Comparison of Diagnostic Performance of Five Different Ultrasound TI-RADS Classification Guidelines for Thyroid Nodules

Ruoning Yang[1,2†], Xiuhe Zou[3†], Hao Zeng[1], Yunuo Zhao[1] and Xuelei Ma[1*]

[1] Department of Biotherapy, State Key Laboratory of Biotherapy, West China Hospital, Cancer Center, Sichuan University, Chengdu, China, [2] West China Hospital, West China School of Medicine, Sichuan University, Chengdu, China, [3] Department of Thyroid Surgery, West China Hospital of Sichuan University, Chengdu, China

Objectives: We aimed to evaluate and compare the diagnostic performance of five ultrasound thyroid imaging reporting and data system (TI-RADS) classification guidelines for thyroid nodules through a review and meta-analysis.

Methods: We searched for relevant studies before February 2020 in PubMed. Then we pooled the sensitivity, specificity, likelihood ratios, diagnostic odds ratios, and area under the summary receiver operating characteristic curves. And the diagnostic odds ratios were used to compare the performance.

Results: We totally included 19 studies with 4,696 lesions in this research. The pooled sensitivity of American College of Radiology (ACR) guidelines, American Thyroid Association (ATA) guidelines, TI-RADS proposed by Kwak (Kwak TI-RADS), Korean Thyroid Association/Korean Society of Thyroid Radiology (KTA/KSThR) guidelines for malignancy risk and European Thyroid Association (ETA) guidelines is between 0.84 and 0.94. The pooled specificity is 0.68, 0.44, 0.62, 0.47, and 0.61, respectively. And the RDOR is 1.57 (ACR *vs* ATA), 1.37 (ACR *vs* ETA), 1.80 (ACR *vs* Kawk), 1.74 (ARC *vs* KTA).

Conclusions: The results suggest that five classification guidelines are all effective methods for differential diagnosis of benign and malignant thyroid nodules and ACR guideline is a better choice.

Keywords: thyroid nodule, meta-analysis, TI-RADS, ultrasound, malignancy, diagnostic performance

INTRODUCTION

Thyroid nodules are easily found in the general population, especially in women (1), and about 10% of patients with thyroid nodules are at risk of malignancy, and the percentage keeps going up (2, 3). Malignant nodules and benign nodules are treated in completely different ways. It's still a big challenge for clinicians to rule out malignancy of the thyroid nodules. At present, ultrasound is a primary, cheap, noninvasive, fast, and valuable tool to identify the thyroid nodules. For suspected thyroid nodules, a surgery or fine-needle aspiration cytology (FNAC) is recommended (4). Benign and malignant nodules have some similar ultrasound features from modulation to size. The ultrasound diagnosis varies with the

experience of radiologists, and operators, image acquisition and interpretation are subjective which can easily lead to misdiagnosis or overtreatment (5).

To conduct an objective detection, the thyroid imaging reporting and data system (TI-RADS) was proposed, which is used to classify thyroid nodules and recommend further treatment (6). Nowadays, there are five common classification systems used in clinic. Among the guidelines, the American College of Radiology (ACR) guidelines, the Korean Thyroid Association/Korean Society of Thyroid Radiology (KTA/KSThR) guidelines, and the European Thyroid Association (ETA) guidelines are recommended by the radiological association, and the American Thyroid Association (ATA) guidelines are in clinical guidelines (1, 7–9).

Although these five guidelines prove to be effective in managing thyroid nodules, there are no guidelines based on a lot of reliable data to prove which is the best (10). And many clinical trials in progress are used to compare their effectiveness, but these results are biased. The primary purpose of this research is to compare the diagnostic effectivity of the five guidelines for thyroid nodules to address the lack of consistency and avoid wasting of medical resources.

METHOD

Literature Search Strategy

We followed the guidelines for the systematic review and meta-analysis of diagnostic studies. Then we too retrieved PubMed for related studies with English language only before February 2020, using the terms as follows: "sensitivity", "specificity", "TI-RADS (or thyroid imaging reporting and data system)", "ACR (or The American Thyroid Association)", "ATA (or American Thyroid Association)", "Kwak (or TI-RADS proposed by Kwak)", "ETA (or EU TI-RADS)", "KTA (or Korean Thyroid Association/Korean Society of Thyroid Radiology)". Two reviewers (RN Yang and YN Zhao) independently reviewed the articles in accordance with the inclusion and exclusion criteria. Disagreements were adjusted by consensus (XL Ma).

