European Thyroid Association

European Thyroid Journal (2024) **13** e230242
https://doi.org/10.1530/ETJ-23-0242

Received 15 November 2023
Accepted 27 February 2024
Available online 27 February 2024
Version of Record published 15 March 2024

SYSTEMATIC REVIEW AND META-ANALYSIS

Head-to-head comparison of American, European, and Asian TIRADSs in thyroid nodule assessment: systematic review and meta-analysis

Tommaso Piticchio[1,2], Gilles Russ[3], Maija Radzina[4,5], Francesco Frasca[1], Cosimo Durante[6] and Pierpaolo Trimboli[2,7]

[1]Endocrinology Section, Department of Clinical and Experimental Medicine, Garibaldi Nesima Hospital, University of Catania, Catania, Italy
[2]Servizio di Endocrinologia e Diabetologia, Ospedale Regionale di Lugano, Ente Ospedaliero Cantonale (EOC), Lugano, Switzerland
[3]Department of Thyroid and Endocrine Tumor Diseases, La Pitie-Salpetriere Hospital, 83 Bd de l'Hopital, Paris, France
[4]Riga Stradins University, Radiology Research Laboratory, Riga, Latvia
[5]University of Latvia, Faculty of Medicine, Riga, Latvia
[6]Department of Translational and Precision Medicine, Sapienza University of Rome, Rome, Italy
[7]Facoltà di Scienze Biomediche, Università della Svizzera Italiana (USI), Lugano, Switzerland

Correspondence should be addressed to T Piticchio: tommaso.piticchio@phd.unict.it

Abstract

Context: Ultrasound-based risk stratification systems (Thyroid Imaging Reporting and Data Systems (TIRADSs)) of thyroid nodules (TNs) have been implemented in clinical practice worldwide based on their high performance. However, it remains unexplored whether different TIRADSs perform uniformly across a range of TNs in routine practice. This issue is highly relevant today, given the ongoing international effort to establish a unified TIRADS (i.e. I-TIRADS), supported by the leading societies specializing in TNs. The study aimed to conduct a direct comparison among ACR-, EU-, and K-TIRADS in the distribution of TNs: (1) across the TIRADS categories, and (2) based on their estimated cancer risk.

Methods: A search was conducted on PubMed and Embase until June 2023. Original studies that sequentially assessed TNs using TIRADSs, regardless of FNAC indication, were selected. General study characteristics and data on the distribution of TNs across TIRADSs were extracted.

Results: Seven studies, reporting a total of 41,332 TNs, were included in the analysis. The prevalence of ACR-TIRADS 1–2 was significantly higher than that of EU-TIRADS 2 and K-TIRADS 2, with no significant difference observed among intermediate- and high-risk categories of TIRADSs. According to malignancy risk estimation, K-TIRADS often classified TNs as having more severe risk, ACR-TIRADS as having moderate risk, and EU-TIRADS classified TNs as having lower risk.

Conclusion: ACR-, EU-, and K-TIRADS assess TNs similarly across their categories, with slight differences in low-risk classifications. Despite this, focusing on cancer risk estimation, the three TIRADSs assess TNs differently. These findings should be considered as a prerequisite for developing the I-TIRADS.

Keywords: TIRADS; thyroid; nodule; risk; malignancy

Published by Bioscientifica Ltd.
https://etj.bioscientifica.com

Introduction

Thyroid nodules (TNs) are frequently found in the general population, especially among women and the elderly. Research has observed TNs in up to 70% of screened adults, with around 5% potentially harboring cancer (1). Considering the epidemiological figures and potential oncological implications, international guidelines recommend an immediate malignancy risk assessment for newly diagnosed TNs, and ultrasound (US) is universally recognized as the first-line diagnostic procedure (2, 3, 4). Despite the worldwide acceptance of US in evaluating the malignancy risk of TN, recent efforts aim to standardize the procedure to further improve its performance (5). In the past decade, prominent scientific societies have proposed US risk stratification systems (RSSs), commonly known as Thyroid Imaging Reporting and Data Systems (TIRADSs) (6). TIRADSs were specifically developed to (1) establish a standard lexicon; (2) define US features associated with specific malignancy risks; (3) assess TNs according to risk classes; and (4) select TNs for fine needle aspiration cytology (FNAC) (7). The advent of TIRADSs has resulted in a substantial increase in published papers, thereby strengthening the evidence in the field. In summary, the accuracy of RSSs in predicting cancers is significantly high, nearly equating to the performance of the highest-risk categories in FNAC (8, 9). Regarding their ability to identify benign or low-risk nodules and prevent ‘unnecessary’ cytological assessments, TIRADSs prompt FNAC at different rates, leading to performance variations across RSSs (10).

However, the full potential of TIRADSs in differentiating between benign and malignant nodules has not yet been fully revealed. Almost all studies in the field assessed the diagnostic efficacy of RSSs by selecting series of TNs suitable for either FNAC or surgery. This type of cohort may not accurately reflect the distribution of TNs across TIRADS categories among patients regularly visiting thyroid disease diagnosis and treatment centers in a real clinical practice context. Only a few of these individuals exhibit nodules appropriate for FNAC or surgery. This implies that numerous published studies, such as original papers and systematic reviews with meta-analysis, may be biased, particularly when assessing the number of potentially avoidable FNAC (i.e. ‘unnecessary’). One might question whether TIRADSs’ ability to differentiate benign from malignant nodules remains consistent in a consecutive series of patients undergoing US, regardless of their suitability for FNAC or further diagnostic procedures.

This systematic review aimed to evaluate the distribution of TNs across the risk categories of different TIRADSs. The focus was on patients primarily referred for a thyroid US evaluation, not for FNAC or preoperative assessment. The secondary objective of the study was to reassess the distribution of TNs within the same TIRADSs categories after consolidating them into a three-tiered scoring system (mild, moderate, severe) based on the estimated risk of malignancy. The most renowned TIRADSs from three different continents, specifically ACR-, EU-, and K-TIRADS (11, 12, 13), were evaluated, and their data were directly compared.

Methods

Conduction of review

This systematic review was conducted following the Meta-analysis of Observational Studies in Epidemiology guidelines (14) (Supplementary Material, see section on supplementary materials given at the end of this article). The study protocol was registered with Prospero under number CRD42023446504.

Search strategy

A literature search was conducted on the PubMed/Medline and Excerpta Medica (Embase) databases using the following search algorithm: (‘EU-TIRADS’ OR ‘EU-TI-RADS’) OR (‘K-TIRADS’ OR ‘K-TI-RADS’) OR (‘ACR-TIRADS’ OR ‘ACR-TI-RADS’). Two investigator authors (TP and PT) independently conducted a duplicate search for papers, screened titles and abstracts, reviewed full texts, and selected studies that met the established inclusion criteria. References from the included studies were further screened for any additional papers. The final electronic search was conducted on June 14, 2023. The search of electronic databases was conducted without any restrictions on date, language, or publication type. Reviewers resolved disagreements through mutual discussion.

Study selection

The study aimed to locate original research reporting on sequential patient series referred to specialized centers for thyroid nodule evaluations, which included a US risk assessment. The present study did not focus on the diagnostic accuracy of TIRADSs, so all data related to FNAC indications and cytological/histological diagnoses were disregarded. The primary data concentrated on the distribution of TNs across the classes of the previously mentioned TIRADSs. The optimal paper selected was essentially an observational study that included consecutive patients referred for thyroid US (regardless of further work-up). The selection criteria for the papers were (a) studies detailing the distribution of TNs series across all ACR-, EU-, and K-TIRADS categories and (b) sequential case enrollment. The exclusion criteria were (a) pediatric patients; (b) studies focusing solely on benign or malignant cases (according to FNAC or pathology report); (c) studies with unclear data or inclusion/exclusion criteria; and (d) studies with overlapping data.

Data extraction

Primary and supplementary data from the included studies were analyzed. The following data were independently collected from all included studies: general study characteristics (author names, year of publication, country of the study), years of enrollment, rates of malignancy and benignancy, number of patients, gender and age of the population, distribution of TNs according to TIRADSs, continent where the study was conducted, US probe frequencies, number of US operators, operator specialization, and reference standard.

The ACR-TIRADS suggests a five-category stratification for TN ultrasound findings, while the EU- and K-TIRADS propose a four-category system. Therefore, considering the similar potential malignancy risk of approximately 2% declared by the ACR-TIRADS for categories 1 and 2, these two classes were combined into one (ACR 1-2) to facilitate a more effective comparison of the three TIRADSs. Relevant authors were contacted to request additional data when necessary. The authors discussed and resolved any discrepancies found during the data cross-checking process.

Assessment of study quality

The risk of bias in the included studies was independently assessed. This study used the National Heart, Lung, and Blood Institute Quality Assessment Tool (15). The evaluated items included: study questions, eligibility criteria, sample size, description and delivery of the intervention, definition of outcome measures, duration of follow-up, blinding, loss to follow-up, and statistical methods. Each domain was assigned a low, high, or not reported score.

Measures

For the initial aim, we evaluated the distribution of TNs across the categories of ACR-, EU-, and K-TIRADS. The second aim involved re-evaluating TNs by categorizing them into TIRADS categories based on their estimated risk of malignancy, which was further divided into a three-tiered score: mild, moderate, and severe risk.

