Diagnostic performance of ultrasound risk stratification systems on thyroid nodules cytologically classified as indeterminate: a systematic review and meta-analysis

Zhichao Xing[1,2], Yuxuan Qiu[3], Jingqiang Zhu[1], Anping Su[1], Wenshuang Wu[1,2]

[1]Center of Thyroid and Parathyroid Surgery, West China Hospital, Sichuan University, Chengdu; [2]Laboratory of Thyroid and Parathyroid Disease, Frontiers Science Center for Disease-related Molecular Network, West China Hospital, Sichuan University, Chengdu; [3]Ultrasound Department, West China Hospital, Sichuan University, Chengdu, China

Correspondence to:
Wenshuang Wu, PhD, Center of Thyroid and Parathyroid Surgery, West China Hospital, Sichuan University, No. 37 Guo Xue Xiang, Chengdu, 610041, Sichuan Province, China and Laboratory of Thyroid and Parathyroid Disease, Frontiers Science Center for Disease-related Molecular Network, West China Hospital, Sichuan University, Chengdu, 610041, Sichuan Province, China

Tel. +86-28-85422467
Fax. +86-28-85422467
E-mail: wenshuang_wu@163.com

Purpose: Ultrasound (US) risk stratification systems (RSSs) are increasingly being utilized for the optimal management of thyroid nodules, including those with indeterminate cytology. The goal of this study was to evaluate the category-based diagnostic performance of US RSSs in identifying malignancy in indeterminate nodules.

Methods: This systematic review and meta-analysis was registered on PROSPERO (CRD42021266195). PubMed, EMBASE, and Web of Science were searched through December 1, 2022. Original articles reporting data on the performance of US RSSs for indeterminate nodules were included. The numbers of nodules classified as true negative, true positive, false negative, and false positive were extracted.

Results: Thirty-three studies evaluating 7,225 indeterminate thyroid nodules were included. The diagnostic accuracy was quantitatively synthesized using a Bayesian bivariate model based on the integrated nested Laplace approximation in R. For the intermediate- to high-risk category, the sensitivity levels of the American College of Radiology, the American Thyroid Association, the European Thyroid Association, the Korean Thyroid Association/Korean Society of Thyroid Radiology, and Kwak et al. were found to be 0.80, 0.72, 0.76, 0.96, and 0.97, respectively. The corresponding specificity measurements were 0.36, 0.50, 0.49, 0.28, and 0.17. Furthermore, for the high-risk category, the sensitivity values were 0.40, 0.46, 0.55, 0.47, and 0.10, while the specificity levels were 0.91, 0.90, 0.71, 0.91, and 0.99, respectively.

Conclusion: The overall diagnostic performance of the US RSSs was moderate in the differentiation of indeterminate nodules.

Keywords: Ultrasonography; Thyroid; Biopsy; Diagnosis; Meta-analysis

Key points: For the intermediate- to high-risk category, the sensitivity levels of the American College of Radiology Thyroid Imaging Reporting and Data System (TIRADS), American Thyroid Association guidelines, European Thyroid Association TIRADS, Korean Society of Thyroid Radiology TIRADS, and Kwak et al. TIRADS ranged from 0.72 to 0.97, while the specificity measurements ranged from 0.17 to 0.49. For the high-risk category, European Thyroid Association TIRADS demonstrated the highest sensitivity at 0.55, while Kwak TIRADS showed the highest specificity at 0.99. This study provided information regarding the performance of each RSS in the context of indeterminate nodules.

Introduction

Cytologically indeterminate thyroid nodules present a consistent challenge in medical management. Currently, the most effective method for determining which nodules require surgical intervention is the fine-needle aspiration biopsy (FNAB). However, cytological results remain indeterminate for 17%–23% of all nodules [1]. The introduction of the six-tiered Bethesda System for Reporting Thyroid Cytopathology (BSRTC) has been helpful in categorizing these results. This system divides indeterminate cytological results into three of the six categories: III (atypia of undetermined significance or follicular lesion of undetermined significance [AUS/FLUS]), IV (follicular neoplasm or suspicious for follicular neoplasm [FN/SFN]), and V (suspicious for malignancy [SM]). These categories correspond to malignancy rates of 5%–15%, 15%–30%, and 60%–75%, respectively [2,3].

The BSRTC and the American Thyroid Association (ATA) guidelines integrate recommendations of repeated FNAB or diagnostic thyroidectomy for indeterminate thyroid nodules [4,5]. However, the best approach for managing these nodules remains a topic of debate, with options ranging from active surveillance and repeated FNAB to core-needle biopsy (CNB), molecular testing, and diagnostic thyroidectomy. The challenge lies in striking a delicate balance between underestimating and undertreating thyroid cancer, and overtreating nodules that are ultimately diagnosed as benign following histological analysis [6,7]. Therefore, it is prudent to identify predictors that can help identify nodules that unequivocally require surgical intervention [8].

Among the diagnostic tools widely available, ultrasound (US) is often the first to be utilized in determining the next steps in such cases. US risk stratification systems (RSSs), more commonly known as thyroid imaging reporting and data systems (TIRADS), have been developed to enhance the selection process of thyroid lesions that necessitate further FNAB or active surveillance [9]. Each category within the US RSS is associated with an escalating likelihood of malignancy, thus warranting more aggressive clinical management [10]. Presently, numerous US RSSs are included in the available guidelines, and several studies have been conducted to evaluate the diagnostic performance of US RSSs on indeterminate nodules. Consequently, the present study was performed to consolidate the diagnostic performance of various US RSSs in detecting thyroid cancer within indeterminate nodules.

Materials and Methods

This systematic review and meta-analysis, registered under PROSPERO with the registration number CRD42021266195,

adheres to the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) extension for diagnostic test accuracy statements [11].

Literature Search

A literature search was conducted across the PubMed, EMBASE, and Web of Science databases through December 1, 2022. The search terms were as follows: thyroid AND (indeterminate OR undetermined OR suspicious OR Bethesda) AND ((thyroid imaging reporting and data system) OR TIRADS OR TI-RADS OR stratification OR classification). The search was restricted to publications in English, but no limitations were implemented based on publication date or whether the studies involved humans or animals.

Inclusion Criteria

First, studies or their subsets that reported data on any US RSS according to the following guidelines were eligible for inclusion: the American Association of Clinical Endocrinologists/American College of Endocrinology/Associazione Medici Endocrinologi (AACE/ACE/AME US RSS) [12], the American College of Radiology (ACR-TIRADS) [13], the American Thyroid Association (ATA US RSS) [4], the British Thyroid Association (BTA US RSS) [14], the European Thyroid Association (EU-TIRADS) [15], the French-TIRADS [16], the TIRADS by Horvath et al. [9], the Korean Society of Thyroid Radiology (K-TIRADS) [17], and the TIRADS by Kwak et al. [18]. These were used as diagnostic criteria for malignant thyroid nodules among patients with a previous indeterminate FNAB report. Next, indeterminate nodules that had at least surgical pathology were included in the meta-analysis. Exclusion criteria included (1) articles not relevant to the subject of this review; (2) review articles, editorials or letters, comments, and conference proceedings; (3) case reports or case series; and (4) articles not written in English.

Data Extraction

One investigator extracted descriptive data, which were then verified by another researcher. This descriptive data encompassed the study and test characteristics. Two separate reviewers independently gathered the numerical data. Any discrepancies in the data extraction were resolved through consensus. If the data could not be extracted, the authors reached out to the authors to request additional data.

Quality Assessment

Two reviewers independently evaluated the risk of bias and potential applicability issues using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool [19]. Discrepancies were resolved through consensus.