Inclusion and Exclusion Criteria

Studies with following inclusion criteria were included: (a) There is enough general information in the article. (b) One or more guidelines are used to evaluate the ultrasound features of thyroid nodules. (c) The study has definite diagnostic criteria. (d) There is sufficient data in the article, whether it is data that can be found directly in the article (sensitivity, specificity, and PPV) or data that can be calculated based on the article [positives (TP), true negatives (TN), false positives (FP), and false negatives (FN)] to fill the diagnostic 2×2 table (FN, FP, TP, and TN). And the exclusion criterion is that data in the article is not enough or the grading system is not designed to evaluate ultrasound features. Finally, a total of 19 articles are included.

Data Extraction

Two reviewers (RN Yang and YN Zhao) picked up some main characters from the studies as following: author, year, country,

number of patients, number of nodules, mean age, involved guideline, gold standard, malignant lesions, and benign lesions. And we obtained the four numbers of TP, TN, FP, and FN for each guideline in different studies by two ways: (1) We got the data from the article directly. (2) Based on the data (sensitivity, specificity, PPV, and NPV) obtained from the articles, we finished the diagnostic 2×2 table. CAL software was use here (11).

Statistical Analysis

On the bases of TP, TN, FP, and FN, we computed the pooled sensitivity, specificity, positive and negative likelihood ratios (PLR and NLR), and diagnostic odds ratio (AUC), with 95% confidence intervals (CI), using the Meta-Disc version 1.4 statistical software (12).

Additionally, using the Meta-Disc version 1.4 statistical software (12), we examined the relationship between sensitivity and specificity by constructing the summary receiver operator characteristic (SROC) curves (13).

At last, we made a head-to-head comparison using R 3.5.1 to calculate the relative diagnostic odds ratio (RDOR) with 95% CI. According to the RDOR, we compared the diagnostic performance among the five guidelines. At comparison, classified into A and B, two guidelines were involved. In A vs B, when the value is greater than 1, A has higher performance. If the value is smaller than 1, B has greater performance. When the value is greater, the performance is better. For all studies, the inconsistency index (I^2) and χ^2 test were used to assess heterogeneity, and it was considered high heterogeneity if the I^2 value was higher than 50% (14). A random-effect model was chosen in this research (15).

Quality of Studies and Publication Bias

We used Quality Assessment of Diagnostic Accuracy Studies version 2 (QUADAS-2) performed in Review Manager 5.2 to assess the quality of the studies included in this analysis. The method mainly evaluated the articles from four domains: (a) patient selection, (b) index test, (c) reference standard, and (d) flow and timing (16). Each domain is rated as three risks (low, high, and unclear). Publication bias was evaluated by the funnel plot asymmetry test using Stata version 11.0 software.

RESULTS

Literature Research and Study Characteristics

At first, we searched 200 articles by reading their abstract, and 166 articles didn't agree with the inclusion criterion. Then we reviewed the rest of the articles further, and 11 articles didn't have enough data to finish the 2×2 table. Another five articles were not related to ultrasound features. Therefore, 19 articles (17–35) were included in this study. The process of including articles is in **Figure 1**.

The studies were published from 2015 to 2020.The number of patients is from 92 to 4,585, and the number of nodules in the

included articles varies from 100 to 4,696, which means some patients have more than one nodule. All data were calculated based on the number of nodules. All thyroid nodules were diagnosed of malignancy through postoperative pathological results or the pathology results of FNAC. We totally included 19 articles, 12 of which involved ACR TI-RADS. 10 articles involved ACR guidelines, and Kwak TI-RADS was mentioned in six articles. The data of the KTA guideline and EU TI-RADS were obtained from four articles respectively.

The above characteristics were shown in **Table 1**.

Diagnostic Accuracy

After pooling all the data of the 19 studies together, we got the final data. The pooled sensitivity of ACR guidelines, ATA guidelines, Kwak TI-RADS, KTA guidelines for malignancy risk, and ETA guidelines is between 0.84 and 0.94. The pooled specificity is 0.68, 0.44, 0.62, 0.47, and 0.61, respectively. We also build SROC curves showing the area under the curve (AUC) with 0.8553, 0.9101, 0.8976, 0.9022, and 0.8810, respectively, for ACR, Kwak TI-RADS, ATA, KTA, and EU TI-RADS guideline groups, on behalf of the accuracy. All pooled sensitivity, specificity, PLR, NLR, DOR, and the AUC values for all the reference standards are shown in detail in **Table 2**. As for RDOR, we found a high result when ACR was compared with other guidelines. The specific results are listed in **Table 3**.

