



Fine needle aspiration biopsy indications for thyroid nodules: compare a point-based risk stratification system with a pattern-based risk stratification system

Jing-liang Ruan^{1,2} · Hai-yun Yang^{1,2} · Rong-bin Liu^{1,2} · Ming Liang^{1,2} · Ping Han^{2,3} · Xiao-lin Xu^{1,2} · Bao-ming Luo^{1,2}

Received: 12 October 2018 / Accepted: 27 November 2018 / Published online: 4 February 2019
© European Society of Radiology 2019

Abstract

Objectives We aim to compare the diagnostic performance to assess thyroid nodules and reliability for recommending fine needle aspiration biopsy (FNAB) between American College of Radiology thyroid image reporting and data system (ACR TI-RADS) and American Thyroid Association (ATA) guidelines.

Methods In total, this retrospective study included 1001 consecutive thyroid nodules in 918 patients from May 2016 to December 2017. US features of the thyroid nodules, including composition, echogenicity, shape, margins, echogenic foci, and size, were reviewed and were classified according to ACR TI-RADS and ATA guidelines, respectively. The diagnostic performance to assess thyroid nodules and reliability for recommending fine needle aspiration biopsy were compared between ACR TI-RADS and ATA guidelines.

Results Of the 1001 thyroid nodules, 609 (60.8%) were benign and 392 (39.2%) were malignant. The sensitivity, specificity, PPV, NPV, and accuracy were 96.7%, 77.3%, 73.3%, 97.3%, and 84.9%, respectively, for ACR TI-RADS and 99.2%, 16.1%, 43.2%, 97.0%, and 48.7%, respectively, for ATA guidelines. AUC of ACR TI-RADS was significantly greater than ATA guidelines (0.935 (0.918, 0.949) vs 0.884 (0.862, 0.903), $p < 0.001$). Biopsy yield of malignancy, biopsy rate of malignancy, and unnecessary FNAB rate were 59.5%, 91.3%, and 40.5%, respectively, for ACR TI-RADS and 38.5%, 97.4%, and 61.5%, respectively, for ATA guidelines.

Conclusions ACR TI-RADS was more accurate than ATA guidelines for differentiating malignant thyroid nodules from benign nodules and more reliable than ATA guidelines for recommending thyroid nodules for FNAB.

Key Points

- Malignant risk of thyroid nodules can be stratified by ultrasound.
- American College of Radiology guidelines were more accurate for differentiating malignant thyroid nodules from benign nodules.
- American College of Radiology guidelines were more reliable for recommending thyroid nodules for biopsy.

Keywords Fine needle aspiration biopsy · Biopsy; Guidelines · Thyroid nodules · Ultrasound

Abbreviations and acronyms

ACR American College of Radiology
ATA American Thyroid Association

AUC Area under the curve
FNAB Fine needle aspiration biopsy
NPV Negative predictive value

Jing-liang Ruan and Hai-yun Yang contributed equally to this work and both should be considered as joint first authors.

✉ Xiao-lin Xu
xuxiaolin@mail.sysu.edu.cn

✉ Bao-ming Luo
luobm@mail.sysu.edu.cn

² Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, China

³ Department of Otolaryngology, Head and Neck Surgery, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, China

¹ Department of Ultrasound, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, No. 107 Yanjiang Road West, Guangzhou 510120, China

PPV	Positive predictive value
PTC	Papillary thyroid cancer
TI-RADS	Thyroid image reporting and data system
US	Ultrasound

