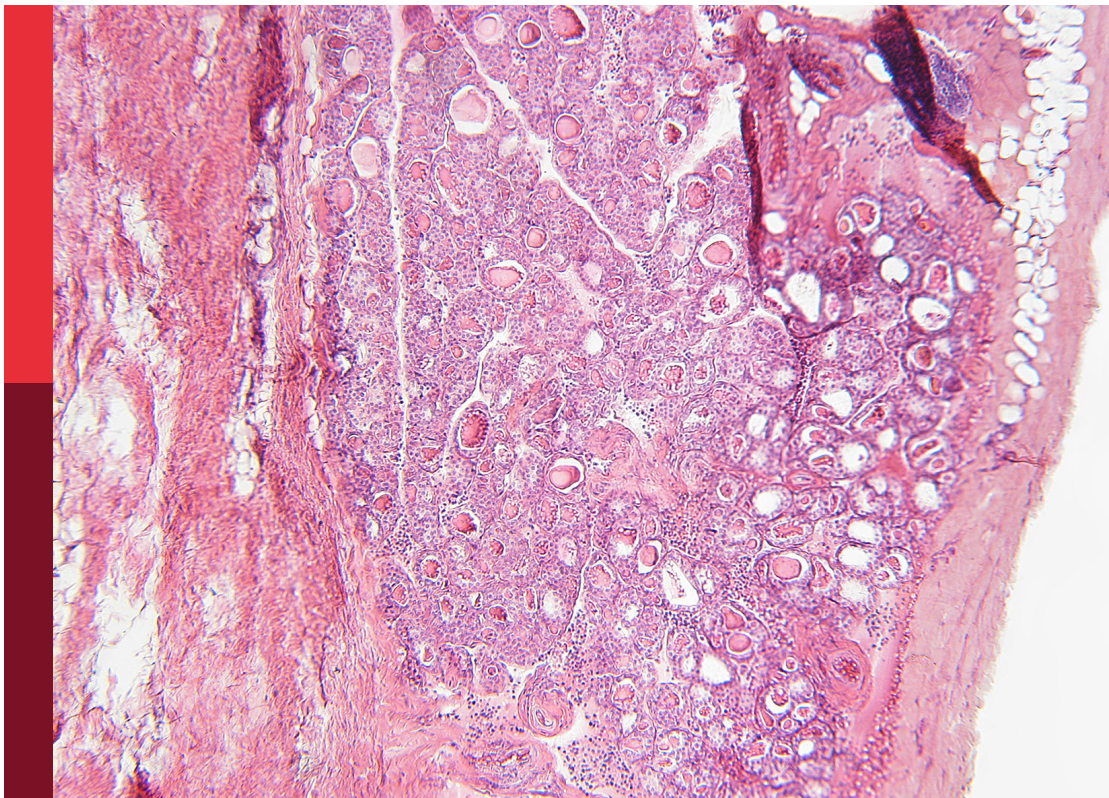
frontiers | Research Topics

Thyroid nodule evaluation: Current, evolving, and emerging tools

Edited by
Jeffrey Garber, Andrea Frasoldati, Enrico Papini and Vivek Patkar

Published in
Frontiers in Endocrinology



ISSN 1664-8714
ISBN 978-2-8325-3516-5
DOI 10.3389/978-2-8325-3516-5

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Thyroid nodule evaluation: Current, evolving, and emerging tools

Topic editors

Jeffrey Garber — Atrius Health, United States
Andrea Frasoldati — Endocrine Unit ASMN, Italy
Enrico Papini — Ospedale Regina Apostolorum, Italy
Vivek Patkar — Deontics Ltd, United Kingdom

Citation

Garber, J., Frasoldati, A., Papini, E., Patkar, V., eds. (2023). *Thyroid nodule evaluation: Current, evolving, and emerging tools*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3516-5

Table of contents

frontiers | Frontiers in Endocrinology

TYPE Editorial
PUBLISHED 05 September 2023
DOI 10.3389/fendo.2023.1276323

OPEN ACCESS

EDITED AND REVIEWED BY
Terry Francis Davies,
Icahn School of Medicine at Mount Sinai,
United States

*CORRESPONDENCE
Jeffrey R. Garber
jgarber@bidmc.harvard.edu

RECEIVED 11 August 2023
ACCEPTED 18 August 2023
PUBLISHED 05 September 2023

CITATION
Garber JR, Frasoldati A, Patkar V and Papini E (2023) Editorial: Thyroid nodule evaluation: current, evolving, and emerging tools.
Front. Endocrinol. 14:1276323.
doi: 10.3389/fendo.2023.1276323

Editorial: Thyroid nodule evaluation: current, evolving, and emerging tools

Jeffrey R. Garber[1,2,3]*, Andrea Frasoldati[4], Vivek Patkar[5] and Enrico Papini[6]

[1]Endocrinology, Atrius Health, Boston, MA, United States, [2]Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States, [3]Department of Medicine, Harvard Medical School, Boston, MA, United States, [4]Endocrinology Unit, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy, [5]Deontics Ltd, London, United Kingdom, [6]Endocrinology and Metabolism Department, Regina Apostolorum Hospital, Albano, Rome, Italy

KEYWORDS

thyroid nodule, clinical practice guideline, clinical calculators clinical decision support, computer interpretable guideline, evidence based care, guideline compliance, CDSS

Editorial on the Research Topic
Thyroid nodule evaluation: current, evolving, and emerging tools

Thyroid nodules are common, predominantly benign, asymptomatic on presentation, and most often remain so (Uppal et al.) (1). Moreover, those that are malignant are principally small low-risk neoplasms with an indolent course and minimal impact on survival (2). Hence, most patients do not benefit from extensive evaluation, treatment, and monitoring (2–5). On the contrary, costly diagnostic techniques and treatment may have a detrimental impact on a patient's physical, emotional, and financial status (Uppal et al.). In the United States well over 500 000 fine-needle aspirations (FNAs) of thyroid nodules are performed yearly with as many as 40% likely unnecessary (6). In European countries, such as Germany and France, as well as well as in the United States, overtreatment is reflected by most thyroidectomies performed for nodular thyroid disease prove to be for benign disease while the minority that are malignant are principally comprised of low-risk thyroid cancers (7). Recommendations from professional societies such as the American Association of Clinical Endocrinologists (AACE), Associazione Medici Endocrinologi (AME) (2), American Thyroid Association (ATA) (3), European Thyroid Association (ETA) (4), and American College of Radiology (5) for reducing the collective burden of evaluating and treating thyroid nodules and low risk thyroid cancers have had limited impact on achieving this goal (6).

Over the past 3 decades narrative or written clinical practice guidelines (CPGs) have emerged as an increasingly important tool to aid clinicians in managing a host of medical conditions. Guidelines are regularly cited in publications and medical education forums and used as a basis for medical decision making in both clinical and administrative settings. Yet, despite their widespread clinical use, there is substantial room for improvement in the following ways:

- Establishing the cost effectiveness and validity of recommendations, which are often based on expert opinion, retrospective studies, and study populations that are not generalizable.
- Evaluating their impact on patient quality of life
- Routinely disseminating, distributing, and implementing guidelines
- Gauging their implementation by tracking their use and applicability
- Creating mechanisms for vetting guideline recommendations in various clinical situations and across different populations and cultures
- Addressing their often-formidable length and the wealth of information they contain, which makes them hard for physicians to navigate as well as absorb and retain (internalize)
- Providing timely updates of narrative multi-authored, highly validated documents

Advanced Clinical Decisions Support Systems (CDSS) addresses all the above-mentioned points by transforming CPGs into computer interpretable guidelines (CIGs). CIGs are derived from CPGs. They employ execution engines (programs) to analyze patient-specific data to electronically generate, document, and track recommendations (Garber and Patkar).

The articles in this Research Topic of Frontiers in Endocrinology are written by a diverse group of authors representing various specialties and regions. They cover the gamut of tools available for evaluating thyroid nodules. They also underscore the challenges in developing a streamlined, cost-effective approach that minimizes unnecessary evaluation and intervention that not only does not benefit patients but may harm them while maximizing the chances for identifying clinically significant disease that if left untreated would lead to significant morbidity.

The diagnostic tools available to clinicians caring for patients with thyroid nodules can be summarized in Table 1. Over time, history and physical examination have taken on a marginal role in evaluating thyroid nodules. This is due to several reasons. Many nodules are discovered incidentally on imaging that is not being performed to evaluate the thyroid gland. Since physical examination is not as reliable as ultrasound in establishing the presence, size, or characteristics of benign or malignant disease it is less frequently or carefully performed. Thus, it is not regularly used to follow patients with benign, inconsequential nodularity. This is not without a downside. Greater reliance on ultrasound as a monitoring tool and to facilitate fine needle aspiration of a candidate nodule leads to surveying the remainder of the thyroid gland for nodularity. While improving the yield and accuracy of fine needle aspiration, ultrasound often leads to the detection and evaluation of clinically inapparent nodules that are either diagnostically indeterminate or an inconsequential malignancy, resulting in surgery without clear benefit.

The current emphasis on evaluating thyroid nodules in addition to standard B-mode ultrasonography, are risk stratification tools such as ACR TIRADS, fine needle aspiration with Bethesda classification and molecular genetic markers (Patel et al.), where available. Additional tools that have not yet been standardized, widely adapted, or extensively studied, include ultrasound elastography (Li et al.), contrast enhanced ultrasound (Zhou et al.), emerging AI tools (Xu D. et al.) (8), and immunocytochemistry (Crescenzi and Baloch; Taccogna et al.). On a positive note, therapeutic advances promoted

TABLE 1 Diagnostic Tools.

*History (see TNAPP)
Radiation
Family History
Symptoms
Thyroid Disease
Disorders of function
Structural abnormalities
*Physical Exam
Initial
Serial
Laboratory determinations
*TSH levels
Calcitonin levels
Imaging:
*Ultrasound
*Standard B mode imaging
Elastography
Contrast Enhanced Ultrasound
**Artificial Intelligence
**Computer Aided Diagnosis
**Radiomics
Nuclear medicine
Radioactive Iodine
PET (FDG)
*Clinical Practice Guidelines (CPGs)
*Risk Score Stratification Tools (Clinical Calculators)
Computer Interactive Guidelines
*Fine Needle Aspiration
*Molecular Marker/Diagnostics (mostly USA)
DNA (Mutations)
Messenger RNA
micro– RNA
Immunocytochemistry
*Principal
**Emerging

by professional societies (9–12) employing minimally invasive ablation procedures have been made. These techniques are being used more frequently. Compared with surgery, they are less expensive, cause less morbidity, and have fewer adverse effects on patient quality of life.

Challenges, however, remain. Tools for assessing risk vary, employing different characteristics underpinning a strong argument for universally accepted risk stratification tools (Majety et al.) (13, 14). Oftentimes, newer technology is complementary rather than substitutive, increasing cost without offering consistently substantial benefit (Uppal et al.). Yet reliable, relatively expensive, new AI tools may ultimately play a key role in resource poor regions that not only lack access to diagnostic tools, but do not have the professional expertise to interpret ultrasound (8).

A promising new development strongly supported by the editors of this Research Topic is the adaptation of CIGs to complement and facilitate the use of clinical practice guidelines and risk stratification systems. Using advanced CDSS to co–develop CIGs that complement conventional society guidelines, may not only increase the use of CPGs, but could serve as testing tools for assessing the efficacy and generalizability of a sequence of diagnostic tests (Triggiani et al.) (15). This could be accomplished by employing a range of assumptions and models for the respective sensitivity, specificity, accuracy, and costs of each tool being employed.

Our challenge is improving our approach to evaluating and managing thyroid nodules by increasingly employing minimally invasive techniques, and developing more specific, less costly molecular tests that not only diagnose malignancy but also provide prognostic information. Doing so will substantially reduce the number of preventable adverse effects of invasive diagnostic and non–surgical therapeutic procedures, surgical morbidity, and financial toxicity at the expense of only detecting the relatively small percentage that prove to be thyroid cancers that require treatment.

Author contributions

JG: Writing – original draft, Writing – review & editing. AF: Writing – review & editing, Writing – original draft. VP: Writing – review & editing. EP: Writing – review & editing.

Conflict of interest

VP is CMO of Deontics Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Durante C, Costante G, Lucisano G, Bruno R, Meringolo D, Paciaroni A, et al. The natural history of benign thyroid nodules. *JAMA* (2015) 313(9):926–35. doi: 10.1001/jama.2015.0956

2. Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedüs L, et al. AACE/ACE/AME task force on thyroid nodules. American association of clinical endocrinologists, american college of endocrinology, and associazione medici endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules–2016 update. *Endocr Pract* (2016) 22(5):622–39. doi: 10.4158/EP161208.GL

3. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the american thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid* (2016) 26(1):1–133. doi: 10.1089/thy.2015.0020

4. Durante C, Hegedus L, Czarniecka A, Paschke R, Russ G, Schmitt F, et al. 2023 European Thyroid Association clinical practice guidelines for thyroid nodule management. *Eur Thyroid J* (2023) 12(5):e230067. doi: 10.1530/ETJ–23–0067

5. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR thyroid imaging, reporting and data system (TI–RADS): white paper of the ACR TI–RADS committee. *J Am Coll Radiol* (2017) 14(5):587–95. doi: 10.1016/j.jacr.2017.01.046

6. White C, Weinstein MC, Fingeret AL, Randolph GW, Miyauchi A, Ito Y, et al. Is less more? A microsimulation model comparing cost–effectiveness of the revised american thyroid association's 2015 to 2009 guidelines for the management of patients with thyroid nodules and differentiated thyroid cancer. *Ann Surg* (2020) 271(4):765–73. doi: 10.1097/SLA.0000000000003074

7. Bartsch DK, Dotzenrath C, Vorländer C, Zielke A, Weber T, Buhr HJ, et al. The stuDoQ/thyroid study group TSS. Current practice of surgery for benign goitre–an analysis of the prospective DGAV stuDoQ|Thyroid registry. *J Clin Med* (2019) 8(4):477. doi: 10.3390/jcm8040477

8. Tessler FN, Thomas J. Artificial intelligence for evaluation of thyroid nodules: A primer. *Thyroid* (2023) 33(2):150–8. doi: 10.1089/thy.2022.0560

9. Perros P, Hegedüs L, Nagy EV, Papini E, Hay HA, Abad–Madroñero J, et al. The impact of hypothyroidism on satisfaction with care and treatment and everyday living: results from E–mode patient self–assessment of thyroid therapy, a cross–sectional, international online patient survey. *Thyroid* (2022) 32(10):1158–68. doi: 10.1089/thy.2022.0324

10. Papini E, Monpeyssen H, Frasoldati A, Hegedüs L. 2020 European thyroid association clinical practice guideline for the use of image–guided ablation in benign thyroid nodules. *Eur Thyroid J* (2020) 9(4):172–85. doi: 10.1159/000508484

11. Orloff LA, Noel JE, Stack BC Jr, Russell MD, Angelos P, Baek JH, et al. Radiofrequency ablation and related ultrasound–guided ablation technologies for treatment of benign and Malignant thyroid disease: An international multidisciplinary consensus statement of the American Head and Neck Society Endocrine Surgery Section with the Asia Pacific Society of Thyroid Surgery, Associazione Medici Endocrinologi, British Association of Endocrine and Thyroid Surgeons, European Thyroid Association, Italian Society of Endocrine Surgery Units, Korean Society of Thyroid Radiology, Latin American Thyroid Society, and Thyroid Nodules Therapies Association. *Head Neck* (2022) 44(3):633–60. doi: 10.1002/hed.26960

12. Papini E, Hegedüs L. Minimally invasive ablative treatments for benign thyroid nodules: current evidence and future directions. *Thyroid* (2023) 30(8):890–3. doi: 10.1089/thy.2023.0263

13. Hoang JK, Asadollahi S, Durante C, Hegedüs L, Papini E, Tessler FN. An international survey on utilization of five thyroid nodule risk stratification systems: A needs assessment with future implications. *Thyroid* (2022) 32(6):675–81. doi: 10.1089/thy.2021.0558

14. Solymosi T, Hegedűs L, Bonnema SJ, Frasoldati A, Jambor L, Karanyi Z, et al. Considerable interobserver variation calls for unambiguous definitions of thyroid nodule ultrasound characteristics. *Eur Thyroid J* (2023) 12(2):e220134. doi: 10.1530/ETJ–22–0134

15. Garber JR, Papini E, Frasoldati A, Lupo MA, Harrell RM, Parangi S, et al. American association of clinical endocrinology and associazione medici endocrinologi thyroid nodule algorithmic tool. *Endocr Pract* (2021) 27(7):649–60. doi: 10.1016/j.eprac.2021.04.007

frontiers | Frontiers in Endocrinology

TYPE Systematic Review
PUBLISHED 08 September 2022
DOI 10.3389/fendo.2022.938961

OPEN ACCESS

EDITED BY
Jeffrey Garber,
Atrius Health, United States

REVIEWED BY
Priyanka Majety,
Beth Israel Deaconess Medical Center and Harvard Medical School, United States
Mariantonia Nacchio,
University of Naples Federico II, Italy
Andrea Frasoldati,
Endocrine Unit ASMN, Italy

*CORRESPONDENCE
Weiwei Zhan
shanghairuijin@126.com

SPECIALTY SECTION
This article was submitted to Thyroid Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 08 May 2022
ACCEPTED 15 August 2022
PUBLISHED 08 September 2022

CITATION
Hu Y, Xu S and Zhan W (2022) Diagnostic performance of C-TIRADS in malignancy risk stratification of thyroid nodules: A systematic review and meta-analysis.
Front. Endocrinol. 13:938961.
doi: 10.3389/fendo.2022.938961

Diagnostic performance of C-TIRADS in malignancy risk stratification of thyroid nodules: A systematic review and meta-analysis

Yan Hu, Shangyan Xu and Weiwei Zhan*

Department of Ultrasound, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Background: Chinese thyroid imaging reports and data systems (C-TIRADS) is a novel malignancy risk stratification used for thyroid nodule diagnosis and guiding thyroid fine needle aspiration (FNA). In this review, we aim to assess the performance of C-TIRADS in malignancy risk stratification of thyroid nodules.

Methods: PubMed, Medline, Web of Science, Embase, CNKI, and Wanfang databases were searched until 1 April 2022. Original articles reporting data about C-TIRADS and setting FNA or histology as reference standards were included. C-TIRADS 4A, 4B, and 4C were set as thresholds, respectively, to obtain pooled sensitivity, specificity, positive likelihood ratio (LR+), negative likelihood ratio (LR-), diagnostic odds ratio (DOR), and the area under the curve (AUC). Integrated nested Laplace approximation was used for Bayesian bivariate meta-analysis of diagnostic tests.

Results: Sixteen studies were included, evaluating 11,506 thyroid nodules. The rate of malignancy in each risk classification is comparable with that in C-TIRADS. C-TIRADS 4B appeared to have better diagnostic performance than C-TIRADS 4A and 4C. The pooled sensitivity, specificity, LR+, LR-, and DOR of C-TI-RADS 4B were 0.94 (95% CI: 0.89–0.97), 0.70 (95% CI: 0.60–0.79), 3.20 (95% CI: 2.28–4.39), 0.09 (95% CI: 0.05–0.15), and 33.71 (95% CI: 25.51–42.40), respectively. The area under the summary ROC curve was 0.94 (95% CI: 0.90-0.96).

Conclusion: C-TIRADS performed well in malignancy risk stratification of thyroid nodules. C-TIRADS 4B showed strong evidence of detecting malignancy.

KEYWORDS

meta-analysis, thyroid, thyroid nodules, ultrasound, risk assessment

Introduction

Thyroid nodules are common. They are detected in 19%–67% of the population (1, 2). With a malignancy rate of less than 5%–10%, the purpose of evaluation for thyroid nodules is to identify malignant nodules (3–5). Ultrasound (US) has been widely applied in the initial evaluation of thyroid nodules and deemed an important standard to distinguish whether they are benign or malignant. A diagnosis based solely on US is not completely reliable (6), and the cytology obtained by fine needle aspiration (FNA) is still considered the gold standard diagnostic tool for thyroid nodules. Yet, the application of US-based risk stratification systems serves as a means to standardize the results of US examination and a tool for deciding which nodules should undergo FNA.

Previously, there have been several thyroid imaging reports and data systems (TIRADS), such as the American College of Radiology (ACR) TIRADS, the Korean (K) TIRADS, Kwak-TIRADS, and the European Thyroid Association (EU) TIRADS (7–10). In 2020, supported by the Superficial Organ and Vascular Ultrasound Group of the Society of Ultrasound in Medicine of the Chinese Medical Association and the Chinese Artificial Intelligence Alliance for Thyroid and Breast Ultrasound, Zhou et al. officially proposed a Chinese version of TIRADS (C-TIRADS) (11). C-TIRADS takes into account both the international standards for the US evaluation and the local conditions of the national health organization in China. Presently, C-TIRADS have been used in some studies to classify thyroid nodules (12–14), but the systematic performance of C-TIRADS has been so far marginally explored.

In this study, we aim to conduct a systematic review and meta-analysis to evaluate the performance of C-TIRADS.

Materials and methods

Search strategy and selection criteria

This meta-analysis was performed based on the Preferred Reporting Items for a Systematic Review and Meta-analysis (PRISMA) reporting guideline (15). We searched PubMed, Medline, Web of Science, Embase, CNKI, and Wanfang databases for studies published before 1 April 2022 using the following search terms: "Chinese-TIRADS," "C-TIRADS," "Chinese thyroid imaging reports and data systems," and related terms.

The studies included in this analysis were based on the following criteria: (1) thyroid nodules were assessed by C-TIRADS classification; (2) reference standards were histopathological and/or cytological examination; (3) studies with sufficient data and without overlapping data were included; and (4) the search was limited to human studies published in English or Chinese. The full text was examined by two reviewers independently. Those that did not meet the criteria were excluded.

Data collection and quality assessment

The following data were extracted from the main paper and supplementary data by two reviewers independently: (1) general information of the study (author, year of publication, study type, number of patients, sex distribution, average age/range of age, and number of nodules); (2) the reference standard for the diagnosis of malignancy; (3) the number of benign and malignant nodules; (4) the number of papillary thyroid carcinoma (PTC), follicular thyroid carcinoma (FTC), medullary thyroid carcinoma (MTC), and other malignancies; and (5) the US model and interpretation.

The risk of bias was assessed independently by two reviewers. The Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool was used for the following aspects: patient selection, index test, reference standard, flow, and timing (16). Risk of bias and concerns about applicability were assessed as low, high, or unclear. All the disagreements were resolved by two reviewers or adjudicated by a third reviewer.

Assessment of thyroid nodules

Thyroid nodule assessment followed the C-TIRADS guideline (11), which excludes US features that have not been fully validated as risk factors for predicting malignancy. C-TIRADS assigned levels of malignancy risk to different patterns, a total of five features, namely solid composition, microcalcifications, markedly hypoechoic, ill-defined/irregular margins or extrathyroidal extensions, and vertical orientation. Each of these features scored +1 point. Comet-tail artifacts were considered as a sign of benign nodule and got -1 point. Every category and malignant rate were based on the points in C-TIRADS (Table 1).

FNA was based on recommendations of C-TIRADS. The results of FNA were determined by the Bethesda system for reporting thyroid cytopathology (17). Class II was defined as benign and class V or VI as malignant. Class III and IV prompted a repeat FNA. When the repeat FNA was benign, the nodule was followed for 24 months or more, and if stable, it was classified as benign. Surgical histopathology, when available, was considered definitive.

Evaluation of diagnostic accuracy

Meta-analysis was performed by R software (version 4.1.3) with the meta4diag and Bayesian bivariate integrated nested Laplace approximation (INLA) package (18). When we defined 4A as the cutoff, a benign nodule was considered as true negative if it was classified as C-TIRADS 2 or 3. A benign nodule was considered as false positive if it was classified as C-TIRADS 4A, 4B, 4C, or 5. A malignant nodule was considered as true positive

if it was classified as C-TIRADS 4A, 4B, 4C, or 5. A malignant nodule was considered as false negative if it was classified as C-TIRADS class 2 or 3. With the same method, true negative, false positive, true positive, and false negative values were defined when setting 4B or 4C as the cutoff.

The diagnostic performance of C-TIRADS for thyroid nodules was analyzed with a random-effects model to calculate estimates of sensitivity, specificity, positive likelihood ratio (LR+), negative likelihood ratio (LR-), and diagnostic odds ratio (DOR) with 95% confidence intervals (95% CI), based on the extracted data of true positive, false positive, true negative, and false negative values. Forest plots of point estimates and 95% CI were provided. The DOR provides a single measure of test performance. Higher DOR values indicate better diagnostic performance. LR+ is the probability of biopsy-proven malignant nodules identified by high C-TIRADS classification (for example, when setting C-TIRADS 4A as the cutoff, C-TIRADS 4A, 4B, 4C, and 5 were regarded as high C-TIRADS classification) compared with that of benign nodules. LR+ higher than 10.0 means strong evidence; 5.0–10.0, moderate evidence; and less than 5.0, weak evidence. LR- is the probability of biopsy-proven benign nodules identified by low C-TIRADS classification (for example, when setting C-TIRADS 4A as the cutoff, C-TIRADS 2 and 3 were regarded as low C-TIRADS classification) compared with that of malignant nodules.

LR- less than 0.1 means strong evidence; 0.1–0.2, moderate evidence; and higher than 0.2, weak evidence. Crosshair plot and summary receiver-operating characteristic (SROC) curves were plotted by R software. Sensitivity analysis was used to evaluate the stability of the result of the meta-analysis *via* the sequential omission of individual studies.

Results

Search results

The initial search identified 111 articles from PubMed, Medline, Web of Science, Embase, CNKI, and Wanfang databases until 1 April 2022. After removing duplicates, we screened 51 articles through the title and the abstract, and 29 articles were deemed irrelevant. Following a full-text assessment, we removed 6 articles due to inadequate or overlapping data and poor quality. Eventually, 16 studies were selected for further analysis (Figure 1) (13, 14, 19–32). QUADAS-2 classification was used to assess the quality of included publications (Figure S1).

Study and patient characteristics

There were 16 studies in total that included the data of 9,052 patients (Table 1). All the studies were retrospective in nature and were published between 2020 and 2022. The number of patients in each study varied between 70 and 2,141 (Table 2). In total, there were 6,820 women and 2,024 men. Of the 11,506 thyroid nodules included, 7,223 were benign and 4,283 were malignant (Table 3). The number of nodules varied from 92 to 2,141 in different studies. Histopathological and/or cytological evidence was regarded as the reference standard in all articles. If both histopathological and cytological examinations were available, the final diagnosis was based on histopathological results. According to nine studies that reported the type of malignant nodules, the most common subtype is papillary thyroid carcinoma.

Diagnostic performance of C-TIRADS in thyroid nodule assessment

Firstly, we calculated the prevalence of malignancy in each risk stratification category. The rate of malignant thyroid nodules was 0% in C-TIRADS 2, 1.37% in C-TIRADS 3, 10.62% in C-TIRADS 4A, 40.02% in C-TIRADS 4B, 77.96% in C-TIRADS 4C, and 94.61% in C-TIRADS 5 (Table 4).

Secondly, C-TIRADS 4A, 4B, and 4C were each analyzed separately to get the diagnostic indicators. The pooled sensitivity of C-TIRADS 4A (1.00, 95% CI: 0.99–1.00) was higher than 4B (0.94, 95% CI: 0.89–0.97) and 4C (0.71, 95% CI: 0.60–0.81), while the pooled specificity of C-TIRADS 4C (0.90, 95% CI: 0.84–0.94) was higher than 4A (0.30, 95% CI: 0.23–0.38) and 4B (0.70, 95% CI: 0.60–0.79) (Figure 2).

TABLE 1 C-TIRADS malignancy risk stratification of thyroid nodules.

Category	US features	Points	Likelihood of malignancy
C-TIRADS 1	No nodule	-	0%
C-TIRADS 2	Benign	-1 point	0%
C-TIRADS 3	Probably benign	0 points	<2%
C-TIRADS 4A	Low suspicion for malignancy	1 point	2%–10%
C-TIRADS 4B	Moderate suspicion for malignancy	2 points	10%–50%
C-TIRADS 4C	High suspicion for malignancy	3–4 points	50%–90%
C-TIRADS 5	Highly suggestive of malignancy	5 points	>90%
C-TIRADS 6	Biopsy proved malignancy	-	100%

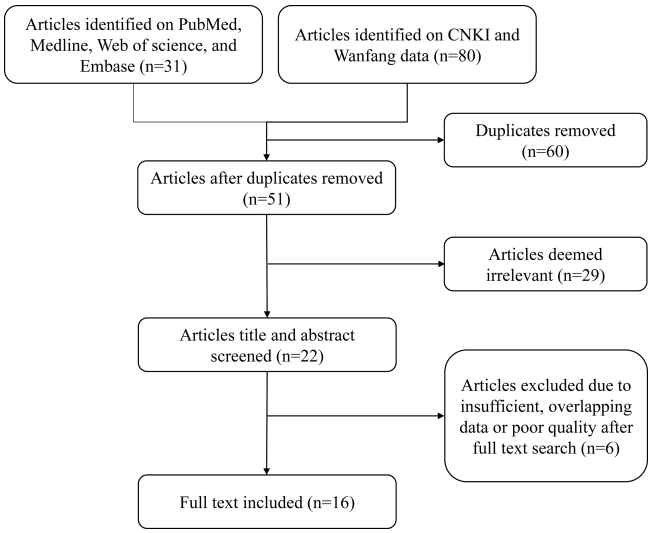


FIGURE 1
Flow diagram of the article selection process.

Thirdly, DOR and the SROC plot were used to determine the optimal one between C-TIRADS 4B and 4C. The DOR of C-TIRADS 4B ranged from 8.37 to 77.92 (summary 33.71, 95% CI: 25.51–42.40), while C-TIRADS 4C ranged from 9.21 to 54.62 (summary 23.77, 95% CI: 17.06–34.37) (Figure 3). The SROC plots suggested that the AUC of 4B (0.94, 95% CI: 0.90–0.96) was higher than that of 4C (0.89, 95% CI: 0.84–0.92) (Figure 4; Table 5). These results indicated that C-TIRADS 4B was superior to 4C in detecting malignancy.

To further evaluate the diagnostic performance of C-TIRADS 4B, LR+ was 3.20 (95% CI: 2.28–4.39) and LR- was 0.09 (95% CI: 0.05–0.15) (Table 5). These provided strong evidence for 4B to differentiate malignant nodules.

TABLE 2 Baseline characteristics of included studies.

First author	Total no. of patients	Men	Women	Mean/median age	Interpretation	
					No. of readers	Average reader experience (y)
Cao 2021	355	99	256	49.7 ± 12.4	2	9
Fan 2021	759	144	615	49.0 ± 12.0	2	NA
Gao 2022	208	NA	NA	NA	2	NA
Li 2021	237	46	191	44.9 ± 11.5	2	NA
Li 2022	481	98	383	45.0 ± 10.4	2	8
Lin 2021	120	27	93	47.8 ± 12.4	1	5
Lin 2022	329	113	216	43.5 ± 14.3	2	NA
Qi 2021	884	203	681	49.26	4	5
Qiao 2021	433	82	351	46.6 ± 12.9	2	NA
Sui 2021	70	13	57	48.1 ± 11.5	2	NA
Wu 2021	104	30	74	NA	2	10+
Zhang 2021	408	93	315	NA	NA	NA
Zhang 2022	560	132	428	47.5 ± 13.0	3	5
Zheng 2021	266	70	196	43.9 ± 12.6	4	10
Zhou 2020	2,141	513	1628	50.3 ± 12.0	4	10
Zhu 2021	1,697	361	1336	49.7 ± 12.2	2	15
Total	9,052	2,024	6820	–	–	–

NA, not applicable.

10.3389/fendo.2022.938961

TABLE 3 Characteristics of thyroid nodules included.

First author	Total no. of nodules	Total no. of malignant nodules	Total no. of benign nodules	Median/mean nodule size (range) (cm)	Reference standard		Type of malignant nodule				
					Surgery	Biopsy	PTC	FTC	MTC	Undifferentiated carcinoma	Other
Cao 2021	388	233 (60.1%)	155 (39.9%)	1.39 ± 0.85 (0.4–4.8)	Yes	NA	229 (98.3%)	1 (1.3%)	3 (1.3%)	0	0
Fan 2021	2213	490 (22.1%)	1723 (77.9%)	1.1 ± 0.8	Yes	NA	NA	NA	NA	NA	NA
Gao 2022	251	132 (52.6%)	119 (47.4%)	NA	Yes	NA	126 (95.5%)	0	4 (3.0%)	1 (0.8%)	1 (0.8%)
Li 2021	237	132 (55.7%)	105 (44.3%)	1.42 ± 0.63 (0.3–3.0)	Yes	Yes	NA	NA	NA	NA	NA
Li 2022	513	206 (40.2%)	307 (59.8%)	2.74 ± 1.14	Yes	NA	187 (90.8%)	13 (6.3%)	2 (1.0%)	2 (1.0%)	2 (1.0%)
Lin 2021	123	67 (54.5%)	56 (45.5%)	1.29 ± 1.15 (0.2–7.5)	Yes	Yes	65 (97.0%)	1 (1.5%)	0	0	1 (1.5%)
Lin 2022	329	67 (20.4%)	262 (79.6%)	3.6 ± 1.7	Yes	NA	0	67 (100%)	0	0	0
Qi 2021	1096	414 (37.8%)	682 (62.2%)	1.9 (0.5–6.4)	Yes	Yes	384 (92.8%)	10 (2.4%)	7 (1.7%)	6 (1.5%)	7 (1.7%)
Qiao 2021	433	202 (46.7%)	231 (53.3%)	1.13 ± 0.55	Yes	Yes	NA	NA	NA	NA	NA
Sui 2021	92	50 (54.3%)	42 (45.7%)	1.42 ± 0.98 (0.32–4.1)	Yes	NA	47 (94.0%)	1 (2.0%)	2 (4.0%)	0	0
Wu 2021	104	66 (63.5%)	38 (36.5%)	NA	NA	Yes	NA	NA	NA	NA	NA
Zhang 2021	434	187 (43.1%)	247 (56.9%)	NA	Yes	Yes	NA	NA	NA	NA	NA
Zhang 2022	560	370 (66.1%)	190 (33.9%)	0.5-5.4	Yes	Yes	NA	NA	NA	NA	NA
Zheng 2021	283	211 (74.6%)	72 (25.4%)	NA	Yes	Yes	NA	NA	NA	NA	NA
Zhou 2020	2,141	565 (26.4%)	1,576 (73.6%)	2.33 ± 1.43 (0.23–8.60)	Yes	NA	529 (93.8%)	14 (2.5%)	21 (3.7%)	0	0
Zhu 2021	2,309	891 (38.6%)	1,418 (61.4%)	1.31 ± 1.06 (0.02–6.9)	Yes	Yes	800 (99.1%)	4 (0.5%)	2 (0.2%)	1 (0.1%)	0
Total	**11,506**	**7,223**	**4,283**	–	–	–	–	–	–	–	–

PTC, papillary thyroid carcinoma; FTC, follicular thyroid carcinoma; MTC, medullary thyroid carcinoma.
NA, not applicable.

Evaluation of study heterogeneity

Study heterogeneity was assessed with crosshair plots and sensitivity analysis. Crosshair plots were made to show the scatter of the study results (Figure 5). There were no significant differences among the sensitivity of 16 included studies, while the specificities were quite different from each other with a wide interval. To investigate the influence of a single study on the overall analysis,

TABLE 4 The prevalence of malignancy in each C-TIRADS classification.

Classification	No. of malignant nodules	Total no. of nodules	Prevalence of malignancy (%)	Suggested malignancy risk (%)
C-TIRADS 2	0	370	0%	0%
C-TIRADS 3	31	2,271	1.37%	<2%
C-TIRADS 4A	301	2,834	10.62%	2–10%
C-TIRADS 4B	854	2,134	40.02%	10–50%
C-TIRADS 4C	2,763	3,544	77.96%	50–90%
C-TIRADS 5	334	353	94.62%	>90%

C-TIRADS, Chinese thyroid imaging reports and data systems.

we omitted one study at a time. The omission of any study did not significantly change the corresponding pooled sensitivity, specificity, LR+, LR-, DOR, and AUC (Table S1). Both sensitivity analysis and crosshair plots indicated that our results were robust and reliable.

Discussion

This systematic review is, to our knowledge, the first to consider all available data using a meta-analytic approach, confirmed by a search of database, thus representing the first review of C-TIRADS internationally. We collected and analyzed 16 articles involving a total of 11,506 nodules (7,223 benign, 4,283 malignant) to assess the diagnostic performance of C-TIRADS in malignancy risk stratification of thyroid nodules. We investigated whether the malignancy rate observed in this analysis was consistent with that of the C-TIRADS guideline. Moreover, a series of diagnostic indicators were used to evaluate the performance of C-TIRADS by setting 4B as the cutoff. We believe that this analysis can provide more convincing evidence and support for wider application and deeper understanding of C-TIRADS.

The rate of malignant thyroid nodules was 0% (0%) in C-TIRADS 2, 1.37% (< 2%) in C-TIRADS 3, 10.62% (2–10%) in C-TIRADS 4A, 40.02% (10–50%) in C-TIRADS 4B, 77.96% (50–90%) in C-TIRADS 4C, and 94.61% (> 90%) in C-TIRADS 5 (Table 4). These results compared favorably with the C-TIRADS guideline designation of "likelihood of malignancy" (11). C-TIRADS should be generally considered as an accurate system to stratify the risk of malignancy of thyroid nodules.

Our results show the high accuracy of the C-TIRADS 4B class in the detection of thyroid malignancies. In facts, C-TIRADS 4B detected 94% of malignant nodules while misdiagnosed 30% of benign nodules as suspicious. Similar with 4C and 5 nodules, C-TIRADS 4B nodules do require FNA as recommended by C-TIRADS guideline (11). In those 4B nodules presenting with a

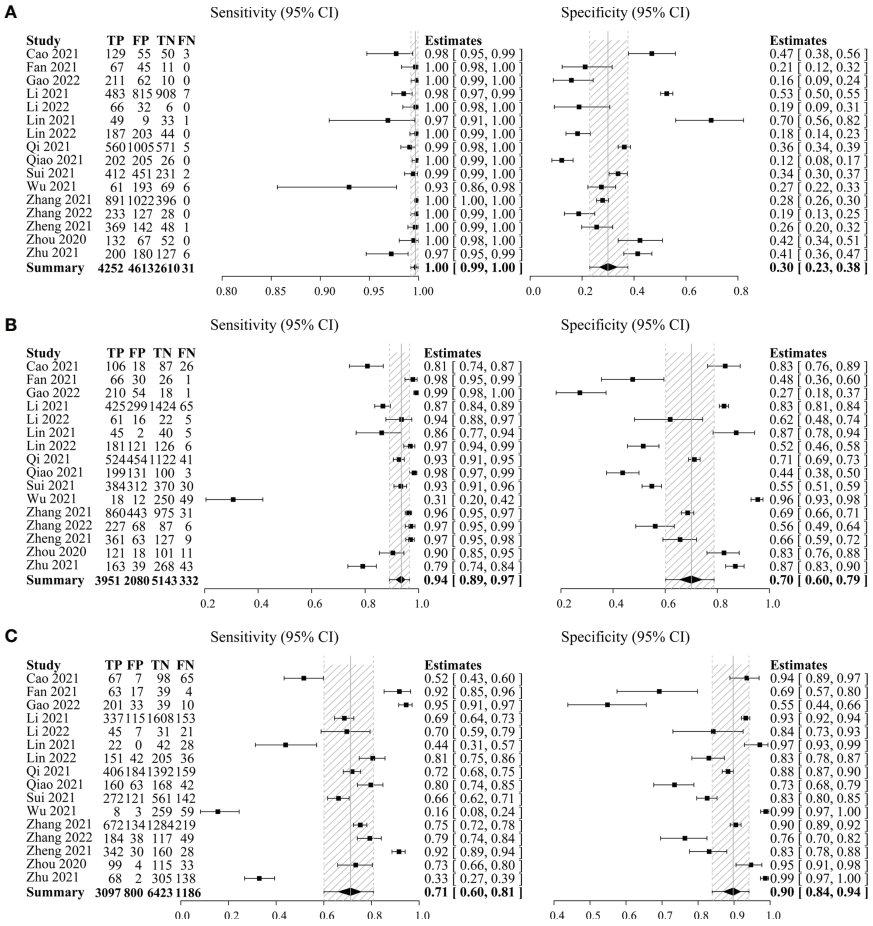


FIGURE 2
Forest plots with individual and pooled sensitivity and specificity of C-TIRADS 4A **(A)**, 4B **(B)**, and 4C **(C)** in the evaluation of thyroid nodules. The estimated accuracy for each study is plotted as a point and the 95% confidence interval (CI) as arrows. TP, true positivity; FP, false positivity; TN, true negativity; FN, false negativity; C-TIRADS, Chinese thyroid imaging reporting and data system.

negative FNA result, a second FNA could be performed to confirm their benign nature. The data obtained in our study may raise the question whether a binary high- *vs.* low-risk stratification of thyroid nodules could be regarded as sufficiently accurate for selecting patients to be referred to FNA and possibly to surgical treatment. With an acceptably low rate (< 2%) of false negative results, C-TIRADS 2 and 3 classes could perhaps both be included in a benign/likely benign single class. Instead, the malignant risk of C-TIRADS 4A (10.62%) is too high to consider the inclusion of nodules belonging to this category in the "benign" subgroup. At the same time, the C-TIRADS 4A class does not qualify a nodule as likely malignant since the risk of malignancy in this category (5–10%) is similar to the one recorded in the general population (3–5). As expected, C-TIRADS 4A included a majority of benign nodules. Hence, given the high proportion of nodules included in this class (2,834/11,506), a substantial burden for the patients and the health care system could be generated if all TIRADS 4A nodules are referred to FNA (and possibly to subsequent surgery). Yet, the frequency of malignant nodules in this class is too high to be neglected. Hence, the management of thyroid nodules classified as 4A should take into account other clinical risk factors such as large size, isthmic or upper lobe location, and positive, family history (33–35), as also recommended in the C-TIRADS guidelines (11). In addition, since the potential of malignancy is higher in iodine-

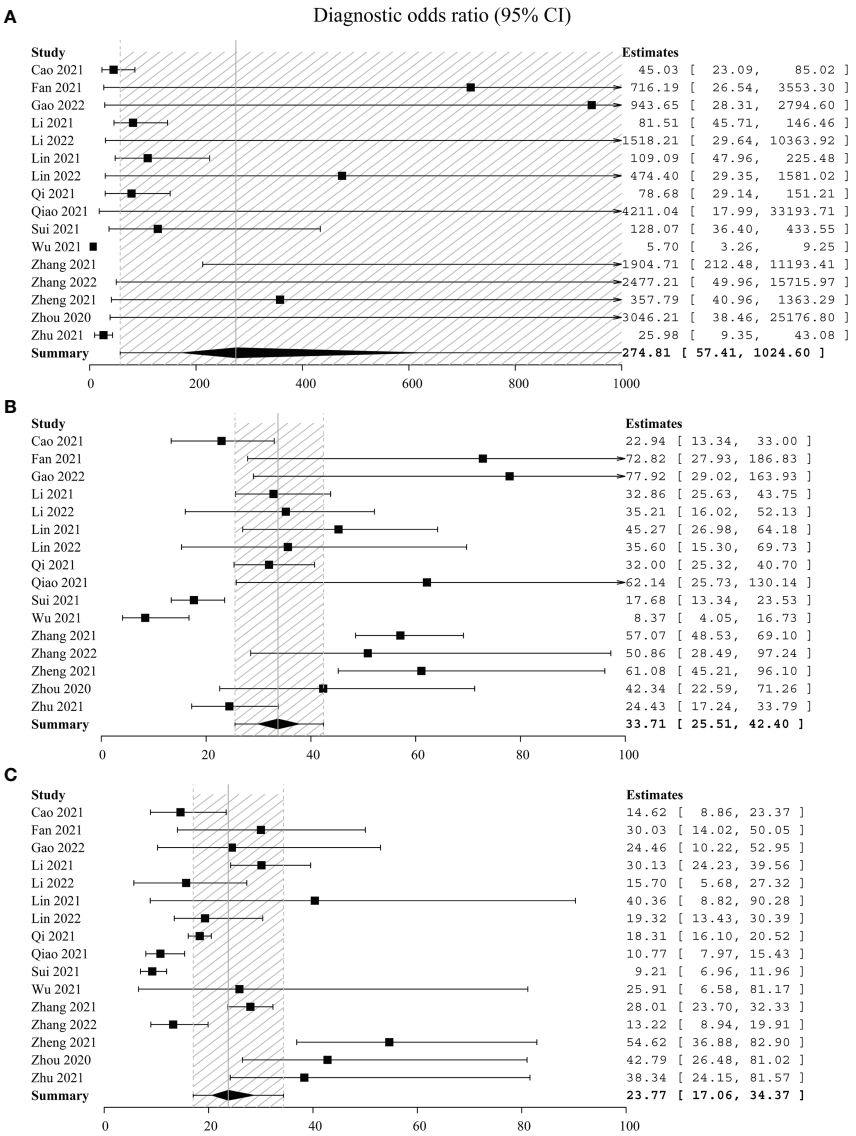


FIGURE 3
Forest plots with individual and pooled DOR of C-TIRADS 4A **(A)**, 4B **(B)**, and 4C **(C)**. The estimated accuracy for each study is plotted as a point and the 95% confidence interval (CI) as arrows. DOR, diagnostic odds ratio; C-TIRADS, Chinese thyroid imaging reporting and data system.

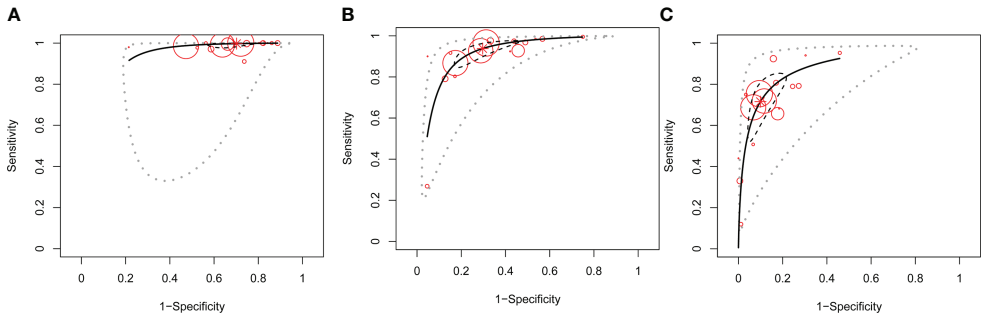


FIGURE 4
SROC of C-TIRADS 4 in detecting malignancy. SROC analysis showing the diagnostic performance of C-TIRADS 4A **(A)**, 4B **(B)**, and 4C **(C)**. The summary point is indicated by the red star; each individual study is represented by red circles (scale = study sample size). The area enclosed by the inner (black line) and outer (gray line) ellipses represents the confidence region and the prediction region of the summary points. SROC, summary receiver operating curve; C-TIRADS, Chinese thyroid imaging reporting and data system.

sufficient areas, the management of C-TIRADS 4A nodules could be determined from region to region based on local iodine sufficiency (or deficiency).

It is obvious that the binary organization of C-TIRADS may not be sufficient to exclude a suspicion of malignancy if a thyroid nodule is diagnosed as C-TIRADS 4A, despite the fact that using C-TIRADS 4B as a cutoff showed excellent diagnostic performance for malignant nodules. Tertiary organization of C-TIRADS, for which C-TIRADS 4A can be considered the intermediate-risk class, may be useful in the management of thyroid nodules. Thyroid nodules classified as intermediate-risk class should be treated more effectively in conjunction with clinical factors, such as more frequent ultrasound surveillance than low-risk stratification (C-TIRADS 2 and 3) and delayed FNA testing than high-risk stratification (C-TIRADS 4B, 4C, and 5). After risk assessment, FNA is the next step in the triage of a thyroid nodule. It should be reserved for lesions that have been determined to be sufficiently suspicious based on C-TIRADS risk stratification. The outcomes are critical in optimizing subsequent management. FNA molecular testing is a new approach that may reduce the need for diagnostic surgery (35). Targeted next-generation sequencing analysis of cancer-related genes for point mutations, gene fusions, copy number alterations, or abnormal gene expression is among the tests developed for this purpose (36). However, molecular testing should unquestionably be taken into consideration if clinical, imaging, and cytology results are insufficient for diagnosis and surgery is the only diagnostic option (5, 37).

This analysis indicates that C-TIRADS performs well in malignancy risk stratification of thyroid nodules and provides more support for appropriate use of FNA recommended by C-TIRADS. Moreover, periodic revisions and updates of C-TIRADS, mainly based on solid evidence and new studies, are necessary to comprehensively reflect the risks and guide FNA. There is no large prospective study evaluating C-TIRADS so far. Further studies are needed to better guide clinical practice.

The diagnostic performance of C-TIRADS was compared with other risk stratification systems in the following 4 publications. Zhou et al. (32) evaluated 2,141 nodules and demonstrated that the diagnostic efficacy of C-TIRADS was significantly greater than that of the American Thyroid Association (ATA) guidelines, the American College of Radiology (ACR) TIRADS, and the Korean TIRADS. Zhu et al. (25) also found that C-TIRADS had better diagnostic performance and a relatively lower unnecessary biopsy rate in detecting thyroid cancer compared to the other three guidelines. On the other hand, the results of Qi et al. (13), which analyzed 3,524 nodules, showed that C-TIRADS had only a little advantage over the ACR TIRADS and the K-TIRADS, and a significant advantage over the EU-TIRADS. This may be due to sample size limitations and bias caused by the fact that not all patients meeting the criteria were included in the study. Furthermore, Zhou et al. (32) found that the EU-TIRADS and ATA guidelines did not apply to 5.1% and 9.9% of nodules, whereas C-TIRADS applied to all nodules.

TABLE 5 Summary estimates of the diagnostic performance of C-TIRADS.

Classification	LR+ (95% CI)	LR- (95% CI)	AUC (95% CI)
C-TIRADS 4A	1.43 (1.28–1.63)	0.01 (0.00–0.03)	1.00 (0.99–1.00)
C-TIRADS 4B	3.20 (2.28–4.39)	0.09 (0.05–0.15)	0.94 (0.90–0.96)
C-TIRADS 4C	7.38 (4.54–12.00)	0.32 (0.20–0.45)	0.88 (0.82–0.92)

C-TIRADS, Chinese thyroid imaging reports and data systems; LR+, positive likelihood ratio; LR-, negative likelihood ratio; AUC, area under curve; 95% CI, 95% confidence interval.

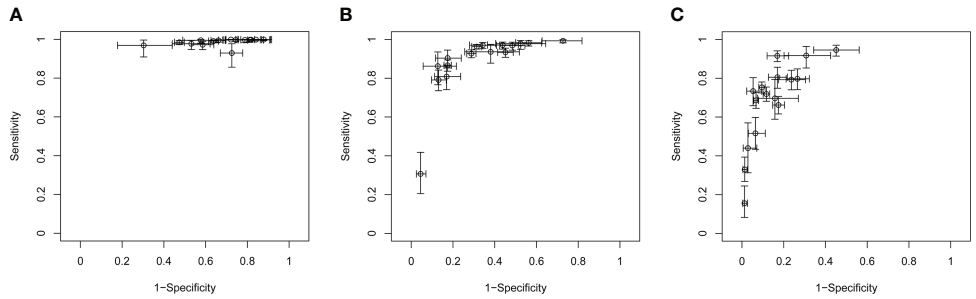


FIGURE 5
Crosshair plots of the sensitivity and specificity. Crosshair plots showing individual sensitivity and specificity of C-TIRADS 4A **(A)**, 4B **(B)**, and 4C **(C)** in the assessment of thyroid nodules. The estimated accuracy for each study is plotted as a circle, and 95% CI is plotted as arrows. C-TIRADS, Chinese thyroid imaging reporting and data system.

There are also several limitations that need to be considered. Firstly, all the studies included were retrospective in nature. There was concern for poor US image quality, and retrospective review may have led to wrong classification. Secondly, the nodule size is another important factor for FNA. However, only two articles reported the nodule size in each classification. Thus, the deviation may affect the risk of FNA in the current study. Thirdly, PTC accounts for more than 90% of current reports. Further research is needed to evaluate the diagnostic performance of C-TIRADS in specific subtypes of thyroid cancer.

Conclusion

In conclusion, C-TIRADS is a good tool for malignancy risk stratification of thyroid nodules. This review provides strong evidence for C-TIRADS 4B in the assessment of malignant thyroid nodules. Further validation of this tool is required.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

WZ conceived the meta-analysis. All authors contributed to the development of the selection criteria, the risk of bias assessment strategy, and data extraction criteria. YH developed the search strategy. YH and SX performed the database search, acquired the data, and analyzed the data. YH and WZ drafted the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The study was funded by the National Natural Science Foundation of China (82071923).

Acknowledgments

The authors would like to thank their colleagues and institutions for their great assistance in this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary Material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2022.938961/full#supplementary-material

References

1. Burman KD, Wartofsky L. Clinical practice. thyroid nodules. *New Engl J Med* (2015) 373(24):2347–56. doi: 10.1056/NEJMcp1415786

2. Durante C, Grani G, Lamartina L, Filetti S, Mandel SJ, Cooper DS. The diagnosis and management of thyroid nodules: A review. *JAMA* (2018) 319 (9):914–24. doi: 10.1001/jama.2018.0898

3. Hegedüs L. Clinical practice. the thyroid nodule. *New Engl J Med* (2004) 351 (17):1764–71. doi: 10.1056/NEJMcp031436

4. Giovanella L, Deandreis D, Vrachimis A, Campenni A, Petranovic Ovcaricek P. Molecular imaging and theragnostics of thyroid cancers. *Cancers (Basel)* (2022) 14(5):1272. doi: 10.3390/cancers14051272

5. Kobaly K, Kim CS, Mandel SJ. Contemporary management of thyroid nodules. *Annu Rev Med* (2022) 73:517–28. doi: 10.1146/annurev-med-042220-015032

6. Hanba C, Khariwala SS. What is the utility of genetic testing in indeterminate thyroid nodules? *Laryngoscope* (2021) 131(11):2399–400. doi: 10.1002/lary.29435

7. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. Acr thyroid imaging, reporting and data system (Ti-rads): White paper of the acr Ti-rads committee. *J Am Coll Radiol* (2017) 14(5):587–95. doi: 10.1016/j.jacr.2017.01.046

8. Shin JH, Baek JH, Chung J, Ha EJ, Kim J-H, Lee YH, et al. Ultrasonography diagnosis and imaging-based management of thyroid nodules: Revised Korean society of thyroid radiology consensus statement and recommendations. *Korean J Radiol* (2016) 17(3):370–95. doi: 10.3348/kjr.2016.17.3.370

9. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European Thyroid association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: The eu-tirads. *Eur Thyroid J* (2017) 6(5):225–37. doi: 10.1159/000478927

10. Kwak JY, Han KH, Yoon JH, Moon HJ, Son EJ, Park SH, et al. Thyroid imaging reporting and data system for us features of nodules: A step in establishing better stratification of cancer risk. *Radiology* (2011) 260(3):892–9. doi: 10.1148/radiol.11110206

11. Zhou J, Yin L, Wei X, Zhang S, Song Y, Luo B, et al. Chinese Guidelines for ultrasound malignancy risk stratification of thyroid nodules: The c-tirads. *Endocrine* (2020) 70(2):256–79. doi: 10.1007/s12020-020-02441-y

12. Hu L, Liu X, Pei C, Xie L, He N. Assessment of perinodular stiffness in differentiating malignant from benign thyroid nodules. *Endocr Connect* (2021) 10 (5):492–501. doi: 10.1530/ec-21-0034

13. Qi Q, Zhou A, Guo S, Huang X, Chen S, Li Y, et al. Explore the diagnostic efficiency of Chinese thyroid imaging reporting and data systems by comparing with the other four systems (Acr Ti-rads, kwak-tirads, ksthr-tirads, and eu-tirads): A single-center study. *Front Endocrinol* (2021) 12:763897. doi: 10.3389/fendo.2021.763897

14. Cao H, Fan Q, Zhuo S, Qi T, Sun H, Rong X, et al. The value of Chinese thyroid imaging report and data system combined with contrast-enhanced ultrasound scoring in differential diagnosis of benign and malignant thyroid nodules. *J Ultrasound Med Off J Am Institute Ultrasound Med* (2021) 9999:1–9. doi: 10.1002/jum.15858

15. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The prisma 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* (2021) 372:n71. doi: 10.1136/bmj.n71

16. Yang B, Mallett S, Takwoingi Y, Davenport CF, Hyde CJ, Whiting PF, et al. Quadas-c: A tool for assessing risk of bias in comparative diagnostic accuracy studies. *Ann Intern Med* (2021) 174(11):1592–9. doi: 10.7326/M21-2234

17. Cibas ES, Ali SZ. The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid* (2017) 27(11):1341–6. doi: 10.1089/thy.2017.0500

18. Guo J, Riebler A, Rue H. Bayesian Bivariate meta-analysis of diagnostic test studies with interpretable priors. *Stat Med* (2017) 36(19):3039–58. doi: 10.1002/sim.7313

19. Fan L, Xue G, Yang F, Wang Y, Wu X. Value of ultrasound c-tirads in malignant risk stratification of thyroid nodules. *Chin Remedies Clinics* (2021) 21 (21):3546–7. doi: 10.11655/zgywylc2021.21.015

20. Li Q, Ding S, Guo W, Liu Y, Wei Y, Ding Q. Evaluation of the efficacy of c-tirads combined with artificial intelligence assisted diagnosis in thyroid nodule differential diagnosis. *Chin J Ultrasonography* (2021) 30(3):231–5. doi: 10.3760/cma.j.cn131148-20201106-00858

21. Lin W, Ding N, Zhu S. Application value of c-tirads combined with elastography in predicting benign and malignant thyroid nodules. *J Nanjing Med Univ* (2021) 41(9):1388–91. doi: 10.7655/nydxbns20210920

22. Zhang G, Zhang L, Hou Y, Chen Y, Weiwei Z. Application value of ultrasound-guided fine needle aspiration biopsy in diagnosis of thyroid nodules in juveniles. *Chin J Med Ultrasound(Electronic Ed)* (2021) 18(9):822–7. doi: 10.3877/cma.j.issn.1672-6448.2021.09.002

23. Zheng L, Li S, Xu L, Zhou L, Yu C, Ma S. Diagnostic performance of ultrasound-based risk stratification systems for thyroid nodules: Comparison of the c-tirads with the acr-tirads and eu-tirads. *Chin J Ultrasonography* (2021) 30 (9):785–91. doi: 10.3760/cma.j.cn131148-20210219-00113

24. Lin Y, Lai S, Wang P, Li J, Chen Z, Wang L, et al. Performance of current ultrasound-based malignancy risk stratification systems for thyroid nodules in patients with follicular neoplasms. *Eur Radiol* (2022) 23(6):3617–30. doi: 10.1007/s00330-021-08450-3

25. Zhu H, Yang Y, Wu S, Chen K, Luo H, Huang J. Diagnostic performance of us-based fnab criteria of the 2020 Chinese guideline for malignant thyroid nodules: Comparison with the 2017 American college of radiology guideline, the 2015 American thyroid association guideline, and the 2016 Korean thyroid association guideline. *Quantitative Imaging Med Surg* (2021) 11(8):3604–18. doi: 10.21037/qims-20-1365

26. Sui Y, Wang B. Two-dimensional Gray scale ultrasound combined with superb microvascular imaging in differential diagnosis of Ti?Rads 3 and 4 thyroid nodules. *Acad J Guangzhou Med Coll* (2021) 49(5):78–82. doi: 10.3969/j.issn.2095-9664.2021.05.17

27. Qiao M, Feng S, Shen D, Xia B. Chinese Thyroid imaging reporting and data system in the differential diagnosis of benign and malignant thyroid nodules. *Chin J Med Imaging* (2021) 29(11):1070–5. doi: 10.3969/j.issn.1005-5185.2021.11.003

28. Wu D, Wang Y. Comparison of the diagnostic efficacy of thyroid ultrasound acr-tirads and c-tirads. *Pharm Weekly* (2021) 30(15):82–4. Available at: http://www.ydzkzzs.net/2115pdf/211580.pdf

29. Gao Y, Deng D, Liu Y, Gan L, Zhang C. Application value of 2020 c-tirads in the diagnosis and follow-up management of thyroid nodules. *Acta Universitatis Med Anhui* (2022) 57(03):497–9. doi: 10.19405/j.cnki.issn1000-1492.2022.03.031

30. Zhang W, Chen T, Liu H, He B, Mao L, Liu Y, et al. Comparisons of the five different ultrasound malignancy risk stratification system of thyroid nodules. *Chin J Ultrasound Med* (2022) 38(02):132–6. doi: 10.3969/j.issn.1002-0101.2022.02.004

31. Li R, Li S. Value of 2020 Chinese Ti -rads in the diagnosis of benign and malignant thyroid nodules. *Chin Imaging J Integrated Traditional Western Med* (2022) 20(02):154–7. doi: 10.3969/j.issn.1672-0512.2022.02.012

32. Zhou J, Song Y, Zhan W, Wei X, Zhang S, Zhang R, et al. Thyroid imaging reporting and data system (Tirads) for ultrasound features of nodules: Multicentric retrospective study in China. *Endocrine* (2020) 72(1):157–70. doi: 10.1007/s12020-020-02442

33. Pappa T, Ahmadi S, Bikas A, Hwang S, Coleman A, Lobon I, et al. Thyroid nodule shape independently predicts risk of malignancy. *J Clin Endocrinol Metab* (2022) 107(7):1865–70. doi: 10.1210/clinem/dgac246

34. Russ G, Trimboli P, Buffet C. The new era of tiradss to stratify the risk of malignancy of thyroid nodules: Strengths, weaknesses and pitfalls. *Cancers (Basel)* (2021) 13(17):4316. doi: 10.3390/cancers13174316

35. Grani G, Sponziello M, Pecce V, Ramundo V, Durante C. Contemporary thyroid nodule evaluation and management. *J Clin Endocrinol Metab* (2020) 105 (9):2869–83. doi: 10.1210/clinem/dgaa322

36. Rossi ED, Locantore P, Bruno C, Dell'Aquila M, Tralongo P, Curatolo M, et al. Molecular characterization of thyroid follicular lesions in the era of "Next-generation" techniques. *Front Endocrinol (Lausanne)* (2022) 13:834456. doi: 10.3389/fendo.2022.834456

37. Livhits MJ, Zhu CY, Kuo EJ, Nguyen DT, Kim J, Tseng C-H, et al. Effectiveness of molecular testing techniques for diagnosis of indeterminate thyroid nodules: A randomized clinical trial. *JAMA Oncol* (2021) 7(1):70–7. doi: 10.1001/jamaoncol.2020.5935

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 20 September 2022
DOI 10.3389/fendo.2022.966572

OPEN ACCESS

EDITED BY
Terry Francis Davies,
Icahn School of Medicine at Mount Sinai, United States

REVIEWED BY
Laura Giacomelli,
Sapienza University of Rome, Italy
Andrea Frasoldati,
Endocrine Unit ASMN, Italy

*CORRESPONDENCE
Huan Zhang
huanzhangy@126.com
Weiwei Zhan
shanghairuijin@126.com
Wei Zhou
zw11468@126.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Thyroid Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 11 June 2022
ACCEPTED 06 September 2022
PUBLISHED 20 September 2022

CITATION
Xu S, Ni X, Zhou W, Zhan W and Zhang H (2022) Development and validation of a novel diagnostic tool for predicting the malignancy probability of thyroid nodules: A retrospective study based on clinical, B-mode, color doppler and elastographic ultrasonographic characteristics. *Front. Endocrinol.* 13:966572. doi: 10.3389/fendo.2022.966572

Development and validation of a novel diagnostic tool for predicting the malignancy probability of thyroid nodules: A retrospective study based on clinical, B-mode, color doppler and elastographic ultrasonographic characteristics

Shangyan Xu[1†], Xiaofeng Ni[1†], Wei Zhou[1*], Weiwei Zhan[1*] and Huan Zhang[2*]

[1]Department of Ultrasound, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [2]Department of Radiology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Background: Clinicians estimate the risk of thyroid nodules and make subsequently decision on the basis of clinical and ultrasonographic findings. Currently, there is no comprehensive diagnostic tool for predicting the malignancy rates of thyroid nodules. Our aim was to develop and validate a novel integrate diagnostic tool for predicting the malignancy probability of thyroid nodules based on clinical, B-mode, Color Doppler and elastographic ultrasonographic characteristics.

Methods: A total of 1016 nodules in 1016 patients who underwent thyroid ultrasonography and surgery from July 2021 to December 2021 were included in this retrospective study. All nodules were confirmed by pathology and randomly classified into the training and validation groups. Clinical, B-mode, Color Doppler and elastographic (CBCE) ultrasonographic characteristics of nodules were recorded. Univariate and multivariate analyses were performed to screen independent predictors associated with thyroid cancer. A multivariate model containing the extracted predictors was constructed and presented in the form of a nomogram. The validation and applicability of the CBCE nomogram was evaluated using the receiver operating characteristic (ROC) curve. Diagnostic performances were calculated to compare the CBCE nomogram with ACR-TIRADS (Thyroid Imaging Reporting Data System by American College of Radiology) and EU-TIRADS (Thyroid Imaging Reporting Data System by European Thyroid Association).

Results: The following factors were included in the CBCE nomogram: patient gender, age, shape, margin, composition and echogenicity, calcification, vascularization distribution, vascularization degree, suspicious lymph node metastases and elastography. The area under the curve (AUC) values were 0.978 and 0.983 for the training and validation groups, respectively. Compared with ACR-TIRADS and EU-TIRADS, the CBCE nomogram showed improved accuracy (0.944) and specificity (0.913) without sacrificing sensitivity (0.963) and showed the highest AUC with an optimal cutoff value of 0.55.

Conclusion: The CBCE nomogram has good and high clinical practicability in predicting the malignancy probability of thyroid nodules.

KEYWORDS

thyroid nodule, ultrasonography, clinical, risk factor, nomogram

Introduction

Thyroid nodules (TNs) are very common (1, 2). There are several factors that can affect their prevalence, including demographic characteristics, iodine sufficiency status and the increasing use of ultrasound (US) examination (3–5). In recent years, overdiagnosis and overtreatment of TNs have become a global problem (6, 7). Correct differentiation between low-risk and high-risk TNs is a crucial starting point for optimal treatment (4). Ultrasonography, as a radiation-free and non-invasive method, is the first-line approach for thyroid examination. The US-based risk stratification systems (RSSs) of TNs, such as ACR-TIRADS (Thyroid Imaging Reporting Data System by American College of Radiology) (8) and EU-TIRADS (Thyroid Imaging Reporting Data System by European Thyroid Association) (9), have remained a research focus for nearly a decade. Numerous studies have verified and compared their diagnostic performances, showing good values in clinical practice but also problems of complex use and weak consistency (10–15).

Moreover, clinicians estimate the risk of TNs and make subsequent decisions based on comprehensive information including clinical and US findings (3, 16, 17). The existing RSSs are limited to B-mode US features and do not address clinical, Color Doppler and elastographic US characteristics. Previous studies have shown that there were gender and age differences in patients with thyroid cancer (TC), but this is still controversial in recent years (16, 18). In addition, routine TSH (Thyroid Stimulating Hormone) measurement is recommended by the ATA (American Thyroid Association) guideline which not only aimed to exclude hyperthyroidism but to better stratify the risk of malignancy as well, since higher serum TSH levels have been correlated with an increased risk of malignancy (19). Furthermore, vascularity information and elastography techniques have been seen as complementary imaging modalities for the diagnosis of TNs (20). However, there is no integrated diagnostic tool for TNs currently.

This research aims to develop and validate a comprehensive and easy-to-use diagnostic tool for TNs based on clinical, B-mode, Color Doppler and elastographic US characteristics, providing more information for clinicians as well as avoiding overdiagnosis and overtreatment.

Abbreviations: AUC, Area Under the Curve; ATA, American Thyroid Association; ACR-TIRADS, Thyroid Imaging Reporting Data System by American College of Radiology; BMI, Body Mass Index; EU-TIRADS, Thyroid Imaging Reporting Data System by European Thyroid Association; KTA-TIRADS, Thyroid Imaging Reporting Data System by Korean Thyroid Association; ROC, Receiver Operating Characteristic; RSS, Risk Stratification System; TSH, Thyroid Stimulating Hormone; TN, Thyroid Nodule; TC, Thyroid Cancer; US, Ultrasound.

Materials and methods

Patients selection

This retrospective study was approved by the Ethics Committee of Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, and the requirement for written informed consent was waived. Between July 2021 and December 2021, a total of 1085 thyoid nodules from 1085 consecutive patients in our hospital were included. Diagnostic thyroid ultrasonography and surgery were performed on all nodules to obtain a definitive pathological diagnosis. Among them, 69 nodules were excluded. The exclusion criteria were as follows (1): patients with insufficient demographic and laboratory data including BMI (Body Mass Index) and preoperative TSH level (n=33) (2); patients with a history of thyroid surgery (n=11) (3); nodules with inadequate

sonographic images including unsatisfied vascularity and elastographic images (n=17) (4); nodules with borderline types of pathological diagnoses including NIFTP (Noninvasive follicular thyroid neoplasm with papillary-like nuclear features) and diagnosis of undetermined malignant potential. (n=8). Ultimately, a total of 1016 nodules from 1016 patients were included. All nodules were randomly divided into the training group (n=712) and the validation group (n=304) at a ratio of 7:3. The training group was used to build the nomogram, while the validation group was used to evaluate the diagnostic performance of this tool. The study flow chart is shown in Figure 1.

Clinical data acquisition

Clinical data were gathered by searching the medical records of patients gender, age, BMI, residence, family history of TC and preoperative TSH level. Residences were divided into inland areas and coastal areas. Family history of TC was defined as a first-degree relative with a history of TC. Preoperative TSH level was defined as the test results within one week before surgery. The normal range of TSH levels in our institution is 0.35–4.94 μIU/ml.

Ultrasound image acquisition

Two radiologists (S.Y.X and X.F.N) with over ten years of experience in thyroid sonography used a 4-15 MHz linear probe (MyLab9, Esaote, Italy) to perform all grayscale, color Doppler and elastography sonographic examinations. Images in the longitudinal and transverse directions of each target nodule were obtained. A picture archive and communication system recorded and uploaded all the images for later retrospective analysis. The grayscale, color Doppler and elastography sonographic features of the target nodules were assessed by two radiologists (S.Y.X and X.F.N) with professional training in thyroid ultrasound in consensus. In the case of disagreement between the two radiologists, the final decision was made by a third radiologist (W.Z) with 20 years of experience in thyroid ultrasound.

The B-mode, color Doppler US features of TNs included shape (wider-than-tall, taller-than-wide), margin (regular, irregular), composition and echogenicity (non-solid hypoechoic, solid hypoechoic), calcification (non-microcalcification, microcalcification), vascularization distribution (internal, non-internal), vascularization (low, high), diffuse lesion (absent, present) and suspicious lymph node metastases (LNM) (absent, present) in the cervical compartment. The vascularization distribution is divided into types 1-4 (type 4, marked intranodular vascularity with or without perinodular vascularity (vascularity≥50%); type 3: mild intranodular vascularity with or without perinodular vascularity (vascularity<50%); type 2: perinodular vascularity only; type 1: no vascularity) (20). In our study, types 3 and 4 were classified as the group with internal vascularity, while types 1 and 2 were classified as the group with non-internal vascularity. The vascularity greater than or similar to that of the surrounding thyroid tissue indicated a high vascularization degree of a nodule (21). The rest of the nodules were relatively low. Diffuse lesions referred to a thyroid US

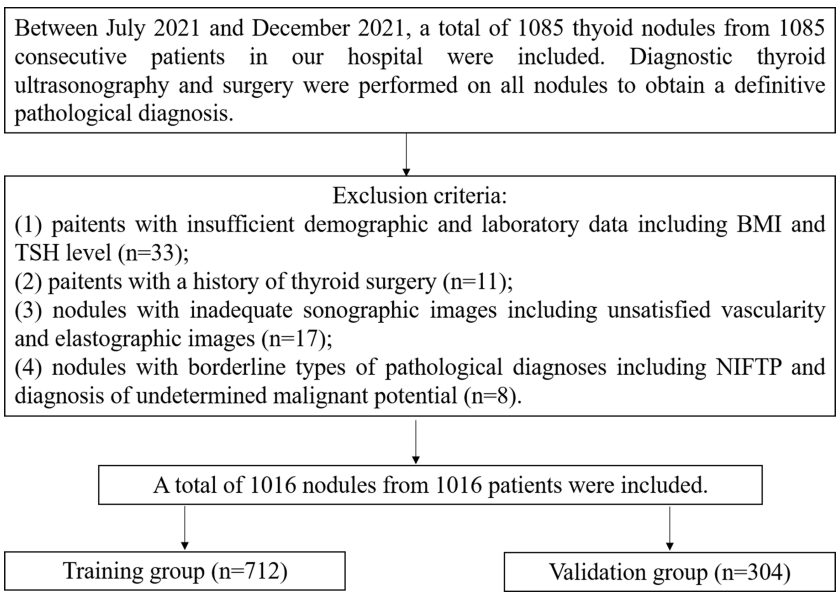


FIGURE 1
The study flow chart.

appearance consistent with a diffuse disease (e.g. chronic thyroiditis or lymphoma). Ultrasonographic features of metastatic lymph nodes included the absence of a hilum, a Solbiati index <2, peripheral vascular flow, hyperechoic, calcifications, and cystic (22). Based on the 2013 European Thyroid Association Guidelines for cervical ultrasound scan (23), cervical LNs can be classified into three groups: normal, indeterminate and suspicious for malignancy. In our study, we classified the latter two as the group with suspicious LNM and the rest as the group with no suspicious LNM. Elastographic US features were classified as soft and hard by the strain elastography technique according to the Asteria criteria (24). Asteria criteria have four scores of tissue stiffness: 1 and 2 for soft nodules and 3 and 4 for hard lesions.

Statistical analysis

Continuous variables are expressed as the mean ± SD. The variances were equal or unequal for continuous data using Student's t test. The Mann–Whitney U test was used for continuous data with a non-normal distribution. Categorical variables were expressed as numbers (%), and categorical data were evaluated by Fisher's exact test and Pearson's $\chi 2$. Univariate and multivariate analyses were performed to screen independent predictors associated with TC. A multivariate logistic model containing the extracted predictors was constructed and presented in the form of a nomogram.

To measure the accuracy of the nomogram, the Hosmer–Lemeshow goodness-of-fit test was applied. To validate the nomogram, two ROC curves were produced for the training and validation groups. Diagnostic values were calculated to assess the applicability of the nomogram and to compare it with ACR-TIRADS (8) and EU-TIRADS (9). Both these two RSSs are divided into five grades according to different US characteristics, which classified as TR1-5 for ACR-TIRADS and EU-TIRADS 1-5 for EU-TIRADS. The comparasion of diagnostic performances for the CBCE nomogram, ACR-TIRADS and EU-TIRADS were based on the AUC, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy. The analysis was performed using the SPSS software (version 23; IBM Corp., Armonk, NY, USA) and R software (version 4.1.0; R Development Core Team, Vienna, Austria). A P value < 0.05 was considered for significant differences.

Results

Baseline for patients with thyroid nodules

Among the 1016 thyroid nodules, 374 were benign and 642 were malignant. The average maximum diameters were 34.15 ± 17.16 mm for benign nodules and 10.23 ± 7.38 mm for malignant nodules. Follicular adenoma (n=22), Hashimoto thyroiditis (n=39), and nodular goiter (n=313) were benign lesions. Malignant lesions included follicular thyroid carcinoma (n=10), medullary thyroid carcinoma (n=15), and papillary thyroid carcinoma (n=617).

Univariate and multivariate analyses of risk factors for patients with thyroid nodules

The univariate and multivariate analyses results were showed in Table 1 and Table 2. Except for the three indicators of BMI, family history of TC and TSH, the remaining indicators showed significant differences in the training group (P < 0.05). Table 2 showed the extracted independent predictors included gender, age, shape, margin, composition and echogenicity, calcification, vascularization distribution, vascularization degree, and suspicious LNM and elastography based on the training group (P < 0.05). These ten features were included into the development of the nomogram associated with TC.

Development and validation of the CBCE nomogram

Figure 2 presented the multivariate logistic model as a nomogram with the ten independent predictors. The risk score ranged from a minimum of 0 to a maximum of 700. The corresponding malignancy probability of each nodule was obtained by adding up the specific point of each predictor. From Figure 3, the nomogram calibration curve showed floating around the baseline by the Hosmer–Lemeshow goodness-of-fit test. Figure 4 presented two ROC curves of the training group (A) and the validation group (B). The AUCs for the two groups were 0.978 (95% CI [0.967,0.989]) and 0.983 (95% CI [0.971,0.994]), respectively.

Diagnostic performances of the CBCE nomogram

Figure 5 showed the CBCE nomogram showed the highest AUC of 0.983 in comparison to ACR-TIRADS (AUC of 0.948) and EU-TIRADS (AUC of 0.889) in the validation group. The optimum cutoff values of the three models were 0.55, TR4 and EU-TIRADS 4, respectively. The numbers of cases according to the cut-off value in the three models are listed in Table 3. Table 4 showed the comparison of diagnostic performances among the CBCE nomogram, ACR-TIRADS and EU-TIRADS

10.3389/fendo.2022.966572

TABLE 1 Univariate analysis of risk factors for thyroid cancer in the training group.

Factors	Training group (n=712)		
	Benign (n=259)	Malignant (n=453)	*P* value
Gender (%)			<0.001
Female	225 (86.9)	282 (62.3)	
Male	34 (13.1)	171 (37.7)	
Age (mean ± SD)	49.25 ± 13.06	41.66 ± 10.73	<0.001
Residence (%)			<0.001
inland area	66 (25.5)	204 (45.0)	
coastal area	193 (74.5)	249 (55.0)	
BMI (mean ± SD)	23.20 ± 3.08	23.52 ± 3.62	0.216
Family History of Thyroid Cancer (%)			0.141
No	238 (91.9)	429 (94.7)	
Yes	21 (8.1)	24 (5.3)	
TSH (mean ± SD), μIU/mL	1.55 ± 1.40	1.85 ± 1.25	0.103
Shape (%)			<0.001
wider-than-tall	242 (93.4)	181 (40.0)	
taller-than-wide	17 (6.6)	272 (60.0)	
Margin (%)			<0.001
regular	234 (90.3)	24 (5.3)	
irregular	25 (9.7)	429 (94.7)	
Composition and Echogenicity (%)			<0.001
non-solid hypoechoic	189 (73.0)	16 (3.5)	
solid hypoechoic	70 (27.0)	437 (96.5)	
Calcification (%)			<0.001
non-microcalcification	227 (87.6)	157 (34.7)	
microcalcification	32 (12.4)	296 (65.3)	
Vascularization Distribution (%)			0.007
non-internal	149 (57.5)	213 (47.0)	
internal	110 (42.5)	240 (53.0)	
Vascularization Degree (%)			<0.001
low	136 (52.5)	373 (82.3)	
high	123 (47.5)	80 (17.7)	
With Diffuse Lesion (%)			<0.001
No	233 (90.0)	356 (78.6)	
Yes	26 (10.0)	97 (21.4)	
Suspicious LNM (%)			<0.001
No	256 (98.8)	312 (68.9)	
Yes	3 (1.2)	141 (31.1)	
Elastography (%)			<0.001
soft	188 (72.6)	45 (9.9)	
hard	71 (17.4)	408 (91.1)	

SD, standard deviation; TSH, Thyroid Stimulating Hormone; LNM, lymph node metastases.

based on the validation group, including sensitivity (96.3% vs 98.4% vs 97.4%), specificity (91.3% vs 62.6% vs 64.3%), PPV (94.8% vs 81.2% vs 81.8%), NPV (93.8% vs 96.0% vs 93.7%), and accuracy (94.4% vs 84.9% vs 84.9%). The CBCE nomogram showed the highest specificity, PPV and accuracy, while ACR-TIRADS and EU-TIRADS showed slightly higher sensitivity and NPV.

Clinical application of the CBCE nomogram

Figure 6 illustrates a typical clinical application of the CBCE nomogram. Images were obtained from a 51-year-old woman with a nodule in the left thyroid. The nomogram scored 0 for woman, 38 for age, 0 for wider than taller shape, 97 for irregular margin, 53

10.3389/fendo.2022.966572

TABLE 2 Multivariate Analysis of risk factors for thyroid cancer in the training group.

Risk Factors	Value assignment	B	Standard Error	*P* Value	OR	OR (95% CI)
Gender	0=Female, 1=Male	1.126	0.449	0.012	3.084	1.280, 7.433
Age	continuous variables	-0.028	0.015	0.049	0.973	0.944, 1.002
Shape	0=wider-than-tall, 1=taller-than-wide	1.462	0.450	0.001	4.314	1.785, 10.426
Margin	0=regular, 1=irregular	2.457	0.422	0.000	11.666	5.099, 26.688
Composition and Echogenicity	0=non-solid hypoechoic, 1=solid hypoechoic	1.353	0.517	0.009	3.870	1.405, 10.661
Calcification	0=non-microcalcification, 1=microcalcification	1.406	0.378	0.000	4.079	1.946, 8.551
Vascularization Distribution	0=non-internal, 1=internal	1.267	0.506	0.012	3.551	1.319, 9.566
Vascularization Degree	0=low, 1=high	-1.739	0.529	0.001	0.176	0.062, 0.496
Suspicious LNM	0=No, 1=Yes	2.543	0.843	0.003	12.714	2.436, 66.352
Elastography	0=soft, 1=hard	1.822	0.431	0.000	6.184	2.655, 14.404

LNM, lymph node metastases; OR, Odds Ratio; CI, Confidence Interval.

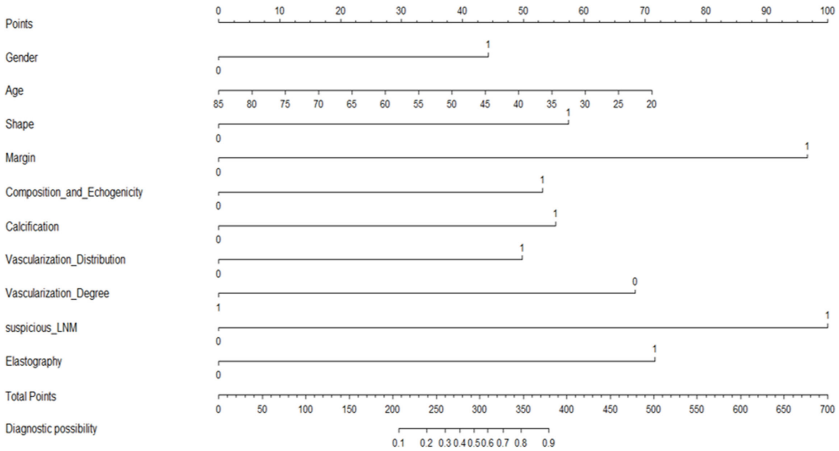


FIGURE 2
A CBCE nomogram to predict the malignancy probability of thyroid nodules based on clinical, B-mode, Color Doppler and elastographic ultrasonographic characteristics.