Data Synthesis

For each study, two-by-two tables were created, with the results demonstrating the highest performance selected if different radiologists separately evaluated the diagnostic performance. The criteria for positive test results were defined as either intermediate to high risk (category 4 or 5) or high risk (category 5). A Bayesian bivariate model of diagnostic test studies was implemented, utilizing the integrated nested Laplace approximation (INLA). This model provided accurate posterior marginal distributions for sensitivity and specificity, along with all hyperparameters, without the need for Markov chain Monte Carlo sampling [20]. Additionally, univariate estimates of sensitivity and specificity, complete with 95% credible intervals (CrIs), were made available for interpretation. The summary receiver operating characteristic curve was also provided. The area under the receiver operating characteristic curve (AUC) values, accompanied by 95% CrIs, were combined. Summary positive and negative likelihood ratios (LR+s and LR−s, respectively) were calculated from the summary sensitivity and specificity estimates. The Bayesian bivariate model incorporated four models for enhanced accuracy. In model 1, both sensitivity and specificity were modeled in the bivariate model. In models 2, 3, and 4, sensitivity and false-negative rate (1−specificity), false-positive rate (1−sensitivity) and specificity, and false-positive rate (1−sensitivity) and false-negative rate (1−specificity) were modeled in the bivariate model, respectively. Model selection was guided by the deviance information criterion (DIC), with a lower DIC indicating a better model fit. To test for publication bias, a Deeks funnel plot was constructed, and statistical significance was assessed using the Deeks asymmetry test. Subgroup analyses were performed according to indeterminate classifications (AUS/FLUS, SN/FSN, and SM). The bivariate meta-regression model considered the following variables: study design (prospective vs. retrospective), sample size (cutoff at 140, which was the median value of the proportions reported by the included studies), proportion of malignancy (cutoff at 35%, which was the median value of the proportions reported by the included studies), and study location (East Asia vs. other countries). All analyses were primarily conducted using R software ver. 4.0.5 (R Foundation for Statistical Computing, Vienna, Austria; https://www.r-project.org) and the R packages meta4diag 2.0.8 and INLA 21.02.23. A P-value of <0.05 was considered to indicate statistical significance.

Results

Literature Search

The study screening procedure is depicted in a PRISMA 2020 flow diagram (Fig. 1). In total, 968 records were identified from PubMed,
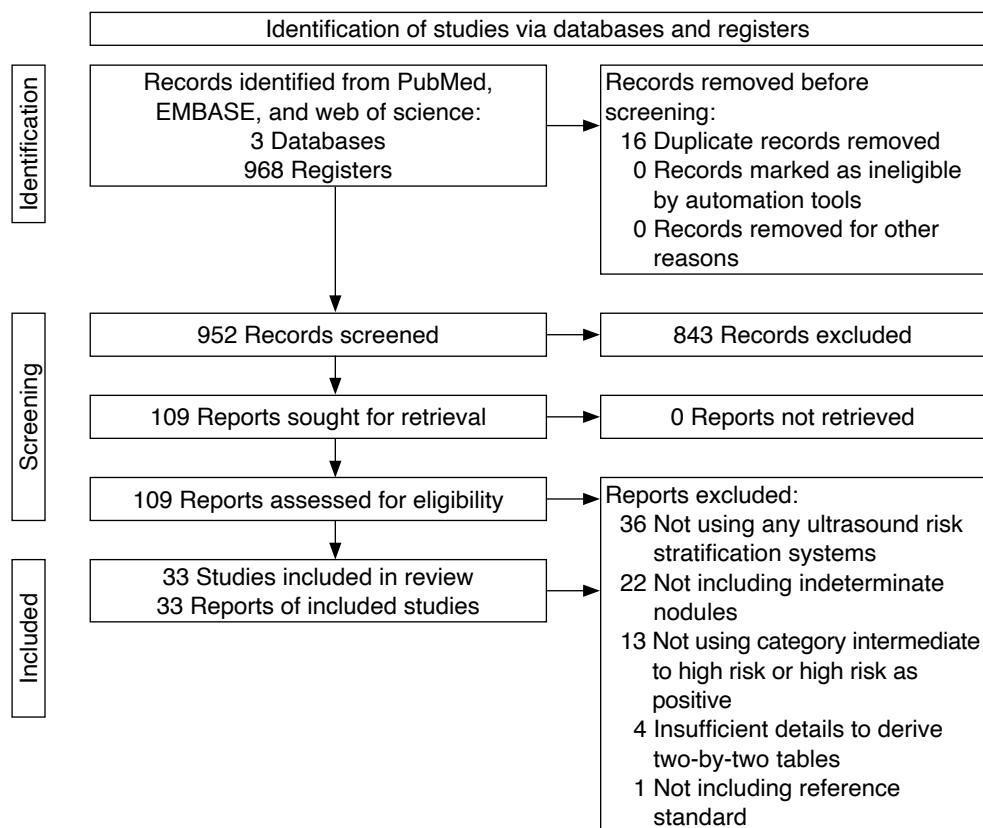

Fig. 1. PRISMA 2020 flow diagram.

Table 1. Study characteristics

Study ID	Country	Study design	Study period	No. of patients (female/male)	Age (year), mean±SD/median (range)	Indeterminate classification	No. of nodules (malignant/benign)	Prevalence of malignance (%)	Reference standard	US RSS
Ahmadi et al. (2019) [21]	United States	Retrospective	2010.1–2017.1	186 (154/32)	57 (NR)	BSRTC (III, IV)	202 (50/152)	24.8	Surgery	ACR and ATA
Al Dawish et al. (2020) [22]	Saudi Arabia	Retrospective	2011.1–2018.12	167 (118/49)	NR	BSRTC (III)	167 (46/121)	27.5	Surgery	ACR and ATA
Barbosa et al. (2019) [23]	Brazil	Retrospective	2012.1–2016.6	139 (118/21)	49 (13)	BSRTC (III, IV, V)	140 (66/74)	47.1	Surgery	ACR and ATA
Baser et al. (2017) [24]	Turkey	Retrospective	NA	618 (492/126)	48.1 (NR)	BSRTC (III)	640 (205/435)	32	Surgery	Kwak et al.
Capezzone et al. (2021) [25]	Italy	Retrospective	2009–2019	73 (54/19)	52 (18–81)	SIAPEC-AIT (TIR3A, 3B)	73 (29/44)	39.7	Surgery	EU
Celletti et al. (2021) [26]	Italy	Prospective	2017.1–2018.2	128 (89/39)[a]	54.3 (18–82)	SIAPEC-AIT (TIR3A, 3B)	96 (28/68)	29.2	Surgery	K
Chaigneau et al. (2018) [27]	France	Retrospective	2010.1–2016.12	602 (444/158)	50.9±14.8	BSRTC (III, IV, V)	602 (210/392)	34.9	Surgery	French
Chirayath et al. (2019) [28]	India	Prospective	2015.8–2017.8	176 (139/37)[a]	47±14	BSRTC (III, IV)	97 (57/40)	58.8	Surgery	Horvath et al.
He et al. (2017) [29]	China	Retrospective	2013.3–2016.9	453 (363/90)	51.2 (10–82)	BSRTC (III, IV, V)	453 (255/198)	56.3	Surgery	Kwak et al.
Hong et al. (2017) [30]	Korea	Retrospective	2010.1–2011.5	1457 (1,126/331)[a]	51±12.1	BSRTC (III, IV, V)	267 (114/153)	42.7	Surgery, repeat FNAB, CNB, and follow-up	K
Hong et al. (2019) [31]	Korea	Retrospective	2010.1–2016.12	683 (568/115)	49.7±11.8	BSRTC (III)	683 (324/359)	47.4	Surgery, repeated FNAB, CNB, and follow-up	ATA and K
Slowinska-Klencka et al. (2020) [32]	Poland	Retrospective	2010–2019	485 (433/52)	54.1 (NR)	BSRTC (III, IV, V)	540 (88/452)	16.3	Surgery	AACE/ACE/AME, ACR, ATA, EU, K, and Kwak et al.
Koh et al. (2016) [33]	Korea	Retrospective	2011.1–2013.12	221 (167/54)	50±13	BSRTC (III)	221 (34/187)	15.4	Surgery, repeated FNAB, and follow-up	Kwak et al.
Maia et al. (2015) [6]	Brazil	Retrospective	2000–2012	242 (208/34)[a]	46.5 (NR)	BSRTC (III, IV, V)	127 (50/77)	39.4	Surgery	French
Mao et al. (2017) [34]	China	Retrospective	2014.1–2015.12	121 (103/18)	55±11	BSRTC (III)	121 (43/78)	35.5	Surgery	Kwak et al.
Marina et al. (2021) [35]	Italy	Prospective	2014.11–2018	90 (65/25)	54±12.9	BSRTC (IV)	91 (34/57)	37.4	Surgery	ACR and EU
Mehta et al. (2020) [36]	India	Prospective	2018.7–2019.12	NR	NR	BSRTC (III, IV, V)	47 (31/16)	66	Surgery	ACR
Park et al. (2015) [37]	Korea	Retrospective	2010.1–2013.1	56 (46/10)	50.6 (23–76)	BSRTC (III)	58 (18/40)	31	Surgery	Kwak et al.
Piccardo et al. (2020) [38]	Italy	Retrospective, multicenter	2015.9–2019.5	111 (93/18)	57.6±15.7	SIAPEC-AIT (TIR3A, 3B)	111 (27/84)	24.3	Surgery	EU
Rho et al. (2017) [39]	Korea	Retrospective	2012.6–2016.12	297 (237/60)[a]	48.9±12.6	BSRTC (III)	78 (49/29)	62.8	Surgery	Kwak et al.
Rocha et al. (2019) [40]	Brazil	Prospective	2014–2017	137 (112/25)	NR	BSRTC (III, IV)	143 (51/92)	35.7	Surgery	ACR