Introduction

Due to the increase in the detection of small papillary thyroid cancers (PTCs), thyroid cancer incidence has nearly increased threefold over a 35-year period in the USA and 15-fold over an 18-year period in Korea, while its mortality has remained stable [1, 2]. The problem is due to the overdiagnosis and overtreatment of PTC [1, 3, 4]. Compared with radiologists' own practice patterns, risk stratification system led to a reduction in the number of nodules recommended for fine needle aspiration biopsy (FNAB) [5]. There were several thyroid nodule malignant risk stratification systems from different international societies to aid in making decision about the use of FNAB [6–9]. On the whole, we can divide these risk stratification systems into two categories of point-based system and pattern-based system according their stratified methods. They can be represented by American College of Radiology Thyroid Imaging Reporting and Data System (ACR TI-RADS) and American Thyroid Association (ATA) guidelines, respectively [7, 9]. Although each of these guidelines has been proved to be helpful in the management of thyroid nodules, no evidence-based guidance is available for which system is best [10–13]. Therefore, rigorous analysis of their diagnostic performance and reliability for recommending FNAB is essential for standardizing the diagnosis and treatment of thyroid cancer and reduce unnecessary and excessive FNAB.

The primary aim of this study was to compare the diagnostic performance for differentiating benign and malignant thyroid nodules between ACR TI-RADS and ATA guidelines. The secondary aim was to compare the reliability for recommending FNAB and reducing unnecessary FNAB for thyroid nodules between two guidelines.

Materials and methods

Study population

From May 2016 to December 2017, 1898 nodules from 1526 consecutive patients had undergone US-guided FNAB in a university hospital. Nodules with following inclusion criteria were included: (a) definitive diagnostic cytologic finding of benignity or malignancy at US-guided FNAB, (b) inconclusive cytologic findings at initial US-guided FNAB but definitive cytologic findings of benignity or malignancy at repeat US-guided FNAB, or (c) underwent surgery. During US-guided FNAB, a thyroid nodule with the most suspicious

US features, if no suspicious US features presented, we selected the largest nodule as biopsied target. Repeat FNAB was usually performed in nodules with prior inconclusive cytopathology. Exclusion criteria were final inconclusive pathology (860 nodules) and US features could not be extracted (37 nodules). Finally, a total of 1001 thyroid nodules with final definite pathology in 918 consecutive patients were included to make a whole database in this study (562 females, 356 males; mean age, 45.7 years; range 14–78 years). This retrospective study was approved by the institutional research ethics committee. The requirement to obtain informed consent was waived.

US examination and image analysis

All US examinations were performed using a variety of commercially available real-time US system. US images with the largest diameter and corresponding orthogonal images, in some cases of necessary, video clips of all biopsied nodules were obtained and stored on hard disk to create a database. Attention was focused on acquisition of optimal imaging to allow evaluation of US features of nodules, such as composition, echogenicity, shape, margins, and echogenic foci.

US images were reviewed by four board-certified radiologists who were experienced in thyroid sonography. To standardize the interpretations, 50 extra cases of nodules were reviewed together by all of the reviewers before the study. All of reviewers, who were blinded to the FNAB result and final diagnosis, evaluated US features of each biopsied nodule on corresponding US images on the basis of a previous report. The composition was categorized as follows: cystic or almost entirely cystic, spongiform, mixed solid and cystic, and solid or almost entirely solid. Echogenicity was categorized as anechoic, isoechoic, hyperechoic, hypoechoic, and very hypoechoic. Shape was categorized as wider than tall and taller than wide. Margins were categorized as smooth, ill-defined, and lobulated or irregular. Echogenic foci were categorized as small non-shadowing echogenic foci with comet-tail artifacts, macrocalcifications (calcifications with shadowing), peripheral calcifications (complete and partial), and punctate echogenic foci (tiny bright reflectors without shadowing). After US features were extracted, two radiologists classified each nodule as high suspicion, intermediate suspicion, low suspicion, or very low suspicion or benign using ATA guidelines, and the other two radiologists classified each nodule as highly suspicious, moderately suspicious, mildly suspicious, not suspicious, or benign on the basis of the total point obtained from each nodule US features using ACR TI-RADS [7, 9].

Statistical analyses

Cytopathology from US-guided FNAB and histopathology from surgery were considered as reference standard.