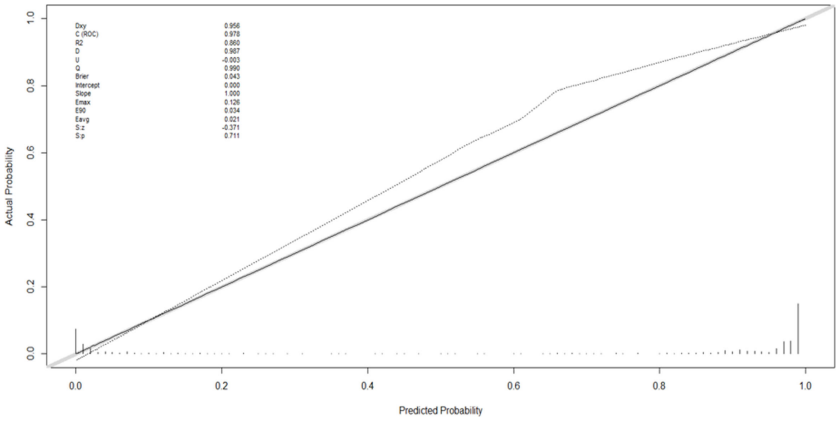


FIGURE 3
The Hosmer-Lemeshow goodness-of-fit test for the CBCE nomogram.

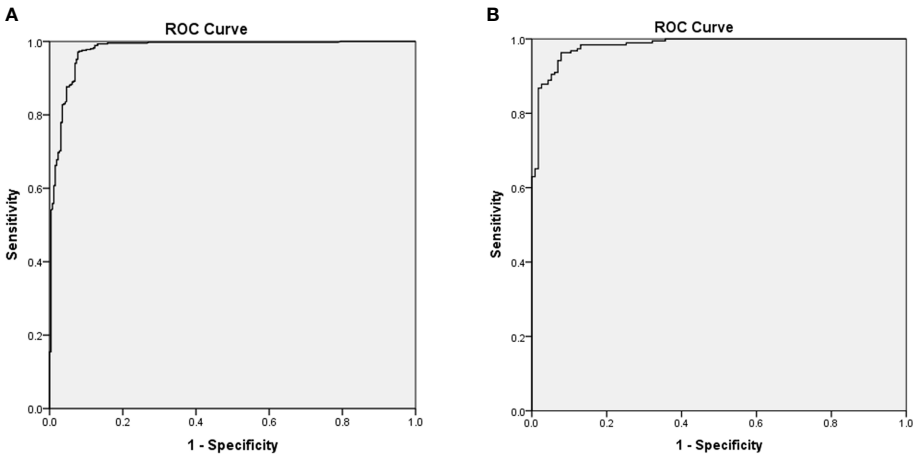


FIGURE 4
Receiver operating characteristic curves of the training group **(A)** and the validation group **(B)** based on the pathological diagnosis. The area under the curves for the two groups were 0.978 (0.967,0.989) and 0.983 (0.971,0.994), respectively.

for solid hypoechoic, 0 for non-microcalcification, 50 for internal vascularization distribution, 69 for low vascularization degree, 0 for no suspicious LNM, and 72 for hard strain elastography, resulting in a total score of 379 points. The corresponding malignancy rate of the nodule was high (>0.90), and the pathology of the nodule was papillary thyroid carcinoma.

Discussion

In this study, a novel integrated diagnostic tool called the CBCE nomogram was developed and validated for predicting the malignancy probability of TNs based on clinical, B-mode, Color Doppler and elastographic US characteristics. The CBCE

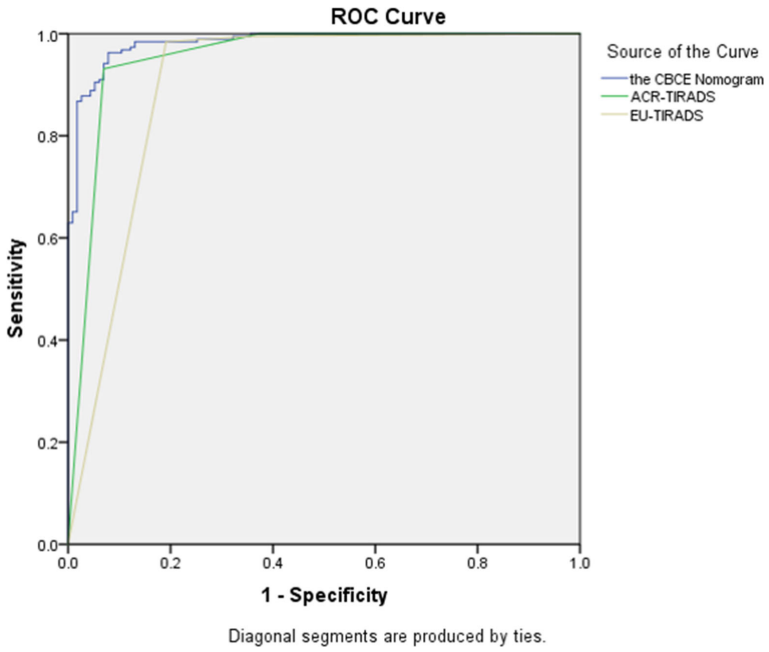


FIGURE 5
The comparison of Receiver operating characteristic curves among the CBCE nomogram, ACR-TIRADS and EU-TIRADS based on the validation group. The area under the curves for the three models were 0.983 (0.971, 0.994), 0.948 (0.918, 0.978) and 0.889 (0.844, 0.934), respectively.

TABLE 3 Numbers of cases according to the cut-off values in the three models based on the validation group.

Diagnostic Models		Malignant	Benign	Total
the CBCE Nomogram	≥0.55	182	10	192
	<0.55	7	105	112
ACR-TIRADS	TR4-5	186	43	229
	TR1-3	3	72	75
EU-TIRADS	EU-TIRADS 4-5	184	41	225
	EU-TIRADS 1-3	5	74	79
Total		189	115	304

TABLE 4 Diagnostic performances of the CBCE Nomogram, ACR-TIRADS and EU-TIRADS in the validation group.

Diagnostic Models	the cut-off values	AUC (95% CI)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)
the CBCE Nomogram	0.55	0.983 (0.971, 0.994)	0.963 (182/189)	0.913 (105/115)	0.948 (182/192)	0.938 (105/112)	0.944 (287/304)
ACR-TIRADS	TR4	0.948 (0.918, 0.978)	0.984(186/189)	0.626 (72/115)	0.812 (186/229)	0.960 (72/75)	0.849 (258/304)
EU-TIRADS	EU-TIRADS 4	0.889 (0.844, 0.934)	0.974 (184/189)	0.643 (74/115)	0.818 (184/225)	0.937(74/79)	0.849 (258/304)

AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value.

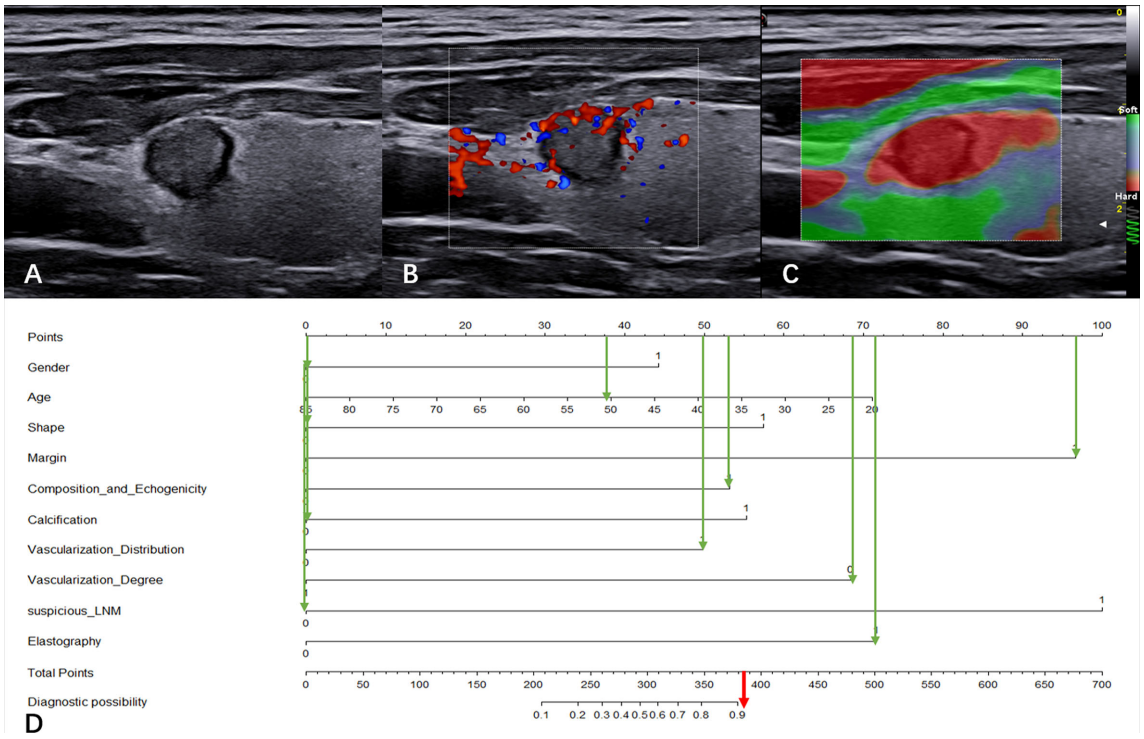


FIGURE 6
Clinical application of the CBCE nomogram. Images from a 51-year-old woman with a nodule (9.1×7.8×8.2 mm) in the left thyroid and the pathology was papillary thyroid carcinoma. **(A)** Greyscale ultrasound imaging of the mass. The lesion showed a wider-than-taller shape, irregular margins and solid hypoechoicity. **(B)** Internal vascularization distribution and low vascularization degree of the nodule in color Doppler. **(C)** Strain elasticity of the nodule was hard. **(D)** In the nomogram, the total score was 379, and the corresponding malignancy rate of the nodule was high (>0.90). The green arrows indicate the score corresponding to each risk factor, and the red arrow indicates the total score of the nodule.

nomogram is not only based on B-mode US features, but also includes clinical, vascularity and elastographic US characteristics, which can provide more information on TNs. Compared with ACR-TIRADS and EU-TIRADS, the CBCE nomogram showed improved accuracy (0.944) and specificity (0.913) without sacrificing sensitivity (0.963) for predicting malignancy. With the optimum cutoff value of 0.55, the CBCE nomogram showed the highest AUC of 0.983 in the validation group. This means that when the malignancy probability of a nodule is higher than 0.55, it is regarded as a high-risk nodule and FNA (fine needle aspiration) is recommended; when the malignant probability of a nodule is lower than 0.55, it is regarded as a low-risk nodule and follow-up ultrasonography is recommended. The significance of this study is to provide a new strategy for clinicians as well as to avoid overdiagnosis and overtreatment.

The clinical characteristics plays an important role in the diagnosis and treatment of patients with TNs. Some studies suggested that thyroid cancer has been more common in women than men, which may related to sexual hormones, although this has been controversial (25, 26). In developed countries, the majority of newly diagnosed thyroid carcinomas correspond to incidentally found papillary microcarcinomas (25). Women are traditionally more keen on undergoing routine clinical examination along their lives as compared to men (27, 28), which could be part of the reasons why thyroid nodules and cancers are more frequently detected in the female population. However, another studies showed that male gender was the clinical factor associated with the high risk of TC (16, 29). Our results showed that the number of TC was more in male than in female, and gender was an independent risk factor associated with TC. These different results may depend on differences in inclusion criteria and in patient's lifestyle. In terms of age, recent evidence suggested that clinically silent thyroid cancer was frequently present in younger age (30), indicating that screening (31, 32) and environmental factors (33) played a key role in the increased incidence. Our results showed that younger age was also a risk factor associated with TC, similar to previous studies. Other clinical risk factors for TC included a family history of TC, obesity and thyroid dysfunctions (19, 34), but the relevant evidence is uncertain. Our results showed that family history, BMI and TSH level were not independent risk factors associated with TC. Discrepancies across different studies may be due to population samples from different countries and cities. Adding clinical characteristics makes our nomogram more comprehensive and more applicable to clinical practice.

Thyroid ultrasonography is the preferred tool for thyroid imaging because it can clarify the location, size, shape, echogenicity, composition, margin, calcification and vascularity of nodules, as well as the features of cervical lymph nodes. Results (9, 35–41) consistent with malignancy contain hypo-echogenicity, solid, irregular margin, taller-than-wide shape, microcalcification and the presence of suspicious lymph nodes. No single characteristic has been proven to reliably distinguish between benign and malignant lesions. The results also showed that the above conventional US indicators are risk factors associated with TC. The value of vascularity for TNs is different in different studies. Internal vascularity was observed in 30.7%-65.3% of benign nodules and 26.7%-91.7% of malignant nodules (42–46). In terms of the vascularity degree, a study showed a high probability (42%) of malignancy in hypervascular nodules (43), while another study indicated that hypovascular vascularity was frequently observed in malignant nodules (47). In our study, internal vascularization distribution and low vascularization degree were more common in malignant nodules, and these two vascularization features were independent risk factors for TC. Different US instruments have different sensitivities for vascularity detection, which may be related to disparate results. US elastography provides conventional US with complementary information in many human organs. The combined use of elastography with conventional US is able to improve the discrimination of malignant and benign nodules (48). Strain elastography is useful to assess malignancy, with an average specificity of 90% and an overall average sensitivity of 92%, according to a meta-analysis of 639 TNs (49). The findings of this study indicated that the strain imaging of elasticity was an independent risk factor associated with TC. It is worth noting that recently, the International Thyroid Nodule Ultrasound Working Group has been trying to develop a uniform international guideline. The organization proposed extending standardized ultrasonography to elastography of the stiffness in TNs (50). Our research established a complementary role of US elastography and vascularity in the risk stratification of TNs.

The US-based RSSs of TNs provides clinicians with an estimated malignancy rate based on the category. Currently, B-mode US-based RSSs include ACR-TIRADS (8), ATA (19), KTA-TIRADS (Thyroid Imaging Reporting Data System by Korean Thyroid Association) (20), and EU-TIRADS (9). However, there are considerable discrepancies among these RSSs regarding the US features used for risk categories, expected malignancy risk, diagnostic performance, and the size threshold for biopsy. A comparative study of 902 nodules showed that both ATA and KTATIRADS had higher sensitivity for the diagnosis of TC but lower specificity than ACR-TIRADS (10). Another comparative study of 3,422 nodules showed that ATA and K-TIRADS had limitations that included no category for all available nodules, and approximately 10% of these uncategorized nodules had malignancy (12). Another study containing the four RSSs mentioned above showed that the interobserver agreement remained only fair to moderate (51). The reasons for these differences may be due to the inconsistent classification of US features and the complexity of the system. Therefore, each predictor was classified into two

categories to establish an easy-to-use model in our study. The CBCE nomogram we developed had good diagnostic performances and calibration for the probability of TC. Compared with ACR-TIRADS and EU-TIRADS, the CBCE nomogram showed improved accuracy and specificity without sacrificing sensitivity. Furthermore, the CBCE nomogram showed the highest AUC with the optimum cutoff value of 0.55. A nodule is considered high-risk and FNA is recommended when its malignancy probability is over 0.55, while a nodule is considered low-risk and US follow-up is recommended when its malignancy probability is lower than 0.55. The CBCE nomogram provided a novel and simple diagnostic tool for clinicians.

There are some limitations in this study. First, as findings from a single center, our results have limited generalizability. Although our study has been internally validated and showed good diagnostic performance, it needs to further external validation. Second, as this is a retrospective study, data bias is inevitable. Further prospective studies will be needed to improve the model. Third, our study does not address FNA thresholds for thyroid nodules, which requires further research in the future. Despite these limitations, **our study has several highlights.** First, to our knowledge, this is the first study to develop and validate an integrated diagnostic tool for predicting the malignancy probability of TNs based on clinical, B-mode, Color Doppler and elastographic US characteristics. Second, the CBCE nomogram has been demonstrated to have better diagnostic performances in predicting malignancy than ACR-TIRADS and EU-TIRADS. Third, the CBCE nomogram is an easy-to-use diagnostic tool for predicting malignancy. The corresponding malignancy probability of each nodule will be obtained by adding up the specific points of each predictor. Fourth, our study obtained an optimal cut-off value of 0.55 for distinguishing low-risk from high-risk nodules. The cut-off value separates the thyroid nodules into two groups, one of which is high risk with a malignancy probability above 0.55 and the other of which is low risk with a malignancy probability below 0.55. FNA and US follow-up are recommended for the two groups, separately. The significance of this study is to provide a new strategy for clinicians.

Conclusions

In conclusion, this study developed and validated a novel, comprehensive and reliable diagnostic tool based on clinical, B-mode, Color Doppler and elastographic ultrasonographic characteristics for thyroid nodules. Our research has two potential benefits, one of which is to identify patients with high-risk nodules who are the candidates for FNA, and the other is to provide a new strategy for clinical practice.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of Ruijin Hospital, Shanghai Jiao Tong University School of Medicine. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by SX and XN. The first draft of the manuscript was written by SX and all authors commented on previous versions of the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors are grateful to all participants of this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2022.966572/full#supplementary-material

References

1. Guth S, Theune U, Aberle J, Galach A, Bamberger CM. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. *Eur J Clin Invest* (2009) 39(8):699–706. doi: 10.1111/j.1365-2362.2009.02162.x

2. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Pineros M, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer* (2019) 144(8):1941–53. doi: 10.1002/ijc.31937

3. Moon JH, Hyun MK, Lee JY, Shim JI, Kim TH, Choi HS, et al. Prevalence of thyroid nodules and their associated clinical parameters: a large-scale, multicenter-based health checkup study. *Korean J Intern Med* (2018) 33(4):753–62. doi: 10.3904/kjim.2015.273

4. Singh Ospina N, Iniguez-Ariza NM, Castro MR. Thyroid nodules: diagnostic evaluation based on thyroid cancer risk assessment. *BMJ* (2020) 368:l6670. doi: 10.1136/bmj.l6670

5. Grani G, Sponziello M, Pecce V, Ramundo V, Durante C. Contemporary thyroid nodule evaluation and management. *J Clin Endocrinol Metab* (2020) 105 (9):2869–83. doi: 10.1210/clinem/dgaa322

6. Ahn HS, Kim HJ, Welch HG. Korea's thyroid-cancer "epidemic"–screening and overdiagnosis. *N Engl J Med* (2014) 371(19):1765–7. doi: 10.1056/NEJMp1409841

7. Vaccarella S, Franceschi S, Bray F, Wild CP, Plummer M, Dal Maso L. Worldwide thyroid-cancer epidemic? the increasing impact of overdiagnosis. *N Engl J Med* (2016) 375(7):614–7. doi: 10.1056/NEJMp1604412

8. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR thyroid imaging, reporting and data system (TI-RADS): White paper of the ACR TI-RADS committee. *J Am Coll Radiol* (2017) 14(5):587–95. doi: 10.1016/j.jacr.2017.01.046

9. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European Thyroid association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: The EU-TIRADS. *Eur Thyroid J* (2017) 6(5):225–37. doi: 10.1159/000478927

10. Ha EJ, Na DG, Moon WJ, Lee YH, Choi N. Diagnostic performance of ultrasound-based risk-stratification systems for thyroid nodules: Comparison of the 2015 American thyroid association guidelines with the 2016 Korean thyroid Association/Korean society of thyroid radiology and 2017 American college of radiology guidelines. *Thyroid* (2018) 28(11):1532–7. doi: 10.1089/thy.2018.0094

11. Zhang WB, Xu HX, Zhang YF, Guo LH, Xu SH , Zhao CK, et al. Comparisons of ACR TI-RADS, ATA guidelines, Kwak TI-RADS, and KTA/KSThR guidelines in malignancy risk stratification of thyroid nodules. *Clin Hemorheol Microcirc* (2020) 75(2):219–32. doi: 10.3233/CH-190778

12. Middleton WD, Teefey SA, Reading CC, Langer JE, Beland MD, Szabunio MM, et al. Comparison of performance characteristics of American college of radiology TI-RADS, Korean society of thyroid radiology TIRADS, and American thyroid association guidelines. *AJR Am J Roentgenol* (2018) 210(5):1148–54. doi: 10.2214/AJR.17.18822

13. Kim PH, Yoon HM, Hwang J, Lee JS, Jung AY, Cho YA, et al. Diagnostic performance of adult-based ATA and ACR-TIRADS ultrasound risk stratification systems in pediatric thyroid nodules: a systematic review and meta-analysis. *Eur Radiol* (2021) 31(10):7450–63. doi: 10.1007/s00330-021-07908-8

14. Kuru B, Kefeli M, Danaci M. Comparison of 5 thyroid ultrasound stratification systems for differentiation of benign and malignant nodules and to avoid biopsy using histology as reference standard. *Endocr Pract* (2021) 27 (11):1093–9. doi: 10.1016/j.eprac.2021.04.411

15. Liu J, Guo Y, Xiao J, Chen L, Liang Z. Comparison of the efficacy and safety of the American thyroid association guidelines and American college of radiology TI-RADS. *Endocr Pract* (2021) 27(7):661–7. doi: 10.1016/j.eprac.2020.11.013

16. Suteau V, Munier M, Briet C, Rodien P. Sex bias in differentiated thyroid cancer. *Int J Mol Sci* (2021) 22(23):12992. doi: 10.3390/ijms222312992

17. Panagiotou G, Komninou D, Anagnostis P, Linardos G, Karoglou E, Somali M, et al. Association between lifestyle and anthropometric parameters and thyroid nodule features. *Endocrine* (2017) 56(3):560–7. doi: 10.1007/s12020-017-1285-6

18. Miranda-Filho A, Lortet-Tieulent J, Bray F, Cao B, Franceschi S, Vaccarella S, et al. Thyroid cancer incidence trends by histology in 25 countries: a population-based study. *Lancet Diabetes Endocrinol* (2021) 9(4):225–34. doi: 10.1016/S2213-8587(21)00027-9

19. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: The American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid* (2016) 26(1):1–133. doi: 10.1089/thy.2015.0020

20. Shin JH, Baek JH, Chung J, Ha EJ, Kim JH, Lee YH, et al. Ultrasonography diagnosis and imaging-based management of thyroid nodules: Revised Korean society of thyroid radiology consensus statement and recommendations. *Korean J Radiol* (2016) 17(3):370–95. doi: 10.3348/kjr.2016.17.3.370

21. Xia Y, Wang L, Jiang Y, Dai Q, Li X, Li W. Sonographic appearance of primary thyroid lymphoma-preliminary experience. *PloS One* (2014) 9(12): e114080. doi: 10.1371/journal.pone.0114080

22. Sipos JA. Advances in ultrasound for the diagnosis and management of thyroid cancer. *Thyroid* (2009) 19(12):1363–72. doi: 10.1089/thy.2009.1608

23. Leenhardt L, Erdogan MF, Hegedus L, Mandel SJ, Paschke R, Rago T, et al. 2013 European Thyroid association guidelines for cervical ultrasound scan and ultrasound-guided techniques in the postoperative management of patients with thyroid cancer. *Eur Thyroid J* (2013) 2(3):147–59. doi: 10.1159/000354537

24. Asteria C, Giovanardi A, Pizzocaro A, Cozzaglio L, Morabito A, Somalvico F, et al. US-Elastography in the differential diagnosis of benign and malignant thyroid nodules. *Thyroid* (2008) 18(5):523–31. doi: 10.1089/thy.2007.0323

25. Davies L, Welch HG. Current thyroid cancer trends in the united states. *JAMA Otolaryngol Head Neck Surg* (2014) 140(4):317–22. doi: 10.1001/jamaoto.2014.1

26. Moleti M, Sturniolo G, Di Mauro M, Russo M, Vermiglio F. Female reproductive factors and differentiated thyroid cancer. *Front Endocrinol (Lausanne)* (2017) 8:111. doi: 10.3389/fendo.2017.00111

27. Bertakis KD, Azari R. Patient gender differences in the prediction of medical expenditures. *J Womens Health (Larchmt)* (2010) 19(10):1925–32. doi: 10.1089/jwh.2009.1448

28. Wang Y, Hunt K, Nazareth I, Freemantle N, Petersen I. Do men consult less than women? an analysis of routinely collected UK general practice data. *BMJ Open* (2013) 3(8):e003320. doi: 10.1136/bmjopen-2013-003320

29. Campanella P, Ianni F, Rota CA, Corsello SM, Pontecorvi A. Quantification of cancer risk of each clinical and ultrasonographic suspicious feature of thyroid nodules: a systematic review and meta-analysis. *Eur J Endocrinol* (2014) 170(5): R203–211. doi: 10.1530/EJE-13-0995

30. Ohtsuru A, Takahashi H, Kamiya K. Incidence of thyroid cancer among children and young adults in fukushima, Japan-reply. *JAMA Otolaryngol Head Neck Surg* (2019) 145(8):770. doi: 10.1001/jamaoto.2019.1102

31. Vergamini LB, Frazier AL, Abrantes FL, Ribeiro KB, Rodriguez-Galindo C. Increase in the incidence of differentiated thyroid carcinoma in children, adolescents, and young adults: a population-based study. *J Pediatr* (2014) 164 (6):1481–5. doi: 10.1016/j.jpeds.2014.01.059

32. Lamartina L, Leboulleux S, Schlumberger M. Thyroid cancer incidence in children and adolescents. *Lancet Diabetes Endocrinol* (2021) 9(3):128–9. doi: 10.1016/S2213-8587(20)30430-7

33. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2018) 68(6):394–424. doi: 10.3322/caac.21492

34. Choi JS, Nam CM, Kim EK, Moon HJ, Han KH, Kwak JY. Evaluation of serum thyroid-stimulating hormone as indicator for fine-needle aspiration in patients with thyroid nodules. *Head Neck* (2015) 37(4):498–504. doi: 10.1002/hed.23616

35. Na DG, Baek JH, Sung JY, Kim JH, Kim JK, Choi YJ, et al. Thyroid imaging reporting and data system risk stratification of thyroid nodules: Categorization based on solidity and echogenicity. *Thyroid* (2016) 26(4):562–72. doi: 10.1089/thy.2015.0460

36. Ha EJ, Moon WJ, Na DG, Lee YH, Choi N, Kim SJ, et al. A multicenter prospective validation study for the Korean thyroid imaging reporting and data system in patients with thyroid nodules. *Korean J Radiol* (2016) 17(5):811–21. doi: 10.3348/kjr.2016.17.5.811

37. Moon WJ, Jung SL, Lee JH, Na DG, Baek JH, Lee YH, et al. Benign and malignant thyroid nodules: US differentiation–multicenter retrospective study. *Radiology* (2008) 247(3):762–70. doi: 10.1148/radiol.2473070944

38. Grant EG, Tessler FN, Hoang JK, Langer JE, Beland MD, Berland LL, et al. Thyroid ultrasound reporting lexicon: White paper of the ACR thyroid imaging, reporting and data system (TIRADS) committee. *J Am Coll Radiol* (2015) 12(12 Pt A):1272–9. doi: 10.1016/j.jacr.2015.07.011

39. Kim SY, Na DG, Paik W. Which ultrasound image plane is appropriate for evaluating the taller-than-wide sign in the risk stratification of thyroid nodules? *Eur Radiol* (2021) 31(10):7605–13. doi: 10.1007/s00330-021-07936-4

40. Zhou J, Yin L, Wei X, Zhang S, Song Y, Luo B, et al. 2020 Chinese Guidelines for ultrasound malignancy risk stratification of thyroid nodules: the c-TIRADS. *Endocrine* (2020) 70(2):256–79. doi: 10.1007/s12020-020-02441-y

41. Kwak JY, Jung I, Baek JH, Baek SM, Choi N, Choi YJ, et al. Image reporting and characterization system for ultrasound features of thyroid nodules:

multicentric Korean retrospective study. *Korean J Radiol* (2013) 14(1):110–7. doi: 10.3348/kjr.2013.14.1.110

42. Papini E, Guglielmi R, Bianchini A, Crescenzi A, Taccogna S, Nardi F, et al. Risk of malignancy in nonpalpable thyroid nodules: predictive value of ultrasound and color-Doppler features. *J Clin Endocrinol Metab* (2002) 87(5):1941–6. doi: 10.1210/jcem.87.5.8504

43. Frates MC, Benson CB, Doubilet PM, Cibas ES, Marqusee E. Can color Doppler sonography aid in the prediction of malignancy of thyroid nodules? *J Ultrasound Med* (2003) 22(2):127–31. doi: 10.7863/jum.2003.22.2.127

44. Rago T, Vitti P, Chiovato L, Mazzeo S, De Liperi A, Miccoli P, et al. Role of conventional ultrasonography and color flow-doppler sonography in predicting malignancy in 'cold' thyroid nodules. *Eur J Endocrinol* (1998) 138(1):41–6. doi: 10.1530/eje.0.1380041

45. Appetecchia M, Solivetti FM. The association of colour flow Doppler sonography and conventional ultrasonography improves the diagnosis of thyroid carcinoma. *Horm Res* (2006) 66(5):249–56. doi: 10.1159/000096013

46. Ma JJ, Ding H, Xu BH, Xu C, Song LJ, Huang BJ, et al. Diagnostic performances of various gray-scale, color Doppler, and contrast-enhanced ultrasonography findings in predicting malignant thyroid nodules. *Thyroid* (2014) 24(2):355–63. doi: 10.1089/thy.2013.0150

47. Moon HJ, Kwak JY, Kim MJ, Son EJ, Kim EK. Can vascularity at power Doppler US help predict thyroid malignancy? *Radiology* (2010) 255(1):260–9. doi: 10.1148/radiol.09091284

48. Sigrist RMS, Liau J, Kaffas AE, Chammas MC, Willmann JK. Ultrasound elastography: Review of techniques and clinical applications. *Theranostics* (2017) 7 (5):1303–29. doi: 10.7150/thno.18650

49. Bojunga J, Herrmann E, Meyer G, Weber S, Zeuzem S, Friedrich-Rust M. Real-time elastography for the differentiation of benign and malignant thyroid nodules: a meta-analysis. *Thyroid* (2010) 20(10):1145–50. doi: 10.1089/thy.2010.0079

50. Tessler FN. Thyroid nodules and real estate: Location matters. *Thyroid* (2020) 30(3):349–50. doi: 10.1089/thy.2020.0090

51. Persichetti A, Di Stasio E, Coccaro C, Graziano F, Bianchini A, Di Donna V, et al. Inter- and intraobserver agreement in the assessment of thyroid nodule ultrasound features and classification systems: A blinded multicenter study. *Thyroid* (2020) 30(2):237–42. doi: 10.1089/thy.2019.0360

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 31 October 2022
DOI 10.3389/fendo.2022.981403

OPEN ACCESS

EDITED BY
Jeffrey Garber,
Atrius Health, United States

REVIEWED BY
Zbigniew Adamczewski,
Medical University of Lodz, Poland
Aldo Bove,
G. D'Annunzio University of Chieti-Pescara, Italy

*CORRESPONDENCE
Liping Wang
wanglp@zjcc.org.cn
Lei Xu
xulei@qsmm.org.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Thyroid Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 11 July 2022
ACCEPTED 17 October 2022
PUBLISHED 31 October 2022

CITATION
Xu D, Wang Y, Wu H, Lu W, Chang W, Yao J, Yan M, Peng C, Yang C, Wang L and Xu L (2022) An artificial intelligence ultrasound system's ability to distinguish benign from malignant follicular-patterned lesions. *Front. Endocrinol.* 13:981403. doi: 10.3389/fendo.2022.981403

An artificial intelligence ultrasound system's ability to distinguish benign from malignant follicular-patterned lesions

Dong Xu[1,2,3,4†], Yuan Wang[5†], Hao Wu[6], Wenliang Lu[5], Wanru Chang[5], Jincao Yao[1], Meiying Yan[1], Chanjuan Peng[1], Chen Yang[1], Liping Wang[1,2*] and Lei Xu[2,7,8*]

[1]Department of Ultrasonography, The Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Institute of Basic Medicine and Cancer, Chinese Academy of Sciences, Hangzhou, China, [2]Ultrasound Branch, Zhejiang Society for Mathematical Medicine, Hangzhou, China, [3]Key Laboratory of Head & Neck Cancer Translational Research of Zhejiang Province, Zhejiang Provincial Research Center for Cancer Intelligent Diagnosis and Molecular Technology, Hangzhou, China, [4]Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai, China, [5]School of Mathematical Sciences, Zhejiang University, Hangzhou, China, [6]Department of Ultrasound, The Second Affiliated Hospital of Zhejiang Chinese Medical University, Hangzhou, China, [7]Group of Computational Imaging and Digital Medicine, Zhejiang Qiushi Institute for Mathematical Medicine, Hangzhou, China, [8]Group of Intelligent Medical Devices, South and North Lake Institute for Medical Artificial Intelligence, Haiyan, China

Objectives: To evaluate the application value of a generally trained artificial intelligence (AI) automatic diagnosis system in the malignancy diagnosis of follicular-patterned thyroid lesions (FPTL), including follicular thyroid carcinoma (FTC), adenomatoid hyperplasia nodule (AHN) and follicular thyroid adenoma (FTA) and compare the diagnostic performance with radiologists of different experience levels.

Methods: We retrospectively reviewed 607 patients with 699 thyroid nodules that included 168 malignant nodules by using postoperative pathology as the gold standard, and compared the diagnostic performances of three radiologists (one junior, two senior) and that of AI automatic diagnosis system in malignancy diagnosis of FPTL in terms of sensitivity, specificity and accuracy, respectively. Pairwise t-test was used to evaluate the statistically significant difference.

Results: The accuracy of the AI system in malignancy diagnosis was 0.71, which was higher than the best radiologist in this study by a margin of 0.09 with a p-value of 2.08×10^{-5}. Two radiologists had higher sensitivity (0.84 and 0.78) than that of the AI system (0.69) at the cost of having much lower specificity (0.35, 0.57 versus 0.71). One senior radiologist showed balanced sensitivity and specificity (0.62 and 0.54) but both were lower than that of the AI system.

Conclusions: The generally trained AI automatic diagnosis system can potentially assist radiologists for distinguishing FTC from other FPTL cases that share poorly distinguishable ultrasonographical features.

KEYWORDS

thyroid adenomas, adenocarcinomas, follicular, ultrasonography, artificial intelligence

Highlights

- The AI automatic diagnosis system exhibited higher accuracy and specificity than radiologists in malignancy diagnosis of FPTL.
- The AI automatic diagnosis system had more balanced performance than radiologists in diagnosis of FPTL cases.

Introduction

Thyroid carcinoma is the most common endocrine tumor in endocrine system. There is growing evidence in support of an increase in the occurrence of thyroid cancer. Lim et al. (1) reported that thyroid cancer incidence increased, on average, 3.6% per year during 1974–2013. Papillary thyroid cancer (PTC) incidence increased for all stages at diagnosis. Overall and distant PTC incidence-based mortality increased respectively 1.1% and 2.9% per year during 1994–2013. The main approaches to identify suspicious thyroid nodules are high-frequency color doppler ultrasonography and ultrasound-guided fine needle aspiration cytology (FNAC) (2). The Thyroid Imaging Reporting and Data Systems proposed by the American College of Radiology (ACR TI-RADS), which is a globally accepted malignancy risk stratification system for classifying thyroid nodules on the basis of their features at ultrasonography (US) imaging (3) shows high accuracy in distinguishing benign and malignant thyroid nodules, and can effectively reduce unnecessary biopsy of thyroid nodules on a large scale (4). However, the malignant ultrasonic features and risk categories of thyroid nodules in ACR TI-RADS are mainly based on papillary thyroid carcinoma (PTC), which accounts for the vast majority of malignant samples.

Abbreviations: AI-CADx, Artificial Intelligent Computer- Aided Diagnosis System; AUC, Area Under the Curve; CNN, Convolutional Neural Network; FNAC, Fine Needle Aspiration Cytology; ROC, Receiver operating Characteristic; SDK, Software Development Kit; TI-RADS, Thyroid Imaging Reporting and Data System.

Currently, there are no clear instructions about how to distinguish benign and malignant thyroid follicular tumors in ACR TI-RADS. Follicular thyroid carcinoma (FTC), accounting for 5%-10% of all thyroid cancer, is the second common thyroid carcinoma (5). Compared with the most commonly occurring malignant PTC, FTC is less prone to lymph node metastasis, but more likely to relapse and metastasize to lungs and bones. When recurrence or distant metastasis occurs, it indicates a poor prognosis. In addition, compared with PTC, FTC is more likely to be locally invasive (6). Thyroid lobectomy alone may be sufficient initial treatment for low-risk follicular carcinomas; however, the treatment team may choose total thyroidectomy to enable RAI therapy for low to intermediate risk patients' follicular carcinomas (2). Therefore, an accurate diagnosis of FTC before the initial operation has a tremendous influence on the surgical procedure and prognosis.

It has been found that, FTC shares similar characteristics in both ultrasound images (7–10) and FNAC (11–14) to other follicular-patterned thyroid lesions (FPTL) such as thyroid follicular adenoma (FTA) and adenomatoid hyperplasia nodule (AHN), hampering the malignancy diagnosis and their differential diagnosis. The gold standard for preoperative diagnosis of thyroid nodules Fine Needle Aspiration (FNA) can only diagnose follicular tumors and cannot distinguish between benign and malignant nodules. The final diagnosis instead relies on the detection of capsule involvement and vascular invasion in the postoperative pathological examinations (15). How to improve the preoperative differentiation of benign and malignant thyroid follicular tumors has important practical significance.

Genetic testing can, in principle, help diagnose thyroid nodules. For FPTL cases specifically, role of RAS mutations may be relevant (16–18). However, the most common thyroid related oncogene, namely the BRAFV600E mutation, is poor for malignancy differentiation of follicular patterned tumors (19, 20). In addition, compared with noninvasive ultrasonography, genetic testing requires more invasive fine needle aspiration biopsy and is also more costly. The advancement of AI technologies and especially the development of deep learning algorithms has brought radiologists new tools during the clinical studies for disease detection and diagnosis (21). Developing ultrasound-based AI technologies for thyroid nodule diagnosis has a potential to reduce invasive examinations. Convolutional neural networks have also been

applied to the automatic detection and diagnosis of thyroid nodules (22–26). However, to our humble knowledge, currently there has not been any study trying to apply deep learning technologies to diagnose malignant nodules among FPTL that have indistinguishable ultrasonographic and cytologic features.

In this study, we applied the software development kit (SDK) of a generally trained thyroid nodule diagnosis system as it is without retraining for malignancy prediction of FPTL that included retrospectively collected FTC, FTA and AHN. This system, trained on an unselected population of nodules as opposed to nodules known to have follicular pattern, is initialized using self-training with noisy student method on ImageNet database, and takes a specially defined focal loss function to resolve the problem of unbalanced sample distribution that is frequently occurring in medical data. Focal Loss (27) increases the weight of rare classes in the loss function, making the minimization of loss function more sensitive to these samples, which is helpful to improve the accuracy of rare classes. In addition, it uses a Sharpness-Aware Minimization (SAM) algorithm to simultaneously minimize loss value and loss sharpness to improve the generalizability of the model (28). In particular, SAM algorithm seeks parameters that lie in neighborhoods with uniformly low loss. We compared its diagnostic performance with that of the radiologists of different experience levels using common evaluation metrics such as sensitivity, specificity and accuracy as well as two-tailed paired t-tests to verify whether if any observed differences were statistically significant.

Materials and methods

Data summary

A total of 607 patients with FTC, FTA and AHN (699 nodules) with complete but anonymized clinical information who underwent preoperative ultrasonography and complete examinations pulled from the provincial database from Zhejiang Society for Mathematical Medicine, with data contributed by 7 member hospitals, in which 263 nodules from The Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), were included in this study. The histopathological diagnoses of all FTC, FTA and AHN were determined surgically. In summary, our study included 167 cases of FTC (23.89%), 241 cases of FTA (34.48%) and 291 cases of AHN (41.63%).

Ultrasound examinations by radiologists and AI software

One junior radiologist A with 10 years of working experience and two senior radiologists, radiologist B and radiologist C, both with 20 years of working experience in ultrasound diagnosis performed the clinical ultrasound examinations on patients without knowing their pathological outcomes.

The ultrasound images were first grouped according to the associated nodules and then analyzed independently by all radiologists and the SDK (version 2.3.1.5) of the AI-SONIC™ Thyroid system with software version 5.3.0.2 (DE-Medicum Petavoxels Co., Ltd), which was developed on the EfficientNet architecture (29) using the proprietary deep learning framework DE-Light, and the system returned the predicted malignancy probability value of each nodule in the ultrasound images. The maximum malignancies predicted from the images associated to each nodule were chosen as the nodule-specific malignancy scores by all the raters, i.e., the radiologists and the AI. The malignancy probability value ranges from 0 to 1, and the cut-off value for the AI system was set by maximizing the mean of the sensitivity and specificity curves. In this study, the cut-off value was set to 0.4 by the AI system according to the analysis in Supplementary Figure 1. If the probability value is ≥ 0.4, the nodule is diagnosed as malignant, otherwise benign.

For further analysis of the nodule cases for which the AI system made correct diagnoses but failed by at least two of the three radiologists participating in the evaluation comparison study, the three radiologists who participated in the evaluation comparison study were asked to assign the ultrasound features according to ACR TI-RADS standards after discussions side-by-side and reviewed those images with a washout time longer than 6 months. We computed the sum of weighted scores by the frequency of nodule cases according to the TI-RADS scoring system for each individual ultrasonographic feature to obtain the average characteristic profile of these nodules.

Statistical analysis

To assess the performance of the AI-SONIC™ system, we computed the Receiver operating Characteristic (ROC) curve and used the Area Under the Curve (AUC) as the evaluation metric. In order to compare its diagnosis with that of the radiologists, we calculated the sensitivity, specificity and accuracy. In addition, two-tailed t-test and McNemar test were used to compute p values for statistical comparisons. In all analyses, a p value less than 0.05 was considered a statistically significant difference. Statistical analysis was performed using Python 3.8 (Python Software Foundation, Delaware, United States).

Results

Comparison between the AI system and radiologists of different experience levels

We calculated the sensitivity, specificity, positive predict value (PPV) and accuracy of the AI system and three radiologists in malignancy diagnosis of FPTL that consisted of FTC, FTA and

AHN. The accuracy and PPV of the AI system was higher than all surveyed radiologists, as shown in Table 1. The sensitivity of the AI system however was lower than that of senior radiologist C (0.69 vs 0.78) and junior radiologist A (0.69 vs 0.84), but higher than that of senior radiologist B (0.69 vs 0.62). The specificity of the AI system in malignancy diagnosis was higher than all surveyed radiologists (0.71 vs 0.35, 0.54, 0.57 respectively). The overall performances were summarized in Figure 1, in which the ROC curve and the associated AUC value of the AI system were computed. Furthermore, we applied the McNemar test to compute p values between the AI system and the three radiologists. The results are as follows: $p_{AI\text{-}A} = 1.71\times10^{-38}$, $p_{AI\text{-}B} = 2.10\times10^{-6}$ and $p_{AI\text{-}C} = 3.62\times10^{-10}$. All p values between the AI system and three radiologists were less than 1×10^{-5}. There were significant statistical differences between the AI system and three radiologists in diagnosing FTC, FTA and AHN. We presented a set of representative ultrasound images that showed the advantages of AI the system over radiologists in malignancy diagnosis of FPTL cases in Figure 2.

To further compare the diagnosis between the AI system and three radiologists, we subdivided the complete dataset to ten randomly divided subsets, summarized as in Table 2.

We calculated each rater's accuracies in each dataset for malignancy diagnosis of FTC, FTA and AHN, and computed their average values and standard deviations over the ten datasets, with the corresponding results summarized in Figure 3A.

To statistically compare the diagnostic accuracies of the AI system and three radiologists in predicting thyroid nodule malignancies of FTC, FTA and AHN, we computed pairwise p values with the results shown in Figure 3B. Note that we skipped the statistical comparisons against oneselves, as in this case the p values were constant 1. All p values were < 0.02, where all p values between the AI system and three radiologists were less than 1×10^{-4}. There were significant statistical differences between the AI system and three radiologists in diagnosing FTC, FTA and AHN.

Comparison the performance of the AI system and radiologists in diagnosis FTC, FTA and AHN respectively

To further compare the AI system and three radiologists' diagnosis, we calculated sensitivity and specificity for these three nodules, including FTC, FTA, and AHN, respectively. The results were shown in Table 3. When we only considered FTC cases, there were no true benign nodule cases in the numerator for the specificity calculation, resulting in a constant 0 for the specificity, which we omitted and showed only the sensitivity in the table. Similarly, for FTA and AHN cases, we presented only the specificity while ignored the sensitivity as there were no true malignant cases. For benign FPTL such as FTA and AHN, the specificity of the AI system was higher than that of the three radiologists. The sensitivity of the AI system was lower than that of senior radiologist C (0.69 vs 0.78) and junior radiologist A (0.69 vs 0.84), but higher than that of senior radiologist B (0.69 vs 0.62) in diagnosis FTC.

For further analysis, we selected the nodule cases for which the AI system made correct diagnoses but failed by at least two of the three radiologists participating in the evaluation comparison study and we got in total 144 benign and 12 malignant nodules. Our summarized results in Table 4 show that for those 144 benign nodules, the sum of weighted scores (3.974) corresponds well to malignant suspicious category 4 nodules with an average characteristic profile of being predominantly solid, hypoechoic, with some but not pronounced echogenic foci and were considered by at least two radiologists to be malignant while they were correctly diagnosed to be benign by the AI system. For the 12 malignant nodules misdiagnosed by at least two radiologists to be benign, the sum of weight scores amounts to 3.916, with an average characteristic profile of being solid, mostly hypoechoic and predominantly without echogenic foci.

TABLE 1 The diagnostic performances of the AI system and three radiologists in thyroid malignancy nodules diagnosis.

	AI	Radiologist A (Junior)	Radiologist B (Senior)	Radiologist C (Senior)
Accuracy	**0.71**	0.47	0.56	0.62
Sensitivity	0.69	**0.84**	0.62	0.78
Specificity	**0.71**	0.35	0.54	0.57
PPV	**0.43**	0.29	0.30	0.36

Bold values are the highest values.

Discussion

FTC is a malignant follicular epithelial thyroid tumor with follicular differentiation but lacking the diagnosis characteristics of PTC. As previously noted, FTC has similar ultrasonic features to FTA and AHN, which have been identified by radiologists for differentiating malignant nodules from benign ones from a general perspective. And this is supported by this study that our three radiologists including two senior ones with experiences of more than 20 years in ultrasound diagnosis could at best reach an overall accuracy slightly more than 60% (62%) for malignancy predictions. It has been reported that the ultrasound diagnostic sensitivity for non-follicular thyroid tumors could reach 86.5%, but the diagnostic sensitivity for follicular tumors was only 18.2%, and the corresponding specificities for non-follicular and follicular tumors were 92.3% and 88.7% respectively (8). In our study, though the sensitivities in malignancy prediction by radiologists were all above 60%, the specificities could be as low as 35%. The poor diagnostic performance of the radiologists for FPTL cases can be most likely attributed to the fact that even senior radiologists lack diagnostic experiences due to the low overall incidence rate of thyroid follicular tumors. The AI-SONIC™ thyroid automatic

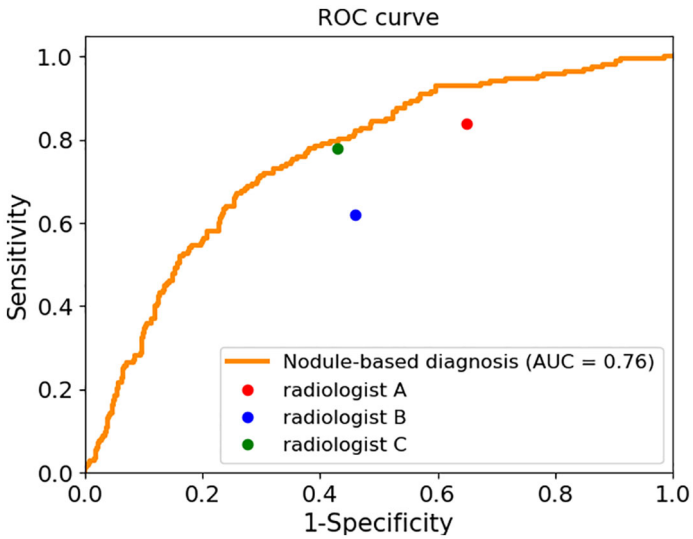


FIGURE 1
The sensitivity and specificity of three radiologists and the ROC curve and AUC value of the AI system.

diagnostic system which is trained for general benign and malignant nodule differentiation however provided a much better diagnostic accuracy of 0.71 for malignancy prediction of FPTL cases, and extremely balanced performance with the sensitivity and specificity being 0.69 and 0.71 respectively on a per-nodule analysis using the maximum malignancy scores computed from the images associated to each nodule. The p value based on two-tail pairwise t-test comparing the AI system and the best radiologist from ten randomly divided subsets in terms of malignancy accuracy in this study was 2.08×10^{-5}, confirming the gap of 9% in accuracy was firmly statistically significant. In diagnosis of FTC, FTA and AHN respectively, the specificity of the AI system (0.74, 0.68) was higher than that of three radiologists, though the sensitivity of the AI system (0.69)

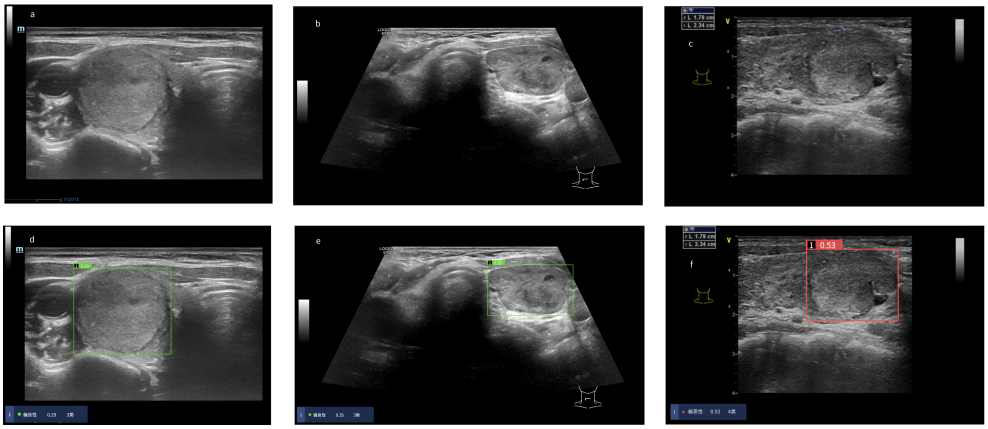


FIGURE 2
Risk coefficient assessment and diagnosis of thyroid nodules in the AI system. When the risk coefficient of thyroid nodules was < 0.4, the AI system diagnosed the thyroid nodules as "benign" as noted in the green display box, otherwise "malignant" as noted in the red display box. **(A–C)** Original ultrasound images of thyroid nodules. **(A)** Pathological diagnosis: thyroid follicular adenoma. **(B)** Pathological diagnosis: adenomatoid hyperplasia nodule. **(C)** Pathological diagnosis: follicular thyroid carcinoma. **(D–F)** Diagnosis of thyroid nodules in the AI system. **(D)** The AI system diagnosed the nodule as "benign". Three radiologists diagnosed the nodule as "malignant". **(E)** The AI system diagnosed the nodule as "benign". Three radiologists diagnosed the nodule as "malignant". **(F)** The AI system diagnosed the nodule as "malignant". Three radiologists diagnosed the nodule as "benign".

10.3389/fendo.2022.981403

TABLE 2 The subdivided datasets for subsequent nodular diagnosis experiment.

Dataset	Total nodules	FTC	FTA	AHN
one	70	19	28	23
two	70	16	25	29
three	70	19	22	29
four	70	15	25	30
five	70	14	15	41
six	70	21	27	22
seven	70	15	23	32
eight	70	16	27	27
nine	70	12	23	35
ten	69	21	24	24

was lower than the radiologist A (0.84) as well as radiologist C (0.78). However, the specificity of the radiologist A was extremely low (0.34, 0.37). This is probably because radiologist A could not distinguish the characteristics of benign and malignant FTPL, but had the tendency to overestimate the malignancy levels, resulting in high sensitivity and low specificity. Radiologist C could not differentiate between FTA and FTC. The AI system provided more balanced performance with the sensitivity and specificity. Our further analysis on the 144 benign cases where AI system made correct diagnoses but failed by at least two of the three radiologists suggests that the radiologists when using the TI-RADS scoring system might have a more conservative concern not to underplay the malignant potentials, consistent with radiologists' higher sensitivity but lower specificity in thyroid nodule diagnosis. For the 12 malignant cases where AI system correctly diagnosed but failed by at least two of the three radiologists, since the number of these cases is small while being also at the boundary of being considered benign or malignant, it is difficult to assess whether this is significant. Nevertheless, the AI system can potentially assist radiologists distinguish FTC from other FPTL cases, given its higher overall accuracy and especially higher specificity. One possibility is to let the radiologists decide whether they would adopt the suggestions by the AI system or not, as long as higher diagnostic accuracy can be expected (30) from the AI system than the radiologists. Another possibility is to set up a rule so that a favorable outcome would be expected (26, 31). In our case, for instance, when the AI system predicts a nodule to be benign which is different from a radiologist's decision given his or her assigned TI-RADS category, one lowers the category assignment by one. Overall, employing an algorithm or workflow that initially uses an AI diagnosis for classes of nodules for which it is superior to a radiologist, is more accurate than enabling a radiologist's subjective decision to accept or reject it.

It shall be noted that in the context of benign thyroid nodular diseases, a high specificity of a diagnostic tool for malignancy detection is desirable. For inconclusive Bethesda categories that may be identified to be PTC follicular variant, noninvasive follicular thyroid neoplasm with papillary-like nuclear features (32), and FTC by histopathology, AI may help with benign and malignancy discrimination. Nevertheless, it was not anticipated that the AI system could manage well for distinguishing FTC from benign FPTL cases given that FTC has relatively rare incidence rate and that the design of the most widely applied ACR TI-RADS lexicon is based on the manifested malignant features of the most dominant PTC of all thyroid cancers. A reasonable explanation would be that the designer of the AI system has defined a focal loss function to resolve the problem of unbalanced sample distribution and likely has paid more attention to FPTL cases with higher learning weights. And the sharpness perception minimization algorithm that has been used could be beneficial for generalizability of their deep learning model.

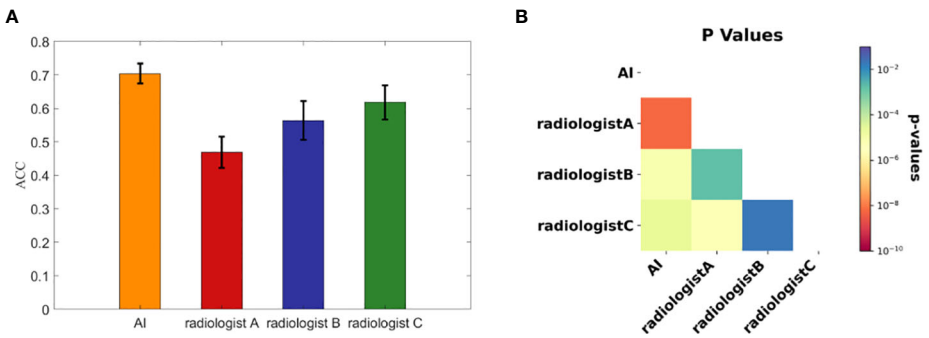


FIGURE 3
The diagnostic performances of three radiologists and the AI automatic diagnosis system. **(A)** The accuracies calculated from the ten subdivided datasets by the AI system and three radiologists. Each bar representing each concerned rater is presented with the average accuracy over the ten subdivided datasets and the standard deviation. **(B)** The associated p value matrix for statistical comparisons of diagnostic accuracies. All p values were <0.02 and self- comparisons were omitted as they were constant 1.

10.3389/fendo.2022.981403

TABLE 3 The respective diagnostic performances of the AI system and three radiologists in diagnosing FTC, FTA and AHN cases.

	No. of cases	AI	Radiologist A(Junior)	Radiologist B(Senior)	Radiologist C(Senior)
FTC	167				
Sensitivity		0.69	**0.84**	0.62	0.78
FTA	241				
Specificity		**0.74**	0.34	0.51	0.5
AHN	291				
Specificity		**0.68**	0.37	0.57	0.63

When we only considered FTC cases, there were no true benign nodule cases, resulting in a constant 0 of the specificity, which we omitted in the table. Similarly, for FTA and AHN cases, we presented only the specificity while ignored the sensitivity for there were no true malignant cases.
Bold values are the highest values.

It is interesting to point out that though FNAC is able to identify follicular tumors with high reliability, it has difficulty in predicting malignancy of FPTL cases (11–14). Therefore, it would be interesting to investigate whether applying the AI system for FPTL cases in combination with FNAC can help reduce the need of surgical excision for malignancy determination in the future. We have not included prospective data in this study because of the low prevalence of FPTL such that it can take a fairly long time to accumulate enough cases for statistically reliable evaluation. Apart from that, it would also be attractive to study how effective it is to train a deep learning model for differentiating follicular from papillary patterns of thyroid nodules based purely on retrospectively collected ultrasound images, which if good enough would help reduce unnecessary fine needle aspirations for future investigation.

The AI automatic diagnosis system can be potentially used as an auxiliary method for screening of FTC from other FPTL cases and may help reduce the need of surgical excision for further characterization given that FNAC has difficulty in determining the malignancy of these cases.

TABLE 4 Correlation of ultrasonographical features with nodular benignity and malignancy for which the AI system made correct diagnoses but failed by at least two of the three radiologists participating in the evaluation comparison study.

ACR TI-RADS Features	ACR Score	Benign (144)			Malignant (12)		
Margin		**Frequency**	**Probability**	**Weighted score**	**Frequency**	**Probability**	**Weighted score**
Smooth	0	120	0.833	0	6	0.500	0
Ill-defined	0	24	0.167	0	6	0.500	0
Shape							
Wider-than-tall	0	140	0.972	0	12	1.000	0
Taller-than-wide	3	4	0.028	0.084	0	0.000	0
Echogenicity							
Anechoic	0	1	0.007	0	0	0.000	0
Very hypoechoic	3	1	0.007	0.021	0	0.000	0
Hypoechoic	2	69	0.479	0.958	9	0.750	1.5
Isoechoic	1	72	0.500	0.5	3	0.250	0.25
Hyperechoic	1	1	0.007	0.007	0	0.000	0
Composition							0
Mixed cystic and solid	1	28	0.194	0.194	0	0.000	0
Cystic or almost completely cystic	0	1	0.007	0	0	0.000	0
Solid or almost completely solid	2	115	0.799	1.598	12	1.000	2
Echogenic foci							
Peripheral	2	8	0.056	0.112	1	0.083	0.166
Macro-calcifications	1	15	0.104	0.104	0	0.000	0
Punctate echogenic foci	3	19	0.132	0.396	0	0.000	0
None	0	102	0.708	0	11	0.917	0
Sum	–	144	1	3.974	12		3.916

Data availability statement

The data that support the findings of this study are available on reasonable request from corresponding author LW after formal approval by the concerned Chinese regulating authorities.

Author contributions

Conceptualization, LW and LX; methodology, YW, JY and WL; software, YW, WL and WC; validation, LW and LX; formal analysis, YW and DX; investigation, HW, MY, CP and CY; resources, LW and LX; data curation, MY, CP and CY; writing—original draft preparation, WY and DX; writing—review and editing, LX; supervision, LW and LX; project administration, LW and LX; funding acquisition, DX and LW. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (No. 82071946) and the Zhejiang Provincial Natural Science Foundation of China (No. LY20H180001, LSD19H180001, LZY21F030001).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2022.981403/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The changes of sensitivity and specificity with respect to the malignancy thresholds. **(A)** Evaluation metrics calculated from all 699 nodules, in which 168 were malignant. **(B)** Evaluation metrics calculated from randomly pulled 100 benign and 100 malignant nodules.

References

1. Lim H, Devesa SS, Sosa JA, Check D, Kitahara CM. Trends in thyroid cancer incidence and mortality in the united states, 1974-2013. *Jama* (2017) 317(13):1338–48. doi: 10.1001/jama.2017.2719

2. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: The American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid* (2016) 26(1):1–133. doi: 10.1089/thy.2015.0020

3. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR thyroid imaging, reporting and data system (TI-RADS): White paper of the ACR TI-RADS committee. journal of the American college of radiology. *J Am Coll Radiol* (2017) 14(5):587–95. doi: 10.1016/j.jacr.2017.01.046

4. Al Dawish M, Alwin Robert A, Al Shehri K, Hawsawi S, Mujammami M, Al Basha IA, et al. Risk stratification of thyroid nodules with Bethesda III category: The experience of a territorial healthcare hospital. *Cureus* (2020) 12(5):e8202. doi: 10.7759/cureus.8202

5. Sobrinho-Simões M, Eloy C, Magalhães J, Lobo C, Amaro T. Follicular thyroid carcinoma. *Mod Pathol* (2011) 24(2):S10–8. doi: 10.1038/modpathol.2010.133

6. Chiofalo MG, D'Anna R, Di Gennaro F, Setola SV, Marotta V. Great veins invasion in follicular thyroid cancer: single-centre study assessing prevalence and clinical outcome. *Endocrine* (2018) 62(1):71–5. doi: 10.1007/s12020-018-1622-4

7. Rick ME, Moll S, Taylor MA, Krizek DM, White GC, Aronson DL. Differential diagnostic ultrasound criteria of papillary and follicular carcinomas: a multivariate analysis. *Rofo* (2014) 186(5):489–95. doi: 10.1055/s-0034-1366282

8. Koike E, Noguchi S, Yamashita H, Murakami T, Ohshima A, Kawamoto H, et al. Ultrasonographic characteristics of thyroid nodules: prediction of malignancy. *Arch Surg* (2001) 136(3):334–7. doi: 10.1001/archsurg.136.3.334

9. Sillery JC, Reading CC, Charboneau JW, Henrichsen TL, Hay ID, Mandrekar JN. Thyroid follicular carcinoma: sonographic features of 50 cases. *Ajr Am J Roentgenol* (2010) 194(1):44–54. doi: 10.2214/AJR.09.3195

10. Rago T, Di Coscio G, Basolo F, Scutari M, Elisei R, Berti P, et al. Combined clinical, thyroid ultrasound and cytological features help to predict thyroid malignancy in follicular and hupsilonrthle cell thyroid lesions: results from a series of 505 consecutive patients. *Clin Endocrinol (Oxf)* (2007) 66(1):13–2. doi: 10.1111/j.1365-2265.2006.02677.x

11. Poller DN, Ibrahim AK, Cummings MH, Mikel JJ, Boote D, Perry M. Fine-needle aspiration of the thyroid. *Cancer* (2000) 90(4):239–44. doi: 10.1002/1097-0142(20000825)90:4<239::aid-cncr7>3.0.co;2-s

12. Busseniers AE, Oertel YC. "Cellular adenomatoid nodules" of the thyroid: Review of 219 fine-needle aspirates. *Diagn Cytopathol* (1993) 9(5):581–89. doi: 10.1002/dc.2840090523

13. Rossi ED, Adeniran AJ, Faquin WC. Pitfalls in thyroid cytopathology. *Surg Pathol Clin* (2019) 12(4):865–81. doi: 10.1016/j.path.2019.08.001

14. Baloch ZW, Fleisher S, Livolsi VA, Gupta PK. Diagnosis of "follicular neoplasm": a gray zone in thyroid fine-needle aspiration cytology. *Diagn Cytopathol* (2002) 26(1):41–4. doi: 10.1002/dc.10043

15. LiVolsi V, Abdulkader NI, Baloch ZW, Bartolazzi A. Follicular thyroid carcinoma. In: Lloyd RV, RY O, Kloppel G, Rosai J, editors. *WHO classification of tumors of endocrine organs, 4th ed.* Lyon: IARC (2017). p. 92–3.

16. Marotta V, Bifulco M, Vitale M. Significance of RAS mutations in thyroid benign nodules and non-medullary thyroid cancer. *Cancers (Basel)* (2021) 13 (15):3785. doi: 10.3390/cancers13153785

17. Marotta V, Sciammarella C, Colao AA, Faggiano A. Application of molecular biology of differentiated thyroid cancer for clinical prognostication. *Endocr Relat Cancer* (2016) 23(11):R499–R15. doi: 10.1530/ERC-16-0372

18. Puzziello A, Guerra A, Murino A, Izzo G, Carrano M, Angrisani E, et al. Benign thyroid nodules with RAS mutation grow faster. *Clin Endocrinol (Oxf)* (2016) 84(5):736–40. doi: 10.1111/cen.12875

19. Guerra A, Di Stasi V, Zeppa P, Faggiano A, Marotta V, Vitale M. BRAF (V600E) assessment by pyrosequencing in fine needle aspirates of thyroid nodules with concurrent hashimoto's thyroiditis is a reliable assay. *Endocrine* (2014) 45 (2):249–55. doi: 10.1007/s12020-013-9994-y

20. Marotta V, Sapio MR, Guerra A, Vitale M. BRAF mutation in cytology samples as a diagnostic tool for papillary thyroid carcinoma. *Expert Opin Med Diagn* (2011) 5(4):277–90. doi: 10.1517/17530059.2011.575058

21. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJ. Artificial intelligence in radiology. *Nat Rev Cancer* (2018) 18(8):500–10. doi: 10.1038/s41568-018-0016-5

22. Ma J, Wu F, Jiang T, Zhao Q, Kong D. Ultrasound image-based thyroid nodule automatic segmentation using convolutional neural networks. *Int J Comput Assist Radiol Surg* (2017) 12(11):1895–910. doi: 10.1007/s11548-017-1649-7

23. Ma J, Wu F, Zhu J, Xu D, Kong D. A pre-trained convolutional neural network based method for thyroid nodule diagnosis. *Ultrasonics* (2017) 73(1):221–30. doi: 10.1016/j.ultras.2016.09.011

24. Wildman-Tobriner B, Buda M, Hoang JK, Middleton WD, Mazurowski MA. Using artificial intelligence to revise acr ti-rads risk stratification of thyroid nodules: diagnostic accuracy and utility. *Radiology* (2019) 292(1):112–9. doi: 10.1148/radiol.2019182128

25. Song W, Li S, Liu J, Qin H, Zhang B, Zhang S, et al. Multi-task cascade convolution neural networks for automatic thyroid nodule detection and recognition. *IEEE J BioMed Health Inform* (2019) 23(3):1215–24. doi: 10.1109/JBHI.2018.2852718

26. Mai W, Zhou M, Li J, Yi W, Li S, Hu Y, et al. The value of the demetics ultrasound-assisted diagnosis system in the differential diagnosis of benign from malignant thyroid nodules and analysis of the influencing factors. *Eur Radiol* (2021) 31(10):7936–44. doi: 10.1007/s00330-021-07884-z

27. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. *IEEE Trans Pattern Anal Mach Intell* (2017) 99:2999–3007. doi: 10.1109/TPAMI.2018.2858826

28. Foret P, Kleiner A, Mobahi H, Neyshabur B. Sharpness-aware minimization for efficiently improving generalization. *Int Conf Learn Representations* (2021). doi: 10.48550/arXiv.2010.01412

29. Tan M, Le QV. Efficientnet: Rethinking model scaling for convolutional neural networks. *Int Conf Mach Learn* (2019) 97:6105–14. doi: 10.48550/arXiv.1905.11946

30. Peng S, Liu Y, Lv W, Liu L, Zhou Q, Yang H, et al. Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study. *Lancet Digit Health* (2021) 3(4):e250–9. doi: 10.1016/S2589-7500(21)00041-8

31. Jia X, Ma Z, Kong D, Li Y, Hu H, Guan L, et al. Novel human artificial intelligence hybrid framework pinpoints thyroid nodule malignancy and identifies overlooked second-order ultrasonographic features. *Cancers* (2022) 14(18):4440. doi: 10.3390/cancers14184440

32. Cibas ES, Ali SZ. The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid* (2017) 27(11):1341–6. doi: 10.1089/thy.2017.0500

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 01 December 2022
DOI 10.3389/fendo.2022.1078019

OPEN ACCESS

EDITED BY
Salvatore Sorrenti,
Department of Surgical Sciences,
Faculty of Medicine and Dentistry,
Sapienza University of Rome, Italy

REVIEWED BY
Eleonora Lori,
Sapienza University of Rome, Italy
Fabio Medas,
University of Cagliari, Italy

*CORRESPONDENCE
Roberto Novizio
robertonovizio@gmail.com

SPECIALTY SECTION
This article was submitted to Thyroid Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 23 October 2022
ACCEPTED 16 November 2022
PUBLISHED 01 December 2022

CITATION
Taccogna S, Papini E, Novizio R, D'Angelo M, Turrini L, Persichetti A, Pontecorvi A and Guglielmi R (2022) An innovative synthetic support for immunocytochemical assessment of cytologically indeterminate (Bethesda III) thyroid nodules.
Front. Endocrinol. 13:1078019.
doi: 10.3389/fendo.2022.1078019

An innovative synthetic support for immunocytochemical assessment of cytologically indeterminate (Bethesda III) thyroid nodules

Silvia Taccogna[1], Enrico Papini[2], Roberto Novizio[3,4]*, Martina D'Angelo[1], Luca Turrini[1], Agnese Persichetti[5], Alfredo Pontecorvi[3,4] and Rinaldo Guglielmi[2]

[1]Pathology, Ospedale Regina Apostolorum, Albano Laziale, Italy, [2]Endocrinology and Metabolism, Regina Apostolorum Hospital, Rome, Italy, [3]Endocrinology and Metabolism, Agostino Gemelli University Polyclinic (IRCCS), Rome, Italy, [4]Catholic University of the Sacred Heart, Rome, Italy, [5]Service of Pharmacovigilance, Regina Elena National Cancer Institute, Hospital Physiotherapy Institutes (IRCCS), Rome, Italy

Background: Fine needle aspiration (FNA) is the procedure of choice in the evaluation of thyroid nodules. Nodules with indeterminate cytological categories, Bethesda III and IV, pose challenges in clinical practice and are frequently submitted to diagnostic surgery. CytoFoam Core (CFCS) uses an absorbent foam device inserted into the needle hub to collect the cytological sample aspirated during FNA. Specimen is formalin-fixed and paraffin-embedded.

Aim of the study: Assessing diagnostic efficacy of CFCS, compared to traditional cytology, in re-evaluating thyroid nodules classified as Bethesda III, using post-surgical histology as reference standard.

Method: Retrospective study on 89 patients with a first indeterminate cytological report who were referred to the Department of Endocrinology of Regina Apostolorum Hospital (Albano L. Rome, Italy) for a second FNA. FNA was performed after at least one month under ultrasound guidance with a 23G needle according to the established procedure. During the second procedure, both traditional cytological (TC) smears and a single-pass CFCS specimen were obtained for each patient. On CFCS samples immunocytochemical staining for Galectin-3, HBME-1, and CK-19 was also performed. 51 patients eventually underwent surgery, and their histological diagnoses were compared to the TC and CFCS reports. Four parameters were evaluated: inadequacy rate, rate of persistent indeterminate (Bethesda III and IV) reports, rate of malignancy in persistently indeterminate nodules, and rate of cancer in lesions cytologically classified as malignant.

Results: Non-diagnostic samples were 6 (11.8%) in TC vs 3 (5.9%) in CFCS (p=0.4). Persistent indeterminate samples were 31 (60.8%) in TC vs 19 (37.2%) in CFCS (p=0.01). Rate of malignancy in persistently indeterminate nodules was 8/19 (42.1%) in CFCS vs 9/31 (29%) in TC group (p=0.3). Nine/51 (17.6%) samples were classified as benign by TC vs 21/51 (41.2%) samples by CFCS (p<0.01). All nodules resulted benign at post-surgical evaluation. Five/51 (9.8%) samples were classified as suspicious for malignancy/malignant in TC group against 8/51 (15.7%) samples in CFCS (p=0.5). Post-surgical evaluation confirmed malignancy in all these cases.

Conclusion: CFCS demonstrated greater diagnostic accuracy than TC in repeat FNA assessment of cytologically indeterminate nodules. CFCS increased the conclusive diagnosis rate and decreased the number of cytologically indeterminate cases.

KEYWORDS

thyroid nodule, fine needle aspirate (FNA), immunoistochemestry, indeterminate thyroid cytology, Bethesda III category, Bethesda IV cytology, CytoFoam Core

Introduction

Thyroid nodules are increasingly detected in clinical practice due to the widespread access to imaging techniques involving the neck. The main issue in their management is to distinguish the minority of malignant lesions, which deserve surgery, from the vast majority of benign thyroid nodules that may be followed over time without intervention (1, 2). Fine needle aspiration (FNA) with ultrasound (US) guidance is the main diagnostic procedure for the assessment of the risk of malignancy of thyroid nodules, being safe, cost-effective and minimally invasive (3, 4). Throughout different studies, from 85% to 90% of US-guided FNA provide a sample adequate for cytological evaluation (5, 6), with a sensitivity ranging from 65% to 98%, a specificity of 72-100% and an accuracy of 84-95% (7). The Bethesda System for Reporting Thyroid Cytopathology (BSRTC) classifies the FNA outcome in 6 diagnostic categories including: (I) non-diagnostic; (II) benign; (III) atypia/follicular lesion of undetermined significance (AUS/FLUS); (IV) follicular neoplasm or suspicious for follicular neoplasm (FN/SFN); (V) suspicious for malignancy; (VI) malignant (8). A non-negligible number of FNA samples are classified as "indeterminate" cytological categories III and IV according to BSRTC, exhibiting a quite wide range of malignancy risk reported in literature, requiring different clinical actions (8, 9). These cytological categories pose a management challenge in clinical practice (9). Even if clinical, laboratory, and US findings offer useful data for refining the risk of thyroid cancer, many of these patients are eventually submitted to diagnostic surgery. According to current thyroid nodule guidelines, either surgery or molecular testing should be considered for patients with Bethesda IV cytology while for Bethesda III nodules a further cytological sampling is recommended. Moreover, managements guidelines are controversial in which surgery, total thyroidectomy and lobectomy, to be performed in AUS/FLUS or FN/SFN, whose management differs among institutions (1, 10, 11). Nevertheless, due to relatively low rate of malignancy revealed by post-surgical histology, the surgical approach may represent an overtreatment in a high number of cases, regardless the type of surgery (12).

Molecular testing can be employed to improve the accuracy of preoperatory diagnosis in thyroid nodules with indeterminate cytological report. Currently, multi-gene classifiers offer relevant sensitivity and high negative predictive value (NPV) but are still limited by a relatively low specificity and positive predictive value (PPV) (13). Most important, the elevated cost of these techniques make their routine use in clinical practice as extremely expensive for the National Health Services (NHS). Thus, at variance with their diffusion in the USA, only few centers in Europe regularly perform molecular testing as a routine complement to the diagnostic work up of class III and IV cytological samples (14, 15).

Immunohistochemical studies may provide complementary information about the nature of thyroid FNA samples (7). These tests are rather inexpensive and can be routinely performed in pathology departments. Main limitations of traditional procedures are the modalities of processing of the samples, which require working time and specific skill from the

operators, the uneven distribution and the possible loss of thyroid cells, and the potentially inadequate staining of intracellular antigens (16).

The CytoFoam Core system (CFCS) is proposed as an innovative technique that can provide optimal formalin-fixed and paraffin-embedded (FFPE) cytologic specimens, obtained with a single FNA pass according to the established sampling procedure (3). The CFCS samples are suitable for high quality immunohistochemical studies, which are performed without destruction of the cytological material that remains available for further studies.

Aim of the present study was to assess the technical feasibility and the cost of CFCS and its diagnostic accuracy in cytologically indeterminate thyroid nodules. Outcomes were compared to the results of traditional cytology, with the use of post-surgical histology as the reference standard.

Methods

Design of the study

Retrospective single center blinded study. Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Patients

From June 2019 to June 2020, 740 patients with solid not hyperfunctioning thyroid nodules were referred for FNA assessment to the thyroid clinic of the Department of Endocrinology of Regina Apostolorum Hospital, Albano, Rome. All sampling procedures were performed by two endocrinologists (EP, RG) and were examined in the Pathology Department of the same hospital by two experienced pathologists (ST and LT). FNA was performed with 23- or 25-gauge needles and the aspirated material was spread on 6 slides which were stained with both Papanicolaou and May-Grunwald Giemsa methods (16). Ninety-three patients (12.5%) had a low-risk indeterminate cytological report (Bethesda class III or Tir3A category, according to the Italian SIAPEC-AME-AIT-SIE classification) (5). According to current thyroid clinical practice guidelines, a second FNA was performed after 1 - 3 months for the definition of patients' clinical management (1). During the second FNA procedure, an additional sampling was performed using the CFCS in 89 patients. Fifty-one patients who eventually underwent surgery because of suspicious cytology at second evaluation, compressive or cosmetic symptoms, anxiety for malignancy risk, or suspicious clinical or US data were included in our retrospective study (Figure 1).

Cytofoam core procedure

Samples for the CFCS procedure were collected with a dedicated US-guided FNA, performed according to the standard procedure (17). An 8 x 3 mm cylinder of synthetic foam with elevated absorbent structure (Diapath, Martinengo company, Italy) was inserted between the hub of a 23 G/25 G needle and the aspirating syringe (Figure 2). The foam structure worked as a terrycloth, holding the cellular material aspirated during the FNA biopsy. After a 10 – 15 seconds aspiration, the foam core was removed by the needle hub, protected with a plastic guard cap, and placed into 10% neutral buffered formalin for 12 hours. Once fixed, the foam core was pulled from the adaptor, automatically processed, and embedded in paraffin blocks. Then, four sections were obtained and prepared to be

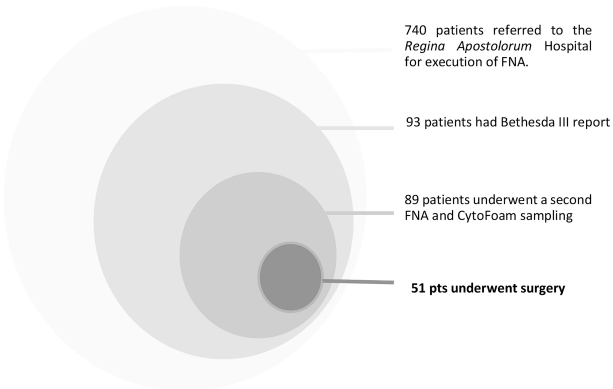


FIGURE 1
Patients Selection.

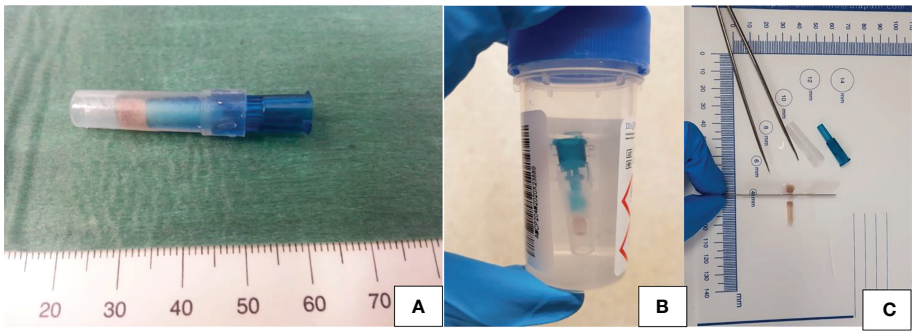


FIGURE 2
(A–C) Cytofoam core sampling: schematic illustration of the sample and of its processation. **(A)** Foam structure of the CFCS protected by a plastic guard cap before fixation; **(B)** CFCS sample fixation in 10% neutral buffered formalin; **(C)** Sections of the CFCS sample ready to be embedded in a paraffin block.

studied: the first section was stained with hematoxylin-eosin for morphologic evaluation, the others automatically treated for immunohistochemistry in a Dako Autostainer (Dako, Carpinteria, CA, USA) with antibodies for Galectin 3 (dilution 1:50, clone 9C4, Cell Marque), Cytokeratin 19 (ready to use, clone RCK 108 Dako Corporation, Carpentaria, California) and HBME1 (dilution 1:50, clone M3505, Dako Corporation, Carpentaria, California). The staining was completed using a streptavidin-biotin-complex detection method (LSAB2). The remaining material was stored for further possible examinations.

Surgical specimens were fixed in 10% buffered formaldehyde and embedded in paraffin, and 5-mm-thick microtome sections were stained with haematoxylin-eosin. Cytologic specimens and histologic sections were separately and blindly reviewed by two experienced pathologists of our center.

Study methodology

Thyroid CFCS specimens were analyzed by the two pathologists, and, for each patient, the results were blindly compared to those of the traditional cytological smears.

A four-class categorization of CFCS cytological findings was arbitrarily built. Class I identified non-diagnostic samples, class II included samples with benign characteristics, class III identified indeterminate samples, and class IV included samples with characteristics suspicious for malignancy (Figure 3). Specifically, CFCS samples were defined as: non diagnostic when the cell number or quality was insufficient for a reliable diagnosis; benign when architectural or cellular atypia were absent and the three immunocytochemical markers were consistently negative; as indeterminate when minimal architectural or cellular atypia were present and the immunocytochemical markers were partially (1-2 out of 3) positive; malignant when architectural or cellular atypia were severe and/or all three immunocytochemical markers were positive.

Post-surgical histology was used as the reference standard for the final diagnosis of the nodules under investigation.

The following parameters of diagnostic efficacy were analyzed: (I) percentage of CFCS non-diagnostic samples; (II) percentage of CFCS samples reported as benign which were confirmed as benign at histology; (III) percentage of persistently indeterminate samples; (IV) percentage of samples reported as indeterminate which resulted as malignant at histology; (V) percentage of samples reported as malignant which were confirmed as malignant at post-surgical histology. The diagnostic efficacy of traditional cytology and CFCS system were compared on the base of the final histologic report.

Statistical analysis

Data were collected on a Microsoft Excel database. X-square and, when appropriate, exact Fisher's test were used to compare results in traditional cytology group vs CFCS group. The level of significance was set at $\alpha < 0.05$. Data analysis was performed using SPSS v22 (IBM). An external monitor independently processed the data.

Results

Traditional cytology reports & results of CytoFoam categories vs. post-surgical histology

Four out of 6 (66.6%) non-diagnostic specimens with the second TC resulted as malignant at histology versus 2 out of 3

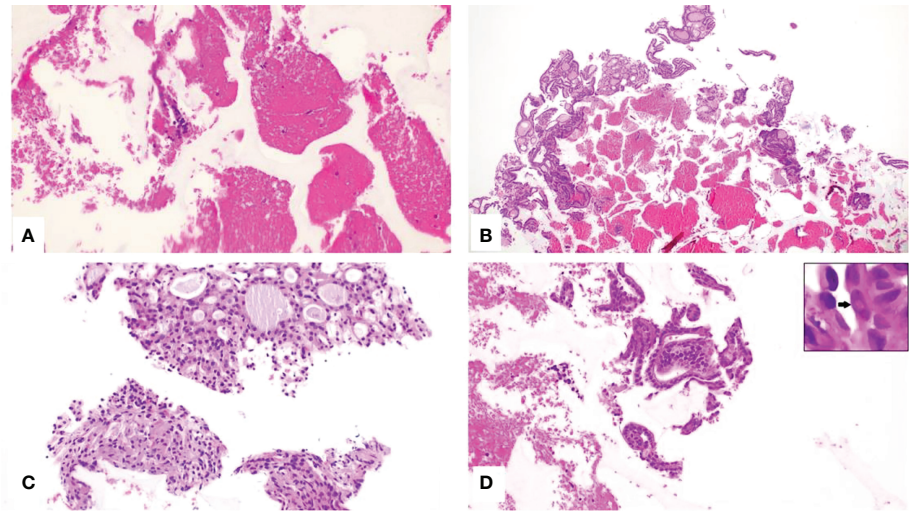


FIGURE 3
(A–D), CytoFoam Core Quality Classification. Hematoxilin & Eosin staining of the CFCS Histology, (H&E)x200: **(A)** Non-diagnostic sample due to the insufficient number of cells (Class I); **(B)** Sample from a benign thyroid nodule showing an adequate number of colloid containing follicles (Class II); **(C)** Indeterminate sample showing an adequate number of thyrocytes, mainly organized in microfollicular structures and with minor cellular atypia (Class III); **(D)** Tissue fragment with groups of irregular cells. The nuclei are variably enlarged with intranuclear inclusions (arrow inset) (Class IV).