Table 1. Continued

Study ID	Country	Study design	Study period	No. of patients (female/male)	Age (year), mean±SD/median (range)	Indeterminate classification	No. of nodules (malignant/benign)	Prevalence of malignance (%)	Reference standard	US RSS
Rosario et al. (2021) [41]	Brazil	Prospective	NR	323 (261/62)[a)]	51 (12−85)	BSRTC (III, IV)	299 (72/227)	24.1	Surgery	ACR and ATA
Sahli et al. (2019) [2]	USA	Retrospective	2012.2−2016.9	131 (94/37)	52.2 (17−80)	BSRTC (III, IV)	133 (30/103)	22.6	Surgery	ACR
Sahli et al. (2019)[a)] [42]	USA	Retrospective	2012.2−2016.9	127 (92/35)	52±14	BSRTC (III, IV)	127 (28/99)	22	Surgery	ACR
Suh et al. (2020) [43]	Korea	Retrospective	2007.1−2017.12	446 (389/57)	49.3 (NR)	BSRTC (III)	446 (193/253)	43.3	Surgery	K
Sultan et al. (2020) [44]	USA	Retrospective	2014.1−2017.9	98 (81/17)[a)]	57.4±12.3	BSRTC (III, IV)	17 (10/7)	58.8	Surgery	ACR, ATA
Trimboli et al. (2017) [45]	Switzerland	Retrospective	Since 2007	101 (68/33)	53.2±13.6	NR	101 (21/80)	20.8	Surgery	AACE/ACE/AME, ATA, BTA, and Horvath et al.
Ulisse et al. (2017) [46]	Italy	Retrospective	2005.1−2013.12	69 (52/17)	58 (13−77)	SIAPEC-AIT (TIR3A, 3B)	69 (17/52)	24.6	Surgery	K and Kwak et al.
Wang et al. (2020) [47]	USA	Retrospective	2012.9−2016.3	281 (228/53)[a)]	51 (NR)	BSRTC (III, IV)	268 (84/184)	31.3	Surgery and follow-up	ACR and ATA
Wu et al. (2019) [48]	China	Retrospective	2017.1−2018.6	43 (34/9)[a)]	47.6±15.5	BSRTC (III, IV)	41 (12/29)	29.3	Surgery	ACR
Yoo et al. (2020) [49]	Korea	Retrospective, multicenter	2010.1−2015.10	382 (297/85)	50.5 (NR)	BSRTC (III)	382 (148/234)	38.7	Surgery, CNB, and follow-up	ACR, ATA, EU, and K
Yoon et al. (2016) [50]	Korea	Retrospective	2011.7−2013.1	188 (145/43)	50.2±11.8	BSRTC (III)	192 (82/110)	42.7	Surgery, repeated FNAB, and follow-up	Kwak et al.
Zhang et al. (2020) [51]	China	Retrospective	2014.1−2019.12	193 (152/41)	46.1±13.1	BSRTC (III, IV, V)	193 (156/37)	80.8	Surgery	ACR

SD, standard deviation; US RSS, ultrasound risk stratification system(s); NR, not reported; BSRTC, Bethesda System for Reporting Thyroid Cytopathology; ACR, American College of Radiology; ATA, American Thyroid Association; SIAPEC-AIT, Italian Society for Anatomic Pathology and Cytology-Italian Division of the International Academy of Pathology; EU, European Thyroid Association; K, Korean Society of Thyroid Radiology; FNAB, fine-needle aspiration biopsy; CNB, core-needle biopsy; AACE/ACE/AME, American Association of Clinical Endocrinologists/American College of Endocrinology/Associazione Medici Endocrinologi; BTA, British Thyroid Association.

[a)]No. of patients were from the whole cohort; however, only indeterminate nodules of the cohort were included in the meta-analysis.

EMBASE, and Web of Science, with an additional six articles retrieved from other sources. Following the selection process, 33 articles were included in the meta-analysis [6,10,21−51].

Study Characteristics

Table 1 and Supplementary Table 1 display the characteristics and two-by-two data presentation of the included articles, respectively. Of the 33 studies, six were prospective in design [23,25,32,33,37,38], and only two were multicenter studies [38,49]. These studies were published between 2015 and 2021, with the number of evaluated indeterminate nodules ranging from 17 to 683. Two studies assessed the AACE/ACE/AME US RSS, 15 evaluated the ACR-TIRADS, 10 examined the ATA US RSS, one looked at the

BTA US RSS, five studied the EU-TIRADS, two analyzed the French-TIRADS, two investigated the TIRADS described by Horvath et al.; seven explored the K-TIRADS, and nine scrutinized the TIRADS delineated by Kwak et al. The prevalence of malignant indeterminate nodules in each study ranged from 15.4% to 80.8%. Most of the studies reported surgical pathology as the reference standard for malignant and benign diagnosis, with the exception of six studies that added repeated FNAB, CNB, or follow-up for reference [30,31,33,47,49,50]. In total, this review included 2,662 malignant and 4,563 benign nodules.

Quality Assessment

The results of the quality assessment using the QUADAS-2 tool are

depicted in Fig. 2. Generally, all studies achieved the required quality standards. However, some were identified as having unclear or high risk of bias or concerns regarding applicability.

