


Systematic Review

Diagnostic Performance of Various Ultrasound Risk Stratification Systems for Benign and Malignant Thyroid Nodules: A Meta-Analysis

Ji-Sun Kim ¹, Byung Guk Kim ¹, Gulnaz Stybayeva ² and Se Hwan Hwang ^{3,*}

¹ Department of Otolaryngology-Head and Neck Surgery, Eunpyeong St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul 06591, Republic of Korea

² Department of Physiology and Biomedical Engineering, Mayo Clinic, Rochester, MN 55902, USA

³ Department of Otolaryngology-Head and Neck Surgery, Bucheon St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul 06591, Republic of Korea

* Correspondence: yellobird@catholic.ac.kr; Tel.: +82-32-340-7044

Simple Summary: In the present study, the sensitivity, specificity, and pooled diagnostic performances according to the cutoff value for diagnosing cancer of five ultrasound risk-stratification systems often used in clinical practice were verified by performing a meta-analysis. Sixty-seven studies involving 76,512 thyroid nodules were included in this research. The highest area under the curve (AUCs) of the K-TIRADS, ACR-TIRADS, ATA classification, EU-TIRADS, and Kwak-TIRADS were 0.904, 0.882, 0.859, 0.843, and 0.929, respectively. Based on the optimal sensitivity and specificity, the AUC or diagnostic odds ratios of K-TIRADS, ACR-TIRADS, ATA, EU-TIRADS, and Kwak-TIRADS were taken as the cutoff values of 4 (intermediate suspicion), TR5 (highly suspicious), high suspicion, 5 (high risk), and 4b, respectively. All ultrasound-based risk-stratification systems had good diagnostic performance.

Abstract: Background: To evaluate the diagnostic performance of ultrasound risk-stratification systems for the discrimination of benign and malignant thyroid nodules and to determine the optimal cutoff values of individual risk-stratification systems. Methods: PubMed, Embase, SCOPUS, Web of Science, and Cochrane library databases were searched up to August 2022. Sensitivity and specificity data were collected along with the characteristics of each study related to ultrasound risk stratification systems. Results: Sixty-seven studies involving 76,512 thyroid nodules were included in this research. The sensitivity, specificity, diagnostic odds ratios, and area under the curves by K-TIRADS (4), ACR-TIRADS (TR5), ATA (high suspicion), EU-TIRADS (5), and Kwak-TIRADS (4b) for malignancy risk stratification of thyroid nodules were 92.5%, 63.5%, 69.8%, 70.6%, and 95.8%, respectively; 62.8%, 89.6%, 87.2%, 83.9%, and 63.8%, respectively; 20.7111, 16.8442, 15.7398, 12.2986, and 38.0578, respectively; and 0.792, 0.882, 0.859, 0.843, and 0.929, respectively. Conclusion: All ultrasound-based risk-stratification systems had good diagnostic performance. Although this study determined the best cutoff values in individual risk-stratification systems based on statistical assessment, clinicians could adjust or alter cutoff values based on the clinical purpose of the ultrasound and the reciprocal changes in sensitivity and specificity.

Keywords: thyroid cancer; thyroid nodules; ultrasonography; meta-analysis; diagnostic imaging



Citation: Kim, J.-S.; Kim, B.G.; Stybayeva, G.; Hwang, S.H.

Diagnostic Performance of Various Ultrasound Risk Stratification Systems for Benign and Malignant Thyroid Nodules: A Meta-Analysis. *Cancers* **2023**, *15*, 424. <https://doi.org/10.3390/cancers15020424>

Academic Editor: Jan Egger

Received: 14 November 2022

Revised: 31 December 2022

Accepted: 4 January 2023

Published: 9 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The thyroid gland is an organ that can be easily inspected by using ultrasound (US). US is an accurate test that can confirm the characteristics of the thyroid and is a highly accessible diagnostic method that can be performed relatively easily in an outpatient setting [1]. Thyroid US is a primary imaging test for the evaluation of thyroid nodules, and the evidence of thyroid cancer has been confirmed through imaging features of thyroid

nodules [2]. The popular use of US has increased the diagnostic rate for thyroid nodules [3]. However, this does not mean that the incidence of thyroid cancer or the need for treatment has increased. There was a report that the more US was performed, the more thyroid cancer was diagnosed [4]. Concerns have been raised about unnecessary biopsies and additional tests for benign thyroid nodules or thyroid cancer with infrequent progression. The low mortality and high diagnostic rates have given rise to a discussion of overdiagnosis.

US-based risk stratification systems (RSSs) have been proposed by several international societies to prevent the overdiagnosis of thyroid nodules and to help determine additional tests and follow-up. RSS is being applied in clinical practice as a method of classifying and scoring characteristic findings of thyroid nodules. Even after the first meta-analysis was performed in 2019 [5], many studies reported the diagnostic accuracy of each RSS. In the present study, the sensitivity, specificity, and pooled diagnostic performances according to the cutoff value for diagnosing cancer of five RSSs often used in clinical practice were verified by performing a meta-analysis. In addition, the clinical implications of the diagnostic accuracy were reviewed.

2. Materials and Methods

2.1. Study Protocol and Literature Search Strategy

This meta-analysis was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines [6]. The study protocol was prospectively registered on the Open Science Framework (<https://osf.io/7cu2y/> (accessed on 27 September 2022)). Clinical studies were retrieved from PubMed, Embase, SCOPUS, Web of Science, and the Cochrane Central Register of Controlled Trials from the start date to August 2022. The search terms were as follows: thyroid, thyroid nodule, thyroid neoplasm, malignancy, thyroid cancer; diagnostic imaging, diagnostic performance diagnostic value, ultrasonography, diagnosis, ultrasound, diagnostic value, ultrasonography, ultrasound classifications, ultrasound risk stratification system, imaging, reporting systems, thyroid imaging reporting and data system (TI-RADS), TI-RADS, TIRADS, Indeterminate, Korean Society of Thyroid Radiology and Korean Thyroid Association guideline (K-TIRADS), American College of Radiology guideline (ACR-TIRADS), American Thyroid Association (ATA) guidelines, European Thyroid Association guideline (EU-TIRADS), and Kwak-TIRADS (Supplementary Table S1). Two independent reviewers removed studies that were not related to the diagnosis or prediction of thyroid malignancy using US classifications by assessments of article titles, abstracts, and full texts.

2.2. Selection Criteria

The inclusion criteria were as follows: articles about patients undergoing US of thyroid nodules, and comparison of US findings with cytologic or histologic findings. Exclusion criteria included review articles, case reports, articles about other neck diseases (e.g., lymphadenitis or neck mass), articles without adequate data to determine the diagnostic value of US, and those not written in English.

2.3. Data Extraction and Risk of Bias Assessment

Data from articles included in the study were extracted in a standardized format [7]. The results of the analysis were diagnostic odds ratio (DOR), summary receiver operating characteristic (SROC) curve, and area under the curve (AUC). DOR was calculated by using the parameters of true positive, true negative, false positive, and false negative. The DOR was assessed with a 95% confidence interval by using a random effects model. The SROC curve and AUC were used as methods to evaluate diagnostic data in meta-analysis. As the discriminant power of the test increases, the SROC curve gets closer to the upper left corner, the point where both sensitivity and specificity are 100% [8]. Higher AUC values range from zero to one, indicating better test performance. The AUC value indicated diagnostic accuracy [9].

The Quality Assessment of Diagnostic Accuracy Studies Version 2 tool (QADAS-2) was used to evaluate methodological quality (risk of bias) [10]. For the definition of true positive and negative, guideline category < cutoff value was regarded as “test negative” and guideline category \geq cutoff value as “test positive.” Therefore, “benign” lesions classified as < cutoff were regarded as true negative, and “non-benign” lesions classified as \geq cutoff value were regarded as true positive. Accordingly, the sensitivity, specificity, and DOR were calculated with reference to the results based on pathological examination, or fine-needle aspiration (FNA) cytology and follow-up. Receiver operating characteristic (ROC) curve analyses and areas under the ROC (AUC) were used to assess the value of guidelines in differentiating benign from malignant thyroid nodules.