(66.6%) of those with CFCS. 9 sample were classified as benign through traditional cytology, while 21 through CFCS: all classified as benign at post-surgical histology. Other diagnostic outcomes of CFCS and of traditional cytology are compared to the results of postsurgical histology in Tables 1, 2.

Comparison of the results of CytoFoam categories vs the traditional cytology classes

The occurrence of non-diagnostic samples (Bethesda I) was 6/51 (11.8%) with TC vs 3/51 (5.9%) with CFCS (p=0.4).

Nine of 51 (17.6%) samples were classified as benign (Bethesda II) with TC vs 21/51 (41.2%) with CFCS (p<0.01).

Persistent indeterminate samples (Bethesda III and IV) were 31/51 (60.8%) with TC vs 19/51 (37.2%) with CFCS (p=0.01).

Five of 51 samples (9.8%) were classified as suspicious for malignancy or malignant (Bethesda V and VI) with TC versus 8/51 (15.7%) samples in the CFCS group (p=0.55). (Table 3)

Traditional cytology vs. cytofoam. Rate of inadequate and indeterminate reports and rate of malignant histology in indeterminate, suspicious or malignant reports

All specimens classified as benign with traditional cytology or with CFCS resulted as benign at post-surgical histological evaluation (benign rate, 100%).

The rate of malignancy in persistently indeterminate nodules was 8/19 (42.1%) in CFCS group vs 9/31 (29%) in TC group (p=0.3).

TABLE 1 Traditional cytology reports vs. post-surgical histology.

Cytological diagnosis	Number of reports	Benign at post-surgical histology N (%)	Malignant at post-surgical histology N (%)
Bethesda I	6	2/6 (33.3%)	4/6 (66.6%)
Bethesda II	9	9/9 (100%)	0/9 (0%)
Bethesda III	22	15/22 (68.2%)	7/22 (31.8%)
Bethesda IV	9	7/9 (77.8%)	2/9 (22.2%)
Bethesda V	4	0/4 (0%)	4/4 (100%)
Bethesda VI	1	0/1 (0%)	1/1 (100%)

TABLE 2 CytoFoam Core System reports vs. post-surgical histology.

CytoFoam Class	Number of reports	Benign at post-surgical histologyN (%)	Malignant at post-surgical histologyN (%)
I	3	1/3 (33.3%)	2/3 (66.6%)
II	21	21/21 (100%)	0/21 (0%)
III	19	11/19 (57.9%)	8/19 (42.1%)
IV	8	0/8 (0%)	8/8 (100%)

All specimens classified as malignant with TC (Bethesda V or VI) or with CFCS (IV) resulted as malignant at post-surgical evaluation (malignancy rate, 100%). (Table 4)

Traditional cytology vs. cytofoam. Rate of conclusive reports

As a whole, out of the 51 nodules classified as Bethesda III at initial cytological assessment, 14 (27.4%%) had a conclusive diagnosis, either benign or malignant, with the second cytological sample versus 29 (56.8%) with the CFCS specimen (p=0.002). (Table 5)

The use of CFCS has statistically significantly increased the number of cytologically benign reports, Moreover, the number of indeterminate was statically reduced, with an increase in the rate of conclusive report, while conserving 0% of rate of malignant histology in cytological benign reports and 100% of rate of malignant histology in cytological suspicious or malignant reports, so not influencing the negative predictive value and positive predictive value among cytologially benign or malignant sample, respectively. Results of all the Bethesda III patients' re-assessment are shown in Table 6. Relevant results are summarized in Table 7.

Immunocytochemical staining

The quality of the immunocytochemically stained samples was arbitrarily classified as follows: 0 when the staining was negative, 1 when positive, and 2 when inadequate for evaluation. Based on this classification, the immunocytochemically stained samples were rated as negative in 14 cases (27.4%), as positive in 33 cases (64.7%), and as inadequate in 4 cases (7.8%).

Complications

No local anaesthesia was requested. No major nor minor complications were reported with the two procedures and no patient required post-procedural clinical or US observation. Pain was described as well tolerated and, according to a 1 to 10 visual analogue scale, was classified as a mean of 2 with the conventional FNA (range 2 to 4) and as 3 (2 to 5) with the CFCS procedure (p = 0.8). No postprocedural medication was necessary with both the procedures.

Time

The time employed for the two sampling procedures was similar, with a range from 15 to 30 seconds.

Discussion

Despite the advances in US imaging, especially through artificial intelligence systems and lately the possible application of contrast in ultrasound, and the promising results of molecular analysis, the main diagnostic step for thyroid nodules still relies on FNA results (5, 18, 19). FNA procedure is minimally invasive, performed with negligible patient discomfort, and offers an elevated diagnostic accuracy (3). However, a considerable percentage of thyroid cytological samples do not reach a

TABLE 3 Comparison of the results of CytoFoam categories vs the traditional cytology classes.

CytoFoam Category	BSRTC	TC	CFCS	P-value	Post -surgical histology
I	Bethesda I	6 (11.8%)	3 (5.9%)	0.4 *	/
II	Bethesda II	9 (17.6%)	21 (41.2%)	<0.01	33
III	Bethesda III - IV	31 (60.8%)	19 (37.2%)	0.01	/
IV	Bethesda V-VI	5 (9.8%)	8 (15.7%)	0.5 *	18

TC: traditional cytology; CFCS: cytofoam.
*Exact Fisher's test.

TABLE 4 Traditional cytology vs cytofoam. Rate of inadequate and indeterminate reports and rate of malignant histology in indeterminate, suspicious or malignant reports.

Report	Traditional Cytology	CytoFoam	P-value
Rate of malignant histology in cytological benign reports	0/9 (0%)	0/21 (0%)	/
Rate of malignant histology in cytological indeterminate reports	9/31 (29%)	8/19 (42.1%)	0.3 *
Rate of malignant histology in cytological suspicious or malignant reports	5/5 (100%)	8/8 (100%)	/

*Exact Fisher's test.

conclusive diagnosis and are classified as non-diagnostic or at indeterminate risk of malignancy (20). According to current thyroid nodule guidelines, either surgery or molecular testing should be considered for patients with Bethesda IV cytology while for Bethesda III nodules a further cytological sampling is recommended (1). In case of a persistent Bethesda III cytological diagnosis either the assessment of molecular markers, active surveillance or surgery are suggested, based on the clinical condition, the local resources, and the patient preferences. Presently, a non-negligible percentage of patients with persistent Bethesda III diagnosis eventually undergo thyroid surgery, generally for diagnostic purpose. The risk of malignancy for Bethesda III thyroid nodules reported in literature shows a a quite wide range, from 19% to 55% in populations with environmental risk factors (e.g. endemic goiter) (4, 21–24). In the majority of these cases, surgery results in a benign lesion at histology but thyroidectomy carries a non-negligible risk of complications, increases healthcare cost and may negatively influence the quality of life of patients (4).. Most patients could then benefit more from a watchful waiting rather than from surgery. Thus, the use of ancillary tests is advocated to allow a more accurate pre-operatory stratification of the risk of malignancy and decrease the still elevated number of diagnostic thyroid surgeries.

Immunohistochemistry (ICC) was introduced in the 1970s as a diagnostic tool for both surgical pathology and cytopathology (16). While ICC plays a relevant role in the differential diagnosis between follicular and C cell-derived neoplasms and in the identification of primary or metastatic thyroid neoplasms, its usefulness in the pre-operatory assessment of risk of malignancy in follicular-patterned lesions is still unsettled (1, 25). Results from histological specimens may be discrepant from those obtained from cytological samples due to differences in fixatives, fixation methods, and/or antigen activation treatment. Immunostaining of histological specimen is carried out using formalin-fixed, paraffin-embedded, tissues with or without antigen retrieval. On the other hand, immunostaining of FNA smears is usually carried out using 95% alcohol-fixation, followed by a further fixation with phosphate-buffered formalin solution, and by an antigen activating treatment (16). Due to the low reliability of alcohol-fixation for the staining of target antigens, an additional formalin fixation (double fixation) is generally needed and it is followed by the antigen activating treatment. Formalin fixation and antigen activation treatment provide reliable results, but the process may result in loss of cell material and deterioration of the sample under examination. These supplementary investigations are best performed on cell blocks, but cell-block preparation is burdensome in high-volume laboratories. All these factors have until now limited the use of immunocytochemistry as a complement to the morphological diagnosis in indeterminate cytological samples.

The combination of CK-19, HBME-1 and galectina-3 immunocytochemistry is the most useful ancillary technique for improving the differential diagnosis in follicular-derived thyroid nodules. Notably, these markers show a diffuse reactivity in true malignant lesions (follicular cancer, classical variant of papillary cancer, and follicular variant of papillary cancer) while a focal staining is observed in benign neoplastic (follicular adenoma) and benign non-neoplastic (nodular goiter) lesions. In various studies, the sensitivity of CK19, galectin-3, HBME-1 was 75.41%, 88.52% and 71.31% respectively, and the specificity of CK19, galectin-3 and HBME-1 was 70.89%, 64.56% and 84.81% (26). The aim of the CFCS testing is to improve the diagnostic accuracy of FNA for guiding clinical action in cytologically indeterminate nodules. Only three diagnostic classes were considered for operative purposes: benign, indeterminate, and probably malignant. Consequently, persistently indeterminate nodules (Bethesda III-IV) are included in the CFCS class III and suspicious/neoplastic nodules (Bethesda V and VI) in the CFCS class IV. This

TABLE 5 Traditional cytology vs cytofoam. Rate of conclusive reports.

Report	Traditional Cytology	CytoFoam	P-value
Rate of conclusive reports	14/51 (27.4%)	29/51 (56.8%)	<0.01

10.3389/fendo.2022.1078019

TABLE 6 Results of the Bethesda III patients' re-assessment. Cytological diagnosis vs. immunocytochemical classes.

ID PTS	Bethesda Category	HMBE 1	Galectin 3	Cytokeratin 19	CytoFoam Class	M vs B
1	III	0	0	0	3	1
2	V	2	2	2	4	1
3	II	1	1	1	2	0
4	I	1	1	1	2	0
5	I	1	1	2	3	1
6	IV	1	1	1	2	0
7	IV	1	1	2	3	0
8	V	2	2	2	4	1
9	IV	1	1	2	3	1
10	IV	1	1	1	2	0
11	II	1	1	1	2	0
12	IV	1	1	2	2	0
13	II	0	0	0	2	0
14	II	1	1	1	2	0
15	IV	1	1	1	2	0
16	IV	1	2	2	3	1
17	IV	2	1	2	3	1
18	IV	2	1	1	3	0
19	V	2	2	2	4	1
20	IV	2	1	2	3	0
21	IV	1	1	1	2	0
22	IV	0	0	0	1	0
23	IV	2	1	2	3	0
24	IV	2	2	2	4	1
25	II	2	1	2	3	0
26	III	2	1	2	3	0
27	III	1	1	2	2	0
28	III	1	2	2	3	0
29	V	2	2	2	4	1
30	III	1	1	2	2	0
31	VI	2	2	2	4	1
32	III	1	1	2	3	1
33	III	1	1	2	3	1
34	III	1	1	1	2	0
35	III	2	1	1	2	0
36	I	2	1	1	3	1
37	II	1	1	1	2	0
38	I	2	2	2	4	1
39	I	1	1	2	2	0
40	III	2	1	2	2	0
41	III	1	1	1	2	0
42	IV	2	1	2	3	0
43	III	1	1	2	1	1
44	I	1	2	1	4	1
45	III	0	0	0	1	1
46	II	1	1	1	2	0
47	II	1	1	1	2	0
48	II	2	1	2	3	0

(Continued)

TABLE 6 Continued

ID PTS	Bethesda Category	HMBE 1	Galectin 3	Cytokeratin 19	CytoFoam Class	M vs B
49	II	2	1	2	3	0
50	III	1	1	2	3	0
51	III	1	1	1	2	0

From left to right: id pts, identification code of the patient; SIAPEC-AME-AIT-SIE report; HMBE1 evaluation, where 0 non-evaluated or non-diagnostic, 1 means negative, 2 means positive; Galectin 3 evaluation, where 0 non-evaluated or non-diagnostic, 1 means negative, 2 means positive; Cytokeratin 19 evaluation, where 0 non-evaluated or non-diagnostic, 1 means negative, 2 means positive; CytoFoam class report; M (malignant) vs B (benign), where 0 means a benign histology report, while 1 means a malignant histology report.

TABLE 7 Statistically relevant results of CytoFoam categories vs traditional cytology classes.

CytoFoam Category	Cytology	TC	CFCS	P-value
II	Bethesda II	9 (17.6%)	21 (41.2%)	<0.01
III	Bethesda III - IV	31 (60.8%)	19 (37.2%)	0.01
Report				
Rate of malignant histology in cytological benign reports		0/9 (0%)	0/21 (0%)	/
Rate of malignant histology in cytological suspicious or malignant reports		5/5 (100%)	8/8 (100%)	/
Rate of conclusive reports		14/51 (27.4%)	29/51 (56.8%)	<0.01

simplified classification can be achieved because the evaluation of thyroid nodules' risk of malignancy is based on both morphological criteria and immunohistochemistry results. The results from our study demonstrate that the use of CFCS provides high quality immunocytochemical staining in the vast majority (92%) of cytological samples. When compared to traditional cytology, CFCS provides an increase, from 27.4 to 56.8%, in the number of conclusive diagnosis obtained with repeat FNA sampling. The predictive value of the immunophenotypic assessment is confirmed by the elevated concordance of the cytological diagnosis with the final post-surgical assessment (100% concordance for both benign and malignant diagnosis). Notably, when all the three markers were negative, the NPV for thyroid cancer was 100% providing a reliable rule-out test. Three suspected cases classified at CFCS class III (cases #25-48-49), resulted at histological examination to be neoplasms, even if non-malignant ones (one follicular adenoma and three NIFTP). So, the surgical indication provided by CFCS examination may be considered as appropriate.

These favorable outcomes are mostly due to the sample characteristics, that are similar to those of a micro-histological specimen and are comparable to those obtained by the more expensive, invasive, and difficult-to-perform core-needle biopsy. Importantly, the immunophenotypic evaluation does not induce any deterioration of the specimen during the staining procedure. On clinical grounds, the CFCS procedure does not require additional time or any observation period when compared to the traditional FNA biopsy and is well tolerated by the patients. The increase in costs of this malignancy rule-out test is modest, as the price of the CFCS device is about 8 euros (27).

In conclusion, this feasibility study demonstrates that cytofoam core technique is a simple, safe, inexpensive, and reproducible procedure. The short processing time and the use of routine technical resources make this modality of immunocytochemical assessment of thyroid FNA samples suitable for routine use in most pathology laboratories. When testing of molecular markers is not accessible, immunocytochemical staining with the use of CFCS may provide - in addition to the clinical, laboratory, and US data - a further relevant element in the multifactorial choice of either surgical resection or follow-up for thyroid nodules with indeterminate cytology. Low numerical sample appears to be the main limitation of the study.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

The authors declare that this study received funding from Diapath S.p.A. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Haugen BR. 2015 American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: What is new and what has changed? *Thyroid* (2016) 26(1):1–33. doi: 10.1002/cncr.30360

2. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European Thyroid association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: The EU-TIRADS. *Eur Thyroid J* (2017) 6(5):225–37. doi: 10.1159/000478927

3. Cappelli C, Pirola I, Gandossi E, de Martino E, Agosti B, Castellano M. Fine-needle aspiration cytology of thyroid nodule: Does the needle matter? *South Med J* (2009) 102(5):498–501. doi: 10.1097/SMJ.0b013e31819c7343

4. Medas F, Erdas E, Gordini L, Conzo G, Gambardella C, Canu GL, et al. Risk of malignancy in thyroid nodules classified as TIR-3A: What therapy? *Int J Surg* (2017) 41(Suppl 1):S60–4. doi: 10.1016/j.ijsu.2017.03.056

5. Cibas ES, Ali SZ. The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid* (2017) 6(6):217–22. doi: 10.1016/j.jasc.2017.09.002

6. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 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* (2015) 2016:26. doi: 10.1089/thy.2015.0020

7. Ulisse S, Baldini E, Lauro A, Pironi D, Tripodi D, Lori E, et al. Papillary thyroid cancer prognosis: An evolving field. *Cancers* (2021) 13(21):5567. doi: 10.3390/cancers13215567

8. Ali S, Cibas ES. *The Bethesda system for reporting thyroid cytopathology*. Cham: Springer Cham (2018). doi: 10.1007/978-3-319-60570-8

9. Nardi F, Basolo F, Crescenzi A, Fadda G, Frasoldati A, Orlandi F, et al. Italian Consensus for the classification and reporting of thyroid cytology. *J Endocrinol Invest* (2014) 37(6):593–9. doi: 10.1007/s40618-014-0062-0

10. Conzo G, Pasquali D, Bellastella G, Esposito K, Carella C, de Bellis A, et al. Total thyroidectomy, without prophylactic central lymph node dissection, in the treatment of differentiated thyroid cancer. clinical retrospective study on 221 cases. *World J Surg Oncol* (2014) 12:152. doi: 10.1186/1477-7819-12-152

11. Wiseman SM, Baliski C, Irvine R, Anderson D, Wilkins G, Filipenko D, et al. Hemithyroidectomy: The optimal initial surgical approach for individuals undergoing surgery for a cytological diagnosis of follicular neoplasm. *Ann Surg Oncol* (2006) 13(3):425–32. doi: 10.1245/ASO.2006.03.089

12. Dobrinja C, Trevisan G, Piscopello L, Fava M, Liguori G. Comparison between thyroidectomy and hemithyroidectomy in treatment of single thyroid nodules identified as indeterminate follicular lesions by fine-needle aspiration cytology. *Ann Ital Chir* (2010) 81(6):403–10.

13. Nikiforov YE, Steward DL, Carty SE, Sippel RS, Yang SP, Sosa JA, et al. Performance of a multigene genomic classifier in thyroid nodules with indeterminate cytology: A prospective blinded multicenter study. *JAMA Oncol* (2019) 5(2):204–12. doi: 10.1001/jamaoncol.2018.4616

14. Rao SN, Bernet V. Indeterminate thyroid nodules in the era of molecular genomics. *Mol Genet Genomic Med* (2020) 8(9):e1288. doi: 10.1002/mgg3.1288

15. Possieri C, Locantore P, Salis C, Bacci L, Aiello A, Fadda G, et al. Combined molecular and mathematical analysis of long noncoding RNAs expression in fine needle aspiration biopsies as novel tool for early diagnosis of thyroid cancer. *Endocrine* (2021) 72(3):711–20. doi: 10.1007/s12020-020-02508-w

16. Rezk S, Khan A. Role of immunohistochemistry in the diagnosis and progression of follicular epithelium-derived thyroid carcinoma. *Appl Immunohistochem Mol Morphol* (2005) 13(3):256–64. doi: 10.1097/01.pai.0000142823.56602.fe

17. Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedüs L, et al. American Association of clinical endocrinologists, American college of endocrinology, and associazione Medici endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules - 2016 update. *Endoc Pract* (2016) 22(5):622–39. doi: 10.4158/EP161208.GL

18. Sorrenti S, Dolcetti V, Radzina M, Bellini MI, Frezza F, Munir K, et al. Artificial intelligence for thyroid nodule characterization: Where are we standing? *Cancers (Basel)* (2022) 14(14):3357. doi: 10.3390/cancers14143357

19. Sorrenti S, Dolcetti V, Fresilli D, del Gaudio G, Pacini P, Huang P, et al. The role of ceus in the evaluation of thyroid cancer: From diagnosis to local staging. *J Clin Med* (2021) 10(19):4559. doi: 10.3390/jcm10194559

20. Davies L, Welch HG. Current thyroid cancer trends in the united states. *JAMA Otolaryngol Head Neck Surg* (2014) 140(4):317–22. doi: 10.1001/jamaoto.2014.1

21. Rosario PW. Thyroid nodules with atypia or follicular lesions of undetermined significance (Bethesda category III): Importance of ultrasonography and cytological subcategory. *Thyroid* (2014) 24(7):1115–20. doi: 10.1089/thy.2013.0650

22. Yoo MR, Gweon HM, Park AY, Cho KE, Kim JA, Youk JH, et al. Repeat diagnoses of Bethesda category III thyroid nodules: What to do next? *PloS One* (2015) 10(6):e0130138. doi: 10.1371/journal.pone.0130138

23. Conzo G, Calò PG, Gambardella C, Tartaglia E, Mauriello C, della Pietra C, et al. Controversies in the surgical management of thyroid follicular neoplasms. retrospective analysis of 721 patients. *Int J Surg* (2014) 12(Suppl 1):S29–S34. doi: 10.1016/j.ijsu.2014.05.013

24. Calò PG, Medas F, Santa Cruz R, Podda F, Erdas E, Pisano G, et al. Follicular nodules (Thy3) of the thyroid: Is total thyroidectomy the best option? *BMC Surg* (2014) 14(1):12. doi: 10.1186/1471-2482-14-12

25. Pusztaszeri M, Rossi ED, Auger M, Baloch Z, Bishop J, Bongiovanni M, et al. The bethesda system for reporting thyroid cytopathology: Proposed modifications and updates for the second edition from an international panel. *Acta Cytol* (2016) 60(5):399–405. doi: 10.1159/000451020

26. Dunderović D, Lipkovski JM, Boričic I, Soldatović I, Cvejić D, Tatić S, et al. Defining the value of CD56, CK19, galectin 3 and HBME-1 in diagnosis of follicular cell derived lesions of thyroid with systematic review of literature. *Diagn Pathol* (2015) 10(1):196. doi: 10.1186/s13000-015-0428-4

27. Taccogna S, Guglielmi R, Persichetti A, Morano C, Angelini F, Ienzi S, et al. Carcinomas of the thyroid with Ewing family tumor elements (CEFTEs): A diagnostic challenge before surgery. *Head Neck Pathol* (2021) 15(1):254–61. doi: 10.1007/s12105-020-01145-z

frontiers | Frontiers in Endocrinology

TYPE Mini Review
PUBLISHED 23 January 2023
DOI 10.3389/fendo.2023.1113977

OPEN ACCESS

EDITED BY
Jeffrey Garber,
Atrius Health, United States

REVIEWED BY
Yasemin Giles Senyürek,
Istanbul University, Türkiye

*CORRESPONDENCE
Benjamin James
bjames1@bidmc.harvard.edu

SPECIALTY SECTION
This article was submitted to Thyroid Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 01 December 2022
ACCEPTED 09 January 2023
PUBLISHED 23 January 2023

CITATION
Uppal N, Collins R and James B (2023) Thyroid nodules: Global, economic, and personal burdens.
Front. Endocrinol. 14:1113977.
doi: 10.3389/fendo.2023.1113977

Thyroid nodules: Global, economic, and personal burdens

Nishant Uppal[1,2], Reagan Collins[3] and Benjamin James[1,3]*

[1]Harvard Medical School, Boston, MA, United States, [2]Department of Medicine, Brigham and Women's Hospital, Boston, MA, United States, [3]Division of Surgical Oncology, Department of Surgery, Beth Israel Deaconess Medical Center, Boston, MA, United States

Thyroid nodules have garnered attention due to changes in population surveillance systems and rising concerns about the associated financial burden on healthcare systems, payers, and patients. In this review, we find that prevalence rates vary widely based on method of detection and may particularly pronounced in asymptomatic patients undergoing routine screening. Incidence rates may be particularly rising in lower-income and middle-income countries and may be declining in higher-income countries. Despite high incidence rates, survival rates continue to be as high as 97% for papillary thyroid cancer. Over the last few decades, thyroid nodule workup and management has grown more sophisticated with the advent of fine-needle aspiration biopsy, specialized biomarkers, and molecular testing. However, gaps remain in risk stratification that can lead to substantial costs of care. Certain molecular tests, such as the Afirma Gene Sequencing Classifier can lead to a cost per diagnosis of $17,873 while achieving only mild decreases in diagnostic lobectomies for patients (11.6% to 9.7% in one study). Out-of-pocket costs associated with thyroid nodule management continue to drive significant financial toxicity for patients, especially for individuals with thyroid cancer. Financial toxicity has been defined as a term that describes how direct and indirect medical costs of cancer care strain patients and households *via* decreased income, assets, and spending on basic necessities. Recent studies suggest that such toxicity can lead to adverse financial outcomes, such as foreclosure and bankruptcy. Additional cost-effectiveness analyses are needed to improve existing thyroid nodule management systems and new clinical tools are needed to avoid unnecessary workup and management.

KEYWORDS

thyroid nodule, cost effectiveness, financial burden, healthcare policy and management, economic impact

Introduction

Increasing rates of thyroid nodule detection has prompted assessment of the global, economic, and patient-borne burden associated with the evaluation and treatment of benign and malignant disease. The global burden of thyroid nodules reflects differences in population surveillance across countries that has led to variation in thyroid nodule

incidence rates. For payers, disparate systems for covering healthcare costs yields unique economic considerations associated with financing the costs of thyroid nodule workup and management. High out-of-pocket costs also lead to patient concerns with managing diagnostic workup and treatment of thyroid nodules. Here, we describe what is known about the global, economic, and patient-borne burdens of thyroid nodule management and outline strategies for mitigating the societal and financial implications associated with potentially unnecessary or extraneous care.

Global burden of thyroid nodules

The prevalence of thyroid nodules among the general population has been estimated as upwards of 67% depending on mode of detection (palpation, ultrasound or autopsy) and varies widely by country (1–3). Prevalence ranges between 34% to 66% depending on ultrasound detection rates or autopsy findings (2, 4). Female sex, higher body mass index (BMI), and older age are associated with an increased prevalence of thyroid nodules (5, 6). Although high prevalence rates suggest a significant burden of disease, most thyroid nodules are benign or have no ultrasound features to suggest malignancy and thus are largely clinically insignificant (7, 8). When evaluating mechanism of nodule diagnosis on a global scale, Sajisevi et al. found variation across participating countries (9). Rates of nodule diagnosis secondary to symptomatic presentation were much higher in South Africa and Denmark at 79% and 54%, respectively, while rates were similar and much lower in the United States and Canada at around 30% (9). However, thyroid surgery was performed more often in asymptomatic patients in the United States and Canada which has substantial impact when considering the effective management of nodules without overtreatment (9).

The widespread adoption of sensitive imaging techniques has contributed to the increasing frequency of detection of incidental thyroid nodules (10). Due to the relatively indolent nature of thyroid nodules, the primary clinical concern is excluding malignancy. Thus, the complex diagnostic assessment of thyroid nodules largely pertains to determining clinical significance while avoiding overdiagnosis and overtreatment (11, 12). Of detected nodules, 10-15% represent malignant disease (11). Globally, incidence rates of thyroid cancer have grown substantially with the adoption of widespread thyroid ultrasound use, which has raised concerns of the overdiagnosis of subclinical thyroid disease. In the United States, the incidence of thyroid cancer tripled from 1975 to 2009 (13), which resembles trends in other countries, such as South Korea (14, 15). Increases are largely due to increasing detection of low-risk subclinical papillary thyroid microcarcinoma (14). Survival rates have remained as high as 97% for papillary thyroid cancer, the most common type of thyroid cancer (16). From 1978 to 2007, mortality rates steadily declined in most countries with reported mortality rate reductions of 43.2% for men, and 50% for women (17). Further, men in China and women in Australia were noted to have the largest decreases in mortality rates during this time (17). Consideration of the falling mortality rates despite the rising incidence further supports the concern for overdiagnosis and overtreatment of thyroid cancer on a global scale (16).

The trend of rising incidence began to regress first in South Korea in 2014 with a 30% reduction in nodule detection in response to less screening, and as a result, less diagnosis (15). Similarly, recent findings from the Global Burden of Disease Study found that incidence rates have started to plateau in EU15+ nations and in the US between 2011 and 2019 (18). However, in low- and middle-income countries, incidence rates have continued to rise (19). Although the incidence in high-income countries may be decreasing, overall rates of thyroid cancer are still highest in these countries with the most incident cases in China, the United States, and India (18, 20). Further, significant variability in reported rates exists globally. For instance, there is a fivefold difference in thyroid cancer incidence rates in women across various regions of the world (17). Despite the regional variation in incidence by sex, the observed female to male ratio is relatively consistent across all regions at 3:1 (17). Possible contributing factors to the regional variation include barriers to access to health care, higher levels of radiation exposure, and iodine deficiency present in certain low- and middle-income countries (20, 21). This suggests that although over screening and overdiagnosis could be contributing to the high incidence in high-income countries, the variability in other parts of the world may be a true rise in incidence due to environmental exposures or modifiable risk factors. However, recent work on US trends in thyroid cancer mortality has suggested that incidence-based mortality may be growing by as much as 1.1% annually for all thyroid cancer patients and 2.9% for advanced-stage papillary thyroid carcinoma (22). Therefore, robustly characterizing the burden of thyroid nodules may require additional research on thyroid cancer incidence and mortality that accounts for demographic and tumor characteristics.

Economic burden of thyroid nodule care

While the global burden of disease has been frequently reported, the economic burden associated with thyroid nodules is only partially understood. For nodules representing malignant disease, the costs of well-differentiated thyroid cancer care in the United States are projected to exceed $3.5 billion by 2030 (23). The plurality (41%) of healthcare expenditures is incurred for newly diagnosed patients. Initial diagnosis and evaluation, including primary care provider visit, endocrinology/surgical consultations, ultrasound imaging, and fine-needle aspiration biopsy (FNA) drive the economic burden of thyroid nodule management for both benign and malignant disease. American Thyroid Association (ATA) guidelines suggest that FNA is the most cost-effective method for evaluating thyroid nodules and further recommends ultrasound guidance, which has been shown to achieve better diagnostic accuracy than palpation alone (24–26). For incidental thyroid nodules that are less than 2 centimeters, the cost-effectiveness of FNA appears poor compared to observation ($542 vs. $412 in direct costs) (27).

Prior cost-effectiveness analyses have estimated that the screening and management of all thyroid nodules in the United States would incur $25.1 billion in costs, and the addition of specialized biomarkers, such as serum calcitonin for medullary thyroid cancer, to current ATA guidelines could add $1.4 billion in costs, which

would represent a mean $11,793 per life-year saved (28). Rather than routine use of adjunct testing, other studies of the cost of thyroid nodule evaluation considered adding molecular testing only for individuals with indeterminate thyroid nodules based on initial FNA cytology results (29), which represent 20-30% of FNA results. Yip et al. found that while molecular testing added $104 per patient to the costs of thyroid nodule diagnostic workup, cost savings were realized by a decrease in the proportion of diagnostic lobectomies compared to standard care (11.6% to 9.7%) (29). The Afirma Gene Sequencing Classifier™ and ThyroSeq™ are two nucleic-acid based molecular tests that use gene expression profiling and/or genotyping of tumor-associated genetic mutations to attempt to determine the likelihood that samples represent malignancy (28–31). The cost-effectiveness of molecular testing also varies between Afirma, which may be more costly than lobectomy (30), and ThyroSeq v3, which was shown in a single-center Canadian study to reduce the number of diagnostic lobectomies (31). A comparative study of both molecular testing options suggested that for indeterminate nodules, both Afirma and ThyroSeq v3 were more cost-effective than lobectomy, but ThyroSeq v3 yielded a cost per diagnosis of $14,277 compared to $17,873 for the Afirma Gene Sequencing Classifier (32). Molecular tests may be used more often in the United States than in other countries. However, due to the relatively recent emergence and evolution of molecular testing, the particular extent to which the use of such tests varies between countries has not yet been fully characterized.

Although not routinely performed, intraoperative frozen section analysis can also be a driver of the economic burden of thyroid nodule management and includes potential costs from testing, labor, extended operating room time, and completion thyroidectomies in some cases (33, 34). One meta-analysis suggested that frozen section analysis offered only moderate diagnostic utility (sensitivity, 95% CI: 43%, 34%-53%) and routine use should be discouraged for follicular neoplasms (34). A separate cost analysis instead suggested that routine use of frozen section for patients with "suspicious for malignancy" cytology during thyroid lobectomy could actually achieve costs of $474 per case, primarily due to a large reduction in rates of subsequent total thyroidectomy compared to standard care (7.7% vs. 26.1%) (35).

Patient-borne financial burden of thyroid nodule management

Finally, the patient-borne financial burden of thyroid disease has been assessed using both out-of-pocket costs and perceived financial toxicity as primary metrics. Out-of-pocket costs are driven by the surgical management of thyroid disease, which are substantial for both benign and malignant conditions and pronounced even for commercial insured patients (15). However out-of-pocket costs for patients who do not undergo surgical management for thyroid nodules remain due to the diagnostic sequelae of incidental detection, including active surveillance which includes lab testing and recurrent imaging. Patients who self-identify with overdiagnosed thyroid cancer but opt for nonintervention are at risk for healthcare disengagement and lower quality-of-life (36). Current estimates of the perceived financial burden rely primarily on cohort and cross-sectional studies of thyroid cancer patients, which have shown that 46.1% of patients endorse a psychological financial burden and 28.1% of patients endorse a material financial burden (37). There is also evidence of household strain associated with thyroid cancer diagnosis and treatment on patients with 48% patients reporting reduced income, 9% losing insurance coverage as well as 18.1% reporting unemployment for at least 6 months (38, 39). Thyroid cancer care has also been associated with adverse financial outcomes, including a higher likelihood of notice of default and foreclosure and bankruptcy compared to other cancer types (40, 41). Notably, bankruptcy rates have been estimated to be as high as 41% at 5 years after diagnosis despite high survival rates (98% at 5 years after diagnosis) (42). Previously, we have summarized the financial burden of thyroid cancer and outlined frameworks for improving research designed to measure and mitigate the financial burden of care (43).

Evidence further suggests that overdiagnosis and overtreatment of thyroid cancer can also impair health-related quality of life (HRQoL) for patients. Thyroid cancer survivors cite declines in psychological and emotional well-being due to anxiety and depression associated with treatment, and these symptoms may persist during remission because patients often fear recurrence of cancer (44). For patients who have undergone thyroidectomy, surveillance costs can also contribute to reduced quality of life and excess out-of-pocket spending, especially since the cost to detect 1 recurrence has been estimated at $147,819 (45, 46). As we have summarized previously, the costs of thyroid cancer diagnosis and treatment lead many patients to delay care and may risk spending on other medical conditions that contribute substantially to improved health, quality of life, and lifespan (43).

Discussion

Recent retrospective analyses have found that 41% of patients undergoing surgical treatment for thyroid nodules have no thyroid-referable symptoms at the time of detection, and the mean tumor size is smaller in asymptomatic patients (2.1 cm) compared to symptomatic patients (3.2 cm) (9). An additional meta-analysis showed that 68.8% of all thyroid nodules undergoing surgical excision represented benign disease (47). This suggests that increasing detection of benign and subclinical disease may be generating excess healthcare costs. The thyroid nodule diagnostic workup routinely involves ultrasound imaging and FNA for nodules considered suspicious for malignancy based on sonographic features. In the United States, the American College of Radiology Thyroid Imaging Reporting and Data System (TIRADS) risk stratification system is used to guide subsequent management of thyroid nodules undergoing sonographic evaluation (48). This points system creates categories and biopsy thresholds for consideration of FNA based on risk of malignancy but does not additionally incorporate cost-effectiveness estimates nor stratifies by thyroid cancer subtype. The latter is particularly important considering the significant differences in 5-year relative survival rates between patients with follicular and medullary thyroid cancer compared to papillary thyroid cancer (49). Thus, there are likely many patients with borderline radiographic features (i.e. TR2 and TR3 classifications) who still undergo unnecessary FNA despite low malignancy risk and potentially little benefit from earlier detection of indolent follicular

thyroid cancer types. In an analogous fashion, the Bethesda classification system for thyroid cytopathology may lead to unnecessary thyroid surgery in many patients with indeterminate typical findings from FNA (50), and there remains debate as to whether molecular testing substantially reduces the costs of care for indeterminate nodules given the high costs of Afirma and ThyroSeq v3 testing. One potential driver is clinical concern that the risk of malignancy for pathologically analyzed samples is underestimated (50). Although thyroid cancer incidence rates have started to plateau in the US after changes in ATA recommendations, disability-adjusted life years have not yet improved which may reflect suboptimal risk stratification and higher average healthcare expenditures relative to other countries (18).

The high costs of care in the United States imply a different risk calculus for assessing the risks and benefits of thyroid cancer care, especially since patients bear substantial out-of-pocket costs for diagnostic workup, surgical management, and surveillance. Importantly, thyroid cancer patients remain at risk for adverse financial outcomes and material and psychological hardship that could impair quality of life more than certain untreated forms of thyroid cancer, such as papillary thyroid microcarcinoma, that are unlikely to produce symptoms or metastasize. Furthermore, these risks do not appear to abate during remission, as patients continue to incur costs due to surveillance and experience burden associated with the fear of recurrence. Therefore, thyroid nodule management may need to be tailored in the United States to the unique healthcare system reimbursement structure and high patient-borne costs of care compared to countries with single-payer systems or alternative payment schemes in which patients pay a smaller proportion of household income towards medical care.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

Author NU reports income from Ironwood Medical Information Technologies, Quantified Ventures, and BrightEdge within the prior 3 years that is outside of the submitted work.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Tamhane S, Gharib H. Thyroid nodule update on diagnosis and management. *Clin Diabetes Endocrinol* (2016) 2:17. doi: 10.1186/s40842-016-0035-7

2. Li Y, Jin C, Li J, Tong M, Wang M, Huang J, et al. Prevalence of thyroid nodules in China: A health examination cohort-based study. *Front Endocrinol* (2021) 12:676144. doi: 10.3389/fendo.2021.676144

3. Tan GH, Gharib H. Thyroid incidentalomas: management approaches to nonpalpable nodules discovered incidentally on thyroid imaging. *Ann Intern Med* (1997) 126(3):226–31. doi: 10.7326/0003-4819-126-3-199702010-00009

4. Sosa JA, Hanna JW, Robinson KA, Lanman RB. Increases in thyroid nodule fine-needle aspirations, operations, and diagnoses of thyroid cancer in the united states. *Surgery* (2013) 154(6):1420–6; discussion 1426–7. doi: 10.1016/j.surg.2013.07.006

5. Kwong N, Medici M, Angell TE, Liu X, Marqusee E, Cibas ES, et al. The influence of patient age on thyroid nodule formation, multinodularity, and thyroid cancer risk. *J Clin Endocrinol Metab* (2015) 100(12):4434–40. doi: 10.1210/jc.2015-3100

6. Guth S, Theune U, Aberle J, Galach A, Bamberger CM. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. *Eur J Clin Invest.* (2009) 39(8):699–706. doi: 10.1111/j.1365-2362.2009.02162.x

7. Durante C, Grani G, Lamartina L, Filetti S, Mandel SJ, Cooper DS. The diagnosis and management of thyroid nodules: A review. *JAMA* (2018) 319(9):914–24. doi: 10.1001/jama.2018.0898

8. Filetti S, Durante C, Torlontano M. Nonsurgical approaches to the management of thyroid nodules. *Nat Clin Pract Endocrinol Metab* (2006) 2(7):384–94. doi: 10.1038/ncpendmet0215

9. Sajisevi M, Caulley L, Eskander A, Du YJ, Auh E, Karabachev A, et al. Evaluating the rising incidence of thyroid cancer and thyroid nodule detection modes: A multinational, multi-institutional analysis. *JAMA Otolaryngol Head Neck Surg* (2022) 14:811–8. doi: 10.1001/jamaoto.2022.1743

10. Dean DS, Gharib H. Epidemiology of thyroid nodules. *Best Pract Res Clin Endocrinol Metab* (2008) 22(6):901–11. doi: 10.1016/j.beem.2008.09.019

11. Alexander EK, Cibas ES. Diagnosis of thyroid nodules. *Lancet Diabetes Endocrinol* (2022) 10(7):533–9. doi: 10.1016/S2213-8587(22)00101-2

12. Alexander EK, Doherty GM, Barletta JA. Management of thyroid nodules. *Lancet Diabetes Endocrinol* (2022) 10(7):540–8. doi: 10.1016/S2213-8587(22)00139-5

13. Davies L, Welch HG. Current thyroid cancer trends in the united states. *JAMA Otolaryngol Head Neck Surg* (2014) 140(4):317–22. doi: 10.1001/jamaoto.2014.1

14. Powers AE, Marcadis AR, Lee M, Morris LGT, Marti JL. Changes in trends in thyroid cancer incidence in the united states, 1992 to 2016. *JAMA* (2019) 322(24):2440–1. doi: 10.1001/jama.2019.18528

15. Ahn HS, Welch HG. South korea's thyroid-cancer "Epidemic"–turning the tide. *N Engl J Med* (2015) 373(24):2389–90. doi: 10.1056/NEJMc1507622

16. Aschebrook-Kilfoy B, James B, Nagar S, Kaplan S, Seng V, Ahsan H, et al. Risk factors for decreased quality of life in thyroid cancer survivors: Initial findings from the north American thyroid cancer survivorship study. *Thyroid* (2015) 25(12):1313–21. doi: 10.1089/thy.2015.0098

17. James BC, Mitchell JM, Jeon HD, Vasilottos N, Grogan RH, Aschebrook-Kilfoy B. An update in international trends in incidence rates of thyroid cancer, 1973-2007. *Cancer Causes Control.* (2018) 29(4-5):465–73. doi: 10.1007/s10552-018-1023-2

18. Schuster-Bruce J, Jani C, Goodall R, Kim D, Hughes W, Salciccioli JD, et al. A comparison of the burden of thyroid cancer among the European union 15+ countries, 1990-2019: Estimates from the global burden of disease study. *JAMA Otolaryngol Head Neck Surg* (2022) 148(4):350–9. doi: 10.1001/jamaoto.2021.4549

19. Deng Y, Li H, Wang M, Li N, Tian T, Wu Y, et al. Global burden of thyroid cancer from 1990 to 2017. *JAMA Netw Open* (2020) 3(6):e208759. doi: 10.1001/jamanetworkopen.2020.8759

20. Kim J, Gosnell JE, Roman SA. Geographic influences in the global rise of thyroid cancer. *Nat Rev Endocrinol* (2020) 16(1):17–29. doi: 10.1038/s41574-019-0263-x

21. Wiltshire JJ, Drake TM, Uttley L, Balasubramanian SP. Systematic review of trends in the incidence rates of thyroid cancer. *Thyroid* (2016) 26(11):1541–52. doi: 10.1089/thy.2016.0100

22. Lim H, Devesa SS, Sosa JA, Check D, Kitahara CM. Trends in thyroid cancer incidence and mortality in the united states, 1974-2013. *JAMA* (2017) 317(13):1338–48. doi: 10.1001/jama.2017.2719

23. Lubitz CC, Kong CY, McMahon PM, Daniels GH, Chen Y, Economopoulos KP, et al. Annual financial impact of well-differentiated thyroid cancer care in the united states. *Cancer* (2014) 120(9):1345–52. doi: 10.1002/cncr.28562

24. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: The American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid* (2016) 26(1):1–133. doi: 10.1089/thy.2015.0020

25. Danese D, Sciacchitano S, Farsetti A, Andreoli M, Pontecorvi A. Diagnostic accuracy of conventional versus sonography-guided fine-needle aspiration biopsy of thyroid nodules. *Thyroid* (1998) 8(1):15–21. doi: 10.1089/thy.1998.8.15

26. Carmeci C, Jeffrey RB, McDougall IR, Nowels KW, Weigel RJ. Ultrasound-guided fine-needle aspiration biopsy of thyroid masses. *Thyroid* (1998) 8(4):283–9. doi: 10.1089/thy.1998.8.283

27. Wong CKH, Liu X, Lang BHH. Cost-effectiveness of fine-needle aspiration cytology (FNAC) and watchful observation for incidental thyroid nodules. *J Endocrinol Invest.* (2020) 43(11):1645–54. doi: 10.1007/s40618-020-01254-0

28. Cheung K, Roman SA, Wang TS, Walker HD, Sosa JA. Calcitonin measurement in the evaluation of thyroid nodules in the united states: a cost-effectiveness and decision analysis. *J Clin Endocrinol Metab* (2008) 93(6):2173–80. doi: 10.1210/jc.2007-2496

29. Yip L, Farris C, Kabaker AS, Hodak SP, Nikiforova MN, McCoy KL, et al. Cost impact of molecular testing for indeterminate thyroid nodule fine-needle aspiration biopsies. *J Clin Endocrinol Metab* (2012) 97(6):1905–12. doi: 10.1210/jc.2011-3048

30. Balentine CJ, Vanness DJ, Schneider DF. Cost-effectiveness of lobectomy versus genetic testing (Afirma®) for indeterminate thyroid nodules: Considering the costs of surveillance. *Surgery* (2018) 163(1):88–96. doi: 10.1016/j.surg.2017.10.004

31. Chen T, Gilfix BM, Rivera J, Sadeghi N, Richardson K, Hier MP, 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(9):1280–7. doi: 10.1089/thy.2019.0539

32. Nicholson KJ, Roberts MS, McCoy KL, Carty SE, Yip L. Molecular testing versus diagnostic lobectomy in Bethesda III/IV thyroid nodules: A cost-effectiveness analysis. *Thyroid* (2019) 29(9):1237–43. doi: 10.1089/thy.2018.0779

33. Alci E, Makay Ö. Impact of healthcare resources on management of indeterminate thyroid tumors. *Ann Thyroid* (2021) 6:3–3. doi: 10.21037/aot-20-44

34. Grisales J, Sanabria A. Utility of routine frozen section of thyroid nodules classified as follicular neoplasm. *Am J Clin Pathol* (2020) 153(2):210–20. doi: 10.1093/ajcp/aqz152

35. Bollig CA, Gilley D, Lesko D, Jorgensen JB, Galloway TL, Zitsch RP3rd, et al. Economic impact of frozen section for thyroid nodules with "Suspicious for malignancy" cytology. *Otolaryngol Head Neck Surg* (2018) 158(2):257–64. doi: 10.1177/0194599817740328

36. Davies L, Hendrickson CD, Hanson GS. Experience of US patients who self-identify as having an overdiagnosed thyroid cancer: A qualitative analysis. *JAMA Otolaryngol Head Neck Surg* (2017) 143(7):663–9. doi: 10.1001/jamaoto.2016.4749

37. Barrows CE, Belle JM, Fleishman A, Lubitz CC, James BC. Financial burden of thyroid cancer in the united states: An estimate of economic and psychological hardship among thyroid cancer survivors. *Surgery* (2020) 167(2):378–84. doi: 10.1016/j.surg.2019.09.010

38. Broekhuis JM, Li C, Chen HW, Chaves N, Duncan S, Lopez B, et al. Patient-reported financial burden in thyroid cancer. *J Surg Res* (2021) 266:160–7. doi: 10.1016/j.jss.2021.03.051

39. Mongelli MN, Giri S, Peipert BJ, Helenowski IB, Yount SE, Sturgeon C. Financial burden and quality of life among thyroid cancer survivors. *Surgery* (2020) 167(3):631–7. doi: 10.1016/j.surg.2019.11.014

40. Gupta A, Morrison ER, Fedorenko C, Ramsey S. *Leverage, default, and mortality: Evidence from cancer diagnoses* (2017). Available at: https://papers.ssrn.com/abstract=2583975.

41. Ramsey S, Blough D, Kirchhoff A, Kreizenbeck K, Fedorenko C, Snell K, et al. Washington State cancer patients found to be at greater risk for bankruptcy than people without a cancer diagnosis. *Health Aff .* (2013) 32(6):1143–52. doi: 10.1377/hlthaff.2012.1263

42. Ramsey SD, Bansal A, Fedorenko CR, Blough DK, Overstreet KA, Shankaran V, et al. Financial insolvency as a risk factor for early mortality among patients with cancer. *J Clin Oncol* (2016) 34(9):980–6. doi: 10.1200/JCO.2015.64.6620

43. Uppal N, Cunningham Nee Lubitz C, James B. The cost and financial burden of thyroid cancer on patients in the US: A review and directions for future research. *JAMA Otolaryngol Head Neck Surg* (2022) 148(6):568–75. doi: 10.1001/jamaoto.2022.0660

44. Roth EM, Lubitz CC, Swan JS, James BC. Patient-reported quality-of-Life outcome measures in the thyroid cancer population. *Thyroid* (2020) 30(10):1414–31. doi: 10.1089/thy.2020.0038

45. Wang LY, Roman BR, Migliacci JC, Palmer FL, Tuttle RM, Shaha AR, et al. Cost-effectiveness analysis of papillary thyroid cancer surveillance. *Cancer* (2015) 121 (23):4132–40. doi: 10.1002/cncr.29633

46. Wu JX, Beni CE, Zanocco KA, Sturgeon C, Yeh MW. Cost-effectiveness of long-term every three-year versus annual postoperative surveillance for low-risk papillary thyroid cancer. *Thyroid* (2015) 25(7):797–803. doi: 10.1089/thy.2014.0617

47. 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):333–9. doi: 10.1159/000339959

48. Middleton WD, Teefey SA, Reading CC, Langer JE, Beland MD, Szabunio MM, et al. Multiinstitutional analysis of thyroid nodule risk stratification using the American college of radiology thyroid imaging reporting and data system. *AJR Am J Roentgenol.* (2017) 208(6):1331–41. doi: 10.2214/AJR.16.17613

49. *Survival rates for thyroid cancer* (2022). Available at: https://www.cancer.org/cancer/thyroid-cancer/detection-diagnosis-staging/survival-rates.html.

50. Cibas ES, Ali SZ. The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid* (2017) 27(11):1341–6. doi: 10.1089/thy.2017.0500

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 27 January 2023
DOI 10.3389/fendo.2022.1080159

OPEN ACCESS

EDITED BY
Vasyl Vasko,
Uniformed Services University of the Health Sciences, United States

REVIEWED BY
Mehmet Haciyanli,
Izmir Katip Celebi University, Türkiye
Eleonora Lori,
Sapienza University of Rome, Italy

*CORRESPONDENCE
Vincenzo Triggiani
vincenzo.triggiani@uniba.it

SPECIALTY SECTION
This article was submitted to Thyroid Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 25 October 2022
ACCEPTED 23 December 2022
PUBLISHED 27 January 2023

CITATION
Triggiani V, Lisco G, Renzulli G, Frasoldati A, Guglielmi R, Garber J and Papini E (2023) The TNAPP web-based algorithm improves thyroid nodule management in clinical practice: A retrospective validation study. *Front. Endocrinol.* 13:1080159. doi: 10.3389/fendo.2022.1080159

The TNAPP web-based algorithm improves thyroid nodule management in clinical practice: A retrospective validation study

Vincenzo Triggiani [1*], Giuseppe Lisco [1], Giuseppina Renzulli [2], Andrea Frasoldati [3], Rinaldo Guglielmi [4], Jeffrey Garber [5] and Enrico Papini [4]

[1]Interdisciplinary Department of Medicine, Section of Internal Medicine, Geriatrics, Endocrinology and Rare Diseases, School of Medicine, University of Bari "Aldo Moro", Bari, Italy, [2]Department of Emergency and Organ Transplantation, Section of Pathological Anatomy, University of Bari "Aldo Moro", Bari, Italy, [3]Endocrinology and Metabolism Department, Arcispedale Santa Maria Nuova Istituto di Ricovero e Cura a Carattere Scientifico-Azienda Sanitaria Locale, Reggio Emilia, Italy, [4]Endocrinology and Metabolism Department, Regina Apostolorum Hospital, Rome, Italy, [5]Endocrine Division, Harvard Vanguard Medical Associates Harvard Medical School, Boston, MA, United States

Background: The detection of thyroid nodules has been increasing over time, resulting in an extensive use of fine-needle aspiration (FNA) and cytology. Tailored methods are required to improve the management of thyroid nodules, including algorithms and web-based tools.

Study aims: To assess the performance of the Thyroid Nodule App (TNAPP), a web-based, readily modifiable, interactive algorithmic tool, in improving the management of thyroid nodules.

Methods: One hundred twelve consecutive patients with 188 thyroid nodules who underwent FNA from January to December 2016 and thyroid surgery were retrospectively evaluated. Neck ultrasound images were collected from a thyroid nodule registry and re-examined to extract data to run TNAPP. Each nodule was evaluated for ultrasonographic risk and suitability for FNA. The sensitivity, specificity, positive and negative predictive values, and overall accuracy of TNAPP were calculated and compared to the diagnostic performance of the other two algorithms by the American Association of Clinical Endocrinology/American College of Endocrinology/Associazione Medici Endocrinologi (AACE/ACE/AME), which it was derived from the American College of Radiology Thyroid Imaging Reporting and Data System (ACR TI-RADS).

Results: TNAPP performed better in terms of sensitivity (>80%) and negative predictive value (68%) with an overall accuracy of 50.5%, which was similar to that found with the AACE/ACE/AME algorithm. TNAPP displayed a slightly better performance than AACE/ACE/AME and ACR TI-RADS algorithms in selectively discriminating unnecessary FNA for nodules with benign cytology (TIR 2 - Bethesda class II: TNAPP 32% vs. AACE/ACE/AME 31% vs. ACR TI-RADS 29%). The TNAPP reduced the number of missed diagnoses of thyroid nodules with suspicious and highly suspicious cytology (TIR 4 + TIR 5 - Bethesda classes V + VI: TNAPP 18% vs. AACE/ACE/AME 26% vs. ACR TI-RADS 20.5%). A total of 14 nodules

that would not have been aspirated were malignant, 13 of which were microcarcinomas (92.8%).

Discussion: The TNAPP algorithm is a reliable, easy-to-learn tool that can be readily employed to improve the selection of thyroid nodules requiring cytological characterization. The rate of malignant nodules missed because of inaccurate characterization at baseline by TNAPP was lower compared to the other two algorithms and, in almost all the cases, the tumors were microcarcinomas. TNAPP's use of size >20 mm as an independent determinant for considering or recommending FNA reduced its specificity.

Conclusion: TNAPP performs well compared to AACE/ACE/AME and ACR-TIRADS algorithms. Additional retrospective and, ultimately, prospective studies are needed to confirm and guide the development of future iterations that incorporate different risk stratification systems and targets for diagnosing malignancy while reducing unnecessary FNA procedures.

KEYWORDS

thyroid nodule, thyroid carcinoma, web-based algorithm, TNAPP, fine-needle aspiration (FNA), retrospective study

Background

The detection and prevalence rates of thyroid nodules have increased over the last six decades, paralleling the growing number of patients undergoing thyroid ultrasound (US) and other imaging modalities involving the neck (1). Accordingly, the number of newly diagnosed thyroid malignancies has increased, with most being microcarcinomas with favorable prognoses, even in the case of delayed treatment (2).

Most patients diagnosed with thyroid nodules after neck US are asymptomatic, and often thyroid nodules are discovered incidentally. Professional societies' guidelines recommend classifying the risk of malignancy of thyroid nodules before recommending fine-needle aspiration biopsy (FNA) to avoid unnecessary procedures (3–5). However, using risk stratification systems for thyroid nodules may be laborious and require expertise. Moreover, currently available recommendations do not incorporate clinical features or exclusion criteria. The indication to FNA should be based not only on US features but also on the integrated evaluation of family and personal history, local symptoms and signs, and laboratory data. Thus, a tailored and easily accessible methodology, including algorithms and web-based tools, is required to reduce costs and improve clinical utility in managing thyroid nodules.

The Thyroid Nodule App (TNAPP) is an integrated web-based algorithm that guides the management of thyroid nodules incorporating clinical factors, laboratory data, US characteristics, and cytology features. The TNAPP algorithm calculates in real-time the indication for FNA or follow-up and, sequentially, the risk of malignancy of thyroid nodules and the indication to surgery, surveillance, or discharge based on the American Association of Clinical Endocrinology/American College of Endocrinology/Associazione Medici Endocrinologi (AACE/ACE/AME) US risk classes (US 1, US 2, and US 3) (4), the American College of Radiology Thyroid Imaging Reporting and Data System (ACR TI-RADS) US categories (TR 1 to 5) (5) and the clinical and laboratory data of the individual patient.

In this original retrospective study, we assessed and compared the clinical accuracy of the TNAPP algorithm with two other frequently used algorithms (AACE/ACE/AME and ACR TI-RADS) in a cohort of unselected patients who had FNA of one or more thyroid nodules in 1 year and went to thyroid surgery, slightly before the publication and dissemination of updated guidelines and algorithms.

Methods

We retrospectively examined a cohort of patients with thyroid nodules who had been referred to the Outpatient Endocrinology Centre of the Azienda Ospedaliero – Universitaria Policlinico of Bari to perform thyroid FNA from January 1, 2016, to December 31, 2016. From a total of 473 patients with 852 nodules who had an FNA in that period, we selected those who undergone thyroid surgery (112 patients with 188 nodules), corresponding to 23.7% of the entire cohort. Detailed cytological (6) and histological reports of thyroid samples with TNM Classification (7) were available for all the nodules. Medical records included a detailed history and laboratory determinations (Thyroid-Stimulating Hormone, free thyroxine, anti-

Abbreviations: AACE/ACE/AME, American Association of Clinical Endocrinology/American College of Endocrinology/Associazione Medici Endocrinologi; ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System; FNA, Fine-Needle Aspiration; SIAPeC-IAP, Società Italiana di Anatomia Patologica e Citologia diagnostica – International Accademy of Pathology; TNAPP, Thyroid Nodule App; US, ultrasound.

thyroperoxidase antibody). Neck US examination and US-assisted FNA were performed by the same operator (V.T.). A set of neck US images was retrieved for each patient from the radiological records and was re-evaluated by two different operators (V.T. and G.L.) to categorize thyroid nodule features according to the current classification systems.

Nodule risk stratification was carried out considering clinical, laboratory, and US hallmarks for each patient. The following six US characteristics employed by TNAPP were based on nomenclature under development by the International Thyroid Nodule Ultrasound Working Group (ITNUWG): nodule composition, echogenicity, shape, margins, and echogenic foci with the addition of vascular patterns. According to the AACE/ACE/AME, thyroid nodules were classified as low risk (US 1), moderate risk (US 2), and high risk (US 3). According to the ACR TI-RADS, nodules were classified into five categories: TR 1 (benign), TR 2 (not suspicious), TR 3 (mildly suspicious), TR 4 (moderately suspicious), and TR 5 (highly suspicious).

First, we calculated risk categories for each thyroid nodule with US classifications and labeled those requiring FNA. According to the AACE/ACE/AME guidelines (4), thyroid biopsy was considered for nodules with a major diameter of at least 5-10 mm when suspicious US signs were present (US 3) or in those associated with pathologic cervical lymph nodes that were not clinically evident. Patients with a personal (none registered in 2016) or family history of thyroid cancer with thyroid nodules>5 mm were also considered suitable for FNA. Patients with thyroid nodules >10 mm with either US 2 or US 3 class of risk and low-risk nodules (i.e., US 1) >20 mm were also included among candidates for FNA. According to the ACR TI-RADS algorithm (5), thyroid biopsy was recommended in patients with TR 3 with a major diameter ≥25 mm, TR 4 with a major diameter ≥15 mm, TR 5 with a major diameter ≥10 mm. After that, nodule malignancy risk was assessed by the TNAPP electronic algorithmic tool, integrating the clinical and laboratory data with the US findings. The principal goal of the study was to evaluate TNAPP's performance as a tool for deciding whether to perform an FNA. Although comprehensive clinical data were available, the TNAPP did not change the decisions based on US data alone to perform FNAs. The "2017 European Union Thyroid Imaging Reporting and Data System Lexicon" (8) was used to categorize US features. The TNAPP is a web-based easy-to-apply tool, accessible for free at the website: https://aace-thyroid.deontics.com.

Surgical histological reports were used as the gold standard for the final diagnosis of thyroid nodules. Thyroid histology was considered the reference value for evaluating the diagnostic performance of the three algorithms as a whole or subdivided according to thyroid nodules' major diameters.

The level of agreement was also calculated, overall and according to thyroid nodules' major diameters, whether to perform FNA.

Results

The median age of patients was 55 years [10-86 yrs], and 21 of 112 were men (19%). Sixty-two of 188 nodules were palpable (33%), and 19 of them had hard consistency. Thyroid nodules were detected in a variety of ways and included cosmetic complaints (14.3%), neck enlargement (13.4%), incidental discovery after a carotid echo-color-doppler examination (11.6%), follow-up of diffuse thyroid diseases (10.7%), or compressive symptoms (8%).

The median TSH value was 1.89 mUI/L [0.3; 9]. Thirty-one patients (28%) had elevated titer of thyroperoxidase antibodies. Unstimulated serum values of calcitonin were available in 38 patients (34%). Among them, 36 had a normal value. The remaining two patients had an elevated unstimulated calcitonin level: a 49-year woman with mild elevation (14.5 pg/mL) and a 44-year woman with marked elevation (784 pg/mL) diagnosed with medullary thyroid cancer and excluded from the enrollment in the study. Twenty-nine patients (26%) were on levothyroxine therapy due to concomitant hypothyroidism; one was on methimazole because of hyperthyroidism. Three excluded patients had suspicious cervical lymph nodes and, as per protocol, underwent FNA irrespective of thyroid US images.

Thyroid surgery was recommended in case of nodules presenting with indeterminate, suspicious, and malignant cytological results (71 patients, 63%) and because of clinical signs or symptoms in large nodules or multinodular goiters (41 patients, 37%). The median diameter of the largest nodule diameter was 14 mm [4; 62]. The histological diameter was available in 133 thyroid nodules with a median of 11 mm after formalin fixation [3; 65]. Histological and ultrasonographic diameters were linearly related ($r = 0.8 \pm 0.04$; F 392.3; $p < .0001$), thus suggesting a high concordance between the two measures. The US characteristics of the nodules under evaluation are described in Table 1, while Table 2 summarizes the cytological findings of biopsied nodules with the corresponding final histology.

A high concordance rate was found between thyroid cytology and histological findings. A complete concordance rate (100%) was found between benign cytological (TIR 2 – Bethesda class II) and non-malignant histology (autoimmune thyroiditis, hyperplastic nodule, goiter, follicular adenoma). A complete concordance rate (100%) was also found between high-risk cytology (TIR 5 - Bethesda class VI) and malignant histology (follicular and papillary thyroid carcinoma). Indeterminate cytology was split into TIR 3A and TIR 3B according to the Società Italiana di Anatomia Patologica e Citologia diagnostica – International Academy of Pathology (SIAPeC-IAP) 2014 classification, corresponding to the classes III and IV, respectively, of the 2017 Bethesda system. The rate of malignant lesions among TIR 3A and TIR 3B (Bethesda classes III and IV) nodules were 29 and 36%, respectively (Figure 1).

The detailed explanation of malignant nodules among TIR 1 (Bethesda class I) cytology is included in Table 2 capitation. Abbreviations: SIAPeC = Società Italiana di Anatomia Patologica e Citologia (diagnostica); IAP = International Academy of Pathology.

Indication of thyroid surgery had been suggested based on cytological results in 105 of 188 thyroid nodules (55.8%). Cytological consistency in driving clinical decisions (as indicated by guidelines) was calculated, considering histological findings as the reference value. The sensitivity and specificity of cytology were 90.9% and 64.6%, respectively. Positive and negative predictive values were 66.7% and 90.1%, respectively. The overall accuracy of cytology in driving clinical decisions was 76.1% (Table 3).

According to the AACE/ACE/AME risk score, 26 thyroid nodules were classified as US 1 (13.8%), 88 US 2 (46.8%), and the remaining 74 (39.4%) US 3. A thyroid biopsy would have been recommended in

10.3389/fendo.2022.1080159

TABLE 1 Ultrasonographic characteristics of examined thyroid nodules (n = 188).

US variables	Prevalence of the leading characteristics of each US variable (n, %)				
Composition	Completely cystic (3; 1.6%)	Mixed cystic and solid (eccentric mural component) (11; 5.9%)	Solid (174; 92.5%)	-	-
Echogenicity	Hyperechoic (5; 2.6%)	Isoechoic (59; 31.5%)	Anechoic (4; 2.1%)	Hypoechoic or slightly hypoechoic (73; 38.8%)	Profoundly hypoechoic (48; 25%)
Shape	Oval or round (167; 88.8%)	"Taller than wide" (21; 11.2%)	–	–	–
Margins	Smooth or regular (134; 71.3%)	Irregular with protrusion into adjacent thyroid (15; 7.9%)	Spiculate or sharp angles (26; 13.8%)	Ill-defined (13; 6.9%)	–
Echogenic foci	Absent (126; 67%)	Difficult to characterize foci (17; 9%)	Intranodal macrocalcifications (10; 5.3%)	Microcalcifications (35; 18.6%)	Peripheral calcifications (4; 2.1%)
Vascular pattern	Peripheral or low vascularity (131; 69.7%)	Intranodular vascularity (57; 30.3%)	–	–	–

A complete description of ultrasonographic variables of examined thyroid nodules (left column, in bold) with a detailed characterization of the hallmarks of each US variable.
US, ultrasonographic.

146 nodules (77.7%). Based upon the ACR TI-RADS risk score, thyroid nodules were classified as follows: TR 1, 3 (1.6%); TR 2, 10 (5.3%); TR 3, 48 (25.5%); TR 4, 74 (39.4%); TR 5, 53 (28.2%). Ninety-two percent of US 1, 67% of US 2, and 34% of US 3 nodules had non-malignant histology (Figure 2A). Non-malignant lesions were in 100% of TR 1, 70% of TR 2, 88% of TR 3, 58% of TR 4, and 25% of TR 5 nodules (Figure 2B). A thyroid biopsy would have been recommended in 100 thyroid nodules (53.2%).

Lastly, according to the TNAPP outputs, thyroid biopsy was suggested (72) or recommended (72) in 144 nodules (76.6%). A concordance between the AACE/ACE/AME recommendations and TNAPP outputs was reached in 172 of 188 thyroid nodules (91.5%), while a lower agreement was found between the ACR TI-RADS recommendations and TNAPP outputs (144 of 188 thyroid nodules, 76.6%).

The concordance rate between ACR TI-RADS and TNAPP ranged between 73.7% and 79.7%, without any relevant differences concerning thyroid nodule diameters. Conversely, the concordance rate between the AACE/ACE/AME algorithm and TNAPP was slightly lower for thyroid nodules ≤10 mm (81.2%) compared to that observed in the case of larger thyroid nodules (Figure 3).

The level of agreement between the AACE/ACE/AME and TNAPP recommendations was similar irrespective of the final indication to perform or avoid FNA/follow-up. On the contrary, the concordance rate between the ACR TI-RADS and TNAPP was profoundly different concerning the final decision to perform rather than avoid FNA, with excellent agreement when the ACR TI-RADS algorithm suggested performing a thyroid biopsy (100%) and considerably lower concordance when the ACR TI-RADS algorithm did not recommend for FNA (50%). Data are reported in detail in Tables 4A, B.

The performance of the TNAPP algorithm was preliminarily calculated by using cytological results as the reference value. The sensitivity and specificity were 77.1% and 26.5%, respectively. The positive and negative predictive values were 60.4% and 44.2%, respectively, with an overall accuracy of 56.5%.

Furthermore, the overall performance of the three algorithms was calculated by using histological results as the reference value. More precisely, the overall accuracy of the AACE/ACE/AME, ACR TI-RADS, and TNAPP algorithms were 50.5%, 61.2%, and 50.5%, respectively. The AACE/ACE/AME and TNAPP algorithms had a better sensitivity (83.5 and 82.5%, respectively) compared to ACR TI-RADS (67.1%) and a lower specificity (26.6%, 27.5%, and 70.5%, respectively). All the algorithms perform better as negative predictors (Tables 5–7).

The performance of the three algorithms was slightly better for nodules ≤10 mm than those between 11 and 20 mm, while it dropped in thyroid nodules between 21 and 40 mm. The accuracy of the AACE/ACE/AME was slightly better for nodules >40 mm (55.5%), whereas the accuracy of both TNAPP and ACR TI-RADS was lower (44.4%). All data are reported in more detail in Supplementary Materials.

In light of the better negative than the positive predictive value of algorithms, we explored the distribution of algorithm-based recommendations according to cytological results (Figure 4). Data showed that TNAPP would have prevented 14 unnecessary FNA with TIR 2 - Bethesda class II cytology (31.8%) with a lower loss in FNA resulting from malignant cytology (TIR 4 - Bethesda class V, 6.8% and TIR 5 – Bethesda class VI, 11.4%). Most importantly, the concordance rate among the three algorithms to avoid thyroid biopsy of TIR 2 (Bethesda class II) nodules was 100% (13 nodules).

By dichotomizing histological results as malignant or benign, for each tool, we calculated the number of aspirates that would not have been performed on benign lesions and done on malignant ones. For the AACE/ACE/AME algorithm, 42 thyroid biopsies would not have been done, with 29 (69%) having non-malignant histology, while for 146 FNA that were recommended, 67 (45.9%) had malignant histology. Similar results were found for TNAPP. For the ACR TI-RADS, sixty-one (69.3%) of 88 that would not have been performed had benign histology, while 53% percent of nodules, for which the ACR TI-RADS recommended thyroid biopsy, were histologically malignant (Figure 5).

10.3389/fendo.2022.1080159

TABLE 2 Descriptive statistics of cytological findings and histological corresponding (n = 188).

SIAPeC-IAP 2014 - Bethesda System 2017	Ultrasonographic diameter (mm)	Histologic diameter (mm)	Benign	Malignant	Variants
TIR 1 - I (12, 6.4%)	15 (7.7)	11 (6)	Adenoma (2; 1.1%) Cystic (1; 0.5%) Goiter (4; 2.1%) Goiter and thyroiditis (1; 0.5%) Thyroiditis (1; 0.5%)	Papillary cancer (1; 0.5%)*	Multicentric follicular (1; 0.5%)
				Follicular cancer (2; 1.1%)**	Oncocytic (1; 0.5%) Multicentric (1; 0.5%)
TIR 2 - II (47; 25%)	18.6 (10.6)	19.1 (12.8)	Adenoma (12; 6.4%) Goiter (18; 9.6%) Goiter and thyroiditis (8; 4.3%) Thyroiditis (9; 4.8%)	–	–
TIR 3A - III (24; 12.8%)	15.8 (6.2)	12.8 (5.6)	Adenoma (6; 3.2%) Goiter (8; 4.3%) Goiter and thyroiditis (1; 0.5%) Thyroiditis (2; 1.1%)	Papillary cancer (4; 2.1%)	Classic intracystic (1; 0.5%) Follicular (2; 1.1%) Solid microfollicular (1; 0.5%)
				Follicular cancer (3; 1.6%)	Microfollicular (1; 0.5%) Multicentric (2; 1.1%)
TIR 3B - IV (44, 23.4%)	22.2 (12.7)	19.1 (13.1)	Adenoma (16; 8.5%) Goiter (9; 4.8%) Goiter and thyroiditis (1; 0.5%) Thyroiditis (3; 1.6%)	Follicular cancer (8; 4.3%)	Oncocytic (5; 2.7%) Microfollicular (2; 1.1%) Oxyphilic (1; 0.5%)
				Papillary cancer (7; 3.7%)	Oncocytic (3; 1.6%) Follicular (3; 1.6%) Solid (1; 0.5%)
TIR 4 - V (19, 10.1%)	13.4 (5.8)	12.4 (10.1)	Adenoma (4; 2.1%) Cystic (1; 0.5%) Goiter (1; 0.5%) Goiter and thyroiditis (1; 0.5%)	Papillary cancer (12; 6.4%)	Purely follicular (5; 2.7%) Follicular, Oncocytic (1; 0.5%) Follicular, Tall cells (3; 1.6%) Follicular, Solid (1; 0.5%) Purely tall cells (1; 0.5%) Microfollicular (1; 0.5%)
TIR 5 - VI (42, 22.3%)	11.6 (8.5)	10.6 (9.0)	–	Medullary cancer (1; 0.5%)	–
				Follicular cancer (1; 0.5%)	Oncocytic (1; 0.5%)
				Papillary cancer (40, 21.3%)	Purely classic (13; 6.9%) Classic cystic (1; 0.5%) Classic solid

(Continued)

10.3389/fendo.2022.1080159

TABLE 2 Continued

SIAPeC-IAP 2014 - Bethesda System 2017	Ultrasonographic diameter (mm)	Histologic diameter (mm)	Benign	Malignant	Variants
					(1; 0.5%) Classic, follicular, tall cells (1; 0.5%) Purely follicular (9; 4.8%) Follicular, Tall cells (3; 1.6%) Follicular, Oncocytic (1; 0.5%) Follicular, Oncocytic; Tall cells (1; 0.5%) Purely tall cells (6; 3.2%) Purely Trabecular (1; 0.5%) Purely Solid (1; 0.5%) Purely Oncocytic (1; 0.5%) Solid, Follicular (1; 0.5%)

* 18-year-old woman with two thyroid nodules with a major diameter of 4 mm; cytological findings were TIR 1 - Bethesda class I and TIR 5 - Bethesda class VI (indication for thyroid surgery), and histological diagnosis was multicentric papillary thyroid cancer.
** 63-year woman with three thyroid nodules underwent FNA with the following cytological findings: TIR 1 - Bethesda class I, TIR 3A - Bethesda class III, and TIR 3A - Bethesda class I. Thyroid surgery was suggested due to compressive symptoms.
** 53-year woman with three thyroid nodules underwent FNA with the following cytological findings: TIR 1 - Bethesda class I (11 mm), TIR 2 - Bethesda class II (7 mm), and TIR 5 - Bethesda class VI (6 mm). Thyroid surgery was suggested due to cytology results (TIR 5 - Bethesda class VI). Histological diagnosis: multifocal papillary cancer (cytology: TIR 5 - Bethesda class VI), oncocytic follicular cancer (cytology: TIR 1 - Bethesda class I), and oxyphilic adenoma (cytology: TIR 2 - Bethesda class II).
SIAPeC, Società Italiana di Anatomia Patologica e Citologia (diagnostica); IAP, International Academy of Pathology.

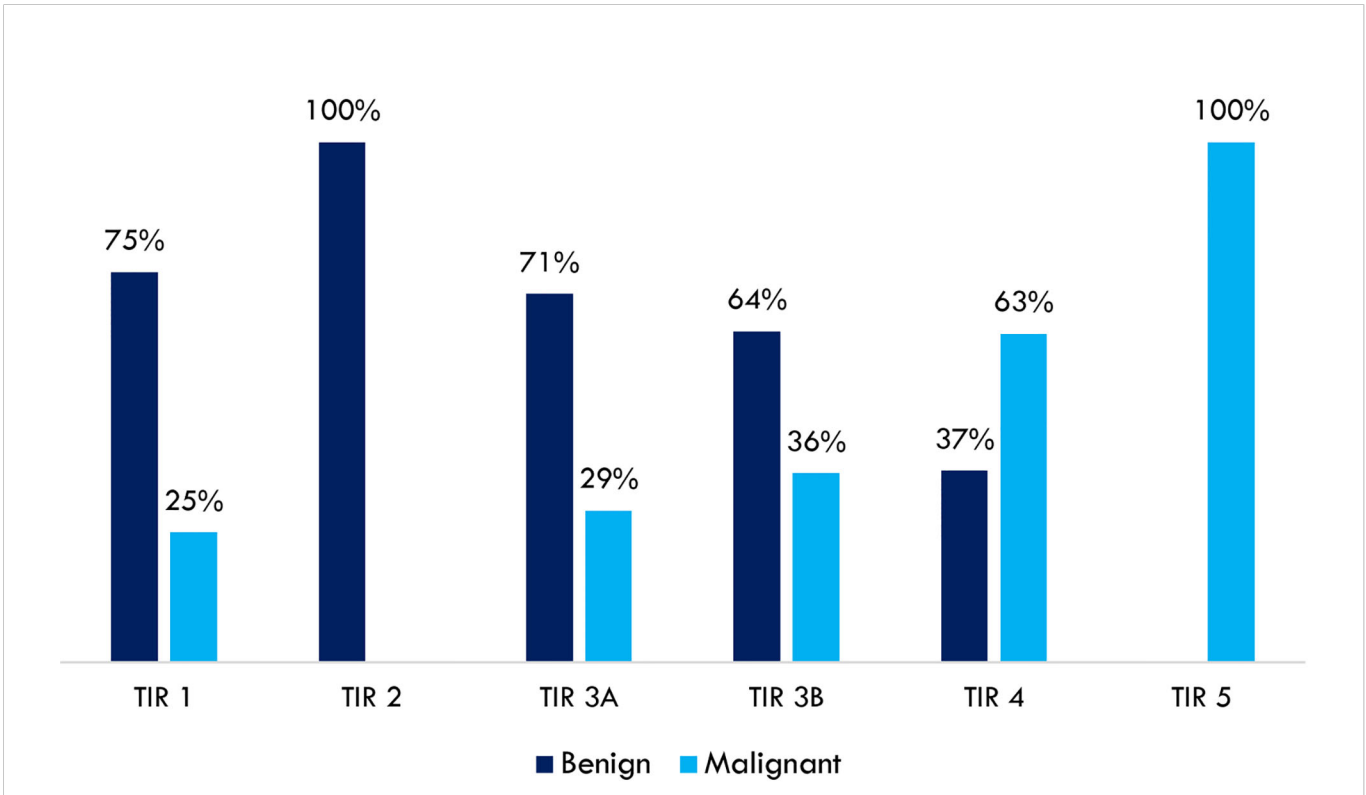


FIGURE 1
Concordance rates between thyroid cytology (SIAPeC-IAP 2014) and histological findings dichotomized as benign or malignant (n = 188).

10.3389/fendo.2022.1080159

TABLE 3 Assessment of cytological consistency in supporting clinical decisions according to guidelines (n = 176, TIR 1 – Bethesda class I excluded).

	Histology			
	Malignant	Benign	Total	
Cytology indicating thyroid surgery (Bethesda system 2017: IV, V, and VI; SIAPeC-IAP 2014 classes: TIR 3B, TIR 4, TIR 5)	70	35	105	Positive predictive value 66.7%
Cytology not suggesting thyroid surgery (Bethesda system 2017: II, and III; SIAPeC-IAP 2014 classes: TIR 2; TIR 3A)	7	64	71	Negative predictive value 90.1%
Total	77	99	176	
Sensitivity 90.9%, specificity 64.6%, overall accuracy 76.1%				

SIAPeC, Società Italiana di Anatomia Patologica e Citologia (diagnostica); IAP, International Academy of Pathology.

Among 12 (28.6%) thyroid nodules with malignant histology for which the AACE/ACE/AME algorithm did not recommend FNA, 11 were microcarcinomas (91.7%). Three of them were diagnosed with indeterminate cytology (one TIR 3A - Bethesda class III and two TIR 3B - Bethesda class IV), one had suspicious cytology (TIR 4 - Bethesda class V), and the remaining eight had positive cytology (TIR 5 - Bethesda class VI).

Twenty-six thyroid nodules with final malignant histology would have been excluded from FNA by the ACR TI-RADS (31.8%). Of them, 14 were microcarcinomas, and the remaining 12 had a major diameter greater than 10 mm. One had non-diagnostic cytology, eleven indeterminate cytology (five TIR 3A - Bethesda class III and six TIR 3B - Bethesda class IV), two suspicious cytology, and twelve positive cytology.

Among 14 (32%) thyroid nodules with malignant histology for which TNAPP suggested avoiding FNA, 13 were microcarcinomas. Among them, one was TIR 1 (Bethesda class I), none TIR 2 (Bethesda class II), three TIR 3A (Bethesda class III), four TIR 3B (Bethesda class IV), one TIR 4 (Bethesda class V), and five TIR 5 (Bethesda class VI). All these tumors were differentiated thyroid carcinomas (Supplementary Table 1).

TNAPP, as opposed to the other algorithms, provides subclassifications based on clinical characteristics in favor of performing FNA (Clinical 2) or against performing FNA (Clinical 1) and exclusion criteria for employing it as a decision tool. Thus, we analyzed the impact of clinical factors on the rate of diagnosing malignancy. Factors against performing FNA include suppressed or low TSH values in patients not taking levothyroxine, limited life expectancy or significant comorbidities making thyroid surgery high risk or low short-term priority, prior lobectomy with ipsilateral vocal cord paralysis, pregnancy, hyperfunctioning autonomous nodule, and at least one prior benign cytology on the same nodule. Factors favoring FNA are nodules with fixed or hard consistency, remote history of head and neck irradiation, compressive symptoms (dyspnea, dysphonia, dysphagia), documented US (nodule) or clinical (neck exam) of sudden enlargement, protocols (such as transplant surgery) that require ruling out cancer, and planned thyroid or parathyroid surgery. Exclusion criteria rendering TNAPP unsuitable for the evaluation of thyroid nodules include a prior history of thyroid cancer or hereditary/familial differentiated thyroid cancer in those with predisposing genetic syndromes (Gardner, Cowden, Adenomatous Familial Polyposis, Werner, Carney's complex), positron emission tomography positive nodules, elevated calcitonin, and suspicious or malignant regional adenopathy.

Among these 14 cases (Supplementary Table 1C), all had normal values of TSH. In three, there was a positive family history of thyroid cancer, and two nodules had hard composition. No other clinical characteristics were found clinical determinants were present. All in all, even after considering the clinical data, the final advice would have been the same as that suggested by the evaluation of US characteristics only: not performing the FNA and re-evaluating at 12 months. Clinical features did not affect guidance for those in whom US criteria alone determined that FNA was not recommended with or without a 12-month re-evaluation.

Discussion

The current overdiagnosis of thyroid nodules may lead to a parallel increased frequency of endocrinological consultations, number of performed FNAs, thyroid surgery procedures, and incidental diagnosis of indolent thyroid carcinomas. Overdiagnosis and overtreatment of thyroid nodules may unfavorably influence patients' quality of life, healthcare provider workload, and the financial status of healthcare systems. For these reasons, the management of the epidemic of thyroid nodules should be customized, providing cost- and risk-effective diagnostic procedures and treatments.

Electronic algorithms and artificial intelligence are currently proposed to improve the quality of care in several medical fields (9–13). The introduction of artificial intelligence is a novelty in thyroid nodule evaluation/management even if further implementation is necessary, including the integration with clinician expertise when composing a decision process, impact on workload and efficiency when using artificial intelligence, and assessment of the overall performance of these systems. In 2020, the TNAPP (14) was developed as an easy-to-use web-based algorithm that provides real-time and updated recommendations for managing thyroid nodules according to clinical factors, laboratory data, US characteristics, and cytology findings. The TNAPP algorithm has been preliminarily validated in a small and retrospective study on 95 thyroid nodules with histology-proven diagnoses (14) and a retrospective review of 59 thyroid nodules with Hurtle cytology (15), providing favorable results.

