Comparison of the C-TIRADS, ACR-TIRADS, and ATA guidelines in malignancy risk stratification of thyroid nodules

Yifeng Cai[1#^], Ruixuan Yang[2#], Shuhui Yang[1], Laishun Lu[3], Rui Ma[3], Zidong Xiao[1], Nie Lin[1], Yuhan Huang[1], Liwen Chen[1]

[1]Department of Endocrinology, Shantou Central Hospital, Shantou, China; [2]Department of Laboratory Medicine, Shantou Central Hospital, Shantou, China; [3]Department of Medicals Ultrasonics, Shantou Central Hospital, Shantou, China

Contributions: (I) Conception and design: Y Cai; (II) Administrative support: Y Cai, R Yang, S Yang; (III) Provision of study materials or patients: R Yang, S Yang, L Lu; (IV) Collection and assembly of data: L Lu, R Ma, Z Xiao, Y Huang, N Lin, L Chen; (V) Data analysis and interpretation: R Yang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work and should be considered as co-first authors.

Correspondence to: Yifeng Cai, MD. Department of Endocrinology, Shantou Central Hospital, No. 114 Waima Road, Shantou 515031, China. Email: gdsstscyf@163.com.

Background: To compare the diagnostic performance in determining the malignancy of thyroid nodules and the fine needle aspiration (FNA) recommendations of the guidelines set forth by the Superficial Organ and Vascular Ultrasound Group of the Society of Ultrasound in Medicine of the Chinese Medical Association in 2020 [2020 Chinese Thyroid Imaging Reporting and Data System (C-TIRADS)], the American College of Radiology in 2017 (2017 ACR-TIRADS) and the American Thyroid Association in 2015 (2015 ATA guidelines).

Methods: From January 2021 to December 2021, 1,228 thyroid nodules with definitive postoperative histopathology and ultrasound (US) examination within 3 months before surgery in Shantou Central Hospital were enrolled in this study. We collected the data in 2022. The participants formed a consecutive series. The clinical and US features of the nodules were retrospectively reviewed and categorized according to the 2020 C-TIRADS, the 2017 ACR-TIRADS and the 2015 ATA guidelines. The diagnostic performance and unnecessary FNA rates of the three guidelines were calculated.

Results: The 2017 ACR-TIRADS had the highest diagnostic performance [area under the receiver operating characteristic curve (AUROC) 0.938], followed by the 2020 C-TIRADS (AUROC 0.933) and the 2015 ATA guidelines (AUROC 0.928). The ATA guidelines had the highest specificity (93.38%), accuracy (92.10%) and positive predictive value (PPV) (80.56%) among the three guidelines. There were no significant differences in the sensitivity and negative predictive value (NPV) among the three guidelines. The sensitivity, specificity, PPV, NPV and accuracy of the FNA recommendations based on the C-TIRADS were 84.25%, 58.76%, 38.92%, 92.28% and 64.82%, respectively, which were higher than those of the ACR-TIRADS (57.53%, 42.94%, 23.93%, 76.43% and 46.42%, respectively) and the ATA guidelines (62.67%, 13.25%, 18.39%, 53.22% and 25.00%, respectively). Compared with the ACR-TIRADS (76.07%) and the ATA guidelines (81.61%), the C-TIRADS showed advantages in the unnecessary FNA rate (61.08%), especially in nodules larger than 20 mm.

Conclusions: The 2020 C-TIRADS, the 2017 ACR-TIRADS and the 2015 ATA guidelines can effectively predict the malignancy risk of thyroid nodules. Compared with the 2017 ACR-TIRADS and the 2015 ATA guidelines, the 2020 C-TIRADS may offer a meaningful reduction in FNA recommendations with the

^ ORCID: 0000-0002-0266-2687.

highest efficacy in distinguishing thyroid carcinoma.

Introduction

Thyroid nodules are detected in 19–68% of randomly selected individuals, and 7–15% of them are malignant (1). With the rising incidence of thyroid cancer and the development of diagnostic effects, the detection rate of thyroid cancer has risen yearly (2). Many ultrasonic examinations and fine needle aspiration (FNA) have been carried out, which has raised concerns about the overtreatment of thyroid nodules.

Many risk stratification systems have been established to standardize the risk stratification of thyroid lesions and propose corresponding management recommendations, including FNA (1,3-5). Two of them, the Thyroid Imaging Reporting and Data System of the American College of Radiology (ACR-TIRADS) set forth by the American College of Radiology in 2017 and the American Thyroid Association (ATA) guidelines set forth by the ATA in 2015, are widely applied for the diagnosis and treatment of thyroid nodules. Recently, the Superficial Organ and Vascular Ultrasound Group of the Society of Ultrasound in Medicine of the Chinese Medical Association established the 2020 Chinese Thyroid Imaging Reporting and Data System (C-TIRADS) guidelines (6). There are many differences in the analysis of ultrasound (US) images among the three guidelines. The ACR-TIRADS takes advantage of a weighted method to calculate the scores of US signs to predict the malignant risk of thyroid nodules, while the C-TIRADS counts the number of positive and negative signs. The ATA guidelines propose a pattern-based qualitative system defining 5 categories with different risks of malignancy. Additionally, the nodule size thresholds for FNA are different in the three guidelines. These different guidelines for assessing the same thyroid nodule may show different diagnostic performance of malignancy and FNA recommendations. Although some studies (7,8) have been conducted to compare the diagnostic performance of malignancy and FNA recommendations of the three guidelines, further exploration on FNA recommendations has not been undertaken, which prevents us from managing thyroid nodules more effectively. The aim of this study was to compare their diagnostic performance for malignancy and FNA recommendations. We presented this article in accordance with the STARD reporting checklist (available at https://qims.amegroups.com/article/view/10.21037/qims-22-826/rc).

