Comparative Efficiency of Diagnostic Performance of Otolaryngologist-Performed Ultrasound in Thyroid Nodules

Jiun-Yi Wu; Ping-Chia Cheng; Chih-Ming Chang; Ming-Hsun Wen; Po-Hsuan Wu; Wu-Chia Lo; Po-Wen Cheng; Li-Jen Liao
Department of Otolaryngology Head and Neck Surgery, Far Eastern Memorial Hospital, New Taipei City 22060, Taiwan

Introduction

Thyroid nodules are highly prevalent, with ultrasound detecting nodules in 40–50% of the general population [1]. Approximately 5% of these nodules are malignant [2], making accurate risk stratification essential for clinical decision-making in otolaryngology–head and neck surgery. Ultrasound is the primary imaging tool for evaluating nodule characteristics, and several ultrasound-based risk stratification systems have been developed to improve diagnostic accuracy. Widely used systems include the American Thyroid Association (ATA) guidelines [3], the American College of Radiology Thyroid Imaging Reporting and Data System (ACR-TIRADS) [4], the Korean Thyroid Imaging Reporting and Data System (K-TIRADS) [5], and the European Thyroid Association system (EU-TIRADS) [6].

Our institution has also developed a real-time scoring system, the FEMH Score [7], which incorporates margin, microcalcifications, echotexture, and shape. The aim of this study was to evaluate and compare the diagnostic performance of these ultrasound risk stratification systems for thyroid nodules.

Material and Methods

This study was approved by the Institutional Review Board of Far Eastern Memorial Hospital (FEMH109074-E). Patients referred to the Department of Otolaryngology–Head and Neck Surgery with suspected thyroid or neck masses were scheduled for ultrasonography in the head and neck ultrasound laboratory. When thyroid nodules were detected, patients were invited to participate in the study and provided written informed consent. Clinical data including age, sex, ultrasound findings, fine-needle aspiration cytology (FNAC) results, and histopathological diagnoses were collected. Only patients who subsequently underwent thyroidectomy and obtained a definitive histological diagnosis were included in the final analysis.

All ultrasound examinations were performed by otolaryngologists with experience in head and neck imaging using high-resolution equipment. The recorded features of each thyroid nodule included the anteroposterior and transverse diameters, margin characteristics (regular or irregular), echogenicity (hyperechoic, isoechoic, or hypoechoic relative to the surrounding thyroid parenchyma), internal structure (solid, predominantly solid, mixed cystic and solid, or cystic, defined as more than 50% cystic component), presence and type of calcification (macrocalcification or microcalcification), and shape (taller-than-wide versus wider-than-tall). Vascularity patterns were assessed by power Doppler ultrasonography. All images and reports were stored in the hospital's PACS system (Marotech, Seoul, South Korea).

Each nodule was classified according to five different ultrasound-based risk stratification systems: the ATA guidelines [3], ACR-TIRADS [4], K-TIRADS [5], EU-TIRADS [6], and the FEMH Score [7]. For ATA, ACR-TIRADS, K-TIRADS, and EU-TIRADS, categories 4 and 5 were defined as malignant. For the FEMH Score, a cutoff value of 3.3 or higher was defined as malignant. The FEMH Score was calculated as follows: 1.25 × margin (regular = 0; irregular = 1) + 2.03 × microcalcification (absent = 0; present = 1) + 1.56 × echotexture (mixed cystic and solid = 0; predominantly solid = 1) + 1.76 × shape (wider-than-tall = 0; taller-than-wide = 1).

All patients underwent US-guided FNAC performed by otolaryngologists. Smears were stained using both Liu's stain and Papanicolaou stain, and the cytology results were classified according to the Bethesda System for Reporting Thyroid Cytopathology (TBSRTC), categories I to VI [8]. Bethesda categories V and VI were considered malignant. The gold standard reference diagnosis was the histopathology of thyroidectomy specimens, which classified nodules as benign or malignant.

Statistical analysis was performed using Fisher's exact test for categorical variables and the Mann–Whitney U test for continuous variables. Diagnostic performance of each system, including ATA, ACR-TIRADS, K-TIRADS, EU-TIRADS, the FEMH Score, and Bethesda cytology, was assessed by calculating sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall accuracy, with corresponding 95% confidence intervals.