General consideration. The AACE/ACE/AME categories were associated with an increased risk of malignancy from US 1 to US 3 score. ACR TI-RADS performed very well with categories 1 (0% malignant) and 5 (75% malignant). However, in this study, a decrease

10.3389/fendo.2022.1080159

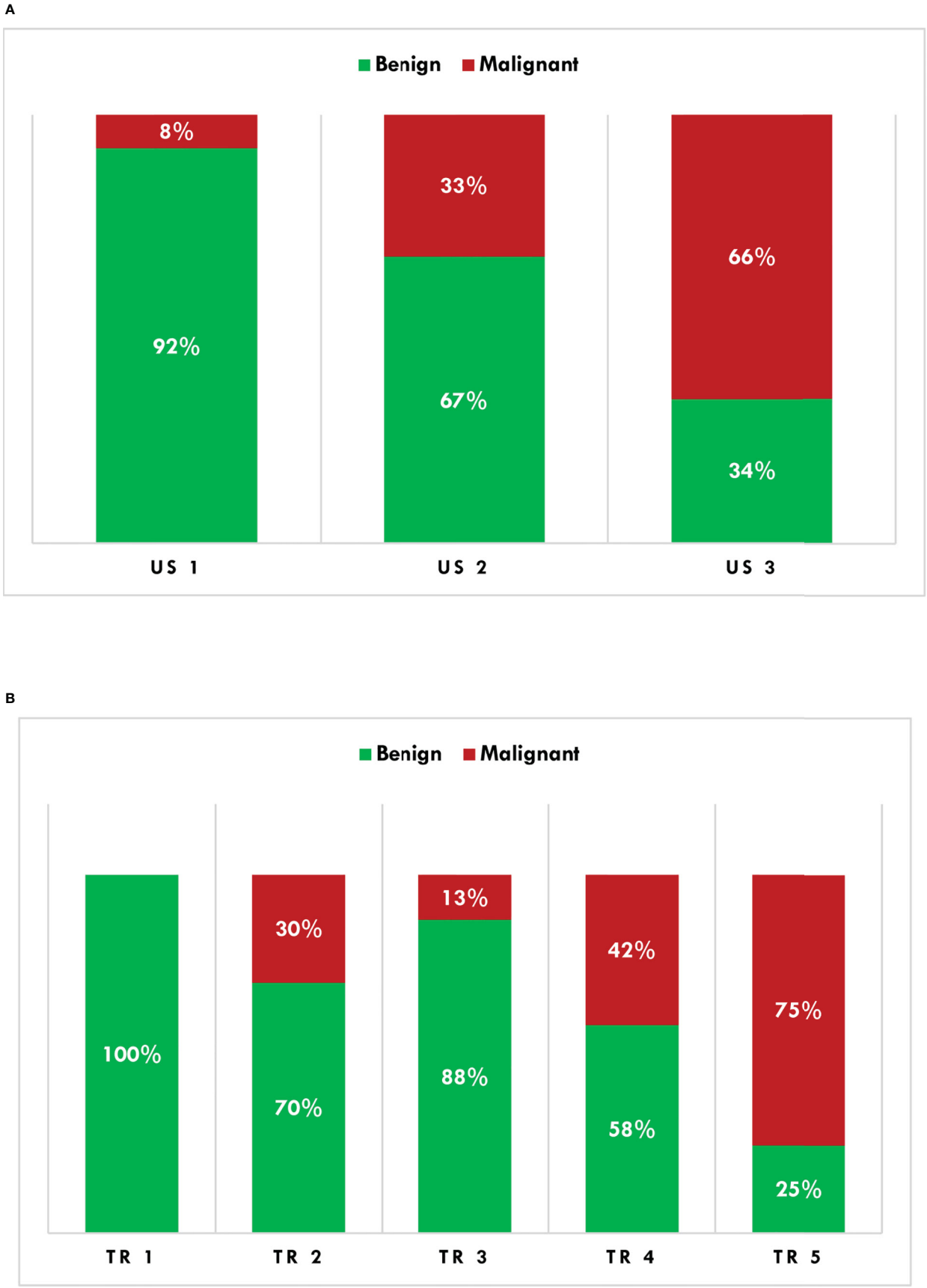


FIGURE 2
Histological characterization (binomial) of thyroid nodules according to the AACE/ACE/AME **(A)** and ACR TI-RADS **(B)** algorithms. AACE/ACE/AME, American Association of Clinical Endocrinology/American College of Endocrinology/Associazione Medici Endocrinologi; ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System.

10.3389/fendo.2022.1080159

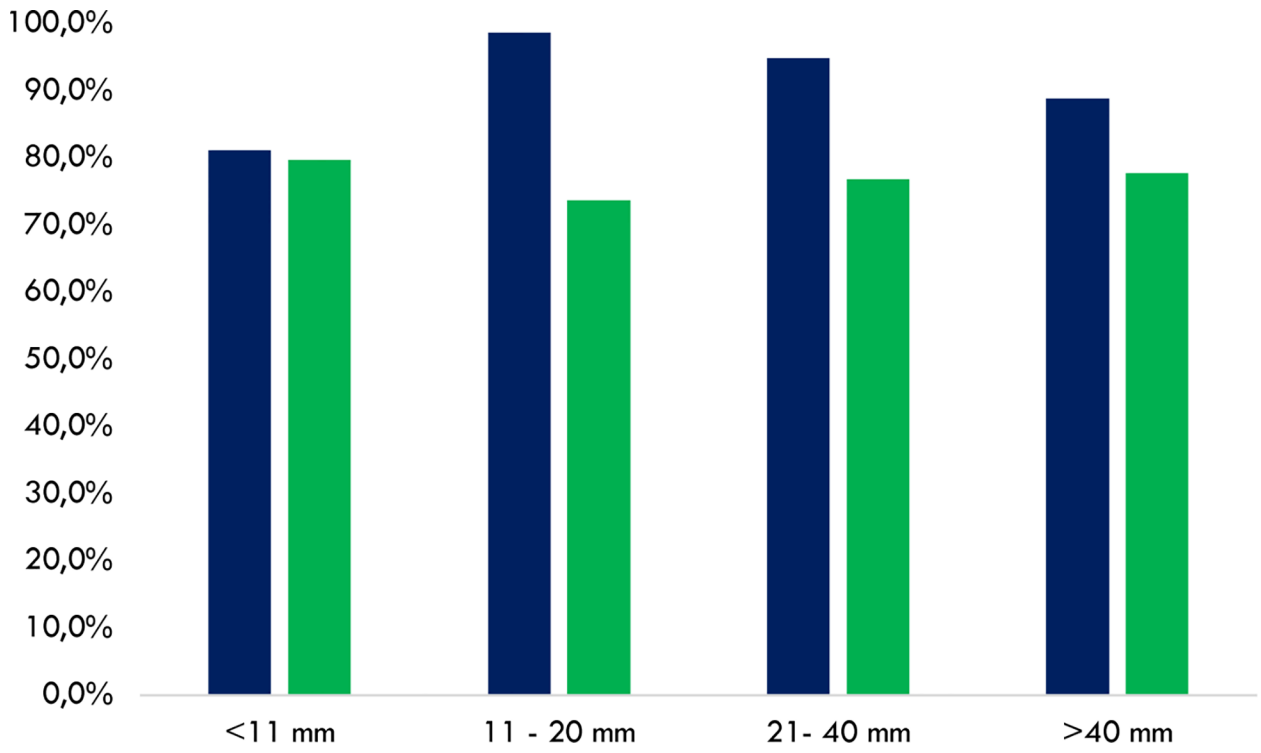


FIGURE 3
Cumulative concordance rates to perform or avoid FNA between the AACE/ACE/AME and TNAPP (blue) and the ACR TI-RADS and TNAPP (green). Data are illustrated by comparing rates among thyroid nodule diameters with different diameters (n = 188). AACE/ACE/AME, American Association of Clinical Endocrinology/American College of Endocrinology/Associazione Medici Endocrinologi. ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System.

in the rate of malignant nodules was found between the categories TR 2 and TR 3, suggesting that the five strata are not continuously discriminatory and could be merged into a single intermediate class. Of note, a discrepancy between the AACE/ACE/AME and ACR TI-RADS was found in 26 nodules, 2 of them with malignant histology. More precisely, the discrepancy concerned nodules classified as US 1 according to the AACE/ACE/AME and TR 3 with the ACR TI-RADS. Nodules with US 1 pattern are at low risk, do not require FNA, and include purely cystic, predominantly cystic with reverberating artifacts not associated with suspicious US signs, and solid spongiform isoechoic nodules. On the other hand, nodules with TR 3 pattern are mildly suspicious, may require FNA in case of major diameter ≥25 mm, and comprise nodules with the ACR TI-RADS score of 3 points [e.g., solid (2 points), isoechoic (1 point) nodules; or mixed cystic and solid (1 point), with the solid component being isoechoic (1 point) and echoic foci attributable to microcalcifications

TABLE 4 Concordance rates between the AACE/ACE/AME and TNAPP (A) and ACR TI-RADS and TNAPP (B) to recommend FNA or follow-up (n = 188).

A	
AACE/ACE/AME recommendation (n, %)	**Concordance rate between AACE/ACE/AME and TNAPP**
No FNA/follow-up (42, 22,3%)	83.3%
Perform FNA (146, 77.7%)	93.8%

B	
ACR TI-RADS recommendation (n, %)	**Concordance rate between ACR TI-RADS and TNAPP**
No FNA/follow-up (88, 46.8%)	50%
Perform FNA (100, 53.2%)	100%

AACE/ACE/AME, American Association of Clinical Endocrinology/American College of Endocrinology/Associazione Medici Endocrinologi; ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System; TNAPP, Thyroid Nodule App; FNA, Fine-Needle Aspiration.

10.3389/fendo.2022.1080159

TABLE 5 Assessment of the overall AACE/ACE/AME performance.

AACE/ACE/AME recommendation	Malignant	Benign	Total	
Perform FNA	66	80	146	Positive predictive value 45.2%
No FNA/follow-up	13	29	42	Positive predictive value 45.2%
Total	79	109	188	
Sensitivity 83.5%, specificity 26.6%, overall accuracy 50.5%				

AACE/ACE/AME, American Association of Clinical Endocrinology/American College of Endocrinology/Associazione Medici Endocrinologi.

(1 point). Thus, US characteristics of thyroid nodules classified as US 1 and TR 3 are dissimilar, and it is not expected to be matched using the same US features, resulting in different risk stratifications, namely, low in the former and moderate in the latter. Indeed, the AACE/ACE/AME US 1 pattern could be compatible with a TR 1 or TR 2 for the ACR TI-RADS; conversely, the TR 3 pattern could be consistent with a US 2. After ruling out possible mistakes in the data input or output reading, we confirmed the discordant results, suggesting that the criteria for defining the US 1 and US 2 pattern of thyroid nodules should be updated in the TNAPP algorithm.

Indication to FNA. In this retrospective study, TNAPP performed well when compared to the AACE/ACE/AME and ACR TI-RADS US risk stratification systems. The level of agreement between TNAPP recommendations and the AACE/ACE/AME algorithm was more significant than that between TNAPP and ACR TI-RADS. While the level of agreement between TNAPP and ACR TI-RADS was similar irrespective of thyroid nodule diameter, the concordance rate between the AACE/ACE/AME algorithm and TNAPP was slightly lower for thyroid nodule diameters ≤10 mm. The level of agreement on the overall indication to perform or avoid FNA was high between the TNAPP and AACE/ACE/AME algorithms. Conversely, the agreement between the TNAPP and ACR TI-RADS algorithms was high when both favored FNA but significantly lower when FNA was not recommended, leading to different guidance about which nodules require FNA.

Malignancy risk. Only 14 thyroid nodules which would have been excluded from FNA according to the TNAPP algorithm, resulted in malignant histology with three follicular and eleven papillary carcinomas. More precisely, thirteen of 14 (92.8%) were microcarcinomas with a diameter of 4 to 10 mm. Although the TNAPP algorithm failed to identify these malignant nodules, risks would have been mitigated by an overall favorable prognosis of these lesions. Similar results were provided by the AACE/ACE/AME algorithm (12 missed diagnoses with 11 microcarcinomas). On the other hand, 26 malignant thyroid nodules would have been excluded from FNA by the ACR TI-RADS algorithm. Of them, 12 (46%) carcinomas would have a major diameter of more than 10 mm, leading to possible concerns in the long-term management of these nodules due to misdiagnosis. Therefore, the TNAPP provided similar results as observed with the AACE/ACE/AME algorithm by reducing the magnitude of loss in thyroid carcinomas while screening the nodules for potential features of malignancy. Thus, the TNAPP missed fewer thyroid carcinomas than ACR TI-RADS that were not microcarcinomas. The main explanation for missing diagnoses was related to the small size (major diameter) of those nodules, as algorithms usually exclude from FNA nodules <5 mm and most nodules of 5-10 mm devoid of clinical or ultrasonographic signs of suspicion. According to the TNAPP, clinical conditions did not change the overall clinical guidance based on the US alone.

Identification of non-malignant nodules. The number of suggested or recommended FNA appeared particularly elevated when assessing the risk stratification of thyroid nodules with TNAPP and AACE/ACE/AME algorithms; thus, significantly lower discrimination of benign nodules could be expected. To improve the discriminative performance of these algorithms, the weight of each leading determinant in the overall risk of malignancy for thyroid nodules could be revised to reduce the chance of unnecessary procedures. This is the case with the thyroid nodule dimension. In fact, despite other relevant US characteristics, such as nodule composition, shape, echogenicity, margins, and echogenic foci, the concomitant evidence of the major nodular diameter of more than 20 mm significantly affects the TNAPP decision in favor of FNA. This matter may considerably increase the number of large, but not necessarily suspicious, thyroid nodules undergoing FNA procedures when using the TNAPP algorithm (false positive results). Thus, future iterations of TNAPP that employ other characteristics for thyroid nodules with a diameter >20 mm with otherwise favorable aspects, such as nodular enlargement over time (e.g., <20% between two consecutive neck US), could be used to determine recommendations for FNA.

TABLE 6 Assessment of the overall ACR TI-RADS performance.

ACR TI-RADS recommendation	Malignant	Benign	Total	
Perform FNA	53	47	100	Positive predictive value 53%
No FNA/follow-up	26	62	88	Negative predictive value 70.5%
Total	79	109	188	
Sensitivity 67.1%, specificity 70.5%, overall accuracy 61.2%				

ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System.

10.3389/fendo.2022.1080159

TABLE 7 Assessment of the overall TNAPP performance.

TNAPP recommendation	Malignant	Benign	Total	
Perform FNA	65	79	144	Positive predictive value 45.1%
No FNA/follow-up	14	30	44	Negative predictive value 68.2%
Total	79	109	188	
Sensitivity 82.3%, specificity 27.5%, overall accuracy 50.5%				

TNAPP, Thyroid Nodule App.

TNAPP is a web-accessible, easy-to-use algorithmic tool based on a narrative clinical practice guideline that incorporates clinical and thyroid nodule ultrasound findings to determine the risk for malignancy, guide whether to perform FNA, parameters for evaluating and following nodules when an FNA is not required, or a diagnosis of malignancy has not been made. Employing TNAPP could enhance the dissemination and implementation in clinical practice of thyroid nodule guidelines, particularly in settings where nodule classification is not routinely carried out. Since TNAPP can readily and rapidly be revised, updated guidance for patients with thyroid nodules can be provided continually as opposed to several years that it presently takes to update narrative clinical practice guidelines.

Although the TNAPP provided less overall accuracy than the ACR TI-RADS, the higher sensitivity compared to the specificity and a more significant negative than the positive predictive value of TNAPP resulted in more thyroid carcinomas, most of which are microcarcinomas, being diagnosed. While the ACR TI-RADS algorithm guidance would reduce the number of FNA procedures, more cases of thyroid carcinoma, with around half having a major diameter exceeding 10 mm, are missed.

Though not part of our study, TNAPP could quickly be regularly revised to offer guidance about the extent of surgery, non-surgical management of thyroid nodules, as well as the duration and type of follow-up. Doing so would keep clinicians informed of updated recommendations for the evaluation and management of thyroid nodules.

This study has some limitations and strengths. Analyses were carried out only in patients who had thyroid surgery, representing only a minority of the cases seen with thyroid nodules. However, definitive histologic diagnoses were used to test the algorithms' accuracy. Another limitation was in the nature of the study (retrospective, single-center, and single-operator for both FNA and pathology), which eliminates heterogeneity but may limit its generalizability to other settings and centers.

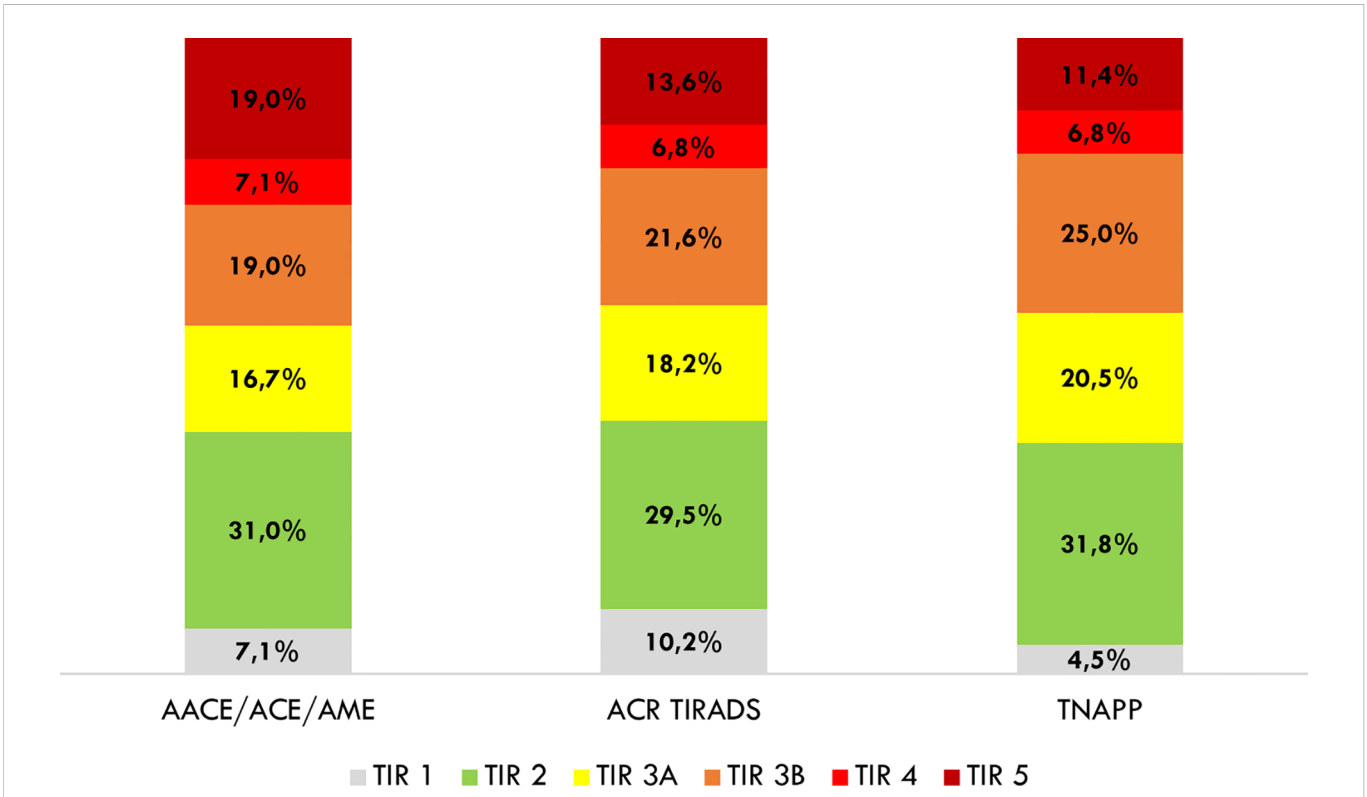


FIGURE 4
Distribution of avoidable thyroid biopsies after a retrospective analysis of cytological findings. AACE/ACE/AME, American Association of Clinical Endocrinology/American College of Endocrinology/Associazione Medici Endocrinologi; ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System; TNAPP, Thyroid Nodule App.

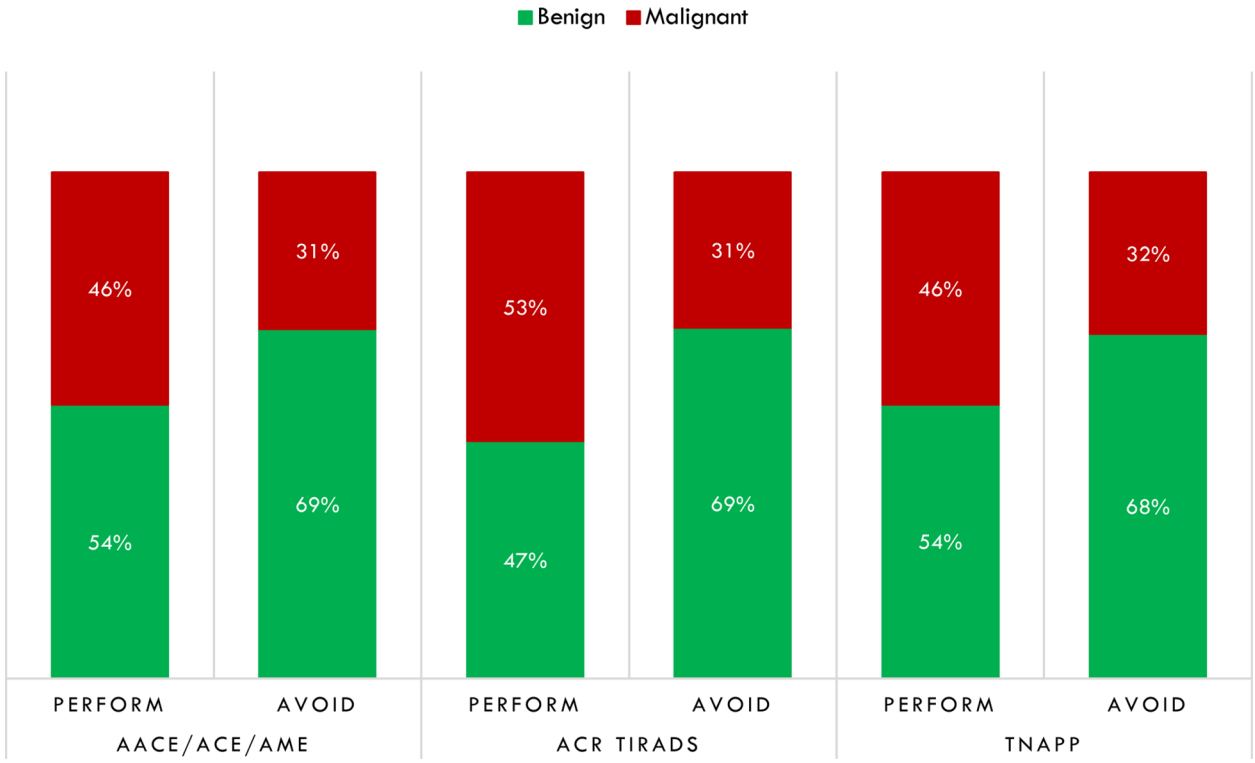


FIGURE 5
Distribution of thyroid biopsies that would have been performed or avoided regarding thyroid histology after dichotomizing histological results in non-malignant (benign) and malignant. AACE/ACE/AME, American Association of Clinical Endocrinology/American College of Endocrinology/Associazione Medici Endocrinologi; ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System; TNAPP, Thyroid Nodule App.

Conclusion

Medical expertise, patient preference, and organization of healthcare facilities to provide adequate diagnosis, treatment, and follow-up are the leading determinants of variation in the medical management of chronic diseases. Easy-to-use and inexpensive tools are needed to improve the quality of care by standardizing and implementing cost-effective clinical decisions for conditions with similar characteristics across different patient populations and clinical settings.

The role of algorithms has been investigated in this retrospective study, suggesting that TNAPP could improve the management of thyroid nodules by facilitating and thereby increasing the implementation of guidelines and recommendations before performing FNA. According to our retrospective results, extensive use of the TNAPP algorithm is expected to reduce the number of thyroid nodules requiring FNA with minimal impact resulting from missing or delaying the diagnosis of well-differentiated thyroid carcinomas, most of which are microcarcinomas, with favorable prognoses.

Our relatively small study indicates that TNAPP's performance could improve if low-risk US characteristics would override recommendations to perform FNA on all thyroid nodules with a major diameter larger than 20 mm. The growth rate, despite limitations of operator performance, machine variations, and establishing a standardized time frame between studies, could be used as an additional determinant. The contrast-enhanced thyroid US may provide more detailed information about parenchymal vascularization. The method could improve the characterization of thyroid nodules and lymph nodes and provide additional information to include in currently available algorithms (16, 17).

Improving and evolving technology that enables future TNAPP web-based versions to store and compare static and video images and artificial intelligence (14) to analyze images hold promise for the future. In the interim, prospective studies of TNAPP are needed to improve its performance and enhance its impact on managing thyroid nodules.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

VT and GL conceived the study. VT provided clinical expertise in performing thyroid ultrasound and ultrasound-guided fine-needle aspiration. GR provided technical expertise in thyroid cytology and pathology. VT and GL re-examined the registries and collected proper data to perform analyses. GL provided formal analyses. GL and VT drafted the manuscript. JG, VT, EP, AF, RG, and GR, read the manuscript and provided criticism and feedback. JG, VT, and EP provided supervision. All authors read the text and approved the final version of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2022.1080159/full#supplementary-material

References

1. Alexander EK, Cibas ES. Diagnosis of thyroid nodules. *Lancet Diabetes Endocrinol* (2022) 10(7):533–9. doi: 10.1016/S2213-8587(22)00101-2

2. Alexander EK, Doherty GM, Barletta JA. Management of thyroid nodules. *Lancet Diabetes Endocrinol* (2022) 10(7):540–8. doi: 10.1016/S2213-8587(22)00139-5

3. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: The American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid* (2016) 26(1):1–133. doi: 10.1089/thy.2015.0020

4. Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedüs L, et al. American Association of clinical endocrinologists, American college of endocrinology, and associazione Medici endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules–2016 update. *Endocr Pract* (2016) 22 (5):622–39. doi: 10.4158/EP161208.GL

5. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR thyroid imaging, reporting and data system (TI-RADS): White paper of the ACR TI-RADS committee. *J Am Coll Radiol* (2017) 14(5):587–95. doi: 10.1016/j.jacr.2017.01.046

6. Cibas ES, Ali SZ. The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid* (2017) 27(11):1341–6. doi: 10.1089/thy.2017.0500

7. Tuttle RM, Haugen B, Perrier ND. Updated American joint committee on Cancer/ Tumor-Node-Metastasis staging system for differentiated and anaplastic thyroid cancer (Eighth edition): What changed and why? *Thyroid* (2017) 27(6):751–6. doi: 10.1089/ thy.2017.0102

8. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European Thyroid association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: The EU-TIRADS. *Eur Thyroid J* (2017) 6(5):225–37. doi: 10.1159/ 000478927

9. Ellahham S. Artificial intelligence: The future for diabetes care. *Am J Med* (2020) 133 (8):895–900. doi: 10.1016/j.amjmed.2020.03.033

10. Lopez-Jimenez F, Attia Z, Arruda-Olson AM, Carter R, Chareonthaitawee P, Jouni H, et al. Artificial intelligence in cardiology: Present and future. *Mayo Clin Proc* (2020) 95 (5):1015–39. doi: 10.1016/j.mayocp.2020.01.038

11. Mekov E, Miravitlles M, Petkov R. Artificial intelligence and machine learning in respiratory medicine. *Expert Rev Respir Med* (2020) 14(6):559–64. doi: 10.1080/ 17476348.2020.1743181

12. Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. *Gastrointest Endosc* (2020) 92(4):807–12. doi: 10.1016/j.gie.2020.06.040

13. Sorrenti S, Dolcetti V, Radzina M, Bellini MI, Frezza F, Munir K, et al. Artificial intelligence for thyroid nodule characterization: Where are we standing? *Cancers (Basel)* (2022) 14(14):3357. doi: 10.3390/cancers14143357

14. Garber JR, Papini E, Frasoldati A, Lupo MA, Harrell RM, Parangi S, et al. American Association of clinical endocrinology and associazione Medici endocrinologi thyroid nodule algorithmic tool. *Endocr Pract* (2021) 27(7):649–60. doi: 10.1016/j.eprac.2021.04.007

15. Spagnuolo GM, Tierney HT, Laver NMV, Eldeiry LS. A retrospective study of clinicopathologic outcomes of nodules with hürthle cell cytology and the thyroid nodule app (TNAPP) ultrasound recommendations. *Endocr Pract* (2022) 28(6):593–8. doi: 10.1016/j.eprac.2022.03.011

16. Sorrenti S, Dolcetti V, Fresilli D, Del Gaudio G, Pacini P, Huang P, et al. The role of CEUS in the evaluation of thyroid cancer: From diagnosis to local staging. *J Clin Med* (2021) 10(19):4559. doi: 10.3390/jcm10194559

17. Fresilli D, David E, Pacini P, Del Gaudio G, Dolcetti V, Lucarelli GT, et al. Thyroid nodule characterization: How to assess the malignancy risk. update of the literature. *Diagn (Basel)* (2021) 11(8):1374. doi: 10.3390/diagnostics11081374

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 01 February 2023
DOI 10.3389/fendo.2023.1080908

OPEN ACCESS

EDITED BY
Andrea Frasoldati,
Endocrine Unit ASMN, Italy

REVIEWED BY
Mehmet Taner Ünlü,
Şişli Hamidiye Etfal Education and Research Hospital, Türkiye
Bo Zhang,
China-Japan Friendship Hospital, China

*CORRESPONDENCE
Wengang Liu
liuwengang312@hotmail.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Thyroid Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 26 October 2022
ACCEPTED 10 January 2023
PUBLISHED 01 February 2023

CITATION
Zhou P, Chen F, Zhou P, Xu L, Wang L, Wang Z, Yu Y, Liu X, Wang B, Yan W, Zhou H, Tao Y and Liu W (2023) The use of modified TI-RADS using contrast-enhanced ultrasound features for classification purposes in the differential diagnosis of benign and malignant thyroid nodules: A prospective and multi-center study. *Front. Endocrinol.* 14:1080908. doi: 10.3389/fendo.2023.1080908

The use of modified TI-RADS using contrast-enhanced ultrasound features for classification purposes in the differential diagnosis of benign and malignant thyroid nodules: A prospective and multi-center study

Ping Zhou[1†], Feng Chen[2†], Peng Zhou[3], Lifeng Xu[3], Lei Wang[4], Zhiyuan Wang[5], Yi Yu[6], Xueling Liu[7], Bin Wang[8], Wei Yan[9], Heng Zhou[9], Yichao Tao[10] and Wengang Liu[1*]

[1]Department of Ultrasound, The Third Xiangya Hospital, Central South University, Changsha, Hunan, China, [2]Department of Ultrasound, Yiyang Central Hospital of Hunan University of Chinese Medicine, Yiyang, Hunan, China, [3]Department of Ultrasound, Shenzhen Second People's Hospital, Shenzhen, Guangdong, China, [4]Department of Ultrasound, Huang Shi Central Hospital, Huang Shi, Hubei, China, [5]Department of Ultrasound, Hunan Cancer Hospital, Changsha, Hunan, China, [6]Department of Ultrasound, The People's Hospital of Liuyang, Changsha, Hunan, China, [7]Department of Ultrasound, The First Affiliated of Guangxi University of Chinese Medicine, Nanning, Guangxi, China, [8]Department of Ultrasound, Yueyang Central Hospital, Yueyang, Hunan, China, [9]Department of Ultrasound, Hubei Provincial Hospital of Traditional Chinese Medicine, Wuhan, Hubei, China, [10]Department of Ultrasound, Xiaogan Central Hospital, Xiaogan, Hubei, China

Objectives: To evaluate the diagnostic efficacy of a modified thyroid imaging reporting and data system (TI-RADS) in combination with contrast-enhanced ultrasound (CEUS) for differentiating between benign and malignant thyroid nodules and to assess inter-observer concordance between different observers.

Methods: This study included 3353 patients who underwent thyroid ultrasound (US) and CEUS in ten multi-centers between September 2018 and March 2020. Based on a modified TI-RADS classification using the CEUS enhancement pattern of thyroid lesions, ten radiologists analyzed all US and CEUS examinations independently and assigned a TI-RADS category to each thyroid nodule. Pathology was the reference standard for determining the diagnostic performance (accuracy (ACC), sensitivity (SEN), specificity (SPN), positive predictive value (PPV), and negative predictive value (NPV)) of the modified TI-RADS for predicting malignant thyroid nodules. The risk of malignancy was stratified for each TI-RADS category-based on the total number of benign and malignant lesions in that category. ROC curve was used to determine the cut-off value and the area under the curve (AUC). Cohen's Kappa statistic was applied to assess the inter-observer agreement of each sonological feature and TI-RADS category for thyroid nodules.

10.3389/fendo.2023.1080908

Results: The calculated malignancy risk in the modified TI-RADS categories 5, 4b, 4a, 3 and 2 nodules was 95.4%, 86.0%, 12.0%, 4.1% and 0%, respectively. The malignancy risk for the five categories was in agreement with the suggested malignancy risk. The ROC curve showed that the AUC under the ROC curve was 0.936, and the cutoff value of the modified TI-RADS classification was >TI-RADS 4a, whose SEN, ACC, PPV, NPV and SPN were 93.6%, 91.9%, 90.4%, 93.7% and 88.5% respectively. The Kappa value for taller than wide, microcalcification, marked hypoechoic, solid composition, irregular margins and enhancement pattern of CEUS was 0.94, 0.93, 0.75, 0.89, 0.86 and 0.81, respectively. There was also good agreement between the observers with regards to the modified TI-RADS classification, the Kappa value was 0.80.

Conclusions: The actual risk of malignancy according to the modified TI-RADS concurred with the suggested risk of malignancy. Inter-observer agreement for the modified TI-RADS category was good, thus suggesting that this classification was very suitable for clinical application.

KEYWORDS

thyroid, thyroid imaging report and data system, contrast-enhanced ultrasound, prospective, multi-center

Introduction

With the development and wide application of high resolution ultrasound, the detection rate of thyroid nodules has increased significantly (1–3). Although there is a high prevalence of thyroid nodules, only 1.6% to 12% of these are malignant (4, 5). According to the bethesda classification system, only 3%–7% of the thyroid nodules undergoing fine-needle aspiration (FNA) have clearly malignant features, at least 60%–70% of thyroid nodules are proven to be benign *via* pathological analysis. The pathological type of thyroid nodules directly affects the treatment and prognosis of patients. Therefore, the accurate judgment of benign and malignant thyroid nodules is of important clinical significance.

Ultrasound (US) is a simple and reproducible non-invasive method and remains the modality of choice for patients with thyroid nodules. US can distinguish benign and malignant tumors by specific ultrasound imaging characteristics (6, 7). Usually, the suspicious signs of malignant nodules on US include a solid composition, a taller shape rather than a wider shape, an irregular margin, micro-calcification, and marked hypo-echogenicity (8, 9). However, the grey scale and Doppler US features of benign and malignant nodules overlap. Furthermore, a single ultrasound sign cannot reliably predict benign and malignant thyroid nodules (10). Therefore, prediction models have been developed for malignancy that combine multiple US features to improve the accuracy of diagnosing benign and malignant thyroid nodules. In 2009, Horvath et al. were the first to classify thyroid nodules based on the principles that have been used in the breast imaging reporting and data system of the American College of Radiology using ten malignant-related ultrasound features, and proposed the first thyroid imaging reporting and data system (TI-RADS) classification system (11). As a quantitative system for the risk stratification of malignant tumors in thyroid nodules, however, this sonographic model is not applicable to all thyroid nodules and is difficult to apply. In the same year, Park et al. proposed a multi-factor logistic regression analysis equation to predict the malignant probability of thyroid nodules based on 12 types of sonographic features (12). However, this prediction equation is more complicated and difficult to apply. To overcome these limitations, Kwak et al. used several suspicious sonographic features and calculated the fitting probability of malignant tumors (13). The Park equation and the Horvath TI-RADS includes more suspicious malignant features and sonographic patterns and are relatively complex. The model proposed by Kwak et al. simplified the number of suspicious malignant signs on US to five. Subsequently, the TI-RADS became widely used in clinical practice.

Over recent years, the development of new US technology has improved the diagnostic accuracy of thyroid nodules, especially the application of contrast-enhanced ultrasound (CEUS). CEUS can be applied non-invasively in real-time and continuously evaluate the perfusion of the microvessels in thyroid nodules under high frequency US (14). Studies have shown that the combination of the TI-RADS classification with CEUS significantly improves the diagnostic accuracy of thyroid nodules (15, 16). However, CEUS enhancement features for thyroid nodules has still not been fully implemented with the TI-RADS system. In 2017, we proposed a new classification standard based on the TI-RADS classification criteria proposed by Horvath et al. (11), Park et al. (12) and Kawk et al. (13), that combined the five ultrasound signs proposed by Kawk TI-RADS (13), and combined it with CEUS enhancement to form a modified version of TI-RADS (17).

We showed that the modified version of TI-RADS significantly increased diagnostic accuracy for the identification of thyroid nodules, particularly for TI-RADS 4a and 4b lesions. This modified

version of TI-RADS was validated by a single center, but all cases were retrospective studies. Thus, the aim of the current study was to confirm the diagnostic efficacy and assess the inter-observer agreement for thyroid nodule characterization using the modified version of TI-RADS in prospective, multi-center trial.

Materials and methods

This prospective multi-center study was approved by the Institutional Review Boards of the ten participating centers

Study population

Between September 2018 and March 2020, we initially collected 3822 consecutive patients from ten centers. The inclusion criteria were as follows: (1) patients with clinically suspected thyroid nodules, (2) patients who consented to undergo CEUS and (3) patients with a final pathological diagnosis as determined by surgical pathology or cytopathological results based on the Bethesda system. Patients were excluded if they refused to undergo final pathological diagnosis or had non-diagnostic or indeterminate cytological results for a lesion without surgical confirmation. Finally, our study featured 3353 patients with 4532 thyroidal nodules. Of the 3353 participants, 729 were male and 2624 were female, patient age ranged from 18 to 82 years with a mean of 46.1 ± 12.2 years). The flow chart of our study is illustrated in Figure 1.

Conventional ultrasound and contrast-enhanced ultrasound examination

All patients underwent conventional ultrasound examinations and CEUS analysis. For conventional US examination, we used a linear, high-frequency probe. Patients were positioned in a supine position with dorsal flexion of the head. Thyroid nodules were then evaluated for location, size, echogenicity, internal composition, margin, shape and the presence/absence of micro-calcification. The internal component of each nodule was classified as solid, mixed or cystic. Echogenicity was classified as hyper-echogenicity, iso-echogenicity, hypo-echogenicity or marked hypo-echogenicity. The margins were classified as irregular or regular. Calcifications, when present, were categorized as micro-calcification (equal to or < 1mm in diameter) or macrocalcification (> 1mm). If a nodule showed both microcalcification and macrocalcification, it was classified as microcalcification. Shape was categorized as taller than wide (greater in its anteroposterior dimension than in its transverse dimension) or wider than tall.

Before starting the multi-center study, all hospitals participating in the center were trained on the specific classification methods and standards to establish a unified approach. Ten experienced radiologists used the TI-RADS classification criteria to classify thyroid nodules according to five ultrasound signs (solid component, marked hypo-echogenicity, taller than wide shape, microcalcification and irregular margin) to evaluate each nodule. This was performed in a blind and independent manner. The TI-RADS classification criteria were as follows (17): TI-RADS score 1: normal thyroid; TI-RADS score 2: no malignant sign, benign lesions; TI-RADS score 3: one malignant sign, high probability of being benign; TI-RADS score 4a: two malignant signs, possibly benign; TI-RADS score 4b: three malignant signs, high probability of malignancy; TI-RADS score 5: four to five malignant signs, highly suggestive of malignancy.

Modified TI-RADS diagnostic criteria in combination with CEUS

The contrast agent used in this study was SonoVue (Bracco, Milan, Italy). A 20-G needle was inserted into the peripheral veins to establish intravenous access. Twenty-five mg of SonoVue was diluted

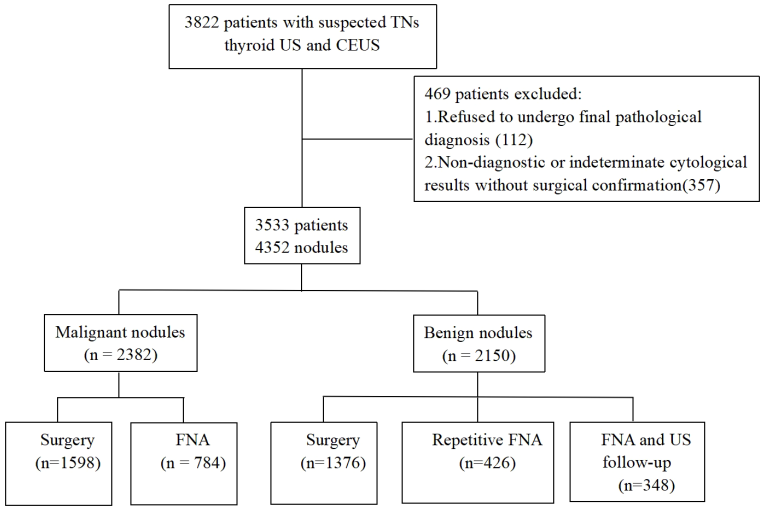


FIGURE 1
Flow chart of the study population. TNs, thyroid nodules; US, ultrasound; CEUS, contrast-enhanced ultrasound; FNA, fine needle aspiration.

in 5 mL of saline and vibrated for 30s to create a microbubble suspension. The suspension was then injected as a bolus and each 2.4 mL injection was then flushed with 5 mL of saline. The dynamic perfusion of the lesion was continuously observed in real time. The CEUS diagnostic criteria were divided into circular enhancement, high enhancement, equal enhancement and low enhancement when compared to the surrounding thyroidal parenchyma.

If CEUS indicated high enhancement or circular enhancement, then the TI-RADS score was reduced by one, if the initial score was 2, then the score remained the same (Figures 2–4). If CEUS indicated low enhancement, then the TI-RADS score was increased by one, a score of 5 remained the same (Figures 5–7). If CEUS indicated equal enhancement, then the TI-RADS classification remained the same. With regards to the modified TI-RADS classification, scores 2-4a were diagnosed as benign and scores 4b-5 were diagnosed as malignant (Table 1).

US-guided FNA procedures

US-guided FNA was performed with 23-gauge needles, each lesion was aspirated three times. Materials obtained from aspiration biopsy were expelled onto glass slides and smeared. All smears were placed immediately in 95% alcohol for Papanicolaou staining. The interpretation of FNA was based on the Bethesda system for reporting thyroid cytopathology.

Statistical analysis

Statistical analyses were performed using SPSS for Windows software (version 23.0, SPSS, Chicago, III, USA). Measurement data are given as mean ± standard deviation and count data are given as percentage and frequency. Analysis of receiver operating characteristic (ROC) curves were used to determine the cut-off value, area under the curve (AUC), and 95% confidence interval (CI). We calculated the ACC, SEN, SPN, PPV and NPV of the modified TI-RADS system to identify malignant thyroid nodules. The level of significance was defined as $p < 0.05$. Cohen's Kappa (κ) coefficient was determined separately to evaluate inter–observer agreement for each of the TI-RADS malignant features. The κ values were interpreted as follows: 0.01–0.20 (poor agreement), 0.21–0.40 (fair agreement), 0.41–0.60 (moderate agreement), 0.61–0.80 (good agreement) and 0.81–1.0 (very good agreement).

Results

Nodule diagnosis

The final diagnosis of the 4532 nodules was benign in 2150 (47.4%) nodules and malignant in 2382 (52.6%) nodules. Final diagnoses were determined by surgical resection in 1598 of the 2382 malignant nodules, including 1529 papillary thyroid carcinomas (PTC), 39 follicular carcinomas, 14 cases of focal canceration of nodular goiter, 7 medullary thyroid carcinomas (MTC), 4 anaplastic thyroid carcinomas (ATC), 3 metastatic carcinomas and 2 lymphomas. In total, 784 of the malignant nodules diagnosed by FNA were PTC. The 1376 surgically confirmed benign nodules included 186 adenomas, 990 nodular goiter, 148 Hashimoto's thyroiditis, and 52 cases of subacute thyroiditis. Overall, 774 benign nodules were diagnosed based on repetitive benign FNA results or benign FNA results by US follow-up studies.

US features of thyroid nodules

The mean maximum diameter of nodules was 13.6 ± 11.8 mm (range: 4.5–59.0 mm) and the mean size of benign nodules was 18.4 ± 14.0 mm, this was significantly larger than that of malignant nodules (9.3 ± 7.1 mm, $p < 0.001$). Of the 4532 thyroid nodules, there were 3673 solid composition nodules and 859 mixed composition nodules, 413 marked hypoechoic nodules, 2021 hypoechoic nodules, 473 isoechoic nodules and 1625 hyperechoic nodules. We identified 1453 nodules with irregular margins and 3079 nodules with regular margins, 1733 nodules with microcalcifications, 341 nodules with macrocalcifications and 2458 nodules with no calcifications. There were 3339 wider than tall nodules and 1193 taller than wide nodules (Table 2).

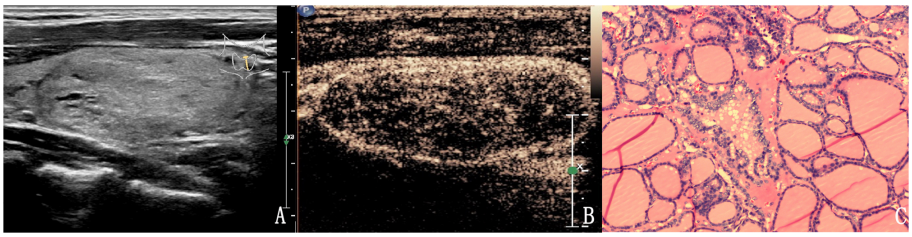


FIGURE 2
The case of a 46-year-old woman with a 37.3 × 18.1 × 24.7 mm solid hyper-echoic nodule in the left lobe of the thyroid. **(A)** Conventional two-dimensional image showed that the nodule had one malignant indicator (solid) and was classified with a TI-RADS score of 3. **(B)** Ultrasound contrast image showing ring enhancement. The modified version of TI-RADS combined with CEUS returned a score of 2 and the patient was diagnosed with a benign nodule. **(C)** Pathological image of the lesion, a nodular goiter.

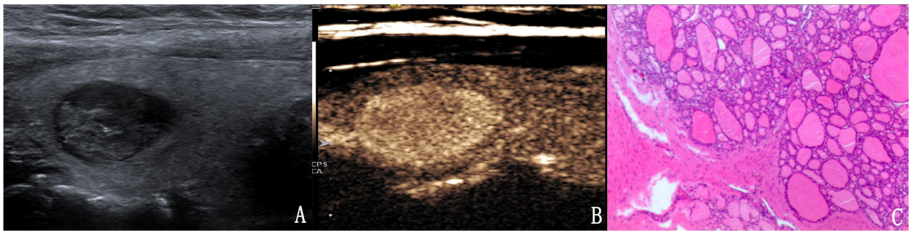


FIGURE 3
The case of a 31-year-old woman with a 15.1 × 10.1 × 9.8 mm solid hypo-echoic nodule in the right lobe of the thyroid. **(A)** Conventional ultrasound showed that the nodule had one malignant indicator (solid) and was classified as a TI-RADS score of 3. **(B)** Contrast-enhanced ultrasound detected high enhancement. The modified TI-RADS resulted in a score of 2 and indicated a benign nodule. **(C)** Pathology of the lesion showed an adenoma.

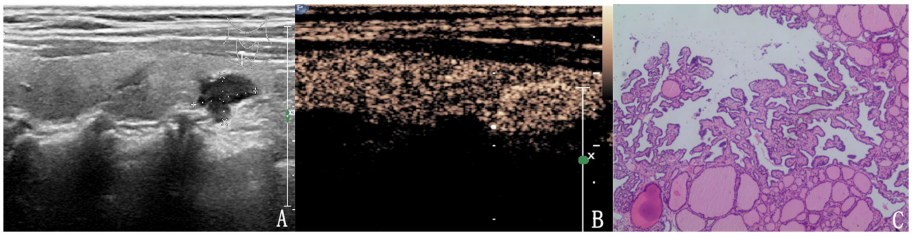


FIGURE 4
The case of a 64-year-old woman with a 10.6 × 6.7 × 5.9 mm solid mark hypo-echoic nodule in the right lobe of the thyroid. **(A)** Conventional two-dimensional image showing that the nodule had three malignant indicators (solid, mark hypo-echoic and irregular margin) and was classified with a TI-RADS score of 4b. **(B)** Ultrasound contrast image showing high enhancement. The modified version of the TI-RADS combined with CEUS returned a score of 4a, and the diagnosis was a benign nodule. **(C)** Pathological image of the lesion, a nodular goiter.

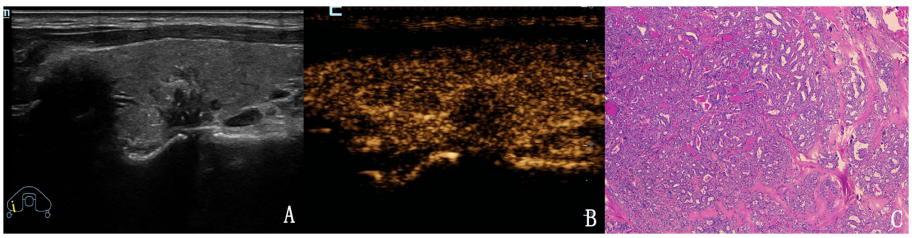


FIGURE 5
The case of a 43-year-old woman with a 15.2 × 11.1 × 12.5 mm solid hypo-echoic nodule in the right lobe of the thyroid. **(A)** Conventional two-dimensional image showing that the nodule had three malignant indicators (solid, irregular margin and microcalcifications) and was classified with a TI-RADS score of 4b. **(B)** Ultrasound contrast image showing low enhancement. The modified version of TI-RADS combined with CEUS returned a score of 5, and the patient was diagnosed with a malignant nodule. **(C)** Pathological image of the lesion, a PTC.

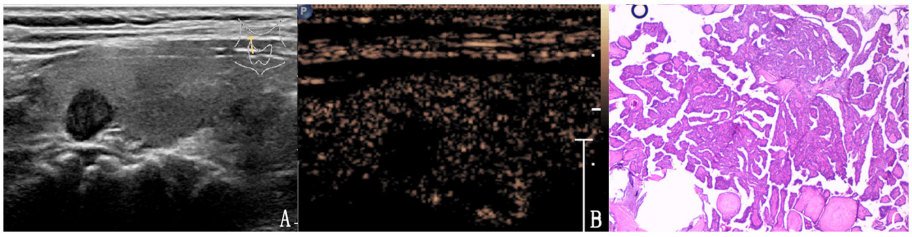


FIGURE 6
The case of a 47-year-old man with a 7.8 × 6.8 × 7.5 mm solid mark hypo-echoic nodule in the right lobe of the thyroid. **(A)** Conventional two-dimensional image showing that the nodule had two malignant indicators (solid, mark hypoechoic) and was classified with a TI-RADS score of 4a. **(B)** Ultrasound contrast image showing low enhancement. The modified version of the TI-RADS combined with CEUS returned a score of 4b and the diagnosis was a malignant nodule. **(C)** Pathological image of the lesion, a PTC.

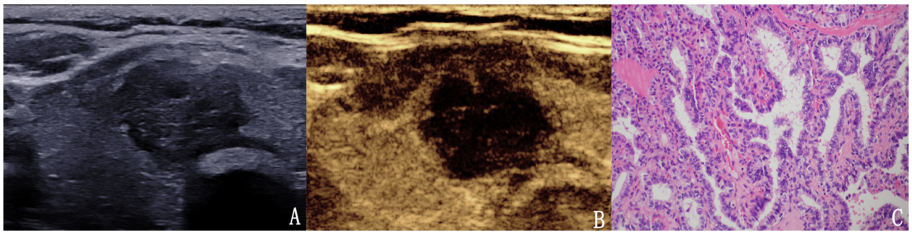


FIGURE 7
The case of a 51-year-old woman with a 20.4 × 15.1 × 16.9 mm solid mark hypo-echoic nodule in the right lobe of the thyroid. **(A)** Conventional two-dimensional image showing that the nodule had three malignant indicators (solid, mark hypo-echoic and irregular margin) and was classified with a TI-RADS score of 4b. **(B)** Ultrasound contrast image showing low enhancement. The modified version of the TI-RADS combined with CEUS returned a score of 5, and the diagnosis indicated a malignant nodule. **(C)** Pathological image of the lesion, a PTC.

TABLE 1 Modified TI-RADS diagnostic criteria in combination with CEUS.

Modified TI-RADS classification	Definition	Risk of malignancy	Recommended
TI-RADS 2	benign lesions	0	Long-term follow-up
TI-RADS 3	high probability of benignity	<5%	Short-term follow-up
TI-RADS 4a	possible benignity	5~15%	FNA
TI-RADS 4b	high probability of malignancy	15~90%	FNA
TI-RADS 5	highly suggestive of malignancy	>90%	Clinical treatment

Malignancy risk according to category in the modified TI-RADS

Of the 4352 thyroid nodules assessed, 159 (3.5%) were classified as TI-RADS 2, 1256 (27.7%) as TI-RADS 3, 649 (14.3%) as TI-RADS 4a, 1323 (29.2%) as TI-RADS 4b and 1145 (25.3%) as TI-RADS 5. Of the 159 thyroid nodules categorized as TI-RADS 2, none were malignant (0%). Of the 1256 thyroid nodules categorized as TI-RADS 3, 52 were malignant (4.1%). Of the 649 thyroid nodules categorized as TI-RADS 4a, 78 were malignant (12.0%). Of the 1323 thyroid nodules categorized as TI-RADS 4b, 1138 were malignant (86.0%). Of the 1145 thyroid nodules categorized as TI-RADS 5, 1092 were malignant (95.4%). The calculated malignancy risk in the modified TI-RADS categories 5, 4b, 4a, 3 and 2 nodules was 95.4%, 86.0%, 12.0%, 4.1% and 0%, respectively, and

TABLE 2 US features of thyroid nodules.

US features	Total lesions	Benign lesions	Malignant lesions	*P value*
Composition				<0.001
Solid	3673	1357	2316	
Mixed	859	813	46	
Echogenicity				<0.001
Marked hypoechoic	413	152	261	
Hyper/iso/hypoechoic	4119	2018	2101	
Margin				<0.001
Irregular	1453	217	1236	
Regular	3079	1953	1126	
Calcification				<0.001
Microcalcifications	1733	439	1294	
Macrocalcifications/no calcifications	2799	1731	1068	
Shape				<0.001
Taller than wide	1193	210	983	
Wider than tall	3339	1960	1379	

all were estimated within the range of the suggested malignancy risk in the modified TI-RADS (Table 3).

Diagnostic performance of the modified version of the TI-RADS for predicting malignant thyroid nodules

ROC curve analysis was used to analyze the diagnostic efficacy of the modified TI-RADS classification for differentiating benign and malignant thyroid nodules Figure 8. The AUC under the ROC curve was 0.936 (95% CI: 0.928–0.943, $p < 0.01$) and the best cut-off value for predicting malignant thyroid nodules was > TI-RADS 4a. Considering TI-RADS 4b and TI-RADS 5 together as predictors for malignancy, the SEN, ACC, PPV, NPV and SPN were 93.6%, 91.9%, 90.4%, 93.7% and 88.5%, respectively.

Inter-observer agreement

We calculated Cohen's Kappa value for each of the five US features and the CEUS enhancement pattern (Table 4). The highest inter-observer agreement was observed for the taller than wide shape and for microcalcification, the Kappa value for these two features was 0.94 and 0.93, respectively. The Kappa value for marked hypoechoic, solid composition, irregular margins and enhancement pattern of the CEUS was 0.75, 0.89, 0.86 and 0.81, respectively. There was also good agreement between the observers for the modified TI-RADS classification, the Kappa value was 0.80, thus implying that the modified TI-RADS system showed comparable results when used for the analysis of thyroid nodules by different radiologists.

Discussion

TI-RADS is a quantitative scoring method that has been developed over recent years. This system can stratify the risk of malignancy for thyroid nodules and standardize the US reports for the thyroid. Consequently, this method is an effective form of communication between clinicians and pathologists. Since Horvath et al. (11) first proposed TI-RADS as a quantitative system for the risk stratification of thyroid nodules in 2009, its format and content have evolved and undergone significant development. Many researchers (12, 13, 17–19) have proposed different TI-RADS classification systems which have been used for the effective management of US for thyroid nodules. However, despite these efforts, there were many different versions and complex models of the TI-RADS classification and there was no unified classification standard. Furthermore, there were different guidelines for TI-RADS in different regions and countries, and there were certain differences between different guidelines this meant that the system was not widely adopted across the world. Therefore, there has been many attempts to develop a practical and standardized risk stratification system for thyroid nodules so as to provide consistent management strategies for assessing thyroid nodules in clinical practice (20).

When facilitated by micro-bubble contrast agents, CEUS can display microvessels, large vessels and dynamic perfusion simultaneously. Compared with conventional ultrasound, CEUS can reveal better characteristics of focal thyroid nodules (21). At present, CEUS is widely used for the differential diagnosis of benign and malignant thyroid nodules. Zhao et al. (15) showed that CEUS has high value for the differentiation of benign and malignant thyroid nodules, and was significantly more useful than conventional ultrasound. The results of the present study showed that benign thyroid nodules mainly showed ring enhancement, high enhancement or equal enhancement, while malignant nodules mainly showed low enhancement, especially non-homogeneous low enhancement, these findings were similar to those of previous studies (22). In this study, low enhancement was used as the standard to judge malignant nodules. The SEN, SPN, ACC, PPV and NPV of benign and malignant thyroid nodules diagnosed by CEUS were 82.8%, 81.6%, 82.3%, 83.0% and 81.4%, respectively. Low enhancement was considered to be the main enhancement mode for thyroid malignant nodule CEUS (23–25). Compared with the low enhancement mode of malignant nodules, benign nodules mainly showed high enhancement, equal enhancement and ring enhancement (26–28).

Many studies have shown that the combination of CEUS and TI-RADS classification for conventional ultrasound can improve the accuracy of diagnosing benign and malignant thyroid nodules. Zhao et al. (15) retrospectively analyzed the conventional ultrasound and CEUS enhancement characteristics of 117 cases of thyroid nodules and compared the diagnostic efficiency of TI-RADS alone against CEUS combined with TI-RADS for predicting benign and malignant thyroid nodules. The results showed that the ACC, SEN, SPN, PPV and NPV for TI-RADS + CEUS were the highest and were significantly higher than that for TI-RADS or CEUS alone. Ruan et al. (29) constructed a CEUS TI-RADS by adding CEUS to widely accepted nonenhanced US features, the CEUS TI-RADS showed the highest AUC under the ROC curve comparison with all other systems (AUC=0.93, P<0.001), the highest biopsy yield of malignancy at 66% (157 of 239 nodules), and the lowest unnecessary biopsy rate at 34% (82 of 239 nodules). Our previous retrospective study (17) of 298

TABLE 3 Malignancy risk according to category in the modified version of the TI-RADS.

Modified TI-RADS category	n	Malignant risk(%)	Calculated malignancy risk(%)	Frequency(%)
TI-RADS2	159	0	0(0/159)	3.5(159/4532)
TI-RADS3	1256	<5	4.1(52/1256)	27.7(1256/4532)
TI-RADS4a	649	5-15	12.0(78/649)	14.3(649/4532)
TI-RADS4b	1323	15-90	86.0(1138/1323)	29.2(1323/4532)
TI-RADS5	1145	>90	95.4(1092/1145)	25.3(1145/4532)

10.3389/fendo.2023.1080908

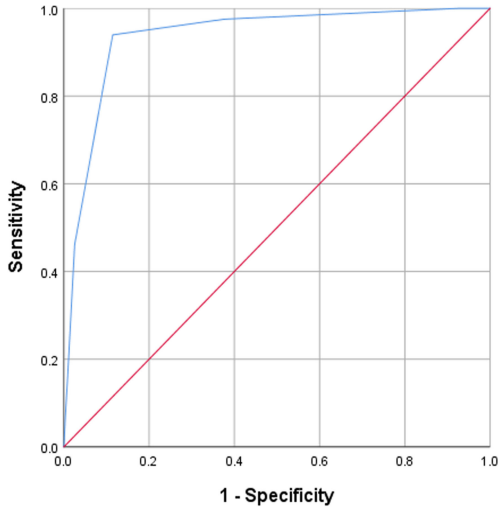


FIGURE 8
ROC analyses for the diagnostic performance of the modified version of the TI-RADS for predicting the malignancy of thyroid nodules. The best cut-off was > TIRADS 4a, resulting in 93.6% SEN and 88.5% SPN.

thyroid nodules in 206 patients showed that the SEN, SPN, ACC, PPV and NPV acquired by CEUS combined with TI-RADS were 96.3%, 94.7%, 95.0%, 80.0% and 99.1%, respectively. Furthermore, the diagnostic efficiency of CEUS for judging benign and malignant thyroid nodules was significantly higher than that of TI-RADS or CEUS alone.

Many studies have stratified the risk of each TI-RADS category separately. Although there were some differences between these studies, they all reported a common pattern, with the risk of malignancy increasing from the TI-RADS 2 to the TI-RADS 5 category. The risk of malignancy described by Horvath et al. (11) was 0, < 5%, 5%–10%, 10%–80% and > 80%, respectively for TI-RADS 2, 3, 4a, 4b, and 5 lesions. Horvath et al. prospectively verified the diagnostic value of this TI-RADS classification for evaluating 1097 benign and malignant thyroid nodules. This previous study showed that the SEN, SPN, PPV, NPV and ACC were 88%, 49%, 49%, 88% and 94%, respectively. The risks of malignancy reported by Park et al. (12) was 0%–7%, 8%–23%, 24%–50%, 51%–90% and 91%–100%, respectively for TI-RADS 1, 2, 3, 4 and 5 lesions. Wang et al. (30) reported that the SEN, SPN, ACC, PPV and NPV when using Park's TI-RADS system were 96.8%, 71.3%, 83.0%, 74.1% and 96.3%, respectively. The risks of malignancy reported by Kwak et al. (13) were 0%, 2–2.8%, 3.6–12.7%, 6.8–37.8%, 21%–91.9% and 88.7%–97.9%, respectively, for TI-RADS2, 3, 4a, 4b, 4c and 5 lesions. Zhang

TABLE 4 The inter-observer agreement for US features and TI-RADS categorization for the diagnosis of thyroid nodules.

Feature	κ coefficients	P Value
Composition	0.89	<0.001
Solid/Mixed		
Echogenicity	0.75	<0.001
Marked hypoeechoic/hyper/iso/hypoechoic		
Margins	0.86	<0.001
Well circumscribed/Irregular		
Calcification	0.93	<0.001
Microcalcifications/macrocalcifications/no calcifications		
Shape	0.94	<0.001
Taller than wide/Wider than tall		
Enhancement mode	0.81	<0.001
Ring/High/Equal/low enhancement		
Modified TI-RADS categorization	0.80	<0.001
TI-RADS 2/3/4a/4b/5		

et al. (17) reported that the SEN, SPN, PPV, NPV and ACC were 94.4%, 69.3%, 40.5%, 98.3% and 73.8%, respectively. In our present study, the calculated risk of malignancy for the modified TI-RADS categories 5, 4b, 4a, 3 and 2 nodules were 95.4%, 86.0%, 12.0%, 4.1% and 0%, respectively. The risk of malignancy risk for the five categories was within the range of the suggested risk of malignancy. Therefore, the modified TI-RADS can be applied for the qualitative diagnosis of benign and malignant thyroid nodules.

Based on ROC curve analyses, our study revealed an AUC of 0.936, the best cut-off value for predicting malignant thyroid nodules was > TI-RADS4a. When considering TI-RADS 4b and TI-RADS 5 together as predictors for malignancy, the SEN, ACC, PPV, NPV and SPN were 93.6%, 91.9%, 90.4%, 93.7% and 88.5%, respectively. In our study, the modified TI-RADS had high diagnostic efficiency for the diagnosis of benign and malignant thyroid nodules. Of the 4352 thyroid nodules assessed, 159 were classified as TI-RADS 2, none were malignant and the diagnostic accordance rate was 100%. In total, 1256 were classified as TI-RADS 3, of which 1204 were benign nodules; the diagnostic accordance rate was 95.9%. In total, 649 were classified as TI-RADS 4a, of which 571 were benign nodules; the diagnostic accordance rate was 88.0%. In total, 1323 were classified as TI-RADS 4b, of which 1138 were malignant nodules; the diagnostic accordance rate was 86.0%. In total, 1145 were classified as TI-RADS 5, of which 1092 were malignant nodules; the diagnostic accordance rate was 95.4%. Except for TI-RADS 2 nodules, there was a certain misdiagnosis rate in other classified nodules. There are several potential reasons for these findings. First, as with conventional two-dimensional ultrasound, CEUS enhancement patterns for some benign and malignant nodules can also overlap. For example, some malignant nodules had a rich blood supply and showed high enhancement, while some benign nodules may exhibit scar hyperplasia or fibrous tissue hyperplasia, thus resulting in low enhancement. Second, some nodules were so small that it was difficult to judge their CEUS enhancement mode. Third, some benign nodules were often associated with focal PTC or FTC carcinogenesis; therefore, it was difficult to correctly diagnose these focal forms of carcinogenesis. Fourth, the presence of hidden PTMC in thyroid glands with diffuse lesions was not typical on conventional two-dimensional ultrasound or CEUS, thus increasing the difficulty of diagnosis. Finally, some thyroid inflammatory lesions were similar to malignant lesions in conventional two-dimensional ultrasound or CEUS enhancement mode; it was difficult to correctly diagnose these inflammatory lesions.

The modified TI-RADS is simple and accurate for the evaluation of benign and malignant thyroid nodules. The radiologists only need to accurately evaluate the conventional two dimensional ultrasound signs and CEUS enhancement mode to classify nodules. Theoretical malignant risk and clinical treatment suggestions were also given for each classification of nodules, thus allowing better communication between clinicians and pathologists. Another important role of the modified TI-RADS was to standardize the criteria for different radiologists to evaluate the signs of thyroid nodules. Therefore, the classification system should have good consistency and repeatability among different radiologists (31–33). Our study revealed strong inter-observer agreement between different radiologists when using the modified TI-RADS categories and features for thyroid nodule characterization. We found that the highest inter-observer agreement was for shape and micro-calcification. The Kappa value for these two features was 0.94 and 0.93, respectively, thus showing excellent consistency among the different observers. There was also good consistency among the observers for echogenicity, composition, margins and enhancement mode; the Kappa values were 0.75, 0.89, 0.86 and 0.81, respectively. In addition, different observers showed good consistency when using the modified TI-RADS classification system; the Kappa value was 0.80. These results showed that the modified TI-RADS produced comparable results for the analysis of suspicious thyroid nodules when used by different radiologists in thyroid imaging.

This study had some limitations that need to be considered. First, all US examinations were performed and analyzed by highly experienced radiologists. Further studies relating to the performance of this reporting system when applied by less experienced radiologists may be needed. Second, we did not compare the modified TI-RADS to the other TI-RADS classification systems. Third, we only evaluated inter-observer consistency, we did not evaluate intra-observer consistency, this needs to be addressed by future studies.

Conclusion

The modified TI-RADS had high diagnostic efficiency for the diagnosis of benign and malignant thyroid nodules. The inter-observer agreement for the modified TI-RADS category was excellent, thus suggesting that this classification is very suitable for clinical application.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

PiZ and FC have equal contributions to this article. Drafting the work or revising: WL, PiZ. Acquisition, analysis, or interpretation of

data: WL, FC, PeZ, LX, LW, ZW, YY, XL, BW, WY, HZ, YT. Conception or design: WL, PiZ. Final approval of the manuscript: PiZ.

Funding

This research was supported by grants from the National Natural Science Foundation of China (Reference: 81871367), the Natural Science Foundation of Hunan Province, China (Reference: 2022JJ30894) and the Project of Hunan Provincial Health Commission (Reference: 202209025123).

Acknowledgments

We thank Prof. Martin Krix, Hanshu Li and Caixiang Xia saw for their constructive suggestions to improve our manuscript. We thank Lifang Li and Yan Liu for assistance in intravenous injection of the contrast agent. We thank Yang Liu, Yu Xin and Sheng Xiao for assistance in cytology. We thank Xinghao Zhang, Xiangqi Zhou and Zhen Huang for assistance in statistical analysis.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Londero SC, Krogdahl A, Bastholt L, Overgaard J, Trolle W, Henrik Baymler Pedersen HB, et al. Papillary thyroid microcarcinoma in Denmark 1996-2008: a national study of epidemiology and clinical significance. *Thyroid* (2013) 23(9):1159–64. doi: 10.1089/thy.2012.0595

2. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. *CA Cancer J Clin* (2010) 60:277–300. doi: 10.3322/caac.20073

3. Magreni A, Bann DV, Schubart JR, Goldenberg D. The effects of race and ethnicity on thyroid cancer incidence. *JAMA Otolaryngol Head Neck Surg* (2015) 141:319–23. doi: 10.1001/jamaoto.2014.3740

4. Smith-Bindman R, Lebda P, Feldstein VA, Sellami D, Goldstein RB, Brasic N, et al. Risk of thyroid cancer based on thyroid ultrasound imaging characteristics: results of a population-based study. *JAMA Intern Med* (2013) 173:1788–96. doi: 10.1001/jamainternmed.2013.9245

5. Uppal A, White MG, Nagar S, Aschebrook-Kilfoy B, Chang PJ, Angelos P, et al. Benign and malignant thyroid incidentalomas are rare in routine clinical practice: a review of 97908 imaging studies. *Cancer Epidemiol Biomarkers Prev* (2015) 24:1327–31. doi: 10.1158/1055-9965.EPI-15-0292

6. Iannuccilli JD, Cronan JJ, Monchik JM. Risk f or malignancy of thyroid nodules as assessed by sonographic crit eria: the need for biopsy. *J Ultrasound Med* (2004) 23 (11):1455–64. doi: 10.7863/jum.2004.23.11.1455

7. Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, Mandel SJ, et al. Revised American thyroid association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid* (2009) 19(11):1167–214. doi: 10.1089/thy.2009.0110

8. Cheng SP, Lee JJ, Lin JL, Chuang SM, Chien MN, Liu CL. Characterization of thyroid nodules using the proposed thyroid imaging reporting and data system (TI-RADS). *Head Neck* (2013) 35:541–7. doi: 10.1002/hed.22985

9. Kim GR, Kim MH, Moon HJ, Chung WY, Kwak JY, Kim EK. Sonographic characteristics suggesting papillary thyroid carcinoma according to nodule size. *Ann Surg Oncol* (2013) 20:906–13. doi: 10.1245/s10434-012-2830-4

10. Remonti LR, Kramer CK, Leitao CB, Pinto LCF, Gross JL. Thyroid ultrasound features and risk of carcinoma: a systematic review and meta-analysis of observational studies. *Thyroid* (2015) 25(5):538–50. doi: 10.1089/thy.2014.0353

11. Horvath E, Majlis S, Rossi R, Franco C, Niedmann JP, Castro A, et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. *J Clin Endocrinol Metab* (2009) 94(5):1748–51. doi: 10.1210/jc.2008-1724

12. Park JY, Lee HJ, Jang HW, Kim HK, Yi JH, Lee W, et al. A proposal for a thyroid imaging reporting and data system for ultrasound features of thyroid carcinoma. *Thyroid* (2009) 19(11):1258–64. doi: 10.1089/thy.2008.0021

13. Kwak JY, Han KH, Yoon JH, Moon HJ, Son EJ, Park SH, et al. Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk. *Radiology* (2011) 260(3):892–9. doi: 10.1148/radiol.11110206

14. Cosgrove D. Future prospects for SonoVue and CPS. *Eur Radiol* (2004) 14:116–24. doi: 10.1007/s10406-004-0084-3

15. Zhao H, Liu X, Lei B, Chen P, Li J, Wu Y, et al. Diagnostic performance of thyroid imaging reporting and data system (TI-RADS) alone and in combination with contrast-enhanced ultrasonography for the characterization of thyroid nodules. *Clin Hemorheol Microcirc* (2019) 72(1):95–106. doi: 10.3233/CH-180457

16. Huang Y, Hong Y, Xu W, Song K, Huang P. Contrast-enhanced ultrasound improves the accuracy of the ACR TI-RADS in the diagnosis of thyroid nodules located in the isthmus. *Ultraschall Med* (2022) 43(06):599–607. doi: 10.1055/a-1543-6033

17. Zhang Y, Zhou P, Tian SM, Zhao YF, Li JL, Li L. Usefulness of combined use of contrast-enhanced ultrasound and TI-RADS classification for the differentiation of benign from malignant lesions of thyroid nodules. *Eur Radiol* (2017) 27:1527–36. doi: 10.1007/s00330-016-4508-y

18. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. American thyroid association management guidelines for adult patients with thyroid nodulesand differentiated thyroid cancer: The American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid: Off J Am Thyroid Association,2016* (2015) 26:1–133. doi: 10.1089/thy.2015.0020

19. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR thyroid imaging, reporting and data system (TI-RADS): White paper of the ACR TI-RADS committee. *J Am Coll Radiol* (2017) 14:587–95. doi: 10.1016/j.jacr.2017.01.046

20. Middleton WD, Teefey SA, Reading CC, Langer J, Beland MD, Szabunio MM, et al. Multi-institutional analysis of thyroid nodule risk stratification using the American college of radiology thyroid imaging reporting and data system. *AJR Am J Roentgenol* (2017) 208(6):1331–41. doi: 10.2214/AJR.16.17613

21. Sidhu PS, Cantisani V, Dietrich CF, Gilja OH, Saftoiu A, Bartels E, et al. The EFSUMB guidelines and recommendations for the clinical practice of contrast-enhanced ultrasound (CEUS) in nonhepatic applications: update 2017 (long version). *Ultraschall Med* (2018) 39(2):e2–e44. doi: 10.1055/a-0586-1107

22. Deng J, Zhou P, Tian SM, Zhang L, Li JL, Qian Y. Comparison of diagnostic efficacy of contrast-enhanced ultrasound, acoustic radiation force impulse imaging, and their combined use in differentiating focal solid thyroid nodules. *PloS One* (2014) 9(3): e90674. doi: 10.1371/journal.pone.0090674

23. Schleder S, Janke M, Agha A, Schacherer D, Hornung M, Schlitt HJ, et al. Preoperative differentiation of thyroid adenomas and thyroid carcinomas using high resolution contrast-enhanced ultrasound(CEUS). *Clin Hemorheol Microcirc* (2015) 61 (1):13–22. doi: 10.3233/CH-141848

24. Zhang B, Jiang YX, Liu JB, Yang M, Dai Q, Zhu QL, et al. Utility of contrast-enhanced ultrasound for evaluation of thyroid nodules. *Thyroid Off J Am Thyroid Assoc* (2010) 20(1):51–7. doi: 10.1089/thy.2009.0045

25. Zhao RN, Zhang B, Yang X, Jiang YX, Lai XJ, Zhang XY. Logistic regression analysis of contrast-enhanced ultrasound and conventional ultrasound characteristics of sub-centimeter thyroid nodules. *Ultrasound Med Biol* (2015) 41:3102–08. doi: 10.1016/j.ultrasmedbio.2015.04.026

26. Li FS, Luo HT. Comparative study of thyroid puncture biopsy guided by contrast-enhanced ultrasonography and conventional ultrasound. *Exp Ther Med* (2013) 5(5):1381–84. doi: 10.3892/etm.2013.1016

27. Bartolotta TV, Midiri M, Galia M, Runza G, Attard M, Savoia G, et al. Qualitative and quantitative evaluation of solitary thyroid nodules with contrast-enhanced ultrasound: initial results. *Eur Radiol* (2006) 16(10):2234–41. doi: 10.1007/s00330-006-0229-y

28. Cantisani V, Consorti F, Guerrisi A, Guerrisi I, Ricci P, Segni MD, et al. Prospective comparative evaluation of quantitative-elastosonography (Q-

elastography) and contrast-enhanced ultrasound for the evaluation of thyroid nodules: Preliminary experience. *Eur J Radiol* (2013) 82(11):1892–8. doi: 10.1016/j.ejrad.2013.07.005

29. Ruan J, Xu X, Cai Y, Zeng H, Luo M, Zhang W, et al. A practical CEUS phyroid reporting system for thyroid nodules. *Radiology* (2022) 305(1):149–59. doi: 10.1148/radiol.212319

30. Wang Y, Lei KR, He YP, Li XL, Ren WW, Zhao CK, et al. Malignancy risk stratification of thyroid nodules: Comparisons of four ultrasound thyroid imaging reporting and data systems in surgically resected nodules. *Sci Rep* (2017) 7(1):11560. doi: 10.1038/s41598-017-11863-0

31. Chung R, Rosenkrantz AB, Bennett GL, Dane B, Jacobs JE, Slywotzky C, et al. Interreader concordance of the TI-RADS: Impact of radiologist experience. *Am J Roentgenol* (2020) 214(5):1152–7. doi: 10.2214/AJR.19.21913

32. Teng D, Fu P, Li W, Guo F, Wang H. Learnability and reproducibility of ACR thyroid imaging, reporting and data system (TI-RADS) in postgraduate freshmen. *Endocrine* (2020) 67(5):643–50. doi: 10.1007/s12020-019-02161-y

33. Basha MAA, Alnaggar AA, Refaat R, EI-Maghraby AM, Refaat MM, Elhamed MEA, et al. The validity and reproducibility of the thyroid imaging reporting and data system (TI-RADS) in categorization of thyroid nodules: Multicentre prospective study. *Eur J Radiol* (2019) 117:184–92. doi: 10.1016/j.ejrad.2019.06.015

frontiers | Frontiers in Endocrinology

TYPE Review
PUBLISHED 24 February 2023
DOI 10.3389/fendo.2023.1101410

OPEN ACCESS

EDITED BY
Jeffrey Garber,
Atrius Health, United States

REVIEWED BY
Sara Ahmadi,
Brigham and Women's Hospital, and Harvard Medical School, United States
David Cohen,
Kaiser Foundation Hospital, United States
Elizabeth Hall,
Beth Israel Deaconess Medical Center, and Harvard Medical School, United States
Phillip Pellitteri,
Geisinger Health System, United States
Anupam Kotwal,
University of Nebraska Medical Center, United States

*CORRESPONDENCE
Elizabeth E. Cottrill
Elizabeth.Cottrill@jefferson.edu

SPECIALTY SECTION
This article was submitted to Thyroid Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 17 November 2022
ACCEPTED 06 February 2023
PUBLISHED 24 February 2023

CITATION
Patel J, Klopper J and Cottrill EE (2023) Molecular diagnostics in the evaluation of thyroid nodules: Current use and prospective opportunities.
Front. Endocrinol. 14:1101410.
doi: 10.3389/fendo.2023.1101410

Molecular diagnostics in the evaluation of thyroid nodules: Current use and prospective opportunities

Jena Patel[1], Joshua Klopper[2] and Elizabeth E. Cottrill[1*]

[1]Department of Otolaryngology – Head & Neck Surgery, Thomas Jefferson University Hospital, Sidney Kimmel Medical College, Philadelphia, PA, United States, [2]Department of Medical Affairs, Veracyte, San Francisco, CA, United States

Thyroid cancer is the most common endocrine malignancy with an estimated 43,800 new cases to be diagnosed in 2022 and representing the 7th most common cancer in women. While thyroid nodules are very common, being identified in over 60% of randomly selected adults, only 5-15% of thyroid nodules harbor thyroid malignancy. Therefore, it is incumbent upon physicians to detect and treat thyroid malignancies as is clinically appropriate and avoid unnecessary invasive procedures in patients with benign asymptomatic lesions. Over the last 15-20 years, rapid advances have been made in cytomolecular testing to aid in thyroid nodule management. Initially, indeterminate thyroid nodules, those with Bethesda III or IV cytology and approximately a 10-40% risk of malignancy, were studied to assess benignity or malignancy. More recently, next generation sequencing and micro-RNA technology platforms have refined the diagnostic capacity of thyroid nodule molecular testing and have introduced opportunities to glean prognostic information from both cytologically indeterminate and malignant thyroid nodules. Therefore, clinicians can move beyond determination of malignancy, and utilize contemporary molecular information to aid in decisions such as extent of surgery and post-therapy monitoring plans. Future opportunities include molecularly derived information about tumor behavior, neo-adjuvant treatment opportunities and response to thyroid cancer therapies.

KEYWORDS

thyroid nodule, thyroid cancer (TC), diagnosis, prognosis, targeted therapy, molecular markers

Introduction

Thyroid cancer is the most common endocrine malignancy with an estimated 43,800 expected new cases diagnosed in 2022 and representing the 7th most common cancer in women (1). Thyroid cancer almost always presents as a thyroid nodule and thyroid nodules are very common with over 60% of the population having one or more by the time patients

reach their 7th and 8th decades of life (2). However, only 5-15% of thyroid nodules harbor thyroid malignancy. Fine needle aspiration cytology (FNAC) is the foundation for diagnosis of nodules that meet criteria for biopsy, and a Bethesda II (BII - benign) or Bethesda VI (BVI - malignant) cytology result has excellent accuracy and correlation with final histopathology upon surgical resection (2–4). BII cytology predicts benign histology 97% of the time or greater and BVI cytology confers a risk of malignancy up to 99% (4). The primary challenge in the evaluation of thyroid nodules occurs in the setting of Bethesda III (BIII) or Bethesda IV (BIV) cytology, often grouped together as indeterminate thyroid nodules (ITN).

Approximately 20-25% of thyroid nodule aspirates result in ITN cytology (5). The risk of malignancy of BIII and BIV ITN ranges from 6-40% depending on the institution and the categorization of noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP) as benign or malignant (4). Historically, consensus guidelines recommended surgery, often in the form of a thyroid lobectomy, for definitive diagnosis of ITN since it is often not possible to differentiate between benign and malignant nodules by cytology alone (6, 7). This approach is sub-optimal given the cost, possible morbidity, and need for thyroid hormone replacement in a subset of patients after lobectomy and all patients after total thyroidectomy; especially since ~75% of ITN will prove to be benign on final histopathology (4, 8, 9). The utilization of transcriptional signatures and discovery of driver mutations promoting thyroid cancer development and influencing its behavior provided the molecular foundation for improved diagnostic accuracy in ITN (10, 11). As will be described, molecular diagnostics has moved beyond aiding in diagnosis and can provide information on tumor prognosis (12).

The goal of this review is to provide an update on commercially available lab developed molecular diagnostic tests for use in nodular thyroid disease. The contemporary clinical use, advantages, and disadvantages, as well as future potential applications will be discussed.

Diagnostic test performance metrics review

A brief review of test sensitivity (SN), specificity (SP), negative predictive value (NPV), and positive predictive value (PPV) is warranted to promote appropriate understanding and scrutiny of molecular diagnostic performance metrics (Figure 1) (13–15). SN is a calculation of the number of true positives (for this topic, the patient has thyroid cancer, and the molecular test reports a positive finding) divided by all the patients with thyroid cancer (who have true positive plus false negative test results). A low SN indicates thyroid cancers have been missed (called negative or benign) by the molecular marker test. Alternatively, SP is a calculation of the true negatives (the patient does not have thyroid cancer and the test is negative) divided by all the patients without thyroid cancer (true negative plus false positive test results) (Figure 1) (15).

Clinically, NPV and PPV are better indicators of a test's ability to rule out or rule in disease, respectively. NPV is a calculation of the true negatives divided by all the patients with a negative test result (true negatives and false negatives). PPV is a calculation of the true positives divided by all the patient with a positive test result (true positives and false positives) (13, 14). At any given SN and SP, both NPV and PPV are affected by the disease prevalence in the population such that a higher disease prevalence will result in a higher PPV and lower NPV than in a population with a lower prevalence of disease (Figure 2).

Other measures of diagnostic performance include overall accuracy, which is the proportion of correctly identified patients (true positive and true negative results) relative to the entire cohort, and likelihood ratios, the probability of the expected test result in those with thyroid cancer as compared to the same result in those without (16).