Statistical analysis

The study endpoints guided the performance of multiple meta-analyses to evaluate the distribution of thyroid nodules, with each analysis separately considering two specific settings within the ACR-, EU-, and K-TIRADS systems: (1) the risk categories proposed by TIRADS, and (2) the malignancy risk estimated in TIRADS. Heterogeneity was evaluated using I^2, and a value of ≥50% indicated its presence. A random-effects model was used. The pooled data were presented with 95% CIs. When heterogeneity is detected, meta-regression and subgroup analyses are performed to investigate the causes using various covariates. The continuous covariates included: sample size, duration and years of screening, case enrollment, age of population, female/male ratio, US probe frequencies, number of operators, size of nodules, and malignancy rate. The meta-regression analysis was significant according to the *P*-value. The dichotomous covariates included the study's continent/country, the operator's specialization, and the cytological or histological reference standard. The subgroup analysis revealed a significant difference when the 95% CI of the two groups did not overlap. Statistical significance was set at $P < 0.05$. Statistical analyses were conducted using OpenMeta[Analyst] software, an open-source platform developed by the Center for Evidence Synthesis in Health at Brown University.

Results

Retrieved studies

The search strategy identified a total of 914 records. After eliminating duplicates and scrutinizing titles and abstracts, 34 papers were selected for full-text retrieval. Ultimately, this systematic review included seven studies (16, 17, 18, 19, 20, 21, 22). Figure 1 depicts the process of article search.

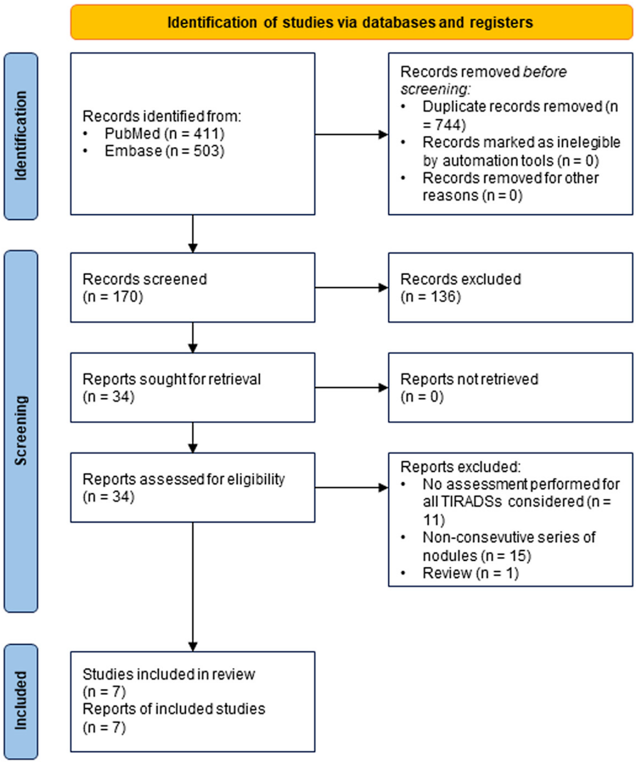

Identification of studies via databases and registers
Identification
Screening
Included
Records identified from:
• PubMed (n = 411)
• Embase (n = 503)
Records removed *before screening:*
• Duplicate records removed (n = 744)
• Records marked as inelegible by automation tools (n = 0)
• Records removed for other reasons (n = 0)
Records screened
(n = 170)
Records excluded
(n = 136)
Reports sought for retrieval
(n = 34)
Reports not retrieved
(n = 0)
Reports assessed for eligibility
(n = 34)
Reports excluded:
• No assessment performed for all TIRADSs considered (n = 11)
• Non-consevutive series of nodules (n = 15)
• Review (n = 1)
Studies included in review
(n = 7)
Reports of included studies
(n = 7)

Figure 1

Flow of article search.

Table 1 General characteristics of the studies included in the present systematic review.

Authors	**Year**	**Country**	**Study period**, y	**Authors' specialization**	**Patients**, *n*	**Nodules**, *n*	**Age**[a], y	**Nodule size**[a] (mm)
Xu *et al.* (16)	2018	China	2	Radiology	2031	2465	47.7	16.63
Qi *et al.* (17)	2021	China	2	Radiology	884	1096	NA	18.86
Hoang *et al.* (18)	2021	USA	8	Radiology	12208	27,933	60.7	15
Seifert *et al.* (19)	2021	Germany	2	Nuclear medicine	849	1211	51	26
Kuru *et al.* (20)	2021	Turkey	8	Radiology	1122	1143	49	NA
Sparano *et al.* (21)	2021	Italy	10	Endocrinology	6401	6474	NA	NA
Orhan Soylemez & Gunduz (22)	2022	Turkey	3	Radiology	977	1010	52	NA

[a]Mean values.

NA, data not available; y, years.

Assessment of study quality

The supplemental material illustrates the risk of bias in the included studies. Overall, 9 out of the 14 items were evaluated as having a low risk of bias in all studies. Considering the study's aim and design, the four items (nos. 7–10) related to the relationship between exposure and outcome were deemed not applicable. No studies provided information regarding power or sample size justification.

Qualitative analysis (systematic review)

This systematic review includes seven papers published between 2018 and 2022. All studies were observational with retrospective data analysis. Four studies were published by European institutions (two from Turkey, one from Italy, and one from Germany), two were from Asia (China), and one from North America (USA). The enrollment period spanned 2–10 years. The study involved 24,472 patients with 41,332 thyroid nodules, all of which underwent US evaluation. The study, conducted between 2008 and 2020, had a sample size ranging from 849 to 12,208 patients. The mean age of patients ranged from 48 to 61 years. Each study reported a range of 1010–27,933 TNs and 849–12,208 patients. Table 1 presents the characteristics of the studies.

All studies were conducted in referral hospital centers by physicians skilled in thyroid US. Radiologists carried out five studies, while endocrinologists and nuclear medicine physicians conducted the remaining two. Most studies referred patients to tertiary hospital centers for comprehensive goiter evaluations. The enrolled cases were independently categorized, regardless of the reason for FNAC and/or surgery.

Quantitative analysis (meta-analysis)

Distribution of TNs across TIRADS categories (from 1 to 5)

Table 2 presents data on the first endpoint (i.e. distribution of TNs across TIRADS categories). Initially, RSSs were evaluated individually, with ACR-TIRADS 4, EU-TIRADS 3 and 5, and K-TIRADS 3 emerging as the categories with the highest pooled call rate (i.e. the highest percentage of nodules). Three statistically significant results were observed: category 4 was most prevalent in ACR-TIRADS (Fig. 2A), while category 2 was least prevalent in both EU-TIRADS and K-TIRADS (Fig. 2B and C). Subsequently, the three RSSs were comparatively analyzed to evaluate the prevalence of thyroid nodules in their respective risk of malignancy categories. The prevalence of ACR-TIRADS 1-2 was significantly higher than that of EU-TIRADS 2 and K-TIRADS 2. K-TIRADS 3 was more prevalent than ACR-TIRADS 3. However, no significant difference was observed among the intermediate- and high-risk categories (Fig. 2A, B, C and Table 2).

Distribution of TNs according to TIRADS risk estimation (mild, moderate, and severe)

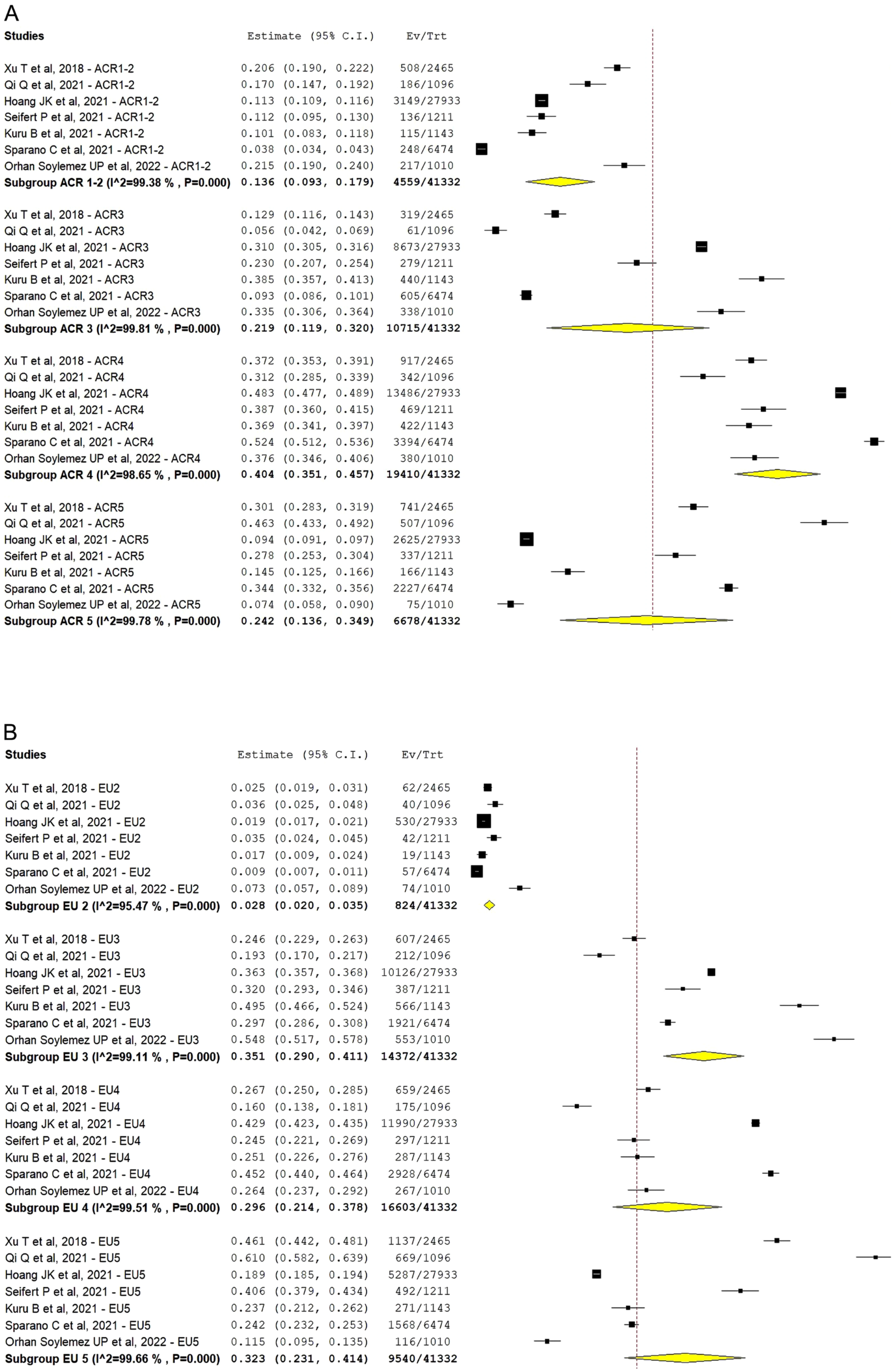
To address the second study aim, the original TIRADS categories were grouped into a three-point system: mild, moderate, and severe risk of malignancy (Fig. 3).