2.4. Statistical Analysis and Outcome Measurements

R statistical software (R Foundation for Statistical Computing, Vienna, Austria) was used for this analysis. To assess heterogeneity, a homogeneity analysis was performed by using the Q statistic. According to the 2016 Korean Society of Thyroid Radiology and Korean Thyroid Association guidelines, thyroid nodules were assigned to be benign, of low suspicion (K-TIRADS 3), intermediate suspicion (K-TIRADS 4), and high suspicion (K-TIRADS 5) [2]. The US features in the ACR TI-RADS are categorized as benign (TR1, 0 point), not suspicious (TR2, 2 points), mildly suspicious (TR3, 3 points), moderately suspicious (TR4, 4–6 points), or highly suspicious (TR5, 7 points or more) for malignancy [11]. Based on the 2015 ATA guidelines, the thyroid nodules were classified according to the malignancy risk as “high”, “intermediate”, “low” or “very low” suspicion [12]. EU-TIRADS classified thyroid nodules as benign and low-, intermediate-, and high-risk nodules according to the malignancy risk [1]. The TI-RADS categories proposed by Kwak et al. classify thyroid nodules as 2 (benign lesions), 3 (no suspicious US features), 4a (one suspicious US feature), 4b (two suspicious US features), 4c (three or four suspicious US features), and 5 (five suspicious US features) according to the risk estimates of malignancy [13]. Diagnostic accuracy in individual risk stratification systems (K-TIRADS, ACR-TIRADS, ATA, EU-TIRADS, and Kwak-TIRADS) was assessed based on the use of different cutoff values. Potential publication bias was assessed by Begg’s funnel plot and Egger’s linear regression test.

3. Results

3.1. Search and Study Selection

After screening 3148 articles through an established process, a total of 746 articles were excluded after reviewing the relevance of titles and abstracts. A full-text review of the remaining 86 articles was performed, and 19 articles were excluded because they analyzed other interventions or lacked results. As a result, 67 studies with 76,512 thyroid nodules were included in the analysis (Figure 1) [14–78]. The characteristics of the study are presented in Supplementary Table S2 and the results of the bias assessment are presented in Supplementary Table S3. Egger’s test yielded a significant result ($p > 0.05$) except sensitivity of ACR TI-RADS TR3 (p -value = 0.009114), K-TIRADS 3 (p -value = 0.0001452), Kwak-TIRADS 4a (p -value = 0.002071), and Kwak-TIRADS 4b (p -value = 0.03186). However, all biased outcomes showed no significance between original and corrected (trim fill method) outcomes. Begg’s funnel plots for each RSS are presented in Supplementary Figures S1–S5.

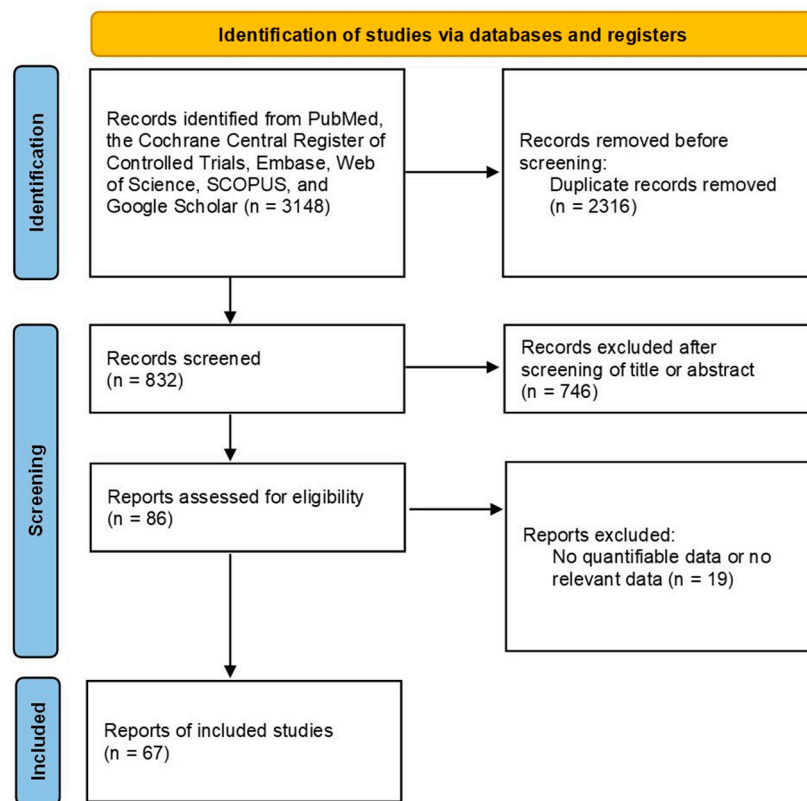


Figure 1. Flowchart of the study selection process for meta-analysis.

3.2. Diagnostic Accuracy in Various US Risk Stratification Systems

In K-TIRADS categories, sensitivity changed from around 66% to 99% (highest in low suspicion) and specificity showed an inverse association (89% to 8%; highest in high suspicion) according to the different cutoff values (categories) (Table 1). ROC analysis and DOR showed that the best diagnostic cutoff values of K-TIRADS were low and intermediate suspicion, respectively (Figures 2 and 3). Although a test with high AUC is statistically considered “better” than one with lower AUC, AUC lacks clinical interpretability because it does not reflect the practical gains and losses of individual patients by diagnostic tests. In addition, AUC can consider a test that increases sensitivity at low specificity superior to one that increases sensitivity at high specificity [79]. If the sensitivity and specificity on a screening test were considered to be too high or too low, they could be adjusted by raising or lowering cutoff values [80]. Based on the above mentioned, the best cutoff value of K-TIRADS was category intermediate suspicion (K-TIRADS 4) with the sensitivity of 92.5% and specificity of 62.8%.

Table 1. Diagnostic efficacy and the ROC curves of K-TIRADS categories.

	Sensitivity [95% CIs]	Specificity [95% CIs]	DOR [95% CIs]	AUC
High (K-TIRADS 5)	0.6644 [0.5488; 0.7632]; I ² = 99.1%	0.8904 [0.8495; 0.9212]; I ² = 98.8%	17.1881 [12.8739; 22.9479]; I ² = 94.7%	0.881
Intermediate (K-TIRADS 4)	0.9251 [0.8783; 0.9548]; I ² = 97.9%	0.6280 [0.5790; 0.6746]; I ² = 98.5%	20.7111 [15.0584; 28.4856]; I ² = 92.6%	0.792
Low (K-TIRADS 3)	0.9991 [0.9955; 0.9998]; I ² = 94.9%	0.0823 [0.0381; 0.1685]; I ² = 99.7%	17.2411 [9.7008; 30.6424]; I ² = 68.3%	0.904

ROC: receiver operating characteristic; CI: confidence interval; AUC: area under the curve; K-TIRADS: Korean Thyroid Imaging Reporting and Data System.