It is critical that a thyroid nodule molecular diagnostic test is validated with a high-quality study that ideally is prospective, multi-

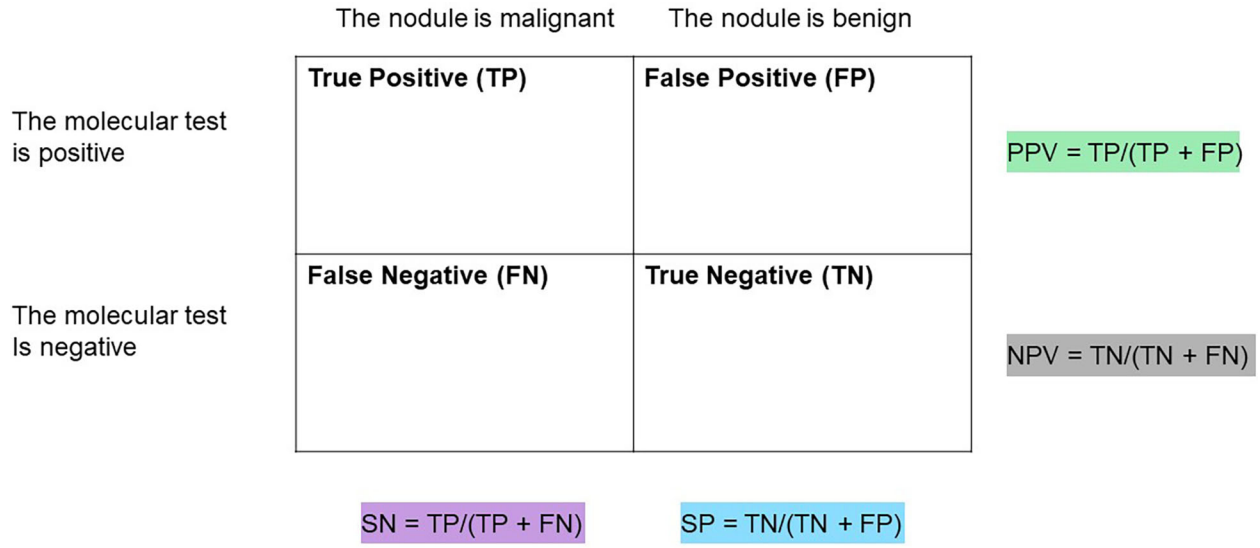


FIGURE 1
Table to assist in calculations of sensitivity (SN), Specificity (SP), positive predictive value (PPV), and negative predictive value (NPV) of a diagnostic test.

FIGURE 2
(A) PPV and NPV for a test with 90% SN and 70% SP at a disease prevalence of 25%. **(B)** PPV and NPV for a test with 90% SN and 70% SP at a disease prevalence of 40%.

center, with blinded central histopathologic review. A prospective validation study reduces clinical decision-making bias regarding who enters the study cohort and who has surgery. A multi-center study with blinded histopathology review confirms the gold standard presence or absence of disease in a broad and representative population which aids in reliable SN and SP calculations. Finally, all patients enrolled in the study must have surgery so the prevalence of thyroid cancer in the studied cohort can be known and utilized to calculate the NPV and PPV.

A brief history of molecular diagnostic laboratory assays

The utilization of molecular diagnostics has rapidly advanced over the last 10-15 years with some older generation tests maintaining a presence for use and others being replaced by next generation sequencing (NGS) platforms. A brief review of older and currently unavailable molecular tests is presented, primarily to provide context for assessing the currently available tests.

The identification of the *BRAF*V600E mutation in papillary thyroid carcinoma (PTC) in 2003 was one of the earliest identified molecular signatures correlating a molecular variant with final histology (17). *BRAF*V600E is a highly specific yet poorly sensitive marker for thyroid cancer, especially in ITN where it is now known that *BRAF*V600E is present in <10% of molecularly tested ITN aspirates (18). Thus, research began into mutation panels that raise test SN to detect more malignant nodules. One of the first studies was a prospective multi-institutional study evaluating *BRAF*V600E, *BRAF*K601E, mutations of *NRAS*, *KRAS* and *HRAS* gene codons as well as *RET/PTC 1/3* rearrangements and *PAX/PPARγ* fusions (19). This panel showed a high specificity with 97% of mutation positive nodules representing histologically malignant tumors yet only a 62% sensitivity as not all malignancies carried variants or fusions detected by the panel.

In 2012, the clinical validation study of the Afirma® Gene Expression Classifier (GEC) was published. The Afirma GEC

combined mRNA expression on a 167-gene microarray platform with machine learning with a goal of predicting benign nodules with ITN cytology to reliably rule out thyroid cancer and avoid unnecessary surgery (10). This was a prospective, multi-center study with blinded central histopathology review and reported a high sensitivity of 90% and a high negative predictive value of 94% [(95% confidence interval (CI)), 87-98)] across BIII and BIV nodules) By virtue of the test design with an emphasis on ruling out thyroid cancer, the specificity and PPV were relatively low. As the first rule-out test, there was caution regarding the possibility of false-negative results among potentially more aggressive cancers. Knowing that the standard treatment of ITN nodules was surgery, a Hurthle cell cassette was included with the GEC to intentionally call most Hurthle samples as GEC suspicious. Resultingly, the overall specificity of the GEC amongst Hurthle cell lesions was only 12% (10, 20). The acceptance and comfort with rule-out testing amongst physicians, the need for a higher benign call rate and PPV, combined with scientific advances and reduced costs of next-generation sequencing prompted development of the Afirma Genomic Sequencing Classifier (GSC) (21).

Thyroseq® has evolved with multiple iterations expanding the number of molecular variants identified from the initial 7 gene panel (targeted variants in 4 genes and 3 gene fusions) to a targeted NGS platform including 12 genes in version 1 to 14 genes analyzed for point mutations and 42 types of gene fusions in version 2 (22, 23). Thyroseq v2 data was published in 2013 and the expanded Thyroseq v2.1 panel data was published in 2015. Thyroseq v2.1 reported test performance was a sensitivity of 90.9% [CI 78.8–100], specificity of 92.1% [CI 86.0–98.2], positive predictive value of 76.9% [CI, 60.7–93.1], and negative predictive value of 97.2% [CI 78.8–100], with an overall accuracy of 91.8% [CI, 86.4–97.3] These earlier versions of Thyroseq NGS panels were not tested in prospective, multi-center studies with blinded histopathologic review (23). As will be described, Thyroseq v3, the current commercially available testing platform, further expanded the number of molecular variants and fusions tested.

Other molecular tests that were used for the preoperative diagnosis of ITN included a combined miRNA and somatic gene mutation panel from Asuragen® (available ~2010-2014) and the micro-RNA (miRNA) classifier RosettaGX® Reveal (available from ~2016-2018) (24, 25). Neither is currently commercially available.

Utilization of molecular diagnostics in clinical practice

The incorporation of thyroid nodule molecular diagnostic testing into clinical practice bears some discussion. Thyroid nodule biopsies can occur in outpatient clinics, pathology departments, radiology suites, and rarely in an inpatient setting. Each practice, institution, and location present opportunities and challenges.

One consideration is whether to utilize a "collect on all" protocol where a sample for molecular marker testing is collected at the time of a thyroid nodule's initial FNA. Alternatively, patients can be asked to return for a repeat FNA for collection of a sample for molecular testing after an indeterminate cytology result. Given most biopsies are read as definitively benign or malignant (approximately 75%), allowing a patient to avoid unnecessary needle passes is reasonable. However, the inconvenience of taking more time from work or away from home, additional copays, and a repeat of the FNA preparation and procedure argues for collecting a molecular marker sample at the time of the initial FNA in the event of an ITN result. Most patients will be in favor of getting all samples collected at once in lieu of returning for a second procedure if given the option. Collecting on all samples does require tracking of specimens, a timely send out of material upon receipt of an ITN result and discarding of unused samples to free up space for future samples. This does require dedicated organization and effort. Currently, the Afirma and Thyroseq testing platforms both allow for centralized cytology diagnosis (at Thyroid Cytology Partners and CBL Path respectively) with reflex send out of collected molecular samples upon an ITN result. ThyGeNEXT/ThyraMIR® (MPTX) offers cytology reads *via* a partnership with Dianon Pathology. In a community practice setting, a transition from decentralized thyroid FNAs in radiology practices, with separate cytology reads at individual centers, to a centralized collection for cytology and molecular markers resulted in a decrease of ITN from 24% to 10% and a reduction in diagnostic surgeries from 24% to 6% (26).

If onsite cytology assessment is available, this may represent the best model. At the time of the FNA, rapid on-site evaluation can be made to determine cytology adequacy, diagnosis, and the need for extra needle passes for dedicated material for molecular testing while a patient is prepped and waits. This practice can reduce nondiagnostic aspirates and improve diagnostic accuracy (27). The logistics of this practice demand an integrated clinic model with enough pathology personnel to create cytology slides and have a rapid read. This is not feasible in many, if not most, clinical settings.

Slide scraping, the collection of thyroid follicular cells from cytology slides with the aid of microscope assisted microdissection, presents a convenient methodology for running some molecular tests on cytology smears when the patient has not had access to molecular diagnostics or there was no collection of a molecular sample at the time of initial FNAC. The Afirma platform does not offer slide scraping while MPTX and Thyroseq do offer this collection method. Though convenient, there are limitations to slide scraping relative to collecting a fresh sample. In the MPTX validation study, 18% of slides failed to provide adequate nucleic acid quantity to run the assay (28). In the Thyroseq validation of slide scraping, Diff-Quik stained smears were inadequate 35% of the time though all Papanicolaou-stained smears were informative (29). Of greater concern than assay failure, are the discordant results between microdissected cytology smears relative to a fresh FNAC placed in its respective nucleic acid protection/storage buffer. There was 11% discordance for miRNA with the ThyraMir portion of MPTX and 14% of copy number alterations along with 17% of fusions were missed (false negatives) on Thyroseq slide scraping compared to a fresh sample from FNAC (29, 30). Clinicians should consider the discussion point regarding the use of slide scraping for Thyroseq, "the collection of a portion of a fresh FNA sample directly into a nucleic acid preservative solution should be

attempted whenever possible because this provides the highest success rate and accuracy of testing" (29).

The role of molecular diagnostics in ITN for benign vs malignant diagnosis

Molecular testing has become a more commonly utilized tool in the clinical setting to help provide additional risk information for ITN. Ideally, the results of the molecular test shift the risk of malignancy (ROM) from ~25% with ITN cytology to risks that help determine which patients will benefit from conservative surveillance versus definitive surgical intervention (31, 32). Molecular testing platforms have evolved with technical advancements coming in the form of expanded genomic information and improved test performance. As of this writing, the three most used molecular tests in the United States include the Afirma Genomic Sequencing Classifier (Afirma GSC), ThyGeNEXT/ThyraMIR (MPTX), and Thyroseqv3 (TSv3). Each molecular test is performed using a different method; however, all three aim to provide the clinician with accurate and precise information concerning patients' risk of nodule malignancy. To our knowledge there is no widespread use of these molecular markers outside of the United States. There is limited use in certain provinces of Canada as well as sporadic use in South America and Europe, almost universally without national healthcare or insurance support.

The Afirma GSC uses next generation RNAseq and whole transcriptome analysis combined with machine learning algorithms to provide a benign or suspicious result in nodules with ITN (21). MPTX is a multiplatform test approach that combines a next generation targeted sequencing panel (ThyGeNEXT) with a microRNA risk classifier test (ThyraMIR) (28). TSv3 is a targeted next generation sequencing test that evaluates point mutations, gene fusions, copy number alterations and abnormal gene expression in 112 thyroid cancer related genes. A high-quality diagnostic test validation study that is prospective, blinded, multi-center and representative of the intended test population is critical to provide confidence in the test performance. Post-validation real-world studies are important for increasing confidence in a test's performance and providing evidence of benefit in clinical practice outside of the controls of a validation study.

MPTX screens samples with the ThyGeNEXT NGS panel that include selected DNA mutations in the following genes: *ALK, BRAF, GNAS, HRAS, KRAS, NRAS, PIK3CA, PTEN, RET* and *TERT* promoter genes. The following gene fusions are detected by analysis of RNA: *ALK, BRAF, NTRK, PPARγ, RET* and *THADA* (28). If there is a strong driver mutation detected, the sample is considered positive. If the sample has a weak oncogenic driver mutation or no mutation, it is further risk stratified using the microRNA classifier (ThyraMir). The initial ThyraMir panel included 5 growth-promoting miRNAs (miR-31, -146, -222, -375, -551) and 5 growth-suppressing miRNAs (miR-29, -138, -139, -155, -204). MPTX results are ultimately reported as one of three categories (negative, moderate, or positive) based on results of the combined ThyGeNEXT mutation panel and ThyraMIR microRNA risk classifier thresholds (28).

The MPTX has been analytically validated and the clinical validation study is a retrospective, blinded multicenter study (28, 33) (Table 1). Unanimous histopathology consensus was not met in 19% of cases which were excluded from analysis. MPTX results for 197 subjects with ITN were categorized as positive, moderate risk or negative for malignancy from a cohort with a 30% disease prevalence. Moderate risk was assigned to 28% of the cohort who are estimated to have the same ROM as the baseline cancer prevalence of 30%. When the moderate risk patients were found to have malignant histology, they were assigned as true positives. When the moderate risk patients were found to have benign histology, they were assigned as true negatives. Thus, the moderate risk groups were categorized in a way that bolsters overall test SN and SP (more true positives or true negatives than defined by the positive or negative groups alone). However, the moderate risk subjects/results were not used in the PPV and NPV calculations. Finally, based on concerns that the proportion of histologic subtypes within the studied cohort were inconsistent with published literature, a prevalence adjustment calculation was made to match the reported proportions of adenomas, malignant subtypes and NIFTP as reported by the TSv3 validation study (28, 34). Bearing these considerations, the results showed 95% SN [CI, 86- 99] and 90% SP [CI, 84-95] for disease. Negative MPTX results ruled out disease with 97% NPV while positive MPTX results ruled in high-risk disease with a 75% PPV. An updated ThyGeNEXT panel improved strong driver mutation detection by 8% with

TABLE 1 Validation study summary of the most used thyroid nodule molecular diagnostic tests in the United States.

	Afirma GSC	Thyroseq v3	ThyGeNEXT/ThyraMIR
Test Type	Whole transcriptome RNA NGS	Targeted DNA and RNA NGS	Targeted NGS + miRNA expression
Validation Study Study Design	Patel et al (20) Prospective	Steward et al (33) Prospective	Lupo et al (27) Retrospective
Sample Size	190	286	178
Specificity	68%	82%	90%
Sensitivity	91%	94%	93%
NPV	96%	97%	95%
PPV	47%	66%	74%

*BRAF*V600E and *TERT* promoters being the most common mutations. Additionally, this newer panel increasingly detected coexisting drivers by 4%, TERT being the most common and often paired with *RAS* (35). A pairwise analysis of miRNA to detect medullary thyroid cancer (MTC) showed 100% accuracy on a study of 4 MTC and 26 non-MTC samples (36). Finally, MPTX has recently been updated with the addition of miR-21 and an interdependent pairwise microRNA expression analysis (MPTXv2). This updated MPTX platform was tested on the same cohort as the original validation study population. The results showed a decrease in the moderate-risk cohort from 28% to 13% ($p < 0.001$) and a reported improvement in PPV to 96% (from 74%) and NPV to 99% (from 95%) (p=NS for both) (37).There have been no completely independent research studies to assess the MPTX performance. In one analysis of pediatric lesions comprised of 66 malignant and 47 benign tumors, MPTXv1, analysis performed with 70% SN and 96% SP (38).

The Afirma GSC samples are initially tested for RNA quantity and quality. Sufficient samples are tested against initial classifiers to detect parathyroid tissue, MTC, *BRAF*V600E variants and *RET/PTC*1 and *RET/PTC*3 fusions. Recently, the validation of the MTC classifier of the Afirma GSC showed 100% SN and 100% SP in a cohort of 21 MTC and 190 non-MTC lesions (39). If all the classifiers are negative and there is adequate follicular content, the GSC ensemble model relies heavily on differential gene expression of > 10,000 genes for sample classification of GSC-B or GSC-S results. The Afirma GSC clinical validation study was based on a cohort of ITN samples collected prospectively from multiple community and academic centers from the Afirma GEC validation (10). All patients underwent surgery without known genomic information and all samples were assigned a histopathology diagnosis by an expert panel blinded to all genomic information. The results showed (at a 24% cancer prevalence): SN - 91% [CI, 79-98], SP - 68% [CI, 60-76], NPV - 96% [CI, 90-99], PPV - 47% [CI, 36-58] (21) (Table 1). Since the validation study, 14 independent real-world studies have been published and in aggregate show a significant improvement in performance over the Afirma GEC, primarily with improved specificity and a higher benign call rate (BCR) of 65% (as compared to 54% with the Afirma GEC) (40–54). As expected, some of these studies have also demonstrated that the implementation of Afirma GSC reduced the rate of surgical intervention by 45-68% (40, 43). A meta-analysis by Vuong et al. including seven studies comparing the performance of Afirma GEC to GSC and found that GSC had a higher BCR (65.3% vs 43.8%; P <0.001), a lower resection rate (26.8% vs 50.1%; P <0.001), and a higher risk of malignancy (60.1% vs 37.6%; P <0.001) in resected specimens (55) (Table 2).

The Afirma GSC incorporates Hurthle/oncocytic and neoplasm classifiers to enhance the diagnostic accuracy in predominately oncocytic ITN relative to the Afirma GEC (20). A review of four independent post validation studies of the Afirma GSC performance in oncocytic cell lesions showed maintenance of a high SN (3 with 100% SN and one with 80% SN) and improved SP (81-100% for GSC compared to 29-43% for GEC) (56). When compared to the GEC, the BCR for oncocytic cell–predominant nodules by the GSC is significantly elevated (73.7% vs 21.4%; P < 0.001) (55).

TSv3 is a genomic classifier (GC) where a value is assigned to each detected genetic alteration based on the strength of association with malignancy: 0 (no association), 1 (low cancer risk), or 2 (high cancer risk). A GC score calculated for each sample is a sum of individual values of all detected alterations, with GC scores 0 and 1 accepted as test negative (score 1 is commercially reported as "currently negative") and scores 2 and above as test positive (57). "Currently negative", low cancer probability alterations, are included in the BCR in TSv3 studies. The clinical validation study for TSv3 by Steward et al. was a prospective, multi-center, blinded study that ultimately analyzed 257 ITN, all with histologic consensus. The test demonstrated a 94% [CI, 86%-98%] SN and 82% [CI, 75%-87%] SP. With a cancer/NIFTP prevalence of 28%, the NPV was 97% [CI, 93%-99%] and PPV was 66% [CI, 56%-75%] (34) (Table 1). There have been 10 independent studies assessing the performance of TSv3 (46, 47, 58–65). A recent meta-analysis by Lee at al. including six studies (total 530 thyroid nodules) evaluating the performance of TSv3 found a similar sensitivity of 95.1% [CI, 91.1–97.4%] but a lower specificity of 49.6% [CI, 29.3–70.1%] when compared to the original validation study; the reported PPV of 70% [CI, 55–83%], and NPV of 92% [CI, 86–97%] remained comparable (66) (Table 2).

Molecular tests can be classified as "rule in" vs "rule out" based on their ability to confirm or exclude malignancy. Vargas-Salas et al. found that with a thyroid cancer prevalence of 20–40%, a robust "rule-out" test requires a minimum NPV of 94% and a minimum

TABLE 2 Meta-analysis data of Afirma GSC and Thyroseq v3.

	Afirma GSC	Afirma GSC	Thyroseq v3
Meta-analysis	Vuong et al (52)	Lee et al (53)	Lee et al (53)
# Included Studies	7 studies	7 studies	6 studies
Sample Size	807	472	530
Specificity	43%	53%	50%
Sensitivity	94.3%	96%	95%
NPV	90%	96%	92%
PPV	63.1%	63%	70%

sensitivity of 90%, whereas to "rule- in" malignancy, a test requires a PPV of at least 60% and a specificity above 80% (67). MPTX, Afirma GSC, and TSv3 all perform well as "rule out" tests for ITN based on their relatively high sensitivities and NPVs, though independent confirmation of MPTX performance is lacking. MPTX has too few studies to compare its performance to other molecular testing platforms and future studies are needed to confirm its clinical efficacy. A study by Silaghi et al. comparing the performance of Afirma GSC and TSv3 found TSv3 to have the best overall diagnostic performance with the lowest negative likelihood ratio (NLR 0.02), followed by Afirma GSC (NLR 0.11). Both TSv3 and Afirma GSC achieved optimal results to exclude malignancy; however, both failed to achieve a higher performance to confirm or "rule in" a malignancy when compared to their predecessor Thyroseqv2 (68). Similarly, Lee et al. found there was no statistically significant difference in diagnostic performances between the Afirma GSC and TSv3 (66) (Table 2). Finally, Livhits et al. performed a randomized clinical trial by using Afirma GSC or TSv3 in routine clinical practice on a rotating monthly basis. They found that both Afirma GSC and TSv3 have a relatively similar specificity (80% and 85%, respectively), and both allowed approximately 49% of patients with indeterminate nodules to avoid diagnostic surgery (46). Given the similar performance, it is no longer accurate to call Afirma a "rule out test" and Thyroseq a "rule in test" as they have been commonly described with earlier iterations of the testing platforms and in a recent review (69).

The role of molecular genetic testing in predicting thyroid cancer prognosis

Molecular genetic testing is a valuable tool in understanding patients' prognosis based on specific mutations detected in thyroid cancer. Various mutations are associated with increased tumor aggressiveness, metastatic lymph node spread, a tendency to de-differentiate, and/or reduced efficiency of radioiodine treatment. The main known genetic causes of thyroid cancer include point mutations in the *BRAF, RAS, TERT* promoter, *RET*, and *TP53* genes and the fusion genes *RET/PTC*, *PAX8/PPARγ*, and *NTRK* (70). Molecular genetic testing of thyroid tissue in the preoperative and/or postoperative period is becoming more common, and therefore detection of genetic changes may serve as a prognostic factor that can help determine the extent of surgical treatment and the use of systemic targeted therapy. The characterization of molecular variants and fusions as *BRAF*-like, *RAS*-like, and non-*BRAF* non-*RAS*-like has helped to group molecular alterations in thyroid cancer that share similar risk of events such as extra-thyroidal extension and lymph node metastases (71, 72). For example, a retrospective analysis by Tang et al. associating pathologic features to the aforementioned molecular classes showed a statistically higher rate of T4 tumor size and N1b nodal metastases in *BRAF*-like mutated tumors (22%) compared to the other classes (≤ 6%) amongst other more aggressive findings (12). Afirma GSC, MPTX, and TSv3, have shown promise in predicting disease recurrence in thyroid cancers and Bethesda V/VI nodules based on the detection of low-risk vs high-risk genetic mutations.

In thyroid nodules with Afirma GSC suspicious results, or thyroid nodules with BV or BVI cytology, Afirma Xpression Atlas (XA) can provide more granular molecular information. The analytical and clinical validation of XA, which identifies thyroid nodule molecular variants and fusions by whole transcriptome sequencing, was published in 2019 (73). In 2020, the panel was expanded to detect molecular alterations in 593 genes allowing XA to report on 905 variants and 235 fusions. Afirma XA results may offer important prognostic insights; for example, nodules with a non-*RAS* and non- *BRAF* molecular profile have lower rates of lymph node metastasis and extrathyroidal extension (74). A large retrospective study by Hu et al. demonstrated that 44% of Bethesda III/IV Afirma GSC-S and most Bethesda V/VI nodules (87% BVI) had at least one genomic variant or fusion identified, which could optimize individual treatment decisions (18). The ability of Afirma XA to demonstrate improved clinical outcomes based on surgery and mutational status is yet to be determined as no randomized trials have been performed; however, the genomic insights provided by XA may predict tumor aggressiveness and provide important information regarding variants for targeted therapy (75).

Labourier et al. found that in a systematic review of the literature, 70%-75% of malignant/Bethesda VI cytology were expected to be positive for the oncogenic *BRAF*V600E substitution with the second most frequent gene alteration being *TERT* promoter mutations (11%) (76). High frequency of oncogenic *BRAF* mutations has important clinical implications and multiple studies have shown that *BRAF*V600E correlates with aggressive features of thyroid cancer such as extrathyroidal extensions, vascular invasion, larger thyroid nodule size, advanced staging, lymph node metastasis and recurrence (77). Additionally, *TERT* promoter mutations are among the most recognized markers associated with aggressive thyroid cancer phenotypes (77).

When specifically evaluating the performance of TSv3 in thyroid nodules with Bethesda V (suspicious for malignancy) cytology, Skaugen et al. found that TSv3 had sensitivity of 89.6% (95% CI, 82.4%- 94.1%) and specificity of 77.3% (95% CI, 56.6%-89.9%). Moreover, when TSv3 positive Bethesda V nodules were sorted into molecular risk groups (low, intermediate, high), disease recurrence was more commonly found in the high-risk group whereas no patients in the low-risk group developed recurrence (78). Another study by Hescot et al. used TSv3 to determine if there were molecular prognostic factors associated with recurrence and overall survival in patients' with poorly differentiated thyroid carcinomas (PDTCs). Of the 40 patients tested with TSv3, high-risk molecular signatures (*TERT*, *TP53* mutations) were found in 24 cases (60%), intermediate-risk signature in 9 cases (22.5%) and low-risk signature in 7 cases (17.5%) with potentially actionable mutations that may be amenable to targeted therapy identified in 10% of cases. Furthermore, the high molecular-risk signature was associated with distant disease metastasis ($P = 0.007$) and with worse overall survival ($P = 0.01$), whereas none of the patients with low-risk molecular signature died due to thyroid cancer (79).

It is important to note that there are no established guidelines addressing management decisions based on the detection of most genetic alterations detected in thyroid nodules regardless of cytology category. In ITN, the most studied value is in the

diagnosis of benignity or malignancy. The value of knowing the molecular alterations in BV and BVI thyroid nodules has yet to be investigated in prospective multi-center studies. Additionally, molecular tests performance metrics are generally assessed independent of other clinically relevant factors such as family history of thyroid cancer, heritable syndromes, radiation exposure, and thyroid ultrasound features. One area of increasing interest is the identification of aggressive thyroid cancers that may be amenable to future systemic targeted therapies as needed, possibly in the neo-adjuvant setting.

Molecular identification of targetable alterations in thyroid cancer

While the use of molecular testing to risk-stratify indeterminate thyroid nodules is encouraging, arguably the most exciting use of this technology is in the setting of advanced and aggressive thyroid disease where identification of targetable mutations can have significant clinical impact (Table 3). In differentiated thyroid cancer, the overall mortality is low, however 15% of cases will be locally invasive and in those with distant metastases which are radioioine (RAI)-refractory, the 10-year overall survival is <50% (86, 87). Conversely, the most aggressive subtypes of thyroid cancer, medullary, poorly differentiated, and anaplastic, have high disease-specific mortality. Especially in these thyroid cancer subsets with high mortality rates, there has been substantial expansion of the therapeutic armamentarium with tumor genome-directed therapies over the past decade (80, 88–90). Studies have identified several targetable (or potentially targetable) alterations in advanced thyroid cancer, including mutations in commonly detected genes such *BRAF*V600E, *RET*, *PIK3CA*, as well as gene fusions including *RET*, *NTRK*, and *ALK*. In addition to therapies targeting specific genetic alterations, immunotherapy shows significant promise in treating tumors with microsatellite instability, high tumor mutational burden (TMB), and high PD-L 1 expression. With the possibility of identifying genomic alterations *via* NGS in advanced thyroid cancers, the study of neoadjuvant therapy for aggressive disease has just begun.

Recently, a multidisciplinary, multi-institutional, multi-national consensus statement was jointly published by the American Head and Neck Society (AHNS) and the International Thyroid Oncology Group (ITOG) defining advanced thyroid cancer and its targeted treatment (91). The group advocates for molecular testing to be "performed in Clinical Laboratory Improvement Amendments (CLIA)- accredited laboratories (or their international equivalent), on appropriate specimens, using clinically validated procedures, which may include laboratory-developed tests or FDA-approved commercial assays" (82). With the support of high-quality evidence, the consensus recommends that "when somatic mutational testing is performed for thyroid cancer, multiplexed NGS-based panels are superior to multiple single-gene tests" and that, "NGS panels that include assays for gene fusions are preferred given the ability to detect multiple mutations and fusions in one assay thereby conserving tissue and limiting expense" (80).

Differentiated Thyroid Cancer (DTC): Accounting for roughly 95% of thyroid cancers, DTC arises from follicular thyroid cells and is often RAI-avid. This allows the vast majority of DTC to be treated with surgery alone for smaller tumors or surgery with RAI and levothyroxine suppression therapy for more advanced or aggressive disease. However, it is reported that 7–23% of patients with DTC will develop distant metastases, and two-thirds of patients with distant metastases become RAI-refractory (86, 90). These patients have poor prognosis with overall 10-year survival of <50% (86, 87). Multicenter, randomized, double-blind, placebo-controlled, phase III studies led to FDA approval of multi-kinase inhibitors (MKIs) Sorafenib and Lenvatinib, for the treatment of RAI-refractory locally advanced (non-operative) or metastatic DTC (80, 90). MKIs block activation of several key receptors that regulate thyroid cancer progression including angiogenesis. While studies showed progression free survival (PFS) benefit in the treatment groups compared to placebo groups (80, 90), because of the non-specific targeting of these drugs, their clinical utility is limited by their substantial toxicity profiles.

In the last decade, recognition of important molecular drivers and signaling pathways has led to the development of molecular-targeted therapies especially for advanced and RAI-refractory differentiated thyroid cancer. Presence of a *BRAF* V600E mutation, the most common driver mutation in the spectrum of follicular cell derived thyroid cancers, can confer susceptibility to selective RAF kinase inhibitors in some cancer lineages. The combination of dabrafenib (*BRAF* inhibitor) and trametinib (*MEK* inhibitor), which was initially FDA-approved in *BRAF*V600E mutated ATC, has also been studied in *BRAF*-

TABLE 3 FDA approved molecularly targeted therapies in thyroid cancer.

Dabrafenib (80) (BRAF inhibitor) & Trametinib (80) (MEK inhibitor)	Larotrectinib (81, 82) (Selective TRK inhibitor)	Entrectinib (83) (multi-kinase inhibitor NTRK1/2/3, ROS1, & ALK)	Selpercatinib (84) (Selective RET kinase inhibitor)	Pralsetinib (85) (Tyrosine Kinase Inhibitor)
• 18 years old • locally advanced, unresectable or metastatic solid tumors • ***BRAFV600E mutant-positive***	• 1 month old • locally advanced or metastatic solid tumors • Tumor agnostic • ***NTRK fusion-positive***	• 18 years old • locally advanced or metastatic solid tumors • penetrate blood-brain barrier • Tumor agnostic • ***NTRK fusion-positive***	• 12 years old • *RET*-driven advanced or metastatic cancer • ***RET mutant-positive*** Medullary Thyroid Cancer • ***RET fusion-positive*** radioactive iodine-refractory thyroid cancers	• 12 years old • *RET*-driven advanced or metastatic cancer • ***RET mutant-positive*** Medullary Thyroid Cancer • ***RET fusion-positive*** radioactive iodine-refractory thyroid cancers

mutated PTC with high response rates (50% single-agent dabrafenib vs. 54% combination, modified RECIST criteria) and median progression free survival 11.4 vs. 15.1 months. This combination of drugs recently garnered approval for treatment of *BRAF*-mutated DTC (83). The FDA-approved drugs selpercatinib and pralsetinib target the oncogenic *RET* gene fusions, detected in approximately 10% of PTC (81, 92). Thyroid cancers harboring genetic rearrangements involving *NTRK*1/3 (~2% of PTC) can respond to treatment with TRK inhibitors, including FDA-approved larotrectinib and entrectinib (72, 93–95). *ALK* fusions are still more rare in well differentiated thyroid cancers (<1% of PTC) but are identified more frequently in PDTC. *ALK*-inhibitors are FDA-approved for solid tumors that harbor *ALK* fusions and a few patients with thyroid cancer have been included in the reported clinical trials and/or case reports, although no ALK-inhibitors are currently FDA-approved for DTC specifically. Therefore, *ALK* fusion testing is currently indicated for advanced DTC only in the context of either "off-label" treatment or clinical trials. Lastly, while microsatellite instability (MSI) and TMB in DTC are often low, MSI-high or TMB-high cancers, may be eligible for treatment with pembrolizumab, a programmed death-1 (PD-1) inhibitor, given the its tissue agnostic approval for MSI-high cancers and the demonstrated responses of TMB-high solid tumors (96, 97).

Anaplastic Thyroid Cancer (ATC), with a median overall survival of 4 months, is considered one of the most aggressive and lethal malignancies and typically presents at a median age of 65-70 years (95–97). This most-aggressive thyroid cancer, with a 6-month OS of 35%, and disease-specific mortality approaching 100% is responsible for over half of the annual thyroid cancer-related deaths despite comprising only 1.5% of all thyroid cancers (98–100). These outcomes are despite aggressive multimodality treatment regimens including surgery (when feasible), traditional cytotoxic chemotherapy and radiation therapy. ATC is postulated to have the potential to arise either *de novo* or from pre-existing DTC. The coexistence of *BRAF*-mutated ATC with PTC described in several studies, suggests the potential of a common DTC origin for most of these tumors (101, 102). ATC has a higher relative tumor mutational burden (TMB) than DTC although overall the TMB for ATC is still lower than many other solid malignancies (100). The mutational profile of ATC tends to include accumulation of variations in tumor suppressor genes such as *TP53 and PTEN*; oncogenes such as *TERT* promoter, *RAS*, *BRAF*, and *PIK3CA*; oncogene-fusions such as *NTRK*, *RET*, and *ALK*; or through mismatch repairs (103). Given the aggressive nature of ATC, most often with surgically unresectable disease at presentation, and resistance to radioactive iodine, chemotherapies, and radiation therapy, all patients with suspected ATC are recommended to undergo expeditious histological confirmation, staging, and molecular testing and if a targetable mutation is identified, treatment should include directed therapies against this actionable target.

The most significant shift in the management of ATC to occur in decades was the afore mentioned combinatorial use of *BRAF/MEK* inhibitors (dabrafenib/trametinib) in ATC patients harboring a *BRAF*V600E mutation (83). Due to the potential for long turn-around times for traditional NGS testing, some centers employ a rapid PCR assay to detect *BRAF*V600E in DNA isolated from paraffin blocks (48–72-hour turnaround) or use peripheral blood NGS (cell-free DNA) which has sensitivity of 75%–90% and turnaround time of 3–7 days. These options may enable slightly earlier initiation of targeted therapies if they exist (104, 105). Mutation-specific immunohistochemistry for *BRAF*V600E can also be useful in expeditiously identifying patients who might benefit from approved targeted therapy, but requires substantial tissue *via* core needle biopsy, FNA cell block, or even surgical specimen due to the potential for false positives (106). When successful, *BRAF*-directed therapy can induce rapid and substantial disease regression and may eventually render previously inoperable disease amenable for surgical resection (107). For these patients with advanced stage ATC who are able to undergo complete locoregional surgical resection, one study has shown some of the highest survival rates ever reported for this disease with a 94% 1-year survival and an unmet median OS in a cohort of 20 patients (8 of 20 having stage IVC disease) having received BRAF-directed therapy followed by surgery (98).

Medullary Thyroid Cancer (MTC) arises from parafollicular C cells which are neuroendocrine in origin and accounts for about 2% of thyroid cancers. Although rare, MTC accounts for about 14% of annual deaths from thyroid cancer (108–110). MTC most often occurs sporadically (80%) with hereditary forms (20%) being associated with the multiple endocrine neoplasia (MEN) type 2 syndromes. These inherited forms of MTC are associated with genomic alterations of the *RET* proto-oncogene and are inherited in an autosomal dominant fashion. Patients diagnosed with MTC, regardless of disease stage, personal history of other endocrinologic disorder, or family history, should have genetic counseling and be tested for germline *RET* mutations (91). About 6% of MTC patients with no family history or other endocrinologic disorder to suggest MEN, are found to harbor a germline *RET* mutation prompting counseling and testing of family members. Somatic *RET* mutations are also found in approximately 50% of patients with sporadic MTC. Somatic mutations in *HRAS* (~25%), *KRAS*, and rarely *NRAS* genes, which are canonically mutually exclusive with *RET* mutations, have also been identified in sporadic MTC (111). About 20% of sporadic MTC harbor neither *RET* nor *RAS* gene alterations (112). Patients with advanced sporadic MTC should be offered molecular testing since somatic *RET* mutations have been shown to lead to more aggressive disease, including higher T- and N-stage, and increase the rate of distant metastasis (84, 108).

Currently, two MKIs, vandetanib and cabozantinib, are approved by the U.S. FDA for the systemic treatment of MTC and show improvement in progression-free survival (78, 79), both MKIs have a narrow therapeutic window and off-target kinase inhibition causes significant toxicities. Additionally, MTC can acquire gatekeeper resistance mutations at RET codon V808 rendering these therapies ineffective (91). Recently however, selective *RET* inhibitors have shown both promising efficacy and more favorable toxicity profiles (85). Selpercatinib (LOXO-292) is a selective *RET* kinase inhibitor potently effective against *RET* alterations, including gene fusions, oncogenic mutations, and even the V804 gatekeeper mutation. Early data from LIBRETTO-001, the phase I/II study of selpercatinib, showed 56% of patients with *RET*-mutant MTC previously treated with vandetanib and/or

cabozantinib achieved objective responses with mostly grade 1 or 2 adverse events, prompting early approval by the FDA (113). Currently, an ongoing randomized trial is evaluating treatment-naïve patients with *RET*-mutant MTC, comparing selpercatinib with standard MKI therapy. Pralsetinib (BLU-667), another selective *RET* inhibitor, has been recently approved by the FDA for the treatment of patients with advanced or metastatic *RET*-mutant MTC (IC50 0.3–5 nM). This approval was based on early data from the phase I/II trial (ARROW) of pralsetinib showing a 65% objective response rate in patients with *RET*- mutant tumors, including patients with MKI resistant tumors and with known gatekeeper mutations (84). In this study, pralsetinib has been well tolerated with most treatment related adverse events being low grade and reversible (114).

In summary, the use of molecular testing in the identification of therapeutic targets can have significant clinical impact. We are undoubtedly only seeing the beginning of this new frontier. Knowledge of molecular mutations, fusions, and gene expression profiles, especially for the most advanced and aggressive forms of thyroid cancer will likely continue to drive drug discovery and development world-wide.

Summary

Molecular testing of thyroid nodules and thyroid cancer has improved the diagnostic accuracy of indeterminate thyroid nodules and provides actionable information regarding tumor prognosis. Additionally, identifiable molecular variants and fusions inform clinicians of a patient's eligibility for targeted systemic therapies in the important subset of thyroid cancer patients with metastatic, progressive, radio-iodine refractory disease. Future research should focus on the clinical utility of molecular information to change the clinical approach to patients with thyroid nodules. For example, prospective studies on the extent of surgery and the assessment of changes in factors such as tumor recurrence. Additionally, novel analyses to predict tumor behavior are warranted. Finally, the investigation of targeted therapies in the neo-adjuvant setting for thyroid cancer that presents aggressively is ongoing and may improve overall outcomes, for example, with improved opportunities for acceptable surgical outcomes in previously unresectable tumors.

Author contributions

JP, JK, and EC wrote equal portions of the first draft of this review. All authors contributed to the article and approved the submitted version.

Conflict of interest

JK is an employee and equity owner of Veracyte, Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be constructed as a potential conflict of interest.

Publisher's note

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin* (2022) 72:7–33. doi: 10.3322/caac.21708

2. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 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;* (2015) 26:1–133. doi: 10.1089/thy.2015.0020

3. Grant EG, Tessler FN, Hoang JK, Langer JE, Beland MD, Berland LL, et al. Thyroid ultrasound reporting lexicon: White paper of the ACR thyroid imaging, reporting and data system (TIRADS) committee. *J Am Coll Radiol* (2015) 12:1272–9. doi: 10.1016/j.jacr.2015.07.011

4. Cibas ES, Ali SZ. The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid* (2017) 27:1341–6. doi: 10.1089/thy.2017.0500

5. Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda system for reporting thyroid cytopathology: A meta-analysis. *Acta Cytologica* (2012) 56:333–9. doi: 10.1159/000339959

6. Baloch ZW, Fleisher S, LiVolsi VA, Gupta PK. Diagnosis of "follicular neoplasm": A gray zone in thyroid fine-needle aspiration cytology. *Diagn Cytopathol* (2002) 26:41–4. doi: 10.1002/dc.10043

7. American Thyroid Association Guidelines Taskforce on Thyroid N, Differentiated Thyroid C, Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, et al. Revised American thyroid association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid* (2009) 19:1167–214. doi: 10.1089/thy.2009.0110

8. Wilson M, Patel A, Goldner W, Baker J, Sayed Z, Finger AL, et al. Postoperative thyroid hormone supplementation rates following thyroid lobectomy. *Am J Surg* (2021) 221:804–8. doi: 10.1016/j.amjsurg.2020.07.001

9. Schumm MA, Lechner MG, Shu ML, Ochoa JE, Kim J, Tseng CH, et al. Frequency of thyroid hormone replacement after lobectomy for differentiated thyroid cancer. *Endocrine Practice: Off J Am Coll Endocrinol Am Assoc Clin Endocrinologists* (2021) 27:691–7. doi: 10.1016/j.eprac.2021.01.004

10. Alexander EK, Kennedy GC, Baloch ZW, Cibas ES, Chudova D, Diggans J, et al. Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. *New Engl J Med* (2012) 367:705–15. doi: 10.1056/NEJMoa1203208

11. Chudova D, Wilde JI, Wang ET, Wang H, Rabbee N, Egidio CM, et al. Molecular classification of thyroid nodules using high-dimensionality genomic data. *J Clin Endocrinol Metab* (2010) 95:5296–304. doi: 10.1210/jc.2010-1087

12. Tang AL, Kloos RT, Aunins B, Holm TM, Roth MY, Yeh MW, et al. Pathologic features associated with molecular subtypes for well-differentiated thyroid cancer. *Endocrine Pract* (2020) 27(3):206–11. doi: 10.1016/j.eprac.2020.09.003

13. Leeflang MM, Bossuyt PM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. *J Clin Epidemiol* (2009) 62:5–12. doi: 10.1016/j.jclinepi.2008.04.007

14. Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. *CMAJ* (2013) 185:E537–544. doi: 10.1503/cmaj.121286

15. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol* (2008) 56:45–50. doi: 10.4103/0301-4738.37595

16. Eusebi P. Diagnostic accuracy measures. *Cerebrovasc Dis* (2013) 36:267–72. doi: 10.1159/000353863

17. Cohen Y, Xing M, Mambo E, Guo Z, Wu G, Trink B, et al. BRAF mutation in papillary thyroid carcinoma. *J Natl Cancer Institute* (2003) 95:625–7. doi: 10.1093/jnci/95.8.625

18. Hu MI, Waguespack SG, Dosiou C, Ladenson PW, Livhits MJ, Wirth LJ, et al. Afirma genomic sequencing classifier & xpression atlas molecular findings in consecutive Bethesda III-VI thyroid nodules. *J Clin Endocrinol Metab* (2021) 106 (8):2198–207. doi: 10.1210/clinem/dgab304

19. Nikiforov YE, Steward DL, Robinson-Smith TM, Haugen BR, Klopper JP, Zhu Z, et al. Molecular testing for mutations in improving the fine-needle aspiration diagnosis of thyroid nodules. *J Clin Endocrinol Metab* (2009) 94:2092–8. doi: 10.1210/jc.2009-0247

20. Hao Y, Duh QY, Kloos RT, Babiarz J, Harrell RM, Traweek ST, et al. Identification of hurthle cell cancers: solving a clinical challenge with genomic sequencing and a trio of machine learning algorithms. *BMC Syst Biol* (2019) 13:27. doi: 10.1186/s12918-019-0693-z

21. Patel KN, Angell TE, Babiarz J, Barth NM, Blevins T, Duh QY, et al. Performance of a genomic sequencing classifier for the preoperative diagnosis of cytologically indeterminate thyroid nodules. *JAMA Surg* (2018) 153:817–24. doi: 10.1001/jamasurg.2018.1153

22. Nikiforova MN, Wald AI, Roy S, Durso MB, Nikiforov YE. Targeted next-generation sequencing panel (ThyroSeq) for detection of mutations in thyroid cancer. *J Clin Endocrinol Metab* (2013) 98:E1852–1860. doi: 10.1210/jc.2013-2292

23. Nikiforov YE, Carty SE, Chiosea SI, Coyne C, Duvvuri U, Ferris RL, et al. 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) 25:1217–23. doi: 10.1089/thy.2015.0305

24. Benjamin H, Schnitzer-Perlman T, Shtabsky A, VandenBussche CJ, Ali SZ, Kolar Z, et al. Analytical validity of a microRNA-based assay for diagnosing indeterminate thyroid FNA smears from routinely prepared cytology slides. *Cancer Cytopathol* (2016) 124:711–21. doi: 10.1002/cncy.21731

25. Wylie D, Beaudenon-Huibregtse S, Haynes BC, Giordano TJ, Labourier E. Molecular classification of thyroid lesions by combined testing for miRNA gene expression and somatic gene alterations. *J Pathol Clin Res* (2016) 2:93–103. doi: 10.1002/cjp2.38

26. Dhingra JK. Office-based ultrasound-guided FNA with molecular testing for thyroid nodules. *Otolaryngology–head Neck Surgery: Off J Am Acad Otolaryngology-Head Neck Surg* (2016) 155:564–7. doi: 10.1177/0194599816652378

27. Muri R, Trippel M, Borner U, Weidner S, Trepp R. The impact of rapid on-site evaluation on the quality and diagnostic value of thyroid nodule fine-needle aspirations. *Thyroid* (2022) 32:667–74. doi: 10.1089/thy.2021.0551

28. Lupo MA, Walts AE, Sistrunk JW, Giordano TJ, Sadow PM, Massoll N, et al. Multiplatform molecular test performance in indeterminate thyroid nodules. *Diagn Cytopathol* (2020) 48:1254–64. doi: 10.1002/dc.24564

29. Nikiforova MN, Lepe M, Tolino LA, Miller ME, Ohori NP, Wald AI, et al. Thyroid cytology smear slides: An untapped resource for ThyroSeq testing. *Cancer Cytopathol* (2021) 129:33–42. doi: 10.1002/cncy.22331

30. Kumar G, Timmaraju VA, Song-Yang JW, Repko B, Narick C, Mireskandari A, et al. Utility of microdissected cytology smears for molecular analysis of thyroid malignancy. *Diagn Cytopathol* (2019) 47:289–96. doi: 10.1002/dc.24100

31. Ohori NP. A decade into thyroid molecular testing: Where do we stand? *J Am Soc Cytopathol* (2022) 11:59–61. doi: 10.1016/j.jasc.2021.11.002

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) 6:115–29. doi: 10.1159/000468519

33. Banizs AB, Silverman JF. The utility of combined mutation analysis and microRNA classification in reclassifying cancer risk of cytologically indeterminate thyroid nodules. *Diagn Cytopathol* (2019) 47:268–74. doi: 10.1002/dc.24087

34. Steward DL, Carty SE, Sippel RS, Yang SP, Sosa JA, Sipos JA, et al. Performance of a multigene genomic classifier in thyroid nodules with indeterminate cytology: A prospective blinded multicenter study. *JAMA Oncol* (2019) 5:204–12. doi: 10.1001/jamaoncol.2018.4616

35. Jackson S, Kumar G, Banizs AB, Toney N, Silverman JF, Narick CM, et al. Incremental utility of expanded mutation panel when used in combination with microRNA classification in indeterminate thyroid nodules. *Diagn Cytopathol* (2020) 48:43–52. doi: 10.1002/dc.24328

36. Ciarletto AM, Narick C, Malchoff CD, Massoll NA, Labourier E, Haugh K, et al. Analytical and clinical validation of pairwise microRNA expression analysis to identify medullary thyroid cancer in thyroid fine-needle aspiration samples. *Cancer Cytopathol* (2021) 129:239–49. doi: 10.1002/cncy.22365

37. Finkelstein SD, Sistrunk JW, Malchoff C, Thompson DV, Kumar G, Timmaraju VA, et al. A retrospective evaluation of the diagnostic performance of an interdependent pairwise MicroRNA expression analysis with a mutation panel in indeterminate thyroid nodules. *Thyroid* (2022) 32(11):1362–71. doi: 10.1089/thy.2022.0124

38. Franco AT, Labourier E, Ablordeppey KK, Surrey LF, Mostoufi-Moab S, Isaza A, et al. miRNA expression can classify pediatric thyroid lesions and increases the diagnostic yield of mutation testing. *Pediatr Blood Cancer* (2020) 67:e28276. doi: 10.1002/pbc.28276

39. Randolph GW, Sosa JA, Hao Y, Angell TE, Shonka DcJr., LiVolsi VA, et al. Preoperative identification of medullary thyroid carcinoma (MTC): Clinical validation of the afirma MTC RNA-sequencing classifier. *Thyroid* (2022) 32:1069–76. doi: 10.1089/thy.2022.0189

40. Harrell RM, Eyerly-Webb SA, Golding AC, Edwards CM, Bimston DN. Statistical comparison of afirma gsc and afirma gec outcomes in a community endocrine surgical practice: Early findings. *Endocrine Practice: Off J Am Coll Endocrinol Am Assoc Clin Endocrinologists* (2019) 25:161–4. doi: 10.4158/EP-2018-0395

41. Angell TE, Heller HT, Cibas ES, Barletta JA, Kim MI, Krane JF, et al. Independent comparison of the afirma genomic sequencing classifier and gene expression classifier for cytologically indeterminate thyroid nodules. *Thyroid* (2019) 29:650–6. doi: 10.1089/thy.2018.0726

42. San Martin VT, Lawrence L, Bena J, Madhun NZ, Berber E, Elsheikh TM, et al. Real-world comparison of afirma GEC and GSC for the assessment of cytologically indeterminate thyroid nodules. *J Clin Endocrinol Metab* (2020) 105. doi: 10.1210/clinem/dgz099

43. Endo M, Nabhan F, Porter K, Roll K, Shirley LA, Azaryan I, et al. Afirma gene sequencing classifier compared with gene expression classifier in indeterminate thyroid nodules. *Thyroid* (2019) 29:1115–24. doi: 10.1089/thy.2018.0733

44. Andrioli M, Carocci S, Alessandrini S, Amini M, Van Doorne D, Pace D, et al. Testing for afirma in thyroid nodules with high-risk indeterminate cytology (TIR3B): First Italian experience. *Endocrine Pathol* (2020) 31:46–51. doi: 10.1007/s12022-020-09604-7

45. Geng Y, Aguilar-Jakthong JS, Moatamed NA. Comparison of afirma gene expression classifier with gene sequencing classifier in indeterminate thyroid nodules: A single-institutional experience. *Cytopathol: Off J Br Soc Clin Cytol* (2021) 32:187–91. doi: 10.1111/cyt.12920

46. Livhits MJ, Zhu CY, Kuo EJ, Nguyen DT, Kim J, Tseng CH, et al. Effectiveness of molecular testing techniques for diagnosis of indeterminate thyroid nodules: A randomized clinical trial. *JAMA Oncol* (2021) 7:70–7. doi: 10.1001/jamaoncol.2020.5935

47. Gortakowski M, Feghali K, Osakwe I. Single institution experience with afirma and thyroseq testing in indeterminate thyroid nodules. *Thyroid* (2021) 31:1376–82. doi: 10.1089/thy.2020.0801

48. Zhang L, Smola B, Lew M, Pang J, Cantley R, Pantanowitz L, et al. Performance of afirma genomic sequencing classifier vs gene expression classifier in Bethesda category III thyroid nodules: An institutional experience. *Diagn Cytopathol* (2021) 49:921–7. doi: 10.1002/dc.24765

49. Yang Z, Zhang T, Layfield L, Esebua M. Performance of afirma gene sequencing classifier versus gene expression classifier in thyroid nodules with indeterminate cytology. *J Am Soc Cytopathol* (2022) 11:74–8. doi: 10.1016/j.jasc.2021.07.002

50. Polavarapu P, Fingeret A, Yuil-Valdes A, Olson D, Patel A, Shivaswamy V, et al. Comparison of afirma GEC and GSC to nodules without molecular testing in cytologically indeterminate thyroid nodules. *J Endocr Soc* (2021) 5:bvab148. doi: 10.1210/jendso/bvab148

51. Babazadeh NT, Sinclair TJ, Krishnamurthy V, Jin J, Heiden KB, Shin J, et al. Thyroid nodule molecular profiling: The clinical utility of afirma xpression atlas for nodules with afirma genomic sequencing classifier-suspicious results. *Surgery* (2022) 171:155–9. doi: 10.1016/j.surg.2021.08.058

52. Wei S, Veloski C, Sharda P, Ehya H. Performance of the afirma genomic sequencing classifier versus gene expression classifier: An institutional experience. *Cancer Cytopathol* (2019) 127:720–4. doi: 10.1002/cncy.22188

53. Kerr CE, Ferrell J, Kitano M, Koek W, Dahia PLM, Velez J, et al. Thyroid nodules of indeterminate cytology in Hispanic/Latinx patients. *Head Neck* (2022) 44:1842–8. doi: 10.1002/hed.27100

54. Nasr CE, Andrioli M, Endo M, Harrell RM, Livhits MJ, Osakwe I, et al. Real world performance of the afirma genomic sequencing classifier (GSC) -a meta-analysis. *J Clin Endocrinol Metab* (2022) 6:dgac688. doi: 10.1210/clinem/dgac688

55. Vuong HG, Nguyen TPX, Hassell LA, Jung CK. Diagnostic performances of the afirma gene sequencing classifier in comparison with the gene expression classifier: A meta-analysis. *Cancer Cytopathol* (2021) 129:182–9. doi: 10.1002/cncy.22332

56. Endo M, Nabhan F, Angell TE, Harrell RM, Nasr C, Wei S, et al. Letter to the Editor: Use of molecular diagnostic tests in thyroid nodules with hurthle cell-dominant cytology. *Thyroid* (2020) 30:1390–2. doi: 10.1089/thy.2020.0021

57. Nikiforova MN, Mercurio S, Wald AI, Barbi de Moura M, Callenberg K, Santana-Santos L, et al. Analytical performance of the ThyroSeq v3 genomic classifier for cancer diagnosis in thyroid nodules. *Cancer* (2018) 124:1682–90. doi: 10.1002/cncr.31245

58. O'Conor CJ, Dash RC, Jones CK, Jiang XS. Performance of repeat cytology with reflex ThyroSeq genomic classifier for indeterminant thyroid cytology. *Cancer Cytopathol* (2022) 130:469–76. doi: 10.1002/cncy.22552

59. Glass RE, Marotti JD, Kerr DA, Levy JJ, Vaickus LJ, Gutmann EJ, et al. Using molecular testing to improve the management of thyroid nodules with indeterminate cytology: an institutional experience with review of molecular alterations. *J Am Soc Cytopathol* (2022) 11:79–86. doi: 10.1016/j.jasc.2021.08.004

60. Torrecillas V, Sharma A, Neuberger K, Abraham D. Utility of mutational analysis for risk stratification of indeterminate thyroid nodules in a real-world setting. *Clin Endocrinol* (2022) 96:637–45. doi: 10.1111/cen.14601

61. Selvaggi SM. The role of ThyroSeq V3 testing in the management of patients with indeterminate thyroid nodules on fine needle aspiration. *Diagn Cytopathol* (2021) 49:838–41. doi: 10.1002/dc.24751

62. Desai D, Lepe M, Baloch ZW, Mandel SJ. ThyroSeq v3 for Bethesda III and IV: An institutional experience. *Cancer Cytopathol* (2021) 129:164–70. doi: 10.1002/cncy.22362

63. Abdelhakam DA, Mojica RE, Huenerberg KA, Nassar A. Impact of a genomic classifier on indeterminate thyroid nodules: An institutional experience. *J Am Soc Cytopathol* (2021) 10:155–63. doi: 10.1016/j.jasc.2020.09.005

64. Jug R, Foo WC, Jones C, Ahmadi S, Jiang XS. High-risk and intermediate-high-risk results from the ThyroSeq v2 and v3 thyroid genomic classifier are associated with neoplasia: Independent performance assessment at an academic institution. *Cancer Cytopathol* (2020) 128:563–9. doi: 10.1002/cncy.22283

65. Chen T, Gilfix BM, Rivera J, Sadeghi N, Richardson K, Hier MP, 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) 30:1280–7. doi: 10.1089/thy.2019.0539

66. Lee E, Terhaar S, McDaniel L, Gorelik D, Gerhard E, Chen C, et al. Diagnostic performance of the second-generation molecular tests in the assessment of indeterminate thyroid nodules: A systematic review and meta-analysis. *Am J Otolaryngol* (2022) 43:103394. doi: 10.1016/j.amjoto.2022.103394

67. Vargas-Salas S, Martinez JR, Urra S, Dominguez JM, Mena N, Uslar T, et al. Genetic testing for indeterminate thyroid cytology: Review and meta-analysis. *Endocrine-related Cancer* (2018) 25:R163–77. doi: 10.1530/ERC-17-0405

68. Silaghi CA, Lozovanu V, Georgescu CE, Georgescu RD, Susman S, Nasui BA, et al. Thyroseq v3, afirma GSC, and microRNA panels versus previous molecular tests in the preoperative diagnosis of indeterminate thyroid nodules: A systematic review and meta-analysis. *Front Endocrinol* (2021) 12:649522. doi: 10.3389/fendo.2021.649522

69. Rossi ED, Locantore P, Bruno C, Dell'Aquila M, Tralongo P, Curatolo M, et al. Molecular Characterization of Thyroid Follicular Lesions in the Era of "Next-Generation" Techniques. *Front Endocrinol* (2022) 13:834456. doi: 10.3389/fendo.2022.834456

70. Hlozek J, Pekova B, Rotnagl J, Holy R, Astl J. Genetic changes in thyroid cancers and the importance of their preoperative detection in relation to the general treatment and determination of the extent of surgical intervention-a review. *Biomedicines* (2022) 10. doi: 10.3390/biomedicines10071515

71. Yoo SK, Lee S, Kim SJ, Jee HG, Kim BA, Cho H, et al. Comprehensive analysis of the transcriptional and mutational landscape of follicular and papillary thyroid cancers. *PloS Genet* (2016) 12:e1006239. doi: 10.1371/journal.pgen.1006239

72. Cancer Genome Atlas Research Network. Integrated genomic characterization of papillary thyroid carcinoma. *Cell* (2014) 159:676–90. doi: 10.1016/j.cell.2014.09.050

73. Angell TE, Wirth LJ, Cabanillas ME, Shindo ML, Cibas ES, Babiarz JE, et al. Analytical and clinical validation of expressed variants and fusions from the whole transcriptome of thyroid FNA samples. *Front Endocrinol* (2019) 10:612. doi: 10.3389/fendo.2019.00612

74. Krane JF, Cibas ES, Endo M, Marqusee E, Hu MI, Nasr CE, et al. The afirma xpression atlas for thyroid nodules and thyroid cancer metastases: Insights to inform clinical decision-making from a fine-needle aspiration sample. *Cancer Cytopathol* (2020) 128:452–9. doi: 10.1002/cncy.22300

75. Ali SZ, Siperstein A, Sadow PM, Golding AC, Kennedy GC, Kloos RT, et al. Extending expressed RNA genomics from surgical decision making for cytologically indeterminate thyroid nodules to targeting therapies for metastatic thyroid cancer. *Cancer Cytopathol* (2019) 127(6):362–9. doi: 10.1002/cncy.22132

76. Labourier E, Fahey TJ3rd. Preoperative molecular testing in thyroid nodules with Bethesda VI cytology: Clinical experience and review of the literature. *Diagn Cytopathol* (2021) 49:E175–80. doi: 10.1002/dc.24637

77. Niciporuka R, Nazarovs J, Ozolins A, Narbuts Z, Miklasevics E, Gardovskis J. Can we predict differentiated thyroid cancer behavior? role of genetic and molecular markers. *Medicina (Kaunas)* (2021) 57. doi: 10.3390/medicina57101131

78. Skaugen JM, Taneja C, Liu JB, Wald AI, Nikitski AV, Chiosea SI, et al. Performance of a multigene genomic classifier in thyroid nodules with suspicious for malignancy cytology. *Thyroid* (2022) 32(12):1500–8. doi: 10.1089/thy.2022.0282

79. Hescot S, Al Ghuzlan A, Henry T, Sheikh-Alard H, Lamartina L, Borget I, et al. Prognostic of recurrence and survival in poorly differentiated thyroid cancer. *Endocrine-related Cancer* (2022) 29:625–34. doi: 10.1530/ERC-22-0151

80. Brose MS, Nutting CM, Jarzab B, Elisei R, Siena S, Bastholt L, et al. Sorafenib in radioactive iodine-refractory, locally advanced or metastatic differentiated thyroid cancer: a randomised, double-blind, phase 3 trial. *Lancet* (2014) 384:319–28. doi: 10.1016/S0140-6736(14)60421-9

81. Drilon AE, Subbiah V, Oxnard GR, Bauer TM, Velcheti V, Lakhani NJ, et al. A phase 1 study of LOXO-292, a potent and highly selective RET inhibitor, in patients with RET-altered cancers. *J Clin Oncol* (2018) 36:102–2. doi: 10.1200/JCO.2018.36.15_suppl.102

82. Febbo PG, Ladanyi M, Aldape KD, De Marzo AM, Hammond ME, Hayes DF, et al. NCCN task force report: Evaluating the clinical utility of tumor markers in oncology. *J Natl Compr Cancer Netw: JNCCN* (2011) 9 Suppl 5:S1–32; quiz S33. doi: 10.6004/jnccn.2011.0137

83. Subbiah V, Kreitman RJ, Wainberg ZA, Cho JY, Schellens JHM, Soria JC, et al. Dabrafenib and trametinib treatment in patients with locally advanced or metastatic BRAF V600-mutant anaplastic thyroid cancer. *J Clin Oncol: Off J Am Soc Clin Oncol* (2018) 36:7–13. doi: 10.1200/JCO.2017.73.6785

84. Moura MM, Cavaco BM, Pinto AE, Domingues R, Santos JR, Cid MO, et al. Correlation of RET somatic mutations with clinicopathological features in sporadic medullary thyroid carcinomas. *Br J Cancer* (2009) 100:1777–83. doi: 10.1038/sj.bjc.6605056

85. Tiedje V, Fagin JA. Therapeutic breakthroughs for metastatic thyroid cancer. *Nat Rev Endocrinol* (2020) 16:77–8. doi: 10.1038/s41574-019-0307-2

86. Durante C, Haddy N, Baudin E, Leboulleux S, Hartl D, Travagli JP, et al. Long-term outcome of 444 patients with distant metastases from papillary and follicular thyroid carcinoma: benefits and limits of radioiodine therapy. *J Clin Endocrinol Metab* (2006) 91:2892–9. doi: 10.1210/jc.2005-2838

87. Ruegemer JJ, Hay ID, Bergstralh EJ, Ryan JJ, Offord KP, Gorman CA. Distant metastases in differentiated thyroid carcinoma: a multivariate analysis of prognostic variables. *J Clin Endocrinol Metab* (1988) 67:501–8. doi: 10.1210/jcem-67-3-501

88. Wells SAJr., Robinson BG, Gagel RF, Dralle H, Fagin JA, Santoro M, et al. Vandetanib in patients with locally advanced or metastatic medullary thyroid cancer: A randomized, double-blind phase III trial. *J Clin Oncol: Off J Am Soc Clin Oncol* (2012) 30:134–41. doi: 10.1200/JCO.2011.35.5040

89. Elisei R, Schlumberger MJ, Muller SP, Schoffski P, Brose MS, Shah MH, et al. Cabozantinib in progressive medullary thyroid cancer. *J Clin Oncol: Off J Am Soc Clin Oncol* (2013) 31:3639–46. doi: 10.1200/JCO.2012.48.4659

90. Schlumberger M, Tahara M, Wirth LJ, Robinson B, Brose MS, Elisei R, et al. Lenvatinib versus placebo in radioiodine-refractory thyroid cancer. *New Engl J Med* (2015) 372:621–30. doi: 10.1056/NEJMoa1406470

91. Shonka DCJr., Ho A, Chintakuntlawar AV, Geiger JL, Park JC, Seetharamu N, et al. American Head and neck society endocrine surgery section and international thyroid oncology group consensus statement on mutational testing in thyroid cancer: Defining advanced thyroid cancer and its targeted treatment. *Head Neck* (2022) 44:1277–300. doi: 10.1002/hed.27025

92. Taylor MH, Gainor JF, Hu MI-N, Zhu VW, Lopes G, Leboulleux S, et al. Activity and tolerability of BLU-667, a highly potent and selective RET inhibitor, in patients with advanced RET-altered thyroid cancers. *J Clin Oncol* (2019) 37:6018–8. doi: 10.1200/JCO.2019.37.15_suppl.6018

93. Hong DS, DuBois SG, Kummar S, Farago AF, Albert CM, Rohrberg KS, et al. Larotrectinib in patients with TRK fusion-positive solid tumours: A pooled analysis of three phase 1/2 clinical trials. *Lancet Oncol* (2020) 21:531–40. doi: 10.1016/S1470-2045(19)30856-3

94. Doebele RC, Drilon A, Paz-Ares L, Siena S, Shaw AT, Farago AF, et al. Entrectinib in patients with advanced or metastatic NTRK fusion-positive solid tumours: integrated analysis of three phase 1-2 trials. *Lancet Oncol* (2020) 21:271–82. doi: 10.1016/S1470-2045(19)30691-6

95. Drilon A, Laetsch TW, Kummar S, DuBois SG, Lassen UN, Demetri GD, et al. Efficacy of larotrectinib in TRK fusion-positive cancers in adults and children. *New Engl J Med* (2018) 378:731–9. doi: 10.1056/NEJMoa1714448

96. Marcus L, Lemery SJ, Keegan P, Pazdur R. FDA Approval summary: Pembrolizumab for the treatment of microsatellite instability-high solid tumors. *Clin Cancer Res: an Off J Am Assoc Cancer Res* (2019) 25:3753–8. doi: 10.1158/1078-0432.CCR-18-4070

97. Genutis LK, Tomsic J, Bundschuh RA, Brock PL, Williams MD, Roychowdhury S, et al. Microsatellite instability occurs in a subset of follicular thyroid cancers. *Thyroid* (2019) 29:523–9. doi: 10.1089/thy.2018.0655

98. Are C, Shaha AR. Anaplastic thyroid carcinoma: biology, pathogenesis, prognostic factors, and treatment approaches. *Ann Surg Oncol* (2006) 13:453–64. doi: 10.1245/ASO.2006.05.042

99. Maniakas A, Dadu R, Busaidy NL, Wang JR, Ferrarotto R, Lu C, et al. Evaluation of overall survival in patients with anaplastic thyroid carcinoma, 2000-2019. *JAMA Oncol* (2020) 6:1397–404. doi: 10.1001/jamaoncol.2020.3362

100. Lin B, Ma H, Ma M, Zhang Z, Sun Z, Hsieh IY, et al. The incidence and survival analysis for anaplastic thyroid cancer: A SEER database analysis. *Am J Trans Res* (2019) 11:5888–96.

101. Xu B, Fuchs T, Dogan S, Landa I, Katabi N, Fagin JA, et al. Dissecting anaplastic thyroid carcinoma: A comprehensive clinical, histologic, immunophenotypic, and molecular study of 360 cases. *Thyroid* (2020) 30:1505–17. doi: 10.1089/thy.2020.0086

102. Oishi N, Kondo T, Ebina A, Sato Y, Akaishi J, Hino R, et al. Molecular alterations of coexisting thyroid papillary carcinoma and anaplastic carcinoma: Identification of TERT mutation as an independent risk factor for transformation. *Modern Pathol: an Off J United States Can Acad Pathol Inc* (2017) 30:1527–37. doi: 10.1038/modpathol.2017.75

103. Maniakas A, Zafereo M, Cabanillas ME. Anaplastic thyroid cancer: New horizons and challenges. *Endocrinol Metab Clinics North America* (2022) 51:391–401. doi: 10.1016/j.ecl.2021.11.020

104. Sandulache VC, Williams MD, Lai SY, Lu C, William WN, Busaidy NL, et al. Real-time genomic characterization utilizing circulating cell-free DNA in patients with anaplastic thyroid carcinoma. *Thyroid* (2017) 27:81–7. doi: 10.1089/thy.2016.0076

105. Khatami F, Tavangar SM. Liquid biopsy in thyroid cancer: New insight. *Int J Hematol Oncol Stem Cell Res* (2018) 12:235–48.

106. Smith AL, Williams MD, Stewart J, Wang WL, Krishnamurthy S, Cabanillas ME, et al. Utility of the BRAF p.V600E immunoperoxidase stain in FNA direct smears and cell block preparations from patients with thyroid carcinoma. *Cancer Cytopathol* (2018) 126:406–13. doi: 10.1002/cncy.21992

107. Wang JR, Zafereo ME, Dadu R, Ferrarotto R, Busaidy NL, Lu C, et al. Complete surgical resection following neoadjuvant dabrafenib plus trametinib in BRAF(V600E)-mutated anaplastic thyroid carcinoma. *Thyroid* (2019) 29:1036–43. doi: 10.1089/thy.2019.0133

108. Wells SAJr., Asa SL, Dralle H, Elisei R, Evans DB, Gagel RF, et al. Revised American thyroid association guidelines for the management of medullary thyroid carcinoma. *Thyroid* (2015) 25:567–610. doi: 10.1089/thy.2014.0335

109. Ceolin L, Duval M, Benini AF, Ferreira CV, Maia AL. Medullary thyroid carcinoma beyond surgery: Advances, challenges, and perspectives. *Endocrine-related Cancer* (2019) 26:R499–518. doi: 10.1530/ERC-18-0574

110. Modigliani E, Cohen R, Campos JM, Conte-Devolx B, Maes B, Boneu A, et al. Prognostic factors for survival and for biochemical cure in medullary thyroid carcinoma: Results in 899 patients. the GETC study group. groupe d'etude des tumeurs a calcitonine. *Clin Endocrinol* (1998) 48:265–73. doi: 10.1046/j.1365-2265.1998.00392.x

111. Boichard A, Croux L, Al Ghuzlan A, Broutin S, Dupuy C, Leboulleux S, et al. Somatic RAS mutations occur in a large proportion of sporadic RET-negative medullary thyroid carcinomas and extend to a previously unidentified exon. *J Clin Endocrinol Metab* (2012) 97:E2031–2035. doi: 10.1210/jc.2012-2092

112. Ciampi R, Romei C, Ramone T, Prete A, Tacito A, Cappagli V, et al. Genetic landscape of somatic mutations in a Large cohort of sporadic medullary thyroid carcinomas studied by next-generation targeted sequencing. *iScience* (2019) 20:324–36. doi: 10.1016/j.isci.2019.09.030

113. Wirth LJ, Sherman E, Robinson B, Solomon B, Kang H, Lorch J, et al. Efficacy of selpercatinib in RET-altered thyroid cancers. *New Engl J Med* (2020) 383:825–35. doi: 10.1056/NEJMoa2005651

114. Gainor JF, Curigliano G, Kim DW, Lee DH, Besse B, Baik CS, et al. Pralsetinib for RET fusion-positive non-small-cell lung cancer (ARROW): A multi-cohort, open-label, phase 1/2 study. *Lancet Oncol* (2021) 22:959–69. doi: 10.1016/S1470-2045(21)00247-3

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 23 March 2023
DOI 10.3389/fendo.2023.1062902

OPEN ACCESS

EDITED BY
Andrea Frasoldati,
Endocrine Unit ASMN, Italy

REVIEWED BY
Adnan Işgör,
Memorial Sisli Hospital, Türkiye
Demet Sengul,
Giresun University, Türkiye

*CORRESPONDENCE
Pengfei Luo
luopengf1987@163.com

SPECIALTY SECTION
This article was submitted to Thyroid Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 16 November 2022
ACCEPTED 13 March 2023
PUBLISHED 23 March 2023

CITATION
Luo P, Mu X, Ma W, Jiao D and Zhang P (2023) Effect of a stylet on specimen sampling in thyroid fine needle aspiration: A randomized, controlled, non-inferiority trial.
Front. Endocrinol. 14:1062902.
doi: 10.3389/fendo.2023.1062902

Effect of a stylet on specimen sampling in thyroid fine needle aspiration: A randomized, controlled, non-inferiority trial

Pengfei Luo*, Xiali Mu, Wei Ma, Dahai Jiao and Peixin Zhang

Department of General Surgery, Fuyang People's Hospital, Fuyang, China

Background: There is a cost advantage in using a needle without stylet over a needle with stylet in thyroid fine needle aspiration (FNA). This study aimed to elucidate the non-inferiority of thyroid FNA without a stylet (S-) to thyroid FNA with a stylet (S+) on specimen sampling.

Methods: In this study, patients with thyroid nodules undergoing FNA were consecutively enrolled between May 2022 and July 2022. One experienced operator performed two punctures of each nodule with a stylet and without a stylet. Specimen adequacy was the primary outcome. Wald test was used for statistical analysis of the primary outcome. The difference in specimen adequacy between the two methods was expressed as a two-sided 95% confidence interval (CI). The S- method was considered non-inferior to the S+ method if the lower bound of the 95% CI of the S- minus S+ adequacy difference was greater than a predetermined non-inferiority margin of -10%.

Results: A total of 149 patients (195 nodules) were enrolled in the study. A total of 167 of 195 nodules (85.64%) and 169 of 195 nodules (86.67%) were obtained adequate specimens using the S+ and S- methods, respectively. The difference in specimen adequacy (S- minus S+) between the two methods was 1.03% (95% CI, -5.83% to 7.88%). The lower bound 95% CI of the difference in specimen adequacy (-5.83%) was greater than the predetermined non-inferiority margin of -10%. The difference in the yield for malignancy was not significantly different between the two methods.

Conclusion: Thyroid FNA without a stylet is non-inferior to thyroid FNA with a stylet on specimen sampling.

KEYWORDS

fine needle aspiration, thyroid nodule, specimen adequacy, stylet, needle

Introduction

Fine needle aspiration (FNA) is a simple, minimally invasive, and highly accurate method widely used in the diagnosis of many organ diseases, including thyroid (1–13). A stylet can theoretically prevent the lumen of the needle from being blocked by other non-lesion components, including blood, before puncturing into the target lesion, thus allowing fuller aspiration of the target tissue into the needle once the stylet is removed. The stylet is routinely used by many puncturing physicians to improve the quality of specimens (1, 2). Although this assumption of a preference for the use of a stylet seems logical, this assumption has not been demonstrated on an empirical basis. Several studies in other areas of FNA, such as gastrointestinal endoscopy and respiratory endoscopy FNA have shown that the stylet does not improve specimen quality and diagnostic efficiency (3–8). Besides, some studies have shown that stylet may be associated with an inferior specimen quality (9).

Furthermore, the value of the stylet has not been systematically evaluated in the FNA of the thyroid. Only one study evaluated the role of stylet using a few selected nodules (hypoechoic vascular type II nodules) (1), representing a small fraction of all types of thyroid nodules (14). This study aimed to compare the specimens obtained by FNA with and without the stylet in all types of unselected thyroid nodules for adequacy and yield for malignancy.

Materials and methods

Study design

A single-center, prospective, randomized, controlled, non-inferiority trial was conducted at Fuyang People's Hospital, a tertiary referral medical center. Written informed consent was obtained from all patients, and the study was approved by the Ethics Committee of Fuyang People's Hospital (NO: 2022-80).

Patients

Patients with thyroid nodules who underwent FNA between May 2022 and July 2022 were prospectively enrolled in the study. The inclusion criteria were: (1) patients aged 18 years and above; (2) patients who provided written informed consent; and (3) patients who underwent ultrasound suggesting the presence of a thyroid nodule. The exclusion criteria were: (1) patients under anticoagulant therapy, such as aspirin and warfarin; (2) patients who could not cooperate with punctures, such as severe cough; and (3) those who could not provide written informed consent.

Randomization

Each nodule was sampled for four passes (twice with a stylet (S+) and twice without a stylet (S-)). The nodules were randomly divided (1:1) into S+ (order of the four passes; S+→S-→S+→S-) and S- (order of the passes; S-→S+→S-→S+) to avoid the effect of bleeding from the previous pass on the later pass specimen. The order of S+ or S- first pass was determined by a preprinted random sequence that was kept in an opaque sealed envelope that the operator opened after the patients confirmed their enrollment.

FNA procedure

FNAs were performed under ultrasound (US) guidance by the same experienced operator. For patients with multiple nodules, FNA was performed on suspicious nodules, otherwise, the largest nodule was selected for sampling. For patients with mixed cystic-solid nodules, FNA was performed from the solid component. An US scanner (M9, Mindray, Shenzhen, China) was used for US with a 7.5-15 MHz linear array transducer. The medical record report contained the number, location, size, echogenicity, composition, vascularity, calcifications, depth, and Thyroid Imaging Reporting and Data System (TIRADS) categories of the punctured nodules. The passes were performed using a 25-gauge disposable puncture needle with a stylet (CCZA, Leapmed, Suzhou, China) without syringes. The stylet in the needle was kept during the puncture for S+ passes and removed before S- passes.

The S+ pass was performed as follows: The patient was placed in a supine position with the neck slightly extended. The patient underwent local anesthesia using 0.1-0.3 ml of 2% lidocaine, and the skin was disinfected with iodophor. The stylet was removed after the needle tip puncture was put into the nodule under ultrasound guidance. The needle was pulled out when the needle was cut back and forth 15-20 times within the nodule or when the sample material was seen in the hub. Suction was not used in the process. The material in the needle lumen was expelled onto one clean glass slide using a 5 ml air-filled syringe after each pass. All the steps were the same for the S+ and S- pass, except for the stylet removal step that was avoided in S- pass.

One slide specimen was made for each pass smear, and these slides were labeled as slides A1, B1, A2, and B2 in the order of pass. Sometimes slides A (A1 and A2) were two S+ specimens or two S- specimens since the choice of - the first pass was random to avoid bias of the pathologist. No on-site assessment was performed, where the specimens were air-dried, fixed in 95% alcohol, and stained with hematoxylin and eosin.

Cytological evaluation

Five pathologists blinded to the stylet status of the passes evaluated cytology specimens. The same pathologist evaluated all specimens of one patient. Every thyroid FNA was first evaluated for specimen adequacy. Each specimen was classified as adequate or inadequate based on the Bethesda System for Reporting Thyroid Cytopathology (TBSRTC). Inadequate specimens (category I specimens in TBSRTC) were defined as specimens that did not meet the criteria for adequacy (presence of at least six groups of well-visualized follicular cells; each group containing at least 10 well-preserved epithelial cells).

The two S+ specimens and two S- specimens of one nodule were evaluated as a separate whole with a separate cytologic result (specimens A1 and A2 as "nodule A" specimens and specimens B1 and B2 as "nodule B" specimens). The S+ and S- specimens of the nodule were considered inadequate when the two S+ specimens and two S- specimens were inadequate.

Outcome variables

Specimen adequacy was the primary outcome, while the yield for malignancy was the secondary outcome. Specimen adequacy was defined as the rate at which adequate specimens were obtained (proportion of non-TBSRTC category I specimens). The yield for malignancy was defined as a percentage of TBSRTC category VI specimens.

Sample size estimation and statistical analysis

Power Analysis and Sample Size (PASS) version 11 was used for sample size estimation. The sample size was determined based on data from a meta-analysis involving 25,000 patients, which claimed an overall specimen adequacy rate of 87.1% for thyroid FNA (15). The specimen adequacy rate by S+ or S- FNA was then assumed to be 87.1%. The non-inferiority margin was then set at 10%, with a class I error of 0.025 (1 margin) and a class II error of 0.8. A total of 195 lesions per group were required assuming 10% dropouts. Besides, this study used its own control, and thus the final sample size was 195 nodules per group (195 nodules in total).

For a patient with more than one nodule undergoing FNA, the nodules were considered independent observations for statistical analysis. Continuous variables were expressed as means and standard deviations. Categorical variables were expressed as frequencies and percentages. The primary outcome was analyzed using Wald test. The difference between the adequacy of the two methods was expressed as a two-sided 95% confidence interval (CI). The S- method was considered non-inferior to the S+ method if the lower bound of the two-sided 95% (equivalent to one-sided 97.5%) CI of the difference in specimen adequacy (S- minus S+) was greater than -10%. Chi-square tests were used to analyze the secondary outcomes. Statistical Package for Social Sciences (SPSS) version 23 and Statistical Analysis System (SAS) version 9.4 were used for all analyses.

Results

Patients and nodules

A total of 154 patients were enrolled from May 2022 to July 2022 (Figure 1). Five patients were excluded, of which 1 was taking aspirin, 1 was taking warfarin, 2 refused to participate, and nodules were not detected in one patient. Finally, 149 patients were included in the analysis (121 (81.21%) females and 28 males with 38 (25.5%) solitary nodules and 111 (74.5%) multiple nodules). The mean age of the included patients was 46.97 years (standard deviation (SD) 13.56). Only one nodule was sampled in 103 patients, and two nodules were sampled in 46 patients. A total of 195 nodules were sampled (mean nodule diameter, 1.42 cm; SD, 1.12 cm). The TIRADS category of the nodules was as follows: 3 in 52 (26.67%), 4a in 105 (53.85%), 4b in 33 (16.92%), and 4c in 5 (2.56%). The composition of the nodules was as follows: solid in 149 (76.41%) and solid-cystic in 46 (23.59%). The echogenicity of the nodules was hypoechoic in 161 (82.56%), isoechoic in 33 (16.92%), and

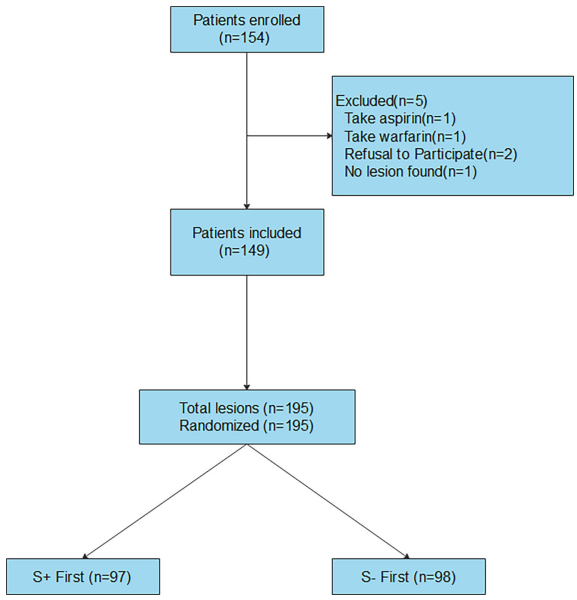


FIGURE 1
Flow diagram for 154 consecutive patients referred for thyroid nodules fine needle aspiration (FNA).

hyperechoic in 1 (0.51%). A total of 101 nodules (51.79%) had calcifications. The vascularity of the nodules was peripheral in 26 (13.33%), central in 63 (32.31%), and no vascularity in 106 (54.36%). A total of 97 and 98 nodules were randomly selected for S+ first puncture and S- first puncture, respectively.

Cytological results

The final cytological diagnosis was malignant (TBSRTC category VI) in 23 nodules (11.79%), suspicious for malignancy (TBSRTC category V) in 33 nodules (16.92%), follicular neoplasm or suspicious for a follicular neoplasm (TBSRTC category IV) in 5 nodules (2.56%), atypia of undetermined significance or follicular lesion of undetermined significance (TBSRTC category III) in 24 nodules (12.31%), benign (TBSRTC category II) in 91 nodules (46.67%), and inadequacy (TBSRTC category I) in 19 nodules (9.74%).