Table 2 Pooled results of TNs assessment according to ACR-, EU-, and K-TIRADS categories.

	TNs, % (95% CI)	I^2 (%)
ACR-TIRADS		
1-2	13.6 (9.3–17.9)[a]	99.38
3	21.9 (11.9–32)[a]	99.81
4	40.4 (35.1–45.7)[b]	98.65
5	24.2 (13.6–34.9)	99.78
EU-TIRADS		
2	2.8 (2–3.5)[b]	95.47
3	35.1 (29–41.1)	99.11
4	29.6 (21.4–37.8)	99.51
5	32.3 (23.1–41.4)	99.66
K-TIRADS		
2	4.9 (3.3–6.5)[b]	97.28
3	39.4 (32.7–46.1)[b,a]	99.2
4	33.8 (27.6–39.9)	99.07
5	21.9 (15–28.7)	99.65

[a]Significantly different with respect to the same categories of the other TIRADSs; [b]Significantly different with respect to the other categories of its TIRADS. .

I^2, heterogeneity; TNs, thyroid nodules.



A

Studies	Estimate (95% C.I.)	Ev/Trt
Xu T et al, 2018 - ACR1-2	0.206 (0.190, 0.222)	508/2465
Qi Q et al, 2021 - ACR1-2	0.170 (0.147, 0.192)	186/1096
Hoang JK et al, 2021 - ACR1-2	0.113 (0.109, 0.116)	3149/27933
Seifert P et al, 2021 - ACR1-2	0.112 (0.095, 0.130)	136/1211
Kuru B et al, 2021 - ACR1-2	0.101 (0.083, 0.118)	115/1143
Sparano C et al, 2021 - ACR1-2	0.038 (0.034, 0.043)	248/6474
Orhan Soylemez UP et al, 2022 - ACR1-2	0.215 (0.190, 0.240)	217/1010
Subgroup ACR 1-2 (I^2=99.38 % , P=0.000)	**0.136 (0.093, 0.179)**	**4559/41332**
Xu T et al, 2018 - ACR3	0.129 (0.116, 0.143)	319/2465
Qi Q et al, 2021 - ACR3	0.056 (0.042, 0.069)	61/1096
Hoang JK et al, 2021 - ACR3	0.310 (0.305, 0.316)	8673/27933
Seifert P et al, 2021 - ACR3	0.230 (0.207, 0.254)	279/1211
Kuru B et al, 2021 - ACR3	0.385 (0.357, 0.413)	440/1143
Sparano C et al, 2021 - ACR3	0.093 (0.086, 0.101)	605/6474
Orhan Soylemez UP et al, 2022 - ACR3	0.335 (0.306, 0.364)	338/1010
Subgroup ACR 3 (I^2=99.81 % , P=0.000)	**0.219 (0.119, 0.320)**	**10715/41332**
Xu T et al, 2018 - ACR4	0.372 (0.353, 0.391)	917/2465
Qi Q et al, 2021 - ACR4	0.312 (0.285, 0.339)	342/1096
Hoang JK et al, 2021 - ACR4	0.483 (0.477, 0.489)	13486/27933
Seifert P et al, 2021 - ACR4	0.387 (0.360, 0.415)	469/1211
Kuru B et al, 2021 - ACR4	0.369 (0.341, 0.397)	422/1143
Sparano C et al, 2021 - ACR4	0.524 (0.512, 0.536)	3394/6474
Orhan Soylemez UP et al, 2022 - ACR4	0.376 (0.346, 0.406)	380/1010
Subgroup ACR 4 (I^2=98.65 % , P=0.000)	**0.404 (0.351, 0.457)**	**19410/41332**
Xu T et al, 2018 - ACR5	0.301 (0.283, 0.319)	741/2465
Qi Q et al, 2021 - ACR5	0.463 (0.433, 0.492)	507/1096
Hoang JK et al, 2021 - ACR5	0.094 (0.091, 0.097)	2625/27933
Seifert P et al, 2021 - ACR5	0.278 (0.253, 0.304)	337/1211
Kuru B et al, 2021 - ACR5	0.145 (0.125, 0.166)	166/1143
Sparano C et al, 2021 - ACR5	0.344 (0.332, 0.356)	2227/6474
Orhan Soylemez UP et al, 2022 - ACR5	0.074 (0.058, 0.090)	75/1010
Subgroup ACR 5 (I^2=99.78 % , P=0.000)	**0.242 (0.136, 0.349)**	**6678/41332**

B

Studies	Estimate (95% C.I.)	Ev/Trt
Xu T et al, 2018 - EU2	0.025 (0.019, 0.031)	62/2465
Qi Q et al, 2021 - EU2	0.036 (0.025, 0.048)	40/1096
Hoang JK et al, 2021 - EU2	0.019 (0.017, 0.021)	530/27933
Seifert P et al, 2021 - EU2	0.035 (0.024, 0.045)	42/1211
Kuru B et al, 2021 - EU2	0.017 (0.009, 0.024)	19/1143
Sparano C et al, 2021 - EU2	0.009 (0.007, 0.011)	57/6474
Orhan Soylemez UP et al, 2022 - EU2	0.073 (0.057, 0.089)	74/1010
Subgroup EU 2 (I^2=95.47 % , P=0.000)	**0.028 (0.020, 0.035)**	**824/41332**
Xu T et al, 2018 - EU3	0.246 (0.229, 0.263)	607/2465
Qi Q et al, 2021 - EU3	0.193 (0.170, 0.217)	212/1096
Hoang JK et al, 2021 - EU3	0.363 (0.357, 0.368)	10126/27933
Seifert P et al, 2021 - EU3	0.320 (0.293, 0.346)	387/1211
Kuru B et al, 2021 - EU3	0.495 (0.466, 0.524)	566/1143
Sparano C et al, 2021 - EU3	0.297 (0.286, 0.308)	1921/6474
Orhan Soylemez UP et al, 2022 - EU3	0.548 (0.517, 0.578)	553/1010
Subgroup EU 3 (I^2=99.11 % , P=0.000)	**0.351 (0.290, 0.411)**	**14372/41332**
Xu T et al, 2018 - EU4	0.267 (0.250, 0.285)	659/2465
Qi Q et al, 2021 - EU4	0.160 (0.138, 0.181)	175/1096
Hoang JK et al, 2021 - EU4	0.429 (0.423, 0.435)	11990/27933
Seifert P et al, 2021 - EU4	0.245 (0.221, 0.269)	297/1211
Kuru B et al, 2021 - EU4	0.251 (0.226, 0.276)	287/1143
Sparano C et al, 2021 - EU4	0.452 (0.440, 0.464)	2928/6474
Orhan Soylemez UP et al, 2022 - EU4	0.264 (0.237, 0.292)	267/1010
Subgroup EU 4 (I^2=99.51 % , P=0.000)	**0.296 (0.214, 0.378)**	**16603/41332**
Xu T et al, 2018 - EU5	0.461 (0.442, 0.481)	1137/2465
Qi Q et al, 2021 - EU5	0.610 (0.582, 0.639)	669/1096
Hoang JK et al, 2021 - EU5	0.189 (0.185, 0.194)	5287/27933
Seifert P et al, 2021 - EU5	0.406 (0.379, 0.434)	492/1211
Kuru B et al, 2021 - EU5	0.237 (0.212, 0.262)	271/1143
Sparano C et al, 2021 - EU5	0.242 (0.232, 0.253)	1568/6474
Orhan Soylemez UP et al, 2022 - EU5	0.115 (0.095, 0.135)	116/1010
Subgroup EU 5 (I^2=99.66 % , P=0.000)	**0.323 (0.231, 0.414)**	**9540/41332**