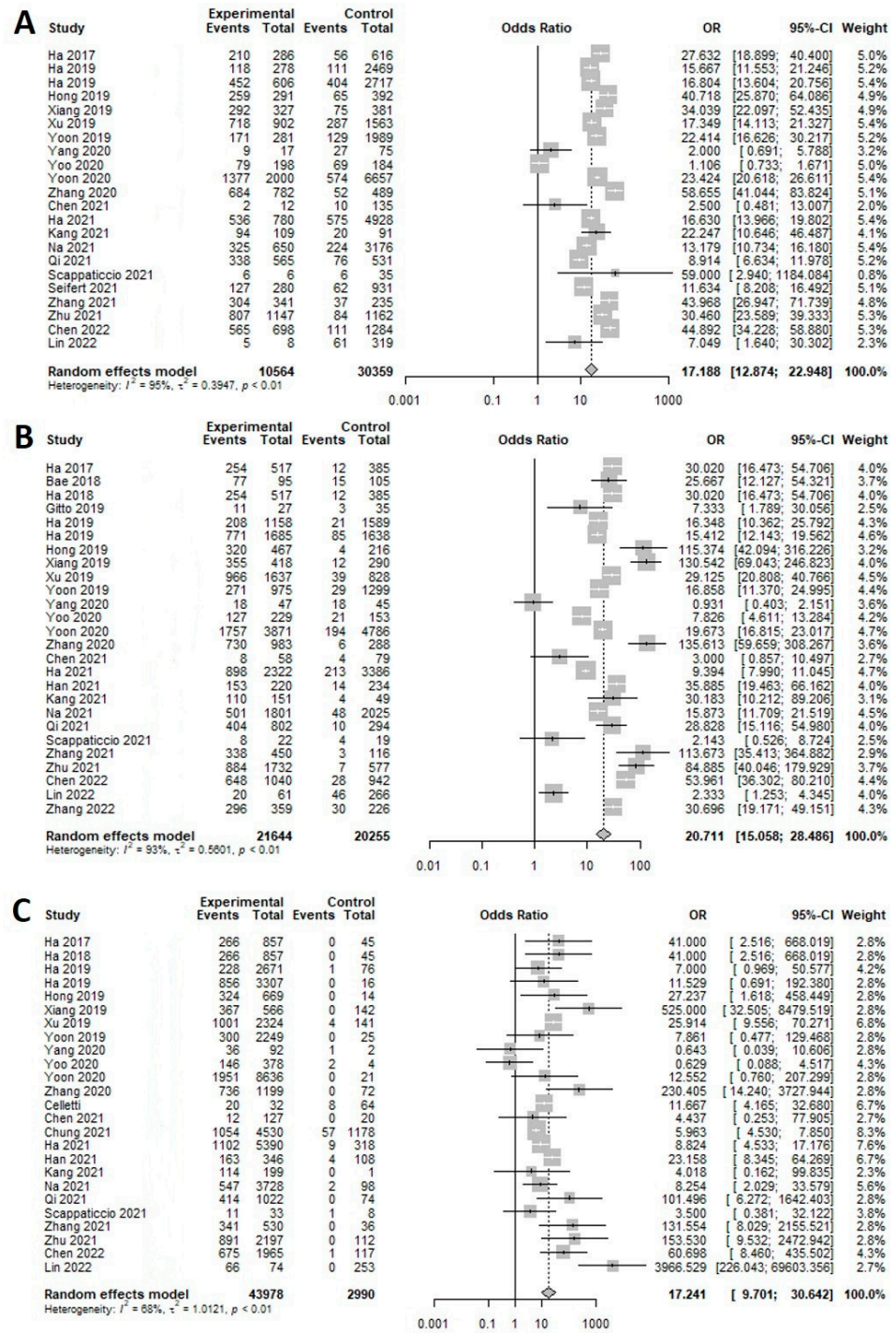


Figure 2. Forest plot of the diagnostic odds ratio for K-TIRAD. (A) High (K-TIRADS 5), (B) intermediate (K-TIRADS 4), (C) low (K-TIRADS 3).

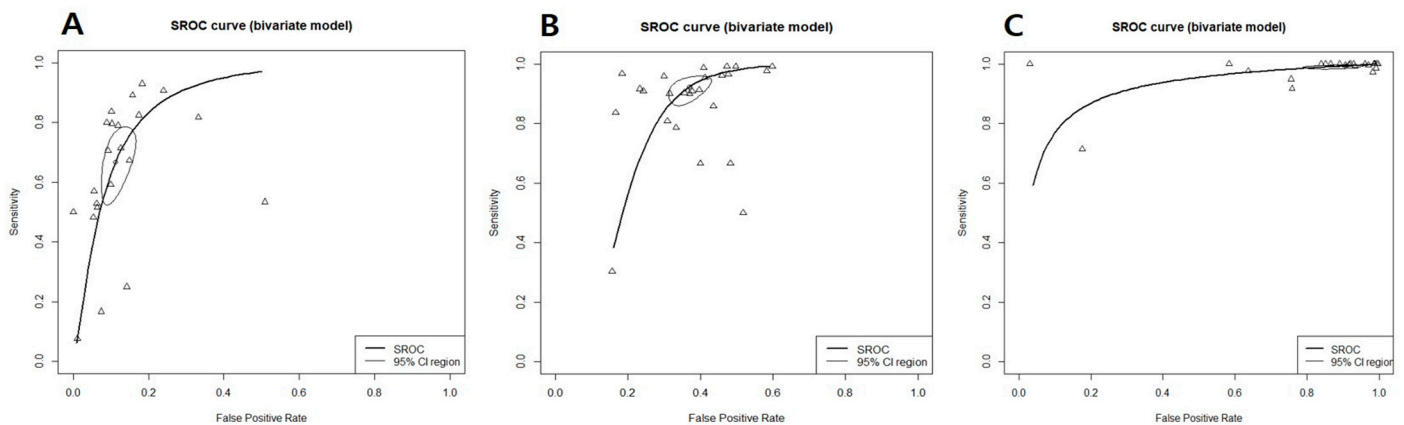


Figure 3. Summary receiver operating characteristic curve for K-TIRAD. (A) High (K-TIRADS 5), (B) intermediate (K-TIRADS 4), (C) low (K-TIRADS 3), thick curved line: summary receiver operating characteristic curve; thin circular line: 95% confident region; small circle: summary estimate; triangle: observed data.

In ACR-TIRADS categories, sensitivity changed from around 63.5% to 98.4% (highest in TR3) and specificity showed an inverse association (89.5% to 22.8%; highest in TR5) according to the different cutoff values (categories) (Table 2). ROC analysis and DOR showed that the best diagnostic cutoff values of ACR-TIRADS were TR5 in common (Supplementary Figures S6 and S7). Based on the statistical results, the best cutoff value of ACR-TIRADS was category TR5 with the sensitivity of 63.5% and specificity of 89.5%. However, TR4 would also be a good cutoff value of ACR-TIRADS based on another clinician’s opinion that high sensitivity could be more suitable than high specificity in the screening test.

Table 2. Diagnostic efficacy and the ROC curves of ACR-TIRADS categories.

	Sensitivity [95% CIs]	Specificity [95% CIs]	DOR [95% CIs]	AUC
TR5 (Suspicious)	0.6350 [0.5309; 0.7279]; I ² = 99.2%	0.8955 [0.8613; 0.9221]; I ² = 98.7%	16.8442 [13.5328; 20.9658]; I ² = 92.5%	0.882
TR4 (Moderately)	0.9249 [0.8808; 0.9535]; I ² = 98.0%	0.5343 [0.4782; 0.5896]; I ² = 98.9%	13.6381 [9.9396; 18.7128]; I ² = 93.6%	0.753
TR3 (Mildly)	0.9843 [0.9698; 0.9919]; I ² = 96.9%	0.2289 [0.1697; 0.3012]; I ² = 99.5%	13.2478 [9.1596; 19.1605]; I ² = 85.9%	0.769

ROC: receiver operating characteristic, CI: confidence interval, AUC: area under the curve; ACR-TIRADS: American College of Radiology-Thyroid Imaging Reporting and Data System.

In ATA categories, sensitivity changed from around 69.7% to 97.6% (highest in low suspicion) and specificity showed an inverse association (87.1% to 22.6%; highest in high suspicion) according to the different cutoff values (categories) (Table 3). ROC analysis and DOR showed that the best diagnostic cutoff values of ATA had high suspicion in common (Supplementary Figures S8 and S9). Statistically, the best cutoff value of ATA was category the “high” with the sensitivity of 69.7% and specificity of 87.1%. However, similar to ACR-TIRADS, intermediate suspicion would also be a good cut-off value of ATA in another clinician’s opinion that high sensitivity could be more suitable than high specificity in the screening test.

Table 3. Diagnostic efficacy and the ROC curves of ATA categories.

	Sensitivity [95% CIs]	Specificity [95% CIs]	DOR [95% CIs]	AUC
High	0.6977 [0.5992; 0.7809]; I ² = 98.8%	0.8715 [0.8082; 0.9161]; I ² = 99.4%	15.7398 [11.5605; 21.4299]; I ² = 95.2%	0.859
Intermediate	0.8800 [0.8239; 0.9199]; I ² = 97.9%	0.6155 [0.5471; 0.6796]; I ² = 99.2%	11.5148 [8.2698; 16.0332]; I ² = 95.0%	0.799
Low	0.9768 [0.9498; 0.9895]; I ² = 98.1%	0.2261 [0.1614; 0.3073]; I ² = 99.5%	6.7781 [4.1264; 11.1339]; I ² = 94.2%	0.694

ROC: receiver operating characteristic; CI: confidence interval; AUC: area under the curve; ATA: American Thyroid Association.

In EU-TIRADS categories, sensitivity changed from around 70.6% to 99.1% (highest in low risk) and specificity showed an inverse association (83.9% to 3%; highest in high risk) according to the different cutoff values (categories) (Table 4). ROC analysis and DOR showed that the best diagnostic cutoff values of EU-TIRADS were high risk and intermediate risk, respectively (Supplementary Figures S10 and S11). Statistically, the best cutoff value of EU-TIRADS was high risk with the sensitivity of 70.6% and specificity of 83.9%. However, like ACR-TIRADS, intermediate risk would also be a good cutoff value for EU-TIRADS in another clinician's opinion that high sensitivity could be more suitable than high specificity as the screening test.