Methods

This is a retrospective study. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the ethics board of the Shantou Central Hospital [ID (2021) scientific research 057], and individual consent for this retrospective analysis was waived. The diagnostic performance of the guidelines is affected by many factors, including the experience of the radiologist and the included sample. In previous studies, the areas under the curve (AUCs) of diagnostic performance in malignant thyroid nodules according to the C-TIRADS, the ACR-TIRADS and the ATA guidelines varied from 0.8 to 0.95, in which the highest remained controversial (9-11). Therefore, we estimated that the AUCs of the guidelines were 0.8 and calculated the sample size that could predict malignancy risk in thyroid nodules. We used PASS v11.0 (NCSS LLC., Kaysville, USA) to estimate the minimal sample size. We chose "Tests for One ROC Curve" in the category, and the analysis was performed with $\alpha=0.05$ and $\beta=0.1$. We set the sample allocation ratio, AUC0 and AUC1 at 1.0, 0.5 and 0.8, respectively. We set the lower false-positive rate and the upper false-positive rate at 0 and 1, respectively. The data were discrete, and the B value was set at 1. A two-sided Z test was adopted. The minimal sample size was 34 and included 17 malignant thyroid nodules.
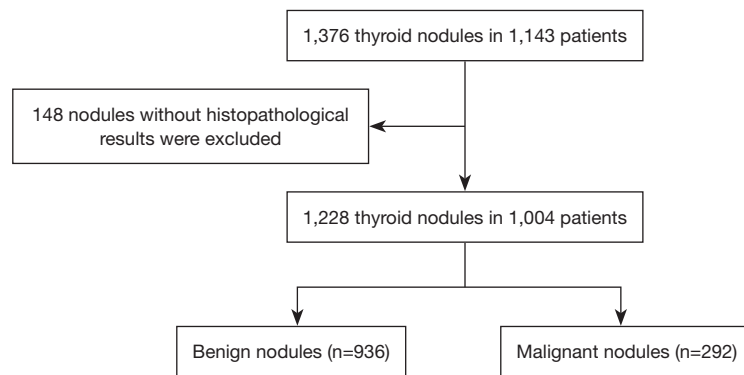
4516

Cai et al. Comparison of the C-TIRADS, ACR-TIRADS and ATA guidelines

Figure 1 Flow of the participants.

Patients

We performed a retrospective review on the database in Shantou Central Hospital for patients with thyroid nodules from January 2021 to December 2021. We collected the data in 2022. The participants formed a consecutive series. Patients with the following criteria were included: (I) underwent surgery; (II) definitive histopathological results of target nodules; and (III) an US examination of the thyroid within 3 months before surgery. The exclusion criteria were as follows: (I) the histopathological results of the target nodules were unclear; (II) the target nodules could not be clearly identified on US examination; (III) the patients had undergone radiotherapy or chemotherapy; (IV) the medical records were incomplete; and (V) the patients had insufficient mental development or gravidas. If the patient had more than one thyroid nodule, the one nodule that met the above criteria and had the most high-risk US features in each thyroid lobe or isthmus was included. We chose the histopathological result as a reference standard to make it more convincing. A total of 1,143 patients with 1,376 thyroid nodules underwent an US examination in the last year, while 139 patients with 148 thyroid nodules were excluded due to the lack of histopathological results. A flowchart of the participants is shown in *Figure 1*.

Analysis of US images

The ultrasonography examinations of the thyroid gland and cervical region were performed with General Electric Voluson E8 (General Electric, Schenectady, USA) and Philips Iu22 (Philips, Amsterdam, Netherlands) equipped with 5–12-MHz linear array transducers. The US images of 1,004 patients with 1,228 nodules who met the above

criteria were independently analyzed by two experienced radiologists (L Lu and R Ma) who had engaged in thyroid US diagnosis for more than 10 years. The radiologists were blinded to the histopathologic results of the thyroid nodules. All thyroid nodules were evaluated and categorized according to the 2020 C-TIRADS, the 2017 ACR-TIRADS and the 2015 ATA guidelines. The 2020 C-TIRADS was calculated by adding the number of malignant US features, including vertical orientation, solid composition, hypoechogenicity, microcalcifications and ill-defined/irregular margins or extrathyroidal extension, and subtracting 1 if a negative feature of the comet-tail artifacts was present (6). Based on the 2020 C-TIRADS, the thyroid nodules were classified as C-TIRADS 2 (–1 point), C-TIRADS 3 (0 point), C-TIRADS 4A (1 point), C-TIRADS 4B (2 points), C-TIRADS 4C (3 points) and C-TIRADS 5 (4 points or more). In the 2017 ACR-TIRADS, mixed cystic and solid composition, solid or almost completely solid composition, hyperechoic or isoechoic, hypoechoic, very hypoechoic, vertical orientation, lobulated or irregular margin, extrathyroidal extension, macrocalcifications, peripheral calcifications and microcalcifications were assigned 1, 2, 1, 2, 3, 3, 2, 3, 1, 2 and 3 points, respectively, and the other US features were assigned 0 points. The score of the ACR-TIRADS was calculated by summing the points of the malignant US features that the thyroid nodule contained (5). Similarly, based on the 2017 ACR-TIRADS, the thyroid nodules were classified as ACR-TIRADS 2 (2 points), ACR-TIRADS 3 (3 points), ACR-TIRADS 4A (4 points), ACR-TIRADS 4B (5 points), ACR-TIRADS 4C (6 points) and ACR-TIRADS 5 (7 points or more). In the 2015 ATA guidelines, thyroid nodules with a high suspicion pattern, intermediate suspicion pattern, low suspicion pattern, very low suspicion

pattern and benign pattern were assigned 5, 4, 3, 2 and 1 point, respectively. The "not special" pattern of the ATA guidelines was considered a subtype of the indeterminate-suspicion pattern as Yoon *et al.* had claimed (12) and was assigned 4 points. Diagnostic FNA in the ATA guidelines is recommended for (I) high or intermediate suspicion sonographic pattern: nodules ≥1 cm; (II) low suspicion sonographic pattern: nodules ≥1.5 cm; and (III) very low suspicion sonographic pattern: nodules ≥2 cm. Diagnostic FNA in the ACR-TIRADS is recommended for (I) ACR-TIRADS 5: nodules ≥1 cm; (II) ACR-TIRADS 4A, 4B or 4C: nodules ≥1.5 cm; and (III) ACR-TIRADS 3: nodules ≥2.5 cm. Diagnostic FNA in C-TIRADS is recommended for (I) C-TIRADS 4A: (i) nodules >15 mm; (ii) if nodules are multiple or immediately adjacent to the trachea or recurrent laryngeal nerve, FNA can be considered if the nodule is >10 mm. (II) C-TIRADS 4B: (i) nodules >10 mm; (ii) if nodules are multiple or immediately adjacent to the capsule, trachea, or recurrent laryngeal nerve, FNA can be considered if the nodule is >5 mm. (III) The recommended management of C-TIRADS 4C and C-TIRADS 5 nodules is similar to that of category 4B nodules. However, if there are typical cervical metastatic lymph nodes, then the most suspicious nodules of any size in the ipsilateral thyroid require FNA. Disagreements in the analysis of the US images between the two radiologists were resolved by consensus. The pathologists did not know the scores of the thyroid nodules.