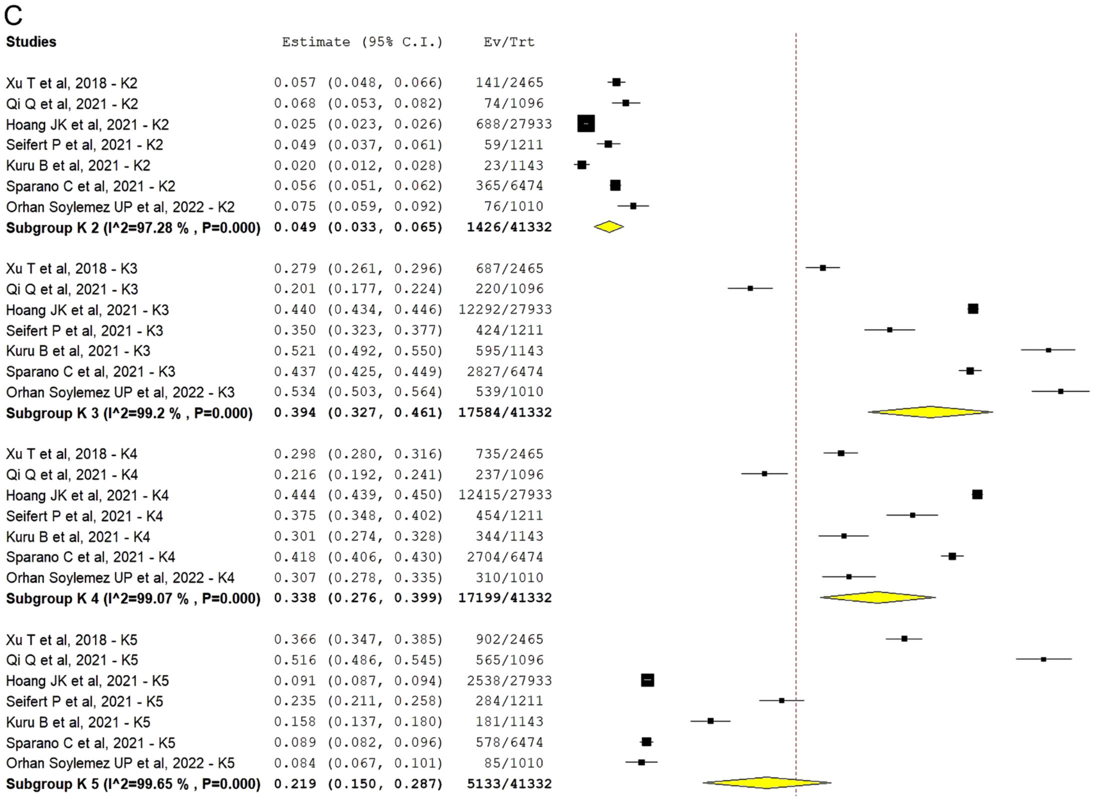


Figure 2

(A) Assessment of TNs according to ACR-TIRADS original category. (B) Assessment of TNs according to EU-TIRADS original category. (C) Assessment of TNs according to K-TIRADS original category. Any square represents a study and its size varies with study effect, while the line represents the 95% CI. Diamond indicates the pooled call rate and its width represents the 95% CI.

In ACR-TIRADS, the moderate class was the most prevalent without a 95% CI overlap with other classes (Fig. 4A). In EU-TIRADS, the three aggregated categories had similar prevalence (Fig. 4B). In K-TIRADS, the mild-, moderate-, and severe-risk categories showed no 95% CI overlap (Fig. 4C). Significant differences were observed when the three TIRADSs were compared (Table 3). A direct comparison of the three systems was conducted based on the three risk assessment categories. Similarly, in the low-risk categories, EU-TIRADS had the highest prevalence, ACR-TIRADS was intermediate, and K-TIRADS was the lowest (Fig. 5A). ACR-TIRADS was the most prevalent among moderate-risk classes (Fig. 5B). K-TIRADS identified the most prevalent severe-risk class (Fig. 5C). Figure 6 graphically illustrates the data from Fig. 5A, B, and C. TNs are usually classified as mild risk by EU-TIRADS, more so than by the other two systems. Conversely, ACR-TIRADS often categorizes them as moderate risk, and K-TIRADS as severe risk.

Exploration of heterogeneity

As previously mentioned, numerous dichotomous/continuous covariates were identified to investigate the heterogeneity of each individual analysis. In total, 180 meta-regression/subgroup analyses were conducted. In general, the heterogeneity in each risk class was resolved by at least one continuous covariate among those considered. The covariates explaining heterogeneity in the most classes include malignancy rate, study period, and population age. The intermediate TIRADS category was defined by the greatest number of covariates. The findings remained significantly consistent when investigating the heterogeneity of the second endpoint's results. Table 4 presents the primary findings of this comprehensive exploration of heterogeneity. All relevant figures and data for each significant meta-regression/subgroup analysis are detailed in the supplemental material.

Discussion

The introduction of US in thyroid disease was significant and now serves as the cornerstone of patient management in this field. Furthermore, the impact of RSSs/TIRADSs has been significant in our daily clinical practice. The primary purpose of introducing RSSs/TIRADSs was to standardize the procedure across different medical specialties, starting with the

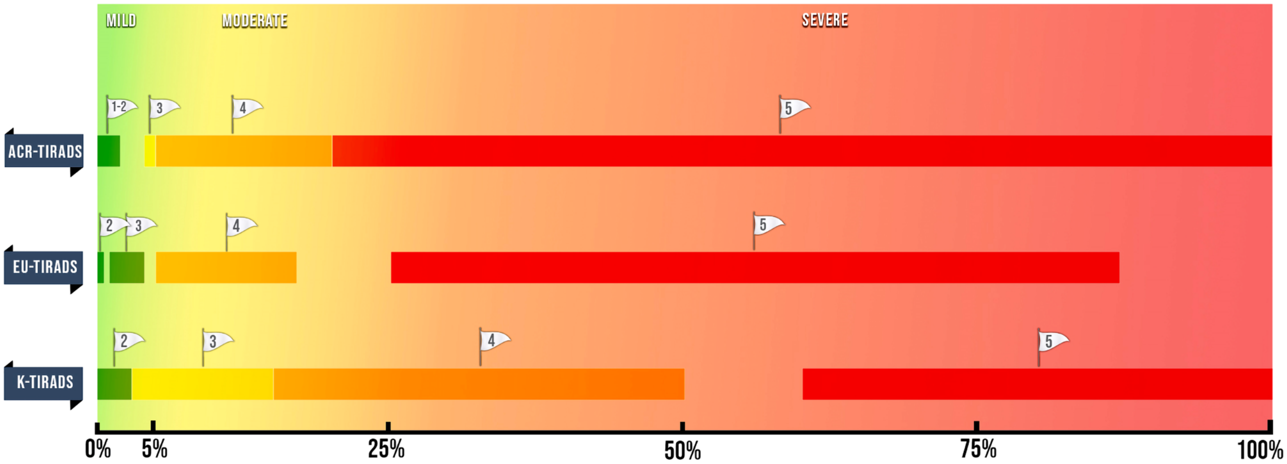
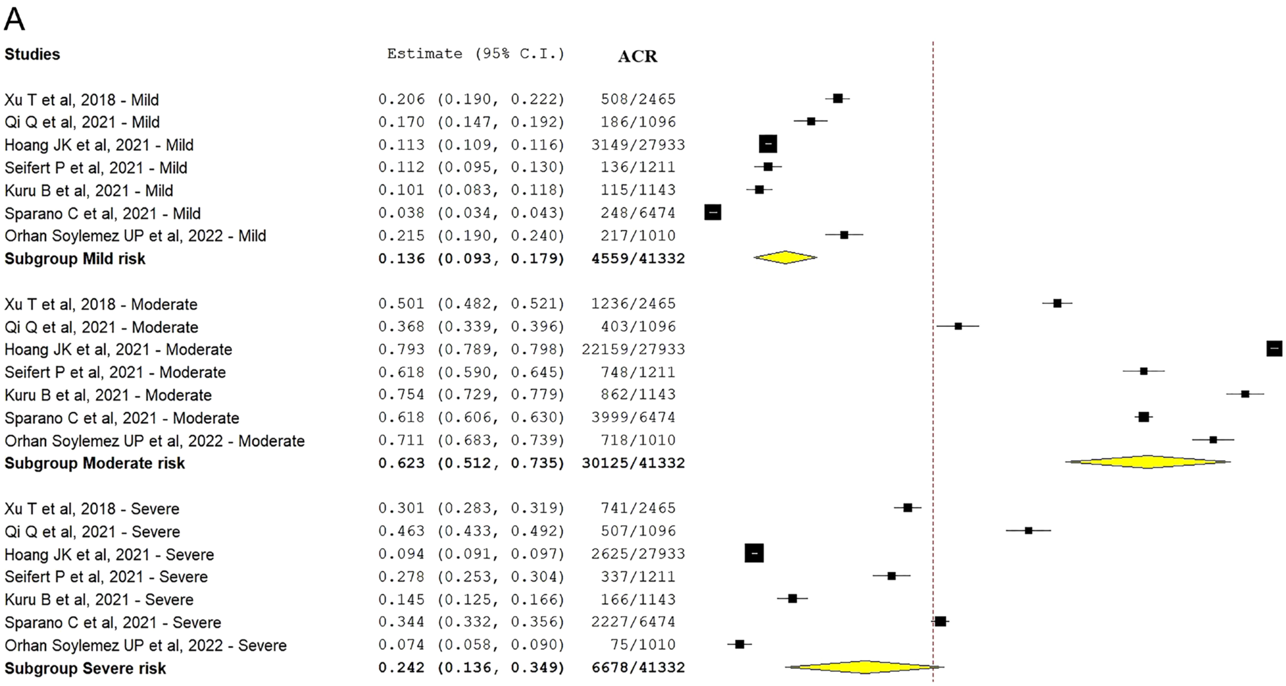