Continuous variables were compared using unpaired *t* test. Categorical variables were compared using chi-squared and Fisher's exact tests where appropriate. Cochran-Mantel-Haenszel test was used to evaluate the relationship between the categories of each stratification system and the final diagnosis. Receiver operating characteristic curve was used to compare the diagnostic performance between ACR TI-RADS and ATA guidelines.

In order to calculate the diagnostic performance of the two stratification systems, TR1 to TR3 nodules were considered negative and TR4 to TR5 nodules were considered positive with ACR TI-RADS. Benign and very low suspicion nodules were considered negative, and low suspicion to high suspicion nodules were considered positive with ATA guidelines. Unclassified nodules with ATA guidelines were considered positive because most of these nodules had suspicious US features, such as micro-lobulated or speculated margins, micro-calcifications, or taller than wide.

In order to compare the reliability of the US-based FNAB indication between ACR TI-RADS and ATA guidelines, thyroid nodules were evaluated to whether that system recommends FNA based on category and size thresholds [7, 9]. Considering the unclassified nodules for ATA guidelines, calculation was under the assumption that these nodules would be biopsied at a size threshold of 1.0 cm or greater. The biopsy yield of malignancy was defined as a percentage of the number of malignant nodules biopsied among total number of biopsy-indicated nodules. The biopsy rate of malignancy was defined as a percentage of the number of malignant nodules biopsied among total number of malignant nodules. The unnecessary biopsy rate was defined as a percentage of the number of benign nodules among biopsy-indicated nodules in the total nodules. Subgroup analysis was performed according to nodules size: 1.0~1.9 cm, 2.0~2.9 cm, and ≥ 3 cm, respectively.

Statistical analyses were performed by using SAS for Windows (version 9.4; SAS Institute). A significant difference was defined as $p < 0.05$.

Results

Of the 1001 thyroid nodules, 609 (60.8%) were benign and 392 (39.2%) were malignant. Of the 359 malignant nodules that were diagnosed via surgical resection, there were 342 PTCs, including 8 follicular variant papillary thyroid carcinomas, seven follicular carcinomas, two medullar carcinomas, two undifferentiated carcinomas, and six other malignant tumors (three lymphomas, one carcinoma showing thymus-like differentiation, one metastasis, and one squamous cell carcinoma). Among the 222 benign nodules that were diagnosed via surgical resection, there were 124 nodular goiters, 87 follicular adenomas, and 11 cases of thyroiditis. For the rest of

included nodules of our database, the final diagnosis was based on definite cytopathology by initial FNAB ($n = 335$) or by repeat FNAB ($n = 85$) after initial FNAB with inconclusive cytopathology.

Demographic of the patients and US features of thyroid nodules are summarized in Table 1. The 609 patients with benign thyroid nodules were significantly older than the 392 patients with malignant nodules ((mean, 47.7 years \pm 12.4; range 14–77 years) vs (mean, 40.7 years \pm 12.2; range 16–78 years), $p < 0.001$). Malignant thyroid nodules were significantly smaller than benign nodules ((mean, 14.5 mm \pm 9.4; range, 5–78 mm) vs (mean, 20.4 mm \pm 13.4; range, 7–79 mm), $p < 0.001$). Malignant nodules had significantly higher rates of solid composition, hypo-echogenicity or marked hypo-echogenicity, taller than wide shape, lobulated or irregular margins, and micro-calcifications than benign nodules ($p < 0.05$ for all). Using ACR TI-RADS, we classified all 1001 nodules into 5 categories. There were significant differences of malignancy rates among categories ($p < 0.001$) (Table 2). Using ATA guidelines, we could not classify 6.6% (66/1001) nodules into any category and thus we classified these nodules into unclassified. There were significant differences of malignancy rates among categories ($p < 0.001$) (Table 2). Area under the curve of ACR TI-RADS was greater than that of ATA guidelines (0.935 vs 0.884, $p < 0.001$) (Table 3).