Diagnostic Performance of the US RSSs

Fig. 3 summarizes the estimates of diagnostic performance, considering intermediate to high risk as positive. The specific results of each study are presented in Supplementary Fig. 1. The overall sensitivity and specificity for US RSSs were found to be 0.86 (95%

CrI, 0.80 to 0.91) and 0.33 (95% CrI, 0.25 to 0.41), respectively. For ACR-TIRADS, the sensitivity and specificity were 0.80 (95% CrI, 0.70 to 0.88) and 0.36 (95% CrI, 0.23 to 0.49), respectively. In the case of the ATA US RSS, the sensitivity and specificity were 0.72 (95% CrI, 0.50 to 0.88) and 0.50 (95% CrI, 0.35 to 0.64), respectively. For the EU-TIRADS, the sensitivity and specificity were 0.76 (95% CrI, 0.58 to 0.88) and 0.49 (95% CrI, 0.32 to 0.66), respectively. For the K-TIRADS, the sensitivity and specificity were 0.96 (95% CrI, 0.81 to 1.00) and 0.28 (95% CrI, 0.02 to 0.73), respectively. For the

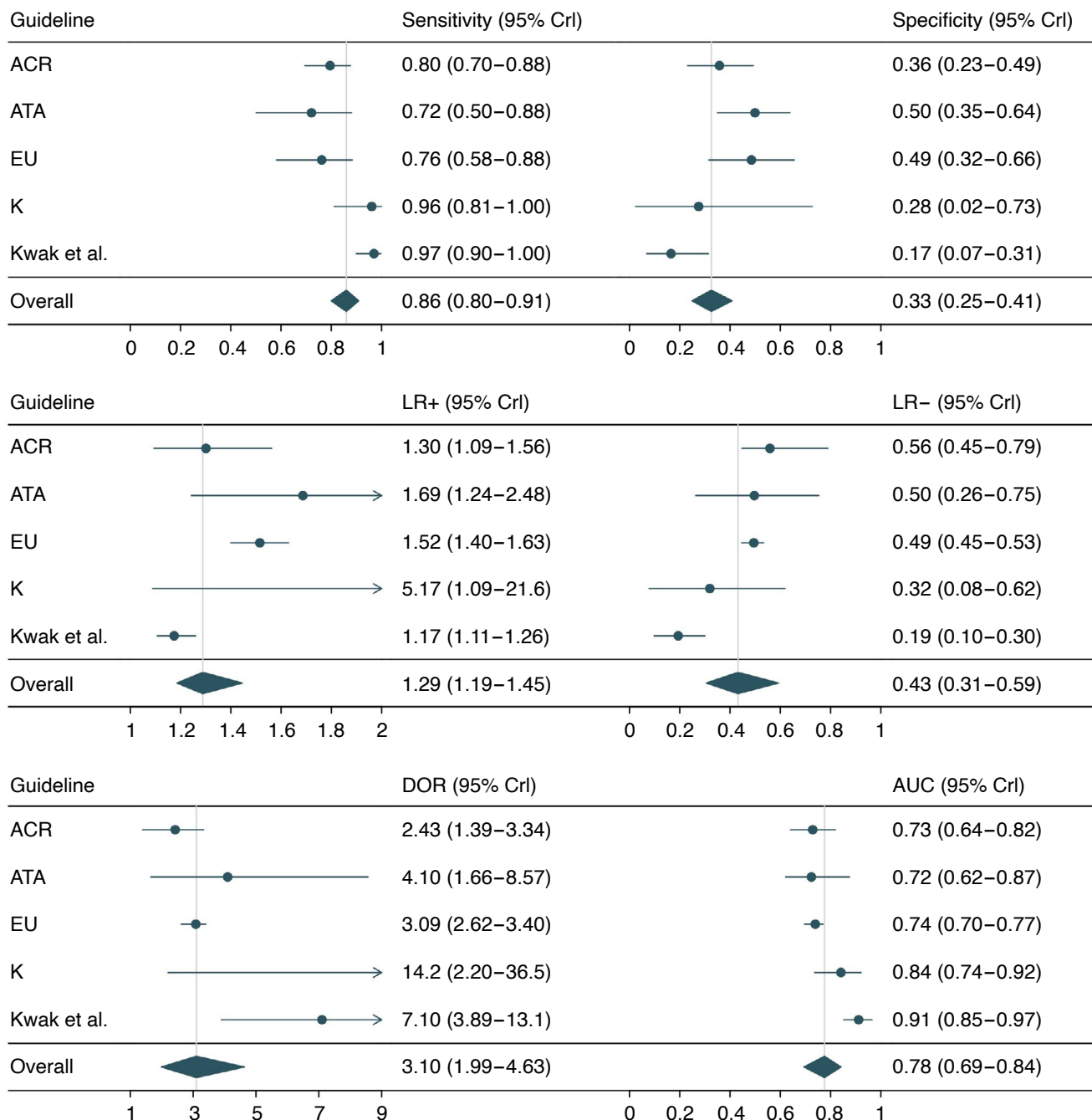Fig. 2. Quality assessment based on the QUADAS-2 tool [6,21–51].

Fig. 3. Estimates of ultrasound risk stratification systems for categorization of intermediate to high risk as positive. CrI, credible interval; ACR, American College of Radiology; ATA, American Thyroid Association; EU, European Thyroid Association; K, Korean Society of Thyroid Radiology; LR+, positive likelihood ratio; LR−, negative likelihood ratio; DOR, diagnostic odds ratio; AUC, area under the receiver operating characteristic curve.

TIRADS described by Kwak et al., the sensitivity and specificity were 0.97 (95% CrI, 0.90 to 1.00) and 0.17 (95% CrI, 0.07 to 0.31), respectively. Furthermore, the Deeks funnel plot and asymmetry test did not indicate a significant probability of publication bias, with the exception of the ATA US RSS (P=0.023). The summary receiver operating characteristic curve of the diagnostic performance

of each ultrasound risk stratification system for categorization of intermediate to high risk as positive was shown in Supplementary Fig. 2.

Fig. 4 summarizes the estimates of diagnostic performance, with high risk considered positive. Specific data related to different US RSSs are presented in Supplementary Fig. 3. The overall sensitivity
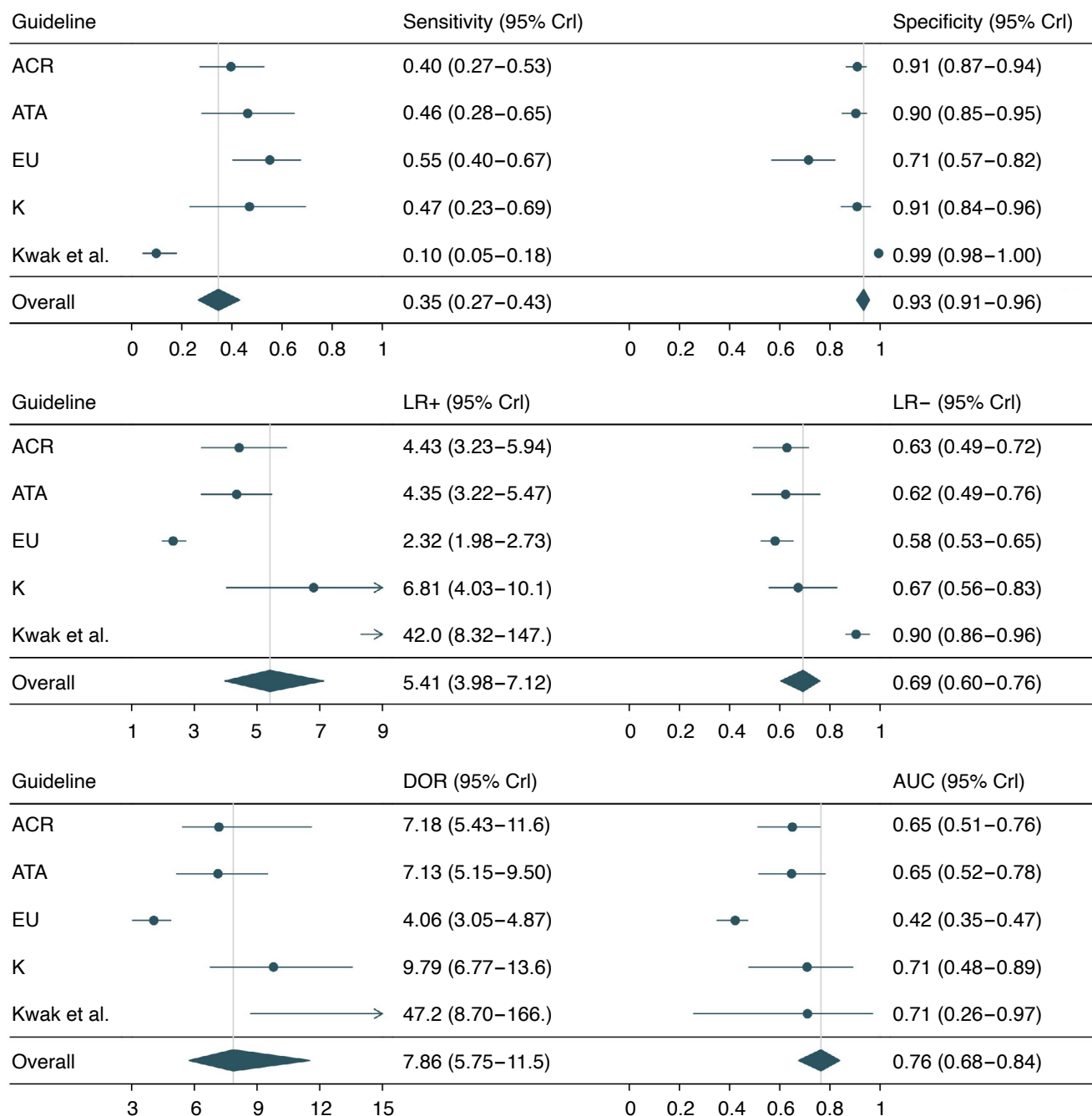
Fig. 4. Estimates of ultrasound risk stratification systems for categorization of high risk as positive. CrI, credible interval; ACR, American College of Radiology; ATA, American Thyroid Association; EU, European Thyroid Association; K, Korean Society of Thyroid Radiology; LR+, positive likelihood ratio; LR−, negative likelihood ratio; DOR, diagnostic odds ratio; AUC, area under the receiver operating characteristic curve.