Results

Between June 2020 and December 2023, a total of 3,328 patients underwent ultrasound-guided FNAC for thyroid nodules at our institution. Among them, 130 patients subsequently underwent thyroidectomy and were included in the final analysis. The mean age was 50.4 years (range, 27–75 years), and the majority were female. Histopathology confirmed 59 benign nodules (45.4%) and 71 malignant nodules (54.6%).

Analysis of ultrasound features demonstrated that irregular margins, hypoechogenicity, predominantly solid composition, microcalcifications, and a taller-than-wide shape were significantly associated with malignancy ($p < 0.01$). When categories 4 or 5 were defined as malignant, the ATA, ACR-TIRADS, K-TIRADS, and EU-TIRADS systems yielded similar diagnostic performance, with sensitivity of 95.6%, specificity of 78.9%, positive predictive value of 84.6%, negative predictive value of 93.7%, and overall accuracy of 88.1%.

In comparison, the FEMH Score demonstrated a sensitivity of 73.9%, specificity of 100%, positive predictive value of 100%, negative predictive value of 76.0%, and accuracy of 85.7%. The diagnostic profile of the FEMH Score was nearly identical to Bethesda category V or VI cytology [8], emphasizing its high specificity and potential utility as a real-time adjunctive tool.

Figure 1. Flow Chart of Patient Enrollment, Ultrasound Evaluation, and Diagnostic Analysis