Quality Assessment

The results of the quality assessment are outlined in **Figure 2**. In conclusion, the quality of the studies was satisfactory.

## Assessment of Publication Bias

There is no clear publication bias for DOR of the five guidelines.

DISCUSSION

TI-RADS classification guidelines classify the thyroid nodules according to the imaging characteristics under ultrasound, including the size, number, calcification, boundary, echoic pattern, aspect ratio, and internal structure. The guidelines are aimed to help determine which thyroid nodules require FNAC to reduce overdiagnosis or missed-diagnosis. The reduction of unnecessary FNAC can prevent the waste of economy and the physical pain of patients. It can also guide further treatment and estimate the risk of recurrence. However, the recommended size thresholds for FNAC are different in different guidelines. At present, there are many studies about the diagnostic efficacy of the five guidelines, but the results vary. These differences between studies may be due in part to differences among observers and study populations, especially in retrospective studies. In this research, we included 19 studies to analyze the diagnostic efficacy of the five diagnostic criteria.

Our meta-analysis systematically estimated the diagnostic efficacy of five different ultrasound classification guidelines in detecting malignancy risk. The pooled sensitivity of the ACR TI-RADS, ATA guidelines, Kwak TI-RADS, KTA guidelines, and ETA is between 0.84 and 0.94. The pooled specificity is 0.68, 0.44, 0.62, 0.47, and 0.61, respectively. The AUC which can represent the diagnostic performance of the ACR TI-RADS, ATA guidelines, Kwak TI-RADS KTA guideline, and ETA is 0.8553, 0.8976, 0.9101, 0.9022, and 0.8810. In theory, AUC above 0.8 is diagnostic (36). The results of our research suggested that all the five guidelines have property. Besides, ACR guidelines showed the best diagnostic performance in the head to head comparison.

Our results were similar with a previous meta-analysis published in 2019 (37). But that article just included 12 studies with 18,750 thyroid nodules, and the data it included was used to
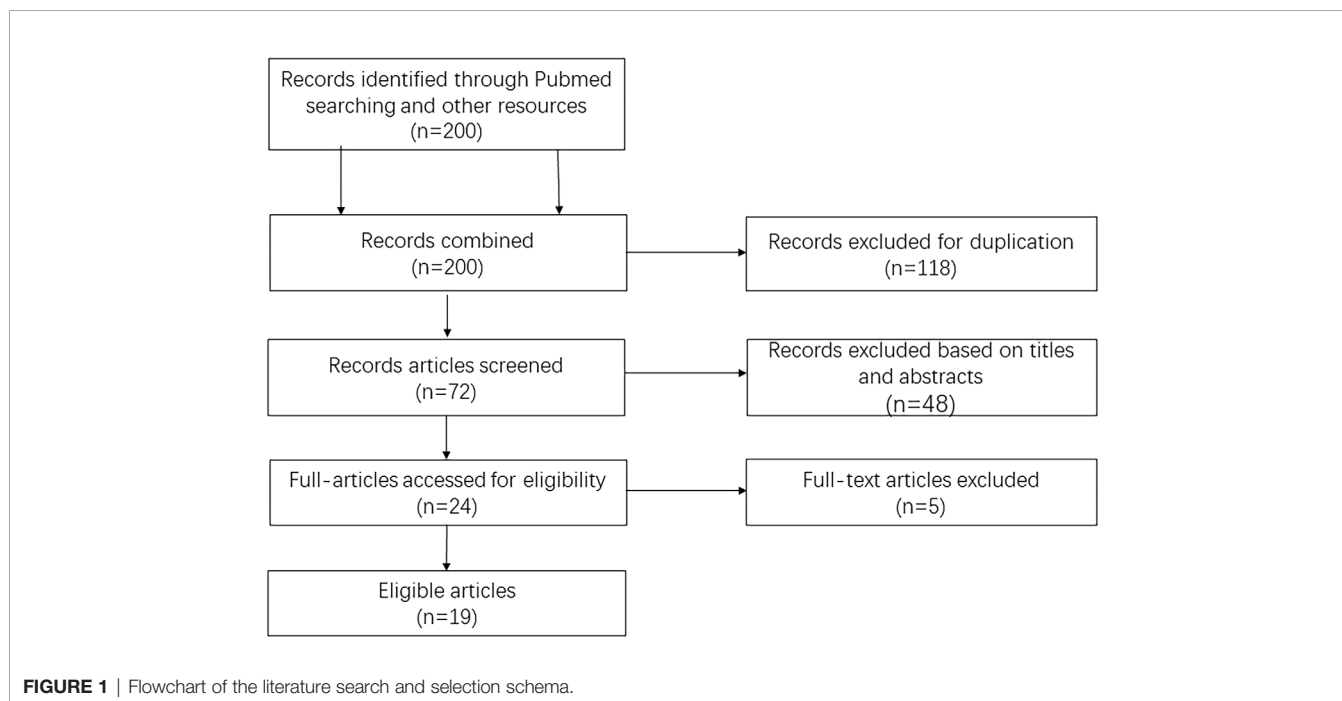
FIGURE 1 | Flowchart of the literature search and selection schema.