Statistical analysis

Statistical analysis was performed at both the patient level and the nodule level. Data were extracted by two groups of reviewers (Z Xiao and Y Huang, N Lin and L Chen) independently and included the sex of the patients, the age of the patients and the histopathological results of the nodules. Adenomas, nodular goiters, thyroiditis and parathyroid diseases were considered benign, while papillary thyroid carcinomas, follicular thyroid carcinomas, medullary thyroid carcinoma and other carcinomas were considered malignant. There were no missing data on the index test or reference standard result in our study. Extracted data were entered into IBM SPSS v12.0 (IBM Corp., Armonk, USA) and STATA v13.0 (Stata Corp., College Station, USA). Continuous variables with a normal distribution are expressed as the mean ± standard deviation (SD), and those with an abnormal distribution are expressed as the median (interquartile range). The nominal and ordinal variables were expressed as frequencies and proportions. Independent two-sample *t*-tests and rank-sum tests were used to compare the results in normal distributions and in abnormal distributions, respectively. The areas under the receiver operating characteristic (AUROC) curves were compared by the Z test. The statistical significance was set at P<0.05. The P value was a two-sided test.

The diagnostic performance of the three guidelines was assessed by ROC curve analysis based on the established cutoff values. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy and AUROC were calculated. The number of thyroid nodules recommended for FNA, the malignant detection rate of the FNA recommendations and the unnecessary FNA rates were also calculated. The unnecessary FNA rates and the malignant detection rates of FNA recommendations were calculated using the following equations: unnecessary FNA rate = the number of benign nodules among the recommended FNA nodules/the number of recommended FNA nodules; malignant detection rates of FNA recommendations = the number of malignant nodules among the recommended FNA nodules/the number of recommended FNA nodules. To compare the unnecessary FNA rates of the three guidelines in thyroid nodules of different sizes, subgroups were established according to whether the thyroid nodules were larger than 20 mm. The sensitivity, specificity, PPV, NPV and accuracy of FNA recommendations were calculated based on the three guidelines.

Results

A total of 1,004 patients with 1,228 thyroid nodules were included in the study. The nodules of 724 patients were all benign, and 280 patients had one or more malignant nodules at least. There were 828 females and 176 males, and the median age was 48.00 (38.00, 57.00) years. There were no significant differences in sex or age between the patients with malignant nodules and the patients with benign nodules. A total of 936 nodules were benign, including 304 adenomas, 576 nodular goiters, 50 thyroiditis and 6 parathyroid diseases, and 292 nodules were malignant, including 280 papillary thyroid carcinomas, 11 follicular thyroid carcinomas and 1 medullary thyroid carcinoma. The median size of the included nodules was 25.00 (13.00, 35.00) mm, and the sizes of the benign nodules were significantly larger than the sizes of the malignant nodules [28.50 (18.25, 37.00) *vs.* 12.00 (8.00, 18.00) mm, P<0.001].

There were many differences in the US features between the malignant nodules and benign nodules. More malignant nodules were solid or purely solid [277 (94.86%) *vs.* 420 (44.87%), P<0.001], very hypoechoic [63 (21.58%) *vs.* 19 (2.03%), P<0.001] or hypoechoic [213 (72.95%) *vs.* 539 (57.59%), P<0.001], immediately adjacent to the capsule, trachea or the recurrent laryngeal nerve [212 (72.60%) *vs.* 480 (51.28%), P<0.001], with a taller-than-wider shape [92 (31.51%) *vs.* 25 (2.67%), P<0.001, 103 (35.27%) *vs.* 29 (3.10%), P<0.001], extrathyroidal extension [28 (9.59%) *vs.* 6 (0.64%), P<0.001], irregular or lobulated margin [198 (67.81%) *vs.* 44 (4.70%), P<0.001], microcalcification [166 (56.85%) *vs.* 45 (4.81%), P<0.001], and typical cervical metastatic lymph nodes (9 (3.08%) *vs.* 0 (0.00%), P<0.001]. More benign nodules were mixed [495 (52.88%) *vs.* 15 (5.14%), P<0.001], cystic, purely cystic or spongiform [21 (2.25%) *vs.* 0 (0.00%), P=0.007], hyperechoic or isoechoic [358 (38.25%) *vs.* 16 (5.48%), P<0.001], and anechoic [20 (2.13%) *vs.* 0 (0.00%), P=0.007], had a wider-than-taller shape [911 (97.33%) *vs.* 200 (68.49%), P<0.001; 907 (96.90%) *vs.* 189 (64.73%), P<0.001], a smooth or ill-defined margin [886 (94.66%) *vs.* 66 (22.60%), P<0.001], and no echogenic foci or only large comet-tail artifacts [801 (85.58%) *vs.* 93 (31.85%), P<0.001]. The basic characteristics of the patients and the US features of the thyroid nodules are shown in *Table 1*.