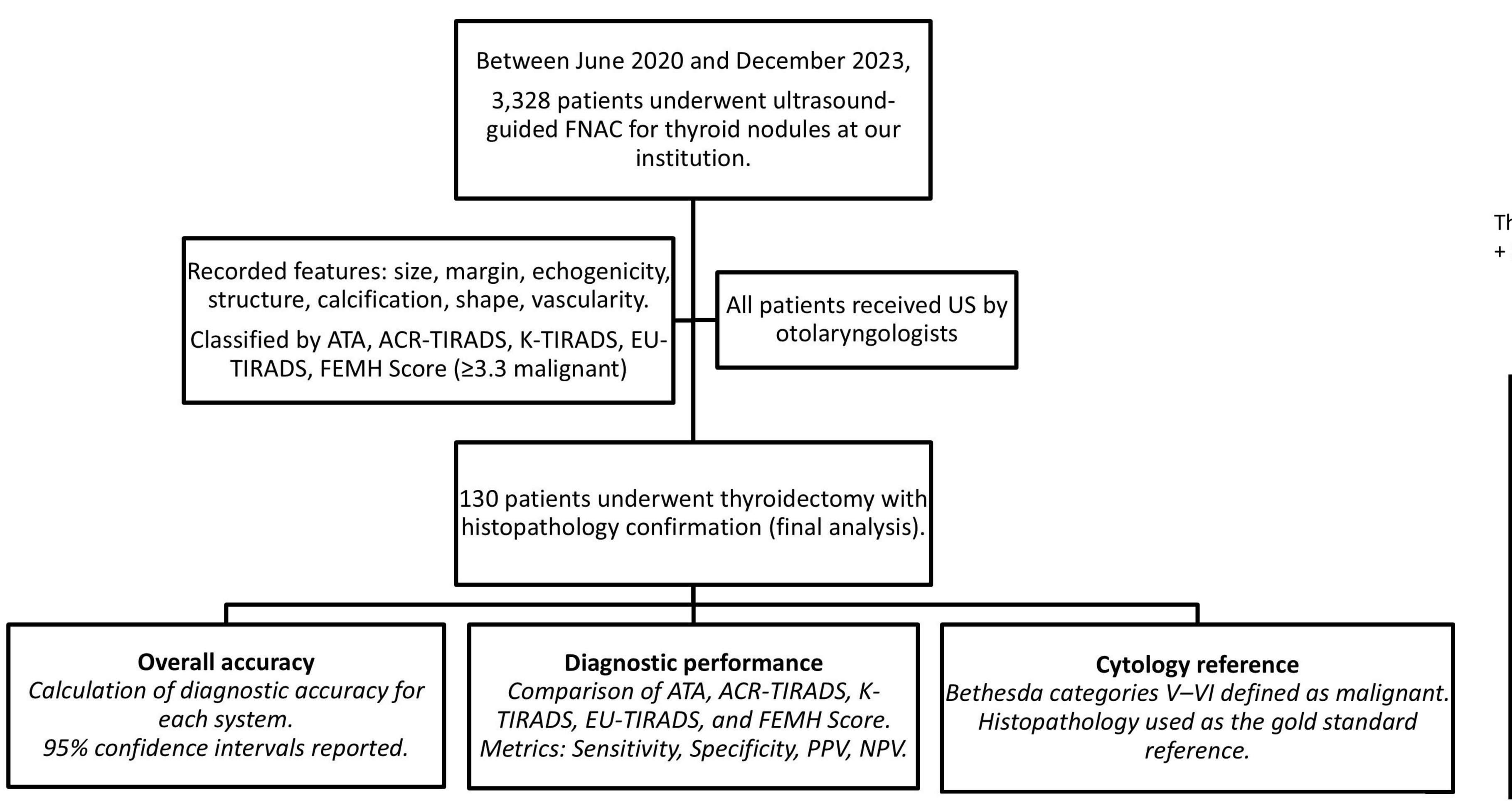


Table 1. Diagnostic performance of ultrasound risk stratification systems and Bethesda cytology

System	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Accuracy % (95% CI)
ATA	95.6 (87.3–104)	78.9 (60.6–97.3)	84.6 (70.7–98.5)	93.7 (81.9–100)	88.1 (78.3–97.9)
ACR-TIRADS	95.6 (87.3–104)	78.9 (60.6–97.3)	84.6 (70.7–98.5)	93.7 (81.9–100)	88.1 (78.3–97.9)
K-TIRADS	95.6 (87.3–104)	78.9 (60.6–97.3)	84.6 (70.7–98.5)	93.7 (81.9–100)	88.1 (78.3–97.9)
EU-TIRADS	95.6 (87.3–104)	78.9 (60.6–97.3)	84.6 (70.7–98.5)	93.7 (81.9–100)	88.1 (78.3–97.9)
FEMH Score	73.9 (56.0–91.9)	100 (100–100)	100 (100–100)	76.0 (59.3–92.7)	85.7 (75.1–96.3)
Bethesda V–VI	73.9 (56.0–91.9)	100 (100–100)	100 (100–100)	76.0 (59.3–92.7)	85.7 (75.1–96.3)

Discussion

This study represents one of the first evaluations of international ultrasound risk stratification systems for thyroid nodules performed exclusively by otolaryngologists in Taiwan. Our results confirm that the ATA, ACR-TIRADS, K-TIRADS, and EU-TIRADS systems all demonstrate high sensitivity and good overall diagnostic performance for malignancy detection, consistent with prior meta-analyses reporting pooled sensitivities above 90% but specificities in the 50–70% range [1,2]. These findings highlight the common trade-off among widely used ultrasound risk stratification systems: they are highly effective in identifying suspicious nodules but often at the cost of false-positive results.

Each system emphasizes different sonographic features. The ATA guidelines adopt a pattern-based approach, stressing hypoechogenicity, irregular or microlobulated margins, microcalcifications, a taller-than-wide shape, and extrathyroidal extension; this design maximizes sensitivity but may reduce specificity [3]. ACR-TIRADS applies a point-based scoring system that integrates composition, echogenicity, shape, margins, and echogenic foci, explicitly aiming to reduce unnecessary biopsies [4]. K-TIRADS is structurally similar to ATA, but it gives additional weight to echogenicity, solidity, and assessment of cervical lymph nodes, which yields excellent sensitivity for malignancy [5]. EU-TIRADS employs a simplified pattern-recognition model, categorizing a nodule as high risk (EU-TIRADS 5) if any major suspicious feature is present, which facilitates application but may decrease specificity compared with ACR-TIRADS [6]. These differences explain the small but clinically relevant variations observed in diagnostic performance among the systems.

Our institution's FEMH Score, incorporating margin, microcalcifications, echotexture, and shape, demonstrated superior specificity compared with international systems and yielded results comparable to Bethesda V/VI cytology. This suggests that the FEMH Score may serve as a valuable adjunct to conventional ultrasound risk stratification systems by providing a highly specific, real-time estimation of malignancy risk at the time of ultrasound. Such an approach can aid physicians in counseling patients immediately after ultrasound-guided FNAC, potentially reducing anxiety associated with waiting several days for cytology reports. The system's specificity makes it particularly useful for avoiding unnecessary interventions and reinforcing clinical decision-making.

This study also underscores the feasibility of otolaryngologists independently performing ultrasound evaluations, FNAC, and surgery within a single department. Continuity of care across the diagnostic and therapeutic pathway minimizes inter-specialty variability and ensures consistent application of stratification criteria. Importantly, our findings show that otolaryngologists can apply standardized ultrasound risk stratification systems with diagnostic accuracy comparable to radiologists and endocrinologists, further broadening the scope of ultrasound practice in head and neck oncology.