TABLE 1 | Baseline characteristics of included studies.

Author	Year	Country	Patients, n	Nodules, n	MeanAge	Guidelines	Standard	Malignant lesions	Benign lesions
Zhang (18)	2020	China	1,271	1,271	48	ACR/ATA/Kwak/KTA	Needle biopsy, surgical resection	736	535
Gao (19)	2019	China	1,764	2,544	—	ACR/ATA/Kwak	Surgical resection	1,681	863
Barbosa (20)	2019	Brazil	139	140	49	ACR/ATA	Needle biopsy, surgical resection	66	74
Jabar (21)	2019	India	127	127	—	ACR/Kwak	Needle biopsy, surgical resection	23	104
Xv (24)	2019	China	370	432	43	ACR	Needle biopsy, surgical resection	258	174
Yoon (25)	2019	Korea	1,836	2,274	55	ACR/KTA	Needle biopsy, surgical resection	300	1,974
Huang (26)	2019	USA	137	250	58	ACR/ATA	Surgical resection	65	185
Ruan (27)	2019	China	918	1,001	46	ATA	Needle biopsy, surgical resection	392	609
Wang (22)	2017	China	1,011	1,011	51	Kwak	Surgical resection	464	547
Liu (23)	2015	China	2,921	3,980	52	Kwak	Needle biopsy	228	3,752
Ha (17)	2018	Korea	1,802	2,000	51	KTA, ATA, ACR	Needle biopsy, surgical resection	1,546	454
Mohammadi (28)	2019	Canada	—	425	—	ATA	Needle biopsy	31	394
Wu (29)	2019	China	894	1,000	—	ATA, ACR	Needle biopsy, surgical resection	530	470
Xu (30)	2019	China	2,031	2,465	48	KTA, ACR, ETA	Surgical resection	885	1,146
Yoon.J (31)	2017	Korea	4,585	4,696	51	ATA, Kwak	Needle biopsy, surgical resection	1,044	3,652
Hoang (32)	2018	USA	92	100	52	ACR	Needle biopsy, surgical resection	15	85
Li (33)	2019	China	128	130	48	ACR	Needle biopsy, surgical resection	73	57
Trimboli (34)	2019	Switzerland	475	1,058	53	ETA	Needle biopsy, surgical resection	—	—
Maino (35)	2018	Italy	340	432	57	ATA. ETA	Needle biopsy	—	—

TABLE 2 | Pooled estimates of the sensitivity, specificity, PLR, NLR, DOR, AUC, and SE (AUC).

Reference guideline	N[a]	Pooled sensitivity (95% CI)	Pooled specificity (95% CI)	Pooled PLR (95% CI)	Pooled NLR (95% CI)	Pooled DOR (95% CI)	AUC	SE (AUC)
ACR	13	0.85(0.84–0.86)	0.68(0.6–0.69)	2.98(2.37–3.75)	0.22(0.16–0.29)	15.23(9.23-25.11)	0.8553	0.0311
Kawk	6	0.94(0.94–0.95)	0.62(0.6–0.63)	3.23(0.90–11.61)	0.08(0.04–0.16)	43.15(19.09–97.53)	0.9101	0.0621
ATA	10	0.94(0–94-0.95)	0.44(0.43–0.45)	2.06(1.54–2.75)	0.16(0.10–0.28)	13.33(5.90-30.14)	0.8976	0.0414
KTA	4	0.85(0.83–0.86)	0.47(0.46-0.48)	2.60(1.2–5.57)	0.18(0.08–0.39)	14.57(5.77-36.84)	0.9022	0.0430
ETA	4	0.85(0.83-0.87)	0.61(0.59-0.62)	2.84(1.43-5.64)	0.21(0.13–0.34)	13.18(4.89-35.5)	0.8810	0.0561

na, number of studies; PLR, positive likelihood ratios; NLR, negative likelihood ratios; DOR, diagnostic odds ratio; AUC, area under the curve.