The two S+ specimens and two S- specimens of each nodule were evaluated as a separate whole, and the cytological results are shown in Table 1.

Specimen adequacy

The rates of obtaining adequate specimens were 167 of 195 nodes (85.64%) and 169 of 195 nodes (86.67%) for S+ and S- methods, respectively. The difference in specimen adequacy (S- minus S+) between the two methods was 1.03% (95% CI, -5.83% to 7.88%). The lower bound 95% CI of the difference in specimen adequacy (-5.83%) was greater than the predetermined non-inferiority margin of -10%, demonstrating the non-inferiority of the S- method to the S+ method.

Analysis of each subgroup showed that the differences in specimen adequacy were not significantly different between the two methods (Table 2).

Yield for malignancy

The yields for malignancy were 21 of 195 nodules (10.77%) and 20 of 195 nodules (10.26%) for S+ method and S- method, respectively (P = 0.869). Similarly, the yields for malignancy were also not significantly different between the two methods when suspicious malignant nodules were included in the malignancy group (S+:50/195(25.64%) vs. S-:51/195(26.15%), P = 0.908).

TABLE 1 Cytological results of specimens obtained with and without stylet.

TBSRTC category	S+(n=195)	S-(n=195)
I	28	26
II	89	90
III	23	23
IV	5	5
V	29	31
VI	21	20

Adverse events

No complications or needle blockage were reported in this study.

Discussion

Puncture needles with a stylet is widely used by many puncturers during thyroid FNA (1, 2) based on the unproven premise that using the stylet improves the specimen's quality by preventing the needle's lumen from being blocked or contaminated by other non-lesion components, including blood, before entering the target lesion. Nevertheless, stylet is time-consuming and labor-intensive and may also increase the risk of accidental needle stick injuries. Meanwhile, a needle with a stylet is expensive compared to a needle without a stylet, such as an ordinary syringe needle, and thus may significantly increase the cost of puncture. However, recent data have shown that using a stylet in other areas of FNA, such as gastrointestinal endoscopic FNA and respiratory endoscopic FNA, does not improve the quality of specimen (3–8). Besides, stylet is sometimes associated with an inferior specimen quality (9).

Besides, there are limited data comparing thyroid FNA with and without a stylet. A published trial comparing the two techniques compared only a subset of selected thyroid nodules (hypoechoic vascular type II nodules) (1), representing only a small fraction of the total types of thyroid nodules (14). No study has evaluated the value of stylet for the puncture of all types of thyroid nodules. To the best of our knowledge, this is the first study assessing the value of the stylet in overall unselective thyroid nodules. In this study, results showed that a stylet does not improve specimen adequacy or yield for malignancy during thyroid FNA, consistent with results in other areas of FNA.

However, a previous published study comparing thyroid FNA with and without a stylet showed that a stylet can improve specimen adequacy (1), by preventing blood or cystic fluid from entering the lumen of the needle during needle insertion into the target nodule. In this study, results showed that stylet may prevent puncture blood from entering the lumen of the needle before removal of the stylet, but this may not enhance specimen quality conducted by experienced puncture operators. Experienced operators can often anticipate the puncture route and angle, and the needle can be punctured into the target nodule with no or only slight adjustment. Nonetheless, puncture bleeding from thyroid tissue often occurs during repeated needle punctures during a long adjustment period. Meanwhile, the process of removal of the stylet generates a negative pressure (11), which depends on the speed and the length of removal. Animal experiments have shown that the maximum negative pressure can reach close to one atmosphere (10), which may affect specimen quality. The needle tip cut has not yet been used to obtain lesion cells when the stylet is withdrawn after the puncture needle has entered the target nodule. The negative pressure generated by the withdrawal of the stylet may draw non-

TABLE 2 Subgroup analysis of specimen adequacy obtained with and without stylet.

Parameter	S+ adequacy no. (%)	S- adequacy no. (%)	*P* value
Size			
<1cm(n=96)	80(83.33%)	81(84.38%)	0.845
1-3cm(n=82)	70(85.37%)	71(86.59%)	0.822
>3cm(n=17)	17(100%)	17(100%)	1.000
Depth in thyroid			
First third(n=132)	115(87.12%)	116(87.88%)	0.852
Middle third(n=52)	42(80.77%)	43(82.69%)	0.800
Last third(n=11)	10(90.91%)	10(90.91%)	1.000
Echogenicity			
Hypoechoic(n=161)	139(86.34%)	142(88.20%)	0.616
Isoechoic(n=33)	27(81.82%)	26(78.79%)	0.757
Hyperechoic(n=1)	1(100%)	1(100%)	1.000
Calcifications			
Microcalcification(n=60)	52(86.67%)	53(88.33%)	0.783
Macrocalcification(n=41)	31(75.61%)	34(82.93%)	0.414
No calcification(n=94)	84(89.36%)	82(87.23%)	0.650
Vascularity			
Central(n=63)	56(88.89%)	55(87.30%)	0.783
Peripheral(n=26)	20(76.92%)	23(88.46%)	0.465
None(n=106)	91(85.85%)	91(85.85%)	1.000
Composition			
Solid(n=149)	128(85.91%)	131(87.92%)	0.606
Solid-cystic(n=46)	39(84.78%)	38(82.61%)	0.778
TIRADS category			
3(n=52)	44(84.62%)	43(82.69%)	0.791
4a(n=105)	92(87.62%)	92(87.62%)	1.000
4b(n=33)	27(81.82%)	30(90.91%)	0.475
4c(n=5)	5(100%)	5(100%)	1.000

lesion components, such as blood and cystic fluid, into the lumen of the needle, which can interfere with the entry of lesion cells obtained by subsequent cutting. Previous gastrointestinal endoscopic FNA studies also highlighted the above when there were significantly higher bloody specimens and lower adequate specimens in the S+ group than in the S- group (9).

Furthermore, only 480 of 2750 patients were screened in the study comparing thyroid FNA with and without a stylet (1) rather than including patients consecutively. Besides, different needles were used in different nodules, and the study lacked true randomization. The above may have contributed to selection bias, possibly by artificially assigning easier satisfactory nodules to the experimental group or excluding more difficult nodules from the experimental group.

There is no doubt that the higher cost of puncture needles with a stylet than those without a stylet, such as an ordinary syringe needle, will significantly increase the puncture cost for patients. Therefore, its higher cost of puncture needles with stylet is unjustified if it does not improve specimen quality. Considering a large number of FNA examinations performed worldwide each year, the cost issue is not irrelevant. Our findings provide a reasonable basis for using low-cost puncture needles without a stylet can be used during thyroid FNA, which would reduce the average per-patient puncture cost and save health insurance funds.

In this study, the patient population was representative of the typical patient population seen by most thyroid surgeons working in tertiary referral centers, with most being women and a wide age range. Therefore, the results may broadly apply to most patients with thyroid nodules. FNA is an operation-dependent procedure. Herein, only one experienced operator performed the procedures, thus excluding the influence of inter-operator differences. Furthermore, a non-inferiority design was used to ensure an adequate sample size. A randomized design was used for the order of S+ first or S- first puncture of each nodule to exclude the effect of bleeding from the first puncture on the adequacy of the second specimen. Patients using anticoagulant therapy, such as aspirin and warfarin, were excluded since these treatments affect specimen adequacy (16). All possible interference factors were eliminated to ensure that other factors, except the stylet, do not affect the results.

This study has some limitations. First, the operator was not blinded to the stylet status of each puncture since this is logically impossible. Second, five cytopathologists interpreted the slides, and interobserver variation may have occurred during the interpretation. However, all slides from one patient were evaluated by the same cytopathologist. Third, only 25-gauge needles were used, and thus these results may not apply to other needle sizes. Finally, the sensitivity, specificity, and accuracy of the two techniques were not explored.

Conclusions

In conclusion, this non-inferiority study demonstrated that the absence of a stylet during thyroid FNA is non-inferiority to using a stylet on specimen sampling. If other studies confirm these results, using a low-cost needle without a stylet during thyroid FNA is justified. This would make the whole process easier, as well as more cost-effective.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

This study was approved by the Ethics Committee of Fuyang People's Hospital (NO: 2022-80). The patients/participants provided their written informed consent to participate in this study.

Author contributions

All authors contributed to the study conception and design. PL, XM, and WM performed material preparation, data collection and analysis. PL wrote the first draft of the manuscript and all authors approved previous versions of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by Self-funded Science and Technology Project of Fuyang City (FK202081015) and Fuyang City Health Commission scientific research project (FY2021-014).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Cappelli C, Tironi A, Mattanza C, Cumetti D, Agosti B, Gandossi E, et al. Cost-effectiveness of fine-needle-aspiration cytology of thyroid nodules with intranodular vascular pattern using two different needle types. *Endocr Pathol* (2005) 16(4):349–54. doi: 10.1385/ep:16:4:349

2. Landau MS, Pearce TM, Carty SE, Wolfe J, Yip L, McCoy KL, et al. Comparison of the collection approaches of 2 large thyroid fine-needle aspiration practices reveals differing advantages for cytology and molecular testing adequacy rates. *J Am Soc Cytopathol* (2019) 8(5):243–9. doi: 10.1016/j.jasc.2019.03.004

3. Rastogi A, Wani S, Gupta N, Singh V, Gaddam S, Reddymasu S, et al. A prospective, single-blind, randomized, controlled trial of EUS-guided FNA with and without a stylet. *Gastrointest Endosc* (2011) 74(1):58–64. doi: 10.1016/j.gie.2011.02.015

4. Wani S, Early D, Kunkel J, Leathersich A, Hovis CE, Hollander TG, et al. Diagnostic yield of malignancy during EUS-guided FNA of solid lesions with and without a stylet: A prospective, single blind, randomized, controlled trial. *Gastrointest Endosc* (2012) 76(2):328–35. doi: 10.1016/j.gie.2012.03.1395

5. Gimeno-García AZ, Paquin SC, Gariépy G, Sosa AJ, Sahai AV. Comparison of endoscopic ultrasonography-guided fine-needle aspiration cytology results with and without the stylet in 3364 cases. *Dig Endosc* (2013) 25(3):303–7. doi: 10.1111/j.1443-1661.2012.01374.x

6. Abe Y, Kawakami H, Oba K, Hayashi T, Yasuda I, Mukai T, et al. Effect of a stylet on a histological specimen in EUS-guided fine-needle tissue acquisition by using 22-gauge needles: A multicenter, prospective, randomized, controlled trial. *Gastrointest Endosc* (2015) 82(5):837–844.e1. doi: 10.1016/j.gie.2015.03.1898

7. Xu Y, Lin J, Jin Y, Wu X, Zheng H, Feng J. Is endobronchial ultrasound-guided transbronchial needle aspiration with a stylet necessary for lymph node screening in lung cancer patients? *Braz J Med Biol Res* (2017) 50(10):e6372. doi: 10.1590/1414-431X20176372

8. Wani S, Gupta N, Gaddam S, Singh V, Ulusarac O, Romanas S, et al. A comparative study of endoscopic ultrasound guided fine needle aspiration with and without a stylet. *Dig Dis Sci* (2011) 56(8):2409–14. doi: 10.1007/s10620-011-1608-z

9. Sahai AV, Paquin SC, Gariépy G. A prospective comparison of endoscopic ultrasound-guided fine needle aspiration results obtained in the same lesion, with and without the needle stylet. *Endoscopy* (2010) 42(11):900–3. doi: 10.1055/s-0030-1255676

10. Welker L, Akkan R, Holz O, Schultz H, Magnussen O. Diagnostic outcome of two different CT-guided fine needle biopsy procedures. *Diagn Pathol* (2007) 2:31. doi: 10.1186/1746-1596-2-31

11. Paik WH, Choi JH, Park Y, Lee JB, Park DH. Optimal techniques for EUS-guided fine-needle aspiration of pancreatic solid masses at facilities without on-site cytopathology: Results from two prospective randomised trials. *J Clin Med* (2021) 10 (20):4662. doi: 10.3390/jcm10204662

12. Sengul D, Sengul I. Reassessing combining real-time elastography with fine-needle aspiration biopsy to identify malignant thyroid nodules. *Am J Med Case Rep* (2021) 9(11):552–3. doi: 10.12691/ajmcr-9-11-9

13. Sengul I, Sengul D. Hermeneutics for evaluation of the diagnostic value of ultrasound elastography in TIRADS 4 categories of thyroid nodules. *Am J Med Case Rep* (2021) 9(11):538–9. doi: 10.12691/ajmcr-9-11-5

14. Cappelli C, Castellano M, Pirola I, Cumetti D, Agosti B, Gandossi E, et al. The predictive value of ultrasound findings in the management of thyroid nodules. *QJM* (2007) 100(1):29–35. doi: 10.1093/qjmed/hcl121

15. 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):333–9. doi: 10.1159/000339959

16. Khan TS, Sharma E, Singh B, Jammu B, Chadha A, Markanday C, et al. Aspirin increases the risk of nondiagnostic yield of fine-needle aspiration and biopsy of thyroid nodules. *Eur Thyroid J* (2018) 7(3):129–32. doi: 10.1159/000488451

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 12 May 2023
DOI 10.3389/fendo.2023.1140816

OPEN ACCESS

EDITED BY
Jeffrey Garber,
Atrius Health, United States

REVIEWED BY
Sina Jasim,
Washington University in St. Louis, United States
Chenbin Liu,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

*CORRESPONDENCE
Ping Zhou
zhouping1000@hotmail.com

RECEIVED 09 January 2023
ACCEPTED 01 May 2023
PUBLISHED 12 May 2023

CITATION
Xia M, Song F, Zhao Y, Xie Y, Wen Y and Zhou P (2023) Ultrasonography-based radiomics and computer-aided diagnosis in thyroid nodule management: performance comparison and clinical strategy optimization.
Front. Endocrinol. 14:1140816.
doi: 10.3389/fendo.2023.1140816

Ultrasonography-based radiomics and computer-aided diagnosis in thyroid nodule management: performance comparison and clinical strategy optimization

Mengwen Xia[1], Fulong Song[2], Yongfeng Zhao[1], Yongzhi Xie[2], Yafei Wen[1] and Ping Zhou[1]*

[1]Department of Ultrasonography, The Third Xiangya Hospital of Central South University, Changsha, China, [2]Department of Radiology, The Third Xiangya Hospital of Central South University, Changsha, China

Objectives: To compare ultrasonography (US) feature-based radiomics and computer-aided diagnosis (CAD) models for predicting malignancy in thyroid nodules, and to evaluate their utility for thyroid nodule management.

Methods: This prospective study included 262 thyroid nodules obtained between January 2022 and June 2022. All nodules previously underwent standardized US image acquisition, and the nature of the nodules was confirmed by the pathological results. The CAD model exploited two vertical US images of the thyroid nodule to differentiate the lesions. The least absolute shrinkage and operator algorithm (LASSO) was applied to choose radiomics features with excellent predictive properties for building a radiomics model. Ultimately, the area under the receiver operating characteristic curve (AUC) and calibration curves were assessed to compare diagnostic performance between the models. DeLong's test was used to analyze the difference between groups. Both models were used to revise the American College of Radiology Thyroid Imaging Reporting and Data Systems (ACR TI-RADS) to provide biopsy recommendations, and their performance was compared with the original recommendations.

Results: Of the 262 thyroid nodules, 157 were malignant, and the remaining 105 were benign. The diagnostic performance of radiomics, CAD, and ACR TI-RADS models had an AUC of 0.915 (95% confidence interval (CI): 0.881–0.947), 0.814 (95% CI: 0.766–0.863), and 0.849 (95% CI: 0.804–0.894), respectively. DeLong's test showed a statistically significant between the AUC values of models ($p < 0.05$). Calibration curves showed good agreement in each model. When both models were applied to revise the ACR TI-RADS, our recommendations significantly improved the performance. The revised recommendations based on radiomics and CAD showed an increased sensitivity, accuracy, positive predictive value, and negative predictive value, and decreased unnecessary

fine-needle aspiration rates. Furthermore, the radiomics model's improvement scale was more pronounced (33.3–16.7% vs. 33.3–9.7%).

Conclusion: The radiomics strategy and CAD system showed good diagnostic performance for discriminating thyroid nodules and could be used to optimize the ACR TI-RADS recommendation, which successfully reduces unnecessary biopsies, especially in the radiomics model.

KEYWORDS

thyroid nodule, radiomics, computer-aided diagnosis, ultrasonography, risk assessment, prediction

1 Introduction

Thyroid nodules are common but often asymptomatic, and guidelines strongly recommend that all patients with known or suspected thyroid nodules undergo thyroid ultrasonography (US) with a survey of the cervical lymph nodes (1). With the widespread use of high-frequency US, the prevalence of thyroid nodules has been reported to be as high as 68%, with a higher proportion among populations with iodine deficiency and the elderly (2). The management of thyroid nodules has shown increased clinical importance due to the high incidence of nodules and soaring healthcare costs. However, operator-specific expertise and the inability to quantify image features frequently restrict the sensitivity and specificity of US diagnoses, which results in a lack of consistency and objectivity (3).

With the presentation and application of various risk-stratification systems, such as the Thyroid Imaging Reporting and Data System released by the American College of Radiology (ACR TI-RADS), standardized terminology has gradually been used to describe the appearance of thyroid nodules (4, 5). US has become a primary diagnostic tool used for the final classification of thyroid nodules and can help in decision-making regarding the use of fine-needle aspiration (FNA). However, due to the subjectivity, diversity, and overlapping risk features between the benign and malignant nodules, data on the interobserver agreement are weak (6).

Recent advances in technology have shown superiority in the differentiation of thyroid nodules. The use of computer-aided diagnosis (CAD) systems in the diagnosis of thyroid nodules seems to be a promising tool (7). Several artificial intelligence tools are commercially available that have received Food and Drug Association approval, such as S-detect, AmCAD-UT, Koios DS, and Medo Thyroid. Previous studies have shown that S-detect could provide second objective decision-making support *via* a semiautomated workflow in differentiating thyroid nodules from US images and reducing the rate of missed diagnoses (8–12). S-detect technology has been iterated several times, and can now identify calcification as an important clue. More recently, a new analysis method called radiomics, which is based on data science, quantifies the characteristics of lesions in medical images to extract a significant number of phenotypic features (13, 14). To our best knowledge, no published study has compared the accuracy of radiomics and CAD systems based on US features in the prediction of thyroid cancer for thyroid nodule management.

Therefore, this study aimed to prospectively evaluate the diagnostic efficiency of benign and malignant thyroid nodules using the US-based radiomics analysis method and CAD system while exploring their potential complementary role to ACR management recommendations.

2 Materials and methods

2.1 Patients

This study was approved by the ethical review committee of the Third Xiangya Hospital of Central South University, and written informed consent was obtained from all patients before they received examinations. Patients and data were collected prospectively randomized and double-blinded by a tertiary hospital.

A total of 301 thyroid nodules from 179 consecutive patients who had undergone regular preoperative gray-scale US imaging of thyroid nodules with clear images and had obtained a pathological diagnosis by FNA or surgical resection for lesions within 2 weeks were included at our institution between January 2022 and June 2022. Among the 301 thyroid nodules, 39 were excluded due to the following reasons (1): biopsy or local treatment before US (n =23) (2); other cancers (n = 2) (3); poor image quality or ill-defined pathological results (n = 8); and (4) multiple nodules could not be conclusively correlated in US images with pathological diagnosis (n = 6). Finally, this study included 148 patients in total with 262 thyroid nodules. Figure 1 shows the flowchart of this study population. The final diagnosis was based on FNA or surgical histopathology.

2.2 Image acquisition and annotation

All US examinations were performed with Hera W10 (Samsung Medison) and a real-time CAD US system (S-Detect for Thyroid;

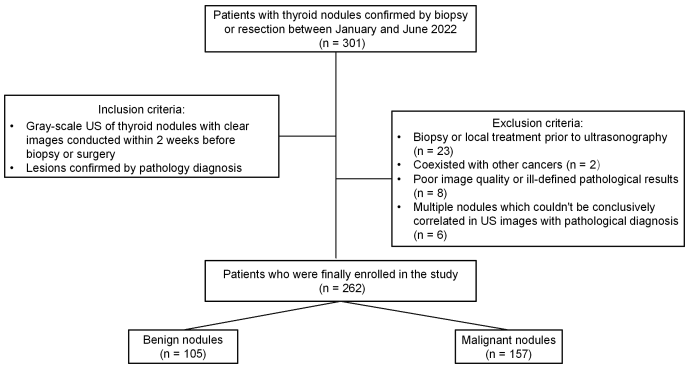


FIGURE 1
Flowchart of the study population.

Samsung Medison) using a 3–12 MHz linear probe. A senior trained radiologist with 25 years of experience in thyroid imaging independently performed all US examinations and numbered the nodules. Meanwhile, the thyroid nodule's largest segment (longitudinal section) and its vertical section (transverse section) were measured for further image annotation.

2.2.1 CAD image acquisition and annotation

The same sonologist analyzed CAD data with S-Detect on transverse and longitudinal sections immediately after image acquisition. After manually confirming the location of the lesion, the software automatically segmented the mass contours. The operator manually readjusted the outline if the contour border was dissatisfactory. The software analyzed US features of the lesion, including composition, echogenicity, orientation, margins, shape, calcifications, and spongiform appearance (8). Finally, S-Detect provided the diagnosis as "possibly benign" or "possibly malignant" in dichotomy form (Figure 2). In addition, if the assessments of two vertical sections were inconsistent, the malignant result was regarded as final.

2.2.2 US image annotation

Sonograms were independently evaluated by an experienced senior thyroid imaging expert with 20 years of experience who was blinded to the pathological result according to ACR TI-RADS for composition, echogenicity, shape, margin, and echogenic foci (1). The reader independently assigned features of every nodule for the five ACR TI-RADS categories. Ultimately, all nodules had feature assignments, resulting in point assignments and corresponding TI-RADS risk classifications for each nodule.

2.3 ROI segmentation, feature extraction and selection

Without any knowledge of the other results, two radiologists (3 and 5 years of experience in thyroid imaging) independently performed the follow-up radiomics analysis. After normalizing the grayscale and voxels, the regions of interest (ROIs) were performed by a 3D Slicer (https://www.slicer.org/) (software version 5.0.2) to manually segment the thyroid nodules on the

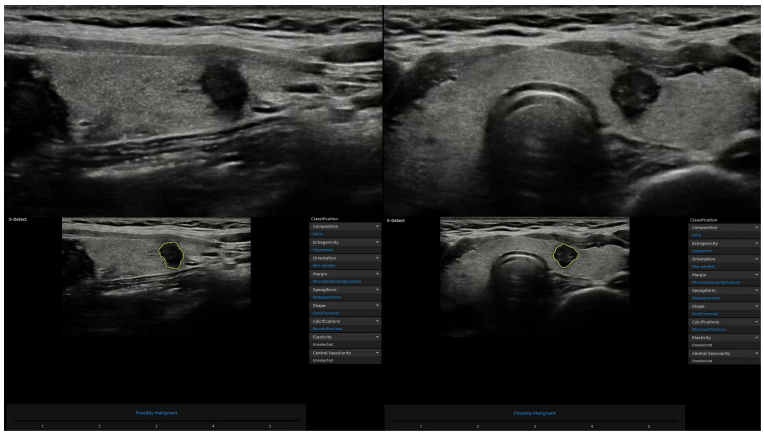


FIGURE 2
Representative thyroid nodule images were acquired with the computer-aided diagnosis (CAD) system.

image in the transverse and longitudinal section. Operators were trained to segment ROIs before the study began. Intra-observer and inter-observer consistency was evaluated with a random cohort of 30 nodules segmented by one of the operators. One month after the first lesion segmentation, two operators completed the re-segmentation of this cohort image. The intraclass correlation coefficient (ICC) was used to assess the reproducibility and robustness of lesion segmentation and feature extraction.

A total of 837 candidate radiomics features were extracted from each ROI using the plug-in "PyRadiomics" package in 3D-Slicer, including features from first-order statistics, gray level co-occurrence matrix (GLCM), gray level dependence matrix (GLDM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), and neighborhood gray level different matrix (NGTDM). All radiomics features were standardized by Z-score transformation to strengthen the data comparability and reduce bias. We only included features with a good agreement (ICC > 0.75). The univariate logistic analysis was performed after the results were obtained to include the features with $p < 0.10$ for further study. Subsequently, the least absolute shrinkage and operator (LASSO) method was used to select radiomics features with excellent predictive properties.

2.4 Establishment of models and performance evaluation

2.4.1 CAD model

S-Detect is a more interactive CAD system based on a specific deep learning algorithm: a convolutional neural network. Deep learning is an intricate multi-layer neural network architecture consisting of input, hidden, and output layers. S-Detect can realize precise decisions and identify benign and malignant nodules by learning a large amount of training data, extracting high-order statistics, and optimizing the balance of input and output data through many hidden layers (15, 16).

2.4.2 Radiomics model

The Rad-Score (radiomics score) for each lesion was computed based on the estimated weighting coefficient of the selected features on each transverse and longitudinal section. Then, the radiomics model was ultimately constructed using this Rad-Score. Moreover, the nomogram was developed by radiomics labels to quantify the possibility of malignancy risk and evaluate high- and low-grade thyroid nodules.

We used the area under the receiver operating characteristic curves (AUCs) and calibration curves to evaluate the performance among the models and the senior radiologist in discriminating between benign and malignant nodules.

2.5 Optimizing the ACR TI-RADS using the models

Based on the nodule's level and maximum diameter, ACR TI-RADS offers three recommendations: no biopsy, US follow-up, or biopsy (4). Both models had binary outputs of high and low malignant risks, and the results were used to upgrade or downgrade ACR TI-RADS management recommendations to explore the possibility of reducing unnecessary biopsies. More specifically, if our assessment indicated a high risk of malignancy, an upgrade was performed, such as no biopsy to follow-up or follow-up to FNA, or FNA remained unchanged; otherwise, when nodules were classified as low risk, we downgraded recommendations. Ultimately, we compared the diagnostic performance of the new risk stratification model with the original ACR TI-RADS recommendations.

2.6 Statistical analysis

The continuous variables were described with the median (interquartile range), and categorical variables were presented as frequencies or percentages. The Student's t-test, chi-square test, and Fisher's exact test were used for the univariate statistical analysis, as appropriate. The AUCs with 95% confidence intervals (CIs) were calculated to assess model performance for classifying benign and malignant thyroid nodules. DeLong's test was employed to analyze between-group differences. Model calibration performance was assessed using calibration curves.

Additionally, the diagnostic value of the management recommendation was evaluated by calculating sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and unnecessary FNA rates (no biopsy and follow-up were considered negative, and the biopsy was positive). We applied the maximum Youden index (sensitivity + specificity − 1) as the optimal cutoff value of the radiomics model to dichotomize all nodules into two groups (high and low risk of malignancy, similar to the CAD model) for discussing potential complementary roles to the ACR guidelines.

Statistical analyses were conducted using the SPSS for Windows version 25.0 (IBM Corporation) and R statistical software version 4.1.30 (R Foundation for Statistical Computing; https://r-project.org). A two-tailed p value < 0.05 was regarded as statistically significant.

3 Results

3.1 Study population

Of the 262 thyroid nodules with complete imaging data and confirmed pathological diagnoses from 148 unique patients (median age, 43 years, 202 women), 157 (59.9%) were malignant, while the remaining 105 (40.1%) were benign (Table 1). Patients of the malignant group were younger and malignant nodules were significantly smaller than benign nodules ($p < 0.001$). There were statistical differences in the ACR TI-RADS level in this cohort. ($p < 0.001$).

3.2 Overall diagnostic performance of the models

Figure 3 demonstrates the receiver operating characteristic (ROC) curves of three models for discriminating malignant and

TABLE 1 Patient demographics and nodule characteristics (stratified by pathologic diagnosis).

Variables	All Nodules (n = 262)	Benign (n = 105)	Malignant (n = 157)	p value
Age (y)	43 (34, 53)	52 (38, 58)	38 (33, 49)	<0.001
Sex	202/60 (77.1/22.9)	83/22 (79.0/21.0)	119/38 (75.8/24.2)	0.643
Location				0.545
Left	126 (48.1)	52 (49.5)	74 (47.1)	
Right	124 (47.3)	50 (47.6)	74 (47.1)	
Isthmus	12 (4.6)	3 (2.9)	9 (5.7)	
Nodule size (mm)	10.0 (6.2, 18.9)	14.7 (7.1, 30.8)	9.0 (5.7, 12.7)	<0.001
ACR TI-RADS level				<0.001
TR1	4 (1.5)	4 (3.8)	0 (0.0)	
TR2	31 (11.8)	30 (28.6)	1 (0.6)	
TR3	14 (5.3)	11 (10.5)	3 (1.9)	
TR4	78 (29.8)	47 (44.8)	31 (19.7)	
TR5	135 (51.5)	13 (12.4)	122 (77.7)	

Data are presented as medians with interquartile ranges in parentheses or number parentheses are persentages. ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System.

benign nodules. The AUCs of the radiomics, ACR TI-RADS, and CAD models were 0.915 (95% confidence interval (CI): 0.881–0.947), 0.849 (95% CI: 0.804–0.894), and 0.814 (95% CI: 0.766–0.863), respectively, as shown in Table 2. Compared with the senior radiologist, the CAD model had a higher sensitivity, and our radiomics model tended towards a higher AUC. The comparative results showed that the radiomics model yielded a higher performance than the ACR TI-RADS and CAD models, and DeLong's test showed that the differences were statistically significant ($p = 0.004$, $p < 0.001$, respectively).

In the radiomics model (Supplementary Material), the two variables, transverse and longitudinal radiomics label scores of thyroid nodules, that were statistically significant in the univariate statistical tests were entered into the model, and then applied to construct the nomogram (Figure 4A). In this visualization, each nodule could obtain predicted risk values for thyroid nodules by summing the scores for each variable. According to the ROC curve, the optimal cutoff value for the "risk of malignant nodules" was 0.656.

Figures 4B, C show the calibration curves of the radiomics and CAD models for predicting thyroid nodules, which illustrates that both models have good agreement between the observed and predicted values.

3.3 The role of management recommendations

The original ACR TI-RADS management recommendations categorized 87 nodules as FNA, and 58 of them were malignant.

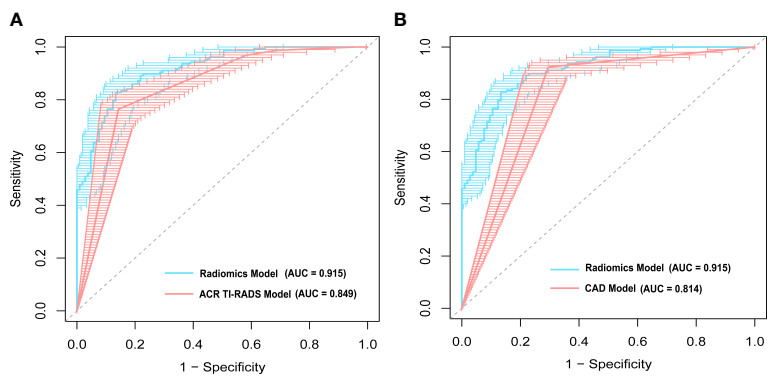


FIGURE 3
A comparison of receiver operating characteristic (ROC) curves between the radiomics model and **(A)** the American College of Radiology Thyroid Imaging Reporting and Data Systems (ACR TI-RADS) model, and **(B)** computer-aided diagnosis (CAD) model. The area under the ROC curve (AUC) was 0.915 for the radiomics model, which was significantly higher than that of the ACR TI-RADS ($p = 0.004$) and CAD models ($p < 0.001$).

TABLE 2 Diagnostic performances comparison of the radiomics, CAD, and ACR TI-RADS models for thyroiuds nodules.

Model	AUC (95%CI)	Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)	p value*
Radiomics	0.915 (0.881-0.947)	83.4 (131/157)	86.7 (91/105)	84.7 (222/262)	90.3 (131/145)	77.8 (91/117)	NA
CAD	0.814 (0.766-0.863)	92,4 (145/157)	70.5 (74/105)	83.6 (219/262)	82.4 (145/176)	86.0 (74/86)	<0.001
ACR TI-RADS	0.849 (0.804-0.894)	77.7 (122/157)	87.6 (92/105)	81.7 (214/262)	90.4 (122/135)	72.4 (92/127)	0.004

Unless otherwise specified, data are presented as AUCs with 95% Cis in brackets, and data are percentages with numerators/denominator in parentheses. CAD, computer-aided diagnosis; ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System; AUC, area under the receiver operating characteristic curve; CI, confidence interval; PPV, positive predictive value; NPV, negatie predictive value; NA, not applicable. *p values reflect the diagnostic efficacy AUC of each model compared to the radiomics model. DeLong's test was used for statistical analysis.

When both models were applied to revise the ACR TI-RADS, our risk stratifications significantly improved the performance. Specifically, the CAD model resulted in the downgrading of 54 nodules (29 from biopsy to follow-up), whereas 118 were upgraded (116 from follow-up to biopsy), and the radiomics model resulted in the downgrading of 83 nodules (39 from biopsy to follow-up), whereas the reassigning from biopsy to follow-up occurred for 97 nodules (Table 3). However, 12 malignant thyroid nodules were missed in the revised CAD model, and 26 tumors were missed in the revised radiomic model. Table 4 shows the diagnostic performance of the original ACR TI-RADS and our revised risk stratification system. Compared with the ACR guidelines, both revised CAD and radiomics recommendations have impressive diagnostic performance, such as higher sensitivity, accuracy, PPV, and NPV, and decreased unnecessary FNA rates. In addition, the improvement scale of the radiomics model in the unnecessary FNA rates was more pronounced (33.3–16.7% vs. 33.3–9.7%). From the perspective of reducing missed diagnoses, the CAD model combined with TI-RADS is more effective.

4 Discussion

In this study, we found that the radiomics model presented with a significantly higher diagnostic accuracy for predicting the malignancy risk of thyroid nodules compared with the CAD model ($p < 0.001$) and a senior radiologist ($p = 0.004$), while the CAD model showed a higher sensitivity (92.4 vs. 83.4, 77.7%). In addition, we applied our systems to revise the ACR TI-RADS management recommendations, especially the radiomics model, successfully optimizing its performance and reducing unnecessary biopsies.

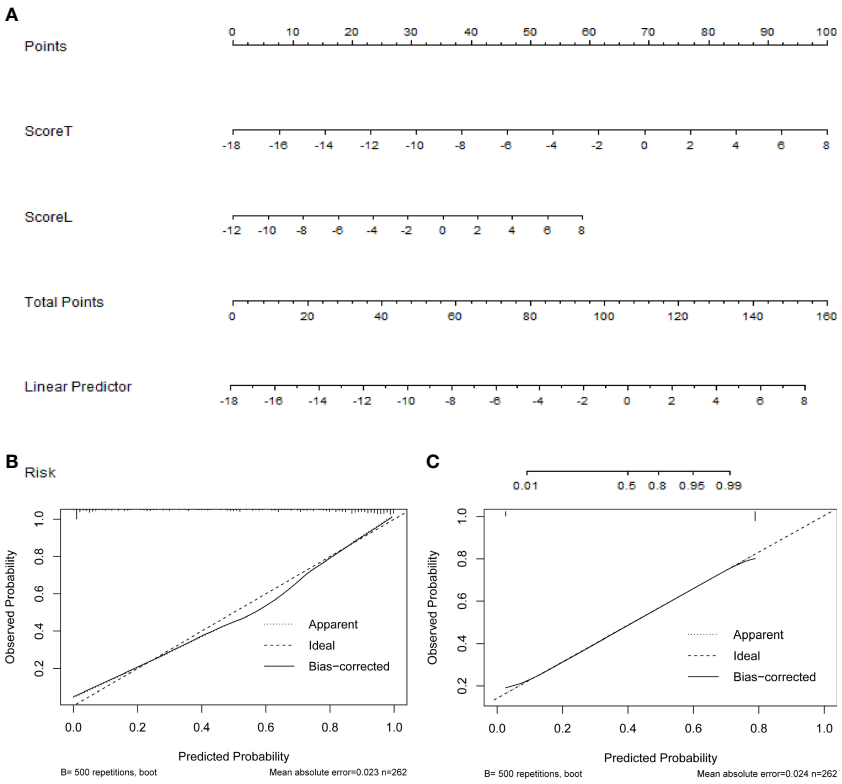


FIGURE 4
(A) The nomogram based on the radiomics model, the calibration curves of **(B)** the radiomics, and **(C)** computer-aided diagnosis (CAD) models.

TABLE 3 Distribution of ACR TI-RADS guidelines revised by our models.

Histopathology	Original ACR TI-RADS			Revised by CAD model			Revised by radiomics model			
	No Biopsy	Follow-up	Biopsy	No Biopsy	Follow-up	Biopsy	No Biopsy	Follow-up	Biopsy	Total
Benign	34	42	29	51	25	29	63	28	14	105
Malignant	1	98	58	6	6	145	15	11	131	157
Total	35	140	87	57	31	174	78	39	145	262

Data are presented as numbers of nodules. ACR TI-Rads, American College of Radiology Thyroid Imaging Reporting and Data System; CAD computer-aided diagnosis.

Our study had some unique characteristics. First, in contrast to most previous studies using the retrospective radiomics strategy, we adopted a prospective design process in which a senior radiologist acquired images and applied strict quality control, thus making all high-quality images more standardized. Second, most S-detect-related studies developed models solely on the transverse section of the lesion for analysis (8, 9, 11, 12). The largest segment (longitudinal section) and its vertical section (transverse section) were chosen in our study to increase the lesion characteristics and reduce the impact of subjective factors. Additionally, the radiomics features of a thyroid nodule were separately extracted from two objective vertical US images, which may provide more detailed information and reflect the tumor heterogeneity.

Radiomics is widely recognized as an important method for medical image analysis in oncology research (17). In this study, we developed a radiomics model for the differentiation of thyroid nodules and constructed a nomogram using the radiomics label. Our model was established through the logistic regression approach, which is the most commonly used supervised learning model in US radiomics (18). The application of radiomics showed adequate diagnostic performance in predicting the malignancy of thyroid nodules with an AUC of 0.915, which was consistent with previously reported studies (19–21). Several studies have reported that the CAD system is a promising approach for solving practical difficulties in clinical diagnosis (8–12, 22). Eun et al. (23) reported a high sensitivity of up to 88.6% and suggested that the CAD system could be useful as decision-making support to rule out cancer. In this study, we also found that the CAD system had a high sensitivity (92.4%) and accuracy (83.6%). The thyroid CAD system used in this study was integrated into the US system, which enabled the use of CAD system in real-time clinical practice. Furthermore, a real-time second opinion on the decision for the necessity of FNA is possible with the present system. Due to its simplicity and reduced analysis time, this system would be simpler to apply in routine practice (8). Therefore, we concluded that the CAD system could reduce the time required for the interpretation process of thyroid nodules and diagnose them as benign or malignant, making it a simple system to screen thyroid nodules for high sensitivity. The radiomics and CAD models constructed in our study showed good robustness and also illustrated the strong generalization ability of our method.

The ACR TI-RADS is based on an expert consensus, literature review, and partial analysis of the database of proven nodules; its core objective is to focus on clinically significant thyroid cancers and reduce the FNA of benign nodules (4). Wildman et al. (7) used genetic algorithms to improve the performance of artificial intelligence TI-RADS by optimization of the points assigned to each TI-RADS feature, which can validate the ACR TI-RADS while improving specificity and maintaining sensitivity. In our study, we attempted to explore the potential complementary role of radiomics and CAD models to ACR TI-RADS Risk Stratification for thyroid nodule management; the results showed that both models could provide additional gains in performance, especially in terms of sensitivity and accuracy (Table 4). Notably, both revised models successfully reduced unnecessary biopsies compared with the ACR TI-RADS, especially the radiomics model. This may support that the radiomics strategy can capture information that is beyond visual interpretation and interpret heterogeneity within lesions. Using quantitative information on radiomics features could be more effective as a complementary tool to management recommendations. On the other hand, although the S-detect model is based on a deep learning algorithm generated using a large database, the algorithm relies on the quality of the annotated US image features, which will inevitably depend on the reader's experience. In addition, the deep learning method may suffer from possible over-fitting. In summary, we recommended that radiologists appropriately optimize the ACR TI-RADS risk stratification system with the assistance of new technologies.

Our study had several limitations. First, this study did not include any large-scale test datasets to validate. Thus, it will be necessary to conduct a more stringent internal and external validation with a larger sample size representing the screening population. Second, this study only used static vertical section

TABLE 4 The effects compared among original ACR TI-RADS management recommendations and revised diagnoses based on our models.

Model	Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)	Unnecessary FNA rate (%)
Original ACR TI-RADS	36.9 (56/157)	72.4 (76/105)	51.1 (134/262)	66.7 (58/87)	43.4 (76/175)	33.3 (29/87)
Revised by CAD model	92.4 (145/157)	72.4 (76/105)	84.4 (221/262)	83.3 (145/174)	86.4 (76/88)	16.7 (29/174)
Revised by radiomics model	83.4 (131/157)	86.7 (91/105)	84.7 (222/262)	90.3 (131/145)	77.8 (91/117)	9.7 (14/145)

Unless otherwise specified, data percentages with numerator/denominator in parentheses. ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System; CAD, computer-aided diagnosis; PPV, positive predictive value; NPV, negatie predictive value, FNA, fine-needle aspiration.

images. In future studies, model evaluation using cine clips that include the entire thyroid nodule and surrounding thyroid parenchyma may be necessary to avoid losing the risk features for malignancy. Third, thyroid nodules have various histological subtypes with different molecular mechanisms, grades of malignancy, clinical aggressiveness, and imaging appearances (24). The low occurrence rate of non-papillary carcinoma determines a relatively low percentage in our study. Future efforts will be warranted to include a larger sample size with varied pathological types to further enhance the generalizability and performance.

In conclusion, our study provides evidence that the radiomics strategy and CAD system both have the potential to predict malignancy in thyroid nodules and suggests a simple method to optimize the ACR TI-RADS recommendation. This approach finds the potential complementary roles of both models to the guidelines, which can more precisely help in the classification of thyroid nodules and successfully reduce unnecessary biopsies, especially the radiomics model, which is recommended due to its lower unnecessary FNA rates.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the ethical review committee of the Third Xiangya Hospital of Central South University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

Author contributions

MX designed the research, analyzed, and drafting of the manuscript. FS contributed to the analysis and provided statistical advice for this manuscript. PZ collected data and reviewed the manuscript. YZ and YW analyzed data. YX provided instructive advice and supervision. All authors contributed to the article and approved the submitted version.

Funding

This study has received funding from the National Natural Science Foundation of China (No. 81871367), the Natural Science Foundation of Hunan Province, China (No. 2021JJ31037 and No. 2022JJ30894), and the Project of Hunan Provincial Health Commission (No. 202209025123).

Acknowledgments

Thanks to all members of the Department of Thyroid Surgery, The Third Xiangya Hospital of Central South University for their support in the manuscript preparation.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2023.1140816/full#supplementary-material

References

1. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid* (2016) 26(1):1–133. doi: 10.1089/thy.2015.0020

2. Guth S, Theune U, Aberle J, Galach A, Bamberger CM. Very high prevalence of thyroid nodules detected by high frequency (13 mhz) ultrasound examination. *Eur J Clin Invest* (2009) 39(8):699–706. doi: 10.1111/j.1365-2362.2009.02162.x

3. Ko SY, Kim EK, Sung JM, Moon HJ, Kwak JY. Diagnostic performance of ultrasound and ultrasound elastography with respect to physician experience. *Ultrasound Med Biol* (2014) 40(5):854–63. doi: 10.1016/j.ultrasmedbio.2013.10.005

4. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS committee. *J Am Coll Radiol* (2017) 14(5):587–95. doi: 10.1016/j.jacr.2017.01.046

5. Tessler FN, Middleton WD, Grant EG. Thyroid imaging reporting and data system (TI-RADS): a user's guide. *Radiology* (2018) 287(1):29–36. doi: 10.1148/radiol.2017171240

6. Persichetti A, Di Stasio E, Coccaro C, Graziano F, Bianchini A, Di Donna V, et al. Inter- and intraobserver agreement in the assessment of thyroid nodule ultrasound features and classification systems: a blinded multicenter study. *Thyroid* (2020) 30 (2):237–42. doi: 10.1089/thy.2019.0360

7. Wildman-Tobriner B, Buda M, Hoang JK, Middleton WD, Thayer D, Short RG, et al. Using artificial intelligence to revise ACR TI-RADS risk stratification of thyroid nodules: diagnostic accuracy and utility. *Radiology* (2019) 292(1):112–9. doi: 10.1148/radiol.2019182128

8. Choi YJ, Baek JH, Park HS, Shim WH, Kim TY, Shong YK, et al. A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of thyroid nodules on ultrasound: initial clinical assessment. *Thyroid* (2017) 27(4):546–52. doi: 10.1089/thy.2016.0372

9. Kim HL, Ha EJ, Han M. Real-world performance of computer-aided diagnosis system for thyroid nodules using ultrasonography. *Ultrasound Med Biol* (2019) 45 (10):2672–8. doi: 10.1016/j.ultrasmedbio.2019.05.032

10. Xia S, Yao J, Zhou W, Dong Y, Xu S, Zhou J, et al. A computer-aided diagnosing system in the evaluation of thyroid nodules-experience in a specialized thyroid center. *World J Surg Oncol* (2019) 17(1):210. doi: 10.1186/s12957-019-1752-z

11. Barczyński M, Stopa-Barczyńska M, Wojtczak B, Czarniecka A, Konturek A. Clinical validation of s-Detect^{Tm} mode in semi-automated ultrasound classification of thyroid lesions in surgical office. *Gland Surg* (2020) 9(Suppl 2):S77–85. doi: 10.21037/gs.2019.12.23

12. Wei Q, Zeng SE, Wang LP, Yan YJ, Wang T, Xu JW, et al. The value of s-detect in improving the diagnostic performance of radiologists for the differential diagnosis of thyroid nodules. *Med Ultrason* (2020) 22(4):415–23. doi: 10.11152/mu-2501

13. Arimura H, Soufi M, Kamezawa H, Ninomiya K, Yamada M. Radiomics with artificial intelligence for precision medicine in radiation therapy. *J Radiat Res* (2019) 60 (1):150–7. doi: 10.1093/jrr/rry077

14. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169

15. Zhao C, Xiao M, Jiang Y, Liu H, Wang M, Wang H, et al. Feasibility of computer-assisted diagnosis for breast ultrasound: the results of the diagnostic performance of s-detect from a single center in China. *Cancer Manag Res* (2019) 11:921–30. doi: 10.2147/cmar.S190966

16. Zhang D, Jiang F, Yin R, Wu GG, Wei Q, Cui XW, et al. A review of the role of the s-detect computer-aided diagnostic ultrasound system in the evaluation of benign and malignant breast and thyroid masses. *Med Sci Monit* (2021) 27:e931957. doi: 10.12659/msm.931957

17. Limkin EJ, Sun R, Dercle L, Zacharaki EI, Robert C, Reuzé S, et al. Promises and challenges for the implementation of computational medical imaging (Radiomics) in oncology. *Ann Oncol* (2017) 28(6):1191–206. doi: 10.1093/annonc/mdx034

18. Jiang M, Li C, Tang S, Lv W, Yi A, Wang B, et al. Nomogram based on shear-wave elastography radiomics can improve preoperative cervical lymph node staging for papillary thyroid carcinoma. *Thyroid* (2020) 30(6):885–97. doi: 10.1089/thy.2019.0780

19. Liang J, Huang X, Hu H, Liu Y, Zhou Q, Cao Q, et al. Predicting malignancy in thyroid nodules: radiomics score versus 2017 American college of radiology thyroid imaging, reporting and data system. *Thyroid* (2018) 28(8):1024–33. doi: 10.1089/thy.2017.0525

20. Galimzianova A, Siebert SM, Kamaya A, Rubin DL, Desser TS. Quantitative framework for risk stratification of thyroid nodules with ultrasound: a step toward automated triage of thyroid cancer. *AJR Am J Roentgenol* (2020) 214(4):885–92. doi: 10.2214/ajr.19.21350

21. Park VY, Lee E, Lee HS, Kim HJ, Yoon J, Son J, et al. Combining radiomics with ultrasound-based risk stratification systems for thyroid nodules: an approach for improving performance. *Eur Radiol* (2021) 31(4):2405–13. doi: 10.1007/s00330-020-07365-9

22. Szczepanek-Parulska E, Wolinski K, Dobruch-Sobczak K, Antosik P, Ostalowska A, Krauze A, et al. S-detect software vs. EU-TIRADS classification: a dual-center validation of diagnostic performance in differentiation of thyroid nodules. *J Clin Med* (2020) 9(8):2495. doi: 10.3390/jcm9082495

23. Jeong EY, Kim HL, Ha EJ, Park SY, Cho YJ, Han M. Computer-aided diagnosis system for thyroid nodules on ultrasonography: diagnostic performance and reproducibility based on the experience level of operators. *Eur Radiol* (2019) 29 (4):1978–85. doi: 10.1007/s00330-018-5772-9

24. Prete A, Borges de Souza P, Censi S, Muzza M, Nucci N, Sponziello M. Update on fundamental mechanisms of thyroid cancer. *Front Endocrinol (Lausanne)* (2020) 11:102. doi: 10.3389/fendo.2020.00102

frontiers | Frontiers in Endocrinology

TYPE Systematic Review
PUBLISHED 12 May 2023
DOI 10.3389/fendo.2023.1187935

OPEN ACCESS

EDITED BY
Andrea Frasoldati,
Endocrine Unit ASMN, Italy

REVIEWED BY
Gerdi Tuli,
Regina Margherita Hospital, Italy
Cristiane Jeyce Gomes-Lima,
MedStar Health Research Institute (MHRI),
United States

*CORRESPONDENCE
Anping Su
suanpingping@126.com
Wenshuang Wu
wenshuang_wu@163.com

SPECIALTY SECTION
RECEIVED 16 March 2023
ACCEPTED 26 April 2023
PUBLISHED 12 May 2023

CITATION
Xing Z, Qiu Y, Zhu J, Su A and Wu W (2023) Diagnostic performance of adult-based ultrasound risk stratification systems in pediatric thyroid nodules: a systematic review and meta-analysis.
Front. Endocrinol. 14:1187935.
doi: 10.3389/fendo.2023.1187935

Diagnostic performance of adult-based ultrasound risk stratification systems in pediatric thyroid nodules: a systematic review and meta-analysis

Zhichao Xing[1,2], Yuxuan Qiu[3], Jingqiang Zhu[1], Anping Su[1]* and Wenshuang Wu[1,2]*

[1]Center of Thyroid and Parathyroid Surgery, West China Hospital, Sichuan University, Chengdu, China, [2]Laboratory of Thyroid and Parathyroid Disease, Frontiers Science Center for Disease-related Molecular Network, West China Hospital, Sichuan University, Chengdu, China, [3]Department of Ultrasound, West China Hospital, Sichuan University, Chengdu, China

Purpose: Ultrasound (US) is the first choice in the detection of thyroid nodules in pediatric and adult patients. The purpose of this study was to evaluate the diagnostic performance of adult-based US risk stratification systems (RSSs) when applied to the pediatric population.

Methods: Medline, Embase, and Cochrane Library (CENTRAL) were searched up to 5 March 2023 for studies about the diagnostic performance of adult-based US RSS in pediatric patients. The pooled sensitivity, specificity, positive likelihood ratio (LR), negative LR, and diagnostic odds ratio (DOR) were calculated. The summary receiver operating characteristic (SROC) curves and area under the curve (AUC) were also analyzed.

Results: The sensitivity was highest in American College of Radiology-Thyroid Imaging Reporting and Data System (ACR-TIRADS) category 4–5 and American Thyroid Association RSS high-intermediate risk (ATA), which was 0.84 [0.79, 0.88] and 0.84 [0.75, 0.90], respectively. The specificity was highest in ACR-TIRADS category 5 and Europe-TIRADS (EU-TIRADS) category 5, which was 0.93 [0.83, 0.97] and 0.93 [0.88, 0.98], respectively. The ACR-TIRADS, ATA, and EU-TIRADS showed moderate diagnostic performance in pediatric thyroid nodule patients. For Korea-TIRADS (K-TRADS) category 5, the summary sensitivity and specificity with a 95% CI were 0.64 [0.40, 0.83] and 0.84 [0.38, 0.99], respectively.

Conclusions: In conclusion, the ACR-TIRADS, ATA, and EU-TIRADS have moderate diagnostic performance in pediatric thyroid nodule patients. The diagnostic efficacy of the K-TIRADS was not as high as expected. However, the diagnostic performance of Kwak-TIRADS was uncertain because of the small sample size and small number of studies included. More studies are needed to evaluate these adult-based RSSs in pediatric patients with thyroid nodules. RSSs specific for pediatric thyroid nodules and thyroid malignancies were necessary.

KEYWORDS

pediatric thyroid nodules, risk stratification systems, ultrasonography, diagnostic performance, meta-analysis

Introduction

Thyroid cancer is the most common pediatric endocrine cancer and presents a diagnostic challenge in pediatric populations. The reported prevalence of thyroid nodules is 3.1% in adolescents (1). However, the malignancy rate is estimated to be 22–26% in children with thyroid nodules and 5–10% in adults (2–5). Furthermore, pediatric patients are more likely to present with cervical lymph node metastases (40–80%) and distant metastases (20–30%) such as pulmonary metastases than adults (6, 7). Therefore, early and accurate diagnosis in children is extremely important.

Neck ultrasound (US) is the first choice in the detection of thyroid nodules in pediatric and adult patients (8–10). Adult-based neck US risk stratification systems (RSSs) have been developed in recent years to integrate US features and improve diagnostic accuracy as an aid in the stratification of the risk of malignancy, such as the American College of Radiology–Thyroid Imaging Reporting And Data System (ACR-TIRADS), American Thyroid Association Ultrasound Risk Stratification Systems (ATA RSS), European Thyroid Imaging and Reporting Data System (EU-TIRADS), Korean Thyroid Imaging Reporting and Data System (K-TIRADS), and Kwak Thyroid Imaging Reporting and Data System (Kwak-TIRADS) (11–15).

Korean Professor Jin Young Kwak was the first in the world to propose a practical TIRADS to categorize thyroid nodules and stratifying their risk of malignancy in 2011, which we called Kwak-TIRADS now (15). He suggested that the following US features showed a significant association with malignancy: solid component, hypoechogenicity, marked hypoechogenicity, microlobulated or irregular margins, microcalcifications, and taller-than-wide shape. Risk stratification of thyroid malignancy by using the number of suspicious US features allows for a practical and convenient Kwak-TIRADS.

However, we do not have any formalized, US-based RSS in pediatrics. Recently, a few studies have reported the utility of these adult-based RSSs in pediatric patients. However, pediatric thyroid cancers are different in clinical, molecular, and pathologic characteristics from those in adults. These RSSs depend significantly on nodule size, while thyroid volume increases with age, and nodule size is not predictive of malignancy in pediatric patients. Therefore, the appropriateness of these RSSs remains to be explored when applied to pediatric patients.

Therefore, the purpose of this study was to evaluate the diagnostic performance of the adult-based RSS when applied to the pediatric population and provide information to guide future clinical practice.

Methods

The meta-analysis was reported in accordance with the instructions of the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) extension statement incorporating network meta-analyses (16, 17).

Search strategy

The Medline, Embase, and Cochrane Controlled Register of Trials (CENTRAL) and Web of Science databases were searched up to March 5, 2023. The search terms to retrieve related studies were as follows: [(thyroid) AND (thyroid imaging reporting and data system)] OR [(thyroid image reporting and data system) OR (TIRADS) OR (TI-RADS) OR (RSS) OR (guideline)] AND [(pediatric) OR (adolescent) OR (child) OR (children)]. Two investigators independently checked retrieved articles blinded to the journal, author, and so on. All abstracts to obtain possible applicable articles and the full text were screened to determine the final eligible articles. Relevant reviews and their reference list were also checked. Discrepancies were resolved by discussion with another investigator.

Inclusion criteria

(a) The study was based on the diagnostic performance of adult-based ultrasound RSS, such as ACR-TIRADS, ATA, EU-TIRADS, K-TIRADS, and Kwak-TIRADS. (b) The patients were pediatric with thyroid nodules. (c) The reference standard was based on pathological diagnosis or imaging follow-up. (d) Data available for sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and diagnostic accuracy. (e) The language was limited to English.

Exclusion criteria

(a) Letters, editorials, conference abstracts, and review articles. (b) The topics of articles were not about the diagnostic performance of adult-based ultrasound RSS. (c) The patients were not pediatric. (d) If studies had an overlapping population, we included the study with the largest population and excluded others.

Data extraction

The eligible articles were reviewed, and the relevant data were extracted using a standardized form. (a) Study characteristics: first author, year of publication, country or region, study period, study design, sample size, and reference standard; (b) Patient characteristics: number of patients, mean age, and male-to-female ratio; (c) Diagnostic performance: numbers of total thyroid nodules, numbers of true positive (TP), true negative (TN), false positive (FP), false negative (FN) thyroid nodules, sensitivity, specificity, PPV, NPV, and diagnostic accuracy; (d) Standard reference: biopsy pathology, surgery pathology, and follow-up; (e) US examinations: US model and vendor, number of readers, and experience.

Quality assessment

Two reviewers assessed the quality of the included articles independently using Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) (18), and disagreement was resolved by discussion. This tool is composed of four domains: patient selection, index test, reference standard, flow, and timing. Each domain is

assessed according to bias. Risk of bias was judged as "low," "high," or "unclear." The first three domains are assessed in terms of concerns regarding applicability.

Statistical analysis

Statistical analysis was mainly performed using Stata version 15.0 software (StataCorp, LLC; College Station, TX). A value of $p < 0.05$ was taken to indicate statistical significance.

The pooled sensitivity, specificity, positive likelihood ratio (LR), negative LR, and diagnosis odds ratio (DOR), each has 95% confidence intervals (95% CI), were calculated using a bivariate random-effects model, and a coupled forest plot was constructed. In addition, a hierarchical summary receiver operating characteristics (HSROC) curve with 95% confidence and prediction regions was plotted and area under the curve (AUC) was also analyzed. The criteria for the positive test results were set to be (a) RSS category 5 or (b) RSS category 4 or 5. For example, if we set category 5 as a cutoff value, TP nodules indicated the nodules classified as category 5 on US and turned out to be malignant. We followed the reference standard set in each study.

Heterogeneity was assessed using the Higgins inconsistency index (I^2) test with a value > 50%, indicating the presence of heterogeneity, and a coupled forest plot was used to graphically assess the presence of a threshold effect (a positive correlation between sensitivity and false-positive rate among the selected studies). We regarded $I^2 > 50\%$ or P-value of Q-test < 0.05 as high heterogeneity. Among the potential covariates such as sample size, region, standard reference of malignant nodules, and standard reference of benign nodules, we compared "sample size more than median" vs. "sample size less than median," "America studies" vs. "Europe studies," "surgery and/or biopsy pathology" vs. "surgery pathology" for malignant nodules, "surgery and/or biopsy pathology and follow-up" vs. "surgery and/or biopsy pathology" for benign nodules.

Results

Literature search

The details of article screening procedures were as Figure 1. A total of 940 articles were generated using search terms mentioned above, and 524 were removed because of duplications. We excluded 387 that did not meet the topic of our study, and four letters, editorials, conference abstracts, review articles after reviewing the titles and abstracts. The remaining 25 articles were screened for eligibility seriously, and five were abandoned because the assessment of diagnosis performence is based on adult patient and one study had an overlapping population. Finally, the remaining 19 studies were included in our meta-analysis.

Characteristics of studies

The characteristics of included studies were detailed in Table 1 (19–37). All the studies were retrospective. The overall study period was from 1996 to 2021. A total of 1,927 pediatric thyroid cancer patients and 2,263 modules were included. Ages ranged from 0.9 to

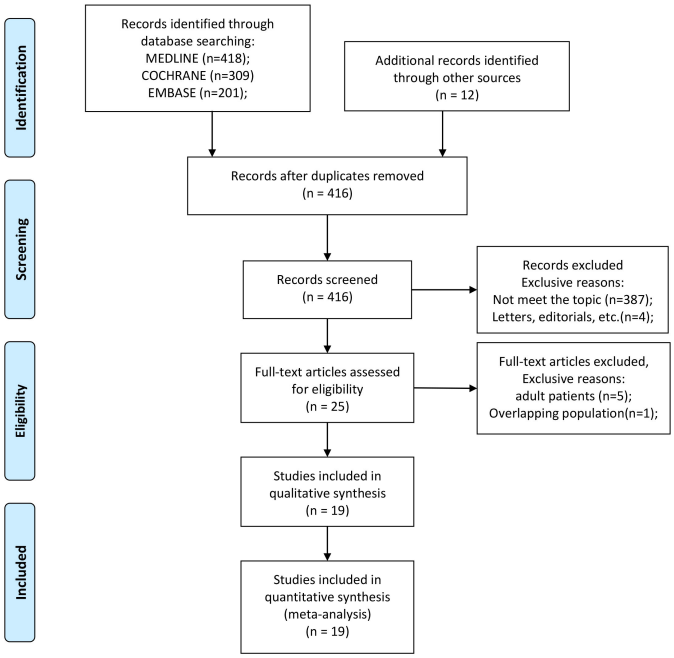


FIGURE 1
PRISMA diagram of study selection.

10.3389/fendo.2023.1187935

TABLE 1 Characteristics of the included studies.

First author	Year	Study period	Country	Study design	No. of patients	M	F	Age	Range/ SD	Inclusion age (years)
Lim-Dunham	2017	1996–2016	USA	R	33	5	28	16 in benign; 16.5 in malignant	<=18	
Creo	2018	1996–2015	USA	R	112	16	96	15.5	15.5 ± 3.2	<=18
Martinez-Rios	2018	Jan. 1992–Oct. 2015	Canada	R	124	40	84	13.6	3.3–17.7	<21
Zaltsberg	2019	Aug. 2007–Aug. 2017	Canada	R	75	16	59	13.4	3–18	<18
Lim-Dunham	2019	1996–2017	USA	R	62	6	56	12.5 in male; 16.5 in female	<=18	
Polat	2019	2015–2018	Turekey	R	106	35	71	11.4	1–17	<18
Uner	2019	NA	Turekey	R	64	10	58	15.2	3–18	<=18
Richman	2020	Jan. 2004–Jul. 2017	USA	R	314	54	260	14.9	14.9 ± 2.7	<=18
Arora	2020	2008–2018	USA	R	20	4	16	14.9	7–22	<=22
Scappaticcio	2021	Jan. 2017–Mar. 2021	Italy	R	36	10	26	15	11–17	<19
Piccardo	2021	Jan. 2012–Dec. 2017	Italy	R	52	20	32	17	15–18	<=18
Ahmad	2021	Jan. 2015–Mar. 2019	USA	R	115	25	90	15.5	5.0–20.2	<=21
Fernández	2021	2005–2020	Spain	R	24	6	18	15.3	13.3–17.3	<18
Lee	2021	Aug. 2007–Feb. 2020	Korea	R	107	24	83	13.9	4–18	<19
Tuli	2022	2000–2020	Italy	R	200	81	119	12	2–18	<18
Yang	2022	Jan. 2004–Sept. 2020	USA	R	139	20	119	17.5	15.3–19.3	<=21
Borysewicz-Sanczyk	2022	NA	Poland	R	17	4	13	15.3	5–18	5–18
Daniels	2022	2007–2018	USA	R	106	20	86	15.6	0.9–18.8	<19

(Continued)

TABLE 1 Continued

First author	Year	Study period	Country	Study design	No. of patients	M	F	Age	Range/ SD	Inclusion age (years)
Kim	2022	Jan. 2000–Apr. 2020	Korea	R	221	49	172	16	13–17	<=21

No. of nodules	Mal	Ben	Median/mean nodule size (cm)		RSS	reference standard		US model (vendor)	Interpretation	
			Ben	Mal		Ben	Mal		No. of readers	Experience (y)
33	12	21	2.1	2.55	ATA	s/b	s	Acuson Sequoia 512, XP128, Aspen (Siemens), Logic E9 (GE)	2	>10
145	50	95	NA	NA	ATA	s/b/f (1 year)	s/b	NA	2	sum>27
123	52	71	2.75		ATA, Kwak	s/b/f (2 years)	s/b	iU22 (Philips), Aplio (Toshiba)	3	2-37
300	52	248	NA	NA	ACR, Kwak	s/b/f (1 year)	s/b	iU22 (Philips), Aplio 500 (Toshiba)	4	5-20
33	12	21	1.9		ACR	s/b	s	Acuson Sequoia 512, XP128, Aspen (Siemens), Logic E9 (GE)	2	>10
105	5	100	0.74		ACR	s/b/f (1 year)	s/b	Aplio 500 (Toshiba), RS80A with Prestige (Samsung Medison)	2	3-7
68	19	49	0.8		ACR	s/b	s/b	iU22 (Philips), Apolio (Toshiba)	NA	NA
404	77	327	NA	NA	ACR	s/b	s/b	Acuson Sequoia (Siemens), Logiq E9 (GE)	4	6-33
20	7	13	NA	NA	ACR	s/b/f (2 years)	s/b	NA	NA	NA
41	12	29	10 (7-13)		ACR, EU, K, ATA	s/b	s	MyLabTMSix, Esaote	3	NA
52	14	38	13 (11-12)		ACR, EU, ATA	s/b	s	LOGIQ S8 (General Electric Medical Systems)	3	NA
138	10	128	NA	NA	ACR, PED, ATA	s/b	s/b	NA	2	NA
19	7	12	19 (9-36)	22.2 (15-34)	EU	s/b	s/b	NA	2	NA
133	62	71	NA	NA	K	s/b	s/b	Aplio XG (Toshiba), iU22 (Philips), Aixplorer (SuperSonic), Logiq E9 (GE)	2	6-8
200	26	174	8 (8-10)	24 (7-60)	ACR, EU	s/b	s/b	NA	3	NA
139	56	83	2.4 (1.6-3.7)		ACR	s/b	s/b	NA	3	1-30
16	5	11	2.0–22.6 (9.9 ± 6.95)	4.5–19.0 (13.1 ± 5.86)	ATA, BTA	s/b	s/b	Apolio (Toshiba)	1	NA

(Continued)

TABLE 1 Continued

First author	Year	Study period	Country	Study design	No. of patients	M	F	Age	Range/ SD	Inclusion age (years)
106	59	47	NA	NA	ACR	s/b/f (2 years)	s/b	NA	2	4-11
221	135	86	NA	NA	ACR, ATA, EU, K, AACE/ACE/ AME	s/b	s/b	iU22 and EPIQ 5 (Philips), Aplio XG (Toshiba)	3	1-8

R, retrospective study; ACR, American College of Radiology TIRADS; ATA, American Thyroid Association Ultrasound Risk Stratification Systems; EU, European TIRADS; K, Korean TIRADS; Kwak, Kwak TIRADS; PED, Pediatric TIRADS; BTA, British Thyroid Association Ultrasound Risk Stratification Systems; s, surgery pathology; b, biopsy pathology; f, follow-up; NA, not available.

22 years. Male patients accounted for 23.1% and female for 76.9%. This study included 660 malignant nodules and 1,603 benign nodules. All malignant nodules and most of benign nodules in the included studies have been diagnosed by surgical pathology or biopsy. Benign nodules included only in six studies were diagnosed by biopsy pathology, surgery pathology, or at least 1 year of follow-up (19–21, 27, 31, 35).

Quality assessment

The overall quality of the included studies assessed by QUADAS-2 was moderate. Five articles satisfied six of the seven items, and nine articles satisfied five items. The details are shown in Figure 2. Thirteen studies had an unclear risk of bias in patient selection. Consecutive enrollment was not clarified in 10 studies (20, 22, 24, 25, 27, 31, 32, 35–37). Martinez-Rios et al. included thyroid nodules measuring more than 10 mm (21). Tuli et al. included thyroid nodules measuring more than 5 mm (33). Piccardo et al. included patients treated with radiotherapy for nonthyroidal cancers (29). No study had an unclear risk of bias in the index test domain because of blinding to the reference standard during the US examinations. All studies had an unclear risk of bias in the reference standard domain because of no or unclear blinding to the index test during pathologic evaluation. Six studies had an unclear risk of bias in the flow and timing domain because of inconsistency or unclear consistency on the reference standard for diagnosing benign nodules across the study population

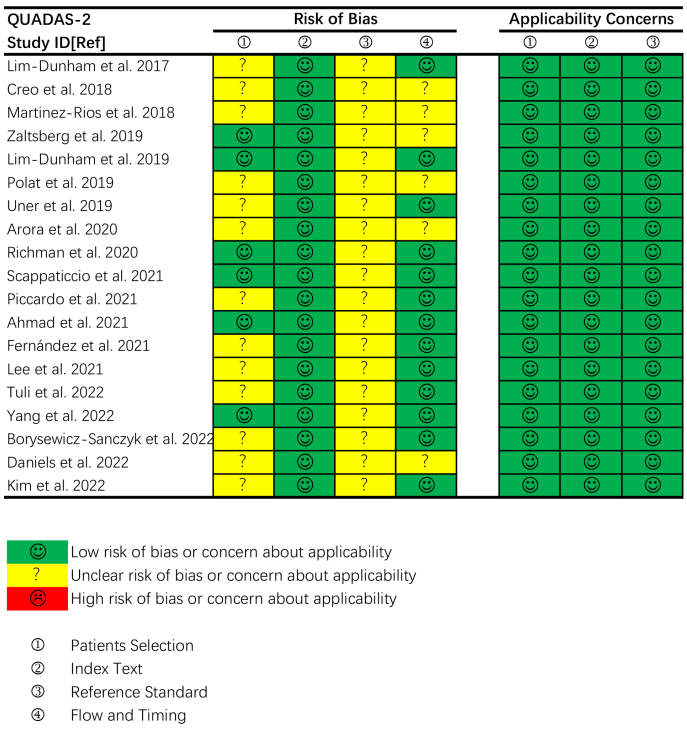


FIGURE 2
Quality assessment of the included studies according to the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) criteria.

(19–21, 24, 27, 35). There were no concerns regarding the applicability of the patient selection, index test and reference standard.

Diagnostic performance

The diagnostic performance and AUC of ACR-TIRADS, ATA system, EU-TIRADS, K-TIRADS and Kwak-TIRADS was synthesized in Table 2.

Diagnostic performance of ACR-TIRADS

Thirteen studies including 1,868 nodules were pooled to analyze the diagnostic performance of ACR-TIRADS category 5 (ACR 5). As shown in Figure 3, the summary sensitivity and specificity with a 95% CI were 0.57 [0.41,0.71] and 0.93 [0.83, 0.97], respectively. For ACR-TIRADS category 4 or 5 (ACR 4-5), 10 studies including 1,486 nodules were pooled and analyzed, and the sensitivity and specificity were 0.84 [0.79, 0.88] and 0.61 [0.49, 0.72], respectively. The AUC was 0.82 [0.79, 0.85] for ACR 5 and 0.85 [0.81, 0.87] for ACR 4-5, shown in Figure 4.

Diagnostic performance of the ATA system

Eight studies including 773 ATA high-risk nodules were pooled and analyzed, and the summary sensitivity and specificity were 0.73 [0.65, 0.79] and 0.73 [0.43, 0.91], respectively. For ATA high-intermediate risk, six studies including 410 nodules were pooled and analyzed. The sensitivity and specificity were 0.84 [0.75, 0.90] and 0.55 [0.40, 0.70], respectively. The details are shown in Figure 5. The AUC was 0.76 [0.72, 0.79] for ATA high risk and 0.82 [0.78, 0.85] for ATA high-intermediate risk, shown in Figure 6.

Diagnostic performance of EU-TIRADS

Three studies including 293 nodules of EU-TIRADS category 5 (EU 5) were pooled and analyzed. The summary sensitivity was 0.45 [0.17, 0.76], and the specificity was 0.93 [0.88, 0.98]. For EU-TIRADS category 4 or 5 (EU 4-5) shown in Figure 7, five studies including 533 nodules were pooled and analyzed. The sensitivity and specificity were 0.78 [0.68, 0.86] and 0.48 [0.36, 0.61], respectively. The AUC of EU TR5 was 0.70 [0.33, 0.94], and the AUC of EU 4-5 (Figure 8) was 0.71 [0.67, 0.75].

Diagnostic performance of K-TIRADS

Only three studies including 385 nodules were pooled to analyze the diagnostic performance of the K-TIRADS category (K 5). The summary sensitivity and specificity with a 95% CI were 0.64 [0.40, 0.83] and 0.84 [0.38, 0.99], respectively. The AUC was 0.56 [0.06, 0.95].

Diagnostic performance of Kwak-TIRADS

Only two studies including 423 nodules were pooled to analyze the diagnostic performance of the Kwak-TIRADS. For Kwak 5, the pooled sensitivity and specificity were 0.10 [0.04, 0.18] and 0.99 [0.97, 0.99], respectively. For Kwak 4-5, the sensitivity and specificity were 0.99 [0.99, 1.00] and 0.33 [0.11, 0.63], respectively. The AUC was 0.09 [0.05, 0.14] for Kwak 5 and 0.48 [0.03, 0.94] for Kwak 4-5.

Meta-regression analysis

The details are shown in Table 3. Sample size and region might be the heterogeneous sources of specificity of the ACR 5 category. The region and standard reference for benign nodules might be the heterogeneous sources of specificity of the ACR 4-5 category. Sample size and standard reference for malignant nodules could lead to the heterogeneous specificity of the ATA high-risk category. Region resulted in the heterogeneous sensitivity of the ATA high-intermediate risk category. No potential heterogeneous source was found in the EU 4-5 category.

TABLE 2 Diagnostic performance of different RSS.

RSS	No. of studies	Sensitivity	Specificity	DOR	AUC	LR+	LR-
ACR 5	13	0.57 [0.41, 0.71]	0.93 [0.83, 0.97]	17 [7, 37]	0.82 [0.79, 0.85]	7.7 [3.5, 17.0]	0.47 [0.34, 0.64]
ACR 4-5	10	0.84 [0.79, 0.88]	0.61 [0.49, 0.72]	8 [5, 14]	0.85 [0.81, 0.87]	2.2 [1.6, 2.9]	0.26 [0.20, 0.35]
ATA high	8	0.73 [0.65, 0.79]	0.73 [0.43, 0.91]	7 [2, 22]	0.76 [0.72, 0.79]	2.7 [1.1, 6.7]	0.37 [0.28, 0.49]
ATA high-intermediate	6	0.84 [0.75, 0.90]	0.55 [0.40, 0.70]	7 [4, 12]	0.82 [0.78, 0.85]	1.9 [1.4, 2.6]	0.28 [0.19, 0.42]
EU 5	3	0.45 [0.17, 0.76]	0.93 [0.88, 0.98]	17 [11, 27]	0.70 [0.33, 0.94]	10.9 [6.3, 17.9]	0.61 [0.56, 0.65]
EU 4-5	5	0.78 [0.68, 0.86]	0.48 [0.36, 0.61]	3 [2, 5]	0.71 [0.67, 0.75]	1.5 [1.2, 1.9]	0.45 [0.33, 0.62]
K 5	3	0.64 [0.40, 0.83]	0.84 [0.38, 0.99]	195 [84, 386]	0.56 [0.06, 0.95]	59.4 [16.8, 123.7]	0.29 [0.16, 0.40]
Kwak 5	2	0.10 [0.04, 0.18]	0.99 [0.97, 0.99]	43 [21, 66]	0.09 [0.05, 0.14]	3.4 [2.9, 3.9]	/
Kwak 4-5	2	0.99 [0.99, 1.00]	0.33 [0.11, 0.63]	1.6 [0.1, 3.1]	0.48 [0.03, 0.94]	/	/

RSS, risk stratification systems; ACR 5, ACR-TIRADS category 5; ACR 4-5, ACR-TIRADS category 4 or 5; ATA high, ATA high risk; ATA high-intermediate, ATA high-intermediate risk; EU 5, EU-TIRADS category 5; EU 4-5, EU-TIRADS category 4 or 5; K 5, K-TIRADS category 5; Kwak 5, Kwak-TIRADS category 5; Kwak 4-5, Kwak-TIRADS category 4 or 5; DOR, diagnostic odds ratio; AUC, area under the curve of receiver operating characteristic; LR+, positive likelihood ratio; LR-, negative likelihood ratio.

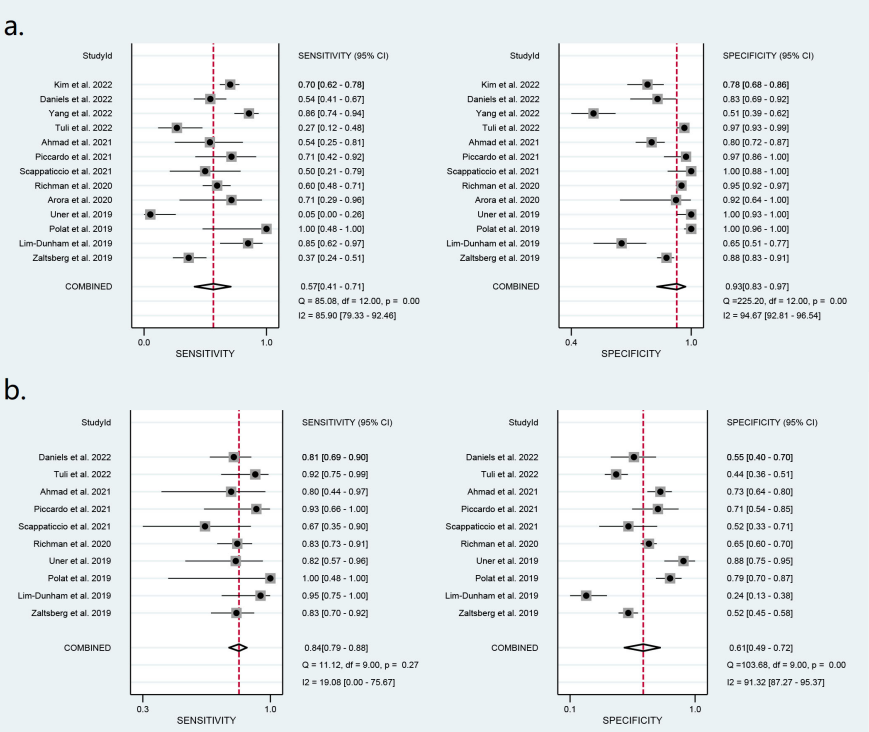


FIGURE 3
Forest plots of the pooled sensitivity and specificity for the diagnosis of malignant thyroid nodules: **(A)** ACR 5 and **(B)** ACR 4-5.

Discussion

The goal of this study was to investigate the reliability and diagnostic performance of the adult-based TI-RADS in the pediatric population. We analyzed the diagnostic performance of the ACR-TIRADS, ATA RSS, EU-TIRADS, K-TIRADS, and Kwak-TIRADS in this study. Since the included studies were not paired studies, we could not directly compare diagnostic performances between different RSSs and calculate p values, which would not be statistically justified.

The sensitivity was highest in ACR category 4–5 and ATA high-intermediate risk, which was 0.84 [0.79, 0.88] and 0.84 [0.75, 0.90], respectively. The specificity was highest in ACR category 5 and EU category 5, which was 0.93 [0.83, 0.97] and 0.93 [0.88, 0.98], respectively. ACR-TIRADS, ATA RSS and EU-TIRADS showed moderate diagnostic performance in pediatric thyroid nodules with

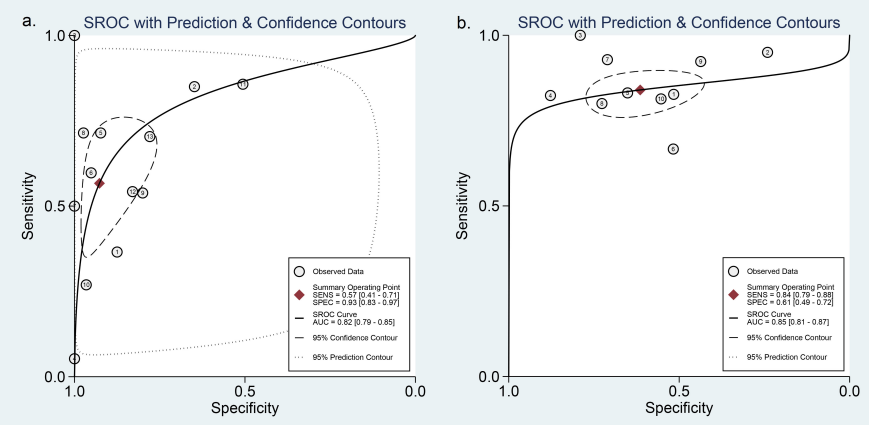


FIGURE 4
HSROC curve of the diagnostic performance: **(A)** ACR 5 and **(B)** ACR 4-5.

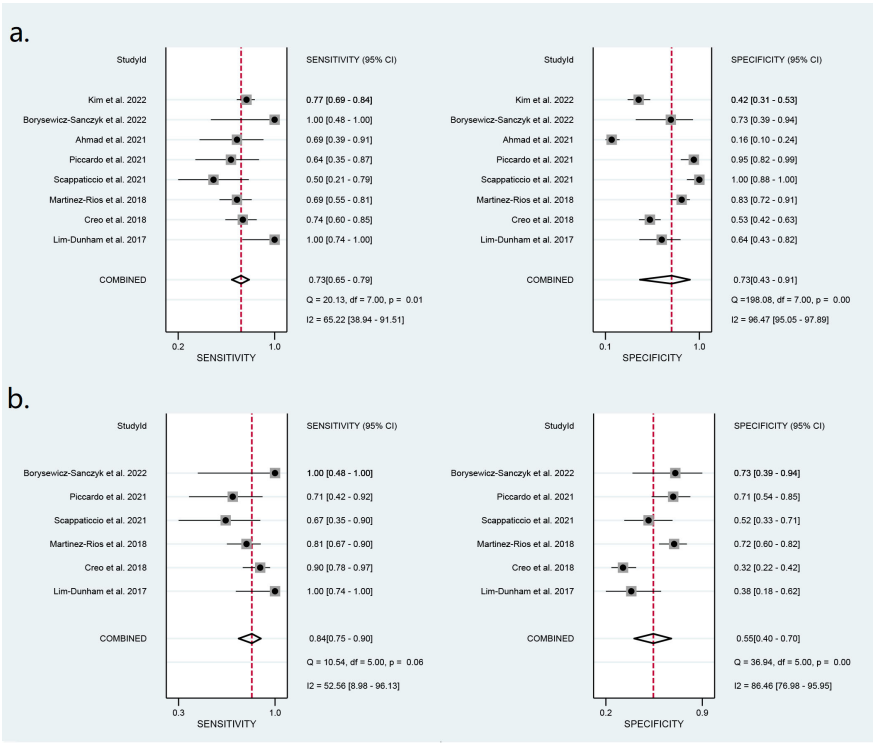


FIGURE 5
Forest plots of the pooled sensitivity and specificity for the diagnosis of malignant thyroid nodules: **(A)** ATA high risk and **(B)** ATA high-intermediate risk.

category 4-5 AUCs of 0.85 [0.81, 0.87], 0.82 [0.78, 0.85], and 0.71 [0.67, 0.75], respectively.

Although ACR-TIRADS, ATA RSS, and EU-TIRADS have moderate diagnostic performance in pediatric thyroid nodule patients. They also have some limitations.

The ACR-TIRADS subdivides features and adds points for composition, echogenicity, shape, margins, and echogenic foci, and stratifies TIRADS level based on the total points of the 5 categories of ultrasound features. This requires a high level of experience and skill, which may be difficult for primary care physicians to master and perform (38).

