Is Chinese Thyroid Imaging Reporting and Data Systems superior to American College of Radiology or American Thyroid Association guidelines for consistency and efficacy in the diagnosis of thyroid cancer?

Na Li[1], Ying He[1], Ruijao Chang[2]

[1]Department of Ultrasound Medicine, 4th (Xing Yuan) Hospital of Yulin, Yulin, Shaanxi 719053, China;
[2]Department of Ultrasound Medicine, General Hospital of Ningxia Medical University, Yinchuan, Ningxia 750001, China.

To the Editor: The incidence of thyroid nodule has significantly increased in the past 30 years. As an important modality for thyroid screening and examination, ultrasound can be used to differentiate malignant from benign thyroid nodules. With the continuous improvement of the resolution of ultrasound, an increasing number of small nodules have been detected, especially nodules <1 cm. However, the misdiagnosis of thyroid nodules increases the incidence of unnecessary biopsies.[1] Therefore, the accurate distinction of malignant from benign thyroid nodules is essential to reduce the rate of unnecessary biopsy. At present, several thyroid imaging reporting and data systems (TI-RADSs) have been used to unify thyroid nodule reporting terms and provide recommendations for ultrasound examination. The American College of Radiology TI-RADS (ACR-TIRADS)[2] and American Thyroid Association (ATA)-2015 guide-lines[3] are widely used in China. In 2020, China issued the Chinese TI-RADS (C-TIRADS)[4] based on Chinese national and medical conditions. Although many studies have compared the TI-RADS guidelines, the interobserver agreement of C-TIRADS and interguideline agreement between C-TIRADS and other guidelines remain unclear. In addition, whether C-TIRADS, as a newly released guideline, is more accurate than ACR-TIRADS and the ATA guidelines in the diagnosis of thyroid nodules in the Chinese population has not been discussed. We compared the diagnostic efficacy and interobserver and interguideline agreement between C-TIRADS and two other guidelines in distinguishing thyroid cancer and to provide a basis for the clinical application of C-TIRADS.

This study was approved by the Ethics Committees of 4th (Xing Yuan) Hospital of Yulin and General Hospital of Ningxia Medical University according to the *Declaration of Helsinki*. Because the privacy information of patients was hidden, the requirement for informed consent was waived. A total of 1000 patients with 1211 lesions who underwent thyroid ultrasound examination in two centers from January 2017 to March 2021 were included in this retrospective study. Patients with at least one lesion, all nodules were confirmed by surgical pathology or core needle biopsy (CNB) pathology, were included in this study. We excluded patients with poor ultrasound image quality, unclear pathological result, previous treatment that may affect the determination of nodule features, and with nodules that could not be evaluated by the guidelines, especially for the ATA guidelines.

All ultrasound examinations were performed by radiologists in two hospitals with >5 years of experience. All nodules were evaluated by ultrasound machines, including Resona7 (Mindray, Shenzhen, China) with an L14-5 linear probe and an Oxana2 or S2000 ultrasonic system (Siemens AG, Erlangen, Germany) with an L9-4 linear transducer. Parameters, such as gain, focus, and depth, were properly adjusted to ensure good-quality images. Nodular features, such as maximum diameter, echogenicity, composition, shape, margin, internal echogenic foci, and abnormal cervical lymph nodes, were assessed and recorded.

If the nodules were suspected of malignancy or were large enough for surgical indications, CNB (performed by radiologists with >10 years of experience) or surgical resection (performed by surgeons with >15 years of experience) was performed at the patients' discretion. The

---

Correspondence to: Ruiiao Chang, MD, Department of Ultrasound Medicine, General Hospital of Ningxia Medical University, 804 Shengli Street, Xingqing District, Yinchuan, Ningxia 750001, China
E-Mail: changruijiao@qq.com

pathologic results (benign or malignant) were used as the gold standard.

The ultrasound images of all nodules were independently evaluated by two radiologists who were blinded to the pathology. The characteristics of nodules were recorded separately and compared by a third investigator. When they reached an agreement, the sonographic characteristics and classifications were determined and recorded. If the results were different, a consensus was achieved by discussing or consulting a specialist for suggestion. After the characteristics of the nodules were determined, all nodules were classified according to ACR-TIRADS,[2] C-TIRADS,[4] and the ATA guidelines.[3] Then, the nodules were reclassified according to the risk of malignancy, and the results of the classification were compared for interguideline agreement. Categories 1 and 2 in ACR-TIRADS (malignancy risk <2%) matched the "benign" and "very low suspicion" in the ATA guidelines (malignancy risk <3%) and categories 1, 2, and 3 in C-TIRADS (malignancy risk <2%). Category 3 in ACR-TIRADS (malignancy risk 5%) matched the "low suspicion" in the ATA guidelines (malignancy risk 5%–10%) and category 4A in C-TIRADS (malignancy risk 2%–10%). Category 4 in ACR-TIRADS (malignancy risk 5% –20%) matched the "intermediate suspicion" in the ATA guidelines (malignancy risk 10% –20%) and category 4B in C-TIRADS (malignancy risk 10% -50%). Category 5 in ACR-TIRADS (malignancy risk >20%) matched the "high suspicion" in the ATA guidelines (malignancy risk 70%–90%) and categories 4C, 5, and 6 in C-TIRADS (malignancy risk >50%). The receiver operating characteristic (ROC) curves of the guidelines were analyzed and compared, and the cut-off values were calculated based on the Youden index.