Table 4. Diagnostic efficacy and the ROC curves of EU-TIRADS categories.

	Sensitivity [95% CIs]	Specificity [95% CIs]	DOR [95% CIs]	AUC
High (EU-TIRADS 5)	0.7060 [0.6034; 0.7912]; I ² = 98.1%	0.8392 [0.7707; 0.8901]; I ² = 99.3%	12.2986 [9.0027; 16.8010]; I ² = 93.6%	0.843
Intermediate (EU-TIRADS 4)	0.9304 [0.8968; 0.9536]; I ² = 94.2%	0.5061 [0.4274; 0.5845]; I ² = 99.2%	13.0061 [9.2913; 18.2062]; I ² = 88.9%	0.819
Low (EU-TIRADS 3)	0.9914 [0.9763; 0.9969]; I ² = 91.8%	0.0303 [0.0112; 0.0795]; I ² = 99.3%	2.9158 [1.4936; 5.6922]; I ² = 74.8%	0.734

ROC: receiver operating characteristic; CI: confidence interval; AUC: area under the curve; EU-TIRADS: European Thyroid Imaging Reporting and Data System.

In Kwak-TIRADS categories, sensitivity changed from approximately 15% to 99% (highest in 4a) and specificity showed an inverse association (99% to 32%; highest in 5) according to the different cutoff values (categories) (Table 5). ROC analysis and DOR showed that the best diagnostic cutoff values of Kwak-TIRADS were 4a and 4b, respectively (Supplementary Figures S12 and S13). A cutoff in the screening test has been chosen to minimize the rate of false negatives rather than reducing false positives, because this would be appropriate for conditions in which misdiagnosing and treating someone as sick is better than missing truly sick individuals [81]. Based on practical and statistical considerations, the best cutoff value of Kwak-TIRADS was category 4b with the sensitivity of 95.8% and specificity of 63.7%.

Table 5. Diagnostic efficacy and the ROC curves of Kwak-TIRADS categories.

	Sensitivity [95% CIs]	Specificity [95% CIs]	DOR [95% CIs]	AUC
5	0.1433 [0.1099; 0.1848]; I ² = 94.7%	0.9961 [0.9908; 0.9983]; I ² = 91.7%	25.8479 [12.8192; 52.1181]; I ² = 87.7%	0.647
4c	0.7538 [0.6426; 0.8391]; I ² = 98.5%	0.8904 [0.8205; 0.9352]; I ² = 99.2%	24.2039 [15.0245; 38.9914]; I ² = 96.6%	0.895
4b	0.9584 [0.9308; 0.9753]; I ² = 94.5%	0.6379 [0.4983; 0.7575]; I ² = 99.5%	38.0578 [22.2904; 64.9785]; I ² = 93.7%	0.929
4a	0.9908 [0.9799; 0.9958]; I ² = 95.3%	0.3286 [0.1986; 0.4914]; I ² = 99.8%	45.6067 [26.6992; 77.9037]; I ² = 88.4%	0.925

ROC: receiver operating characteristic; CI: confidence interval; AUC: area under the curve; Kwak-TIRADS: Kwak-Thyroid Imaging Reporting and Data System.

4. Discussion

The US-based RSSs have been useful for diagnosing thyroid nodules in clinical practice over the past decade (Supplementary Table S4). There have been many previous studies confirming the usefulness of each RSS. However, a comprehensive analysis of the cutoff value for diagnosing thyroid cancer has been lacking. Therefore, we confirmed the diagnostic accuracy and cutoff value including the most recent clinical studies for each RSS. This study analyzed the results of 76,512 thyroid nodules from 67 studies. The highest AUC of the K-TIRADS was 0.904 for low suspicion, but the false positive rate was high with a low specificity of 8%. Based on DOR, intermediate suspicion (K-TIRADS 4) showed the highest diagnostic accuracy. ACR-TIRADS showed the highest accuracy with AUC 0.882 in TR 5 and high sensitivity of 92.5% in TR 4. ATA classification demonstrated the highest diagnostic accuracy with an AUC of 0.859 in high suspicion and a high sensitivity of 88% in intermediate suspicion. In EU-TIRADS, EU-TIRADS 5 showed the highest diagnostic accuracy of 0.843, and EU-TIRADS 4 showed a high sensitivity of 93%. In Kwak-TIRADS, 4b showed a high

AUC of 0.925 and sensitivity of 95.8%. All values except low suspicion of ATA classification and EU-TIRADS showed good diagnostic accuracy of more than DOR 10 [82].

When compared with a meta-analysis study of the diagnostic performance of the four RSSs performed in 2019 [5], the sensitivity and AUC of K-TIRADS cutoff values 4 and 5 were higher in the present study. The sensitivity, specificity, and AUC of ACR-TIRADS in TR5 were similar to present study results or slightly higher in the previous meta-analysis. When TR4 was used as the cutoff value, sensitivity and AUC were lowered in the present study (95% vs. 92.5%, 0.88 vs. 0.75, respectively). In ATA and EU-TIRADS, both high and intermediate categories showed lower diagnostic accuracy in the present study, which included an additional 34 studies published after 2020. It was interesting that the sensitivity and diagnostic accuracy of K-TIRADS was higher than those of the meta-analysis performed in 2019, and that many studies on K-TIRADS were added to the present study.

Thyroid nodules are relatively common, and the incidence rate confirmed through palpation is approximately 4%, but after the popularization of US examination, the incidence rate was reported to be as high as 70% [3]. US can screen for thyroid cancer by confirming the imaging characteristics such as composition, echogenicity, shape, and margin of the thyroid nodule [73]. It is also the basis for deciding whether to proceed with additional diagnostic tests such as FNA biopsy and core needle biopsy [83]. According to a population study, the number of patients diagnosed with papillary thyroid cancers (PTCs) increased rapidly at about 3% per year [84]. Meanwhile, mortality from thyroid cancer was found to remain stable, inferring that a large proportion of PTCs is due to the overdiagnosis of low-risk tumors [85,86]. Various US-based RSSs have been proposed to avoid unnecessary additional examination of incidental thyroid nodules and to systematically evaluate and report the findings of thyroid nodules. The diagnostic performance of the five representative RSSs included in this study was presented in several studies. Analysis results of the present study revealed the highest AUCs of the ACR-TIRADS, EU-TIRADS, Kwak-TIRADS, K-TIRADS, and ATA classification of 0.882, 0.843, 0.929, 0.904, and 0.859, respectively, showing high diagnostic accuracy.

ROC is an integrated result showing the performance of a diagnostic test at various thresholds by using sensitivity and specificity [87]. The value showing the highest AUC can be used as a cutoff value for diagnosis. However, AUC alone cannot draw definitive conclusions about the cutoff value. Because AUC is the result of measuring performance for all thresholds, it includes both clinically meaningful and nonsignificant values [79]. Therefore, for AUC to be clinically meaningful, it must be understood in terms of gains and losses for individual patients. Higher false-positives lead to complications and increased costs due to unnecessary additional tests, and higher false-negatives increase mortality due to disease [88]. Low sensitivity in thyroid US means that cancer may be missed and treatment may be delayed, whereas low specificity means that many unnecessary biopsies could be performed. Therefore, the cutoff value of RSS should be evaluated by the clinician on a case-by-case basis by considering both specificity and sensitivity. In other words, for patients with thyroid cancer risk factors, a high sensitivity value can be selected as a cutoff value, and in situations where overdiagnosis is concerned, it is acceptable to select a value with high specificity and AUC rather than sensitivity. On the other hand, a new thyroid ultrasound technology image such as elastography, which reflects tissue deformation when an external force is applied to the thyroid nodule, has recently been used for thyroid nodule diagnosis along with conventional ultrasound findings [89]. More clinical studies are expected to be reported on additional diagnostic methods to increase sensitivity and specificity for thyroid nodules.

This study had several limitations. First, individual characteristics of patients at high risk of developing thyroid cancer, such as sex and age, were not considered. In addition, the countries and health care facility levels in which US was conducted were not considered. Secondly, the size of the thyroid nodule was not considered. Nodule size is an important factor in follow-up and treatment decisions for thyroid nodules. Recent studies have

suggested that criteria for the size of the nodule to be biopsied should be raised to avoid unnecessary procedures [88]. In 2021, modified K-TIRADS with revised biopsy criteria was proposed [90]. It was reported that modified K-TIRADS significantly reduced the unnecessary biopsy rate for small (≤ 2 cm) nodules while maintaining high sensitivity [27]. Therefore, when long-term clinical results are obtained for small nodules, an additional integrated cutoff value can be confirmed. Thirdly, studies were primarily retrospective in design. It seems that many prospective studies are needed for more accurate verification.