**TABLE 3** | Relative diagnostic odds ratio (RDOR) with 95% confidence limit.

B A	ACR	ATA	ETA	Kawk	KTA
ACR	–	0.6387(0.3678–1.1090)	0.7308(0.3000–1.7803)	0.5564(0.2552–1.2131)	0.5734(0.2759–1.1919)
ATA	1.5658(0.9017–2.7189)	–	1.1443(0.4532–2.8897)	0.8713(0.3995–1.8999)	0.8979(0.4072–1.9802)
ETA	1.3683(0.5617–3.3332)	0.8739(0.3461–2.2067)	–	0.7614(0.2498–2.3208)	0.7846(0.3075–2.0020)
Kawk	1.7972(0.8243–3.9183)	1.3134(0.5264–2.5028)	1.3138(0.4309–4.0035)	–	1.0306(0.3927–2.7048)
KTA	1.7439(0.8390–3.6247)	1.1137(0.5050–2.4561)	1.2745(0.4995–3.2518)	0.9703(0.3697–2.5466)	–

SROC, summary receiver operator characteristics.

describe the unnecessary FNA rates. There are 19 articles of 24,325 thyroid nodules in our research. Compared with the published article, we can include the articles with indirect data and finish the diagnostic 2 × 2 table using Cal software. Besides, in an article we both included, our article included the data of

diagnostic performance for malignant thyroid nodules which better describes the diagnostic efficiency. It can influence the pooled results.

The TI-RADS guidelines based on ultrasound have been widely used in clinics, providing recommendation for further
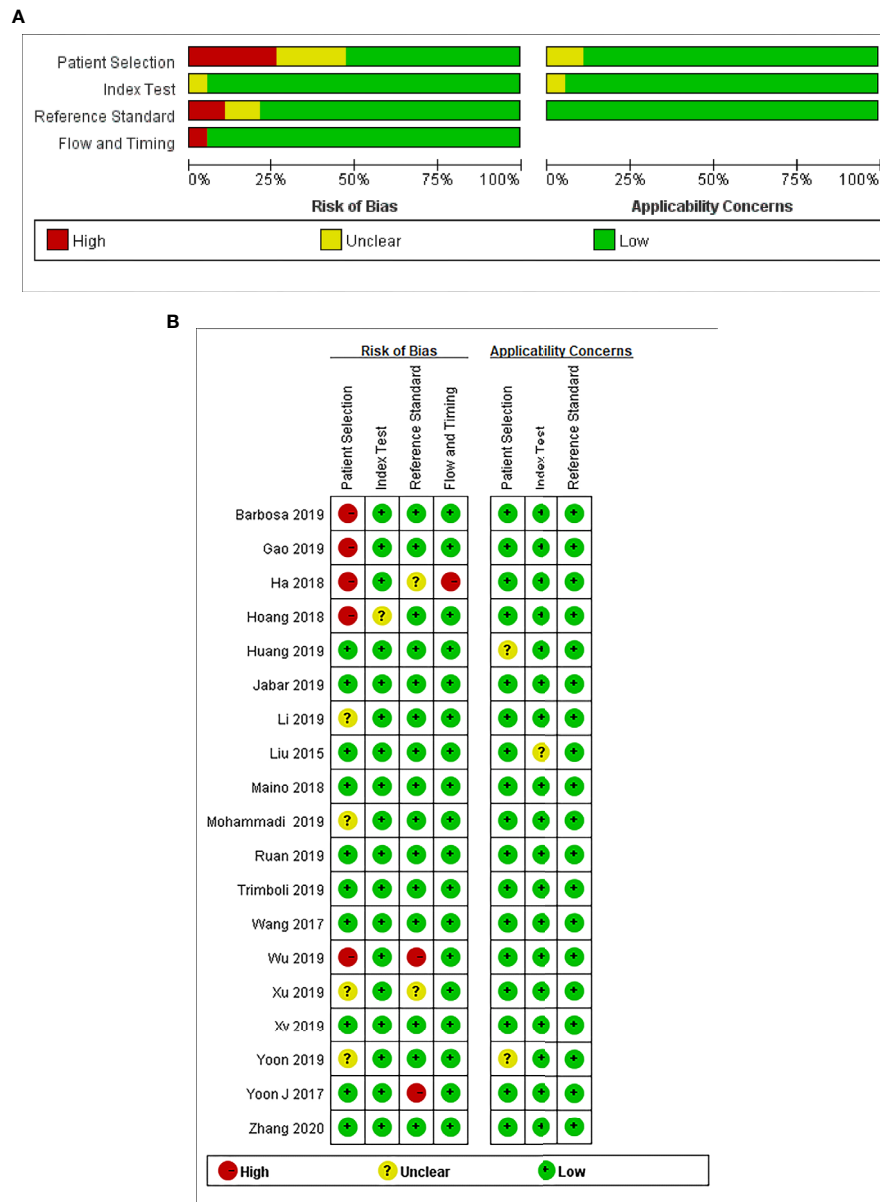
FIGURE 2 | (A) Risk of bias and applicability concerns graph: review authors' judgments about each domain presented as percentages across the included studies. **(B)** Risk of bias and applicability concerns summary: review authors' judgments about each domain for each included study.