Comparison of diagnostic performance of the three guidelines

All thyroid nodules were classified following the three guidelines. The ROC curve shows the diagnostic performance of the three guidelines in *Figure 2*. As the ROC curve suggested, the C-TIRADS, ACR-TIRADS and ATA guidelines could distinguish malignant nodules from benign nodules with cutoff values of 1.5, 5.5 and 4.5 points, respectively. According to the cutoff value, thyroid nodules were considered malignant nodules if they were classified as 4B, 4C and 5 in the C-TIRADS, 4C and 5 in the ACR-TIRADS and highly suspicious in the ATA guidelines. When the thyroid nodules were classified as 4A or below in the C-TIRADS, 4B or below in the ACR-TIRADS and intermediate suspicion or below in the ATA guidelines, the nodules were regarded as benign. Based on the criteria, the AUROC of the ACR-TIRADS was significantly larger than that of the ATA guidelines (0.938 *vs.* 0.928, P=0.02). However, the C-TIRADS had no significant difference in the AUROC with the ACR-TIRADS and the ATA

guidelines (0.933 *vs.* 0.938, P=0.24; 0.933 *vs.* 0.928, P=0.32). The comparison of the diagnostic performance of the three guidelines is shown in *Table 2*. The ATA guidelines had a higher specificity (93.38% *vs.* 89.42%, P=0.002; 93.38% *vs.* 88.68%, P<0.001), accuracy (92.10% *vs.* 89.41%, P=0.02; 92.10% *vs.* 89.25%, P=0.02) and PPV (80.56% *vs.* 72.50%, P=0.01; 80.56% *vs.* 71.51%, P=0.006) than the C-TIRADS and the ACR-TIRADS. There were no significant differences in the sensitivity (89.38% *vs.* 91.10%, P=0.49; 91.10% *vs.* 88.01%, P=0.22; 89.38% *vs.* 88.01%, P=0.60) and NPV (96.43% *vs.* 96.96%, P=0.54; 96.96% *vs.* 96.15%, P=0.35; 96.43% *vs.* 96.15%, P=0.76) among the three guidelines.

Comparison of FNA recommendations

According to the C-TIRADS, the number of FNA recommendations was 632, which was significantly less than those based on the ACR-TIRADS (51.47% *vs.* 57.17%, P<0.001) and the ATA guidelines (51.47% *vs.* 81.03%, P<0.001). The malignant detection rates (PPVs) of the FNA recommendations of the three guidelines are shown in *Table 3*. The sensitivity of the FNA recommendations of the C-TIRADS was significantly higher than that of the ACR-TIRADS (84.25% *vs.* 57.53%, P<0.001) and the ATA guidelines (84.25% *vs.* 62.67%, P<0.001), while there was no significant difference in the sensitivity of FNA recommendations between the ACR-TIRADS and the ATA guidelines (57.53% *vs.* 62.67%, P=0.21). The specificity, PPV, NPV and accuracy of FNA recommendations of the C-TIRADS were highest among the three guidelines (58.76%, 38.92%, 92.28% and 64.82%, respectively), followed by the ACR-TIRADS (42.94%, 23.93%, 76.43% and 46.42%, respectively) and the ATA guidelines (13.25%, 18.39%, 53.22% and 25.00%, respectively). The comparison of FNA recommendations of the three guidelines is shown in *Table 4*. The unnecessary FNA rate based on the C-TIRADS was significantly lower than those based on the ACR-TIRADS and the ATA guidelines (61.08% *vs.* 76.07%, P<0.001; 61.08% *vs.* 81.61%, P<0.001). The unnecessary FNA rate of the ACR-TIRADS was significantly lower than that of the ATA guidelines (76.07% *vs.* 81.61%, P=0.006). In the subgroup (<20 mm), the unnecessary FNA rates of the C-TIRADS and ACR-TIRADS were lower than those of the ATA guidelines (39.16% *vs.* 54.23%, P<0.001; 32.28% *vs.* 54.23%, P<0.001), but there was no significant difference in the unnecessary FNA rates between the C-TIRADS and ACR-

Table 1 Demographic characteristics

Basic characteristics	Total	Benign	Malignant	P value
Patient	1,004 (100.0)	724 (72.11)	280 (27.89)	–
Sex				0.10
Female	828 (82.47)	606 (83.70)	222 (79.29)	
Male	176 (17.53)	118 (16.30)	58 (20.71)	
Age (years)	48.00 (38.00, 57.00)	49.00 (38.00, 58.00)	47.00 (37.50, 55.50)	0.07
Nodule	1,228 (100.0)	936 (76.22)	292 (23.78)	–
Size (mm)	25.00 (13.00, 35.00)	28.50 (18.25, 37.00)	12.00 (8.00, 18.00)	<0.001
Composition				
Solid or purely solid	697 (56.76)	420 (44.87)	277 (94.86)	<0.001
Mixed	510 (41.53)	495 (52.88)	15 (5.14)	<0.001
Cystic, purely cystic or spongiform	21 (1.71)	21 (2.25)	0 (0.00)	0.007
Echogenicity				
Very hypoechoic	82 (6.68)	19 (2.03)	63 (21.58)	<0.001
Hypoechoic	752 (61.24)	539 (57.59)	213 (72.95)	<0.001
Hyperechoic or isoechoic	374 (30.46)	358 (38.25)	16 (5.48)	<0.001
Anechoic	20 (1.62)	20 (2.13)	0 (0.00)	0.007
Shape				
Taller than wider (on transverse section)	117 (9.53)	25 (2.67)	92 (31.51)	0.91*
Wider than taller (on transverse section)	1,111 (90.47)	911 (97.33)	200 (68.49)	<0.001
Taller than wider (on transverse or longitudinal section)	132 (10.75)	29 (3.10)	103 (35.27)	
Wider than taller (on transverse or longitudinal section)	1,096 (89.25)	907 (96.90)	189 (64.73)	<0.001
Margin				
Extrathyroidal extension	34 (2.77)	6 (0.64)	28 (9.59)	<0.001
Irregular or lobulated	242 (19.71)	44 (4.70)	198 (67.81)	<0.001
Smooth or ill-defined	952 (77.52)	886 (94.66)	66 (22.60)	<0.001
Echogenic foci				
Microcalcifications and macrocalcifications	1 (0.08)	0 (0)	1 (0.34)	0.24
Microcalcifications	211 (17.18)	45 (4.81)	166 (56.85)	<0.001
Peripheral calcifications	29 (2.36)	24 (2.56)	5 (1.71)	0.40
Macrocalcifications	93 (7.56)	66 (7.05)	27 (9.25)	0.22
None or large comet-tail artifacts	894 (72.80)	801 (85.58)	93 (31.85)	<0.001
Multifocal high-suspicion nodules[†]				0.08
Yes	293 (23.86)	212 (22.65)	81 (27.74)	
No	935 (76.14)	724 (77.35)	211 (72.26)	