and specificity for US RSSs were found to be 0.35 (95% CrI, 0.27 to 0.43) and 0.93 (95% CrI, 0.91 to 0.96), respectively. For the ACR-TIRADS, the sensitivity and specificity were 0.40 (95% CrI, 0.27 to 0.53) and 0.91 (95% CrI, 0.87 to 0.94), respectively. The ATA US RSS showed a sensitivity and specificity of 0.46 (95% CrI, 0.28 to

0.65) and 0.90 (95% CrI, 0.85 to 0.95), respectively. The EU-TIRADS demonstrated a sensitivity and specificity of 0.55 (95% CrI, 0.40 to 0.67) and 0.71 (95% CrI, 0.57 to 0.82), respectively. For the K-TIRADS, the sensitivity and specificity were 0.47 (95% CrI, 0.23 to 0.69) and 0.91 (95% CrI, 0.84 to 0.96), respectively. The TIRADS

by Kwak et al. showed a sensitivity and specificity of 0.10 (95% CrI, 0.05 to 0.18) and 0.99 (95% CrI, 0.98 to 1.00), respectively. Furthermore, no significant probability of publication bias was detected. The summary receiver operating characteristic curve of the diagnostic performance of each ultrasound risk stratification system for categorization of high risk as positive was shown in Supplementary Fig. 4.

The results of model selection guided by the DIC were shown in Supplementary Table 2.

Subgroup Analysis

A subgroup analysis was conducted based on various indeterminate categories, namely AUS/FLUS, FN/SFN, and SM. The results for AUS/FLUS are displayed in Table 2 and Supplementary Table 3. When considering intermediate to high risk as positive, the highest sensitivity was observed in TIRADS by Kwak et al., while the highest specificity was found in ATA US RSS. The overall sensitivity and

specificity were 0.90 (95% CrI, 0.82 to 0.96) and 0.40 (95% CrI, 0.24 to 0.57), respectively. When high risk was considered positive, the highest sensitivity was seen in the EU-TIRADS, and the highest specificity was found in the TIRADS delineated by Kwak et al. The overall sensitivity and specificity in this case were 0.33 (95% CrI, 0.23 to 0.44) and 0.94 (95% CrI, 0.90 to 0.97), respectively.

For FN/SFN and SM, the number of studies was insufficient to conduct quantitative analysis for each US RSS. In the FN/SFN subgroup, the overall sensitivity and specificity were 0.64 (95% CrI, 0.44 to 0.81) and 0.40 (95% CrI, 0.26 to 0.56), respectively, when categorizing intermediate to high risk as positive. When categorizing high risk as positive, the sensitivity and specificity were 0.22 (95% CrI, 0.11 to 0.36) and 0.89 (95% CrI, 0.79 to 0.96), respectively. For SM, the overall sensitivity and specificity were 0.89 (95% CrI, 0.78 to 0.96) and 0.23 (95% CrI, 0.11 to 0.38), respectively, when categorizing intermediate to high risk as positive. When categorizing high risk as positive, the sensitivity and specificity were 0.49 (95%

Table 2. Diagnostic performance of each US RSS for AUS/FLUS (sensitivity, specificity)

US RSS	Intermediate to high-risk as positive		High-risk as positive	
	Sensitivity (95% CrI)	Specificity (95% CrI)	Sensitivity (95% CrI)	Specificity (95% CrI)
ACR	0.76 (0.49−0.93)	0.57 (0.36−0.77)	0.33 (0.17−0.50)	0.91 (0.82−0.97)
ATA	0.75 (0.34−0.96)	0.74 (0.36−0.96)	0.48 (0.22−0.73)	0.90 (0.78−0.98)
EU	0.88 (0.70−0.95)	0.55 (0.24−0.84)	0.59 (0.41−0.73)	0.77 (0.64−0.88)
K	0.95 (0.85−1.00)	0.75 (0.13−1.00)	0.54 (0.32−0.73)	0.89 (0.84−0.93)
Kwak et al.	0.95 (0.86−0.99)	0.17 (0.06−0.34)	0.14 (0.10−0.20)	0.99 (0.97−1.00)
Overall	0.90 (0.82−0.96)	0.40 (0.24−0.57)	0.33 (0.23−0.44)	0.94 (0.90−0.97)

US RSS, ultrasound risk stratification system; AUS/FLUS, atypia of undetermined significance or follicular lesion of undetermined significance; CrI, credible interval; ACR, American College of Radiology; ATA, American Thyroid Association; EU, European Thyroid Association; K, Korean Society of Thyroid Radiology.

Table 3. Meta-regression analysis for all US RSSs

Covariate	Intermediate to high-risk as positive				High-risk as positive			
	Sensitivity (95% CrI)	P-value	Specificity (95% CrI)	P-value	Sensitivity (95% CrI)	P-value	Specificity (95% CrI)	P-value
Study design								
Prospective	0.85 (0.69−1.00)	0.332	0.47 (0.23−0.72)	0.151	0.42 (0.18−0.66)	0.319	0.88 (0.77−0.99)	0.011
Retrospective	0.86 (0.80−0.93)		0.30 (0.21−0.39)		0.33 (0.23−0.43)		0.94 (0.92−0.97)	
No. of nodules								
>140	0.90 (0.83−0.96)	0.404	0.36 (0.23−0.49)	0.090	0.37 (0.26−0.48)	0.128	0.94 (0.90−0.97)	0.014
<140	0.82 (0.72−0.92)		0.29 (0.18−0.40)		0.29 (0.15−0.44)		0.93 (0.88−0.98)	
Prevalence of malignance								
>35%	0.92 (0.87−0.97)	0.793	0.32 (0.19−0.44)	0.331	0.45 (0.32−0.58)	0.010	0.91 (0.86−0.96)	<0.001
<35%	0.78 (0.67−0.88)		0.33 (0.21−0.45)		0.24 (0.14−0.34)		0.95 (0.93−0.98)	
Study location								
East Asia	0.95 (0.92−0.99)	0.792	0.29 (0.15−0.43)	0.523	0.47 (0.31−0.63)	0.031	0.93 (0.88−0.98)	<0.001
Others	0.78 (0.69−0.87)		0.34 (0.23−0.45)		0.28 (0.19−0.38)		0.94 (0.91−0.97)	

US RSSs, ultrasound risk stratification systems; CrI, credible interval.