The mean $\pm$ standard deviation is used to describe the distribution characteristics of continuous variables conforming to normal distribution, and the t-test was used to compare the differences. Categorical variables are described as the frequency and percentage, and the chi-square test was used to compare the differences. Weighted kappa test was used to check the interguideline and interobserver agreement. The data were analyzed by using the statistical software SPSS 26.0 (IBM, Somers, NY, USA). The area under the ROC curve (AUC) was compared by MedCalc software (ver.19.5.6; Ostend, Belgium), and the sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), and 95% confidence intervals (CIs) were calculated. $P < 0.05$ was considered statistically significant. An adjusted $P < 0.016$ was considered statistically significant when comparing the three variables.

Among all patients enrolled in this study, 731 (73.10%) were males and 269 (26.90%) were females. The average age was 45.58 ± 11.78 years. There were 539 (44.50%) benign nodules, including 357 nodular goiters, 142 thyroid adenomas, and 40 localized Hashimoto's thyroiditis, and 672 (55.5%) malignant nodules, including 630 papillary thyroid carcinomas, 28 medullary thyroid carcinomas, 13 follicular carcinomas, and 1 squamous cell carcinoma. The mean maximum diameter of nodules was 1.46 ± 1.30 cm. There were statistically significant differences in age between patients with benign (48.58 ± 11.97 years) and malignant nodules (43.17 ± 11.06 years) ($P < 0.001$) and in maximum diameter between patients with benign (1.92 ± 1.50 cm) and malignant nodules (1.09 ± 0.97 cm) ($P < 0.001$).

In C-TIRADS, the interobserver agreement of classification was almost excellent agreement, with a Kappa value of 0.824 (95% CI: 0.797, 0.851), which was better than that of the ATA guidelines and ACR-TIRADS, with Kappa values of 0.714 (95% CI: 0.675, 0.753) and 0.798 (95% CI: 0.767, 0.829) (both classified into substantial agreement), respectively.

As a newly issued guideline, C-TIRADS was evaluated for interguideline agreement with the other two guidelines. The interguideline agreement between C-TIRADS and ACR-TIRADS was moderate, with a Kappa value of 0.627, higher than that between C-TIRADS and the ATA guidelines, with a Kappa value of 0.494 (fair agreement).

ROC curves were plotted based on the classifications of the three guidelines. For ACR-TIRADS, the ATA guidelines, and C-TIRADS, the AUROCs were 0.782 (95% CI: 0.758, 0.805), 0.737 (95% CI: 0.711, 0.761), and 0.846 (95% CI: 0.824, 0.866), respectively. The AUROCs of these three guidelines were significantly different from each other. Based on the Youden index in ROC curve analysis, the cut-off values for ACR-TIRADS, the ATA guidelines, and C-TIRADS were determined to be ACR-TR5, ATA high suspicion, C-TIRADS 4C.

For ACR-TIRADS in malignant nodules, the sensitivity was 89.43% (601/672) (95% CI: 0.8680, 0.9161), the specificity was 63.08% (340/539) (95% CI: 0.5883, 0.6714), the accuracy was 77.70% (941/1211) (95% CI: 0.7527, 0.7996), the PPV was 75.13% (601/800) (95% CI: 0.7195, 0.7806), and the NPV was 82.73% (340/411) (95% CI: 0.7864, 0.8618).

For the ATA guidelines in malignant nodules, the sensitivity was 96.73% (650/672) (95% CI: 0.9500, 0.9789), the specificity was 49.72% (268/539) (95% CI: 0.4543, 0.5402), the accuracy was 75.81% (918/326) (95% CI: 0.7331, 0.7813), the PPV was 70.58% (650/921) (95% CI: 0.6750, 0.7348), and the NPV was 92.41% (268/290) (95% CI: 0.8858, 0.9508).