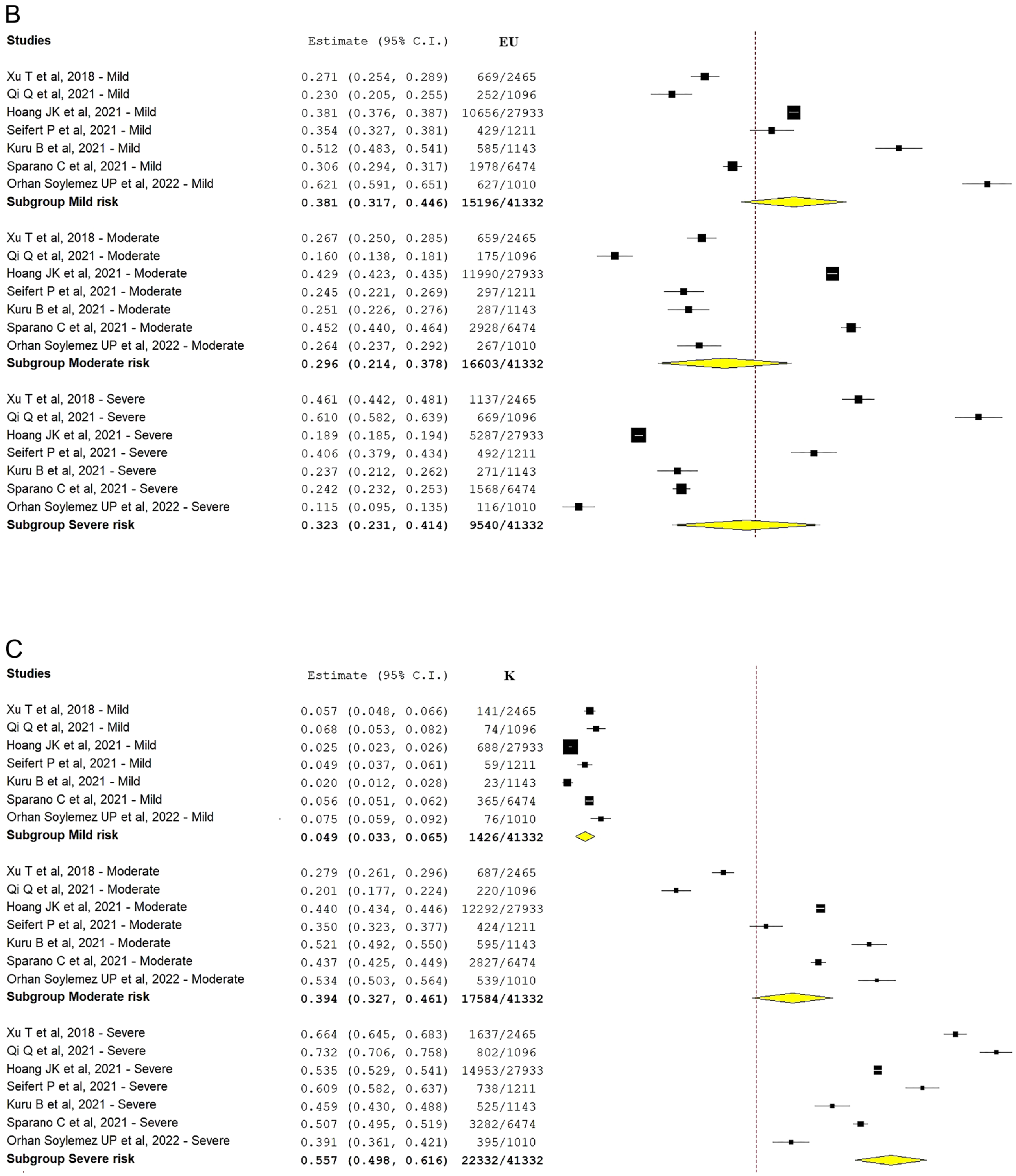
Figure 3

Graphical illustration of the risk assigned to categories of ACR-, EU-, and K-TIRADS leading to aggregate categories and building a three-scale system of mild, moderate, and severe risk of malignancy. The numerical heading of original categories is indicated by triangle flag. The colored bands indicate the malignancy risk range originally assigned to each category of TIRADS. *X*-axis represents the estimated risk of malignancy from 0 to 100%. The background schematically illustrates the increasing risk from green (mild) to moderate (yellow) and severe (red).

appropriate terminology. Evidence-based research has proven that these systems accurately detect cancer, resulting in their swift global adoption (6). Despite significant achievements, the current challenge is to construct a universal TIRADS that can resolve the differences among the existing TIRADSs (23). Although TIRADSs appear to have a similar structure (i.e. 4 and 5 categories with an escalating risk of malignancy) and diagnostic performance (8, 9), their foundational differences, which remain largely unexplored, distinguish them. The aim of this study was to evaluate the distribution of risk assessment scores by the three major TIRADSs in a population referred for US, but not for FNA or preoperative assessment, through a head-to-head comparison.

First, our search algorithm identified approximately 1000 articles. After the selection process, we included





B

Studies	Estimate (95% C.I.)	EU
Xu T et al, 2018 - Mild	0.271 (0.254, 0.289)	669/2465
Qi Q et al, 2021 - Mild	0.230 (0.205, 0.255)	252/1096
Hoang JK et al, 2021 - Mild	0.381 (0.376, 0.387)	10656/27933
Seifert P et al, 2021 - Mild	0.354 (0.327, 0.381)	429/1211
Kuru B et al, 2021 - Mild	0.512 (0.483, 0.541)	585/1143
Sparano C et al, 2021 - Mild	0.306 (0.294, 0.317)	1978/6474
Orhan Soylemez UP et al, 2022 - Mild	0.621 (0.591, 0.651)	627/1010
Subgroup Mild risk	**0.381 (0.317, 0.446)**	**15196/41332**
Xu T et al, 2018 - Moderate	0.267 (0.250, 0.285)	659/2465
Qi Q et al, 2021 - Moderate	0.160 (0.138, 0.181)	175/1096
Hoang JK et al, 2021 - Moderate	0.429 (0.423, 0.435)	11990/27933
Seifert P et al, 2021 - Moderate	0.245 (0.221, 0.269)	297/1211
Kuru B et al, 2021 - Moderate	0.251 (0.226, 0.276)	287/1143
Sparano C et al, 2021 - Moderate	0.452 (0.440, 0.464)	2928/6474
Orhan Soylemez UP et al, 2022 - Moderate	0.264 (0.237, 0.292)	267/1010
Subgroup Moderate risk	**0.296 (0.214, 0.378)**	**16603/41332**
Xu T et al, 2018 - Severe	0.461 (0.442, 0.481)	1137/2465
Qi Q et al, 2021 - Severe	0.610 (0.582, 0.639)	669/1096
Hoang JK et al, 2021 - Severe	0.189 (0.185, 0.194)	5287/27933
Seifert P et al, 2021 - Severe	0.406 (0.379, 0.434)	492/1211
Kuru B et al, 2021 - Severe	0.237 (0.212, 0.262)	271/1143
Sparano C et al, 2021 - Severe	0.242 (0.232, 0.253)	1568/6474
Orhan Soylemez UP et al, 2022 - Severe	0.115 (0.095, 0.135)	116/1010
Subgroup Severe risk	**0.323 (0.231, 0.414)**	**9540/41332**

C

Studies	Estimate (95% C.I.)	K
Xu T et al, 2018 - Mild	0.057 (0.048, 0.066)	141/2465
Qi Q et al, 2021 - Mild	0.068 (0.053, 0.082)	74/1096
Hoang JK et al, 2021 - Mild	0.025 (0.023, 0.026)	688/27933
Seifert P et al, 2021 - Mild	0.049 (0.037, 0.061)	59/1211
Kuru B et al, 2021 - Mild	0.020 (0.012, 0.028)	23/1143
Sparano C et al, 2021 - Mild	0.056 (0.051, 0.062)	365/6474
Orhan Soylemez UP et al, 2022 - Mild	0.075 (0.059, 0.092)	76/1010
Subgroup Mild risk	**0.049 (0.033, 0.065)**	**1426/41332**
Xu T et al, 2018 - Moderate	0.279 (0.261, 0.296)	687/2465
Qi Q et al, 2021 - Moderate	0.201 (0.177, 0.224)	220/1096
Hoang JK et al, 2021 - Moderate	0.440 (0.434, 0.446)	12292/27933
Seifert P et al, 2021 - Moderate	0.350 (0.323, 0.377)	424/1211
Kuru B et al, 2021 - Moderate	0.521 (0.492, 0.550)	595/1143
Sparano C et al, 2021 - Moderate	0.437 (0.425, 0.449)	2827/6474
Orhan Soylemez UP et al, 2022 - Moderate	0.534 (0.503, 0.564)	539/1010
Subgroup Moderate risk	**0.394 (0.327, 0.461)**	**17584/41332**
Xu T et al, 2018 - Severe	0.664 (0.645, 0.683)	1637/2465
Qi Q et al, 2021 - Severe	0.732 (0.706, 0.758)	802/1096
Hoang JK et al, 2021 - Severe	0.535 (0.529, 0.541)	14953/27933
Seifert P et al, 2021 - Severe	0.609 (0.582, 0.637)	738/1211
Kuru B et al, 2021 - Severe	0.459 (0.430, 0.488)	525/1143
Sparano C et al, 2021 - Severe	0.507 (0.495, 0.519)	3282/6474
Orhan Soylemez UP et al, 2022 - Severe	0.391 (0.361, 0.421)	395/1010
Subgroup Severe risk	**0.557 (0.498, 0.616)**	**22332/41332**

Figure 4

(A) Assessment of TNs according to the risk of malignancy estimated in ACR-TIRADS. (B) Assessment of TNs according to the risk of malignancy estimated in EU-TIRADS. (C) Assessment of TNs according to the risk of malignancy estimated in K-TIRADS. Any square represents a study and its size varies with study effect, while the line represents a 95% CI. Diamond indicates the pooled call rate and its wideness represents the 95% CI.

Table 3 Pooled results of TNs assessment according to ACR-, EU-, and K-TIRADS risk-aggregated categories.

	ACR-TIRADS		EU-TIRADS		K-TIRADS	
	% (95% CI)	I^2 (%)	% (95% CI)	I^2 (%)	% (95% CI)	I^2 (%)
Risk-aggregated category						
Mild	13.6 (9.3–17.9)[a]	99.38	38.1 (31.7–44.6)[a]	99.19%	4.9 (3.3–6.5)[b,a]	97.28
Moderate	62.3 (51.2–73.5)[b,a]	99.73	29.6 (21.4–37.8)	99.51%	39.4 (32.7–46.1)[b]	99.2
Severe	24.2 (13.6–34.9)	99.78	32.3 (23.1–41.4)	99.66%	55.7 (49.8–61.6)[b,a]	98.92

[a]Significantly different with respect to the same categories of the other TIRADSs; [b]Significantly different with respect to the other categories of its TIRADS. I^2, heterogeneity; TNs, thyroid nodules.

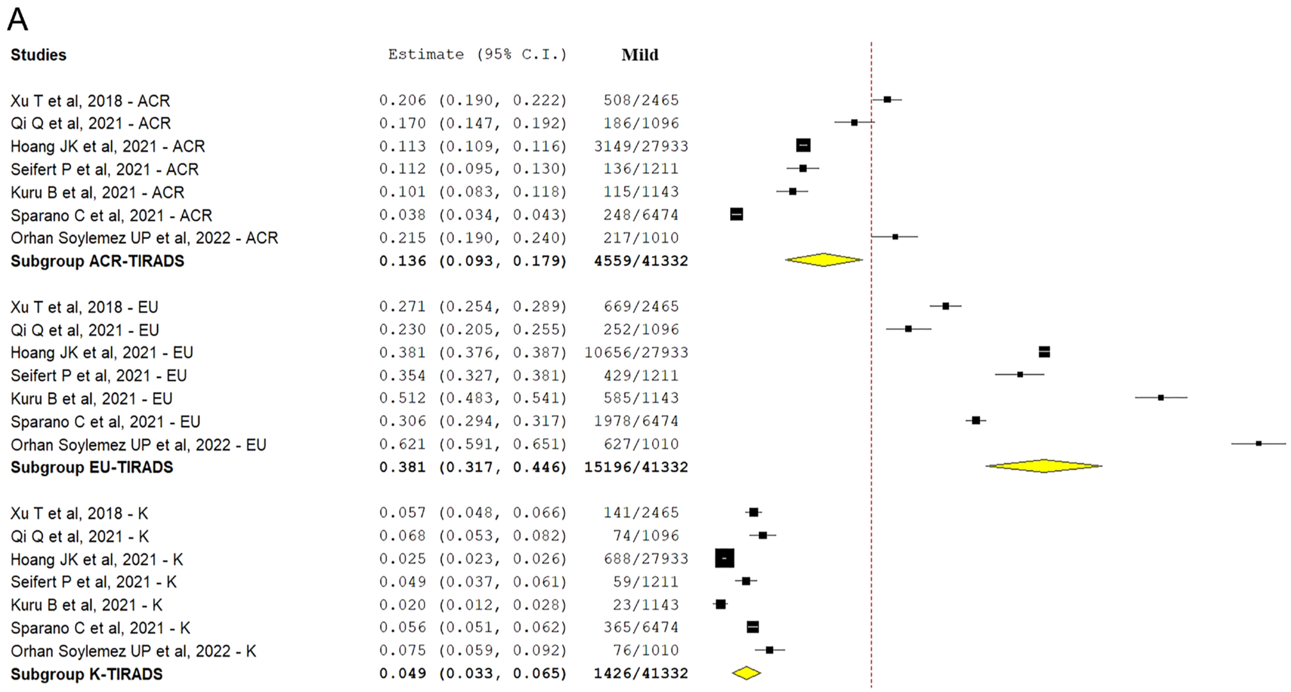
only seven recently published studies. This indicates that during these years, the authors primarily focused on evaluating and comparing the diagnostic performance of TIRADS and the unnecessary FNAC rate in selected populations. Despite this, we have pooled over 40,000 nodules, enabling a comprehensive and detailed analysis for the study. These seven reports came from various departments (imaging, endocrinology, and surgery) and from three different continents (Europe, Asia, and America). Thus, the diagnostic profile of the centers varied and seems to be quite a representative sample of the location where patients are daily taken care of.

Secondly, the initial comprehensive view of TN distribution across categories proposed by the three TIRADSs suggests that: (1) The American system typically assigns nodules to the intermediate-risk category with a 40% call rate (Fig. 2A); (2) The European system demonstrates a similar distribution across categories 3, 4, and 5 (Fig. 2B); (3) The Korean system categorizes 40% of TNs as category 3 (Fig. 2C); and (4) all three systems exhibit the lowest call rate for the low/very-low-risk class (Fig. 2A, B, and C).

Thirdly, a comparative analysis of TIRADSs category call rates shows that ACR-TIRADS significantly exceeds EU- and K-TIRADS 2 in category 1–2 call rates, but is lower in category 3 call rates compared to K-TIRADS. Notably, there is no significant difference in the call rates for intermediate- and high-risk categories (i.e. 4 and 5) across all three systems (Table 2).

Fourthly, upon analyzing the risk-adjusted classes (Fig. 3), it was observed that TNs are often classified as severe risk by K-TIRADS, as moderate risk by ACR-TIRADS, and as mild risk more frequently by EU-TIRADS than the other two systems (Figs. 5A, B, C and 6).

The high-level evidence findings warrant a thorough discussion. First, the small number of studies investigating the evaluation of TNs according to TIRADSs, regardless