CrI, 0.31 to 0.68) and 0.99 (95% CrI, 0.95 to 1), respectively (Supplementary Table 4).

Meta-Regression

The results of the meta-regression are outlined in Table 3 (all US RSSs) and Supplementary Table 5 (each US RSS). Overall, no significant covariates were identified when the risk was set to intermediate or high. However, the sensitivity of the high-risk category was influenced by variations in malignant prevalence (P=0.010) and study location (P=0.031). The specificity, in contrast, was potentially affected by all four covariates: study design (P=0.011), number of nodules (P=0.014), prevalence of malignancy (P<0.01), and study location (P<0.01).

Discussion

To the best of the authors' knowledge, the present study is the first in the literature to investigate the utility of US RSSs in patients with cytologically indeterminate nodules. The current meta-analysis examined the diagnostic performance of various US RSSs, using 33 studies that included 7,225 indeterminate thyroid nodules. Limited data were available on the AACE/ACE/AME, BTA, French, and Horvath et al. TIRADS. However, more studies were found evaluating the ACR TIRADS, ATA US RSS, EU-TIRADS, K-TIRADS, and TIRADS outlined by Kwak et al. Most US RSSs are pattern-based systems. For instance, the K-TIRADS incorporates solidity, echogenicity, and suspicious features (nonparallel orientation, spiculated/ microlobulated margin, and microcalcifications) to stratify nodules [17]. Other examples include the ATA US RSS and the EU-TIRADS. In contrast, some US RSSs are scoring systems. For example, with the ACR-TIRADS, all US characteristics are integrated and scored from 0 to 3 based on their malignant potential [13]. The Kwak TIRADS also employs a score-based system. The advantage of pattern-based systems is that they are intuitive and practical for clinical application, while a scoring system may provide a more objective evaluation of each nodule [52].

In the present meta-analysis, individual system meta-analyses were used to identify the threshold categories with the highest accuracy for indeterminate nodules. These categories included TR5 (highly suspicious) for the ACR TIRADS, high suspicion for the ATA system, EU-TIRADS 5 (high risk) for the EU-TIRADS, K-TIRADS 5 (high suspicion) for the K-TIRADS, and category 5 (highly suggestive of malignancy) for the Kwak TIRADS. At these category thresholds, the RSSs demonstrated a sensitivity of 10%–55%, a specificity of 71%–99%, and an accuracy of 69%–79% (Fig. 4, Supplementary Table 6). Kim et al. [53] reported similar results for thyroid nodules across all categories, with a higher sensitivity of 65%–77% and a

higher specificity of 82%–90%. However, the difference lay in the threshold categories with the highest accuracy for the Kwak TIRADS, which was category 4c in the study by Kim et al. Overall, the clinical application of US RSSs in indeterminate nodules provides valuable information for deciding between surgical treatment or active surveillance.

The diagnostic performance for indeterminate nodules varied among US RSSs. For the category deemed intermediate to high risk, the highest sensitivity was observed with the Kwak TIRADS (0.97; 95% CrI, 0.90 to 1.00), while the lowest was seen with the ATA US RSS (0.72; 95% CrI, 0.50 to 0.88). Conversely, the specificity was highest for the ATA US RSS (0.50; 95% CrI, 0.35 to 0.64) and lowest for the Kwak TIRADS (0.17; 95% CrI, 0.07 to 0.31). For the high-risk category, the highest and lowest sensitivity values were observed for the EU-TIRADS (0.55; 95% CrI, 0.40 to 0.67) and the Kwak TIRADS (0.10; 95% CrI, 0.05 to 0.18), respectively. The specificity was highest for the EU-TIRADS (0.71; 95% CrI, 0.57 to 0.82) and lowest for the Kwak TIRADS (0.99; 95% CrI, 0.98 to 1.00). However, due to the absence of studies directly comparing different US RSSs, these differences should be interpreted with caution. The variation in diagnostic performance was not solely due to the overlapping US appearance of benign and malignant nodules, but also to substantial variability in thyroid nodule reporting and recommendations for further workup [35]. Limited evidence was available of differences in interobserver agreement among US RSSs, with only Sahli et al. [42] reporting moderate agreement for ACR-TIRADS among the three participating radiologists. Compared to US features, the use of US RSSs may improve interobserver agreement, and when selecting nodules for FNAB, the interobserver agreement can approach perfection [54,55]. US practitioners can adapt each RSS to their clinical setting, considering the proportion of malignant thyroid nodules and other factors. In primary hospitals, most patients present due to thyroid nodules detected during routine physical examinations. However, in tertiary hospitals, many patients are referred due to an initial diagnosis and surgical recommendation from a primary hospital. Consequently, these tertiary hospitals tend to have a higher proportion of malignant nodules. Table 3 indicates that a higher proportion of malignant nodules can increase sensitivity and decrease specificity, leading to a high proportion of false positive cases. In such cases, clinicians can opt for noninvasive strategies such as active surveillance for nodules of similar categories to avoid unnecessary FNAB. Conversely, in situations with lower proportions of malignant nodules, repeat FNAB or surgery may be chosen over active surveillance [56].

AUS/FLUS accounts for the majority of indeterminate nodules, yet the actual incidence of malignancy within AUS/FLUS remains uncertain due to the lack of pathologic confirmation in every case

[57]. Research into AUS/FLUS has revealed a broad spectrum of malignant incidence, ranging from 5%–27% in all cases and 6%–48% in surgical cases [58]. In this meta-analysis, the K-TIRADS demonstrated the highest sensitivity (0.95; 95% CrI, 0.85 to 1.00) and specificity (0.75; 95% CrI, 0.13 to 1.00) when intermediate to high risk was categorized as positive. However, the K-TIRADS results could be impacted by an excess of zeros in the two-by-two table, as it had the lowest AUC among all of the US RSSs. For the high-risk category, the EU-TIRADS (0.59; 95% CrI, 0.41 to 0.73) and Kwak TIRADS (0.99; 95% CrI, 0.97 to 1.00) exhibited the highest sensitivity and specificity, respectively. Despite variations among US RSSs, AUS/FLUS could still benefit from US RSSs in determining the need for repeated FNAB, as opposed to diagnostic thyroidectomy [43]. Numerous studies have highlighted the advantages of repeated FNAB in reclassifying an AUS/FLUS result into a category with a more definitive malignancy rate and management strategy [59,60]. Due to insufficient data on FN/SFN and SM, only overall effects were analyzed. Generally, US RSSs were more effective in identifying malignancy in SM (AUC, 0.95; 95% CrI, 0.93 to 0.98) than in AUS/FLUS or FN/SFN when considering the intermediate to high-risk category as positive, likely due to the substantially higher malignancy rate in SM. Ultimately, in the meta-regression, factors such as sample size, the proportion of malignant nodules, and study location were identified as common sources of study heterogeneity.

This meta-analysis had several limitations. First, while category-based comparisons of diagnostic performance are intuitively interpretable, they are inherently limited due to the varying malignancy risks of the categories suggested in the guidelines. Second, most of the included studies had retrospective and single-center designs. Furthermore, despite the use of a Bayesian model to fit estimates and mitigate heterogeneity, substantial between-study heterogeneity persisted, particularly due to the mixed indeterminate components. Third, the diagnosis of both benign and malignant lesions typically relied on surgical pathology, potentially introducing a reference standard bias. Fourth, actual recommendations for FNAB are based on a combination of risk categories and nodule size, a factor not assessed in this study. Finally, insufficient studies were available to conduct quantitative analyses on all of the included US classification systems.