diagnosis and treatment while reducing the influence of subjective factors in diagnosing. In our study, we can see the five guidelines all have great diagnostic performance with high AUC above 0.8. However, there are similarities and differences among the five guidelines in structure, risk stratification, size thresholds, and diagnostic performance. More studies need to be done. The structure of the five classification guidelines is internally different. The ACR guidelines and Kwak TI-RADS are point-based systems, and the other three guidelines are based on the pattern. Compared with point-based guidelines, the

simplified pattern-based guidelines are more intuitive and feasible clinically but with decreased accuracy. Although the point-based guidelines are cumbersome, they're easy to control by clinical doctors, especially estimating individual nodules, which are with great accuracy. However, in clinical application, complex analyses and calculations always require the help of computers (10, 38). Every guideline has been divided into several categories to evaluate the thyroid nodules. As the risk stratification categories rise, the risk of the malignancy is increased, but the five guidelines have differences in the

classification. For example, a category five or four thyroid nodule in ETA may be classified as ACR T4/3 or KTA T4/3, and a nodule of KTA T3 (low suspicion) and ETA category 3 (low risk) may be classified as ACR T2, which means not suspicious. Different classification criteria like the above may lead to different specificities, and as the results in our research, the ACR guidelines surely had the highest specificity. It also means less recommendation for FNAC, but the rate of misdiagnosis increases. We need more studies to discuss. As for the performance of recommendation for FNAC, the five guidelines have different size thresholds, and the thresholds also change with categories in different guidelines. For example, for ACR TI-RADS, the threshold of categories three, four, and five is 2.5, 1.5, and 1cm, respectively (39). Some studies (17) have shown ACR TI-RADS have the most effective criteria which can avoid the unnecessary biopsies effectively. Our results also confirmed this, with the highest RDOR for ACR TI-RADS. Nodule size is an important standard for guidelines and further treatment. The too large thyroid nodules with low malignancy risk will suggest surgery or FNAC.

In addition, there are several limitations in this research. Firstly, the final diagnosis was determined by cytology or pathology. It may be influenced by the operators or observers, with possible bias. Especially in retrospective studies, we are not sure whether subjective factors affect the diagnosis. This influence can't be avoided. The second limitation is caused by the patient selection of included studies. Some studies have included more patients with malignant nodules which could influence the sensitivity and specificity. Thirdly, we didn't have enough data for KTA guidelines and ETA to analyze. Lastly, all analyses are based on the ultrasound; the intra-observer and inter-observer variability still exists.

In conclusion, our research indicates that the five classification guidelines are all effective methods for differential diagnosis of benign and malignant thyroid nodules. They can be used before further diagnosis or treatment as an effective recommendation. In head to head comparison, the result suggests ACR guideline is a better choice in the benign and malignant diagnosis with high diagnostic accuracy. However, we still need more studies to prove our findings.