Table 1 (continued)

Table 1 (*continued*)

Basic characteristics	Total	Benign	Malignant	P value
Immediately adjacent to the capsule, trachea or the recurrent laryngeal nerve				<0.001
Yes	692 (56.35)	480 (51.28)	212 (72.60)	
No	536 (43.65)	456 (48.72)	80 (27.40)	
Cervical metastatic lymph nodes				<0.001
Yes	9 (0.73)	0 (0.00)	9 (3.08)	
No	1,219 (99.27)	936 (100.00)	283 (96.92)	

Data are presented as number (%) or median (interquartile range). *, the P value reflects the difference between the ratio of thyroid nodules with a taller-than-wider shape only on transverse section and that on transverse or longitudinal section; [†], the nodules were considered multifocal high-suspicion nodules if they were multiple and classified as 4A, 4B, 4C or 5 based on C-TIRADS. C-TIRADS, Chinese Thyroid Imaging Reporting and Data System.
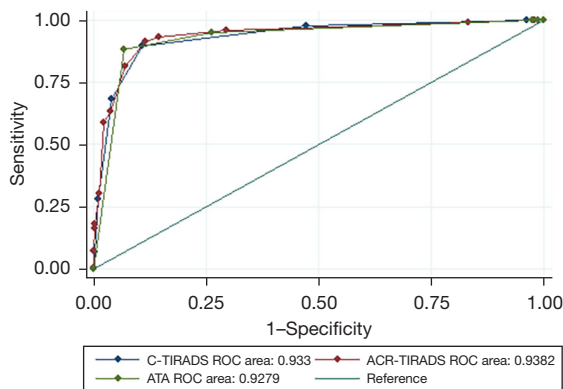

Figure 2 ROC curve of the 2020 C-TIRADS, the 2017 ACR-TIRADS and the 2015 ATA guidelines. The area under ROC: χ^2=6.02, P value =0.05; the cutoff value of C-TIRADS =1.5 points; the cutoff value of ACR-TIRADS =5.5 points; the cutoff value of ATA =4.5 points; C-TIRADS, Chinese Thyroid Imaging Reporting and Data System; ACR-TIRADS, Thyroid Imaging Reporting and Data System of the American College of Radiology; ATA, American Thyroid Association guidelines; ROC, receiver operating characteristic.

Discussion

In our study, we found some low-risk and high-risk features, as previous studies reported. The ratio of malignant thyroid nodules with mixed composition was below that of benign thyroid nodules with mixed composition (5.14% *vs.* 52.88%, P<0.001), which was concurrent with previous studies (8,13). The mixed composition in the US signs of thyroid nodules was considered to be at risk and assigned 1 point according to the ACR-TIRADS. The results of our study showed that the mixed composition in the US sign of thyroid nodules might have no strong association with malignancy, and its significance in the risk stratification system should be reassessed. In addition, we found no significant difference in the malignant rate between thyroid nodules with a taller-than-wider shape on transverse or longitudinal sections and those with a taller-than-wider shape only on transverse sections (35.27% *vs.* 31.51%, P=0.91), which contradicts the results of Moon *et al.* (14). It was necessary to reevaluate whether to assess the vertical orientation of thyroid nodules on both transverse and longitudinal sections. Therefore, future risk stratification systems may need to reassess the significance of these features in distinguishing malignant nodules from benign nodules.

Our study showed that the diagnostic performance of the C-TIRADS was closely comparable to that of the other two guidelines, and the best cutoff value of the C-TIRADS was 1.5 points, which is still controversial. A recent study with a small sample size conducted by Dong *et al.* (9) demonstrated that the ACR-TIRADS had slightly better diagnostic performance than the C-TIRADS for thyroid nodules. In their study, the cutoff value of the C-TIRADS

TIRADS (39.16% *vs.* 32.28%, P=0.15). In the subgroup (≥20 mm), the unnecessary FNA rate of the C-TIRADS was significantly lower than those of the ACR-TIRADS (82.04% *vs.* 88.79%, P=0.005) and the ATA guidelines (82.04% *vs.* 91.29%, P<0.001), but there was no significant difference in the unnecessary FNA rates between the ACR-TIRADS and the ATA guidelines (88.79% *vs.* 91.29%, P=0.14). The comparison of the unnecessary FNA rates of the three guidelines is shown in *Table 5*.

Table 2 Diagnostic performance of the 2020 C-TIRADS, the 2017 ACR-TIRADS and the 2015 ATA guidelines

Parameter	C-TIRADS	ACR-TIRADS	ATA guidelines	P value		
				C-TIRADS *vs.* ACR-TIRADS	ACR-TIRADS *vs.* ATA guidelines	C-TIRADS *vs.* ATA guidelines
AUROC (95% CI)	0.933 (0.916, 0.949)	0.938 (0.921, 0.955)	0.928 (0.911, 0.945)	0.24	0.02	0.32
Sensitivity (%)	89.38	91.10	88.01	0.49	0.22	0.60
Specificity (%)	89.42	88.68	93.38	0.60	<0.001	0.002
PPV (%)	72.50	71.51	80.56	0.76	0.006	0.01
NPV (%)	96.43	96.96	96.15	0.54	0.35	0.76
Accuracy (%)	89.41	89.25	92.10	0.90	0.02	0.02

The AUROC: χ^2=6.02, P=0.0492; the cutoff value of C-TIRADS =1.5 points; the cutoff value of ACR-TIRADS =5.5 points; the cutoff value of ATA =4.5 points. C-TIRADS, Chinese Thyroid Imaging Reporting and Data System; ACR-TIRADS, Thyroid Imaging Reporting and Data System of the American College of Radiology; ATA guidelines, American Thyroid Association guidelines; AUROC, area under the receiver operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value; CI, confidence interval.