5. Conclusions

In this study, valid diagnostic accuracy for each RSS was confirmed, but superiority among RSS was not verified. The study confirmed the sensitivity and specificity change for each cutoff value and explained that the cutoff value can be set based on the clinical situation. When applying RSS to actual clinical practice, the pros and cons should be judged between additional examination and follow-up, with consideration of patient characteristics such as age and sex and based on the diagnostic accuracy of each cut-off value assessed.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers15020424/s1>, Supplementary Table S1. Search terms by database; Supplementary Table S2. Study characteristics; Supplementary Table S3. Methodological quality of the included studies; Supplementary Table S4. Brief description of the Risk Stratification Systems; Supplementary Figure S1. Begg's funnel plot for K-TIRAD; Supplementary Figure S2. Begg's funnel plot for ATA; Supplementary Figure S3. Begg's funnel plot for ACR; Supplementary Figure S4. Begg's funnel plot for EU; Supplementary Figure S5. Begg's funnel plot for Kwak-TIRAD; Supplementary Figure S6. Forest plot of the diagnostic odds ratio for ACR; Supplementary Figure S7. Summary receiver operating characteristic curve for ACR guidelines; Supplementary Figure S8. Forest plot of the diagnostic odds ratio for ATA; Supplementary Figure S9. Summary receiver operating characteristic curve for ATA; Supplementary Figure S10. Forest plot of the diagnostic odds ratio for EU; Supplementary Figure S11. Summary receiver operating characteristic curve for EU; Supplementary Figure S12. Forest plot of the diagnostic odds ratio for Kwak-TIRAD; Supplementary Figure S13. Summary receiver operating characteristic curve for Kwak-TIRAD.

Author Contributions: Conceptualization, J.-S.K., B.G.K., G.S. and S.H.H.; methodology, J.-S.K. and S.H.H.; software, S.H.H.; validation, S.H.H.; formal analysis, J.-S.K. and S.H.H.; investigation, J.-S.K. and S.H.H.; data curation, J.-S.K. and S.H.H.; writing—original draft preparation, J.-S.K. and S.H.H.; writing—review and editing, J.-S.K. and S.H.H.; visualization, J.-S.K. and S.H.H.; supervision, J.-S.K., B.G.K., G.S. and S.H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2022R1F1A1066232). The sponsors had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Russ, G.; Bonnema, S.J.; Erdogan, M.F.; 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. [[CrossRef](#)]
2. Shin, J.H.; Baek, J.H.; Chung, J.; Ha, E.J.; Kim, J.H.; Lee, Y.H.; Lim, H.K.; Moon, W.J.; Na, D.G.; Park, J.S.; 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. [[CrossRef](#)]
3. Ezzat, S.; Sarti, D.A.; Cain, D.R.; Braunstein, G.D. Thyroid incidentalomas. Prevalence by palpation and ultrasonography. *Arch. Intern. Med.* **1994**, *154*, 1838–1840. [[CrossRef](#)]
4. Li, R.; Wang, Y.; Du, L. A rapidly increasing trend of thyroid cancer incidence in selected East Asian countries: Joinpoint regression and age-period-cohort analyses. *Gland Surg.* **2020**, *9*, 968–984. [[CrossRef](#)]
5. Kim, P.H.; Suh, C.H.; Baek, J.H.; Chung, S.R.; Choi, Y.J.; Lee, J.H. Diagnostic performance of four ultrasound risk stratification systems: A systematic review and meta-analysis. *Thyroid* **2020**, *30*, 1159–1168. [[CrossRef](#)]
6. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ* **2009**, *339*, b2535. [[CrossRef](#)] [[PubMed](#)]

7. Kim, D.H.; Kim, S.W.; Basurrah, M.A.; Hwang, S.H. Clinical and laboratory features for various criteria of eosinophilic chronic rhinosinusitis: A systematic review and meta-analysis. *Clin. Exp. Otorhinolaryngol.* **2022**, *15*, 230–246. [[CrossRef](#)]
8. Reitsma, J.B.; Glas, A.S.; Rutjes, A.W.; Scholten, R.J.; Bossuyt, P.M.; Zwinderman, A.H. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J. Clin. Epidemiol.* **2005**, *58*, 982–990. [[CrossRef](#)]
9. Hoeboer, S.H.; van der Geest, P.J.; Nieboer, D.; Groeneveld, A.B. The diagnostic accuracy of procalcitonin for bacteraemia: A systematic review and meta-analysis. *Clin. Microbiol. Infect.* **2015**, *21*, 474–481. [[CrossRef](#)] [[PubMed](#)]
10. Whiting, P.F.; Rutjes, A.W.; Westwood, M.E.; Mallett, S.; Deeks, J.J.; Reitsma, J.B.; Leeflang, M.M.; Sterne, J.A.; Bossuyt, P.M. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann. Intern. Med.* **2011**, *155*, 529–536. [[CrossRef](#)] [[PubMed](#)]
11. Tessler, F.N.; Middleton, W.D.; Grant, E.G.; Hoang, J.K.; Berland, L.L.; Teefey, S.A.; Cronan, J.J.; Beland, M.D.; Desser, T.S.; Frates, M.C.; 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*, 587–595. [[CrossRef](#)] [[PubMed](#)]
12. Haugen, B.R.; Alexander, E.K.; Bible, K.C.; Doherty, G.M.; Mandel, S.J.; Nikiforov, Y.E.; Pacini, F.; Randolph, G.W.; Sawka, A.M.; 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. [[CrossRef](#)]
13. Kwak, J.Y.; Han, K.H.; Yoon, J.H.; Moon, H.J.; Son, E.J.; Park, S.H.; Jung, H.K.; Choi, J.S.; Kim, B.M.; Kim, E.K. Thyroid imaging reporting and data system for US features of nodules: A step in establishing better stratification of cancer risk. *Radiology* **2011**, *260*, 892–899. [[CrossRef](#)] [[PubMed](#)]
14. Xu, T.; Wu, Y.; Wu, R.-X.; Zhang, Y.-Z.; Gu, J.-Y.; Ye, X.-H.; Tang, W.; Xu, S.-H.; Liu, C.; Wu, X.-H. Validation and comparison of three newly-released Thyroid Imaging Reporting and Data Systems for cancer risk determination. *Endocrine* **2019**, *64*, 299–307. [[CrossRef](#)]
15. Shi, Y.X.; Chen, L.; Liu, Y.C.; Zhan, J.; Diao, X.H.; Fang, L.; Chen, Y. Differences among the Thyroid Imaging Reporting and Data System proposed by Korean, the American College of Radiology and the European Thyroid Association in the diagnostic performance of thyroid nodules. *Transl. Cancer Res.* **2020**, *9*, 4958–4967. [[CrossRef](#)] [[PubMed](#)]
16. Zhang, W.B.; Xu, W.; Fu, W.J.; He, B.L.; Liu, H.; Deng, W.F. 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*, 163–174. [[CrossRef](#)] [[PubMed](#)]
17. 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. *Quant. Imaging Med. Surg.* **2021**, *11*, 3604–3618. [[CrossRef](#)] [[PubMed](#)]
18. Alqahtani, S.M.; Alanesi, S.F.; Mahmood, W.S.; Moustafa, Y.M.; Moharram, L.M.; Alharthi, N.F.; Alzahrani, A.M.; Alalawi, Y.S. Clinical and ultrasonographic features in cancer risk stratification of indeterminate thyroid nodules. *Saudi Med. J.* **2022**, *43*, 473–478. [[CrossRef](#)]
19. 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.* **2022**, *13*, 899575. [[CrossRef](#)]
20. Lin, Y.; Lai, S.; Wang, P.; Li, J.; Chen, Z.; Wang, L.; Guan, H.; Kuang, J. Performance of current ultrasound-based malignancy risk stratification systems for thyroid nodules in patients with follicular neoplasms. *Eur. Radiol.* **2022**, *32*, 3617–3630. [[CrossRef](#)]
21. Qi, T.Y.; Chen, X.; Liu, H.; Mao, L.; Chen, J.; He, B.L.; Zhang, W.B. Comparison of thyroid nodule FNA rates recommended by ACR TI-RADS, Kwak TI-RADS and ATA guidelines. *Eur. J. Radiol.* **2022**, *148*, 110152. [[CrossRef](#)] [[PubMed](#)]
22. Thedinger, W.; Raman, E.; Dhingra, J.K. Comparative Study of ACR TI-RADS and ATA 2015 for Ultrasound Risk Stratification of Thyroid Nodules. *Otolaryngol. Head Neck Surg.* **2022**, *167*, 35–40. [[CrossRef](#)] [[PubMed](#)]
23. Zhang, J.; Gong, Z.; Li, S.; Fan, P.; Yue, G.; Zou, G.; He, S.; Wang, J.; Xu, J. The value of neutrophil-to-lymphocyte ratio combined with the thyroid imaging reporting and data system in the diagnosis of the nature of thyroid nodules. *J. Clin. Lab. Anal.* **2022**, *36*, e24429. [[CrossRef](#)]
24. Huh, S.; Lee, H.S.; Yoon, J.; Kim, E.K.; Moon, H.J.; Yoon, J.H.; Park, V.Y.; Kwak, J.Y. Diagnostic performances and unnecessary US-FNA rates of various TIRADS after application of equal size thresholds. *Sci. Rep.* **2020**, *10*, 10632. [[CrossRef](#)]
25. Chung, S.R.; Ahn, H.S.; Choi, Y.J.; Lee, J.Y.; Yoo, R.E.; Lee, Y.J.; Kim, J.Y.; Sung, J.Y.; Kim, J.H.; Baek, J.H. Diagnostic Performance of the Modified Korean Thyroid Imaging Reporting and Data System for Thyroid Malignancy: A Multicenter Validation Study. *Korean J. Radiol.* **2021**, *22*, 1579–1586. [[CrossRef](#)] [[PubMed](#)]
26. Freire da Silva, P.; Corrêa de Araújo Arcoverde, L.; de Siqueira Barbosa Arcoverde, L.; Tenório Wanderley Fernandes Lima, G.; Paes de Medeiros Lima, T.; José do Amaral, F.; Bandeira, F. Agreement Between American and European Thyroid Imaging, Reporting, and Data System (TIRADS) in the Diagnosis of 473 Thyroid Nodules From a Single Center in Brazil. *Endocr. Pract.* **2021**, *27*, 1108–1113. [[CrossRef](#)]
27. Ha, E.J.; Shin, J.H.; Na, D.G.; Jung, S.L.; Lee, Y.H.; Paik, W.; Hong, M.J.; Kim, Y.K.; Lee, C.Y. Comparison of the diagnostic performance of the modified Korean Thyroid Imaging Reporting and Data System for thyroid malignancy with three international guidelines. *Ultrasonography* **2021**, *40*, 594–601. [[CrossRef](#)]