ATA RSS assesses the malignancy of thyroid nodules based on the performance of ultrasound features with high diagnostic weight, which improves the detection rate of malignant nodules, but has the disadvantage that the assessment of the risk of nodule malignancy is overly dependent on the stratification of suspicious ultrasound features. A small number of pediatric patients cannot be categorized according to ATA RSS because of their specific

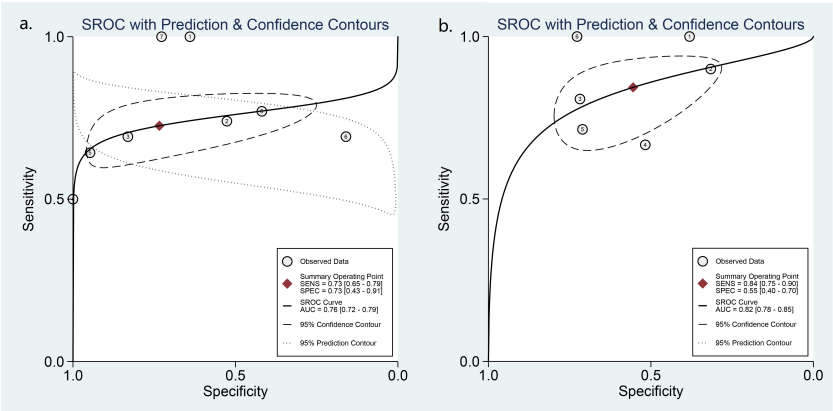


FIGURE 6
HSROC curve of the diagnostic performance: **(A)** ATA high risk and **(B)** ATA high-intermediate risk.

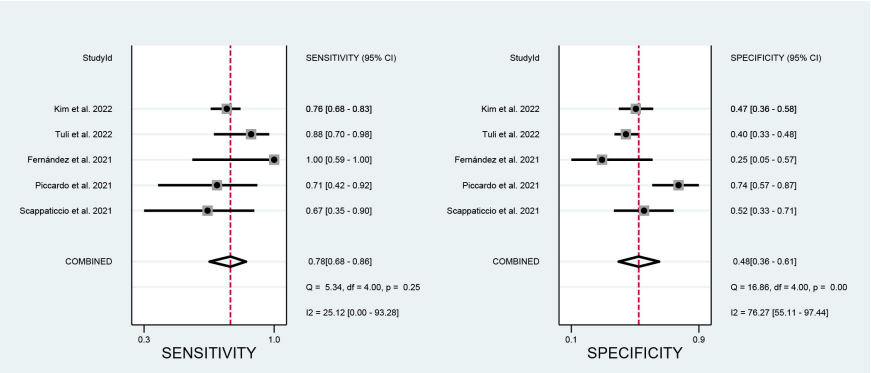


FIGURE 7
Forest plots of the pooled sensitivity and specificity for the diagnosis of malignant thyroid nodules in EU 4-5.

imaging presentation, and such poorly classified nodes could lead to misdiagnosis or underdiagnosis of malignant nodes (39).

ACR-TIRADS, ATA considers FNAB only for nodules greater than or equal to 10 mm, which may miss some malignant nodes in pediatric patients because thyroid volume increases with age, and nodule size is not predictive of malignancy in pediatric patients.

The EU-TIRADS concept of malignancy stratification of thyroid nodules has some similarities with the ATA guidelines. Comparatively, EU-TIRADS has a more streamlined classification of diagnostic weights for malignant nodule features, focusing on the diagnostic weights of highly specific suspicious malignant features, and has a better specificity in identifying benign and malignant nodules. However, the classification of low- and intermediate-risk nodules (EU-TIRADS 3 and 4) by EU-TIRADS explicitly requires the ultrasound features of ovoid shape and smooth margins, while some pediatric patients in the clinic do not have the above two ultrasound features meanwhile cannot be clearly classified in EU-TIRADS 5 categories. This may result in unclassifiable or subjective empirical misjudgment of risk level and is an important reason for the low sensitivity (40).The diagnostic performance of the K-TIRADS was not as expected, with an AUC of only 0.56 [0.06, 0.95]. The K-TIRADS was first proposed by the Korean Society of Thyroid Radiology and Korean Thyroid Association in 2016. Although it shows respectable diagnostic performance for thyroid nodules in adults, recent adult-based studies revealed that in comparison with ACR-TIRADS, the 2016 K-TIRADS demonstrated higher sensitivity (94.5 [92.4, 96.6] vs. 74.7 [70.7, 78.7]) but lower specificity (26.4 [24.2, 28.6] vs. 67.3 [65.0, 69.7]) (41). In this context, the modified K-TIRADS was published in 2021 (42). For pediatric populations, 2021 K-TIRADS newly recommends biopsy of nodules of 0.5–1.0 cm with high suspicion. Compared with the 2016 K-TIRADS, the 2021 K-TIRADS (biopsy cutoffs, 0.5 cm for K-TIRADS 5; 1.0–1.5 cm for K-TIRADS 4) showed higher sensitivity (34.0% vs. 67.3%; $p < 0.001$) while maintaining specificity (89.4% vs. 88.2%; $p = 0.790$) in small nodules of pediatric patients and higher specificity (5.9% vs. 25.4%; $p < 0.001$) while maintaining sensitivity (100% vs. 98.7%; $p = 0.132$) in large nodules of pediatric patients (43).

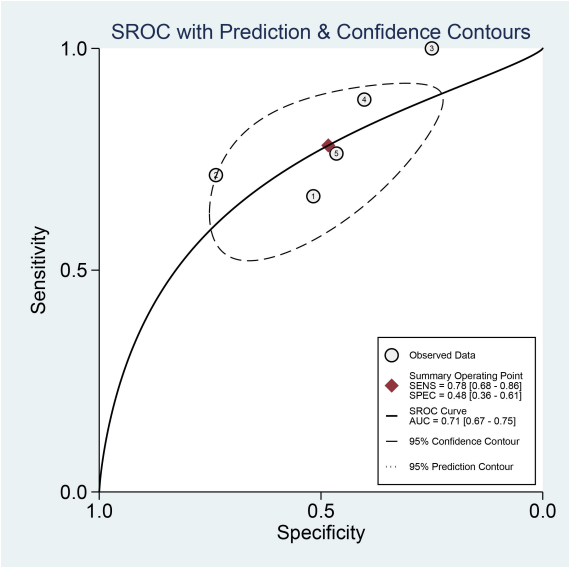


FIGURE 8
HSROC curve of the diagnostic performance of EU 4-5.

In addition, two articles investigated the diagnostic performance of the Kwak-TIRADS (19, 21). Shapira et al. reported an AUC of 0.74 [0.67–0.82] for the diagnostic performance of Kwak-TIRADS compared with 0.72 [0.61–0.82] for ACR-TIRADS. No significant difference was obtained when comparing the Kwak-TIRADS to the ACR TI-RADS (19). Martinez-Rios et al. evaluated the performance of the Kwak-TIRADS and the ATA RSS in assessing thyroid nodules in children. They showed that the test characteristics of both methods were similar to those in adults (21). However, probably because only two studies on Kwak-TIRADS were included, the results of diagnostic performance that we pooled for analysis in our study were not very meaningful.

Additionally, Borysewicz-Sanczyk et al. evaluated the ATA RSS and British Thyroid Association (BTA) ultrasound RSS in the management of thyroid nodules in pediatric patients. The sensitivity and specificity of ATA high risk were (5/5) 100% and

10.3389/fendo.2023.1187935

TABLE 3 Meta-regression analysis.

RSS	Parameter	Category	N	Sensitivity	P	Specificity	P
ACR 5	Total nodes	yes	6	0.60 [0.36, 0.85]	0.96	0.98 [0.95, 1.00]	0.03*
		no	7	0.58 [0.38, 0.78]		0.86 [0.72, 1.00]	
	Region	yes	5	0.43 [0.18, 0.68]	0.21	0.99 [0.97, 1.00]	0.01*
		no	7	0.65 [0.48, 0.83]		0.82 [0.72, 0.92]	
	Standard reference for benign thyroid nodules	yes	9	0.57 [0.40, 0.75]	0.99	0.92 [0.84, 1.00]	0.95
		no	4	0.55 [0.25, 0.85]		0.93 [0.83, 1.00]	
	Standard reference for malignant thyroid nodules	yes	3	0.72 [0.46, 0.98]	0.41	0.95 [0.84, 1.00]	0.41
		no	10	0.52 [0.36, 0.68]		0.92 [0.85, 1.00]	
ACR 4-5	Region	yes	5	0.86 [0.78, 0.94]	0.03*	0.69 [0.54, 0.83]	0.67
		no	5	0.84 [0.78, 0.89]		0.54 [0.38, 0.70]	
	Standard reference for benign thyroid nodules	yes	7	0.85 [0.80, 0.91]	0.02*	0.61 [0.46, 0.75]	0.59
		no	3	0.82 [0.73, 0.90]		0.63 [0.42, 0.84]	
	Standard reference for malignant thyroid nodules	yes	3	0.87 [0.77, 0.97]	0.15	0.48 [0.26, 0.70]	0.17
		no	7	0.83 [0.78, 0.88]		0.66 [0.54, 0.78]	
ATA high	Total nodes	yes	4	0.77 [0.62, 0.92]	0.31	0.91 [0.78, 1.00]	0.02*
		no	4	0.76 [0.66, 0.85]		0.48 [0.14, 0.82]	
	Standard reference for malignant thyroid nodules	yes	3	0.72 [0.55, 0.89]	0.17	0.94 [0.83, 1.00]	0.01*
		no	5	0.75 [0.67, 0.83]		0.52 [0.22, 0.82]	
ATA high-intermediate	Total nodes	yes	3	0.86 [0.72, 1.00]	0.56	0.52 [0.29, 0.75]	0.64
		no	3	0.83 [0.74, 0.93]		0.58 [0.39, 0.78]	
	Region	yes	3	0.74 [0.58, 0.91]	0.01*	0.64 [0.45, 0.83]	0.44
		no	3	0.88 [0.81, 0.95]		0.47 [0.29, 0.66]	
	Standard reference for benign thyroid nodules	yes	4	0.82 [0.69, 0.95]	0.11	0.58 [0.39, 0.76]	0.84
		no	2	0.86 [0.77, 0.95]		0.52 [0.28, 0.75]	
	Standard reference for malignant thyroid nodules	yes	3	0.80 [0.65, 0.94]	0.05	0.54 [0.33, 0.76]	0.83
		no	3	0.86 [0.77, 0.95]		0.56 [0.35, 0.77]	
EU 4-5	Total nodes	yes	2	0.81 [0.61, 1.00]	0.86	0.40 [0.16, 0.63]	0.57
		no	3	0.76 [0.66, 0.87]		0.52 [0.36, 0.68]	
	Standard reference for malignant thyroid nodules	yes	2	0.69 [0.51, 0.88]	0.11	0.64 [0.52, 0.76]	0.08
		no	3	0.80 [0.68, 0.92]		0.42 [0.35, 0.48]	

RSS, risk stratification systems, $*p < 0.05$
Total nodes "yes," sample size more than median; total nodes "no," sample size less than median;
Region "yes," American studies; Region "no," European studies;
Standard reference of malignant thyroid nodules "yes," surgery and/or biopsy pathology; standard reference of malignant thyroid nodules "no," surgery pathology;
Standard reference of benign thyroid nodules "yes," surgery and/or biopsy pathology and follow-up; standard reference of benign thyroid nodules "no," surgery and/or biopsy pathology.

(8/11) 72.7%, respectively, while they were (4/5) 80% and (9/11) 81.8% for BTA category 5 (37). Both RSSs showed good diagnostic performance.

We acknowledge that there were certain limitations. First, all the included studies were retrospective. Second, the number of studies on K-TIRADS and Kwak-TIRADS was small, which resulted in the pooled analyzed diagnostic performance not being very informative.

In conclusion, the ACR-TIRADS, ATA, and EU-TIRADS have moderate diagnostic performance in pediatric thyroid nodule patients. The diagnostic efficacy of the K-TIRADS was not as high as expected. However, the diagnostic performance of Kwak-TIRADS was uncertain because of the small sample size and small number of studies included. More studies are needed to evaluate these adult-based RSSs in pediatric patients with thyroid nodules.

RSS specific for pediatric thyroid nodules and thyroid malignancies were necessary.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Author contributions

ZX conceived the meta-analysis. All authors contributed to the development of the selection criteria, the risk of bias assessment strategy, and data extraction criteria. ZX and YQ developed the search strategy, performed database search, acquired the data, analyzed the data, and drafted the manuscript. AS and WW did statistical analysis. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Miller KD, Fidler-Benaoudia M, Keegan TH, Hipp HS, Jemal A, Siegel RL. Cancer statistics for adolescents and young adults, 2020. *CA Cancer J Clin* (2020) 70(6):443–59. doi: 10.3322/caac.21637

2. Al Nofal A, Gionfriddo MR, Javed A, Haydour Q, Brito JP, Prokop LJ, et al. Accuracy of thyroid nodule sonography for the detection of thyroid cancer in children: systematic review and meta-analysis. *Clin Endocrinol (Oxf)* (2016) 84(3):423–30. doi: 10.1111/cen.12786

3. Canfarotta M, Moote D, Finck C, Riba-Wolman R, Thaker S, Lerer TJ, et al. McGill Thyroid nodule score in differentiating benign and malignant pediatric thyroid nodules: a pilot study. *Otolaryngol Head Neck Surg* (2017) 157(4):589–95. doi: 10.1177/0194599817715629

4. Gupta A, Ly S, Castroneves LA, Frates MC, Benson CB, Feldman HA, et al. A standardized assessment of thyroid nodules in children confirms higher cancer prevalence than in adults. *J Clin Endocrinol Metab* (2013) 98(8):3238–45. doi: 10.1210/jc.2013-1796

5. Francis GL, Waguespack SG, Bauer AJ, Angelos P, Benvenga S, Cerutti JM, et al. Management guidelines for children with thyroid nodules and differentiated thyroid cancer. *Thyroid* (2015) 25(7):716–59. doi: 10.1089/thy.2014.0460

6. Qaisi M, Eid I. Pediatric head and neck malignancies. *Oral Maxillofac Surg Clin North Am* (2016) 28(1):11–9. doi: 10.1016/j.coms.2015.07.008

7. Dinauer CA, Breuer C, Rivkees SA. Differentiated thyroid cancer in children: diagnosis and management. *Curr Opin Oncol* (2008) 20(1):59–65. doi: 10.1097/CCO.0b013e3282f30220

8. Essenmacher AC, Joyce PH, Kao SC, Epelman M, Pesce LM, D'Alessandro MP, et al. Sonographic evaluation of pediatric thyroid nodules. *Radiographics* (2017) 37(6):1731–52. doi: 10.1148/rg.2017170059

9. Ogle S, Merz A, Parina R, Alsayed M, Milas M. Ultrasound and the evaluation of pediatric thyroid malignancy: current recommendations for diagnosis and follow-up. *J Ultrasound Med* (2018) 37(10):2311–24. doi: 10.1002/jum.14593

10. Iakovou I, Giannoula E, Sachpekidis C. Imaging and imaging-based management of pediatric thyroid nodules. *J Clin Med* (2020) 9(2):384. doi: 10.3390/jcm9020384

11. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid* (2016) 26(1):1–133. doi: 10.1089/thy.2015.0020

12. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European Thyroid association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: the EU-TIRADS. *Eur Thyroid J* (2017) 6(5):225–37. doi: 10.1159/000478927

13. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS committee. *J Am Coll Radiol* (2017) 14(5):587–95. doi: 10.1016/j.jacr.2017.01.046

14. Shin JH, Baek JH, Chung J, Ha EJ, Kim JH, Lee YH, et al. Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean society of thyroid radiology consensus statement and recommendations. *Korean J Radiol* (2016) 17(3):370–95. doi: 10.3348/kjr.2016.17.3.370

15. Kwak JY, Han KH, Yoon JH, Moon HJ, Son EJ, Park SH, et al. Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk. *Radiology* (2011) 260(3):892–9. doi: 10.1148/radiol.11110206

16. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Internal Med* (2015) 162(11):777–84. doi: 10.7326/M14-2385

17. Vrabel M. Preferred reporting items for systematic reviews and meta-analyses. *Oncol Nurs Forum* (2015) 42(5):552–4. doi: 10.1188/15.ONF.552-554

18. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* (2011) 155(8):529–36. doi: 10.7326/0003-4819-155-8-201110180-00009

19. Shapira-Zaltsberg G, Miller E, Martinez-Rios C, Bass J, Goldbloom EB, Tang K, et al. Comparison of the diagnostic performance of the 2017 ACR TI-RADS guideline to the kwak guideline in children with thyroid nodules. *Pediatr Radiol* (2019) 49(7):862–8. doi: 10.1007/s00247-019-04385-6

20. Creo A, Alahdab F, Al Nofal A, Thomas L, Kolbe A, Pittock ST. Ultrasonography and the American thyroid association ultrasound-based risk stratification tool: utility in pediatric and adolescent thyroid nodules. *Horm Res Paediatr* (2018) 90(2):93–101. doi: 10.1159/000490468

21. Martinez-Rios C, Daneman A, Bajno L, van der Kaay DCM, Moineddin R, Wasserman JD. Utility of adult-based ultrasound malignancy risk stratifications in pediatric thyroid nodules. *Pediatr Radiol* (2018) 48(1):74–84. doi: 10.1007/s00247-017-3974-y

22. Lim-Dunham JE, Erdem Toslak I, Alsabban K, Aziz A, Martin B, Okur G, et al. Ultrasound risk stratification for malignancy using the 2015 American thyroid association management guidelines for children with thyroid nodules and differentiated thyroid cancer. *Pediatr Radiol* (2017) 47(4):429–36. doi: 10.1007/s00247-017-3780-6

23. Lim-Dunham JE, Toslak IE, Reiter MP, Martin B. Assessment of the American college of radiology thyroid imaging reporting and data system for thyroid nodule malignancy risk stratification in a pediatric population. *AJR Am J Roentgenol* (2019) 212(1):188–94. doi: 10.2214/AJR.18.20099

24. Polat YD, Öztürk VS, Ersoz N, Anık A, Karaman CZ. Is thyroid imaging reporting and data system useful as an adult ultrasonographic malignancy risk stratification method ın pediatric thyroid nodules? *J Med Ultrasound* (2019) 27(3):141–5. doi: 10.4103/JMU.JMU_35_19

25. Uner C, Aydin S, Ucan B. Thyroid image reporting and data system categorization: effectiveness in pediatric thyroid nodule assessment. *Ultrasound Q* (2020) 36(1):15–9. doi: 10.1097/RUQ.0000000000000476

26. Richman DM, Benson CB, Doubilet PM, Wassner AJ, Asch E, Cherella CE, et al. Assessment of American college of radiology thyroid imaging reporting and data system (TI-RADS) for pediatric thyroid nodules. *Radiology* (2020) 294(2):415–20. doi: 10.1148/radiol.2019191326

27. Arora S, Khoury J, Trout AT, Chuang J. Improving malignancy prediction in AUS/FLUS pediatric thyroid nodules with the aid of ultrasound. *Horm Res Paediatr* (2020) 93(4):239–44. doi: 10.1159/000509118

28. Scappaticcio L, Maiorino MI, Iorio S, Docimo G, Longo M, Grandone A, et al. Exploring the performance of ultrasound risk stratification systems in thyroid nodules of pediatric patients. *Cancers (Basel)* (2021) 13(21):5304. doi: 10.3390/cancers13215304

29. Piccardo A, Fiz F, Bottoni G, De Luca C, Massollo M, Catrambone U, et al. Facing thyroid nodules in paediatric patients previously treated with radiotherapy for non-thyroidal cancers: are adult ultrasound risk stratification systems reliable? *Cancers (Basel)* (2021) 13(18):4692. doi: 10.3390/cancers13184692

30. Ahmad H, Al-Hadidi A, Bobbey A, Shah S, Stanek J, Nicol K, et al. Pediatric adaptions are needed to improve the diagnostic accuracy of thyroid ultrasound using TI-RADS. *J Pediatr Surg* (2021) 56(6):1120–5. doi: 10.1016/j.jpedsurg.2021.02.034

31. Yeste Fernández D, Vega Amenabar E, Coma Muñoz A, Arciniegas Vallejo L, Clemente León M, Planes-Conangla M, et al. Ultrasound criteria (EU-TIRADS) to identify thyroid nodule malignancy risk in adolescents. correlation with cyto-histological findings. *Endocrinol Diabetes Nutr (Engl Ed)* (2021) 68(10):728–34. doi: 10.1016/j.endinu.2020.11.009

32. Lee SB, Cho YJ, Lee S, Choi YH, Cheon JE, Kim WS. Korean Society of thyroid radiology guidelines for the management of pediatric thyroid nodules: suitability and risk factors. *Thyroid* (2021) 31(10):1472–80. doi: 10.1089/thy.2020.0875

33. Tuli G, Munarin J, Scollo M, Quaglino F, De Sanctis L. Evaluation of the efficacy of EU-TIRADS and ACR-TIRADS in risk stratification of pediatric patients with thyroid nodules. *Front Endocrinol (Lausanne)* (2022) 13:1041464. doi: 10.3389/fendo.2022.1041464

34. Yang J, Page LC, Wagner L, Wildman-Tobriner B, Bisset L, Frush D, et al. Thyroid nodules on ultrasound in children and young adults: comparison of diagnostic performance of radiologists' impressions, ACR TI-RADS, and a deep learning algorithm. *AJR Am J Roentgenol* (2023) 220(3):408–17. doi: 10.2214/AJR.22.28231

35. Daniels KE, Shaffer AD, Garbin S, Squires JH, Vaughan KG, Viswanathan P, et al. Validity of the American college of radiology thyroid imaging reporting and data system in children. *Laryngoscope* (2022). doi: 10.1002/lary.30425

36. Kim PH, Yoon HM, Baek JH, Chung SR, Choi YJ, Lee JH, et al. Diagnostic performance of five adult-based US risk stratification systems in pediatric thyroid nodules. *Radiology* (2022) 305(1):190–8. doi: 10.1148/radiol.212762

37. Borysewicz-Sańczyk H, Sawicka B, Karny A, Bossowski F, Marcinkiewicz K, Rusak A, et al. Suspected malignant thyroid nodules in children and adolescents according to ultrasound elastography and ultrasound-based risk stratification systems-experience from one center. *J Clin Med* (2022) 11(7):1768. doi: 10.3390/jcm11071768

38. Alexander AA. US-Based risk stratification "guidelines" for thyroid nodules: quō vādis? *J Clin Ultrasound* (2020) 48(3):127–33. doi: 10.1002/jcu.22803

39. Al Maawali A, Matheson C, Baird R, Blair G. The thyroid nodules in kids study (ThyNK study): an evaluation of clinical practice variation. *J Pediatr Surg* (2020) 55 (5):950–3. doi: 10.1016/j.jpedsurg.2020.01.046

40. Dobruch-Sobczak K, Adamczewski Z, Szczepanek-Parulska E, Migda B, Wolański K, Krauze A, et al. Histopathological verification of the diagnostic performance of the EU-TIRADS classification of thyroid nodules-results of a multicenter study performed in a previously iodine-deficient region. *J Clin Med* (2019) 8(11):1781. doi: 10.3390/jcm8111781

41. Ha EJ, Na DG, Baek JH, Sung JY, Kim JH, Kang SY. US Fine-needle aspiration biopsy for thyroid malignancy: diagnostic performance of seven society guidelines applied to 2000 thyroid nodules. *Radiology* (2018) 287(3):893–900. doi: 10.1148/radiol.2018171074

42. Ha EJ, Chung SR, Na DG, Ahn HS, Chung J, Lee JY, et al. 2021 Korean thyroid imaging reporting and data system and imaging-based management of thyroid nodules: Korean society of thyroid radiology consensus statement and recommendations. *Korean J Radiol* (2021) 22(12):2094–123. doi: 10.3348/kjr.2021.0713

43. Kim PH, Yoon HM, Baek JH, Chung SR, Choi YJ, Lee JH, et al. Diagnostic performance of the 2021 Korean thyroid imaging reporting and data system in pediatric thyroid nodules. *Eur Radiol* (2023) 33(1):172–80. doi: 10.1007/s00330-022-09037-2

frontiers | Frontiers in Endocrinology

TYPE Review
PUBLISHED 31 May 2023
DOI 10.3389/fendo.2023.1198099

OPEN ACCESS

EDITED BY
Marco António Campinho,
University of Algarve, Portugal

REVIEWED BY
Clara Ugolini,
University of Pisa, Italy
Gianmaria Pennelli,
University of Padua, Italy

*CORRESPONDENCE
Zubair Baloch
baloch@pennmedicine.upenn.edu

†These authors have contributed equally to this work

RECEIVED 31 March 2023
ACCEPTED 10 May 2023
PUBLISHED 31 May 2023

CITATION
Crescenzi A and Baloch Z (2023) Immunohistochemistry in the pathologic diagnosis and management of thyroid neoplasms.
Front. Endocrinol. 14:1198099.
doi: 10.3389/fendo.2023.1198099

Immunohistochemistry in the pathologic diagnosis and management of thyroid neoplasms

Anna Crescenzi[1†] and Zubair Baloch[2*†]

[1]Pathology, University Campus Bio-Medico of Rome, Fondazione Policlinico, Rome, Italy, [2]Pathology & Laboratory Medicine, University of Pennsylvania Medical Center, Perelman School of Medicine, Philadelphia, PA, United States

The use of immunohistochemistry cannot be underestimated in the everyday practice of thyroid pathology. It has evolved over the years beyond the traditional confirmation of thyroid origin to molecular profiling and the prediction of clinical behavior. In addition, immunohistochemistry has served to implement changes in the current thyroid tumor classification scheme. It is prudent to perform a panel of immunostains, and the immunoprofile should be interpreted in light of the cytologic and architectural features. Immunohistochemistry can also be easily performed in the limited cellularity specimen preparation generated from thyroid fine-needle aspiration and core biopsy; however, it will require laboratory validation of immunostains specific to these preparations to avoid diagnostic pitfalls. This review discusses the application of immunohistochemistry in thyroid pathology with a focus on limited cellularity preparations.

KEYWORDS

thyroid, pathology, immunohistochemistry, molecular, biomarkers, cytology

1 Introduction

Thyroid carcinoma is the most common malignancy of endocrine organs and accounts for approximately 1% of all cancers. As per the new WHO classification scheme, the neoplasms of the thyroid gland are stratified into the following main categories: follicular cell-derived neoplasms, C-cell derived neoplasms, mixed medullary and follicular cell-derived neoplasms, salivary gland type carcinomas, thyroid tumors of uncertain histogenesis, thymic tumors within the thyroid, and embryonal thyroid neoplasms. Even though the majority of thyroid neoplasms can be diagnosed on the basis of cellular and architectural features, difficulties in the diagnosis can occur due to overlapping histomorphologic features between primary and secondary thyroid neoplasms and partial or complete loss of differentiation (1).

The well-differentiated thyroid carcinomas originating from the thyroid follicular cells show either follicular or papillary growth patterns or an admixture of both. The presence of colloid within follicles and complex papillary structures with diagnostic nuclear cytology in

these neoplasms facilitate the diagnosis of these neoplasms (1–6). The solid and "insular" growth pattern of poorly differentiated carcinoma, especially in cases with an inconspicuous or lack of a well-differentiated component, can be mistaken for C-cell-derived medullary thyroid carcinoma or metastatic neuroendocrine neoplasm arising at other body sites. Anaplastic/undifferentiated carcinoma can show varying cytologic features and growth patterns mimicking lymphoma, mesenchymal tumors, and secondary tumors of the thyroid gland (2, 4).

In the abovementioned scenarios, employing a panel of immunostains can help solve diagnostic conundrums. In addition, immunohistochemistry (IHC) has proven to be helpful in the diagnosis of the following rare tumors: mixed follicular and medullary thyroid carcinoma, salivary gland type carcinomas, tumors of uncertain histogenesis, and intra-thyroidal thymic neoplasms (2). Employing mutation-specific antibodies can serve to distinguish between papillary carcinomas harboring the BRAFV600E mutation from RAS-like neoplasms. The utility of proliferation markers such as Ki67 cannot be underestimated in the grading of thyroid carcinoma, which has been shown to be a predictor of clinical behavior in both follicular and C-cell-derived neoplasms (1, 2, 7).

2 Immunochemistry: basic concepts

Immunostaining is an easy, cheap, and widely available technique for selectively identifying specific molecules in tissue sections and cytological preparations. This technique is based on the use of antibodies (also called immunoglobulins) that are Y-shaped globular proteins formed by two light chains and two heavy chains, held together by disulphide bonds (8). The molecular recognition abilities of the antibodies allow for various applications in diagnostic pathology. Each antibody is capable of binding only to a specific antigen (9); therefore, they are currently applied in pathology to identify the cell lineage, examine the expression of biomarkers, characterise tumors, and more recently, determine the expression of targets for tailored therapies. For diagnostic purposes, antibodies are labelled, directly or with a multistep chain, with a visible molecule that allows the recognition of their binding reaction in tissue sections or cytological samples. The immunostaining protocol is mostly automated in many laboratories, which improves the reproducibility of the reaction product, although this standardization is usually developed for IHC on histological sections, whereas dedicated recommendations and practice paradigms are still lacking for cytological samples. This is probably because of the large variability in cellular preparations (conventional smears, thin layer cytology, and cell-blocks), different treatment of the specimens for immunostaining, and interpretative cutoffs (7, 8, 10–13).

As stated above, there are three main reasons for the application of immunocytochemistry in thyroid pathology: ***determining cell and site of origin, differentiating benign from malignant neoplasms, and influencing clinical management.***

3 Determining cell and site of origin

IHC is an indispensable tool that complements routine histologic techniques for elucidating differential diagnosis in histologic and cytologic preparations. The use of IHC in thyroid pathology is based on knowledge regarding cell of origin and further characterization. It is mainly suggested for lesions that are suspected of non-follicular or non-thyroidal origin (e.g., parathyroid lesions/neoplasms, medullary thyroid carcinoma, lymphoma, metastases from other organs—secondary tumors, etc.) (14, 15).

3.1 Thyroid follicular cell lineage markers

3.1.1 Thyroid follicular cell origin

Is usually confirmed by a panel of immunohistochemical markers that can identify metastasis to the thyroid gland from other organs, thyroid cancer metastasis to extra-thyroidal sites, and thyroid carcinoma arising in ectopic thyroid tissue. The most important markers of thyroid follicular cell derivation are thyroglobulin (TG), thyroid transcription factor 1 (TTF1), and paired-box gene 8 (PAX8); antibodies against these are often used in a panel to overcome the limits of a single antibody (Figures 1, 2). It is relevant to know some details about these antibodies (7, 13, 16–18).

3.1.2 Thyroglobulin

Is the most specific marker of thyroid follicular cell derivation. It is a glycoprotein manufactured by thyrocytes, from which it is secreted into thyroid follicles, forming a major constituent of colloid. Normal thyrocytes show diffuse cytoplasmic staining by TG; this staining pattern is maintained in well-differentiated follicular-derived thyroid carcinomas, such as papillary thyroid carcinoma (PTC) and follicular thyroid carcinoma (FTC), and is completely absent in medullary thyroid carcinoma (MTC) and metastasis to the thyroid gland (Figure 3). Focal expression of TG in the follicular component is seen in cases of mixed medullary and follicular thyroid carcinoma. Focal diffuse staining has been reported in >50% of high-grade follicular-cell-derived non-anaplastic carcinomas and is often lost in the foci of necrotic tumor and anaplastic thyroid carcinoma (ATC). In oncocytic lesions, TG usually shows punctate and dot-like perinuclear staining pattern (1, 7, 19–23).

The low or absent production of thyroglobulin by some tumors can lead to diagnostic conundrums, especially in the following clinical scenarios: the use of TG FNA washout evaluation for regional and distant metastasis and the role of serum TG measurement for the follow up of patients with thyroid carcinomas (21, 24, 25).

3.1.3 Thyroid transcription factor 1

Also termed as thyroid-specific enhancer binding protein (NKx2.1), belongs to the family of homologous transcription factors in the NKx2 gene, and is located in the q12–q21 region of chromosome 14. The *TTF-1* gene translates a nuclear protein with

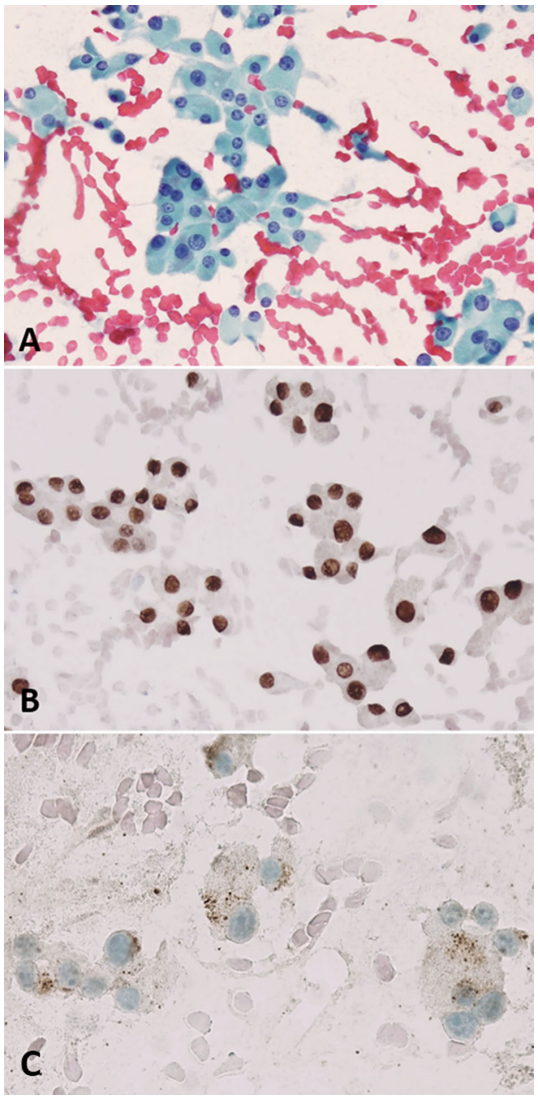


FIGURE 1
An oncocytic proliferative lesion, direct smear. **(A)** Papanicolaou stain shows numerous oncocytic cells with variable nuclear dimensions. **(B)** PAX8 immunostaining performed on a destained slide shows strong nuclear positivity in the cells. **(C)** Thyroglobulin immunostaining performed on a destained slide shows a punctate dot-like positive cytoplasmic reaction.

an approximate mass of 38 kDa, comprising a single polypeptide of 371 amino acid polypeptides. TTF-1 expression is regulated during embryonic development and appears early in the foregut endoderm and then in the tracheal precursor cells. After birth, TTF-1 expression is confined to the pulmonary type II alveolar cells (26). In the thyroid gland, *TTF-1* expression occurs earlier than the expression of genes related to follicular cell differentiation, such as TG, thyroid peroxidase (TPO), and thyrotropin receptor (TSHR) (27).

TTF-1 shows nuclear expression by IHC in thyroid follicular and parafollicular cells and lung. TTF-1 is diffusely expressed in PTC, FTC, high-grade follicular-derived non-anaplastic thyroid carcinoma and MTC. TTF-1 expression is retained in less than 20% of ATCs (7). TTF-1 is expressed in more than 80% of lung adenocarcinomas, a subset of squamous cell carcinoma of pulmonary origin, small cell carcinoma, neuroendocrine carcinomas, and also rarely in adenocarcinoma of genitourinary and gastrointestinal tracts and breast (26–29).

TTF-1 can be useful in confirming the diagnosis of a thyroid primary lacking a well-differentiated growth pattern (papillary or follicular) or unusual cytology, such as poorly differentiated thyroid carcinoma, mucoepidermoid carcinoma, and secondary tumors (7, 30).

3.1.4 Paired box gene 8

Is a transcription factor that belongs to the paired-box family of genes; it plays a critical role in the development of the thyroid gland, kidney, and Mullerian tract (31). With IHC, its expression is seen in thyroid, renal, and urinary bladder neoplasms and malignancies of Mullerian origin, including ovarian primaries. Several studies have added the following to the repertoire of PAX8 positive tumors: carcinomas of the breast, lung, prostate, gastrointestinal tract, liver and pancreas, testicular tumors, mesothelioma, melanoma, and rhabdomyosarcoma (31–34).

3.1.5 PAX8

Gives a nuclear staining in normal and neoplastic thyrocytes, and usually maintains this expression pattern also in cases of high-grade follicular-cell-derived carcinoma, anaplastic carcinoma, and its squamous subtype (35–37). Among thyroid tumors, a majority of non-follicular cell-derived thyroid carcinomas stain negative for the PAX8 antibody (36). Rare cases of medullary thyroid carcinoma can show PAX8 expression. Intrathyroid thymic carcinoma (ITC) shows a nuclear positive reaction with polyclonal PAX8 antibody but does not react with the monoclonal form. Therefore, monoclonal PAX8 antibody is more specific for thyroid follicular cell origin (36). As noted above, as PAX8 is also expressed in a wide variety of neoplasms from other organs, an initial panel of TTF-1, TG, and PAX8 is needed to confirm or exclude distant metastases from a thyroid primary (35, 37). The other markers used to confirm thyroid follicular cell differentiation include TTF-2 (FOXE1) and thyroid peroxidase (1, 7, 22, 35).

3.2 Parafollicular C-cell specific markers

Medullary thyroid carcinoma (MTC) originates from parafollicular C-cells of the thyroid gland. The C-cells mainly secrete calcitonin hormone, which plays a minor role in calcium metabolism compared with parathyroid hormone (PTH) (38–40). Most MTCs (>95%) secrete calcitonin and show patchy to diffuse cytoplasmic expression of this biomarker with IHC (Figure 4). It is well-known that MTC, in addition to its typical nesting growth, tumor cells with nuclear chromatin typical of neuroendocrine tumors (salt and pepper), and amyloid rich tumor stroma, can demonstrate a variety of architectures and cellular features that can be mistaken for other primary thyroid tumors (41–43). In such cases, IHC for calcitonin in pathologic preparations confirms the diagnosis of MTC. This also holds true for rare cases of mixed

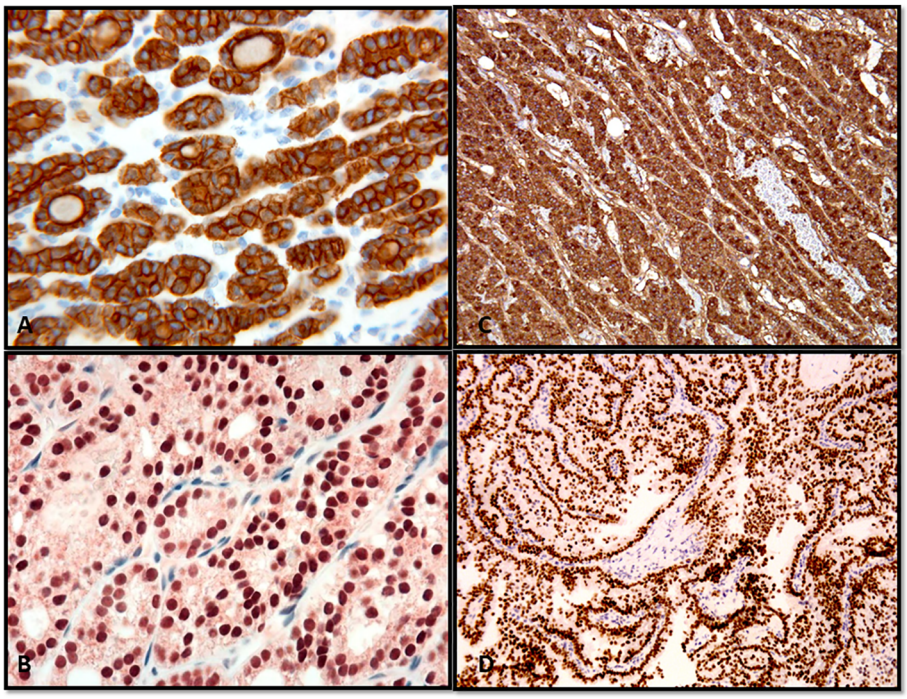


FIGURE 2
Confirmation of the thyroid follicular cell origin through the expression of thyroglobulin and TTF in follicular adenoma **(A)**, thyroglobulin; **(B)**, TTF-1) and papillary thyroid carcinoma **(C)**, thyroglobulin; **(D)**, TTF-1).

medullary and follicular thyroid carcinoma, in which calcitonin specifically highlights the MTC components (43) (Figure 5).

Owing to its architecture and cellular features, MTC can be difficult to distinguish from metastases to the thyroid from neuroendocrine carcinoma arising in other organs, especially the lung and gastrointestinal tracts (41, 43). It is well known that calcitonin is also expressed in other neuroendocrine tumors besides MTC. In cases in which the diagnostic differential includes MTC and metastatic neuroendocrine carcinoma, clinical correlation and serum calcitonin level, which is often quite increased in MTC, can help determine the correct diagnosis. MTC also shows cytoplasmic expression of monoclonal carcinoembryonic antigen (mCEA), which can also serve as biomarker for disease surveillance in addition to calcitonin (44, 45). This proves to be helpful in rare cases of calcitonin negative MTC (43, 45–47). Additionally, MTC shows expression of other neuroendocrine markers, such as chromogranin, synaptophysin, and rarely CD56. The second-generation neuroendocrine markers insulinoma-associated protein 1 (INSM1), ISL1, and secretagogin show high sensitivity and specificity for neuroendocrine differentiation and maintain the expression even in poorly differentiated neuroendocrine carcinomas. In particular, INSM1 has been reported to be a highly sensitive and specific neuroendocrine marker and is useful in the diagnosis of MTC and C-cell hyperplasia (48–50).

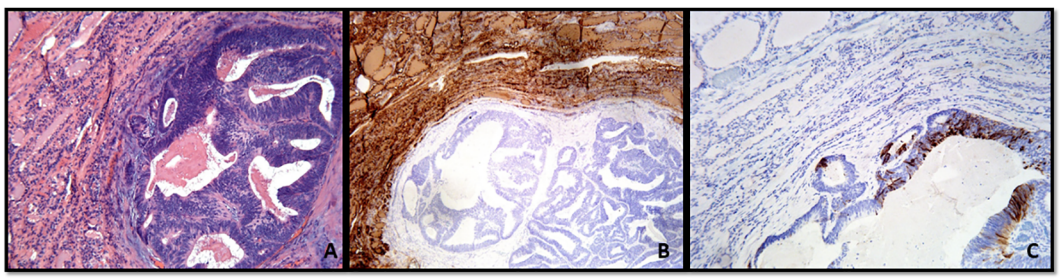


FIGURE 3
(A–C) Metastatic colonic adenocarcinoma to the thyroid gland showing the neoplastic gland in the background thyroid parenchyma **(A)**, Hematoxylin and Eosin stain). **(B, C)** With immunohistochemistry, thyroglobulin expression is only present in the thyroid parenchyma **(B)**, and the tumor shows cytokeratin 20 expression **(C)**.

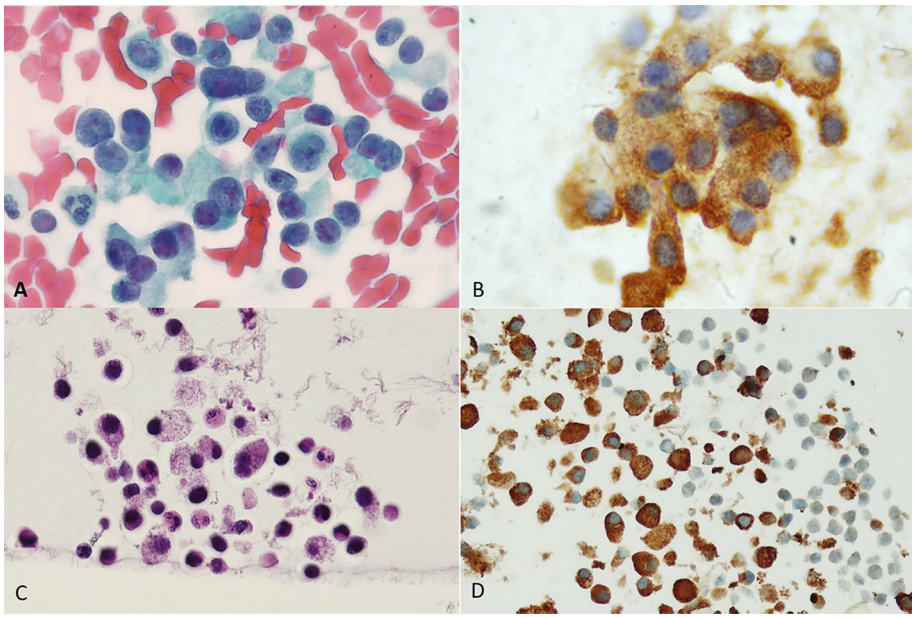


FIGURE 4
Medullary carcinoma. **(A)** Direct smear. Papanicolaou stain shows a cellular proliferation of polymorphous sometimes plasmacytoid appearing cells with oval nuclei of variable sizes. **(B)** Calcitonin immunostain performed on a destained smear. A granular brown positive reaction is evident in the cytoplasm of the cells. **(C)** Cellblock preparation using the agar gel method. Hematoxylin and Eosin stain shows discohesive cells with granular cytoplasm. Bi-nucleated cells are also evident. **(D)** Calcitonin immunostain shows strong cytoplasmic granular expression.

3.2.1 Rare thyroid neoplasms

The value of IHC cannot be underemphasized in the diagnosis of uncommon thyroid neoplasms and those of uncertain histogenesis (1, 2, 51) (see Table 1).

4 Differentiating benign from malignant thyroid neoplasms

Most thyroid neoplasms are diagnosed based on architectural and cellular features and a lack or presence of invasive features. However, in some instances, it may be difficult to distinguish between follicular adenoma and non-invasive follicular tumor with papillary-like nuclear features (NIFTP), an encapsulated follicular variant of papillary thyroid carcinoma, follicular carcinoma, and follicular adenoma with papillary architecture from papillary thyroid carcinoma.

The diagnosis of follicular carcinoma and the encapsulated follicular variant of papillary thyroid carcinoma requires the evaluation of the tumor-capsule-thyroid interface. Invasion *of* the capsule, invasion *through* the capsule, and invasion *into* veins in or beyond the capsule represent the diagnostic criteria for carcinoma in a follicular-patterned thyroid neoplasm. To this date, what constitutes "*capsular*" invasion in a follicular-patterned thyroid neoplasm remains controversial. Some require penetration *through* the capsule of the tumor and others invasion *into* the capsule to render a diagnosis of either follicular carcinoma or an encapsulated follicular variant of papillary thyroid carcinoma, while others believe that the diagnosis of minimally invasive follicular carcinoma should only be rendered when vascular invasion is present. However, studies have shown that metastatic disease can occur in cases of follicular carcinoma in which only capsular invasion occurred. Thus, "capsular invasion is a sufficient criterion to diagnose malignancy" (1, 2, 4, 6, 52–57).

Despite the controversy regarding capsular invasion as a criterion for malignancy, all agree that angioinvasion is a definite feature of malignancy. It has been shown that encapsulated angioinvasive follicular-patterned tumors carry a significant propensity for clinically malignant behavior. The following histomorphologic criteria have been proposed for the diagnosis of angioinvasion: the invasive tumor should form a plug or polyp in a subendothelial location, enveloping of tumor thrombus by the endothelium, and the tumor thrombus does not have to be attached to the vessel wall to be accepted as an invasion (52–57).

Immunostaining for Factor VIII–related antigen and other endothelial markers, such as CD31, CD34, and ERG, can confirm the foci of angioinavsion. Rarely, histiocytes intermixed with fibrin and inflammatory cells within capsular vessels can mimic foci of angioinvasion. In such instances, macrophage markers, such as CD68 or CD163, and markers for follicular cell lineage, TTF-1, PAX8, and thyroglobulin, can help to confirm the presence of tumor cells within a vessel lumen (52, 53, 55).

The diagnostic conundrum of differentiating a benign from a malignant thyroid lesion is often encountered in limited cellularity fine-needle aspiration (FNA) and core-biopsy specimens. The use of immunohistochemical markers for differentiating benign from malignant thyroid neoplasms in FNA specimens classified as indeterminate is often debated in the literature (11, 12, 58–69).

The combination of HBME-1, GAL-3, and CK19 is by far the most common panel for distinguishing benign from malignant

TABLE 1 Immunoprofile of key thyroid primary and secondary tumors.

Site of origin	Immunostaining profile*
Thyroid tumors - primary	
1. Follicular cell	CK7+, CK20-, TTF1+, PAX8+, thyroglobulin+
2. C-Cell	CK7+, TTF-1+, PAX8-, calcitonin+, CEA+, synaptophysin+, chromogranin+, thyroglobulin -
Thyroid tumors – primary others	
1. Hyalinizing trabecular tumor	TTF1+, PAX8+, thyroglobulin+, MIB1 (membranous)
2. Mucoepidermoid carcinoma	AE1/AE3+, pan-cytokeratin +, p63+, TTF1**+, PAX8+**, thyroglobulin +**
3. Sclerosing mucoepidermoid carcinoma with eosinophilia	TTF1+, PAX8-, thyroglobulin -
4. Cribriform morular thyroid carcinoma	β-catenin, TTF1+ (mainly in cribriform components), PAX8-, thyroglobulin-,
5. Intrathyroidal thymic tumors	AE1/AE3+, TTF1-, thyroglobulin-, CD5+, p63+, bcl-2+
Parathyroid	TTF1-, PAX8-, thyroglobulin-, calcitonin-, PTH+, GATA3+, chromogranin+
Thyroid tumors - secondary	
1. Pulmonary	CK7+, TTF1+, napsin+, PAX8-, thyroglobulin-
2. Gastrointestinal tract	
I. Esophagus	CK7+, CK20-, TTF-1 -, CDX2 +/-, CEA+, MUC1-/+, MUC5AC -/+, SATB2-
II. Stomach	CK7+, CK20+, TTF-1 -, CEA+, CDX2** MUC1 -/+, MUC5AC-/+
III. Colorectal	CK7-, CK20+, CDX2+, SATB2+, MOC31+
3. Breast	CK7+, CK20-, GATA3+, mammoglobin+/-, GCDFP15-/+, ER+, PR+, TTF-1 -, TG-
4. Melanoma	SOX10+, Melan-A+, S100+, HMB45+, CK7-, CK20-
5. Kidney	
I. Clear cell	CK7-, PAX8+, PAX2+, CAIX+, CD10+, RCC+, AE1/AE3+, CAM5.2+, EMA+, AMACR+/-, GATA3 -, TTF-1-
II. Clear cell papillary	CK7+, PAX8+, CAIX +, CD10-, RCC+/-, AMACR-, GATA3 -/+ (rare cases)
III. Papillary renal cell	CK7+, PAX8+, CA1X+/-, CD10+, RCC+, AMACR +, GATA3 -

*it is preferable that all immunostains should be validated with cytology preparations; **some cases show negative expression; CK, cytokeratin.

thyroid neoplasms, as no individual biomarker has sufficient sensitivity or specificity to accomplish this task. Combined immunopositivity for Gal-3, CK19, and HBME-1 shows high sensitivity (95%) and specificity (97%) for the diagnosis of papillary thyroid carcinoma (11, 12, 58–64, 66, 68, 69) (Figure 6). Combined immunoexpression of Gal-3 and CK19 had 92% sensitivity and 99% specificity while combined positivity for Gal-3 and HBME-1 had 95% sensitivity and 95% specificity for papillary thyroid carcinoma (64). It should be noted that the expression of HBME-1, GAL-3, and CK19 is not predictive of the clinical aggressiveness of the tumor and cannot be used to guide the surgical excision (64).

Galectin 3 (Gal-3) has received significant attention for its utility as a diagnostic marker for thyroid cancer, being positive in thyroid carcinoma and negative in benign neoplasms and normal thyroid tissue (70). In a meta-analysis of 8,172 thyroid nodules with histologic evaluation, Gal-3 IHC was reported to be positive in 87% of thyroid cancers, confirmed by histopathologic follow-up. This information confirms that many thyroid carcinomas have overexpression of this marker. A Gal-3 test on thyroid FNA samples (cellblock preparation) has a sensitivity lower than that observed in histologic preparations (pooled histologic sensitivity of Gal-3 was 96%, while sensitivity with FNA was 90%); mainly due to the different methods used for Gal-3 evaluation in thyroid cytological specimens, technical variability in antibody clones and immunostaining protocols, and relevant differences in staining interpretation (i.e. nuclear, cytoplasmic, or membranous positivity) (12). In summary, the use of Gal-3 in cellblock preparation from thyroid FNA may support a diagnosis of malignancy in thyroid nodules classified as indeterminate. In addition to galectin-3, other markers such as Hector Battifora mesothelial cell-1 (HBME-1), cytokeratin-19 (CK19), and cluster differentiation antigen 56 (CD56) can facilitate the diagnosis of thyroid carcinoma in both histologic and cytologic preparations (62–64, 71) (Figure 6).

Ki-67 is the protein product of the gene *MKI67* and is a commonly used IHC marker for cell proliferation. Recently, the Ki-67 index has been proposed for the stratification of PTC, FTC, and MTC into different risk categories. The proposed Ki-67 indices show that differentiated thyroid carcinomas can be stratified into low-, moderate-, and high-risk groups using the cutoff values of

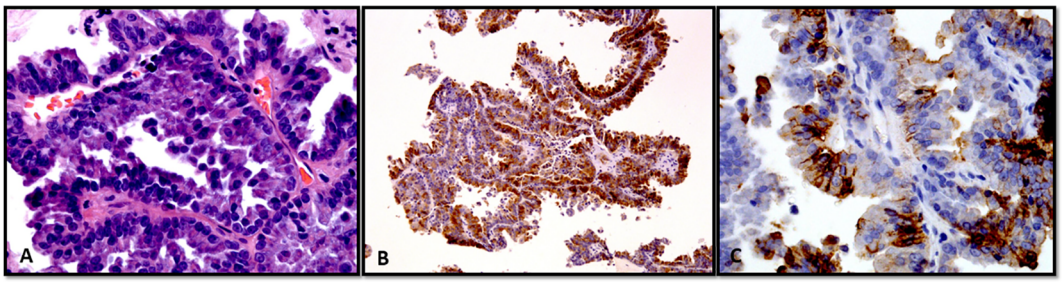


FIGURE 6
(A–C) Fine-needle aspiration cellblock preparation showing a case of papillary thyroid carcinoma. **(A)** Classic subtype (Hematoxylin and Eosin stain). **(B, C)** With immunohistochemistry, the tumor cells show positive expression of thyroglobulin **(B)** and HBME-1 **(C)**.

<5%, 5–10%, and 10–30% (72–75). A two-tiered system is recommended for Ki67 evaluation in medullary thyroid carcinoma, employing a cutoff of 5% (76).

With light microscopy, at least one of three features, mitotic index of ≥5 per 2 mm^2, Ki67 proliferative index of ≥5%, or tumor necrosis, is required to define high-grade MTCs, and these criteria have been integrated in the 5th edition of the World Health Organization classification of thyroid tumors (1, 2). Additionally, the use of Ki67 IHC with cytological samples from thyroid FNA has been investigated. In a recent study, the authors applied a scoring evaluation for calculating the percentage of positive cells by counting at least 200 tumor cells; they concluded that the Ki-67 index determined in cytology specimens significantly correlates with the Ki-67 index obtained by immunohistochemical analyses of histologic specimens (77). This analysis was performed on air-dried smears that were formalin fixed before immunostaining.

IgG4-thyroid-related disease (TRD), although uncommon, is a spectrum of diseases with a clinical presentation that can often mimic malignancy. The threshold to confirm increased IgG4-positive plasma cells ranges from more than 20 to more than 30 IgG4-positive plasma cells per high-power field by microscopic examination (78–81). The FNA specimens show lymphoplasmacytic infiltrates and oncocytes and the cytological features are usually classified as benign (80). Clinical history, radiological characteristics, and cytological features, such as abundant plasma cells, fibroblast, and epithelial atypia, should raise the suspicion of IgG4-related disease (82). If IgG4 TRD is clinically suspected at the time of FNA, IHC might confirm the predominance of IgG4-secreting plasma cells in the cytological sample, leading to additional clinical workup.

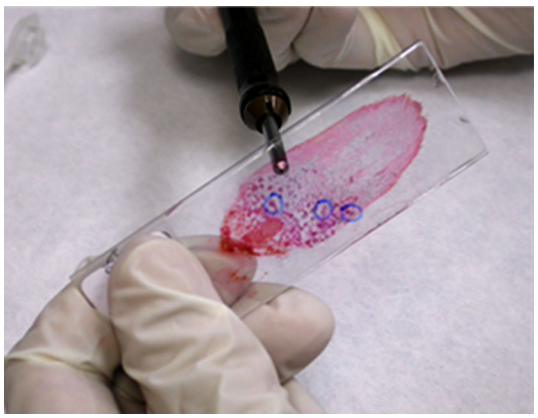


FIGURE 7
A smeared slide is prepared for destaining and treatment for immunocytochemistry. The cell groups of interest are circled by a pen on the coverslip under the microscope as reference, and then they are circled with a glass etching pen on the back of the slide. This step will make it easier to identify the lesional cells after immunostaining.

5 Molecular immunohistochemistry

Molecular profiling of thyroid carcinomas with aggressive clinical behavior has become a standard of care. Modern immunohistochemistry has proven to be an easily practiced approach in the everyday practice of histopathology to triage advanced tumors for further mutation testing. Specific IHC is available for BRAFV600E mutation, RASQ61R mutation, NTRK rearrangement, and ALK rearrangement. Of note, among the IHC for these altered proteins derived from molecular changes, only IHC for BRAFV600E is approved to be of value in the clinical management of malignant thyroid neoplasms; other mutation-specific IHC only confirms the presence or absence of mutation or rearrangement (83). IHC using mutation-specific antibodies against BRAFV600E (VE1 clone, Spring Bioscience, Pleasanton, CA) provides an alternative inexpensive method for the rapid identification of BRAFV600E mutation-positive thyroid tumors. The overall reported sensitivity and specificity of *BRAF p.V600E* immunostaining with cellblock preparation is 94.4% and 100%, respectively; however, this approach is not recommended for FNA smears and monolayer preparations (84).

Consensus guidelines drawn up by an international expert panel do not recommend IHC for NTRK fusion confirmation; however, in some cases, IHC can be used for preliminary screening (85). Similarly, ALK IHC is suggested as a screening procedure, and FISH analysis is recommended for the final confirmation of ALK rearrangement (86). The IHC for RET should not be considered as an option for pre-screening (71).

IHC also allows the characterization of tumor microenvironment (PD-L1 and CD markers) and its role in predicting the response of

thyroid cancer to immunotherapy (13). Different scoring systems for PD-L1 immunostaining have been approved by the FDA as companion diagnostic tests for patient selection for the treatment of various tumors, such as melanoma and lung cancer. PD-L1 expression in thyroid cancer has been shown to be similar to that in other solid tumors. A study of 407 primary thyroid cancers showed PD-L1 expression in 6.1% of papillary thyroid carcinomas, 7.6% of follicular thyroid carcinomas, and 22.2% of anaplastic thyroid carcinomas at a threshold of 1% (87). In a recent meta-analysis, the frequency of PD-L1 positivity in thyroid tumor cells for different histological types ranged from 7% to 90%. This study also demonstrates the role of PD-L1 expression as a potential prognostic marker of disease recurrence in patients with papillary thyroid carcinoma (88).

The clinical utility of determining PD-L1 expression supports the use of PD-L1 immunotherapy as a part of combination therapy in metastatic and RAI-refractory thyroid cancer (89). The optimal cutoff value for immunohistochemical positivity of PD-L1 immunostaining has not yet been validated in thyroid cancer, and the variability in different studies depends on the selected clone, the immunostaining method, and the morphological interpretation. When PD-L1 immunoreaction is used for treatment purposes, it is mandatory that only membranous staining of PD-L1 is considered positive and not the cytoplasmic expression (13).

A recent review has shown that cytological samples constitute a reliable source for PDL-1 IHC analysis, as evidenced by the tumor-rich specimens and concordant results between cytological and histological specimens (90). This study emphasizes that the fixatives used in today's cytology laboratories do not compromise PD-L1 staining, attesting to the utility of cytological specimens for PD-L1 testing in routine clinical practice. PD-L1 IHC may predict the success of PD-1 blockade therapy in a subset of patients with an anaplastic carcinoma PD-1 tumor proportion score of >=1% (91).

5.1 Technical considerations for immunohistochemistry

Destained smear slides prepared during the rapid onsite evaluation of FNA specimens have been used for IHC in cytology. This technique only allows the use of either one or two immunostains and helps the characterization of the cell of origin (follicular, parafollicular/C-cells, metastatic disease, and hematologic neoplasms). In specimens with limited cellularity, areas of interest on the smear can be circled with a glass pen on the reverse slide to easily locate the cell cluster after immunostaining (Figure 7). As a general rule, international guidelines recommend that cellblock is the cytology preparation of choice for performing a panel of immunostains (15, 92).

Cytology cell blocks are made by employing different fixatives and paraffin to obtain morphology similar to histologic preparations and can be routinely used for IHC. Different techniques can be used to prepare cellblocks either by automatic instruments or ready to use gels and matrices. Quality control assessment for immunostaining with cytological samples is a mandatory requirement for each laboratory. UK NEQAS ICC for the quality of immunocytochemical staining reported that cellblock sections achieved the highest score (93). Cellblock preparations are also recommended for "molecular immunohistochemistry", in which markers are designed to recognize the presence of altered proteins from mutated genes (71, 85, 94).

6 Conclusion

Currently, immunohistochemical and molecular analysis are integral to the diagnosis and management of thyroid neoplasms. Accurate diagnosis and classification of thyroid tumors according to the recent classification scheme can be achieved by employing specific immunostains in both histologic and cytologic specimens (1). New multiplex chromogenic and multiplex fluorescent IHC are emerging technologies that enable the simultaneous detection of multiple biomarkers in a single tissue section. Their development for preclinical research and clinical application has increased extraordinarily in the last 5 years, paving the way for a better understanding of tumorigenesis and clinical behavior, and they are expected to improve the personalized treatment of patients with malignant tumors of the thyroid gland (95).

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Baloch ZW, Asa SL, Barletta JA, Ghossein RA, Juhlin CC, Jung CK, et al. Overview of the 2022 WHO classification of thyroid neoplasms. *Endocr Pathol* (2022) 33:27–63. doi: 10.1007/s12022-022-09707-3

2. Christofer Juhlin C, Mete O, Baloch ZW. The 2022 WHO classification of thyroid tumors: novel concepts in nomenclature and grading. *Endocr Relat Cancer* (2022) 30 (2):e220293. doi: 10.1530/ERC-22-0293

3. Baloch ZW, Cooper DS, Gharin H, Alexander EK. Overview of diagnostic terminology and reporting. In: The Bethesda System for Reporting Thyroid Cytopathology. Definition, Criteria, and Explanatory Note. Eds. Syed A, Cibas E Switzerland: Springer, Verlag.

4. Asa SL. The current histologic classification of thyroid cancer. *Endocrinol Metab Clin North Am* (2019) 48:1–22. doi: 10.1016/j.ecl.2018.10.001

5. Baloch Z, LiVolsi VA. The Bethesda system for reporting thyroid cytology (TBSRTC): from look-backs to look-ahead. *Diagn Cytopathol* (2020) 48:862–6. doi: 10.1002/dc.24385

6. Baloch Z, LiVolsi VA. Fifty years of thyroid pathology: concepts and developments. *Hum Pathol* (2020) 95:46–54. doi: 10.1016/j.humpath.2019.09.008

7. Baloch Z, Mete O, Asa SL. Immunohistochemical biomarkers in thyroid pathology. *Endocr Pathol* (2018) 29:91–112. doi: 10.1007/s12022-018-9532-9

8. Liu H, May K. Disulfide bond structures of IgG molecules: structural variations, chemical modifications and possible impacts to stability and biological function. *MAbs* (2012) 4:17–23. doi: 10.4161/mabs.4.1.18347

9. Spiers JA. Goldberg's theory of antigen-antibody reactions in vitro. *Immunology* (1958) 1:89–102.

10. Liu H, Lin F. Application of immunohistochemistry in thyroid pathology. *Arch Pathol Lab Med* (2015) 139:67–82. doi: 10.5858/arpa.2014-0056-RA

11. Rossi ED, Straccia P, Palumbo M, Stigliano E, Revelli L, Lombardi CP, et al. Diagnostic and prognostic role of HBME-1, galectin-3, and beta-catenin in poorly differentiated and anaplastic thyroid carcinomas. *Appl Immunohistochem Mol Morphol* (2013) 21:237–41. doi: 10.1097/PAI.0b013e3182688d0f

12. Trimboli P, Guidobaldi L, Amendola S, Nasrollah N, Romanelli F, Attanasio D, et al. Galectin-3 and HBME-1 improve the accuracy of core biopsy in indeterminate thyroid nodules. *Endocrine* (2016) 52:39–45. doi: 10.1007/s12020-015-0678-7

13. Agarwal S, Bychkov A, Jung CK. Emerging biomarkers in thyroid practice and research. *Cancers (Basel)* (2021) 14(1):204. doi: 10.3390/cancers14010204

14. Garber JR, Papini E, Frasoldati A, Lupo MA, Harrell RM, Parangi S, et al. American Association of clinical endocrinology and associazione Medici endocrinologi thyroid nodule algorithmic tool. *Endocr Metab Immune Disord Drug Targets* (2021) 21:2104–15. doi: 10.2174/1871530321111230225617

15. Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedus L, et al. Nodules, American association of clinical endocrinologists, American college of endocrinology, and associazione Medici endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules–2016 update. *Endocr Pract* (2016) 22:622–39. doi: 10.4158/EP161208.GL

16. Gavrailov M, Petkov R, Mladenov B, Todorov G, Kutev N. The routine immunohistochemical diagnosis and surgical treatment of medullary carcinoma of the thyroid. *Khirurgiia* (1995) 48:17–9.

17. Fischer S, Asa SL. Application of immunohistochemistry to thyroid neoplasms. *Arch Pathol Lab Med* (2008) 132:359–72. doi: 10.5858/2008-132-359-AOITTN

18. Song S, Kim H, Ahn SH. Role of immunohistochemistry in fine needle aspiration and core needle biopsy of thyroid nodules. *Clin Exp Otorhinolaryngol* (2019) 12(2):224–30. doi: 10.21053/ceo.2018.01011

19. Lin JD. Thyroglobulin and human thyroid cancer. *Clin Chim Acta* (2008) 388:15–21. doi: 10.1016/j.cca.2007.11.002

20. Di Jeso B, Arvan P. Thyroglobulin from molecular and cellular biology to clinical endocrinology. *Endocr Rev* (2016) 37:2–36. doi: 10.1210/er.2015-1090

21. Indrasena BS. Use of thyroglobulin as a tumour marker. *World J Biol Chem* (2017) 8:81–5. doi: 10.4331/wjbc.v8.i1.81

22. Bejarano PA, Nikiforov YE, Swenson ES, Biddinger PW. Thyroid transcription factor-1, thyroglobulin, cytokeratin 7, and cytokeratin 20 in thyroid neoplasms. *Appl Immunohistochem Mol Morphol* (2000) 8:189–94. doi: 10.1097/00129039-200009000-00004

23. Rosai J, DeLellis RA, Carcangui ML, Frable WJ, Tallini G. *Tumors of the thyroid gland*. Washington, DC: Armed Forces Institute of Pathology (2014).

24. Salmaslioglu A, Erbil Y, Citlak G, Ersoz F, Sari S, Olmez A, et al. Diagnostic value of thyroglobulin measurement in fine-needle aspiration biopsy for detecting metastatic lymph nodes in patients with papillary thyroid carcinoma. *Langenbecks Arch Surg* (2011) 396:77–81. doi: 10.1007/s00423-010-0723-1

25. Wartofsky L. Management of low-risk well-differentiated thyroid cancer based only on thyroglobulin measurement after recombinant human thyrotropin. *Thyroid* (2002) 12:583–90. doi: 10.1089/105072502320288438

26. Guan L, Zhao X, Tang L, Chen J, Zhao J, Guo M, et al. Thyroid transcription factor-1: structure, expression, function and its relationship with disease. *BioMed Res Int* (2021) 2021:9957209. doi: 10.1155/2021/9957209

27. Liu L, Wu HQ, Wang Q, Zhu YF, Zhang W, Guan LJ, et al. Association between thyroid stimulating hormone receptor gene intron polymorphisms and autoimmune thyroid disease in a Chinese han population. *Endocr J* (2012) 59:717–23. doi: 10.1507/endocrj.EJ12-0024

28. Chen S, Guan L, Zhao X, Yang J, Chen L, Guo M, et al. Optimized thyroid transcription factor-1 core promoter-driven microRNA-7 expression effectively inhibits the growth of human non-small-cell lung cancer cells. *J Zhejiang Univ Sci B* (2022) 23:915–30. doi: 10.1631/jzus.B2200116

29. Agoff SN, Lamps LW, Philip AT, Amin MB, Schmidt RA, True LD, et al. Thyroid transcription factor-1 is expressed in extrapulmonary small cell carcinomas but not in other extrapulmonary neuroendocrine tumors. *Modern Pathol* (2000) 13:238–42. doi: 10.1038/modpathol.3880044

30. Abram M, Huhtamella R, Kalfert D, Hakso-Makinen H, Ludvikova M, Kholova I. The role of cell blocks and immunohistochemistry in thyroid atypia of undetermined Significance/Follicular lesion of undetermined significance Bethesda category. *Acta Cytol* (2021) 65:257–63. doi: 10.1159/000514906

31. Di Palma T, Filippone MG, Pierantoni GM, Fusco A, Soddu S, Zannini M. Pax8 has a critical role in epithelial cell survival and proliferation. *Cell Death Dis* (2013) 4: e729. doi: 10.1038/cddis.2013.262

32. Di Palma T, Lucci V, de Cristofaro T, Filippone MG, Zannini M. A role for PAX8 in the tumorigenic phenotype of ovarian cancer cells. *BMC Cancer* (2014) 14:292. doi: 10.1186/1471-2407-14-292

33. Huang H, Shi Y, Liang B, Cai H, Cai Q. Iodinated TG in thyroid follicular lumen regulates TTF-1 and PAX8 expression via TSH/TSHR signaling pathway. *J Cell Biochem* (2017) 118:3444–51. doi: 10.1002/jcb.26001

34. Mansouri A, Chowdhury K, Gruss P. Follicular cells of the thyroid gland require Pax8 gene function. *Nat Genet* (1998) 19:87–90. doi: 10.1038/ng0598-87

35. Nonaka D, Tang Y, Chiriboga L, Rivera M, Ghossein R. Diagnostic utility of thyroid transcription factors Pax8 and TTF-2 (FoxE1) in thyroid epithelial neoplasms. *Mod Pathol* (2008) 21:192–200. doi: 10.1038/modpathol.3801002

36. Suzuki A, Hirokawa M, Takada N, Higuchi M, Tanaka A, Hayashi T, et al. Utility of monoclonal PAX8 antibody for distinguishing intrathyroid thymic carcinoma from follicular cell-derived thyroid carcinoma. *Endocr J* (2018) 65:1171–5. doi: 10.1507/endocrj.EJ18-0282

37. Bishop JA, Sharma R, Westra WH. PAX8 immunostaining of anaplastic thyroid carcinoma: a reliable means of discerning thyroid origin for undifferentiated tumors of the head and neck. *Hum Pathol* (2011) 42:1873–7. doi: 10.1016/j.humpath.2011.02.004

38. Findlay DM, Sexton PM. Calcitonin. *Growth Factors* (2004) 22:217–24. doi: 10.1080/08977190410001728033

39. Davey RA, Findlay DM. Calcitonin: physiology or fantasy? *J Bone Miner Res* (2013) 28:973–9. doi: 10.1002/jbmr.1869

40. Algeciras-Schimnich A, Preissner CM, Theobald JP, Finseth MS, Grebe SK. Procalcitonin: a marker for the diagnosis and follow-up of patients with medullary thyroid carcinoma. *J Clin Endocrinol Metab* (2009) 94:861–8. doi: 10.1210/jc.2008-1862

41. Albores-Saavedra J, LiVolsi VA, Williams ED. Medullary carcinoma. *Semin DIAGN Pathol* (1985) 2:137–46.

42. Suzuki A, Hirokawa M, Takada N, Higuchi M, Ito A, Yamao N, et al. Fine-needle aspiration cytology for medullary thyroid carcinoma: a single institutional experience in Japan. *Endocr J* (2017) 64:1099–104. doi: 10.1507/endocrj.EJ17-0238

43. Schmid KW. Histopathology of c cells and medullary thyroid carcinoma. *Recent Results Cancer Res* (2015) 204:41–60. doi: 10.1007/978-3-319-22542-5_2

44. Abraham D, Delbridge L, Clifton-Bligh R, Clifton-Bligh P, Grodski S, Robinson BG, et al. Medullary thyroid carcinoma presenting with an initial CEA elevation. *ANZ J Surg* (2010) 80:831–3. doi: 10.1111/j.1445-2197.2010.05350.x

45. Giovanella L, Crippa S, Cariani L. Serum calcitonin-negative medullary thyroid carcinoma: role of CgA and CEA as complementary markers. *Int J Biol Markers* (2008) 23:129–31. doi: 10.1177/172460080802300212

46. Chernyavsky VS, Farghani S, Davidov T, Ma L, Barnard N, Amorosa LF, et al. Calcitonin-negative neuroendocrine tumor of the thyroid: a distinct clinical entity. *Thyroid* (2011) 21:193–6. doi: 10.1089/thy.2010.0299

47. DeLellies RA, Matias-Guiu X. Medullary thyroid carcinoma: a 25-year perspective. *Endocr Pathol* (2014) 25:21–9. doi: 10.1007/s12022-013-9287-2

48. Adel Hakim S, Mohamed Abd Raboh N. The diagnostic utility of INSM1 and GATA3 in discriminating problematic medullary thyroid carcinoma, thyroid and parathyroid lesions. *Pol J Pathol* (2021) 72:11–22. doi: 10.5114/pjp.2021.106440

49. Maleki Z, Abram M, Dell'Aquila M, Kilic I, Lu R, Musarra T, et al. Insulinoma-associated protein 1 (INSM-1) expression in medullary thyroid carcinoma FNA: a multi-institutional study. *J Am Soc Cytopathol* (2020) 9:185–90. doi: 10.1016/j.jasc.2020.01.005

50. Seok JY, Kang M, De Peralta-Venturina M, Fan X. Diagnostic utility of INSM1 in medullary thyroid carcinoma. *Int J Surg Pathol* (2021) 29:615–26. doi: 10.1177/1066896921995935

51. Rossi ED, Baloch Z. The impact of the 2022 WHO classification of thyroid neoplasms on everyday practice of cytopathology. *Endocr Pathol* (2023) 34:23–33. doi: 10.1007/s12022-023-09756-2

52. Baloch ZW, LiVolsi VA. Prognostic factors in well-differentiated follicular-derived carcinoma and medullary thyroid carcinoma. *Thyroid* (2001) 11:637–45. doi: 10.1089/105072501750362709

53. Cracolici V, Ritterhouse LL, Segal JP, Puranik R, Wanjari P, Kadri S, et al. Follicular thyroid neoplasms: comparison of clinicopathologic and molecular features of atypical adenomas and follicular thyroid carcinomas. *Am J Surg Pathol* (2020) 44:881–92. doi: 10.1097/PAS.0000000000001489

54. Shimbashi W, Sugitani I, Kawabata K, Mitani H, Toda K, Yamada K, et al. Thick tumor capsule is a valuable risk factor for distant metastasis in follicular thyroid carcinoma. *Auris Nasus Larynx* (2018) 45:147–55. doi: 10.1016/j.anl.2017.05.002

55. Vuong HG, Kondo T, Duong UNP, Pham TQ, Oishi N, Mochizuki K, et al. Prognostic impact of vascular invasion in differentiated thyroid carcinoma: a systematic review and meta-analysis. *Eur J Endocrinol* (2017) 177:207–16. doi: 10.1530/EJE-17-0260

56. Xu B, Ghossein R. Evolution of the histologic classification of thyroid neoplasms and its impact on clinical management. *Eur J Surg Oncol* (2018) 44(3):338–47. doi: 10.1016/j.ejso.2017.05.002

57. Xu B, Tuttle RM, Sabra MM, Ganly I, Ghossein R. Primary thyroid carcinoma with low-risk histology and distant metastases: clinicopathologic and molecular characteristics. *Thyroid* (2017) 27:632–40. doi: 10.1089/thy.2016.0582

58. Coli A, Bigotti G, Parente P, Federico F, Castri F, Massi G. Atypical thyroid nodules express both HBME-1 and galectin-3, two phenotypic markers of papillary thyroid carcinoma. *J Exp Clin Cancer Res* (2007) 26:221–7.

59. Park YJ, Kwak SH, Kim DC, Kim H, Choe G, Park DJ, et al. Diagnostic value of galectin-3, HBME-1, cytokeratin 19, high molecular weight cytokeratin, cyclin D1 and p27(kip1) in the differential diagnosis of thyroid nodules. *J Korean Med Sci* (2007) 22:621–8. doi: 10.3346/jkms.2007.22.4.621

60. de Matos LL, Del Giglio AB, Matsubayashi CO, de Lima Farah M, Del Giglio A, da Silva Pinhal MA. Expression of CK-19, galectin-3 and HBME-1 in the differentiation of thyroid lesions: systematic review and diagnostic meta-analysis. *Diagn Pathol* (2012) 7:97. doi: 10.1186/1746-1596-7-97

61. Mataraci EA, Ozguven BY, Kabukcuoglu F. Expression of cytokeratin 19, HBME-1 and galectin-3 in neoplastic and nonneoplastic thyroid lesions. *Pol J Pathol* (2012) 63:58–64.

62. Nechifor-Boila A, Catana R, Loghin A, Radu TG, Borda A. Diagnostic value of HBME-1, CD56, galectin-3 and cytokeratin-19 in papillary thyroid carcinomas and thyroid tumors of uncertain malignant potential. *Rom J Morphol Embryol* (2014) 55:49–56.

63. Dunderovic D, Lipkovski JM, Boricic I, Soldatovic I, Bozic V, Cvejic D, et al. Defining the value of CD56, CK19, galectin 3 and HBME-1 in diagnosis of follicular cell derived lesions of thyroid with systematic review of literature. *Diagn Pathol* (2015) 10:196. doi: 10.1186/s13000-015-0428-4

64. Arcolia V, Journe F, Renaud F, Leteurtre E, Gabius HJ, Remmelink M, et al. Combination of galectin-3, CK19 and HBME-1 immunostaining improves the diagnosis of thyroid cancer. *Oncol Lett* (2017) 14:4183–9. doi: 10.3892/ol.2017.6719

65. Murtezaoglu AR, Gucer H. Diagnostic value of TROP-2 expression in papillary thyroid carcinoma and comparison with HBME-1, galectin-3 and cytokeratin 19. *Pol J Pathol* (2017) 68:1–10. doi: 10.5114/pjp.2017.67610

66. Xin Y, Guan D, Meng K, Lv Z, Chen B. Diagnostic accuracy of CK-19, galectin-3 and HBME-1 on papillary thyroid carcinoma: a meta-analysis. *Int J Clin Exp Pathol* (2017) 10:8130–40.