AUTHOR CONTRIBUTIONS

All authors directly participated in the planning, execution, or analysis of the study and wrote the manuscript. RY conducted the literature review, planned and performed all statistical analyses. HZ and XZ provided input and direction for the analytic strategy and editing of the manuscript. YZ reviewed the included articles and provided editing of the manuscript. XM provided technical quality control to ensure accuracy of reported results. All authors contributed to the article and approved the submitted version.

REFERENCES

1. 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):1–133. doi: 10.1089/thy.2015.0020

2. 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

3. Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, et al. Cancer treatment and survivorship statistics, 2016. *CA Cancer J Clin* (2016) 66(4):271–89. doi: 10.3322/caac.21349

4. Ozdemir D, Bestepe N, Faki S, Kilicarslan A, Parlak O, Ersoy R, et al. Comparison of thyroid fine needle aspiration biopsy results before and after implementation of Bethesda classification. *Cytopathology* (2017) 28(5):400–6. doi: 10.1111/cyt.12437

5. 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(3):892–9. doi: 10.1148/radiol.11110206

6. An JY, Unsdorfer KML, Weinreb JC. BI-RADS, C-RADS, CAD-RADS, LI-RADS, Lung-RADS, NI-RADS, O-RADS, PI-RADS, TI-RADS: Reporting and Data Systems. *Radiographics* (2019) 39:1435–6. doi: 10.1148/rg.2019190087

7. 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) 17:370–95. doi: 10.3348/kjr.2016.17.3.370

8. 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(5):225–37. doi: 10.1159/000478927

9. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey S, 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(5):587–95. doi: 10.1016/j.jacr.2017.01.046

10. Ha SM, Ahn HS, Baek JH, Ahn HY, Chung YJ, Cho BY, et al. Validation of Three Scoring Risk-Stratification Models for Thyroid Nodules. *Thyroid* (2017) 27(12):1550–7. doi: 10.1089/thy.2017.0363

11. Ma X, Zeng H, Zhang J, Chen L, Jia H. New practical methods to obtain crucial data in performing diagnostic meta-analysis of the published literature. *J Evid Based Med* (2018) 11:56–63. doi: 10.1111/jebm.12281

12. Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol* (2006) 6:31. doi: 10.1186/1471-2288-6-31

13. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med* (1993) 12:1293–316. doi: 10.1002/sim.4780121403

14. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *Bmj* (2003) 327:557–60. doi: 10.1136/bmj.327.7414.557

15. Dinnes J, Deeks J, Kirby J, Roderick P. A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. *Health Technol Assess* (2005) 9:1–113, iii. doi: 10.3310/hta9120

16. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma J, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* (2011) 155(8):529–36. doi: 10.7326/0003-4819-155-8-201110180-00009

17. Ha EJ, Na DG, Baek JH, Sung JY, Kim JH, Kang SY. US Fine-Needle Aspiration Biopsy for Thyroid Malignancy: Diagnostic Performance of Seven Society Guidelines Applied to 2000 Thyroid Nodules. *Radiology* (2018) 287(3):893–900. doi: 10.1148/radiol.2018171074

18. Zhang WB, et al. Comparisons of ACR TI-RADS, ATA guidelines, Kwak TI-RADS, and KTA/KSThR guidelines in malignancy risk stratification of thyroid nodules. *Clin Hemorheol Microcirc* (2020). doi: 10.3233/ch-190778

19. Gao L, Xi X, Jiang Y, Yang X, Wang Y, Zhu S, et al. Comparison among TIRADS (ACR TI-RADS and KWAK- TI-RADS) and 2015 ATA Guidelines in the diagnostic efficiency of thyroid nodules. *Endocrine* (2019) 64(1):90–6. doi: 10.1007/s12020-019-01843-x

20. Barbosa TLM, Junior COM, 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 (1):112. doi: 10.1186/s12902-019-0429-5

21. Jabar ASS, Koteshwara P, Andrade J. Diagnostic reliability of the Thyroid Imaging Reporting and Data System (TI-RADS) in routine practice. *Pol J Radiol* (2019) 84:e274–80. doi: 10.5114/pjr.2019.86823

22. Wang Y, Lei KR, He YP, Li XL, Ren WW, Zhao CK, et al. Malignancy risk stratification of thyroid nodules: comparisons of four ultrasound Thyroid Imaging Reporting and Data Systems in surgically resected nodules. *Sci Rep* (2017) 7(1):11560. doi: 10.1038/s41598-017-11863-0