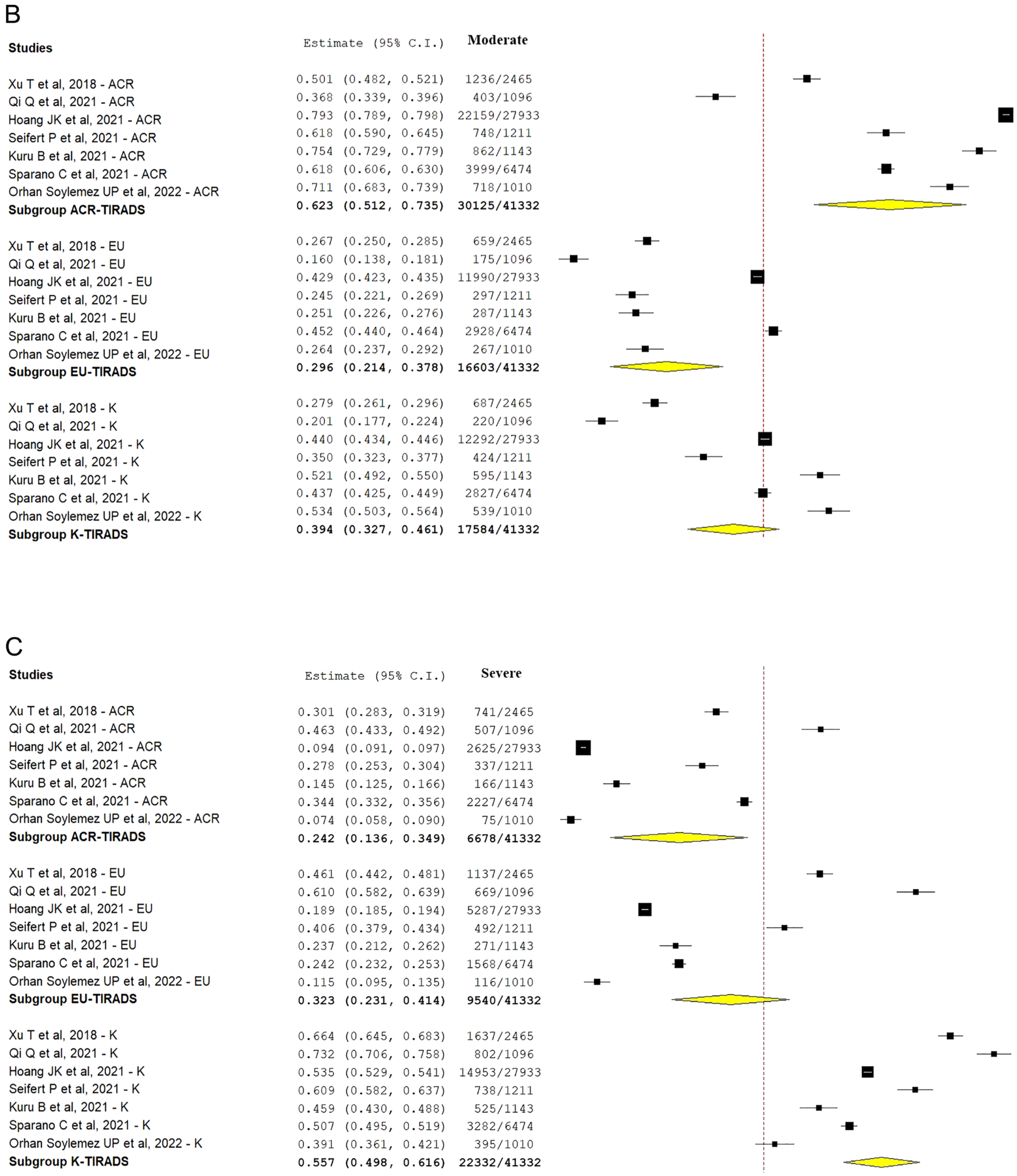


Figure 5

(A) Head-to-head comparison of TIRADSs according to TNs assessment in mild risk class. (B) Head-to-head comparison of TIRADSs according to TNs assessment in moderate risk class. (C) Head-to-head comparison of TIRADSs according to TNs assessment in severe-risk class. Any square represents a study and its size varies with study effect, while the line represents a 95% CI. Diamond indicates the pooled call rate and its wideness represents the 95% CI.

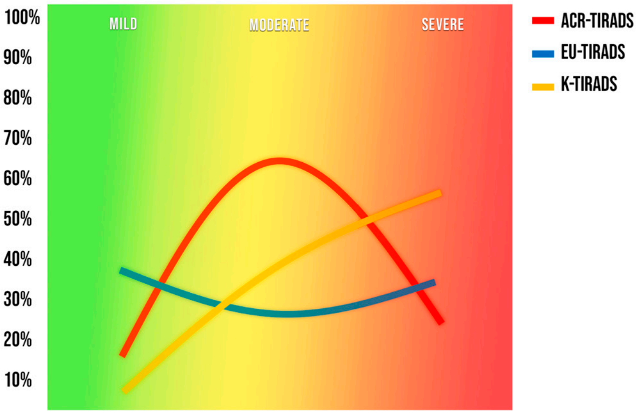


Figure 6

Graphical schematic representation of the meta-analysis of TIRADSs assessment according to risk-aggregated categories. This figure represents a sort of graphical abstract of the entire article. It illustrates schematically the different behavior exhibited by ACR-, EU-, and K-TIRADS when considering the risk of malignancy that they assign to each their category. According to malignancy risk estimation, K-TIRADS often assesses thyroid nodules as severe risk, ACR-TIRADS as moderate risk, and EU-TIRADS as mild risk) The details of these results are reported in Table 3 and Fig. 5.

of their accuracy, is an interesting topic. The latter is probably influenced by the need to assess the reliability of TIRADS shortly after its introduction. Regardless, TIRADS users need to understand how the RSS performs in assessing TNs, irrespective of their indication for FNAC/surgery. Secondly, Table 2 shows the pooled results of TNs grading according to ACR-, EU-, and K-TIRADS, revealing a significant issue that has not been addressed in previous studies. The distribution of nodules in categories 1–2 and 3 ranges from 35.5% to 44.3%. The distribution of category 5 nodules ranges from 21.9% to 32.3%. This result is noteworthy, as many studies indicate that the expected malignancy rate for unselected thyroid nodules is between 5% and 10% (1). Thus, the developed RSSs appear to overestimate this percentage by a factor of two to three. Furthermore, in at least half of the cases, 90% of benign nodules are not correctly classified. These facts are alarming because they consistently lead to patients undergoing unnecessary testing and follow-up. Thirdly, RSS users should be aware that the distribution of nodules in TIRADS risk classes is influenced not only by potential geographic differences (e.g. iodine sufficiency, genetic), but also by the classification systems themselves. This data may influence a clinical user's decision in favor of one TIRADS over the others. Furthermore, understanding this distribution can serve as a self-assessment tool for users. In other words, each healthcare setting can assess if there are significant deviations from the expected figures found in the selected literature. In this scenario, over-scoring or under-scoring may occur and require correction. Under-scoring can lead to missed carcinomas, while over-scoring may result in unnecessary examinations and heightened anxiety. Fourthly, it is reassuring for clinical users that no difference was found in the call rate for intermediate- and high-risk categories across the three TIRADSs. We are all quite prone to recommend these patients for surgery, especially those in category 5. This data somewhat aligns with the literature on the accuracy of TIRADS in predicting malignancy (8, 9). Fifthly, the malignancy risk estimates across each RSSs TIRADS categories are not always comparable. Thus, when the risk-adjusted system is developed to align the risks, it becomes clear that TIRADSs evaluate TNs differently. K-TIRADS appears to increase the risk of TNs, ACR-TIRADS reduces the call rate for the two extremes, and EU-TIRADS divides TNs into three equal parts and has the highest rate in the mild-risk category. This represents essential new information in this field that requires comprehensive discussion. Different TIRADSs, such as ACR-, EU-, and K-TIRADS, have been developed and conceptualized in three distinct geographic contexts. Numerous factors can lead to the development of ‘aggressive’ systems, prompting clinicians to proceed with diagnostic work-up, including invasive tools (i.e. FNAC, large biopsy, surgery). Key factors to consider in this context include the health system, cost charges, national programs, patient expectations, anxiety about living with the disease, hospital accessibility, availability of diagnostic procedures, physician's inclination to investigate further, and doctor–patient communication, among others. Indeed, it has been previously reported that the rate of unnecessary FNAC varies among different RSSs/TIRADSs (10, 24). These figures suggest significant differences among TIRADSs, primarily due to specific ambiguous ultrasound pattern definition such as hypoechogenicity (25). These issues need to be resolved before establishing a globally accepted TIRADS, endorsed by leading societies focused on TNs (I-TIRADS) (23).

Significant heterogeneity was consistently observed in pooled analyses. The heterogeneity in the intermediate-risk category of TIRADSs may be due to a large number of covariates. The system's intermediate class is recognized as the weakest (2, 26). Conversely, the recorded large heterogeneity confirms that the assessment of TNs according to TIRADSs can significantly vary based on numerous characteristics.

A discussion of practical conclusions is necessary to assist readers and TIRADS users to be mindful of the present findings, while waiting for the introduction of I-TIRADS. If we are working in a context with a high incidence of thyroid cancers or facing patients with significant anxiety about living with the disease, we should consider using the K-TIRADS conceptualization. This system is more likely to recommend further diagnostic work-up and is associated with a higher risk estimation (and, presumably, higher diagnostic sensitivity). If our healthcare system is disciplined by laws that encourage cost-saving measures, or if we operate in an iodine-deficient region with a high incidence of benign goiter,

Table 4 Summary of findings of exploration of heterogeneity.

Risk of malignancy[a]/ feature	General explanation	Figure number in supplemental material	TIRADS category	
			Description	Category
Mild				
Malignancy rate	The higher the cancer rate in the study series, the lower the proportion of calls for this class.[b]	1, 2, 10, 11, 12, 38	Low-risk/ suspicion	ACR-TIRADS 13, EU-TIRADS 2–3, and K-TIRADS 2–3
Study period	The longer the study period, the lower the proportion of calls for this class.[c]	3, 4, 5		
Moderate				
Malignancy rate	The higher the cancer rate in the study series, the lower the proportion of calls for this class.	16, 17, 18, 39	Intermediate-risk/ suspicion	ACR-TIRADS 4, EU-TIRADS 4, and K-TIRADS 4
Study period	The longer the study period, the higher the proportion of calls for this class.	19, 20, 21, 40		
Age	The higher the age of the population enrolled, the higher the call for this class.	22, 23, 24, 41		
Sample size	The higher the size, the higher the proportion of calls for this class.	26, 27, 28		
Severe				
Malignancy rate	The higher the cancer rate in the study series, the higher the proportion of calls for this class.	30, 31, 32, 42	High-risk/ suspicion	ACR-TIRADS 5, EU-TIRADS 5, and K-TIRADS 5
Age	The higher the age of the population enrolled, the lower the call for this class.[d]	33, 34		
Female-to-male ratio	The higher the ratio in the population enrolled, the lower the call for this class.[e]	35, 36		

[a]These categories are defined in the text; [b]Not significant for ACR-TIRADS 1–2; [c]Not significant for ACR-, EU-, and K-TIRADS 3; [d]Not significant for EU-TIRADS 5; [e]Not significant for K-TIRADS 5.

we should consider using ACR-TIRADS or EU-TIRADS to classify nodules. ACR-TIRADS tends to classify nodules as intermediate risk more often than other systems, while EU-TIRADS more frequently assess TNs as low-risk. For all the above reasons, we are not able to conclude whether any of the three TIRADSs can be universalized and adapted to all geographical contexts. In addition, the specific setting of the institution (e.g. oncological center, internal medicine department, radiology division, endocrinology service, private office) to which patients are referred may affect the pre-test risk, resulting in a different risk assessment result. Each institution and each user should adopt the system that is better suited to its specific setting. In any case, all these issues themselves confirm the need to achieve a generalizable I-TIRADS. Finally, uniform classification aspects are also essential for the implementation of artificial intelligence algorithms in routine thyroid ultrasound (27).