67. Toy H, Etli O, Celik ZE, Sezgin Alikanoglu A. Associations between nucleus size, and immunohistochemical galectin-3, cytokeratine-19 and hbme-1 markers in thyroid papillary carcinoma: a morphometric analyze. *Pathol Oncol Res* (2019) 25:401–8. doi: 10.1007/s12253-017-0337-9

68. Zargari N, Mokhtari M. Evaluation of diagnostic utility of immunohistochemistry markers of TROP-2 and HBME-1 in the diagnosis of thyroid carcinoma. *Eur Thyroid J* (2019) 8:1–6. doi: 10.1159/000494430

69. Ramkumar S, Sivanandham S. The combined utility of HBME-1 and galectin-3 immunohistochemistry and BRAF V600E mutations in the diagnosis of papillary thyroid carcinoma. *Cureus* (2021) 13:e20339. doi: 10.7759/cureus.20339

70. Bartolazzi A, Orlandi F, Saggiorato E, Volante M, Arecco F, Rossetto R, et al. Galectin-3-expression analysis in the surgical selection of follicular thyroid nodules with indeterminate fine-needle aspiration cytology: a prospective multicentre study. *Lancet Oncol* (2008) 9:543–9. doi: 10.1016/S1470-2045(08)70132-3

71. Saleh HA, Jin B, Barnwell J, Alzohaili O. Utility of immunohistochemical markers in differentiating benign from malignant follicular-derived thyroid nodules. *Diagn Pathol* (2010) 5:9. doi: 10.1186/1746-1596-5-9

72. Kakudo K, Wakasa T, Ohta Y, Yane K, Ito Y, Yamashita H. Prognostic classification of thyroid follicular cell tumors using ki-67 labeling index: risk stratification of thyroid follicular cell carcinomas. *Endocr J* (2015) 62:1–12. doi: 10.1507/endocrj.EJ14-0293

73. Huang L, Wang X, Huang X, Gui H, Li Y, Chen Q, et al. Diagnostic significance of CK19, galectin-3, CD56, TPO and Ki67 expression and BRAF mutation in papillary thyroid carcinoma. *Oncol Lett* (2018) 15:4269–77. doi: 10.3892/ol.2018.7873

74. Saglietti C, La Rosa S, Sykiotis GP, Letovanec I, Bulliard JL, Piana S, et al. Expression of Prox1 in medullary thyroid carcinoma is associated with chromogranin a and calcitonin expression and with Ki67 proliferative index, but not with prognosis. *Endocr Pathol* (2019) 30:138–45. doi: 10.1007/s12022-019-9576-5

75. Tang J, Gui C, Qiu S, Wang M. The clinicopathological significance of Ki67 in papillary thyroid carcinoma: a suitable indicator? *World J Surg Oncol* (2018) 16:100. doi: 10.1186/s12957-018-1384-8

76. Xu P, Wu D, Liu X. A proposed grading scheme for predicting recurrence in medullary thyroid cancer based on the Ki67 index and metastatic lymph node ratio. *Endocrine* (2023). doi: 10.1007/s12020-023-03328-4

77. Mu N, Juhlin CC, Tani E, Sofiadis A, Reihner E, Zedenius J, et al. High ki-67 index in fine needle aspiration cytology of follicular thyroid tumors is associated with increased risk of carcinoma. *Endocrine* (2018) 61:293–302. doi: 10.1007/s12020-018-1627-z

78. Benitez Valderrama P, Castro Calvo A, Rodriganez Riesco L, Regojo Zapata R, Parra Ramirez P. Fibrous variant of hashimoto's thyroiditis as a sign of IgG4-related disease, mimicking thyroid lymphoma: case report. *Endocrinol Diabetes Nutr (Engl Ed)* (2022). doi: 10.1016/j.endien.2022.11.034

79. Li Y, Wang X, Liu Z, Ma J, Lin X, Qin Y, et al. Hashimoto's thyroiditis with increased IgG4-positive plasma cells: using thyroid-specific diagnostic criteria may identify early phase IgG4 thyroiditis. *Thyroid* (2020) 30:251–61. doi: 10.1089/thy.2019.0063

80. Matos T, Almeida MM, Batista L, do Vale S. IgG4-related disease of the thyroid gland. *BMJ Case Rep 14* (2021) 14(3):e238177. doi: 10.1136/bcr-2020-238177

81. Takeshima K, Li Y, Kakudo K, Hirokawa M, Nishihara E, Shimatsu A, et al. Proposal of diagnostic criteria for IgG4-related thyroid disease. *Endocr J* (2021) 68:1–6. doi: 10.1507/endocrj.EJ20-0557

82. Kaur R, Mitra S, Rajwanshi A, Das A, Nahar Saikia U, Dey P. Fine needle aspiration cytology of IgG4-related disease: a potential diagnostic pitfall? *Diagn Cytopathol* (2017) 45:14–21. doi: 10.1002/dc.23617

83. Shonka DCJr., Ho A, Chintakuntlawar AV, Geiger JL, Park JC, Seetharamu N, et al. American Head and neck society endocrine surgery section and international thyroid oncology group consensus statement on mutational testing in thyroid cancer: defining advanced thyroid cancer and its targeted treatment. *Head Neck* (2022) 44:1277–300. doi: 10.1002/hed.27025

84. Smith AL, Williams MD, Stewart J, Wang WL, Krishnamurthy S, Cabanillas ME, et al. Utility of the BRAF p.V600E immunoperoxidase stain in FNA direct smears and cell block preparations from patients with thyroid carcinoma. *Cancer Cytopathol* (2018) 126:406–13. doi: 10.1002/cncy.21992

85. Yoshino T, Pentheroudakis G, Mishima S, Overman MJ, Yeh KH, Baba E, et al. JSCO-ESMO-ASCO-JSMO-TOS: international expert consensus recommendations for tumour-agnostic treatments in patients with solid tumours with microsatellite instability or NTRK fusions. *Ann Oncol* (2020) 31:861–72. doi: 10.1016/j.annonc.2020.03.299

86. Park G, Kim TH, Lee HO, Lim JA, Won JK, Min HS, et al. Standard immunohistochemistry efficiently screens for anaplastic lymphoma kinase rearrangements in differentiated thyroid cancer. *Endocr Relat Cancer* (2015) 22:55–63. doi: 10.1530/ERC-14-0467

87. Ahn S, Kim TH, Kim SW, Ki CS, Jang HW, Kim JS, et al. Comprehensive screening for PD-L1 expression in thyroid cancer. *Endocr Relat Cancer* (2017) 24:97–106. doi: 10.1530/ERC-16-0421

88. Girolami I, Pantanowitz L, Mete O, Brunelli M, Marletta S, Colato C, et al. Programmed death-ligand 1 (PD-L1) is a potential biomarker of disease-free survival in papillary thyroid carcinoma: a systematic review and meta-analysis of PD-L1 immunoexpression in follicular epithelial derived thyroid carcinoma. *Endocr Pathol* (2020) 31:291–300. doi: 10.1007/s12022-020-09630-5

89. Shin MH, Kim J, Lim SA, Kim J, Lee KM. Current insights into combination therapies with MAPK inhibitors and immune checkpoint blockade. *Int J Mol Sci 21* (2020) 21(7):2531. doi: 10.3390/ijms21072531

90. Iaccarino A, Salatiello M, Migliatico I, De Luca C, Gragnano G, Russo M, et al. PD-L1 and beyond: immuno-oncology in cytopathology. *Cytopathology* (2021) 32:596–603. doi: 10.1111/cyt.12982

91. Capdevila J, Wirth LJ, Ernst T, Ponce Aix S, Lin CC, Ramlau R, et al. PD-1 blockade in anaplastic thyroid carcinoma. *J Clin Oncol* (2020) 38:2620–7. doi: 10.1200/JCO.19.02727

92. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, 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. doi: 10.1089/thy.2015.0020

93. Kirbis IS, Maxwell P, Flezar MS, Miller K, Ibrahim M. External quality control for immunocytochemistry on cytology samples: a review of (cytology module) results. *Cytopathology* (2011) 22:230–7. doi: 10.1111/j.1365-2303.2011.00867.x

94. Crescenzi A, Fulciniti F, Bongiovanni M, Giovanella L, Trimboli P. Detecting n-RAS Q61R mutated thyroid neoplasias by immunohistochemistry. *Endocr Pathol* (2017) 28:71–4. doi: 10.1007/s12022-016-9466-z

95. Sheng W, Zhang C, Mohiuddin TM, Al-Rawe M, Zeppernick F, Falcone FH, et al. Multiplex immunofluorescence: a powerful tool in cancer immunotherapy. *Int J Mol Sci 24* (2023) 24(4):3086. doi: 10.3390/ijms24043086

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 12 June 2023
DOI 10.3389/fendo.2023.1161424

OPEN ACCESS

EDITED BY
Andrea Frasoldati,
ASMN, Italy

REVIEWED BY
Alessia Cozzolino,
Sapienza University of Rome, Italy
Evren Üstüner,
Ankara University School of Medicine, Türkiye

*CORRESPONDENCE
Chunsong Kang
kcsdr_09sea@163.com

RECEIVED 08 February 2023
ACCEPTED 18 May 2023
PUBLISHED 12 June 2023

CITATION
Li H, Xue J, Zhang Y, Miao J, Jing L and Kang C (2023) Diagnostic efficacy of a combination of the Chinese thyroid imaging reporting and data system and shear wave elastography in detecting category 4a and 4b thyroid nodules. *Front. Endocrinol.* 14:1161424. doi: 10.3389/fendo.2023.1161424

Diagnostic efficacy of a combination of the Chinese thyroid imaging reporting and data system and shear wave elastography in detecting category 4a and 4b thyroid nodules

Huizhan Li[1], Jiping Xue[1], Yanxia Zhang[1], Junwang Miao[1], Liwei Jing[2] and Chunsong Kang[1*]

[1]Department of Ultrasonography, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Tongji Shanxi Hospital, Third Hospital of Shanxi Medical University, Taiyuan, Shanxi, China, [2]Department of Health Statistics, Shanxi Medical University, Taiyuan, Shanxi, China

Objectives: Differential diagnosis of benign and malignant thyroid imaging reporting and data system (TIRADS) category 4a and 4b nodules can be difficult using conventional ultrasonography (US). The objective of this study was to evaluate the diagnostic efficacy of a combination of the Chinese-TIRADS (C-TIRADS) and shear wave elastography (SWE) in detecting malignant nodules among category 4a and 4b thyroid nodules.

Methods: Among 409 thyroid nodules in 332 patients that we included in this study, 106 thyroid nodules were diagnosed as category 4a and 4b using C-TIRADS. We used SWE to measure the maximum Young's modulus (Emax) values of category 4a and 4b thyroid nodules. We calculated the diagnostic efficacy of only the C-TIRADS, only SWE, and a combination of C-TIRADS with SWE, and compared these, while taking the pathology results as the gold standard.

Results: The area under the ROC curve (AUC), sensitivity, and accuracy values of the combination of C-TIRADS and SWE (0.870, 83.3%, and 84.0%, respectively) were all higher when compared with the values of only the C-TIRADS (0.785, 68.5%, and 78.3%, respectively) or only SWE (0.775, 68.5%, and 77.4%, respectively) in the diagnosis of category 4a and 4b thyroid nodules.

Conclusion: In this study, we found that the combination of C-TIRADS and SWE significantly improved the diagnostic efficacy in detecting malignant nodules among category 4a and 4b thyroid nodules, and this could provide a reference for further use of this combination by clinicians for diagnosis and treatment.

KEYWORDS

papillary thyroid carcinoma, shear wave elastography, thyroid, thyroid imaging reporting and data system, thyroid nodule, ultrasonography

Introduction

Thyroid nodules are very common worldwide, and are detected in approximately 19%–68% of the general population. The majority of these thyroid nodules are benign (1, 2). Ultrasound is the best imaging method for the diagnosis of thyroid nodules. However, it is still difficult to diagnose atypical benign and malignant nodules, which are often classified as category 4a or 4b as per the Thyroid Imaging Reporting and Data System (TIRADS) (3, 4). Category 4a and 4b thyroid nodules usually need to be referred for fine needle aspiration (FNA) biopsy to rule out or confirm malignancy. However, a wide range of malignancy rates for these nodules (3.3%–72.4%) is reported in literature (5). It is, therefore, necessary to identify complementary investigations that can improve the diagnostic efficacy of detecting category 4a and 4b thyroid nodules.

Shear wave elastography (SWE) is used to quantitatively measure tissue stiffness based on Young's modulus values. The maximum value measured using SWE (Emax) is the most commonly used parameter, and it is used extensively in the diagnosis of benign and malignant thyroid nodules. The sensitivity and specificity of SWE for differentiating benign from malignant thyroid nodules are 0.79–0.86 and 0.84–0.90, respectively (6).

The purpose of this study was to evaluate the diagnostic efficacy of using a combination of conventional ultrasonography (US) and SWE to differentiate between benign and malignant category 4a and 4b thyroid nodules. The findings of this research have implications for improving the diagnostic efficacy of detecting such nodules and their clinical management. Considering that there are many different versions of thyroid ultrasound classification systems to diagnose benign and malignant thyroid nodules (7), we chose the Chinese Thyroid Imaging Reporting and Data System (C-TIRADS), which was recently released and more practical given the current status of medical treatment (8).

Materials and methods

Ethics declaration

This study was approved by the Medical Ethics Committee of the Shanxi Bethune Hospital(Taiyuan, Shanxi Province, China). This research was conducted in accordance with the relevant regulations and guidelines, and all participants or their legal guardians gave their signed written informed consent.

Participants

In this study, we enrolled 332 consecutive patients (214 women and 118 men) with 409 thyroid nodules, who were treated in Shanxi Bethune Hospital (Taiyuan, Shanxi Province, China) between January 2019 to October 2021. Their median age was 45 years (range: 28–69 years).

The inclusion criteria were as follows: (1) Patients who underwent thyroid surgery and had positive pathology findings; (2) Patients with complete data, including US indicators and SWE data; and (3) Patients who had not been previously treated for thyroid nodules.

Among the total enrolled patients, 245 patients (73.8%) presented with a single nodule, and 87 patients (26.2%) had multiple nodules. The size of the 409 thyroid nodules ranged from 0.5–3.4 cm.

Histological findings after thyroid surgery were used as a reference for the diagnosis of malignant thyroid nodules.

Ultrasonography examinations

Thyroid US and SWE examinations were performed with an Aixplorer US system (SuperSonic Imagine, Aix-en-Provence, France), which was equipped with an SL15-4 multifrequency linear array transducer. All nodules were examined by the same radiologist who was proficient in performing the SWE imaging procedure with more than 10 years of ultrasound work experience.

Patients were placed in the supine position with the neck fully exposed before the US examination began. As per the C-TIRADS, we assessed six features of each nodule, namely, internal structure, echogenicity, margin, calcification, aspect ratio, and comet-tail artifact. We assigned a corresponding score for each feature, and then the nodules were assigned different C-TIRADS classifications according to their total scores. Additionally, we measured the maximum diameter of each nodule.

SWE was performed with the same US machine and transducer after the US examination. The target nodule was displayed on the long-axis section of the thyroid, and the image was switched to SWE mode (display Young's modulus scale: 0–100 kPa). A region of interest, including the whole target lesion and the surrounding normal thyroid tissue, was identified on the nodule, and the SWE image was captured and stored after stabilizing the image. Subsequently, the Emax value of the nodule was measured using the Q-box in three independent measurements, and the mean of the three Emax values was recorded for analysis.

Scoring system

Two physicians with at least 5 years of ultrasound work experience independently evaluated all the ultrasonic images. In case of a disagreement, a third associate chief physician with more than 10 years of ultrasound work experience evaluated the image, it was discussed among the three physicians, and a consensus was reached.

Abbreviations: AUC, the area under the ROC curve; C-TIRADS, Chinese Thyroid Imaging Reporting and Data System; Emax, the maximum value of Young's modulus;ROC, receiver operating characteristic; SWE, shear wave elastography; TIRADS, Thyroid Imaging Reporting and Data System; US, ultrasonography.

We rated all thyroid nodules according to the C-TIRADS scoring system (8): Solid composition, microcalcifications, markedly hypoechoic, ill-defined or irregular margins, or extrathyroidal extensions, and vertical orientation were considered as malignant ultrasound features, while comet-tail artifacts were considered as indicating benign status. Risk stratification was calculated by adding the number of the above-mentioned malignant ultrasound features and then subtracting one (1) if negative features of the comet-tail artifacts were present.

TIRADS 1 (Score 0): no nodule;

TIRADS 2 (Score-1): benign nodules, including the so-called "white knight" nodules, which are referred to as uniform hyperechoic nodules that appear on a background of Hashimoto's thyroiditis;

TIRADS 3 (Score 0): probably benign nodules, including nodular goiter;

TIRADS 4a (Score 1): low suspicious nodules (malignancy between 2% and 10%), including nodules with macrocalcifications or peripheral calcifications with strong acoustic shadowing;

TIRADS 4b (Score 2): moderately suspicious nodules (malignancy between 10% and 50%);

TIRADS 4c (Score 3, 4): highly suspicious nodules (malignancy between 50% and 90%);

TIRADS 5(Score 5): highly suggestive of malignancy (malignancy >90%), including nodules with a "snowstorm" pattern of microcalcifications;

TIRADS 6: biopsy-proven malignant nodules.

TIRADS 1 to TIRADS 4a were classified as benign, and TIRADS 4b to TIRADS 6 were classified as malignant.

In this study, 106 nodules were diagnosed as category 4a or 4b, which included 63 patients with category 4a nodules and 43 patients with category 4b nodules.

SWE classification standard

These nodules were also diagnosed using SWE, and the diagnostic criteria were based on our previous research results (9):

According to the size of the nodules, we used different cutoff points to diagnose the nodules.

Maximum diameter ≤ 1 cm: Emax ≥ 33.7 kPa, the nodule was diagnosed as malignant;

Maximum diameter 1–2 cm: Emax ≥ 37.7 kPa, the nodule was diagnosed as malignant;

Maximum diameter ≥2 cm: Emax ≥55.1 kPa, the nodule was diagnosed as malignant.

C-TIRADS + SWE classification standard

Then, we diagnosed the nodules using C-TIRADS + SWE:

If Emax ≥ cutoff points, nodules were regarded as having a higher TIRADS category.

Statistical analysis

We used the R software package (R Foundation for Statistical Computing, Vienna, Austria) for all statistical analyses in our study. Two-tailed P <0.05 was considered to be statistically significant. We used the Shapiro–Wilk test for evaluating normality of the distribution. Descriptive statistics were expressed as medians (25^{th} and 75^{th} percentiles) or mean values ± standard deviations for continuous data. We assessed the diagnostic efficacy of each method in detecting category 4a and 4b thyroid nodules using the receiver operating characteristic (ROC) curve analysis. We calculated the area under the ROC curve (AUC), and the AUC values were compared using Z test. The accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated. We used the McNemar test for comparison of sensitivity and specificity between the methods.

Results

Diagnostic efficacy of C-TIRADS in detecting malignant nodules among category 4a and 4b thyroid nodules

106 thyroid nodules were diagnosed using US as category 4a or 4b as per the C-TIRADS. Among them, 63 cases were of category 4a nodules, and the pathology findings identified 17 malignant nodules as papillary thyroid carcinomas, the other 46 cases were benign nodules which included 42 nodular goiters and 4 adenomas. The remaining 43 cases out of 106 were of category 4b nodules, and the pathology findings identified 37 malignant nodules which were papillary thyroid carcinomas, the other 6 cases had benign nodules which were nodular goiters.

The conventional US characteristics of 106 thyroid nodules are presented in Table 1.

The diagnostic efficacy of the C-TIRADS in detecting malignant nodules among category 4a and 4b nodules is presented in Table 2.

Diagnostic efficacy of SWE in detecting malignant nodules among category 4a and 4b nodules

The distribution of Emax values of 106 TIRADS category 4a and 4b nodules is shown in Figure 1. The Emax values of the 106 thyroid nodules were non-normally distributed.

Among category 4a nodules, 33 cases had a maximum diameter ≤ 1 cm, 19 cases had a maximum diameter of 1–2 cm, and 11 cases had a maximum diameter ≥ 2 cm. Among category 4b nodules, there were 27 nodules with a maximum diameter ≤ 1 cm, 12 nodules with a maximum diameter of 1–2 cm, and 4 nodules with a maximum diameter ≥ 2 cm.

According to the above Emax diagnostic criteria, 106 cases of benign and malignant nodules were diagnosed. Of the 63 cases of category 4a nodules, 47 were diagnosed as benign, and 16 cases as malignant. Among the category 4b nodules, 30 cases were diagnosed as malignant, and 13 as benign.

The diagnostic efficacy of SWE in detecting malignant nodules among category 4a and 4b thyroid nodules is presented in Table 2.

10.3389/fendo.2023.1161424

TABLE 1 Conventional US characteristics of 106 category 4a and 4b thyroid nodules.

Characteristics	4a	4b
Nodules(n=106)	63	43
Maximum diameter		
≤1 cm	33	27
1-2 cm	19	12
≥2 cm	11	4
Internal structure		
Solid	61	41
Non solid	2	2
Echogenicity		
Markedly hypoechoic	28	38
Isoechoic or Mixed echoic	35	5
Margin		
ill-defined or irregular	3	35
defined or regular	60	8
Calcification		
Microcalcification	2	4
None or Macrocalcification	61	39
Aspect ratio		
>1	0	6
<1	63	37
Comet-tail artifact		
Present	3	0
None	60	43

Diagnostic efficacy of C-TIRADS + SWE in detecting malignant nodules among category 4a and 4b nodules

According to the diagnostic criteria of C-TIRADS + SWE, 16 cases of C-TIRADS category 4a nodules were reclassified to category 4b, and 30 cases of C-TIRADS category 4b nodules were reclassified to category 4C.

The diagnostic efficacy of C-TIRADS+SWE in detecting malignant nodules among category 4a and 4b nodules is presented in Table 2.

Comparison of diagnostic efficacy of the three diagnostic methods

We drew ROC curves to evaluate the efficacy of three methods in the diagnosis of category 4a and 4b nodules (Figure 2), while taking the pathology results as the gold standard. The AUC values of C-TIRADS, SWE, and C-TIRADS+SWE in the diagnosis of category 4a and 4b nodules were 0.785, 0.775, and 0.870, respectively. The AUC value of C-TIRADS + SWE was significantly higher compared with that of C-TIRADS ($P < 0.05$) or SWE ($P < 0.05$). The diagnostic efficacy parameters are shown in Table 2.

We also compared the sensitivity and specificity values for the diagnosis of category 4a and 4b thyroid nodules among the three diagnostic methods. The sensitivity value of C-TIRADS + SWE (83.3%) was higher than that of C-TIRADS (68.5%) or SWE (68.5%) alone ($P < 0.05$). There was no significant difference among the specificity values of the three diagnostic methods ($P>0.05$).

Discussion

US is the best imaging method for the thyroid, and plays an important role in the diagnosis and management of thyroid nodules. However, there are often some inconsistencies in terms of the terminology used for reporting, or recommendations for management due to subjective interpretation of the images. In view of this, clinicians across many countries set up the Thyroid Imaging Reporting and Data System (TIRADS) which is specific to thyroid nodules, patterned on the Breast Imaging Reporting and Data System (BIRADS) published by the American College of Radiology (ACR). Since 2009, various versions of TIRADS were successively established (10–18), including the Eu-TIRADS (19) of the European Thyroid Association, and ACR-TIRADS published by the ACR (20). However, the difference in the diagnosis and treatment of thyroid nodules across different countries makes it difficult for many doctors to adopt the risk stratification system of ACR-TIRADS or Eu-TIRADS in their respective countries. Therefore, in this study, we chose the recently released C-TIRADS.

TABLE 2 Comparison of diagnostic efficacy of three diagnostic methods.

Methods	AUC	Sensitivity (%)	Specificity (%)	Accuracy(%)	PPV (%)	NPV(%)
C-TIRADS	0.785*	68.5	88.5	78.3	86.0	73.0
SWE	0.775**	68.5	86.5	77.4	84.1	72.6
C-TIRADS+SWE	0.870	83.3$^{\#}$	84.6$^{\#\#}$	84.0	84.9	83.0

* indicates the AUC of C-TIRADS compared with that of SWE, z = 0.18, P > 0.05, the AUC of C-TIRADS compared with that of C-TIRADS + SWE, z = 2.76, P < 0.05; ** indicates the AUC of SWE compared with that of C-TIRADS + SWE, z = 2.25, P < 0.05; $^{\#}$ indicates that the comparison of sensitivity values among the three diagnostic methods had statistical significance (P < 0.05); $^{\#\#}$ indicates that the comparison of specificity values among the three diagnostic methods had no statistical significance (P > 0.05).

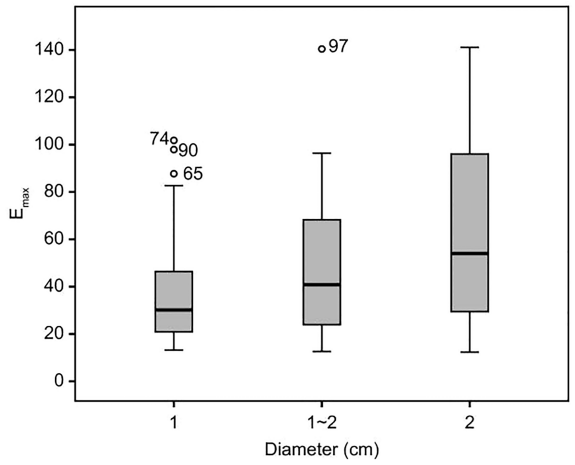


FIGURE 1
Emax values distribution of 106 category 4a and 4b thyroid nodules. The figure was created using R software (version 3.4.4, url: https://www.R-project.org).

The AUC, sensitivity, specificity, and accuracy of US with C-TIRADS for the diagnosis of malignant nodules among category 4a and 4b nodules were 0.785, 68.5%, 88.5%, and 78.3%, respectively, in this study. Compared with previous studies, the diagnostic efficacy of our results was lower than those of US in the diagnosis of benign and malignant thyroid nodules (7, 21). This can be due to the type of thyroid nodules selected for the investigation. In our study, US was only used to evaluate C-TIRADS category 4a and 4b thyroid nodules and not all thyroid nodules. The malignant features of these nodules were often not obvious, and it was difficult to differentiate between benign and malignant nodules using US alone. In addition, we did not study category 4c thyroid nodules because the malignant features of these nodules were relatively obvious and these were easier to diagnose than category 4a and 4b nodules. However, compared with similar studies that evaluated only category 4 thyroid nodules (3), the diagnostic efficacy of our results was higher, and this may be related to C-TIRADS, the ultrasonic diagnostic standard that we selected for this study.

In our study, there was no significant difference in diagnostic efficacy between SWE and US with the C-TIRADS. The AUC, sensitivity, specificity, and accuracy of SWE in the diagnosis of category 4a and 4b nodules were 0.775, 68.5%, 86.5%, and 77.4%, respectively. Diagnosis of TIRADS category 4a and 4b nodules using SWE had high specificity and low sensitivity, and this was consistent with previous studies (22, 23). In our previous study, we found that the size of thyroid nodules had a great impact on the Emax value of SWE (9). Using different diagnostic cut-off points for different sizes of nodules improved the diagnostic efficacy significantly. Therefore, in this study, we used different cut-off points for different sizes of thyroid nodules when SWE was used to diagnose category 4a and 4b thyroid nodules, and hence, the diagnostic efficacy of SWE was better when compared with other similar studies.

In recent years, there have been many reports on the combination of ultrasound classification systems with elastography or contrast-enhanced ultrasound for the diagnosis of thyroid nodules. Most of them believed that combined methods were helpful for the differential diagnosis of thyroid nodules (5, 24–26). Some studies reported that combining SWE or Virtual Touch Tissue Imaging and Quantification (VTIQ) with TI-RADS could improve the diagnostic specificity of thyroid nodules (27). Some researches have shown that the modified TI-RADS based on ACR TI-RADS+ SWE+ CEUS could reduce the frequency of FNA for benign nodules and implement consistent follow-up in clinical practice (28). Some studies have shown that the combination of TI-RADS and CEUS could improve the diagnostic accuracy of thyroid nodules, especially for TI-RADS 4 nodules (29). As we found in the present study too, the combined diagnostic method (C-TIRADS + SWE) significantly improved the diagnostic efficacy in detecting malignant nodules among category 4a and 4b nodules, and the AUC, sensitivity, specificity and accuracy were 0.870, 83.3%, 84.6%, and 84.0%, respectively, which might provide a new standard for diagnosis of such nodules. The

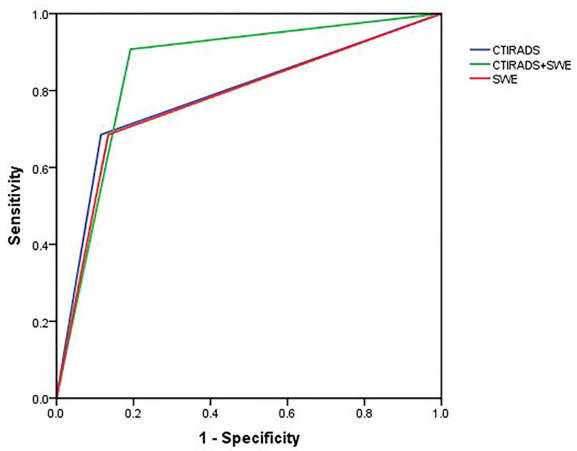


FIGURE 2
ROC curves to evaluate the efficacy of three diagnostic methods for the diagnosis of category 4a and 4b thyroid nodules. The AUC value of C-TIRADS + SWE was higher significantly compared with that of C-TIRADS (z = 2.76, $P < 0.05$) or SWE (z = 2.25, $P < 0.05$). There was no significant difference in AUC value between C-TIRADS and SWE (z = 0.18, $P > 0.05$). The figure was created using R software (version 3.4.4, url: https://www.R-project.org).

improvement in diagnostic efficacy effectively reduced the false positive rate and false negative rate, thereby reducing unnecessary fine-needle aspiration (FNA) or surgery.

Our study had some limitations. First, all patients in our study underwent surgery. Therefore, there might be a bias in the selection of this sample which had an increased proportion of malignancy. Second, in this study, the pathological types were relatively singular and most of them were papillary carcinomas and nodular goiters. The diagnostic performance of the above methods for other thyroid pathological types requires further investigation. Last, the sample size in this study was not large enough and further research with larger samples is required.

Conclusions

In conclusion, in this study, we found that a combination of SWE and US with C-TIRDS was an effective diagnostic method for the differential diagnosis of category 4a and 4b thyroid nodules. While the diagnostic efficacy of these two methods used separately was similar, the combination of SWE and US with the C-TIRADS significantly enhanced the diagnostic efficacy of detecting malignant nodules among category 4a and 4b nodules. This provides a reference for its further use by clinicians in diagnosis and treatment.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Shanxi Bethune Hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Conception and design of the research: CK and HL; Acquisition of data: JX and YZ; Analysis and interpretation of the data: YZ and JM; Statistical analysis: LJ and JM; Obtaining financing: JX; Writing of the manuscript: HL; Critical revision of the manuscript for intellectual content: CK, HL, and LJ. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by the Key Research and Development project of Shanxi Province (grant number 201803D31143).

Acknowledgments

We thank the highly qualified native English speaking editors at AJE for editing the English text of a draft of this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Durante C, Grani G, Lamartina L, Filetti S, Mandel SJ, Cooper DS. The diagnosis and management of thyroid nodules: a review. *JAMA* (2018) 319(9):914–24. doi: 10.1001/jama.2018.0898

2. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid* (2016) 26(1):1–133. doi: 10.1089/thy.2015.0020

3. Pei S, Cong S, Zhang B, Liang C, Zhang L, Liu J, et al. Diagnostic value of multimodal ultrasound imaging in differentiating benign and malignant TI-RADS category 4 nodules. *Int J Clin Oncol* (2019) 24(6):632–9. doi: 10.1007/s10147-019-01397-y

4. Xu T, Gu JY, Ye XH, Xu SH, Wu Y, Shao XY, et al. Thyroid nodule sizes influence the diagnostic performance of TIRADS and ultrasound patterns of 2015 ATA guidelines: a multicenter retrospective study. *Sci Rep* (2017) 7:43183. doi: 10.1038/srep43183

5. Du YR, Ji CL, Wu Y, Gu XG. Combination of ultrasound elastography with TI-RADS in the diagnosis of small thyroid nodules (≤10 mm): a new method to increase the diagnostic performance. *Eur J Radiol* (2018) 109:33–40. doi: 10.1016/j.ejrad.2018.10.024

6. Xu HX, Yan K, Liu BJ, Liu WY, Tang LN, Zhou Q, et al. Guidelines and recommendations on the clinical use of shear wave elastography. *Clin Hemorheol Microcirc* (2019) 72(1):39–60. doi: 10.3233/CH-180452

7. Zhang WB, Xu W, Fu WJ, He BL, Liu H, Deng WF. Comparison of ACR TI-RADS, kwak TI-RADS, ATA guidelines and KTA/KSThR guidelines in combination with SWE in the diagnosis of thyroid nodules. *Clin Hemorheol Microcirc* (2021) 78 (2):163–74. doi: 10.3233/CH-201021

8. Zhou J, Yin L, Wei X, Zhang S, Song Y, Luo B, et al. 2020 Chinese Guidelines for ultrasound malignancy risk stratifification of thyroid nodules: the c-TIRADS. *Endocrine* (2020) 70(2):256–79. doi: 10.1007/s12020-020-02441-y

9. Li HZ, Kang CS, Xue JP, Jing LW, Miao JW. Influence of lesion size on shear wave elastography in the diagnosis of benign and malignant thyroid nodules. *Sci Rep* (2021) 11(1):216216. doi: 10.1038/s41598-021-01114-8

10. Horvath E, Majlis S, Rossi R, Franco C, Niedmann JP, Castro A, et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. *J Clin Endocrinol Metab* (2009) 94(5):1748–51. doi: 10.1210/jc.2008-1724

11. Park JY, Lee HJ, Jang HW, Kim HK, Yi JH, Lee W, et al. A proposal for a thyroid imaging reporting and data system for ultrasound features of thyroid carcinoma. *Thyroid* (2009) 19(11):1257–64. doi: 10.1089/thy.2008.0021

12. Kwak JY, Han KH, Yoon JH, Moon HJ, Son EJ, Park SH, et al. Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratifification of cancer risk. *Radiology* (2011) 260(3):892–9. doi: 10.1148/radiol.11110206

13. Russ G, Bigorgne C, Royer B, Rouxel A, Bienvenu Perrard M. The thyroid imaging reporting and data system (TIRADS) for ultrasound of the thyroid. *J Radio* (2011) 92(7–8):701–13. doi: 10.1016/j.jradio.2011.03.022

14. Maia FF, Matos PS, Pavin EJ, Zantut-Wittmann DE. Thyroid imaging reporting and data system score combined with Bethesda system for malignancy risk stratification in thyroid nodules with indeterminate results on cytology. *Clin Endocrinol (Oxf)* (2015) 82(3):439–44. doi: 10.1111/cen.12525

15. Zayadeen AR, Abu-Yousef M, Berbaum K. JOURNAL CLUB: retrospective evaluation of ultrasound features of thyroid nodules to assess malignancy risk: a step toward TIRADS. *AJR Am J Roentgenol* (2016) 207(3):460–9. doi: 10.2214/AJR.15.15121

16. Songsaeng D, Soodchuen S, Korpraphong P, Suwanbundit A. Siriraj thyroid imaging reporting and data system and its effcacy. *Siriraj Med J* (2017) 69(5):262–7. doi: 10.14456/smj.2017.52

17. Xu SY, Zhan WW, Wang WH. Evaluation of thyroid nodules by a scoring and categorizing method based on sonographic features. *J Ultrasound Med* (2015) 34 (12):2179–85. doi: 10.7863/ultra.14.11041

18. Shin JH, Baek JH, Chung J, Ha EJ, Kim JH, Lee YH, et al. Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean society of thyroid radiology consensus statement and recommendations. *Korean J Radiol* (2016) 17(3):370–95. doi: 10.3348/kjr.2016.17.3.370

19. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European Thyroid association guidelines for ultrasound malignancy risk stratifification of thyroid nodules in adults: the EU-TIRADS. *Eur Thyroid J* (2017) 6(5):225–37. doi: 10.1159/000478927

20. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS committee. *J Am Coll Radiol* (2017) 14(5):587–95. doi: 10.1016/j.jacr.2017.01.046

21. Zhou J, Song Y, Zhan W, Wei X, Zhang S, Zhang R, et al. Thyroid imaging reporting and data system (TIRADS) for ultrasound features of nodules: multicentric retrospective study in China. *Endocrine* (2021) 72(1):157–70. doi: 10.1007/s12020-020-02442-x

22. Huang ST, Zhang B, Yin HL, Li B, Liao JT, Wang YB. Incremental diagnostic value of shear wave elastography combined with contrast-enhanced ultrasound in TI-RADS category 4a and 4b nodules. *J Med Ultrason* (2020) 47(3):453–62. doi: 10.1007/s10396-020-01016-8

23. Chang N, Zhang X, Wan W, Zhang C, Zhang X. The preciseness in diagnosing thyroid malignant nodules using shear-wave elastography. *Med Sci Monit* (2018) 24:671–7. doi: 10.12659/msm.904703

24. Hang J, Li F, Qiao XH, Ye XH, Li A, Du LF. Combination of maximum shear wave elasticity modulus and TIRADS improves the diagnostic specificity in characterizing thyroid nodules: a retrospective study. *Int J Endocrinol* (2018) 2018:4923050. doi: 10.1155/2018/4923050

25. Wang F, Chang C, Chen M, Gao Y, Chen YL, Zhou SC, et al. Does lesion size affect the value of shear wave elastography for differentiating between benign and malignant thyroid nodules? *J Ultrasound Med* (2018) 37(3):601–9. doi: 10.1002/jum.14367

26. Kim H, Kim JA, Son EJ, Youk JH. Quantitative assessment of shear wave ultrasound elastography in thyroid nodules: diagnostic performance for predicting malignancy. *Eur Radiol* (2013) 23(9):2532–7. doi: 10.1007/s00330-013-2847-5

27. Li X, Hou XJ, Du LY, Wu JQ, Wang L, Wang H, et al. Virtual touch tissue imaging and quantifification (VTIQ) combined with the American college of radiology thyroid imaging reporting and data system (ACR TI-RADS) for malignancy risk stratifification of thyroid nodules. *Clin Hemorheol Microcirc* (2019) 72(3):279–91. doi: 10.3233/CH-180477

28. Jin ZQ, Yu HZ, Mo CJ, Su RQ. Clinical study of the prediction of malignancy in thyroid nodules: modifified score versus 2017 American college of radiology's thyroid imaging reporting and data system ultrasound lexicon. *Ultrasound Med Biol* (2019) 45 (7):1627–37. doi: 10.1016/j.ultrasmedbio.2019.03.014

29. Zhang Y, Zhou P, Tian SM, Zhao YF, Li JL, Li L. Usefulness of combined use of contrast-enhanced ultrasound and TI-RADS classifification for the differentiation of benign from malignant lesions of thyroid nodules. *Eur Radiol* (2017) 27(4):1527–36. doi: 10.1007/s00330-016-4508-y

frontiers | Frontiers in Endocrinology

TYPE Opinion
PUBLISHED 21 July 2023
DOI 10.3389/fendo.2023.1209631

OPEN ACCESS

EDITED BY
Andrea Frasoldati,
Endocrine Unit ASMN, Italy

REVIEWED BY
Magdalena Stasiak,
Polish Mother's Memorial Hospital Research Institute, Poland

*CORRESPONDENCE
Priyanka Majety
priyanka.majety@vcuhealth.org

RECEIVED 21 April 2023
ACCEPTED 30 June 2023
PUBLISHED 21 July 2023

CITATION
Majety P (2023) Thyroid nodules: need for a universal risk stratification system. *Front. Endocrinol.* 14:1209631. doi: 10.3389/fendo.2023.1209631

Thyroid nodules: need for a universal risk stratification system

Priyanka Majety*

Department of Endocrinology, Diabetes and Metabolism, Virginia Commonwealth University Health, Richmond, VA, United States

KEYWORDS

thyroid nodule, ultrasound scoring systems, risk stratification, sonographic features, risk calculators

Introduction

Thyroid nodules are common and are one of the most common reasons for endocrinology clinic encounters. The widespread use of various imaging modalities and improved healthcare access have resulted in a significant increase in the discovery of incidental thyroid nodules. About half of the population develops a thyroid nodule by age 60 that can be found either through physical examination or imaging. Thankfully, 85% to 90% prove benign (1–3). However, in the United States, every year over 500 000 fine-needle aspirations (FNAs) are conducted, with about 200 000 of them being unnecessary. Thus, identifying the nodules at the highest risk of malignancy is critical.

Evaluation of patients with a suspected thyroid nodule must include a thorough medical history and physical examination and a thyroid-stimulating hormone (TSH) level and ultrasound (US) evaluation. The sonographic characteristics of these nodules are used to better assess the risk of malignancy (RoM). Based on large studies, US features that are associated with an increased risk of malignancy (hypoechogenicity, solid composition, microcalcifications/punctate echogenic foci, irregular margins, taller than wide shape) and decreased risk of malignancy (isoechoic nodules, spongiform appearance, simple cystic nodules, comet tail artifacts) have been identified (4–6). No single US feature satisfactorily identifies malignant nodules. Over the years, several risk stratification systems (RSSs) that use a combination of these features to help clinicians identify high-risk nodules have been developed. An ideal RSS would minimize the number of unnecessary FNAs and identify all *clinically significant* thyroid cancers, leading to lower healthcare costs and morbidity.

Ultrasound scoring systems

Currently available tools to help clinicians risk-stratify thyroid nodules are:

1) Clinical practice guidelines (CPG) from various professional societies,

2) Scoring systems (qualitative or quantitative),

3) Web-based calculators and

4) An interactive algorithm.

In recent years, artificial intelligence (AI) has shown significant promise in the evaluation of thyroid ultrasounds and in stratifying thyroid nodules.

Several professional organizations have developed ultrasound-based RSSs and management guidelines for thyroid nodules, namely, the American College of Radiology Thyroid Imaging Reporting and Data System (ACR TI-RADS), the American Thyroid Association (ATA) guidelines, the European Thyroid Association (ETA, EU-TIRADS), the Korean Society of Thyroid Radiology/Korean Thyroid Association (KSThR/KTA, K-TIRADS), the Chinese Medical Association (C-TIRADS), the American Association of Clinical Endocrinology (AACE), the American College of Endocrinology (ACE), and the Associazione Medici Endocrinologi (AME) (7–13). There are additional RSSs developed by groups of investigators who do not represent professional organizations.

The characteristics of the commonly used RSSs are outlined in Table 1. The most commonly used ultrasound RSSs are based on the presence of one or more discrete features with one exception. The ATA system uses discrete features and patterns comprised of a combination of these discrete features.

Risk calculators and computer-interpretable guidelines (CIG) are interactive tools where unambiguous, sequential recommendations are made and can be used to engage patients. Table 2 summarizes the various risk calculators available.

Comparison of risk stratification systems

There are considerable differences between the various RSSs. They differ in their formats (pattern recognition versus point systems), risk categories, FNA size thresholds, and in the recommended surveillance intervals (if present). Multiple studies have compared various risk stratification tools, most of them retrospective. No single system has consistently demonstrated

TABLE 1 Characteristics of major ultrasound risk stratification systems [adapted from reference (14)].

RSS	Classification format	Number of categories	Categories and estimated RoM
2021 AACE/ACE-AME tool/ TNAPP	Electronic algorithmic tool that uses history, labs, and combinations of US features	Clinical 2; US features 3	US1 – 1% US2 – 5-15% US3 – 50-90%
2015 ATA	Pattern recognition	5	Benign - <1% Very low - <3% Low - 5–10% Intermediate - 10–20% High - 70–90%
2017 ACR-TIRADS	Point-based system	5	TR1 - <2% TR2 - <2% TR3 - <5% TR4 - 5–20% TR5 - >20%
2017 EU-TIRADS	Algorithmic (combinations of US features)	5	TR1 – None TR2 – 0% TR3 - 2-4% TR4 - 6-17% TR5 - 26-87%
2016 K-TIRADS	Algorithmic (combinations of US features)	5	K-TIRADS 1 – None K-TIRADS 2 – < 3% K-TIRADS 3 – 3-15% K-TIRADS 4 – 15-50% K-TIRADS 5 - > 60%
2020 C-TIRADS	Point-based system	6	C-TR 1 – None C-TR 2 – 0% C-TR 3 – <2% C-TR 4 A– 2-10% C-TR 4 B– 10-50% C-TR 4 C– 50-90% C-TR 5 – >90% C-TR 6 – Proven malignancy

AACE/ACE/AME, American Association of Clinical Endocrinology, American College of Endocrinology, Associazione Medici Endocrinologi; TNAPP, The Thyroid Nodule App; ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System; ATA, American Thyroid Association; EU-TIRADS, European Thyroid Association Thyroid Imaging Reporting and Data System; K-TIRADS, Korean Society of Thyroid Radiology/Korean Thyroid Association Thyroid Imaging Reporting and Data System; RSS, risk stratification system; RoM, risk of malignancy.

TABLE 2 Summary—thyroid nodule risk calculators [Adapted from reference (14)].

Thyroid nodule risk calculators				
Inputs	Inputs		Outputs	Comments
ACR TI-RADS & AI TI-RADS: ***Websites:*** https://deckard.duhs.duke.edu/~ai-ti-rads/index.html	-Composition -Echogenicity -Shape -Margin -Echogenic foci		- Total points - TI-RADS score - FNA recommendation	Most widely used risk calculator, particularly among radiologists. It is restricted to thyroid US features and size. Clinical factors are not taken into consideration.
Malignancy risk estimation of lesions with AUS/FLUS: ***Website:*** http://www.gap.kr/thyroidnodule_b3.php	- Biopsy results (nuclear vs. architectural atypia) - Diameter - Internal content - Shape - Margin - Echogenicity - Calcification		- Total score - RoM in %	It is restricted to nodules with AUS/FLUS. It provides statistics about the RoM but not guidance about whether to perform FNA.
The BWH thyroid nodule risk estimator: ***Website:*** https://thyroidcancerrisk.brighamandwomens.org/	-Age at time of diagnosis -Sex -Largest diameter -Cystic content -Additional nodules (≥1cm)		RoM in %	***Strengths:*** Simple, reproducible (due to relatively objective data used as inputs), and best suited for populations. ***Weaknesses:*** It is best suited for evaluating RoM in populations rather than individual patients (employs only a limited amount of reproducible data). It provides statistics about the RoM but not guidance about whether to perform FNA.
The thyroid nodule malignancy risk calculator - Spain: ***Website:*** https://obgynreference.shinyapps.io/calccdt/	Patient characteristics: -Age -Sex -Family history of thyroid cancer (1st degree relatives) -TSH -Autoimmune thyroiditis (positive antibodies)	Nodule characteristics: -Maximum diameter -Content -Echogenicity -Margins -Calcifications -Shape -Suspicious lymph node	- RoM in % - FNA recommendations	Requires data such as anti-thyroid antibodies, which are not routinely performed in the evaluation of thyroid nodules.
Cleveland Clinic calculator: ***Website:*** https://riskcalc.org/ThyroidCancer/	a) FNA – No: -Shape -Vascularity -TSH -Echo texture -Age -Margin -Tumor size -Calcification	b) FNA – Yes: -Shape -Vascularity -TSH -Echo texture -Calcification -Grooves -Pseudo-inclusions -Cellularity -Colloid (Scant or abundant)	RoM in %	Employs vascularity, which is no longer recognized as a key determinant of thyroid malignancy.
TNAPP: ***Website:*** https://aace-thyroid.deontics.com/dwe/int/public/welcome.jsp a) Clinical features b) US features c) Cytology features			-Eligibility for using TNAPP -AACE US category -AACE clinical category -FNA recommendations -ACR TI-RADS risk category -ACR TI-RADS biopsy advice - RoM in % -If FNA available, recommendations on	***Strengths:*** Interactive, comprehensive, paralleling clinical practice guidelines (CPG) guidance. Integrates clinical, sonographic, and cytologic variables together to assess risk. Limited data required for each recommendation. Guides the clinician at various stages: eligibility to use the application, FNA and follow-up advice, and post-FNA advice. Modifies recommendations as more information is provided.

(Continued)

TABLE 2 Continued

Thyroid nodule risk calculators			
Inputs	Inputs	Outputs	Comments
		molecular testing, surgery, and follow-up	***Weaknesses:*** Requires familiarity with the user interface. Creation of an account login is necessary.

ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System; RoM, risk of malignancy; AUS: atypia of undetermined significance; FLUS: follicular lesion of undetermined significance; BWH: Brigham and Women's Hospital; TNAPP, The Thyroid Nodule App.

superiority over the others (possibly due to differences in the patient populations, inclusion and exclusion criteria, and analytic methods).

A meta-analysis compared five major RSSs, namely, AACE/ACE/AME, ATA, K-TIRADS, ACR TI-RADS, and EU-TIRADS. It included 12 studies with 28,750 nodules (15.2% malignant). In order to avoid the bias arising from the different methodologies of the published studies, summary operating measures that are assumed to be independent of disease prevalence were used, such as the diagnostic odds ratio (DOR). The DOR is the odds of a positive test in those with disease relative to the odds of a positive test in those without disease. The diagnostic odds ratio ranged from 2.2 to 4.9 among the different RSSs. A head-to-head comparison showed a higher relative DOR (RDOR) [1.9, 95% CI (1.3-2.9); P = .002] for ACR-TIRADS [DOR: 5.6, 95% CI (3.4–9.0)] versus ATA [DOR: 2.9, 95% CI (1.3–6.5)] due to a higher relative likelihood ratio for positive results. Similarly, a comparison between ACR-TIRADS [DOR: 4.5, 95% CI (2.5–7.9)] and K-TIRADS [DOR: 2.5 95% CI (1.1-5.6)] showed a higher RDOR [1.8, 95% CI (1.2 – 2.6); P = .002] (15).

Ha et al. studied a total of 2000 consecutive thyroid nodules (≥ 1 cm) in 1802 patients and compared seven society guidelines. Overall, the ACR TI-RADS recommended the fewest "unnecessary" (benign) thyroid nodule FNAs at 25.3%, followed by the 2016 AACE/ACE/AME guidelines (32.5%), ATA (51.7%), and K-TIRADS (56.9%). While the K-TIRADS (94.5%) and ATA (89.6%) guidelines were more sensitive compared with the AACE/ACE/AME (80.4%) and ACR (74.7%), the latter were more specific (ACR 67.3%, AACE/ACE/AME 58%, and ATA 33.2%) (16).

Another meta-analysis compared four RSSs, namely, ACR-TIRADS, EU-TIRADS, ATA, and K-TIRADS. This analysis included 29 different studies with a total of 33,748 nodules with pathological or imaging follow-up. The respective pooled sensitivity and specificity of the various RSSs were:

- ACR-TIRADS: 66% and 91% for category 5 and 95% and 55% for category 4 or 5
- ATA: 74% and 88% for category 5 and 91% and 64% for category 4 or 5
- K-TIRADS: 55% and 95% for category 5 and 89% and 64% for category 4 or 5
- EU-TIRADS: 82% and 90% for category 5 and 96% and 52% for category 4 or 5.

When high-risk categories (categories 4-5) were evaluated, no difference was found between the RSSs (17).

A prospective, observational study from a single thyroid cancer unit of a large hospital analyzed 832 thyroid nodules referred for FNA and compared the performance of five RSSs (ATA, AACE/ACE/AME, ACR TIRADS, EU-TIRADS, and K-TIRADS). All the nodules were classified based on US features and stratified using each of the five RSSs, and the recommendation for FNA was evaluated with the final pathologic diagnosis. After excluding nodules with indeterminate cytology, a total of 502 nodules were included in the final cohort. It was concluded that consistently adhering to any of the RSS guidelines would have reduced the number of FNAs by 17.1% and that ACR-TIRADS allowed the largest reduction (268 of 502) in the number of FNAs with the lowest false-negative rate of 2.2% (95% CI, 95.2% to 99.2%). Although the discriminatory capacities of all the RSSs (except for K-TIRADS) were comparable to that of ACR-TIRADS, they recommended more FNAs (18).

Discussion

With multiple risk stratification tools available, clinicians choose their tools informed by their geography and specialization. Both these factors select for involvement with particular professional societies, many of which have their own validated risk stratification systems. As discussed above, studies comparing the performance of various RSSs have had inconsistent results. This makes it difficult for clinicians to consistently implement an RSS. The wide variety of systems may often lead to confusion on the part of both patients and physicians due to a lack of uniformity. This is relevant, especially in the era of "open notes", where patients can access their health records. It can be a puzzling experience when radiologists and clinicians use multiple RSSs with differing management recommendations. It can also be a time-consuming exercise for clinicians to re-evaluate all the nodules using a different RSS, particularly in the fast-paced clinics.

This also poses a challenge to endocrinologists and other clinicians in training. During clinical training, trainees work with several teaching attendings, and many of them have a different approach to thyroid nodule evaluation, the biggest difference being the RSS in use. Some senior clinicians do not use any specific RSS but go with their *intuition*, while others use different RSSs, reflective of the differences in their training and experience. Some radiologists

include the ACR-TIRADS classification of nodules in their reports, while others do not. Although this system enables clinicians in training to learn and use one of several RSSs to justify a specific recommendation based on the patient's medical history, comorbidities, and preferences, it can be an overwhelming and confusing experience.

Another challenge of US-based RSSs is inter- and intra-observer variability (19). When comparing various RSSs, studies have shown that inter-observer agreement is better for intermediate- and high-suspicion nodules than for low-suspicion nodules (20). In another blinded, multi-center study, 100 electronically recorded thyroid nodule US images were analyzed, and the evaluation was repeated four months later after randomization. The analysis was performed by radiologists and endocrinologists. They were also classified according to the ATA, AACE/ACE/AME, EU-TIRADS, and ACR-TIRADS classifications. The aim of this study was to assess inter- and intra-observer agreement between different thyroid centers and different specialists. They concluded that while the intra-observer reproducibility for thyroid nodule US classification appears fairly adequate, the inter-observer agreement between the different centers is lower than in single-center trials (21). There are still inconsistencies in thyroid US examiners' reporting and rating abilities. A potential solution to this problem is a unified lexicon of thyroid US features and dedicated training. This may increase inter-observer agreement and improve the predictive value of the classification system.

There is a compelling need for a universal risk stratification system that would help not only clinicians but also patients in understanding ultrasound reports and making appropriate recommendations in identifying the nodules that require further evaluation including a biopsy. A grassroots initiative, managed by the steering committee of the International Thyroid Nodule Ultrasound Working Group (ITNUWG), is currently working to develop an international RSS, termed I-TIRADS, that integrates the leading RSSs (22). A recent multidisciplinary international survey conducted by the ITNUWG on RSS-use patterns and practitioner characteristics and preferences confirmed this notion. There were 875 respondents from 52 countries from more than seven specialties. About one-third of the respondents indicated the use of more than one RSS in their practice, potentially leading to confusion, and another third of the respondents reported not using an RSS for various reasons. Most of them supported a comprehensive points-based RSS with no more than five risk categories (23). The majority of them (62% of the respondents) indicated that a universal lexicon paired with illustrative images of ultrasound features would improve inter-observer variability. They also supported the idea of a comprehensive atlas of thyroid US images and videos and dedicated training on the universal lexicon.

There is a strong need for a universal RSS with a lexicon to harmonize all the current systems and standardize the evaluation of thyroid nodules with the aim of reducing unnecessary thyroid biopsies without jeopardizing the detection of clinically significant malignancies. The development of I-TIRADS is a step towards this vision, but we would still need to wait for validation in large population studies.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Yassa L, Cibas ES, Benson CB, Alexander EK, Krane JF, Barletta JA, et al. Long-term assessment of a multidisciplinary approach to thyroid nodule diagnostic evaluation. *Cancer* (2007) 111(6):508–16. doi: 10.1002/cncr.23116

2. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin* (2012) 62(1):10–29. doi: 10.3322/caac.20138

3. Popoveniuc G, Jonklaas J. Thyroid nodules. *Med Clin North Am* (2012) 96 (2):329–49. doi: 10.1016/j.mcna.2012.02.002

4. Angell TE, Maurer R, Wang Z, Robbins J, Sosa BE, Gofnung Y, et al. A cohort analysis of clinical and ultrasound variables predicting cancer risk in 20,001 consecutive thyroid nodules. *J Clin Endocrinol Metab* (2019) 104(11):5665–72. doi: 10.1210/jc.2019-00664

5. Cappelli C, Castellano M, Pirola I, Gandossi L, Agosti L, Cimino L, et al. The predictive value of ultrasound findings in the management of thyroid nodules. *QJM* (2007) 100(1):29–35. doi: 10.1093/qjmed/hcl121

6. Sipos JA. Advances in ultrasound for the diagnosis and management of thyroid cancer. *Thyroid* (2009) 19(12):1363–72. doi: 10.1089/thy.2009.1608

7. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Eberhardt JDH, et al. ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS committee. *J Am Coll Radiol* (2017) 14(5):587–95. doi: 10.1016/j.jacr.2017.01.046

8. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid* (2016) 26(1):1–133. doi: 10.1089/thy.2015.0020

9. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European Thyroid association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: the EU-TIRADS. *Eur Thyroid J* (2017) 6(5):225–37. doi: 10.1159/000478927

10. Shin JH, Baek JH, Chung J, Ha EJ, Na DG, Jung SL, et al. Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean society of thyroid radiology consensus statement and recommendations. *Korean J Radiol* (2016) 17(3):370–95. doi: 10.3348/kjr.2016.17.3.370

11. Zhou J, Yin L, Wei X, Xue S, Zhang X, Liu C, et al. 2020 Chinese Guidelines for ultrasound malignancy risk stratification of thyroid nodules: the c-TIRADS. *Endocrine* (2020) 70(2):256–79. doi: 10.1007/s12020-020-02441-y

12. Gharib H, Papini E, Garber JR, Duick DS, Hamilton RL, Harrell RM, et al. American association of clinical endocrinologists, american college of endocrinology, and associazione medici endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules–2016 update. *Endocr Pract* (2016) 22 (5):622–39. doi: 10.4158/EP161208.GL

13. Garber JR, Papini E, Frasoldati A, Bartalena L, Hegedüs L, Hansen JC, et al. American Association of clinical endocrinology and associazione Medici endocrinologi thyroid nodule algorithmic tool. *Endocr Pract* (2021) 27(7):649–60. doi: 10.1016/j.eprac.2021.04.007

14. Majety P, Garber JR. Ultrasound scoring systems, clinical risk calculators, and emerging tools. In: *Handbook of thyroid and neck ultrasonography: an illustrated case compendium with clinical and pathologic correlation*. (Cham, Germany: Springer International Publishing) (2023). p. 25–52. doi: 10.1007/978-3-031-18448-2_2

15. Castellana M, Castellana C, Treglia G, Giovanella L, Bruno R, Trimboli P, et al. Performance of five ultrasound risk stratification systems in selecting thyroid nodules for FNA. *J Clin Endocrinol Metab* (2020) 105(5):dgz170. doi: 10.1210/clinem/dgz170

16. Ha EJ, Na DG, Baek JH, Sung JY, Kim JH, Kang SY. US Fine-needle aspiration biopsy for thyroid malignancy: diagnostic performance of seven society guidelines applied to 2000 thyroid nodules. *Radiology* (2018) 287(3):893–900. doi: 10.1148/radiol.2018171074

17. Hoang JK, Middleton WD, Langer JE, Tabár L, Zhang Z, Haas BR, et al. Comparison of thyroid risk categorization systems and fine-needle aspiration recommendations in a multi-institutional thyroid ultrasound registry. *J Am Coll Radiol* (2021) 18(12):1605–13. doi: 10.1016/j.jacr.2021.07.019

18. Grani G, Lamartina L, Ascoli V, Filetti S, Elisei R, Durante C, et al. Reducing the number of unnecessary thyroid biopsies while improving diagnostic accuracy: toward the "Right" TIRADS. *J Clin Endocrinol Metab* (2019) 104(1):95–102. doi: 10.1210/jc.2018-01674

19. Russ G, Trimboli P, Buffet C. The new era of TIRADSs to stratify the risk of malignancy of thyroid nodules: strengths, weaknesses and pitfalls. *Cancers (Basel)* (2021) 13(17):4316. doi: 10.3390/cancers13174316

20. Yim Y, Na DG, Ha EJ, Lim HK, Kim JH, Shin JH, et al. Concordance of three international guidelines for thyroid nodules classified by ultrasonography and diagnostic performance of biopsy criteria. *Korean J Radiol* (2020) 21(1):108–16. doi: 10.3348/kjr.2019.0215

21. Persichetti A, Di Stasio E, Coccaro C, Mirabella R, Campanella L, Giacomelli L, et al. Inter- and intraobserver agreement in the assessment of thyroid nodule ultrasound features and classification systems: a blinded multicenter study. *Thyroid* (2020) 30(2):237–42. doi: 10.1089/thy.2019.0360

22. Tessler F. *I-TIRADS (International thyroid imaging, reporting, and data system) project: roadmap and status*. American Thyroid Association, Chicago,IL, USA. (2019).

23. Hoang JK, Asadollahi S, Durante C, Hegedüs L, Papini E, Tessler FN. An international survey on utilization of five thyroid nodule risk stratification systems: a needs assessment with future implications. *Thyroid* (2022) 32(6):675–81. doi: 10.1089/thy.2021.0558

frontiers | Frontiers in Endocrinology

TYPE Perspective
PUBLISHED 15 August 2023
DOI 10.3389/fendo.2023.1228834

OPEN ACCESS

EDITED BY
Antonino Belfiore,
University of Catania, Italy

REVIEWED BY
Shuhang Xu,
Nanjing University of Chinese Medicine, China

*CORRESPONDENCE
Jeffrey R. Garber
Jgarber@bidmc.harvard.edu
Vivek Patkar
vivek.patkar@deontics.com

RECEIVED 25 May 2023
ACCEPTED 11 July 2023
PUBLISHED 15 August 2023

CITATION
Garber JR and Patkar V (2023) Computer-interpretable guidelines: electronic tools to enhance the utility of thyroid nodule clinical practice guidelines and risk stratification tools.
Front. Endocrinol. 14:1228834.
doi: 10.3389/fendo.2023.1228834

Computer-interpretable guidelines: electronic tools to enhance the utility of thyroid nodule clinical practice guidelines and risk stratification tools

Jeffrey R. Garber[1]* and Vivek Patkar[2]*

[1]Atrius Health, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States, [2]Deontics, London, United Kingdom

Clinicians seeking guidance for evaluating and managing thyroid nodules currently have several resources. The principal ones are narrative clinical guidelines and clinical risk calculators. This paper will review the strengths and weaknesses of both. The paper will introduce a concept of computer interpretable guideline, a novel way of transforming narrative guidelines in to a clinical decision support tool that can provide patient specific recommendations at the point of care. The paper then describes an experience of developing an interactive web based computer interpretable guideline for thyroid nodule management, called Thyroid Nodule Management App (TNAPP). The advantages of this approach and the potential barriers for widespread adaptation are discussed.

KEYWORDS

thyroid nodule, clinical practice guideline, clinical calculators, clinical decision support, computer interpretable guideline, evidence based care, guideline compliance, CDSS

Background

Thyroid nodules are a common clinical problem. Increasing availability and the use of ultrasensitive imaging modalities have resulted in the over-detection of incidental thyroid nodules. A meta-analysis showed that 68.8% of all thyroid nodules undergoing surgical excision represented benign disease (1). Over diagnosis and over-treatment of thyroid nodules is a well-known challenge and has economic as well as individual health consequences (2). Deciding on the optimal management of a thyroid nodule and avoiding both unnecessary evaluation and treatment of benign nodules as well as missing thyroid cancer could be a challenging task for a nonspecialist. Clinicians seeking guidance for evaluating and managing thyroid nodules have a number of resources currently available to help them in their clinical decision-making. Available clinical

resources fall into two broad formalisms: (A) Narrative clinical practice guidelines (CPGs)[1] and (B) Clinical risk calculators [2]. This paper provides a brief overview of both, highlighting the strengths and weaknesses of each formalism. The paper then introduces a lesser-known formalism known as computer-interpretable guideline (CIG), a derivative of conventional CPG, which harnesses the power of computational logic, workflow engines[3], and artificial intelligence (AI) to deliver patient-specific recommendations at the point of care. CIGs can overcome many of the limitations of narrative guidelines and clinical calculators. Lastly, the paper discusses the advantages and the barriers to the adaptation of CIGs into clinical practice.

Clinical practice guidelines

The CIGs are "systematically developed statements to assist practitioner and patient decision-making (3) and are usually published in the form of narrative documents. The CPGs provide a number of actionable recommendations of varying degrees of evidence strength, ranging from high-quality evidence (RCTs and meta-analysis) to expert opinions. CPGs are typically developed by a variety of specialist bodies such as professional associations, healthcare providers, or the national bodies entrusted with the task of overseeing clinical standards. A typical guideline-developing group is often multidisciplinary in nature, and the guideline-development process requires an exhaustive literature review, evaluation of evidence, and a consensus process. CPGs may include "clinical algorithms" in the form of flowcharts or a risk stratification model; however, these are usually intended for humans to read, internalize, and apply their recommendations when the appropriate situation arises. Once the guideline is written and published, it is disseminated using various paper and electronic dissemination routes. A systematic review specifically looking at CPGs on the management of thyroid nodules identified 10 guidelines published by different professional organizations, and the overall quality ranged from 3.0 to 6.25 on a seven-point AGREE-II scale (4). The study found that CPGs varied in methodological quality, and increased efforts are required to improve the quality of recommendations on the diagnosis and management of thyroid nodules and cancer.

The primary intention of a CPG is to reduce unjustified variation, standardize clinical practice, improve the quality of care, and decrease the cost of care. While the intention is laudable, the evidence suggests that the effort that goes into creating them may not be matched by the level of usage and adherence in practice (5–7). There are a number of reasons contributing to the underutilization of CPGs.

- **Dissemination barriers:** The target clinicians are often unaware of the availability of CPGs, and even when they are aware, it is difficult to access, read, and apply the relevant recommendations embedded within a lengthy guideline document.
- **Workflow integration barrier:** The inability to integrate a narrative CPG into the clinical workflow through an electronic medical record (EMR) means the usage is entirely dependent upon clinicians' initiative to remember guideline recommendations and apply them to appropriate clinical scenarios. This is an unrealistic expectation in a busy clinical practice, especially in a generalist environment where the clinician is managing a diverse group of conditions.
- **Ambiguity:** The conventional narrative format of CPGs may introduce the inevitable ambiguity associated with language, and different readers could interpret the same recommendation differently. Consistent with this observation, Huang et al., in their systematic review, found the thyroid nodule guidelines' score on the clarity of presentation varied widely from 39% to 82% (4).
- **Oversimplification:** Another drawback of narrative CPGs is the oversimplification of the clinical logic underpinning the recommendations. A typical CPG generally presents recommendations in the form of narrative statements, flowcharts, and tables, limiting its ability to embody complex clinical logic in order to preserve the legibility of the guideline. As a result, guideline recommendations often do not cover complex clinical scenarios.
- **Lack of validation:** Moreover, there are no standard mechanisms to reliably measure guideline usage and adherence in different situations across different populations and cultures.
- **Lack of mechanism for feedback and refinements:** The lack of workflow integration makes it difficult to collect any user feedback on the validity of guideline recommendations in real-world clinical practice, thus missing an important opportunity to close the loop and continue the refinement of CPG content.

Clinical calculators for risk stratifications

Clinical calculators have become ubiquitous and are used by practitioners in a variety of clinical activities, such as calculating risks, scores, and probabilities, classifying patients into prognostic categories, and calculating derived data such as BMI. Fracture Risk Assessment Tool (FRAX) for calculating the risk of osteoporotic fracture in those with osteopenia (8) and the American College of Cardiology (ACC) Atherosclerotic Cardiovascular Disease (ASCVD) risk estimator (9) for predicting cardiovascular events in those

1 CPG is a systematically developed narrative statement to assist clinicians and patients.

2 Clinical risk calculator is a computable model encoded in a software tool that takes discrete and nondiscrete data elements as its input and provides risk stratification and management advice as its output.

3 Workflow engine is a term commonly used to describe types of clinical decision support systems that can model and enact clinical processes or workflows at the point of care.

without known ASCVID calculators are the ones most frequently utilized. At the time this manuscript was written, FRAX had been employed over 11 million times in the USA and over 40 million times across the globe. The ACC risk estimator, which employs standard, discrete data that are recorded in medical records, can be interfaced with electronic health records such as the EPIC medical record version that one of the author's healthcare system uses. In addition to providing a risk estimation of cardiac events, it provides guidance about the use of aspirin, statins, and blood pressure goals. Similarly, a thyroid nodule calculator determines the risk of cancer to guide patient management decisions.

The most commonly employed thyroid nodule calculator is the American College of Radiology (ACR) TI-RADS (10). An updated artificial intelligence version of this tool, AI-TIRADS (11, 12), that uses a modified scoring system, has recently been developed. There are a number of other thyroid nodule risk calculators. A Korean calculator (13) uses ultrasound features for the evaluation of thyroid nodules with the AUS/FLUS Bethesda III cytology classification. The Brigham and Women's Hospital tool is based on 20,000 cases (14). It provides estimates for populations of patients that are based on relatively objective, reproducible data. Given the limited amount of detailed information, specifically excluding high-risk characteristics, it is not well tailored for an individual patient at high risk for having thyroid cancer. In contradistinction, a calculator from Spain (15) calls for a substantial amount of specific information, including whether there is a history of autoimmune disease or a family history of thyroid cancer. The Cleveland Clinic calculator (16), which was among the earliest thyroid nodule risk calculators developed over a decade ago, serves as an example of the evolution of risk estimation. It uses vascularity, which is no longer used for risk stratification or as a "scoreable item."

Calculators have several advantages in terms of computability and automation, workflow integration, decidability, proven validity, and usage data. They are easy to use, readily accessible, and require little time to employ. Hence, calculators can serve as a " point- of-service " tool. Additionally, they are suited to engaging patients by illustrating how data impact decision-making. For example, would the approach be different if the nodule were bigger, grew larger, or the patient was 10 years older?

However, there are many limitations to thyroid nodule calculators in terms of their overall applicability in the wider management decisions

- **Limited input variables:** It may limit data evaluation to the most reproducible and therefore limited number of items. The Brigham and Women's calculator is an example of this (14). It may restrict data evaluation to thyroid ultrasound features alone, such as TI-RADS and AI TI-RADS (10–12). It may focus on a single FNA result, such as AUS/FLUS, as seen with the Korean version (13).
- **Exclusion of symptoms, signs, and patient preferences:** Nearly all thyroid nodule calculators omit features impacting clinical decision-making such as symptoms, physical examination findings such as a firm or fixed nodule, patient anxiety, or cosmetic concerns.
- **Lack of explainability and actionable advice:** Most clinical calculators are black boxes from the end users' point of view. They do not explain the reasoning used by the algorithm to provide the output to the user. Thus, they do not serve as a tool for teaching clinicians. Some may provide risk statistics but not guidance (13, 14), while others may provide guidance but not statistics (10). They may not provide guidance about follow-up and simply provide a statistic about malignancy risk, leaving decision-making to the clinician using the calculator (12, 13).
- **Dissemination:** Stand-alone clinical calculators that are not integrated into the clinical workflow face the same dissemination barriers as CPGs, as many target users may not be aware of their existence.

We anticipate that calculators will continue to evolve, play a role as a clinician aid, proliferate in number, and serve as a tool to assist clinicians in managing thyroid nodules or other conditions. However, for the better adoption of the calculators, they are required to be automated through integration into clinical workflow and should be a part of the broader digital ecosystem within an EMR (17).

Computer -interpretable guideline: a formalism that enhances narrative guidelines

Clinical decision support system (CDSS) is a term used to describe a diverse group of computer technologies designed to assist clinical decision-making at the point of care. CDSSs have evolved over the last four decades, starting from simple rule-based expert systems to more advanced knowledge representation and workflow management systems (18). Advanced CDSS technologies have made it possible to encode and transform complex narrative guidelines, written primarily for human understanding, into an executable, automated CIG. Many different CIG formalisms have been developed in academia (19) to represent different aspects of clinical guidelines, such as recommendations, evidence, and workflow. The CIG format mitigates many of the limitations of a CPG described earlier. Formal semantics that underpins CIG enforces disambiguation of the clinical guideline logic. For example, a guideline may call for obtaining a serum TSH value and performing an ultrasound. However, it may not specify whether the recommendation is to do so simultaneously or sequentially. While a CIG will clarify and automate the workflow and track recommendations. Also, when a CIG is deployed *via* CDSS and integrated in to an electronic medical record, it can automatically pull the investigation results and clinical data to generate relevant patient-specific recommendations and drive the clinical workflow. Studies have shown them to be effective in the management of chronic disease by improving adherence to CPGs (20, 21). CIGs may also facilitate the testing and validation of guideline recommendations, both prospectively and retrospectively, by comparing recommendation acceptance and outcome data. Potential benefits of CIG, in addition to being trackable, include their use as a stand-alone medical education tool, including, for example, instructing the trainees about the impact of varying data. They can be integrated into

electronic health records, and their use may range from a single patient decision tool with no retained data to storing data on multiple patients. They ultimately hold promise to serve as a registry platform for the entire spectrum of practice sizes, large multispecialty delivery systems, regional and national databases, or research consortiums, thus ultimately becoming a key tool for studying the impact of algorithms and recommendations on clinical outcomes.

Typically, a narrative clinical practice guideline development process and the process of transforming the CPG into a computer-interpretable guideline are disconnected and sequential rather than a joint co-development process. A completed and published CPG is used as input to develop a CIG. Peleg et al., in their 2014 paper (22), describe their experience developing computer-interpretable guidelines based on already published narrative and evidence-based AACE, AME, and ETA guidelines for the diagnosis and management of thyroid nodules (23). One of the learning lessons from this exercise was that the narrative guideline development process may miss potential refinements and improvements of the guideline recommendations identified during the validation and vetting process of CIG development by the time the narrative guideline is finished, dusted, and already disseminated. The section below describes a novel approach to using retrospective data and the CIG toolset to define, validate, and refine the guideline recommendations.

TNAPP: a novel experiment using CIG and CDS technology to vet and validate thyroid nodule diagnostic and management recommendations

The Thyroid Nodule App (TNAPP) (24) is a novel web-based, readily modifiable, interactive algorithmic tool developed to provide thyroid nodule recommendations using the PROforma CIG formalism (25) and Deontics® commercially available advanced AI-based CDDS technology . The Deontic CDS platform comes with an authoring toolset and a CDS execution engine. The authoring tool converts language or "human understandable guidelines " to a " computer-interpretable logical model." A CDSS engine then runs the logical model using individual patient data to generate patient-specific recommendations. A goal-based cognitive argumentation framework (26) underpins the inference logic of the engine to come up with recommendations. An example of inference logic is illustrated by the following common-practice example. It is raining outside, and you want to stay dry. Variables are how hard it is raining, wind intensity, and the time that will be spent in the rain. The resources to keep you from getting wet are an umbrella, a raincoat that has a hood, and a rain hat. The "data" from the following two examples determine the programmed recommendations about what resources to employ. It is drizzling, and you will only be stepping outside to pick up your morning paper. You may opt out of using any resources or just a rain hat. There is a monsoon, and you are headed on foot to a destination one mile away. Parallelism would provide elected resources, e.g., all or just a raincoat and hood, or a raincoat with a hood and hat. The same approach can be applied to a narrative guideline. In the case of thyroid nodularity, there are clinical factors and ultrasound findings that influence the decision of whether to proceed with a biopsy. If it is done, what actions do the results call for? If a biopsy is not done, does the patient need any follow-up? What follow-up is recommended, and when should it happen?

A prototype CIG in the early stages of being vetted to evaluate a patient with a thyroid nodule (27, 28) provides a comprehensive approach for patients who meet inclusion criteria for employing the tool for decision-making. The variables are clinical factors supporting or not supporting performing a biopsy; ultrasound finding categorization as either low, intermediate, or high risk per AACE/AME guidelines; or a more stratified approach employing ACR TI-RADS. The initial recommendation is whether to perform a fine-needle aspiration (Figure 1). If not, the recommendations are

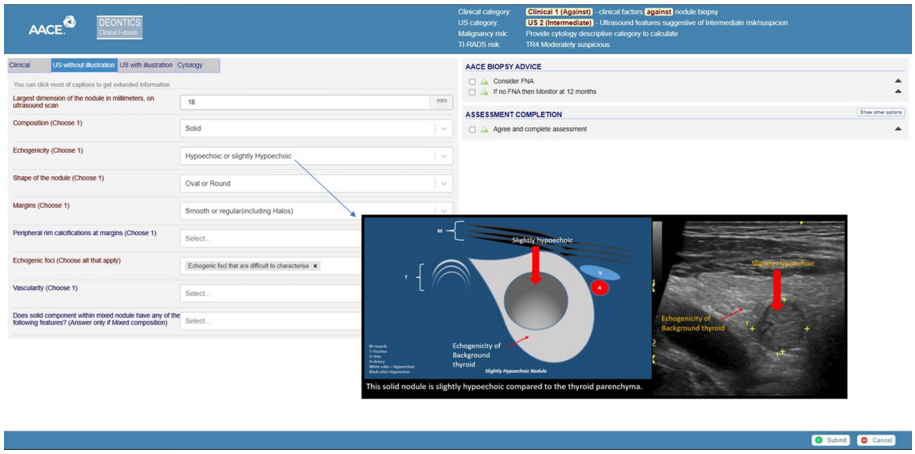


FIGURE 1
Screen capture of a demonstrator tool. The header shows various risk stratification calculator results: clinical and ultrasound risk stratification as clinical 1 and US2 (intermediate). The ACR-TIRADS risk is calculated as TR4. The left side of the screen shows the clinical, ultrasound (US), and cytology data capture tabs. The right side of the screen shows biopsy advice. The link to interactive TNAPP is https://aace-thyroid.deontics.com/.

whether follow-up is at all required and, if so, when it should be done. If an FNA is done, the cytologic Bethesda classification serves as the next determining variable for advice about follow-up and care.

The major challenge for employing CIG will be making it easy to use in a clinical setting. To be implemented in a time-constrained clinical setting, it must require minimal time to employ. It needs to be applicable in settings where resources are not identical. For example, do practitioners have access to or can patients afford molecular markers for evaluating indeterminate cytology? Are highly skilled surgeons available who can perform bilateral thyroidectomy with minimal morbidity when compared with surgery limited to unilateral lobectomy? More than one guideline can be used alone or alongside another, for example, AACE/AME or ACR TI-RADS can be used alone or together. Integration with electronic health records that provide substantial, if not complete, auto-population of requisite data, eliminating the time constraints clinicians face, will facilitate embracing and utilizing CIGs.

Conclusion

CIG should become an adjunct rather than a replacement for clinical practice guideline development. They should be flexible tools that can be customized, readily accessed electronically, and easily modified as new guidance emerges. They have all these potential advantages in addition to facilitating expedited dissemination to a community of users whose feedback can accelerate their refinement, study outcomes, and influence how best to deliver cost-effective patient care when algorithmic approaches apply.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Author contributions

Both authors have equally and substantially contributed to the concept, design and drafting of this paper and have approved the version to be published.

Conflict of interest

Dr. VP is employed by Deontics.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

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):333–9. doi: 10.1159/000339959

2. Uppal N, Collins R, James B. Thyroid nodules: Global, economic, and personal burdens. *Front Endocrinol (Lausanne)* (2023) 14:1113977. doi: 10.3389/fendo.2023.1113977

3. Institute of Medicine (US) Committee to Advise the Public Health Service on Clinical Practice Guidelines. *Clinical Practice Guidelines: Directions for a New Program* Vol. 1990. Field MJ, Lohr KN, editors. Washington (DC: National Academies Press (US) (1990).

4. Huang TW, Lai JH, Wu MY, Chen SL, Wu CH, Tam KW. Systematic review of clinical practice guidelines in the diagnosis and management of thyroid nodules and cancer. *BMC Med* (2013) 11:191. doi: 10.1186/1741-7015-11-191

5. Latosinsky S, Fradette K, Lix L, Hildebrand K, Turner D. Canadian breast cancer guidelines: have they made a difference? *CMAJ* (2007) 176(6):771–6. doi: 10.1503/cmaj.060854

6. LaGrone L, Riggle K, Joshipura M, Quansah R, Reynolds T, Sherr K, et al. Uptake of the World Health Organization's trauma care guidelines: a systematic review. *Bull World Health Organ* (2016) 94(8):585–598C. doi: 10.2471/BLT.15.162214

7. Fox J, Patkar V, Chronakis I, Begent R. From practice guidelines to clinical decision support: closing the loop. *J R Soc Med* (2009) 102(11):464–73. doi: 10.1258/jrsm.2009.090010

8. Kanis JA. Diagnosis of osteoporosis and assessment of fracture risk. *Lancet* (2002) 359(9321):1929–36. doi: 10.1016/S0140-6736(02)08761-5

9. Goff DCJr, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines [published correction appears in Circulation. *Circulation* (2014) 129(25 Suppl 2):S49–73. doi: 10.1161/01.cir.0000437741.48606.98

10. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS committee. *J Am Coll Radiol* (2017) 14(5):587–95. doi: 10.1016/j.jacr.2017.01

11. Wildman-Tobriner B, Buda M, Hoang JK, Middleton WD, Thayer D, Short RG, et al. Using artificial intelligence to revise ACR TI-RADS risk stratification of thyroid nodules: diagnostic accuracy and utility. *Radiology* (2019) 292(1):112–9. doi: 10.1148/radiol.2019182128

12. Weng J, Wildman-Tobriner B, Buda M, Yang J, Ho LM, Allen BC, et al. Deep learning for classification of thyroid nodules on ultrasound: validation on an independent dataset. *Clin Imaging* (2023) 99(11):60–66. doi: 10.1016/j.clinimag.2023.04.010

13. Chen Q, Lin M, Wu S. Validating and comparing C-TIRADS, K-TIRADS and ACR-TIRADS in stratifying the malignancy risk of thyroid nodules. *Front Endocrinol (Lausanne)* (2022) 13:899575. doi: 10.3389/fendo.2022.899575

14. Angell TE, Maurer R, Wang Z, Kim MI, Alexander CA, Barletta JA, et al. A cohort analysis of clinical and ultrasound variables predicting cancer risk in 20,001 consecutive thyroid nodules. *J Clin Endocrinol Metab* (2019) 104(11):5665–72. doi: 10.1210/jc.2019-00664

15. San Laureano FC, Alba JJF, Heras JMJ, Millán AIJ, Fernández-Ladreda MT, Ortega MDCA. Development and internal validation of a predictive model for individual cancer risk assessment for thyroid nodules. *Endocr Pract* (2020) 26 (10):1077–84. doi: 10.4158/EP-2020-0004

16. Nixon IJ, Ganly I, Hann LE, Lin O, Yu C, Brandt S, et al. Nomogram for predicting malignancy in thyroid nodules using clinical, biochemical, ultrasonographic, and cytologic features. *Surgery* (2010) 148(6):1120–8. doi: 10.1016/j.surg.2010.09.030

17. Green TA, Shyu CR. Developing a taxonomy of online medical calculators for assessing automatability and clinical efficiency improvements. *Stud Health Technol Inform* (2019) 264:601–5. doi: 10.3233/SHTI190293

18. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med* (2020) 3:17. doi: 10.1038/s41746-020-0221-y

19. Peleg M. Computer-interpretable clinical guidelines: a methodological review. *J BioMed Inform* (2013) 46(4):744–63. doi: 10.1016/j.jbi.2013.06.009

20. Patkar V, Acosta D, Davidson T, Jones A, Fox J, Keshtgar M. Using computerised decision support to improve compliance of cancer multidisciplinary meetings with evidence-based guidance. *BMJ Open* (2012) 2(3):e000439. doi: 10.1136/bmjopen-2011-000439

21. Patkar V, Hurt C, Steele R, Love S, Purushotham A, Williams M, et al. Evidence-based guidelines and decision support services: A discussion and evaluation in triple assessment of suspected breast cancer. *Br J Cancer* (2006) 95(11):1490–6. doi: 10.1038/sj.bjc.6603470

22. Peleg M, Fox J, Patkar V, Glasspool D, Chronakis I, South M, et al. A computer-interpretable version of the AACE, AME, ETA medical guidelines for clinical practice for the diagnosis and management of thyroid nodules. *Endocr Pract* (2014) 20(4):352–9. doi: 10.4158/EP13271.OR

23. Gharib H, Papini E, Duick DS, Valcavi R, Hegedüs L, Vitti P. AACE, ETA, AME. Guidelines for clinical practice for the diagnosis and management of thyroid nodules. *Endocr Pract* (2010) 16(Suppl 1):1–43.

24. Garber JR, Papini E, Frasoldati A, Lupo MA, Harrell RM, Parangi S, et al. American association of clinical endocrinology and associazione medici endocrinologi thyroid nodule algorithmic tool. *Endocr Pract* (2021) 27(7):649–60. doi: 10.1016/j.eprac.2021.04.007

25. Sutton DR, Fox J. The syntax and semantics of the PROforma guideline modeling language. *J Am Med Inform Assoc* (2003) 10(5):433–43. doi: 10.1197/jamia.M1264

26. Fox J, Cooper RP, Glasspool DW. A canonical theory of dynamic decision-making. *Front Psychol* (2013) 4:150. doi: 10.3389/fpsyg.2013.00150

27. Triggiani V, Lisco G, Renzulli G, Frasoldati A, Guglielmi R, Garber J, et al. The TNAPP web-based algorithm improves thyroid nodule management in clinical practice: A retrospective validation study. *Front Endocrinol (Lausanne)* (2023) 13:1080159. doi: 10.3389/fendo.2022.1080159

28. Majety P, Garber JR. Ultrasound Scoring Systems, Clinical Risk Calculators, and Emerging Tools. In: *Handbook of Thyroid and Neck Ultrasonography: An Illustrated Case Compendium with Clinical and Pathologic Correlation*. Cham: Springer International Publishing (2023). p. 25–52.

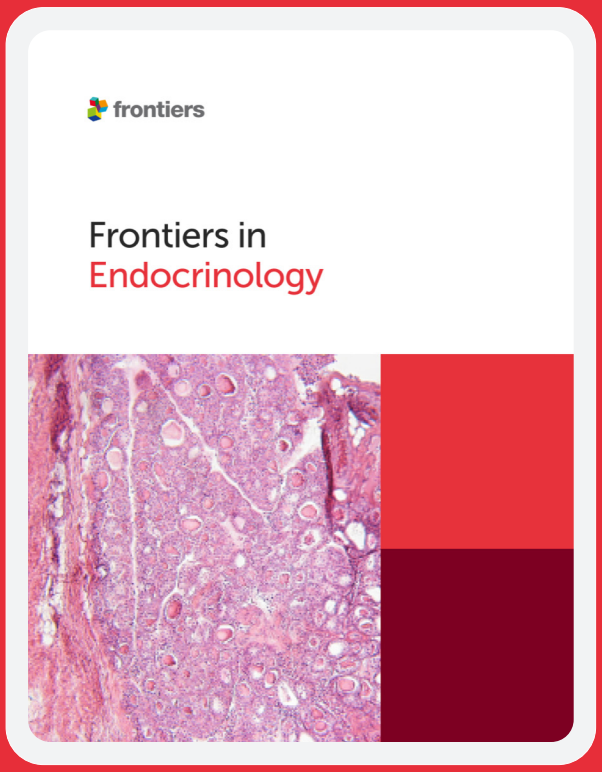
Frontiers in Endocrinology

Explores the endocrine system to find new therapies for key health issues

The second most-cited endocrinology and metabolism journal, which advances our understanding of the endocrine system. It uncovers new therapies for prevalent health issues such as obesity, diabetes, reproduction, and aging.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact