



# Systematic Review and Meta-Analysis of American College of Radiology TI-RADS Inter-Reader Reliability for Risk Stratification of Thyroid Nodules

Wei Li<sup>1†</sup>, Yuan Sun<sup>2†</sup>, Haibing Xu<sup>1</sup>, Wenwen Shang<sup>1</sup> and Anding Dong<sup>1\*</sup>

<sup>1</sup> Department of Medical Imaging, Jiangsu Vocational College of Medicine, Yancheng, China, <sup>2</sup> Department of Burn and Plastic Surgery, Affiliate Huaihai Hospital of Xuzhou Medical University, Xuzhou, China

## OPEN ACCESS

### Edited by:

Jan Baptist Vermorken,  
University of Antwerp, Belgium

### Reviewed by:

Christian Albert Koch,  
Fox Chase Cancer Center,  
United States  
Elisa Giannetta,  
Sapienza University of Rome, Italy

### \*Correspondence:

Anding Dong  
11368@jmc.edu.cn

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Head and Neck Cancer,  
a section of the journal  
Frontiers in Oncology

Received: 21 December 2021

Accepted: 11 April 2022

Published: 13 May 2022

### Citation:

Li W, Sun Y, Xu H, Shang W and  
Dong A (2022) Systematic  
Review and Meta-Analysis of  
American College of Radiology  
TI-RADS Inter-Reader Reliability for  
Risk Stratification of Thyroid Nodules.  
Front. Oncol. 12:840516.  
doi: 10.3389/fonc.2022.840516

**Purpose:** To investigate the inter-reader agreement of using the American College of Radiology (ACR) Thyroid Imaging Reporting and Data System (TI-RADS) for risk stratification of thyroid nodules.

**Methods:** A literature search of Web of Science, PubMed, Cochrane Library, EMBASE, and Google Scholar was performed to identify eligible articles published from inception until October 31, 2021. We included studies reporting inter-reader agreement of different radiologists who applied ACR TI-RADS for the classification of thyroid nodules. Quality assessment of the included studies was performed with the Quality Assessment of Diagnostic Accuracy Studies-2 tool and Guidelines for Reporting Reliability and Agreement Studies. The summary estimates of the inter-reader agreement were pooled with the random-effects model, and multiple subgroup analyses and meta-regression were performed to investigate various clinical settings.

**Results:** A total of 13 studies comprising 5,238 nodules were included in the current meta-analysis and systematic review. The pooled inter-reader agreement for overall ACR TI-RADS classification was moderate ( $\kappa = 0.51$ , 95% CI 0.42–0.59). Substantial heterogeneity was presented throughout the studies, and meta-regression analyses suggested that the malignant rate was the significant factor. Regarding the ultrasound (US) features, the best inter-reader agreement was composition ( $\kappa = 0.58$ , 95% CI 0.53–0.63), followed by shape ( $\kappa = 0.57$ , 95% CI 0.41–0.72), echogenicity ( $\kappa = 0.50$ , 95% CI 0.40–0.60), echogenic foci ( $\kappa = 0.44$ , 95% CI 0.36–0.53), and margin ( $\kappa = 0.34$ , 95% CI 0.24–0.44).

**Conclusions:** The ACR TI-RADS demonstrated moderate inter-reader agreement between radiologists for the overall classification. However, the US feature of margin only showed fair inter-reader reliability among different observers.

**Keywords:** thyroid nodule, ultrasonography, reproducibility of results, classification, meta-analysis

## INTRODUCTION

Thyroid nodules are very common in the general population and affect approximately half the population older than 40 years (1–3). Even though most thyroid nodules are benign, 5%–15% are proved to be malignant (4, 5). In recent years, the detection rate of thyroid nodules is increasing with the widespread use of ultrasound (US), and the incidence of thyroid cancer has increased about 5 times in the past 5 decades (6). Currently, the US is accepted as the most effective modality for routine thyroid nodule detection and evaluation and also is used as a guide during fine-needle aspiration biopsy (FNAB) (7–9). However, the risk stratification of thyroid nodules is primarily dependent on radiologists' personal experience and deficiency of reproducibility and agreement, which results in many overdiagnoses and unnecessary FNABs (10).

To address this problem, various guidelines and scoring systems were developed to standardize the risk stratification and management of thyroid nodules, including the American Thyroid Association (ATA) guideline (10, 11), Kwak Thyroid Imaging Reporting and Data System (TI-RADS) (12), American Association of Clinical Endocrinologists (AACE) guidelines (13), and European Union (EU) TI-RADS (14). In 2017, the American College of Radiology (ACR) released the ACR TI-RADS, in which a thyroid nodule is scored 1–3 points according to five US features (composition, echogenicity, shape, margin, and echogenic foci). Afterward, the calculated sum of these points is used to classify a nodule into one of five malignancy risk groups: benign (TR1, 0 points), minimally suspicious for malignancy (TR2, 2 points), mildly suspicious for malignancy (TR3, 3 points), moderately suspicious for malignancy (TR4, 4–6 points), or highly suspicious for malignancy (TR5,  $\geq 7$  points). Since the ACR TI-RADS was proposed, it has been one of the most reported guidelines in recent years (15, 16). In a recently published study, Liu et al. reported that the inter-observer agreement for ATA guidelines was 0.628 and for ACR TI-RADS was 0.748 (17). However, the inter-reader agreement of this scoring system has not been systematically evaluated. Therefore, in this study, we aimed to assess the inter-reader agreement of the ACR TI-RADS for the classification of thyroid nodules.

## METHODS

For this meta-analysis and systematic review, a standardized review and data extraction protocol was used, performed in compliance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (18). The primary outcome was the inter-reader agreement using ACR TI-RADS for risk stratification of thyroid nodules.

### Search Strategy and Selection Criteria

A thorough literature search was performed of PubMed, EMBASE, Cochrane Library, Web of Science, and Google Scholar from inception until October 31, 2021, with no

language restrictions applied. Original research articles reporting inter-reader agreement of ACR TI-RADS were identified, and the terms combined synonyms used for searching were as follows: ([ACR] OR [American College of Radiology]) AND ([TI-RADS] OR [TIRADS] OR [Thyroid Imaging Reporting and Data System]) AND ([thyroid nodule] or [nodule]). An additional search was performed by manually screening the bibliographies of the included articles and reviews. Studies identified by literature search were assessed by two independent reviewers (WL and YS, with 6 and 8 years of experience, respectively, in performing systematic reviews and meta-analyses) for potential inclusion, and disagreements were resolved by consensus *via* discussion with a third reviewer (AD).

### Inclusion and Exclusion Criteria

Studies that satisfy all of the following criteria were included: 1) studies that reported the inter-reader agreement of ACR TI-RADS for risk stratification of thyroid nodules; 2) studies that provided Cohen's kappa ( $\kappa$ ) values or intraclass correlation coefficients (ICCs) and their 95% CIs, or other measurements for assessment of inter-reader agreement; and 3) cytology results from US-guided FNAB or surgical resection pathology results as the reference standard. Studies that met any of the following criteria were excluded: 1) with a small sample size that involved less than 20 participants; 2) studies using other scoring systems rather than the ACR TI-RADS; 3) studies that did not provide detailed data to evaluate the inter-reader agreement; and 4) not original articles such as reviews, letters, guidelines, conference abstracts, or editorials.

### Data Extraction and Quality Assessment

A predefined standardized form was used to extract relevant information as follows: 1) demographic characteristics such as the number of patients and nodules, patient age, and male/female ratio; and 2) study characteristics such as first author, publication year, location, and institution, period of the study conducted, the mean or median size of nodules, number of readers and their experience, whether blinded to final results, reference standard, inter-reader agreement regarding ACR TI-RADS, and US features (composition, echogenicity, shape, margin, and echogenic foci). The quality assessment of included studies was performed according to the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) (19), in which the following items were used to score each study: description of the title and abstract, methods, results, discussion, and auxiliary material. For individual studies, these categories were scored as high quality if it was described in sufficient detail in the article with no potential bias.

### Data Synthesis and Statistical Analysis

The summary estimates of  $\kappa$  or ICC values were calculated with the random-effects model (Sidik–Jonkman method) (20, 21) and categorized as follows: a  $\kappa$  value of  $<0.20$  indicates slight agreement; a  $\kappa$  value between 0.21 and 0.40, fair agreement; a  $\kappa$  value between 0.41 and 0.60, moderate agreement; a  $\kappa$  value between 0.61 and 0.80, substantial agreement; and a  $\kappa$  value of between 0.81 and 1.00, almost perfect agreement. Aside from

overall ACR TI-RADS classification, the  $\kappa$  values of five US features included in this guideline were pooled, namely, composition, echogenicity, shape, margin, and echogenic foci. The forest plots were created to graphically present the results. Multiple subgroup analyses regarding the following variables were performed: 1) for readers with experience of more than 10 years, 2) for readers with experience of less than 10 years or not dedicated in such area, 3) for studies providing details on nodule size  $\geq 1$  cm, and 4) for studies providing details on patients recommended for FNAB. Considering that several studies reported the head-to-head comparisons between the ACR TI-RADS with other scoring systems such as ATA guidelines and the EU TI-RADS, these guidelines were compared in available studies.

Sensitivity analyses were performed to evaluate the contribution of the individual study to the pooled estimates, by excluding each study at a time and recalculating the pooled estimates for the remaining studies. The meta-regression analysis was performed to explore the causes of heterogeneity by adding the following covariates: 1) study design (prospective vs. retrospective), 2) reference standard (FNAB vs. FNAB+thyroid resection pathology), 3) number of readers (readers = 2 vs. readers > 2), 4) number of nodules (<150 vs.  $\geq 150$ ), 5) malignant rate (<20% vs.  $\geq 20\%$ ), and 6) inter-reader agreement (Cohen  $\kappa$  vs. ICC). Heterogeneity throughout studies was determined with the Q statistics and the inconsistency index ( $I^2$ ), as follows: for values between 0% and 40%, unimportant; between 30% and 60%, moderate; between 50% and 90%, substantial; and between 75% and 100%, considerable (22). Funnel plots and the rank test were used for the assessment of any possible publication bias. All analyses were performed with STATA 16.0 (StataCorp, Texas, USA), with  $p$ -values <0.05 considered as statistical significance. Two reviewers (WL and YS) independently conducted the data extraction and quality assessment, and disagreements were resolved by discussion, arbitrated by a third reviewer (AD).

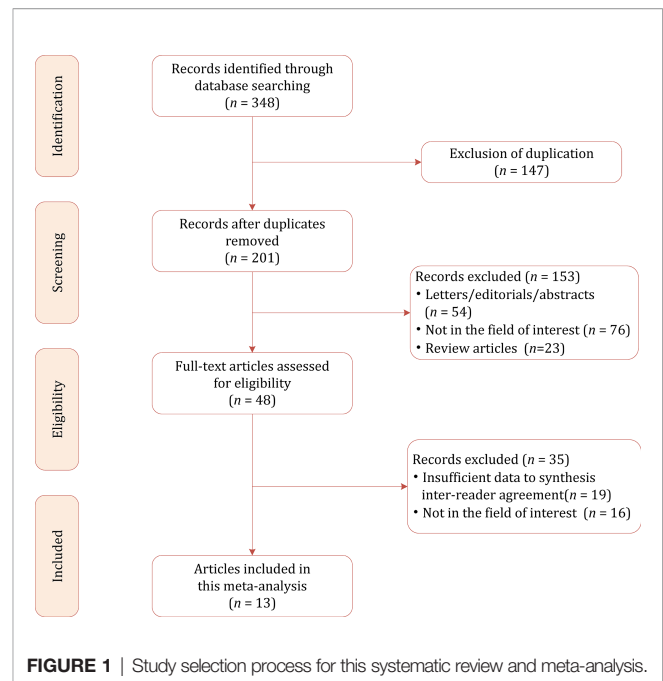
## RESULTS

### Literature Search and Data Extraction

**Figure 1** shows the flowchart of the publication selection process. The literature search identified 348 references initially, of which 147 were excluded due to duplicates. After the titles and abstracts were inspected, 153 articles were excluded. The full-text review was conducted among the remaining 48 potential articles, then 19 articles were excluded due to insufficient data to synthesize inter-reader agreement, and 16 articles were excluded because they were not in the field of interest. Finally, a total of 13 articles comprising 5,238 nodules were included in the current meta-analysis and systematic review (23–35).

### Characteristics of the Included Studies

The detailed demographic and study characteristics are summarized in **Tables 1, 2**. Concerning study design, the



majority of studies were retrospective, and only two studies were prospective. The patient sample ranged from 62 to 1,947, with a nodule number of 74–1,947. In all studies, women were dominant and account for 70%–93% of the total study population. In most studies, patients were adults, whereas in one study, the participants were adolescents (30). In 3 studies, the reported statistics of the inter-reader agreement were ICCs (29, 31, 33); in 1 study, Krippendorff's  $\alpha$  (26); and in the remaining 9 studies, Cohen's kappa. In four studies, the US images were interpreted by 2 radiologists (26, 30, 32, 35), whereas in the remaining 9 studies, the inter-reader agreements were produced by at least 3 radiologists. Most studies reported details on the experience level of readers, which ranged from 1 to 32 years; however, in 1 study, the exact level of readers' experience was not provided but merely described as "experienced" (33). In addition to ACR TI-RADS, there were 3 studies that provided the inter-reader agreement of ATA guidelines and EU TI-RADS. The number of radiologists across studies ranged from 2 to 8. The mean nodule size among included studies ranged from 1.46 to 2.9 cm. Aside from overall ACR TI-RADS classification, 8 studies reported the inter-reader agreement regarding in detail five US features; moreover, 3 studies reported the inter-reader agreement recommended by the radiologists for FNAB (24, 26, 27).

### Quality Assessment

All included studies were scored as high quality according to the criteria evaluated. The majority of studies reported that they included the consecutive patient population, however, which was not reported explicitly in one study (34). In one study, the readers' experience levels were unavailable (33). Further details on the study quality are provided in **Table S1**.

**TABLE 1** | Demographic characteristics of included studies.

First author	Publication year	Location	Patient number	Nodule number	Malignant	M/F	Age (year, mean $\pm$ SD/IQR, range)	Size (cm, mean $\pm$ SD/IQR, range)
Chung et al. (24)	2020	USA	150	150	10	34/116	58 $\pm$ 15	2.0 (1.0–6.1)
Daniels et al. (25)	2020	USA	121	136	24	21/115	55.8 $\pm$ 14.1	2.65 $\pm$ 1.5
Grani et al. (26)	2018	Italy	265	501/554	NA	NA	NA	NA
Huang et al. (35)	2021	China	198	203	58	43/156	42.8 $\pm$ 11.2	0.8 (0.6–1.2)
Itani et al. (28)	2019	USA	131	180	10	31/100	51.1 $\pm$ 14.4	2.58 $\pm$ 1.21
Lim-Dunham et al. (30)	2019	USA	62	74	20	6/56	12.5 (9–17)/16.5 (15–18) <sup>†</sup>	1.90 (1.30–2.80)
Hoang et al. (27)	2018	USA	92	100	15	NA	52 $\pm$ 14	2.8 $\pm$ 1.3/2.2 $\pm$ 1.3
Pandya et al. (31)	2020	USA	1947	1947	193	475/1472	56 $\pm$ 15	1.68 $\pm$ 0.88
Phuttharak et al. (32)	2019	Thailand	94	108	53	7/87	51.6 $\pm$ 13.08	2.12 $\pm$ 1.2
Seifert et al. (34)	2019	Germany	80	80	14	24/56	50.8 $\pm$ 13.1	NA
Basha et al. (23)	2019	Egypt	100	948	136	66/314	45.3 $\pm$ 14.2	2.9 $\pm$ 1.37
Li et al. (29)	2019	China	128	130	73	27/101	47.8 $\pm$ 10.5	1.46 $\pm$ 1.0
Sahli et al. (33)	2019	USA	127	127	26	35/92	52 $\pm$ 14	2.3 (0.5–8.0)

IQR, interquartile range; NA, not available.

<sup>†</sup> Male/female.