was 2.5 points, and that of the ACR-TIRADS was 6.5 points, which were different from ours. In their study, the AUROCs of the C-TIRADS and the ACR-TIRADS were 0.806 and 0.843, respectively, which were much lower than those of our study. Some previous studies claimed that the AUROC of the C-TIRADS was significantly higher than those of the ACR-TIRADS and the ATA guidelines (8,15). They set the cutoff value of the C-TIRADS, the ACR-TIRADS and the ATA guidelines at 2.5, 6.5 and 4.5 points, while we set the cutoff value at 1.5, 5.5 and 4.5 points, respectively. All the AUROCs of the three guidelines in their studies were lower than those of our study. We thought the differences in the diagnostic performance between their studies and our study might be associated with the population group being examined, the different cutoff values and the experience of the radiologists. We thought 1.5 points might be a better cutoff value for the C-TIRADS according to our results, which was consistent with the conclusion drawn by Hu *et al.* (10). In addition, some researchers have recently taken advantage of computer-aided diagnosis technology, which has no operator-dependent inference, to distinguish malignant thyroid nodules from benign thyroid nodules (16-18). The paper published by Bian *et al.* (16) showed that there was no significant difference in the AUROC between the TIRADS and the ultrasonic S-Detect model (0.94 *vs.* 0.91, P=0.19), and the ultrasonic S-Detect model performed well in malignancy risk stratification of thyroid nodules. The application of computer-aided diagnosis technology may eliminate subjective factors and help young radiologists predict malignancy risk accurately. All the studies showed

that the AUROCs of the C-TIRADS, ACR-TIRADS and ATA guidelines were high, and we could draw a conclusion that the three guidelines could effectively predict the malignancy risk of thyroid nodules.

Thyroid nodules are very common. However, the malignant rate of thyroid nodules is so low that unnecessary FNA should be avoided if possible. Some previous studies claimed that the ACR-TIRADS reduced unnecessary FNA with a higher malignancy detection rate compared with the ATA guidelines, which was consistent with our results (11,19). However, the paper published by Zhu *et al.* claimed that the fewest thyroid nodules were recommended for FNA according to the ACR-TIRADS, followed by those according to the ATA guidelines and the C-TIRADS (8). In their study, the unnecessary FNA rates of the C-TIRADS were lower than those of the ACR-TIRADS and the ATA guidelines. Our study found that the sensitivity, specificity, PPV, NPV and accuracy of FNA recommendations according to the C-TIRADS were higher than those according to the other two guidelines. The NPV of FNA recommendations according to the C-TIRADS was 92.28%, which was much higher than those according to the other two guidelines. Our results implied that the FNA recommendations of the C-TIRADS could be used to exclude malignancy effectively. On the other hand, lesions that were not recommended for FNA according to the C-TIRADS might be only followed up without biopsy. In our study, the unnecessary FNA rate of the C-TIRADS was even lower than that of the ACR-TIRADS in all nodules and one of the subgroups (≥20 mm), and we found that the

Table 3 Distribution of thyroid nodules among the malignancy risk stratification systems

Category	Nodule, n (%)	Malignancy, n (%)	Size (mm)			P value	No. of FNA (%)	Malignant detection rate of FNA (%)
			Total	Benign	Malignant			
C-TIRADS	1,228 (100.00)	292 (23.78)	25.00 (13.00, 35.00)	28.50 (18.25, 37.00)	12.00 (8.00, 18.00)	<0.001	632 (51.47)	38.92
5	21 (1.71)	20 (95.24)	12.00 (10.00, 14.00)	12.00[†]	12.00 (10.00, 14.50)	0.93	19 (90.48)	94.74
4C	215 (17.51)	179 (83.26)	11.00 (7.00, 17.00)	11.00 (8.00, 15.25)	10.00 (7.00, 17.00)	0.97	182 (84.65)	82.97
4B	124 (10.10)	62 (50.00)	13.00 (9.00, 21.75)	15.00 (10.75, 27.00)	12.00 (7.00, 17.00)	0.006	112 (90.32)	50.00
4A	367 (29.89)	24 (6.54)	27.00 (16.00, 37.00)	27.00 (16.00, 37.00)	26.50 (13.00, 38.50)	0.64	319 (86.92)	6.58
3	464 (37.79)	7 (1.51)	31.00 (23.00, 40.00)	31.00 (23.00, 40.00)	42.00 (30.00, 60.00)	0.04	0 (0.00)	0.00
2	37 (3.01)	0 (0.00)	33.00 (27.50, 38.00)	33.00 (27.50, 38.00)	[‡]	–	0 (0.00)	0.00
ACR-TIRADS	1,228 (100.00)	292 (23.78)	–	–	–	–	702 (57.17)	23.93
5	301 (24.51)	237 (78.74)	11.00 (8.00, 17.00)	12.00 (8.00, 18.75)	11.00 (8.00, 17.00)	0.32	185 (61.46)	77.30
4C	71 (5.78)	29 (40.85)	13.00 (9.00, 22.00)	14.00 (9.00, 26.25)	12.00 (6.50, 18.00)	0.13	28 (39.44)	32.14
4B	34 (2.77)	6 (17.65)	15.50 (9.75, 31.00)	20.50 (9.25, 34.00)	13.50 (11.25, 19.75)	0.56	18 (52.94)	11.11
4A	149 (12.13)	7 (4.70)	24.00 (7.00, 29.00)	24.50 (13.00, 32.25)	24.00 (13.00, 35.00)	0.48	106 (71.14)	4.72
3	512 (41.69)	10 (1.95)	30.00 (23.00, 40.00)	30.00 (23.00, 40.00)	46.00 (28.50, 56.25)	0.02	365 (71.29)	2.47
2	161 (13.11)	3 (1.86)	32.00 (24.50, 40.50)	32.00 (24.00, 40.00)	41.00[§]	0.20	0 (0.00)	0.00
ATA guidelines	1,228 (100.00)	292 (23.78)	–	–	–	–	995 (81.03)	18.39
High suspicion	319 (25.98)	257 (80.56)	11.00 (8.00, 16.00)	12.00 (8.00, 16.25)	11.00 (7.00, 16.00)	0.44	192 (60.19)	79.17
Intermediate suspicion	160 (13.03)	10 (6.25)	19.00 (10.25, 30.00)	19.00 (10.75, 30.00)	17.00 (8.50, 26.75)	0.44	129 (80.63)	5.43
Low suspicion	683 (55.62)	15 (2.20)	30.00 (23.00, 40.00)	30.00 (23.00, 40.00)	42.00 (29.00, 55.00)	0.02	629 (92.09)	2.23
Very low suspicion	11 (0.90)	0 (0.00)	37.00 (22.00, 39.00)	37.00 (22.00, 39.00)	[¶]	–	10 (90.91)	0.00
Benign	13 (1.06)	0 (0.00)	33.00 (24.50, 39.50)	33.00 (24.50, 39.50)	[§]	–	0 (0.00)	0.00
Not specified	42 (3.42)	10 (23.81)	22.50 (12.75, 30.25)	22.50 (10.25, 29.75)	24.00 (13.75, 31.75)	0.50	35 (83.33)	28.57