28. Han, M.; Ha, E.J.; Park, J.H. Computer-Aided Diagnostic System for Thyroid Nodules on Ultrasonography: Diagnostic Performance Based on the Thyroid Imaging Reporting and Data System Classification and Dichotomous Outcomes. *AJNR Am. J. Neuroradiol.* **2021**, *42*, 559–565. [[CrossRef](#)]
29. Hekimsoy, İ.; Öztürk, E.; Ertan, Y.; Orman, M.N.; Kavukçu, G.; Özgen, A.G.; Özdemir, M.; Özbek, S.S. Diagnostic performance rates of the ACR-TIRADS and EU-TIRADS based on histopathological evidence. *Diagn. Interv. Radiol.* **2021**, *27*, 511–518. [[CrossRef](#)]
30. Kang, S.; Kwon, S.K.; Choi, H.S.; Kim, M.J.; Park, Y.J.; Park, D.J.; Cho, S.W. Comparison of Korean vs. American Thyroid Imaging Reporting and Data System in Malignancy Risk Assessment of Indeterminate Thyroid Nodules. *Endocrinol. Metab.* **2021**, *36*, 1111–1120. [[CrossRef](#)]
31. Na, D.G.; Paik, W.; Cha, J.; Gwon, H.Y.; Kim, S.Y.; Yoo, R.E. Diagnostic performance of the modified Korean Thyroid Imaging Reporting and Data System for thyroid malignancy according to nodule size: A comparison with five society guidelines. *Ultrasonography* **2021**, *40*, 474–485. [[CrossRef](#)] [[PubMed](#)]
32. 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. *Front. Endocrinol.* **2021**, *12*, 763897. [[CrossRef](#)] [[PubMed](#)]
33. Scappaticcio, L.; Maiorino, M.I.; Iorio, S.; Docimo, G.; Longo, M.; Grandone, A.; Luongo, C.; Cozzolino, I.; Piccardo, A.; Trimboli, P.; et al. Exploring the Performance of Ultrasound Risk Stratification Systems in Thyroid Nodules of Pediatric Patients. *Cancers* **2021**, *13*, 5304. [[CrossRef](#)]
34. Seifert, P.; Schenke, S.; Zimny, M.; Stahl, A.; Grunert, M.; Klemenz, B.; Freesmeyer, M.; Kreissl, M.C.; 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. [[CrossRef](#)]
35. Seminati, D.; Capitoli, G.; Leni, D.; Fior, D.; Vacirca, F.; Di Bella, C.; Galimberti, S.; L’Imperio, V.; Pagni, F. Use of Diagnostic Criteria from ACR and EU-TIRADS Systems to Improve the Performance of Cytology in Thyroid Nodule Triage. *Cancers* **2021**, *13*, 5439. [[CrossRef](#)] [[PubMed](#)]
36. Chen, X.; Kutaiba, N.; Pearce, S.; Digby, S.; Van Gelderen, D. Application of Thyroid Imaging Reporting and Data System (TIRADS) guidelines to thyroid nodules with cytopathological correlation and impact on healthcare costs. *Intern. Med. J.* **2022**, *52*, 1366–1373. [[CrossRef](#)]
37. Xiao, J.; Xiao, Q.; Cong, W.; Li, T.; Ding, S.; Shao, C.; Zhang, Y.; Liu, J.; Wu, M.; Jia, H. Discriminating Malignancy in Thyroid Nodules: The Nomogram Versus the Kwak and ACR TI-RADS. *Otolaryngol. Head Neck Surg.* **2020**, *163*, 1156–1165. [[CrossRef](#)]
38. Yang, W.; Fananapazir, G.; LaRoy, J.; Wilson, M.; Campbell, M.J. Can the American Thyroid Association, K-Tirads, and Acr-Tirads Ultrasound Classification Systems Be Used to Predict Malignancy in Bethesda Category IV Nodules? *Endocr. Pract.* **2020**, *26*, 945–952. [[CrossRef](#)]
39. Yoo, W.S.; Ahn, H.Y.; Ahn, H.S.; Chung, Y.J.; Kim, H.S.; Cho, B.Y.; Seo, M.; Moon, J.H.; Park, Y.J. Malignancy rate of Bethesda category III thyroid nodules according to ultrasound risk stratification system and cytological subtype. *Medicine* **2020**, *99*, e18780. [[CrossRef](#)]
40. Yoon, J.H.; Lee, H.S.; Kim, E.K.; Moon, H.J.; Park, V.Y.; Kwak, J.Y. Pattern-based vs. score-based guidelines using ultrasound features have different strengths in risk stratification of thyroid nodules. *Eur. Radiol.* **2020**, *30*, 3793–3802. [[CrossRef](#)]
41. Zhang, W.B.; Li, J.J.; Chen, X.Y.; He, B.L.; Shen, R.H.; Liu, H.; Chen, J.; He, X.F. SWE combined with ACR TI-RADS categories for malignancy risk stratification of thyroid nodules with indeterminate FNA cytology. *Clin. Hemorheol. Microcirc.* **2020**, *76*, 381–390. [[CrossRef](#)]
42. Celletti, I.; Fresilli, D.; De Vito, C.; Bononi, M.; Cardaccio, S.; Cozzolino, A.; Durante, C.; Grani, G.; Grimaldi, G.; Isidori, A.M.; et al. TIRADS, SRE and SWE in INDETERMINATE thyroid nodule characterization: Which has better diagnostic performance? *Radiol. Med.* **2021**, *126*, 1189–1200. [[CrossRef](#)]
43. Watkins, L.; O’Neill, G.; Young, D.; McArthur, C. Comparison of British Thyroid Association, American College of Radiology TIRADS and Artificial Intelligence TIRADS with histological correlation: Diagnostic performance for predicting thyroid malignancy and unnecessary fine needle aspiration rate. *Br. J. Radiol.* **2021**, *94*, 20201444. [[CrossRef](#)] [[PubMed](#)]
44. Yoon, S.J.; Na, D.G.; Gwon, H.Y.; Paik, W.; Kim, W.J.; Song, J.S.; Shim, M.S. Similarities and Differences Between Thyroid Imaging Reporting and Data Systems. *AJR Am. J. Roentgenol.* **2019**, *213*, W76–W84. [[CrossRef](#)] [[PubMed](#)]
45. Huang, B.L.; Ebner, S.A.; Makkar, J.S.; Bentley-Hibbert, S.; McConnell, R.J.; Lee, J.A.; Hecht, E.M.; Kuo, J.H. A Multidisciplinary Head-to-Head Comparison of American College of Radiology Thyroid Imaging and Reporting Data System and American Thyroid Association Ultrasound Risk Stratification Systems. *Oncologist* **2020**, *25*, 398–403. [[CrossRef](#)] [[PubMed](#)]
46. Koc, A.M.; Adibelli, Z.H.; Erkul, Z.; Sahin, Y.; Dilek, I. Comparison of diagnostic accuracy of ACR-TIRADS, American Thyroid Association (ATA), and EU-TIRADS guidelines in detecting thyroid malignancy. *Eur. J. Radiol.* **2020**, *133*, 109390. [[CrossRef](#)] [[PubMed](#)]
47. Peng, J.Y.; Pan, F.S.; Wang, W.; Wang, Z.; Shan, Q.Y.; Lin, J.H.; Luo, J.; Zheng, Y.L.; Hu, H.T.; Ruan, S.M.; et al. Malignancy risk stratification and FNA recommendations for thyroid nodules: A comparison of ACR TI-RADS, AACE/ACE/AME and ATA guidelines. *Am. J. Otolaryngol.* **2020**, *41*, 102625. [[CrossRef](#)]