Strengths and potential limitations of the paper should be addressed. Occasionally, the research aim was somewhat unclear during the study selection. We initially set very strict and rigorous study selection criteria, leading to the exclusion of most papers. The seven included papers were performed in tertiary centers with varying study periods. This may constitute a selection bias because patients assessed at risk are generally referred to tertiary centers. Also, data on the setting of institutions could not be fully analyzed, but can at least be considered as fairly representative of daily practice. This aspect cannot be fully explored and, in any case, patients may have been referred to tertiary centers regardless of the complexity of their clinical condition. They presented data from a retrospective analysis, thus minimizing the risk of inter-observer variability. The three continents had equal representation. Furthermore, the major strength was the ability to directly compare the three systems, ensuring data consistency.

Conclusion

To our knowledge, this is the first systematic review and meta-analysis aimed at analyzing and comparing the distribution of thyroid nodules across the categories of different TIRADS, regardless of the need for further diagnostic or therapeutic work-up. The results suggest that ACR-, EU-, and K-TIRADS exhibit different behaviors. Clinicians and researchers need to be fully aware of these findings in order to self-assess and understand the underlying assumptions of the system they are using. These figures have to be seen as essential prerequisites for developing the I-TIRADS.

Supplementary materials
This is linked to the online version of the paper at https://doi.org/10.1530/ETJ-23-0242.

Declaration of interest
The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the study reported.

Funding
This work did not receive any specific grant from any funding agency in the public, commercial, or not-for-profit sector.

References

1 Russ G, Leboulleux S, Leenhardt L & Hegedüs L. Thyroid incidentalomas: epidemiology, risk stratification with ultrasound and workup. *European Thyroid Journal* 2014 **3** 154–163. (https://doi.org/10.1159/000365289)

2 Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, Pacini F, Randolph GW, Sawka AM, Schlumberger M, *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. (https://doi.org/10.1089/thy.2015.0020)

3 Gharib H, Papini E, Paschke R, Duick DS, Valcavi R, Hegedüs L, Vitti P, AACE/AME/ETA Task Force on Thyroid Nodules & American Association of Clinical Endocrinologists. American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and European Thyroid Association Medical guidelines for clinical practice for the diagnosis and management of thyroid nodules: executive summary of recommendations. *Endocrine Practice* 2010 **16** 468–475. (https://doi.org/10.4158/EP.16.3.468)

4 Durante C, Hegedüs L, Czarniecka A, Paschke R, Russ G, Schmitt F, Soares P, Solymosi T & Papini E. 2023 European Thyroid Association Clinical Practice Guidelines for thyroid nodule management. *European Thyroid Journal* 2023 **12** e230067. (https://doi.org/10.1530/ETJ-23-0067)

5 Rago T & Vitti P. Risk stratification of thyroid nodules: from ultrasound features to TIRADS. *Cancers* 2022 **14** 717. (https://doi.org/10.3390/cancers14030717)

6 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* 2021 **13** 4316. (https://doi.org/10.3390/cancers13174316)

7 Trimboli P & Durante C. Ultrasound risk stratification systems for thyroid nodule: between lights and shadows, we are moving towards a new era. *Endocrine* 2020 **69** 1–4. (https://doi.org/10.1007/s12020-020-02196-6)

8 Castellana M, Castellana C, Treglia G, Giorgino F, Giovanella L, Russ G & Trimboli P. Performance of five ultrasound risk stratification systems in selecting thyroid nodules for FNA. *Journal of Clinical Endocrinology and Metabolism* 2020 **105** dgz170. (https://doi.org/10.1210/clinem/dgz170)

9 Kim PH, Suh CH, Baek JH, Chung SR, Choi YJ & Lee JH. Diagnostic performance of four ultrasound risk stratification systems: a systematic review and meta-analysis. *Thyroid* 2020 **30** 1159–1168. (https://doi.org/10.1089/thy.2019.0812)

10 Kim PH, Suh CH, Baek JH, Chung SR, Choi YJ & Lee JH. Unnecessary thyroid nodule biopsy rates under four ultrasound risk stratification systems: a systematic review and meta-analysis. *European Radiology* 2021 **31** 2877–2885. (https://doi.org/10.1007/s00330-020-07384-6)

11 Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, Cronan JJ, Beland MD, Desser TS, Frates MC, *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* 2017 **14** 587–595. (https://doi.org/10.1016/j.jacr.2017.01.046)

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. *European Thyroid Journal* 2017 **6** 225–237. (https://doi.org/10.1159/000478927)

13 Shin JH, Baek JH, Chung J, Ha EJ, Kim JH, Lee YH, Lim HK, Moon WJ, Na DG, Park JS, *et al.* Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean Society of Thyroid Radiology Consensus Statement and Recommendations. *Korean Journal of Radiology* 2016 **17** 370–395. (https://doi.org/10.3348/kjr.2016.17.3.370)

14 Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA & Thacker SB. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000 **283** 2008–2012. (https://doi.org/10.1001/jama.283.15.2008)

15 National Heart, Lung, and Blood Institute. Study quality assessment tools. Available at: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools. (Accessed 18 June 2023).

16 Xu T, Wu Y, Wu RX, Zhang YZ, Gu JY, Ye XH, Tang W, Xu SH, Liu C & Wu XH. Validation and comparison of three newly-released thyroid Imaging Reporting and Data Systems for cancer risk determination. *Endocrine* 2019 **64** 299–307. (https://doi.org/10.1007/s12020-018-1817-8)

17 Qi Q, Zhou A, Guo S, Huang X, Chen S, Li Y & Xu P. 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. *Frontiers in Endocrinology* 2021 **12** 763897. (https://doi.org/10.3389/fendo.2021.763897)

18 Hoang JK, Middleton WD, Langer JE, Schmidt K, Gillis LB, Nair SS, Watts JA, Snyder RW 3rd, Khot R, Rawal U, *et al.* Comparison of thyroid risk categorization systems and fine-needle aspiration recommendations in a multi-institutional thyroid ultrasound registry. *Journal of the American College of Radiology* 2021 **18** 1605–1613. (https://doi.org/10.1016/j.jacr.2021.07.019)

19 Seifert P, Schenke S, Zimny M, Stahl A, Grunert M, Klemenz B, Freesmeyer M, Kreissl MC, Herrmann K & Görges R. Diagnostic performance of Kwak, EU, ACR, and Korean TIRADS as well as ATA guidelines for the ultrasound risk stratification of non-autonomously functioning thyroid nodules in a region with long history of iodine deficiency: a German multicenter trial. *Cancers* 2021 **13** 4467. (https://doi.org/10.3390/cancers13174467)

20 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. *Endocrine Practice* 2021 **27** 1093–1099. (https://doi.org/10.1016/j.eprac.2021.04.411)

21 Sparano C, Verdiani V, Pupilli C, Perigli G, Badii B, Vezzosi V, Mannucci E, Maggi M & Petrone L. Choosing the best algorithm among five thyroid nodule ultrasound scores: from performance to cytology sparing-a single-center retrospective study in a large cohort. *European Radiology* 2021 **31** 5689–5698. (https://doi.org/10.1007/s00330-021-07703-5)

22 Orhan Soylemez UP & Gunduz N. Diagnostic accuracy of five different classification systems for thyroid nodules: a prospective, comparative study. *Journal of Ultrasound in Medicine* 2022 **41** 1125–1136. (https://doi.org/10.1002/jum.15802)

23 Durante C, Hegedüs L, Na DG, Papini E, Sipos JA, Baek JH, Frasoldati A, Grani G, Grant E, Horvath E, *et al.* International expert consensus on US lexicon for thyroid nodules. *Radiology* 2023 **309** e231481. (https://doi.org/10.1148/radiol.231481)

24 Ruan JL, Yang HY, Liu RB, Liang M, Han P, Xu XL & Luo BM. Fine needle aspiration biopsy indications for thyroid nodules: compare a point-based risk stratification system with a pattern-based risk stratification system. *European Radiology* 2019 **29** 4871–4878. (https://doi.org/10.1007/s00330-018-5992-z)

25 Popova NM, Radzina M, Prieditis P, Liepa M, Rauda M & Stepanovs K. Impact of the hypoechogenicity criteria on thyroid nodule malignancy risk stratification performance by different TIRADS systems. *Cancers* 2021 **13** 5581. (https://doi.org/10.3390/cancers13215581)

26 Cibas ES & Ali SZ. The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid* 2017 **27** 1341–1346. (https://doi.org/10.1089/thy.2017.0500)

27 Sorrenti S, Dolcetti V, Radzina M, Bellini MI, Frezza F, Munir K, Grani G, Durante C, D'Andrea V, David E, *et al.* Artificial intelligence for thyroid nodule characterization: where are we standing? *Cancers* 2022 **14** 3357. (https://doi.org/10.3390/cancers14143357)