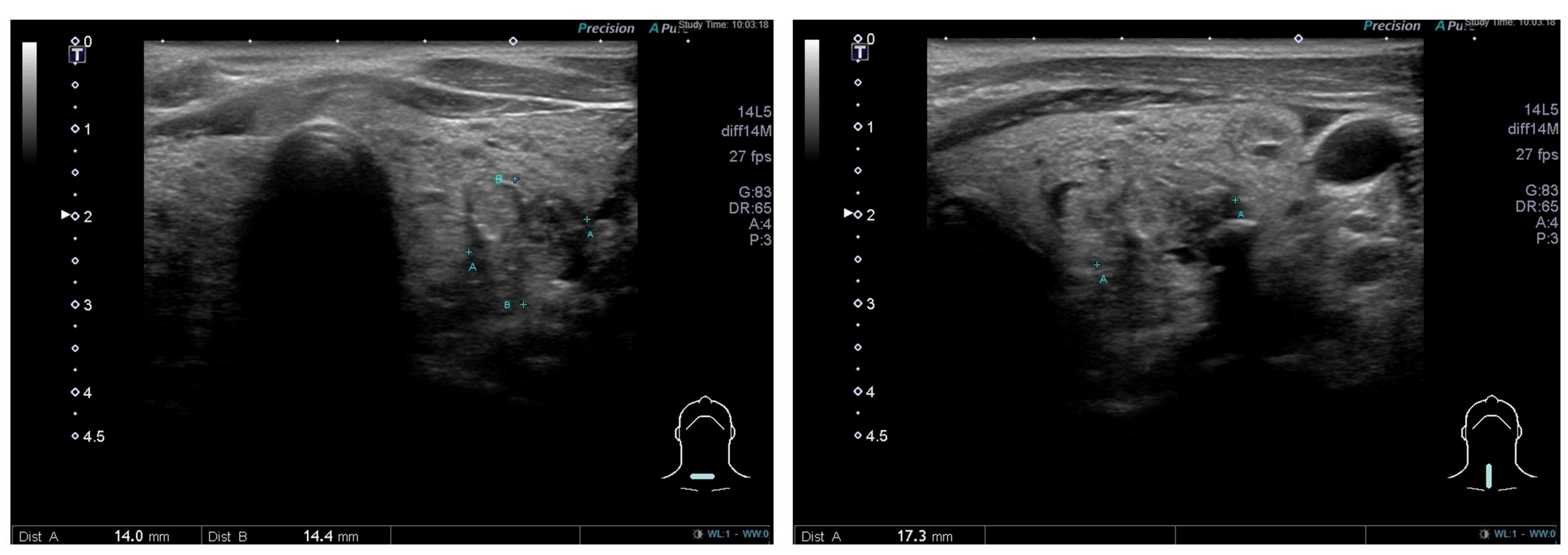
Nevertheless, limitations should be acknowledged. This was a single-center study, and inclusion of only surgically treated patients may have inflated specificity, particularly for the FEMH Score and Bethesda cytology. As sample size increases and multicenter data are incorporated, specificity values will likely align more closely with real-world performance. Operator experience and ultrasound equipment quality remain critical factors affecting diagnostic accuracy. Standardized training and the integration of artificial intelligence to assist in feature recognition, such as microcalcifications or shape analysis, may further enhance reproducibility and precision in the future.

In summary, while international ultrasound risk stratification systems remain highly sensitive tools for malignancy triage, the FEMH Score provides complementary value by offering high specificity and immediate predictive utility. Combining these approaches with FNAC optimizes the diagnostic pathway for thyroid nodules and supports more confident and efficient patient management.

Conclusion

All evaluated ultrasound-based risk stratification systems demonstrated reliable diagnostic performance for thyroid nodules. The FEMH Score developed at our institution showed superior specificity (100%), making it a valuable adjunct for predicting malignancy and guiding patient counseling. A unique feature of this study is that all ultrasound examinations, FNAC, and surgeries were performed exclusively by otolaryngologists, ensuring continuity of care and enhancing physician–patient communication.