For C-TIRADS in malignant nodules, the sensitivity was 84.08% (565/672) (95% CI: 0.8104, 0.8672), the specificity was 78.85% (425/539) (95% CI: 0.7511, 0.8217), the accuracy was 81.75% (990/ 1211) (95% CI: 0.7947, 0.8383), the PPV was 83.21% (565/679) (95% CI: 0.8014, 0.8590), and the NPV was 79.89% (425/532) (95% CI: 0.7617, 0.8316).

TI-RADS were first proposed and used in the clinic in 2009. Since then, various TI-RADS guidelines for thyroid ultrasound have been applied and compared, but their indications and diagnostic efficacy remain controversial. In this study, C-TIRADS was first introduced to compare the consistency and diagnostic efficacy with ACR-TIRADS and the ATA guidelines in the Chinese population. The results

showed that the interobserver agreement and the diagnostic efficacy of C-TIRADS were better than that of ACR-TIRADS and the ATA guidelines.

Although only two observers participated in the image interpretation, the interobserver agreement was also of certain clinical value. Our results showed that the interobserver agreement of C-TIRADS was better than that of ACR-TIRADS and the ATA guidelines. The reason may be that only vertical orientation, solid composition, markedly hypoechoic, microcalcifications, and irregular margin or extrathyroidal extension were included in C-TIRADS. Meanwhile, among nodules that needed discussion or were identified by the third observer, the interobserver agreement was poor only when judging the margin of the nodule. Other features, such as shape, composition, echo, and calcification, were less controversial between the two observers. In the interguideline analysis, the result showed that C-TIRADS had better agreement with ACR-TIRADS than with the ATA guidelines, which may be due to the similar malignant features in their classification criteria, and the definition and weight of malignant characteristics are different in different guidelines.

Diagnostic efficacy is the universally acknowledged standard for evaluating diagnostic guidelines. Our results showed that C-TIRADS had the best diagnostic efficacy, including the highest specificity, which may be helpful to reduce unnecessary thyroid nodule biopsy. Although the ATA guidelines had higher sensitivity, the lower specificity may increase the probability of unnecessary thyroid nodule biopsy and even unnecessary treatment. Previous studies have shown that the ATA guidelines have a higher rate of unnecessary biopsies than the ACR guidelines,[1] which is similar to the results of this study. The ATA guidelines define a nodule as highly suspicious when there are malignant features without considering multiple malignancy risks combined, which is not appropriate for thyroid cancer with a relatively low mortality rate. Before the advent of C-TIRADS, ACR-TIRADS was considered b[5y] many studies to have the best clinical diagnostic value.[5] However, in our study, C-TIRADS showed better diagnostic performance than ACR-TIRADS. Differences in malignant features and weights may lead to differences in diagnostic efficacy. In ACR-TIRADS, the malignant weights vary greatly, whereas in C-TIRADS, the weight differences are small, and there is a benign characteristic with negative weight. Moreover, the weights of features, including hypoechoic, peripheral, or macrocalcification and mixed components, were excluded. Adjusting the malignant features and their weights may be conducive to improve the diagnostic efficiency.

This study is a retrospective study, and the reviewed data may be inconsistent in image storage. In addition, only subjectively suspected malignant nodules that required biopsy and benign nodules that required surgery were included. Thus, there may be a selection bias. Moreover, only two radiologists were included in analyzing interobserver agreement, further verification is needed for the result.

In conclusion, C-TIRADS has good interobserver agreement and accuracy in the diagnosis of thyroid nodules in the Chinese population. It may have better application prospects in reducing overdiagnosis and overtreatment and provide a basis for ultrasonic radiologists to diagnose thyroid nodules.

Acknowledgments

Conflicts of interest

None.

References

1. Kim PH, Suh CH, Baek JH, Chung SR, Choi YJ, Lee JH. Unnecessary thyroid nodule biopsy rates under four ultrasound risk stratification systems: a systematic review and meta-analysis. Eur Radiol 2021;31:2877–2885. doi: 10.1007/s00330-020-07384-6.
2. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS committee. J Am Coll Radiol 2017;14:587–595. doi: 10.1016/j.jacr.2017.01.046.
3. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on thyroid nodules and differentiated thyroid cancer. Thyroid 2016;26:1–133. doi: 10.1089/thy.2015.0020.
4. 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;70:256–279. doi: 10.1007/s12020-020-02441-y.
5. Peng JY, Pan FS, Wang W, Wang Z, Shan QY, Lin JH, et al. Malignancy risk stratification and FNA recommendations for thyroid nodules: a comparison of ACR TI-RADS, AACE/ACE/AME and ATA AQ4 guidelines. Am J Otolaryngol 2020;41:102625. doi: 10.1016/j.amjoto.2020.102625.