48. Szczepanek-Parulska, E.; Wolinski, K.; Dobruch-Sobczak, K.; Antosik, P.; Ostalowska, A.; Krauze, A.; Migda, B.; Zylka, A.; Lange-Ratajczak, M.; Banasiewicz, T.; 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*, 2495. [[CrossRef](#)]
49. Wu, H.; Zhang, B.; Cai, G.; Li, J.; Gu, X. American College of Radiology thyroid imaging report and data system combined with K-RAS mutation improves the management of cytologically indeterminate thyroid nodules. *PLoS ONE* **2019**, *14*, e0219383. [[CrossRef](#)]
50. Wu, X.L.; Du, J.R.; Wang, H.; Jin, C.X.; Sui, G.Q.; Yang, D.Y.; Lin, Y.Q.; Luo, Q.; Fu, P.; Li, H.Q.; et al. Comparison and preliminary discussion of the reasons for the differences in diagnostic performance and unnecessary FNA biopsies between the ACR TIRADS and 2015 ATA guidelines. *Endocrine* **2019**, *65*, 121–131. [[CrossRef](#)]
51. Xiang, P.; Chu, X.; Chen, G.; Liu, B.; Ding, W.; Zeng, Z.; Wu, X.; Wang, J.; Xu, S.; Liu, C. Nodules with nonspecific ultrasound pattern according to the 2015 American Thyroid Association malignancy risk stratification system: A comparison to the Thyroid Imaging Reporting and Data System (TIRADS-Na). *Medicine* **2019**, *98*, e17657. [[CrossRef](#)]
52. Phuttharak, W.; Boonrod, A.; Klungboonkrong, V.; Witsawapaisan, T. Interrater Reliability of Various Thyroid Imaging Reporting and Data System (TIRADS) Classifications for Differentiating Benign from Malignant Thyroid Nodules. *Asian Pac. J. Cancer Prev.* **2019**, *20*, 1283–1288. [[CrossRef](#)] [[PubMed](#)]
53. Ruan, J.L.; Yang, H.Y.; Liu, R.B.; Liang, M.; Han, P.; Xu, X.L.; Luo, B.M. Fine needle aspiration biopsy indications for thyroid nodules: Compare a point-based risk stratification system with a pattern-based risk stratification system. *Eur. Radiol.* **2019**, *29*, 4871–4878. [[CrossRef](#)]
54. Shen, Y.; Liu, M.; He, J.; Wu, S.; Chen, M.; Wan, Y.; Gao, L.; Cai, X.; Ding, J.; Fu, X. Comparison of Different Risk-Stratification Systems for the Diagnosis of Benign and Malignant Thyroid Nodules. *Front. Oncol.* **2019**, *9*, 378. [[CrossRef](#)] [[PubMed](#)]
55. Li, J.; Li, H.; Yang, Y.; Zhang, X.; Qian, L. The KWAK TI-RADS and 2015 ATA guidelines for medullary thyroid carcinoma: Combined with cell block-assisted ultrasound-guided thyroid fine-needle aspiration. *Clin. Endocrinol.* **2020**, *92*, 450–460. [[CrossRef](#)] [[PubMed](#)]
56. Gao, L.; Xi, X.; Jiang, Y.; Yang, X.; Wang, Y.; Zhu, S.; Lai, X.; Zhang, X.; Zhao, R.; Zhang, B. Comparison among TIRADS (ACR TI-RADS and KWAK- TI-RADS) and 2015 ATA Guidelines in the diagnostic efficiency of thyroid nodules. *Endocrine* **2019**, *64*, 90–96. [[CrossRef](#)]
57. Gitto, S.; Grassi, G.; De Angelis, C.; Monaco, C.G.; Sdao, S.; Sardanelli, F.; Sconfienza, L.M.; Mauri, G. A computer-aided diagnosis system for the assessment and characterization of low-to-high suspicion thyroid nodules on ultrasound. *Radiol. Med.* **2019**, *124*, 118–125. [[CrossRef](#)] [[PubMed](#)]
58. Ha, S.M.; Baek, J.H.; Choi, Y.J.; Chung, S.R.; Sung, T.Y.; Kim, T.Y.; Lee, J.H. Malignancy risk of initially benign thyroid nodules: Validation with various Thyroid Imaging Reporting and Data System guidelines. *Eur. Radiol.* **2019**, *29*, 133–140. [[CrossRef](#)]
59. Ha, S.M.; Baek, J.H.; Na, D.G.; Suh, C.H.; Chung, S.R.; Choi, Y.J.; Lee, J.H. Diagnostic Performance of Practice Guidelines for Thyroid Nodules: Thyroid Nodule Size versus Biopsy Rates. *Radiology* **2019**, *291*, 92–99. [[CrossRef](#)]
60. Hong, H.S.; Lee, J.Y. Diagnostic Performance of Ultrasound Patterns by K-TIRADS and 2015 ATA Guidelines in Risk Stratification of Thyroid Nodules and Follicular Lesions of Undetermined Significance. *AJR Am. J. Roentgenol.* **2019**, *213*, 444–450. [[CrossRef](#)]
61. Persichetti, A.; Di Stasio, E.; Guglielmi, R.; Bizzarri, G.; Taccogna, S.; Misischi, I.; Graziano, F.; Petrucci, L.; Bianchini, A.; Papini, E. Predictive Value of Malignancy of Thyroid Nodule Ultrasound Classification Systems: A Prospective Study. *J. Clin. Endocrinol. Metab.* **2018**, *103*, 1359–1368. [[CrossRef](#)] [[PubMed](#)]
62. Ahmadi, S.; Herbst, R.; Oyekunle, T.; Jiang, X.; Strickland, K.; Roman, S.; Sosa, J.A. Using the ata and acr ti-rads sonographic classifications as adjunctive predictors of malignancy for indeterminate thyroid nodules. *Endocr. Pract.* **2019**, *25*, 908–917. [[CrossRef](#)]
63. Barbosa, T.L.M.; Junior, C.O.M.; Graf, H.; Cavalvanti, T.; Trippia, M.A.; da Silveira Ugino, R.T.; de Oliveira, G.L.; Granella, V.H.; de Carvalho, G.A. ACR TI-RADS and ATA US scores are helpful for the management of thyroid nodules with indeterminate cytology. *BMC Endocr. Disord.* **2019**, *19*, 112. [[CrossRef](#)]
64. Macedo, B.M.; Izquierdo, R.F.; Golbert, L.; Meyer, E.L.S. Reliability of Thyroid Imaging Reporting and Data System (TI-RADS), and ultrasonographic classification of the American Thyroid Association (ATA) in differentiating benign from malignant thyroid nodules. *Arch. Endocrinol. Metab.* **2018**, *62*, 131–138. [[CrossRef](#)]
65. Chng, C.L.; Tan, H.C.; Too, C.W.; Lim, W.Y.; Chiam, P.P.S.; Zhu, L.; Nadkarni, N.V.; Lim, A.Y.Y. Diagnostic performance of ATA, BTA and TIRADS sonographic patterns in the prediction of malignancy in histologically proven thyroid nodules. *Singapore Med. J.* **2018**, *59*, 578–583. [[CrossRef](#)]
66. Bae, J.M.; Hahn, S.Y.; Shin, J.H.; Ko, E.Y. Inter-exam agreement and diagnostic performance of the Korean thyroid imaging reporting and data system for thyroid nodule assessment: Real-time versus static ultrasonography. *Eur. J. Radiol.* **2018**, *98*, 14–19. [[CrossRef](#)] [[PubMed](#)]
67. Yoon, J.H.; Han, K.; Kim, E.K.; Moon, H.J.; Kwak, J.Y. Diagnosis and Management of Small Thyroid Nodules: A Comparative Study with Six Guidelines for Thyroid Nodules. *Radiology* **2017**, *283*, 560–569. [[CrossRef](#)]
68. Trimboli, P.; Deandrea, M.; Mormile, A.; Ceriani, L.; Garino, F.; Limone, P.P.; Giovanella, L. American Thyroid Association ultrasound system for the initial assessment of thyroid nodules: Use in stratifying the risk of malignancy of indeterminate lesions. *Head Neck* **2018**, *40*, 722–727. [[CrossRef](#)]