In conclusion, the diagnostic performance of the US RSS in accordance with the representative society guidelines was found to be moderate. This study aims to equip readers and physicians with insights into the performance of each RSS in the context of indeterminate nodules. This information could be instrumental in making decisions about system implementation. Further prospective studies that evaluate all of the most common US RSSs and utilize histology as the standard of reference are necessary.

ORCID: Zhichao Xing: https://orcid.org/0000-0003-0638-4597; Yuxuan Qiu: https://orcid.org/0000-0002-5467-6466; Jingqiang Zhu: https://orcid.org/0000-0002-0794-9169; Anping Su: https://orcid.org/0000-0003-2289-9339; Wenshuang Wu: https://orcid.org/0000-0002-1210-4865

Author Contributions
Conceptualization: Xing Z, Qiu Y, Zhu J, Wu W. Data acquisition: Xing Z, Qiu Y, Wu W. Data analysis or interpretation: Xing Z, Qiu Y, Su A. Drafting of the manuscript: Xing Z, Qiu Y. Critical revision of the manuscript: Xing Z, Qiu Y, Zhu J, Su A, Wu W. Approval of the final version of the manuscript: all authors.

Conflict of Interest
No potential conflict of interest relevant to this article was reported.

Acknowledgments

Supplementary Material
Supplementary Table 1. Two-by-two tables of included studies (https://doi.org/10.14366/usg.23055).

Supplementary Table 2. Deviance information criterion for all analyses (https://doi.org/10.14366/usg.23055).

Supplementary Table 3. Diagnostic performance of each US RSS for AUS/FLUS (LR+, LR−, DOR, and AUC) (https://doi.org/10.14366/usg.23055).

Supplementary Table 4. Diagnostic performance for FN/SFN and SM (LR+, LR−, DOR, and AUC) (https://doi.org/10.14366/usg.23055).

Supplementary Table 5. Meta-regression analysis for each US RSS (https://doi.org/10.14366/usg.23055).

Supplementary Table 6. Diagnostic performance of each US RSS for indeterminate nodules (accuracy) (https://doi.org/10.14366/usg.23055).

Supplementary Fig. 1. Forest plots of sensitivity and specificity of each ultrasound risk stratification system for categorization of intermediate to high risk as positive (https://doi.org/10.14366/usg.23055).

Supplementary Fig. 2. Summary receiver operating characteristic curve of the diagnostic performance of each ultrasound risk stratification system for categorization of intermediate to high risk as positive (https://doi.org/10.14366/usg.23055).

Supplementary Fig. 3. Forest plots of sensitivity and specificity of each ultrasound risk stratification system for categorization of high risk as positive (https://doi.org/10.14366/usg.23055).

Supplementary Fig. 4. Summary receiver operating characteristic curve of the diagnostic performance of each ultrasound risk stratification system for categorization of high risk as positive (https://doi.org/10.14366/usg.23055).

References

1. Russ G, Royer B, Bigorgne C, Rouxel A, Bienvenu-Perrard M, Leenhardt L. Prospective evaluation of thyroid imaging reporting and data system on 4550 nodules with and without elastography. Eur J Endocrinol 2013;168:649-655.

2. Cibas ES, Ali SZ. The Bethesda System for Reporting Thyroid Cytopathology. Thyroid 2009;19:1159-1165.

3. Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda System for Reporting Thyroid Cytopathology: a meta-analysis. Acta Cytol 2012;56:333-339.

4. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid 2016;26:1-133.

5. Alexander EK, Faquin WC, Krane JF. Highlights for the cytology community from the 2015 American Thyroid Association clinical guidelines on the management of thyroid nodules and well-differentiated thyroid cancer. Cancer Cytopathol 2016;124:453-456.

6. Maia FF, Matos PS, Pavin EJ, Zantut-Wittmann DE. Thyroid imaging reporting and data system score combined with Bethesda system for malignancy risk stratification in thyroid nodules with indeterminate results on cytology. Clin Endocrinol (Oxf) 2015;82:439-444.

7. Grani G, Lamartina L, Ascoli V, Bosco D, Nardi F, D'Ambrosio F, et al. Ultrasonography scoring systems can rule out malignancy in cytologically indeterminate thyroid nodules. Endocrine 2017;57:256-261.

8. Migda B, Migda M, Migda MS. A systematic review and meta-analysis of the Kwak TIRADS for the diagnostic assessment of indeterminate thyroid nodules. Clin Radiol 2019;74:123-130.

9. Horvath E, Majlis S, Rossi R, Franco C, Niedmann JP, Castro A, et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. J Clin Endocrinol Metab 2009;94:1748-1751.

10. Sahli ZT, Karipineni F, Hang JF, Canner JK, Mathur A, Prescott JD, et al. The association between the ultrasonography TIRADS classification system and surgical pathology among indeterminate thyroid nodules. Surgery 2019;165:69-74.

11. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM; PRISMA-DTA Group, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. JAMA 2018;319:388-396.

12. Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedus L, et al. American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules--2016 update. Endocr Pract 2016;22:622-639.

13. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): white paper of the ACR TI-RADS Committee. J Am Coll Radiol 2017;14:587-595.

14. Perros P, Boelaert K, Colley S, Evans C, Evans RM, Gerrard Ba G, et al. Guidelines for the management of thyroid cancer. Clin Endocrinol (Oxf) 2014;81 Suppl 1:1-122.

15. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European Thyroid Association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: the EU-TIRADS. Eur Thyroid J 2017;6:225-237.

16. Russ G. Risk stratification of thyroid nodules on ultrasonography with the French TI-RADS: description and reflections. Ultrasonography 2016;35:25-38.

17. Shin JH, Baek JH, Chung J, Ha EJ, Kim JH, Lee YH, et al. Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean Society of Thyroid Radiology consensus statement and recommendations. Korean J Radiol 2016;17:370-395.

18. Kwak JY, Han KH, Yoon JH, Moon HJ, Son EJ, Park SH, et al. Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk. Radiology 2011;260:892-899.

19. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529-536.

20. Guo J, Riebler A, Rue H. Bayesian bivariate meta-analysis of diagnostic test studies with interpretable priors. Stat Med 2017;36:3039-3058.

21. Ahmadi S, Herbst R, Oyekunle T, Jiang X, Strickland K, Roman S, et al. Using the ATA and ACR TI-RADS sonographic classifications as adjunctive predictors of malignancy for indeterminate thyroid nodules. Endocr Pract 2019;25:908-917.

22. Al Dawish M, Alwin Robert A, Al Shehri K, Hawsawi S, Mujammami M, Al Basha IA, et al. Risk stratification of thyroid nodules with Bethesda III category: the experience of a territorial healthcare

hospital. Cureus 2020;12:e8202.

23. Barbosa TL, Mesa Junior CO, Graf H, Cavalvanti T, Trippia MA, da Silveira Ugino RT, et al. ACR TI-RADS and ATA US scores are helpful for the management of thyroid nodules with indeterminate cytology. BMC Endocr Disord 2019;19:112.

24. Baser H, Cakir B, Topaloglu O, Alkan A, Polat SB, Dogan HT, et al. Diagnostic accuracy of Thyroid Imaging Reporting and Data System in the prediction of malignancy in nodules with atypia and follicular lesion of undetermined significance cytologies. Clin Endocrinol (Oxf) 2017;86:584-590.

25. Capezzone M, Cantara S, Di Santo A, Sagnella A, Pilli T, Brilli L, et al. The combination of sonographic features and the seven-gene panel may be useful in the management of thyroid nodules with indeterminate cytology. Front Endocrinol (Lausanne) 2021;12:613727.

26. Celletti I, Fresilli D, De Vito C, Bononi M, Cardaccio S, Cozzolino A, et al. TIRADS, SRE and SWE in INDETERMINATE thyroid nodule characterization: which has better diagnostic performance? Radiol Med 2021;126:1189-1200.

27. Chaigneau E, Russ G, Royer B, Bigorgne C, Bienvenu-Perrard M, Rouxel A, et al. TIRADS score is of limited clinical value for risk stratification of indeterminate cytological results. Eur J Endocrinol 2018;179:13-20.

28. Chirayath SR, Pavithran PV, Abraham N, Nair V, Bhavani N, Kumar H, et al. Prospective study of Bethesda categories III and IV thyroid nodules: outcomes and predictive value of BRAF(V600E) mutation. Indian J Endocrinol Metab 2019;23:278-281.

29. He YP, Xu HX, Zhao CK, Sun LP, Li XL, Yue WW, et al. Cytologically indeterminate thyroid nodules: increased diagnostic performance with combination of US TI-RADS and a new scoring system. Sci Rep 2017;7:6906.

30. Hong MJ, Na DG, Baek JH, Sung JY, Kim JH. 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.

31. Hong HS, Lee JY. 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.

32. Slowinska-Klencka D, Wysocka-Konieczna K, Klencki M, Popowicz B. Diagnostic value of six thyroid imaging reporting and data systems (TIRADS) in cytologically equivocal thyroid nodules. J Clin Med 2020;9:2281.

33. Koh J, Kim EK, Kwak JY, Yoon JH, Moon HJ. Repeat fine-needle aspiration can be performed at 6 months or more after initial atypia of undetermined significance or follicular lesion of undetermined significance results for thyroid nodules 10 mm or larger. Eur Radiol 2016;26:4442-4448.

34. Mao F, Xu HX, Zhao CK, Bo XW, Li XL, Li DD, et al. Thyroid

imaging reporting and data system in assessment of cytological Bethesda Category III thyroid nodules. Clin Hemorheol Microcirc 2017;65:163-173.

35. Marina M, Zatelli MC, Goldoni M, Del Rio P, Corcione L, Martorana D, et al. Combination of ultrasound and molecular testing in malignancy risk estimate of Bethesda category IV thyroid nodules: results from a single-institution prospective study. J Endocrinol Invest 2021;44:2635-2643.

36. Mehta S, Kannan S. Approaching indeterminate thyroid nodules in the absence of molecular markers: "The BETH-TR score". Indian J Endocrinol Metab 2020;24:170-175.

37. Park VY, Kim EK, Kwak JY, Yoon JH, Moon HJ. Malignancy risk and characteristics of thyroid nodules with two consecutive results of atypia of undetermined significance or follicular lesion of undetermined significance on cytology. Eur Radiol 2015;25:2601-2607.

38. Piccardo A, Puntoni M, Dezzana M, Bottoni G, Foppiani L, Marugo A, et al. Indeterminate thyroid nodules: the role of ^{18}F-FDG PET/CT in the "era" of ultrasonography risk stratification systems and new thyroid cytology classifications. Endocrine 2020;69:553-561.

39. Rho M, Kim EK, Moon HJ, Yoon JH, Park VY, Han K, et al. Clinical parameter for deciding the BRAFV600E mutation test in atypia of undetermined significance/follicular lesion of undetermined significance thyroid nodules: US features according to TIRADS. Ultrasound Q 2017;33:284-288.

40. Rocha TG, Rosario PW, Silva AL, Nunes MB, Calsolari MR. Thyroid Imaging Reporting and Data System (TI-RADS) of the American College of Radiology (ACR) for predicting malignancy in thyroid nodules >1 cm with indeterminate cytology. Diagn Cytopathol 2019;47:523-525.

41. Rosario PW, Silva AL, Calsolari MR. The ATA classification and TI-RADS ACR predict not only benignity but also the histology of nonbenign tumors in thyroid nodules with indeterminate cytology. Diagn Cytopathol 2021;49:165-167.

42. Sahli ZT, Sharma AK, Canner JK, Karipineni F, Ali O, Kawamoto S, et al. TIRADS interobserver variability among indeterminate thyroid nodules: a single-institution study. J Ultrasound Med 2019;38:1807-1813.

43. Suh YJ, Choi YJ. Strategy to reduce unnecessary surgeries in thyroid nodules with cytology of Bethesda category III (AUS/FLUS): a retrospective analysis of 667 patients diagnosed by surgery. Endocrine 2020;69:578-586.

44. Sultan R, Levy S, Sulanc E, Honasoge M, Rao SD. Utility of afirma gene expression classifier for evaluation of indeterminate thyroid nodules and correlation with ultrasound risk assessment: single institutional experience. Endocr Pract 2020;26:543-551.

45. Trimboli P, Fulciniti F, Zilioli V, Ceriani L, Giovanella L. Accuracy of international ultrasound risk stratification systems in thyroid lesions cytologically classified as indeterminate. Diagn Cytopathol

2017;45:113-117.

46. Ulisse S, Bosco D, Nardi F, Nesca A, D'Armiento E, Guglielmino V, et al. Thyroid Imaging Reporting and Data System score combined with the new Italian classification for thyroid cytology improves the clinical management of indeterminate nodules. Int J Endocrinol 2017;2017:9692304.

47. Wang MM, Beckett K, Douek M, Masamed R, Patel M, Tseng CH, et al. Diagnostic value of molecular testing in sonographically suspicious thyroid nodules. J Endocr Soc 2020;4:bvaa081.

48. 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.

49. Yoo WS, Ahn HY, Ahn HS, Chung YJ, Kim HS, Cho BY, et al. Malignancy rate of Bethesda category III thyroid nodules according to ultrasound risk stratification system and cytological subtype. Medicine (Baltimore) 2020;99:e18780.

50. Yoon JH, Kwon HJ, Kim EK, Moon HJ, Kwak JY. Subcategorization of atypia of undetermined significance/follicular lesion of undetermined significance (AUS/FLUS): a study applying Thyroid Imaging Reporting and Data System (TIRADS). Clin Endocrinol (Oxf) 2016;85:275-282.

51. Zhang WB, Li JJ, Chen XY, He BL, Shen RH, Liu H, et al. 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.

52. Kim DH, Chung SR, Choi SH, Kim KW. Accuracy of Thyroid Imaging Reporting and Data System category 4 or 5 for diagnosing malignancy: a systematic review and meta-analysis. Eur Radiol 2020;30:5611-5624.

53. Kim DH, Kim SW, Basurrah MA, Lee J, Hwang SH. Diagnostic performance of six ultrasound risk stratification systems for thyroid nodules: a systematic review and network meta-analysis. AJR Am J Roentgenol 2023;220:791-803.

54. Grani G, Lamartina L, Cantisani V, Maranghi M, Lucia P, Durante C. Interobserver agreement of various thyroid imaging reporting and data systems. Endocr Connect 2018;7:1-7.

55. Sych YP, Fadeev VV, Fisenko EP, Kalashnikova M. Reproducibility and interobserver agreement of different Thyroid Imaging and Reporting Data Systems (TIRADS). Eur Thyroid J 2021;10:161-167.

56. 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.

57. Chen JC, Pace SC, Chen BA, Khiyami A, McHenry CR. Yield of repeat fine-needle aspiration biopsy and rate of malignancy in patients with atypia or follicular lesion of undetermined significance: the impact of the Bethesda System for Reporting Thyroid Cytopathology. Surgery 2012;152:1037-1044.

58. Dincer N, Balci S, Yazgan A, Guney G, Ersoy R, Cakir B, et al. Follow-up of atypia and follicular lesions of undetermined significance in thyroid fine needle aspiration cytology. Cytopathology 2013;24:385-390.

59. McElroy MK, Mahooti S, Hasteh F. A single institution experience with the new Bethesda System for Reporting Thyroid Cytopathology: correlation with existing cytologic, clinical, and histological data. Diagn Cytopathol 2014;42:564-569.

60. Nayar R, Ivanovic M. The indeterminate thyroid fine-needle aspiration: experience from an academic center using terminology similar to that proposed in the 2007 National Cancer Institute Thyroid Fine Needle Aspiration State of the Science Conference. Cancer 2009;117:195-202.