## Pooled Inter-Reader Agreement of American College of Radiology Thyroid Imaging Reporting and Data System

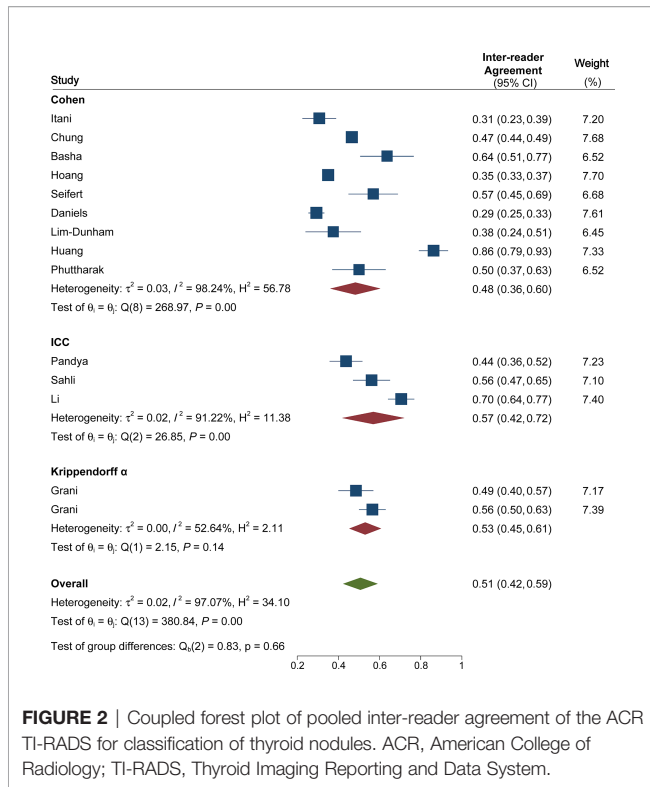
The pooled summary estimates of inter-reader agreement of the ACR TI-RADS are summarized in **Figure 2**. For individual studies, the  $\kappa$  values ranged from 0.29 to 0.86, and the pooled

summary estimates of the  $\kappa$  value were 0.51 (95% CI 0.42–0.59) for the overall ACR TI-RADS classification. For 5 major US features, the highest inter-reader agreement was composition, with pooled  $\kappa$  of 0.58 (95% CI 0.53–0.63), followed by shape ( $\kappa$  = 0.57, 95% CI 0.41–0.72). The poorest inter-reader agreement was margin ( $\kappa$  = 0.34, 95% CI 0.24–0.44). For the other 2 features, echogenicity and echogenic foci, the pooled  $\kappa$  values were 0.50

**TABLE 2** | Study characteristics of included studies.

First author	Reference standard	Study design	Period	Blinding	Reader number	Readers' experience (year)	Agreement statistic	Agreement value
Chung et al. (24)	US-guided FNAB	Retrospective	Jan. 2017–Apr. 2018	Yes	6	3 readers $\geq$ 22 3 readers $\geq$ 1	Cohen $\kappa$	0.31
Daniels et al. (25)	FNAB+pathology	Retrospective	Jan. 2017–Aug. 2018	Yes	4	11 (5–16)	Cohen $\kappa$	0.49
Grani et al. (26)	US-guided FNAB	Retrospective	Jan. 2006–Jan. 2008	Yes	2	$\geq$ 6	Krippendorff's $\alpha$	0.57
Huang et al. (35)	FNAB+pathology	Retrospective	Jan. 2015–Dec. 2018	Yes	2	17/22	Cohen $\kappa$	0.47
Itani et al. (28)	FNAB+pathology	Retrospective	Jan. 2014–Jun. 2014	Yes	4	5/15/20/20	Cohen $\kappa$	0.64
Lim-Dunham et al. (30)	US-guided FNAB	Retrospective	1996–2017	Yes	2	$\geq$ 10	Cohen $\kappa$	0.29
Hoang et al. (27)	FNAB+pathology	Retrospective	Apr. 2009–May 2010	Yes	8	3–32	Cohen $\kappa$	0.38
Pandya et al. (31)	US-guided FNAB	Retrospective	Oct. 2009–Feb. 2016	Yes	5	4–32	ICC	0.86
Phuttharak et al. (32)	US-guided FNAB	Retrospective	Jan. 2015–Oct. 2017	Yes	2	1 reader $\geq$ 2 1 reader $\geq$ 10	Cohen $\kappa$	0.71
Seifert et al. (34)	US-guided FNAB	Prospective	NA	Yes	4	1 reader $\geq$ 3 3 readers $\geq$ 5	Cohen $\kappa$	0.56
Basha et al. (23)	US-guided FNAB	Prospective	May 2017–Dec. 2018	Yes	3	$\geq$ 15	Cohen $\kappa$	0.44
Li et al. (29)	FNAB+pathology	Retrospective	Dec. 2016–Mar. 2018	Yes	4	$\geq$ 5	ICC	0.51
Sahli et al. (33)	US-guided FNAB	Retrospective	Feb. 2012–Sep. 2016	Yes	3	Experienced	ICC	0.57

FNAB, fine-needle aspiration biopsy; ICC, intraclass correlation coefficient; NA, not available; US, ultrasound.

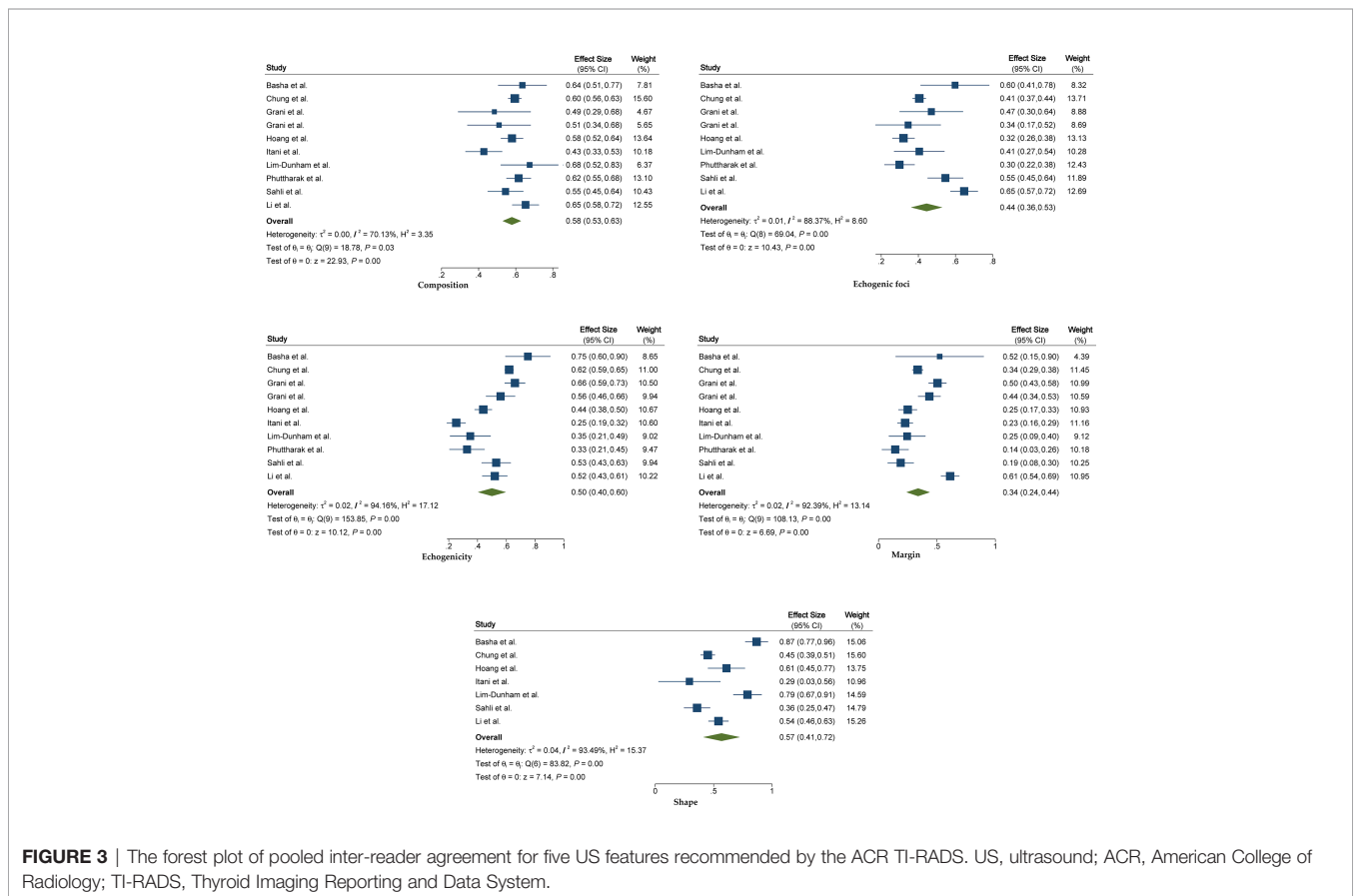


**FIGURE 2 |** Coupled forest plot of pooled inter-reader agreement of the ACR TI-RADS for classification of thyroid nodules. ACR, American College of Radiology; TI-RADS, Thyroid Imaging Reporting and Data System.

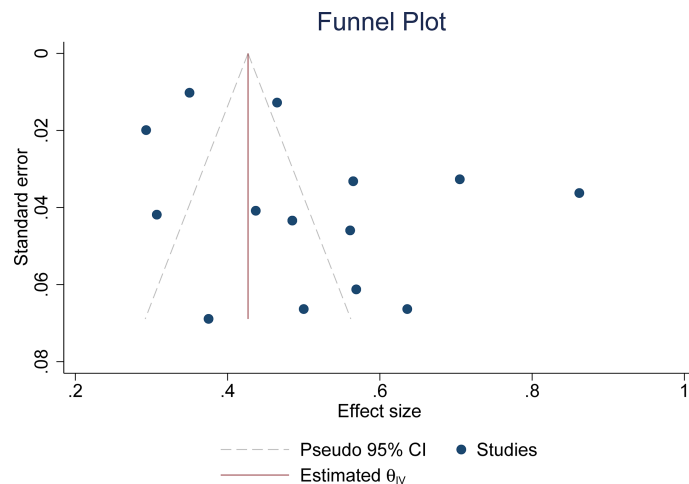
(95% CI 0.40–0.60) and 0.44 (95% CI 0.36–0.53), respectively (Figure 3). We did not find substantial difference between the ACR TI-RADS with EU TI-RADS (0.54, 95% CI 0.49–0.58 vs. 0.54, 95% CI 0.32–0.77,  $p = 0.95$ ) and ATA guidelines (0.51, 95% CI 0.45–0.58 vs. 0.54, 95% CI 0.45–0.64,  $p = 0.60$ ). There was no significant publication bias regarding the ACR TI-RADS classification among included studies ( $p = 0.47$ , Figure 4). Likewise, there was no significant publication bias regarding the five US features, with  $p$ -values of 0.43–0.97.

### Subgroup Analyses and Meta-Regression Analysis

For experienced readers, the pooled  $\kappa$  value based on 3 studies was 0.46 (95% CI 0.30–0.62), whereas for inexperienced readers, the inter-reader agreement was higher, with pooled  $\kappa$  value of 0.54 (95% CI 0.32–0.76). However, the difference did not reach statistical significance. For 3 studies reporting inter-reader agreement for thyroid nodules  $\geq 1$  cm, the pooled  $\kappa$  value of 0.53 (95% CI 0.44–0.63) was suggested to be higher than the overall  $\kappa$  value. The pooled summary estimates of inter-reader agreement in sensitivity analysis were similar to those of all studies, with substantial heterogeneity across included studies (Table S2). Meta-regression analysis was performed to investigate the cause of heterogeneity; among the various potential factors, we found that only the malignant rate ( $<0.2$  vs.  $\geq 0.2$ ) was significantly associated with the



**FIGURE 3 |** The forest plot of pooled inter-reader agreement for five US features recommended by the ACR TI-RADS. US, ultrasound; ACR, American College of Radiology; TI-RADS, Thyroid Imaging Reporting and Data System.