23. Zhang J, Liu BJ, Xu HX, Xu JM, Zhang YF, Liu C, 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(4):5911–7.

24. Xu Y, Qi X, Zhao X, Ren W, Ding W. Clinical diagnostic value of contrast-enhanced ultrasound and TI-RADS classification for benign and malignant thyroid tumors: One comparative cohort study. *Med (Baltimore)* (2019) 98: e14051. doi: 10.1097/md.0000000000014051

25. Yoon SJ, Na DG, Gwon HY, Paik W, Kim WJ, Song JS, et al. Similarities and Differences Between Thyroid Imaging Reporting and Data Systems. *AJR Am J Roentgenol* (2019) 213(2):W76–84. doi: 10.2214/AJR.18.20510

26. Huang BL, Ebner SA, Makkar JS, Bentley-Hibbert S, McConnell RJ, Lee JA, et al. 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(5):398–403. doi: 10.1634/theoncologist.2019-0362

27. Ruan JL, Yang HY, Liu RB, Liang M, Han P, Xu XL, et al. 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(9):4871–8. doi: 10.1007/s00330-018-5992-z

28. Mohammadi M, Betel C, Burton KR, Higgins KM, Ghorab Z, Halperin IJ. Retrospective Application of the 2015 American Thyroid Association Guidelines for Ultrasound Classification, Biopsy Indications, and Follow-up Imaging of Thyroid Nodules: Can Improved Reporting Decrease Testing? *Can Assoc Radiol J* (2019) 70(1):68–73. doi: 10.1016/j.carj.2018.09.001

29. Wu XL, Du JR, Wang W, Jin CX, Sui GQ, Yang DY, 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(1):121–31. doi: 10.1007/s12020-019-01886-0

30. Xu T, Wu Y, Wu RX, Zhang YZ, Gu JY, Ye XH, et al. Validation and comparison of three newly-released Thyroid Imaging Reporting and Data Systems for cancer risk determination. *Endocrine* (2019) 64(2):299–307. doi: 10.1007/s12020-018-1817-8

31. Yoon JH, Han K, Kim EK, Moon HJ, Kwak JY. Diagnosis and Management of Small Thyroid Nodules: A Comparative Study with Six Guidelines for Thyroid Nodules. *Radiology* (2017) 283:560–9. doi: 10.1148/radiol.2016160641

32. Hoang JK, Middleton WD, Farjat AE, Langer JE, Reading CC, Teefey S, et al. Reduction in Thyroid Nodule Biopsies and Improved Accuracy with American College of Radiology Thyroid Imaging Reporting and Data System. *Radiology* (2018) 287(1):185–93. doi: 10.1148/radiol.2018 172572

33. Li X, Hou XJ, Du LY, Wu JQ, 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(3):279–91. doi: 10.3233/ch-180477

34. Trimboli P, Ngu R, Royer B, Giovanella L, Bigorgne C, Simo R, et al. A multicentre validation study for the EU-TIRADS using histological diagnosis as a gold standard. *Clin Endocrinol (Oxf)* (2019) 91(2):340–7. doi: 10.1111/cen.13997

35. Maino F, Forleo R, Martinelli M, Fralassi N, Barbato F, Pilli T, et al. Prospective Validation of ATA and ETA Sonographic Pattern Risk of Thyroid Nodules Selected for FNAC. *J Clin Endocrinol Metab* (2018) 103 (6):2362–8. doi: 10.1210/jc.2018-00274

36. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* (1978) 8:283–98. doi: 10.1016/s0001-2998(78)80014-2

37. Castellana M, Castellana C, Treglia G, Giorgino F, Giovanella L, Russ G, et al. Performance of Five Ultrasound Risk Stratification Systems in Selecting Thyroid Nodules for FNA. *J Clin Endocrinol Metab* (2020) 105(5):dgz170. doi: 10.1210/clinem/dgz170

38. Choi YJ, Baek JH, Baek SH, Shim WH, Lee KD, Lee HS, et al. Web-Based Malignancy Risk Estimation for Thyroid Nodules Using Ultrasonography Characteristics: Development and Validation of a Predictive Model. *Thyroid* (2015) 25(12):1306–12. doi: 10.1089/thy.2015.0188

39. Tessler FN, Middleton WD, Grant EG. Thyroid Imaging Reporting and Data System (TI-RADS): A User's Guide. *Radiology* (2018) 287:29–36. doi: 10.1148/radiol.2017171240