Data are presented as number (%) or median (interquartile range). [†], only one benign nodule was classified as C-TIRADS 5; [‡], no malignant nodule was classified as C-TIRADS 2; [§], only 3 malignant nodules were classified as ACR-TIRADS 2, and the interquartile range could not be calculated; [¶], no malignant nodule was classified as very low suspicion based on the 2015 ATA guidelines; [§], no malignant nodule was classified as benign based on the 2015 ATA guidelines; the malignant detection rates of FNA recommendations were calculated using the following equation: malignant detection rates of FNA recommendations = the number of malignant nodules among the recommended FNA nodules/the number of the recommended FNA nodules. C-TIRADS, Chinese Thyroid Imaging Reporting and Data System; ACR-TIRADS, Thyroid Imaging Reporting and Data System of the American College of Radiology; ATA guidelines, American Thyroid Association guidelines; FNA, fine needle aspiration.

unnecessary FNA rates of the three guidelines increased as the thyroid nodules became larger. We concluded that FNA recommendations according to the C-TIRADS might be more efficient than those according to the ACR-TIRADS and the ATA guidelines, especially among larger thyroid nodules. More unnecessary FNA could be avoided based on the FNA recommendations of the C-TIRADS. We thought the efficient FNA recommendations of C-TIRADS were associated with attention to high-suspicion nodules. In our study, the malignant nodules were significantly smaller than

Table 4 FNA recommendations of the 2020 C-TIRADS, 2017 ACR-TIRADS and 2015 ATA guidelines

Parameter	C-TIRADS	ACR-TIRADS	ATA guidelines	P value		
				C-TIRADS vs. ACR-TIRADS	C-TIRADS vs. ATA guidelines	ACR-TIRADS vs. ATA guidelines
Sensitivity (%)	84.25	57.53	62.67	<0.001	<0.001	0.21
Specificity (%)	58.76	42.94	13.25	<0.001	<0.001	<0.001
PPV (%)	38.92	23.93	18.39	<0.001	<0.001	0.006
NPV (%)	92.28	76.43	53.22	<0.001	<0.001	<0.001
Accuracy (%)	64.82	46.42	25.00	<0.001	<0.001	<0.001

C-TIRADS, Chinese Thyroid Imaging Reporting and Data System; ACR-TIRADS, Thyroid Imaging Reporting and Data System of the American College of Radiology; ATA guidelines, American Thyroid Association guidelines; AUROC, area under the receiver operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value; FNA, fine needle aspiration.

Table 5 Unnecessary FNA rates of the 2020 C-TIRADS, 2017 ACR-TIRADS and 2015 ATA guidelines

Guidelines	Nodules for recommending FNA (n)	Benign nodules (n)	Unnecessary FNA rate (%)	P value
Total				
C-TIRADS	632	386	61.08	<0.001*
ACR-TIRADS	702	534	76.07	0.006**
ATA guidelines	995	812	81.61	<0.001***
<20 mm				
C-TIRADS	309	121	39.16	0.15*
ACR-TIRADS	158	51	32.28	<0.001**
ATA guidelines	260	141	54.23	<0.001***
≥20 mm				
C-TIRADS	323	265	82.04	0.005*; <0.001****
ACR-TIRADS	544	483	88.79	0.14**; <0.001****
ATA guidelines	735	671	91.29	<0.001***; <0.001****

The P value reflects the difference in the unnecessary FNA rate: *, between C-TIRADS and ACR-TIRADS; **, between ACR-TIRADS and ATA; ***, between C-TIRADS and ATA; ****, between the two subgroups based on the same guidelines. The unnecessary FNA rates were calculated using the following equation: unnecessary FNA rate = the number of benign nodules among the recommended FNA nodules/ the number of recommended FNA nodules. C-TIRADS, Chinese Thyroid Imaging Reporting and Data System; ACR-TIRADS, Thyroid Imaging Reporting and Data System of the American College of Radiology; ATA guidelines, American Thyroid Association guidelines; FNA, fine needle aspiration.

the benign nodules. The C-TIRADS pays more attention to high-suspicion nodules that are relatively small and pays less attention to low-suspicion nodules, even though more efficient FNAs can be achieved with them because they are large, especially nodules classified as 4B, 4C or 5, which are recommended for FNA even if they are larger than 5 mm under certain conditions. The paper published by Lin *et al.* also indicated that the C-TIRADS showed an advantage in

reducing unnecessary FNA among patients with follicular neoplasms compared with the ACR-TIRADS and the ATA guidelines (20). However, more randomized controlled trials are needed to confirm our conclusion.

Our study had several limitations. First, all the included nodules were from patients in our hospital who were admitted to surgery, which could not perfectly represent the thyroid nodules in the social population. Second, we

chose histopathological results as the diagnostic criteria to improve diagnostic accuracy and excluded nodules that only had cytopathological results, which may lead to selection bias. Third, this study is a retrospective study rather than a randomized trial, which may lead to selection bias. Finally, the FNA results are influenced by many factors, such as nodule size and FNA skill. All the results of the FNAs in our paper were only calculated in theory.