**FIGURE 4** | The funnel plot for testing publication bias.

heterogeneity ( $p = 0.03$ ). All of the other covariates were not substantial factors, with  $p$ -values ranging from 0.08 to 0.96; the details are presented in **Table 3**.

## DISCUSSION

In this study, we systematically assessed the inter-reader agreement for the ACR TI-RADS for the classification of thyroid nodules. Based on 13 studies, the pooled  $\kappa$  value of 0.51 (95% CI 0.42–0.59) revealed that the ACR TI-RADS has moderate inter-reader agreement among the varied experience of

radiologists. Aside from overall ACR TI-RADS classification, most studies reported the inter-reader agreement regarding the detailed five US features recommended by the guideline. Of them, the highest was composition ( $\kappa = 0.58$ ), whereas the lowest was margin ( $\kappa = 0.34$ ). For other features, the pooled  $\kappa$  values ranging from 0.44 to 0.57 indicated moderate inter-reader agreement. In light of several studies that provided a head-to-head comparison between ACR TI-RADS and EU TI-RADS as well as ATA guidelines, we made a direct comparison in available studies, and the results suggested that there was no substantial difference between these 3 guidelines ( $p = 0.95$  and  $p = 0.60$ , respectively). However, both comparisons were based on 3

**TABLE 3** | Meta-regression analysis.

Item	Value	Pooled inter-reader agreement	$p$
Study design	Prospective	0.60 (95% CI 0.51–0.69)	0.08
	Retrospective	0.49 (95% CI 0.41–0.58)	
Size	$\geq 1$	0.54 (95% CI 0.37–0.70)	0.71
	All	0.50 (95% CI 0.40–0.60)	
Reference	FNAB	0.51 (95% CI 0.46–0.55)	0.96
	FNAB+pathology	0.50 (95% CI 0.32–0.68)	
Readers	=2	0.56 (95% CI 0.39–0.73)	0.36
	>2	0.47 (95% CI 0.39–0.55)	
Patient number	<150	0.46 (95% CI 0.36–0.56)	0.15
	$\geq 150$	0.57 (95% CI 0.45–0.70)	
Malignant rate	<0.2	0.42 (95% CI 0.35–0.49)	0.03
	>0.2	0.61 (95% CI 0.45–0.76)	
Nodule number	$\leq 150$	0.47 (95% CI 0.39–0.56)	0.43
	>150	0.55 (95% CI 0.39–0.71)	
Scoring system	ACR vs. EU	0.54 (95% CI 0.49–0.58)	0.95
	ACR vs. ATA	0.54 (95% CI 0.32–0.77)	
	ATA vs. EU	0.51 (95% CI 0.45–0.58)	
Agreement statistic	Cohen	0.54 (95% CI 0.45–0.64)	0.60
	ICC	0.49 (95% CI 0.41–0.57)	
	ICC	0.57 (95% CI 0.41–0.73)	

ACR, American College of Radiology; ATA, American Thyroid Association guidelines; EU, European Union; FNAB, fine-needle aspiration biopsy; ICC, intraclass correlation coefficient; TI-RADS, Thyroid Imaging Reporting and Data System.

studies; therefore, the results should be interpreted with caution. In a previous meta-analysis by Liu et al., which included 7 studies, the pooled inter-reader agreement was slightly higher, with 0.54 (95% CI 0.49–0.5). However, this only provided an indirect comparison; moreover, only 3 studies included in their meta-analysis involved the ACR TI-RADS (36).

As substantial heterogeneity throughout studies was observed, we performed meta-regression analysis to look into the sources. Among the various factors, we found that only the malignant rate ( $<0.2$  vs.  $\geq 0.2$ ) was significantly associated with heterogeneity ( $p = 0.03$ ), whereas all other covariates were not significant factors that contributed to the heterogeneity. Nevertheless, we found a substantial difference in inter-reader agreement regarding study design (0.60 for prospective vs. 0.49 for retrospective), even though it was not found to be statistically significant ( $p = 0.08$ ). In the current study, multiple subgroup analyses were performed to account for several clinical settings. We noted that there was no significant difference between experienced readers and inexperienced readers ( $p = 0.55$ ); moreover, our analysis showed that the inter-reader agreement in inexperienced readers was even higher than in experienced readers (0.53 vs. 0.45). One explanation may be that those with less experience were more likely to strictly use the ACR TI-RADS, while those observers with more experience would be influenced by their prior experience (24). In the study by Daniels et al., they observed that most experienced readers had the lowest intra-reader agreement (25). To address this problem, Seifert et al. and Grani et al. conducted training and demonstrated that the reproducibility and inter-reader agreement were improved on overall ACR TI-RADS classification (24, 26).

Since Kwak et al. proposed the first TI-RADS in 2005 (12), several versions of scoring systems based on this guideline were developed, including the EU, ACR, and Korean TI-RADS (14, 16, 37). Different from other classification systems based on patterns, the ACR TI-RADS is a point-based risk stratification system in which nodules are scored according to five US features. According to this guideline, the more features that a nodule possesses, the more points will accumulate, and the total points determine the nodule's final ACR TI-RADS classification. We performed direct comparison in studies providing a head-to-head comparison between ACR TI-RADS with ATA guidelines and EU TI-RADS, and the results demonstrated that there was no significant difference among these guidelines even though these comparisons were based on a few studies. Aside from inter-reader agreement, previous studies revealed no significant difference in diagnostic performance between these guidelines, irrespective of sensitivity and specificity (38).

In view of most included studies reporting details on US features of composition, echogenicity, shape, margins, and echogenic foci that are recommended by the ACR TI-RADS, we thus investigated the inter-reader agreement regarding these US features. According to our analysis, the poorest inter-reader agreements were margin and echogenic foci, which had only fair ( $\kappa = 0.34$ ) and moderate ( $\kappa = 0.44$ ) reproducibility, respectively. One possible explanation was that a suspicious margin or

echogenic foci may have been apparent on only a few of the static images or only could be observed in the video clip. By comparison, composition and echogenicity had higher reproducibility, as findings present throughout the nodule and are likely to be seen on more images, according to our analysis, which had a nearly substantial agreement, with pooled  $\kappa$  values of 0.57 and 0.58.

## Limitations

There are some limitations to our study. First, most included studies were retrospective in study design, leading to a high risk of bias for the patient selection domain. Nevertheless, because of insufficient data, it is unfeasible to pool the summary estimates from merely two prospective studies. Second, substantial heterogeneity existed among included studies, which affected the general applicability of this systematic review. We performed multiple subgroup analyses and meta-regression, and the results indicated that the malignancy rate was the only significant factor that contributed to the heterogeneity. However, these analyses only explained part of the heterogeneity, and these analyses were based on only a few studies; thus, the results should be interpreted cautiously.

## CONCLUSION

The ACR TI-RADS demonstrated moderate inter-reader agreement between radiologists for the overall classification. However, the US feature of margin recommended by this guideline only showed fair inter-reader reliability among different observers.

## DATA AVAILABILITY STATEMENT

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

## AUTHOR CONTRIBUTIONS

Guarantor of the article: AD. Conception and design: WL and YS. Collection and assembly of data: HX, WS. Data analysis and interpretation: WL and HX. All authors contributed to the article and approved the submitted version.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2022.840516/full#supplementary-material>

## REFERENCES

- Remonti LR, Kramer CK, Leitão CB, Pinto LCF, Gross JL. Thyroid Ultrasound Features and Risk of Carcinoma: A Systematic Review and Meta-Analysis of Observational Studies. *Thyroid* (2015) 25:538–50. doi: 10.1089/thy.2014.0353
- Hegedüs L. The Thyroid Nodule. *N Engl J Med* (2004) 351:1764–71. doi: 10.1056/NEJMcp031436
- Lim H, Devesa SS, Sosa JA, Check D, Kitahara CM. Trends in Thyroid Cancer Incidence and Mortality in the United States, 1974–2013. *JAMA* (2017) 317:1338–48. doi: 10.1001/jama.2017.2719
- Mandel SJ. Diagnostic Use of Ultrasonography in Patients With Nodular Thyroid Disease. *Endocr Pract* (2004) 10:246–52. doi: 10.4158/EP.10.3.246
- Papini E, Guglielmi R, Bianchini A, Crescenzi A, Taccogna S, Nardi F, et al. Risk of Malignancy in Nonpalpable Thyroid Nodules: Predictive Value of Ultrasound and Color-Doppler Features. *J Clin Endocrinol Metab* (2002) 87:1941–6. doi: 10.1210/jcem.87.5.8504
- Brito JP, Morris JC, Montori VM. Thyroid Cancer: Zealous Imaging Has Increased Detection and Treatment of Low Risk Tumours. *BMJ* (2013) 347: f4706. doi: 10.1136/bmj.f4706
- Essenmacher AC, Joyce PH, Kao SC, Epelman M, Pesce LM, D'Alessandro MP, et al. Sonographic Evaluation of Pediatric Thyroid Nodules. *RadioGraphics* (2017) 37:1731–52. doi: 10.1148/rg.2017170059
- Cantisani V, Maceroni P, D'Andrea V, Patrizi G, Di Segni M, De Vito C, et al. Strain Ratio Ultrasound Elastography Increases the Accuracy of Colour-Doppler Ultrasound in the Evaluation of Thy-3 Nodules. A Bi-Centre University Experience. *Eur Radiol* (2016) 26:1441–9. doi: 10.1007/s00330-015-3956-0
- Isidori AM, Cantisani V, Giannetta E, Diacinti D, David E, Forte V, et al. Multiparametric Ultrasonography and Ultrasound Elastography in the Differentiation of Parathyroid Lesions From Ectopic Thyroid Lesions or Lymphadenopathies. *Endocrine* (2017) 57:335–43. doi: 10.1007/s12020-016-1116-1
- Cooper DS, Doherty GM, Haugen BR, Hauger BR, Kloos RT, Lee SL, et al. Revised American Thyroid Association Management Guidelines for Patients With Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid* (2009) 19:1167–214. doi: 10.1089/thy.2009.0110
- Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. American Thyroid Association Management Guidelines for Adult Patients With Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid* (2016) 26:1–133. doi: 10.1089/thy.2015.0020
- 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) 260:892–9. doi: 10.1148/radiol.11110206
- Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedüs L, et al. AACE/ACE/AME Task Force on Thyroid Nodules. American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi Medical Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules—2016 Update. *Endocr Pract* (2016) 22:622–39. doi: 10.4158/EP161208.GL
- 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) 6:225–37. doi: 10.1159/000478927
- Grant EG, Tessler FN, Hoang JK, Langer JE, Beland MD, Berland LL, et al. Thyroid Ultrasound Reporting Lexicon: White Paper of the ACR Thyroid Imaging, Reporting and Data System (TIRADS) Committee. *J Am Coll Radiol JACR* (2015) 12:1272–9. doi: 10.1016/j.jacr.2015.07.011
- 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 JACR* (2017) 14:587–95. doi: 10.1016/j.jacr.2017.01.046
- Liu J, Guo Y, Xiao J, Chen L, Liang Z. Comparison of the Efficacy and Safety of the American Thyroid Association Guidelines and American College of Radiology TI-RADS. *Endocr Pract* (2021) 27:661–7. doi: 10.1016/j.eprac.2020.11.013
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Healthcare Interventions: Explanation and Elaboration. *Epidemiol Biostat Public Health* (2009) 6:e1–e34. doi: 10.1016/j.jclinepi.2009.06.006
- Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) Were Proposed. *J Clin Epidemiol* (2011) 64:96–106. doi: 10.1016/j.jclinepi.2010.03.002
- Sidik K, Jonkman JN. A Simple Confidence Interval for Meta-Analysis. *Stat Med* (2002) 21:3153–9. doi: 10.1002/sim.1262
- Sidik K, Jonkman JN. On Constructing Confidence Intervals for a Standardized Mean Difference in Meta-Analysis. *Commun Stat - Simul Comput* (2003) 32:1191–203. doi: 10.1081/SAC-120023885
- Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's Tool for Assessing Risk of Bias in Randomised Trials. *BMJ* (2011) 343:889–93. doi: 10.1136/bmj.d5928
- Basha MAA, Alnaggar AA, Refaat R, El-Maghraby AM, Refaat MM, Elhamed MEA, et al. The Validity and Reproducibility of the Thyroid Imaging Reporting and Data System (TI-RADS) in Categorization of Thyroid Nodules: Multicentre Prospective Study. *Eur J Radiol* (2019) 117:184–92. doi: 10.1016/j.ejrad.2019.06.015
- Chung R, Rosenkrantz AB, Bennett GL, Dane B, Jacobs JE, Slywotzky C, et al. Interreader Concordance of the TI-RADS: Impact of Radiologist Experience. *Am J Roentgenol* (2020) 214:1152–7. doi: 10.2214/AJR.19.21913
- Daniels KE, Xu J, Liu J-B, Chen X, Huang K, Patel J, et al. Diagnostic Value of TI-RADS Classification System and Next Generation Genetic Sequencing in Indeterminate Thyroid Nodules. *Acad Radiol* (2021) 28:1685–91. doi: 10.1016/j.acra.2020.07.037
- 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. doi: 10.1530/EC-17-0336
- Hoang JK, Middleton WD, Farjat AE, Teefey SA, Abinanti N, Boschini FJ, et al. Interobserver Variability of Sonographic Features Used in the American College of Radiology Thyroid Imaging Reporting and Data System. *Am J Roentgenol* (2018) 211:162–7. doi: 10.2214/AJR.17.19192
- Itani M, Assaker R, Moshiri M, Dubinsky TJ, Dighe MK. Inter-Observer Variability in the American College of Radiology Thyroid Imaging Reporting and Data System: In-Depth Analysis and Areas for Improvement. *Ultrasound Med Biol* (2019) 45:461–70. doi: 10.1016/j.ultrasmedbio.2018.09.026
- Li X, Hou X-J, Du L-Y, Wu J-Q, Wang L, Wang H, et al. Virtual Touch Tissue Imaging and Quantification (VTIQ) Combined With the American College of Radiology Thyroid Imaging Reporting and Data System (ACR TI-RADS) for Malignancy Risk Stratification of Thyroid Nodules. *Clin Hemorheol Microcirc* (2019) 72:279–91. doi: 10.3233/CH-180477
- Lim-Dunham JE, Toslak IE, Reiter MP, Martin B. Assessment of the American College of Radiology Thyroid Imaging Reporting and Data System for Thyroid Nodule Malignancy Risk Stratification in a Pediatric Population. *Am J Roentgenol* (2018) 212:188–94. doi: 10.2214/AJR.18.20099
- Pandya A, Caoili EM, Jawad-Makki F, Wasnik AP, Shankar PR, Bude R, et al. Retrospective Cohort Study of 1947 Thyroid Nodules: A Comparison of the 2017 American College of Radiology TI-RADS and the 2015 American Thyroid Association Classifications. *Am J Roentgenol* (2020) 214:900–6. doi: 10.2214/AJR.19.21904
- 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 APJCP* (2019) 20:1283–8. doi: 10.31557/APJCP.2019.20.4.1283
- 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) 38:1807–13. doi: 10.1002/jum.14870
- Seifert P, Görges R, Zimny M, Kreissl MC, Schenke S. Interobserver Agreement and Efficacy of Consensus Reading in Kwak-, EU-, and ACR-Thyroid Imaging Reporting and Data Systems and ATA Guidelines for the Ultrasound Risk Stratification of Thyroid Nodules. *Endocrine* (2020) 67:143–54. doi: 10.1007/s12020-019-02134-1



35. Huang Y, Hong Y, Xu W, Song K, Huang P. Contrast-Enhanced Ultrasound Improves the Accuracy of the ACR TI-RADS in the Diagnosis of Thyroid Nodules Located in the Isthmus. *Ultraschall Med - Eur J Ultrasound* (2021). doi: 10.1055/a-1543-6033
36. Liu H, Ma A-L, Zhou Y-S, Yang D-H, Ruan J-L, Liu X-D, et al. Variability in the Interpretation of Grey-Scale Ultrasound Features in Assessing Thyroid Nodules: A Systematic Review and Meta-Analysis. *Eur J Radiol* (2020) 129:109050. doi: 10.1016/j.ejrad.2020.109050
37. Shin JH, Baek JH, Chung J, Ha EJ, Kim J-H, Lee YH, et al. Ultrasonography Diagnosis and Imaging-Based Management of Thyroid Nodules: Revised Korean Society of Thyroid Radiology Consensus Statement and Recommendations. *Korean J Radiol* (2016) 17:370–95. doi: 10.3348/kjr.2016.17.3.370
38. Zhang Q, Ma J, Sun W, Zhang L. Comparison of Diagnostic Performance Between the American College of Radiology Thyroid Imaging Reporting and Data System and American Thyroid Association Guidelines: A Systematic Review. *Endocr Pract* (2020) 26:552–63. doi: 10.4158/EP-2019-0237

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

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

Copyright © 2022 Li, Sun, Xu, Shang and Dong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.