69. Chng, C.L.; Kurzwinski, T.R.; Beale, T. Value of sonographic features in predicting malignancy in thyroid nodules diagnosed as follicular neoplasm on cytology. *Clin. Endocrinol.* **2015**, *83*, 711–716. [[CrossRef](#)]
70. Yoon, J.H.; Lee, H.S.; Kim, E.K.; Moon, H.J.; Kwak, J.Y. Malignancy Risk Stratification of Thyroid Nodules: Comparison between the Thyroid Imaging Reporting and Data System and the 2014 American Thyroid Association Management Guidelines. *Radiology* **2016**, *278*, 917–924. [[CrossRef](#)] [[PubMed](#)]
71. Zhang, J.; Liu, B.J.; Xu, H.X.; Xu, J.M.; Zhang, Y.F.; Liu, C.; Wu, J.; Sun, L.P.; Guo, L.H.; Liu, L.N.; et al. Prospective validation of an ultrasound-based thyroid imaging reporting and data system (TI-RADS) on 3980 thyroid nodules. *Int. J. Clin. Exp. Med.* **2015**, *8*, 5911–5917. [[PubMed](#)]
72. Srinivas, M.N.; Amogh, V.N.; Gautam, M.S.; Prathyusha, I.S.; Vikram, N.R.; Retnam, M.K.; Balakrishna, B.V.; Kudva, N. A Prospective Study to Evaluate the Reliability of Thyroid Imaging Reporting and Data System in Differentiation between Benign and Malignant Thyroid Lesions. *J. Clin. Imaging Sci.* **2016**, *6*, 5. [[CrossRef](#)] [[PubMed](#)]
73. Ha, E.J.; Moon, W.J.; Na, D.G.; Lee, Y.H.; Choi, N.; Kim, S.J.; Kim, J.K. A Multicenter Prospective Validation Study for the Korean Thyroid Imaging Reporting and Data System in Patients with Thyroid Nodules. *Korean J. Radiol.* **2016**, *17*, 811–821. [[CrossRef](#)]
74. Mao, F.; Xu, H.X.; Zhao, C.K.; Bo, X.W.; Li, X.L.; Li, D.D.; Liu, B.J.; Zhang, Y.F.; Xu, J.M.; Qu, S. Thyroid imaging reporting and data system in assessment of cytological Bethesda Category III thyroid nodules. *Clin. Hemorheol. Microcirc.* **2017**, *65*, 163–173. [[CrossRef](#)]
75. Xu, T.; Gu, J.Y.; Ye, X.H.; Xu, S.H.; Wu, Y.; Shao, X.Y.; Liu, D.Z.; Lu, W.P.; Hua, F.; Shi, B.M.; 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. [[CrossRef](#)]
76. Ha, E.J.; Na, D.G.; Baek, J.H.; Sung, J.Y.; Kim, J.H.; Kang, S.Y. US Fine-Needle Aspiration Biopsy for Thyroid Malignancy: Diagnostic Performance of Seven Society Guidelines Applied to 2000 Thyroid Nodules. *Radiology* **2018**, *287*, 893–900. [[CrossRef](#)]
77. Ha, E.J.; Na, D.G.; Moon, W.J.; Lee, Y.H.; 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*, 1532–1537. [[CrossRef](#)]
78. Zhang, Z.; Lin, N. Clinical diagnostic value of American College of Radiology thyroid imaging report and data system in different kinds of thyroid nodules. *BMC Endocrine Disorders* **2022**, *22*, 145. [[CrossRef](#)] [[PubMed](#)]
79. Halligan, S.; Altman, D.G.; Mallett, S. Disadvantages of using the area under the receiver operating characteristic curve to assess imaging tests: A discussion and proposal for an alternative approach. *Eur. Radiol.* **2015**, *25*, 932–939. [[CrossRef](#)]
80. Treveltham, R. Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. *Front. Public Health* **2017**, *5*, 307. [[CrossRef](#)]
81. Labrique, A.B.; Pan, W.K. Diagnostic tests: Understanding results, assessing utility, and predicting performance. *Am. J. Ophthalmol.* **2010**, *149*, 878–881.e872. [[CrossRef](#)] [[PubMed](#)]
82. Deeks, J.J. Systematic reviews of evaluations of diagnostic and screening tests. *BMJ* **2001**, *323*, 157–162. [[CrossRef](#)] [[PubMed](#)]
83. Hong, M.J.; Na, D.G.; Baek, J.H.; Sung, J.Y.; Kim, J.H. Cytology-Ultrasonography Risk-Stratification Scoring System Based on Fine-Needle Aspiration Cytology and the Korean-Thyroid Imaging Reporting and Data System. *Thyroid* **2017**, *27*, 953–959. [[CrossRef](#)] [[PubMed](#)]
84. Brito, J.P.; Al Nofal, A.; Montori, V.M.; Hay, I.D.; Morris, J.C. The Impact of Subclinical Disease and Mechanism of Detection on the Rise in Thyroid Cancer Incidence: A Population-Based Study in Olmsted County, Minnesota During 1935 Through 2012. *Thyroid* **2015**, *25*, 999–1007. [[CrossRef](#)] [[PubMed](#)]
85. Ho, A.S.; Davies, L.; Nixon, I.J.; Palmer, F.L.; Wang, L.Y.; Patel, S.G.; Ganly, I.; Wong, R.J.; Tuttle, R.M.; Morris, L.G. Increasing diagnosis of subclinical thyroid cancers leads to spurious improvements in survival rates. *Cancer* **2015**, *121*, 1793–1799. [[CrossRef](#)] [[PubMed](#)]
86. Walgama, E.; Sacks, W.L.; Ho, A.S. Papillary thyroid microcarcinoma: Optimal management versus overtreatment. *Curr. Opin. Oncol.* **2020**, *32*, 1–6. [[CrossRef](#)] [[PubMed](#)]
87. Mandrekar, J.N. Receiver operating characteristic curve in diagnostic test assessment. *J. Thorac. Oncol.* **2010**, *5*, 1315–1316. [[CrossRef](#)] [[PubMed](#)]
88. Ha, E.J.; Na, D.G.; Baek, J.H. Korean thyroid imaging reporting and data system: Current status, challenges, and future perspectives. *Korean J. Radiol.* **2021**, *22*, 1569. [[CrossRef](#)] [[PubMed](#)]
89. Sigrist, R.M.; Liao, J.; El Kaffas, A.; Chammas, M.C.; Willmann, J.K. Ultrasound elastography: Review of techniques and clinical applications. *Theranostics* **2017**, *7*, 1303. [[CrossRef](#)]
90. Ha, E.J.; Chung, S.R.; Na, D.G.; Ahn, H.S.; Chung, J.; Lee, J.Y.; Park, J.S.; Yoo, R.-E.; Baek, J.H.; Baek, S.M. 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*, 2094. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.