Conclusions

In conclusion, the 2020 C-TIRADS, the 2017 ACR-TIRADS and the 2015 ATA guidelines can efficiently predict malignancy risk in thyroid nodules. Specifically, the C-TIRADS shows the highest sensitivity, specificity, PPV, NPV and accuracy in its FNA recommendations among the three guidelines and offers a reduction in unnecessary FNA and the highest efficacy in distinguishing thyroid carcinoma theoretically. However, the limitations of this study prevent us from reaching a confirmed and general conclusion. More multicenter RCTs with larger sample sizes, long-term follow-up and higher quality are needed to confirm and update the findings of our study.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the STARD reporting checklist (available at https://qims.amegroups.com/article/view/10.21037/qims-22-826/rc).

Conflicts of Interest: All the authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-22-826/coif). All authors report that this study was supported by the Science and Technology Planning Project of Shantou (No. 211114116491861). The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the ethics board of the Shantou Central Hospital [ID (2021) scientific research 057], and individual consent for this retrospective analysis was waived.

References

1. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, Pacini F, Randolph GW, Sawka AM, Schlumberger M, Schuff KG, Sherman SI, Sosa JA, Steward DL, Tuttle RM, Wartofsky L. 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.
2. Megwalu UC, Moon PK. Thyroid Cancer Incidence and Mortality Trends in the United States: 2000-2018. Thyroid 2022;32:560-70.
3. 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.
4. 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.
5. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, Cronan JJ, Beland MD, Desser TS, Frates MC, Hammers LW, Hamper UM, Langer JE, Reading CC, Scoutt LM, Stavros AT. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. J Am Coll Radiol 2017;14:587-95.
6. Zhou J, Yin L, Wei X, Zhang S, Song Y, Luo B, et al.

2020 Chinese guidelines for ultrasound malignancy risk stratification of thyroid nodules: the C-TIRADS. Endocrine 2020;70:256-79.

7. Chen Q, Lin M, Wu S. Validating and Comparing C-TIRADS, K-TIRADS and ACR-TIRADS in Stratifying the Malignancy Risk of Thyroid Nodules. Front Endocrinol (Lausanne) 2022;13:899575.

8. Zhu H, Yang Y, Wu S, Chen K, Luo H, Huang J. Diagnostic performance of US-based FNAB criteria of the 2020 Chinese guideline for malignant thyroid nodules: comparison with the 2017 American College of Radiology guideline, the 2015 American Thyroid Association guideline, and the 2016 Korean Thyroid Association guideline. Quant Imaging Med Surg 2021;11:3604-18.

9. Dong W, Wu Y, Cai T, Wang X. Comparison of diagnostic performance and FNA management of the ACR-TIRADS and Chinese-TIRADS based on surgical histological evidence. Quant Imaging Med Surg 2023;13:1711-22.

10. Hu Y, Xu S, Zhan W. Diagnostic performance of C-TIRADS in malignancy risk stratification of thyroid nodules: A systematic review and meta-analysis. Front Endocrinol (Lausanne) 2022;13:938961.

11. Peng JY, Pan FS, Wang W, Wang Z, Shan QY, Lin JH, Luo J, Zheng YL, Hu HT, Ruan SM, Liang JY, Xie XY, Lu MD. Malignancy risk stratification and FNA recommendations for thyroid nodules: A comparison of ACR TI-RADS, AACE/ACE/AME and ATA guidelines. Am J Otolaryngol 2020;41:102625.

12. Yoon JH, Lee HS, Kim EK, Moon HJ, Kwak JY. Malignancy Risk Stratification of Thyroid Nodules: Comparison between the Thyroid Imaging Reporting and Data System and the 2014 American Thyroid Association Management Guidelines. Radiology 2016;278:917-24.

13. Şahin M, Oguz A, Tuzun D, Akkus G, Törün GI, Bahar AY, Şahin H, Gül K. Effectiveness of TI-RADS and ATA classifications for predicting malignancy of thyroid nodules. Adv Clin Exp Med 2021;30:1133-9.

14. Moon HJ, Kwak JY, Kim EK, Kim MJ. A taller-than-wide shape in thyroid nodules in transverse and longitudinal ultrasonographic planes and the prediction of malignancy. Thyroid 2011;21:1249-53.

15. Qi Q, Zhou A, Guo S, Huang X, Chen S, Li Y, Xu P. Explore the Diagnostic Efficiency of Chinese Thyroid Imaging Reporting and Data Systems by Comparing With the Other Four Systems (ACR TI-RADS, Kwak-TIRADS, KSThR-TIRADS, and EU-TIRADS): A Single-Center Study. Front Endocrinol (Lausanne) 2021;12:763897.

16. Bian J, Wang R, Lin M. Ultrasonic S-Detect mode for the evaluation of thyroid nodules: A meta-analysis. Medicine (Baltimore) 2022;101:e29991.

17. Zhong L, Wang C. Diagnostic accuracy of S-Detect in distinguishing benign and malignant thyroid nodules: A meta-analysis. PLoS One 2022;17:e0272149.

18. Szczepanek-Parulska E, Wolinski K, Dobruch-Sobczak K, Antosik P, Ostalowska A, Krauze A, Migda B, Zylka A, Lange-Ratajczak M, Banasiewicz T, Dedecjus M, Adamczewski Z, Slapa RZ, Mlosek RK, Lewinski A, Ruchala M. S-Detect Software vs. EU-TIRADS Classification: A Dual-Center Validation of Diagnostic Performance in Differentiation of Thyroid Nodules. J Clin Med 2020.

19. Zhang WB, Xu HX, Zhang YF, Guo LH, Xu SH, Zhao CK, Liu BJ. 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;75:219-32.

20. Lin Y, Lai S, Wang P, Li J, Chen Z, Wang L, Guan H, Kuang J. Performance of current ultrasound-based malignancy risk stratification systems for thyroid nodules in patients with follicular neoplasms. Eur Radiol 2022;32:3617-30.