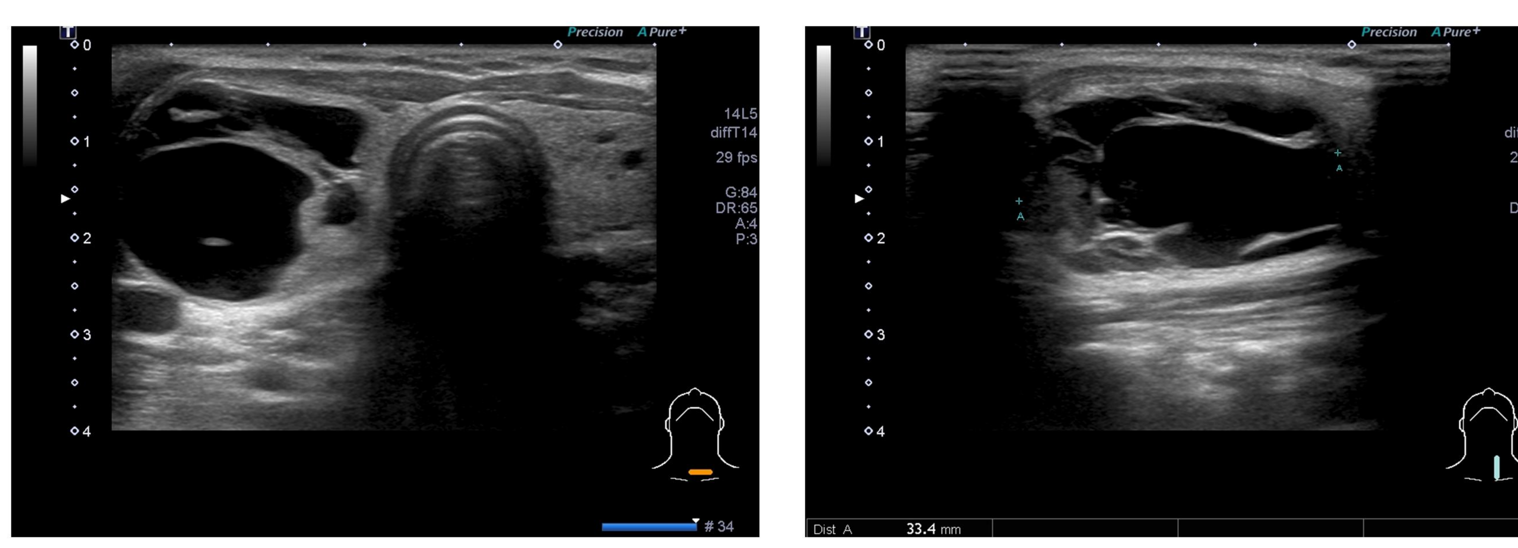
Figure 2. High-Suspicion Thyroid Nodule Classified by Respective Risk Stratification Systems



System	Classification
ATA	5
Korean-TIRADS	5
ACR-TIRADS	5
EU-TIRADS	5
FEMH Score	6.6(≥3.3)

The FEMH Score was calculated as follows: 1.25 × margin (regular = 0; irregular = 1) + 2.03 × microcalcification (absent = 0; present = 1) + 1.56 × echotexture (mixed cystic and solid = 0; predominantly solid = 1) + 1.76 × shape (wider-than-tall = 0; taller-than-wide = 1).

Figure 3. Low-Suspicion Thyroid Nodule Classified by Respective Risk Stratification Systems



System	Classification
ATA	2
Korean-TIRADS	2
ACR-TIRADS	2
EU-TIRADS	2
FEMH Score	0(≤3.3)

Referrence
[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;39:699–706.
[2] Davies L, Welch HG. Increasing incidence of thyroid cancer in the United States, 1973–2002. JAMA. 2006;295:2164–2167.
[3] Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer. Thyroid. 2016;26:1–133.
[4] Tessler FN, Middleton WD, Grant EG, et al. ACR Thyroid Imaging, Reporting and Data System (ACR-TIRADS): White paper of the ACR TIRADS Committee. J Am Coll Radiol. 2017;14:587–595.
[5] Shin JH, Baek JH, Chung J, 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;17:370–395.
[6] 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;6:225–237.
[7] Liao LJ, et al. Real-time scoring system for thyroid nodules (FEMH Score). Unpublished institutional study, Far Eastern Memorial Hospital.
[8] Cibas ES, Ali SZ. The Bethesda System for Reporting Thyroid Cytopathology. Thyroid. 2009;19:1159–1165.