In the overall group, the biopsy yield of malignancy, biopsy rate of malignancy, and unnecessary biopsy rate were 59.5%, 91.3%, and 40.5%, respectively, for ACR TI-RADS and 38.5%, 97.4%, and 61.5%, respectively, for ATA guidelines. In the subgroup of 1.0~1.9 cm, the biopsy yield of malignancy, biopsy rate of malignancy, and unnecessary biopsy rate were 81.1%, 88.7%, and 18.9%, respectively, for ACR TI-RADS and 57.4%, 96.0%, and 42.6%, respectively, for ATA guidelines. In the subgroup of 2.0~2.9 cm, the biopsy yield of malignancy, biopsy rate of malignancy, and unnecessary biopsy rate were 56.5%, 96.0%, and 43.5%, respectively, for ACR TI-RADS and 27.2%, 100.0%, and 72.8%, respectively, for ATA guidelines. In the subgroup of ≥ 3.0 cm, the biopsy yield of malignancy, biopsy rate of malignancy, and unnecessary biopsy rate were 27.5%, 96.6%, and 72.5%, respectively, for ACR TI-RADS and 20.1%, 100.0%, and 79.9%, respectively, for ATA guidelines (Table 4).

Discussion

Overdiagnosis and overtreatment of thyroid cancer have brought about serious problems. It had increased thyroid cancer patients with surgical complications, including hypoparathyroidism and vocal cord paralysis [4]. It also had expanded societal medical cost for thyroid cancer [14]. Thyroid nodule risk stratification system was developed

Table 1 Summary of demographic and US features

Parameter	Final diagnosis		Total	Malignancy rate (%)	<i>p</i> value
	Benign	Malignant			
No. of nodules	609	392	1001		
No. of patients	542	376	918		
Age (year)					< 0.001
Mean	47.7 ± 12.4	40.7 ± 12.2			
Range	14–77	16–78			
Gender					0.002
Male	106	100			
Female	503	292			
Size (mm)					< 0.001
Mean	20.4 ± 13.4	14.5 ± 9.4			
Range	7–79	5–78			
Composition					< 0.001
Predominantly cystic	7	0	7	0.0	
Predominantly solid	249	9	258	3.5	
Solid	353	383	736	52.0	
Echogenicity					< 0.001
Anechoic	7	0	7	0.0	
Hyper- to isoechoic	496	71	567	12.5	
Hypoechoic	92	266	358	74.3	
Markedly hypoechoic	14	55	69	79.7	
Shape					< 0.001
Wider than tall	585	259	844	30.6	
Taller than wide	24	133	157	84.7	
Margin					< 0.001
Smooth or ill-defined	577	91	668	13.6	
Lobulated or irregular	32	301	333	90.4	
Echogenic foci					< 0.001
None or large comet-tail artifact	501	113	614	18.4	
Macro-calcifications	68	24	92	26.1	
Peripheral calcifications	16	14	30	46.7	
Punctate echogenic foci	24	241	265	90.6	

to minimize potential harm from overuse of FNAB [7, 9, 15–17]. However, different stratification systems have recommended different criteria for suspicious US patterns and size thresholds for FNAB. The reliability of US-based FNAB criteria are not well understood. In order to reduce unnecessary FNAB, we retrospectively included 1001 thyroid nodules of 918 patients to compare the reliability of FNAB indications between ACR TI-RADS and ATA guidelines. Diagnostic performance of ACR TI-RADS was greater than that of ATA guidelines. Biopsy yield of malignant of ACR TI-RADS was greater than that of ATA guidelines and unnecessary biopsy rate of ACR TI-RADS was lower than that of ATA guidelines. However, biopsy rate of malignancy of ACR TI-RADS was lower than that of ATA guidelines.

An ideal guideline should recommend as many malignant nodules as possible for FNAB while minimizing unnecessary biopsy. Unnecessary FNAB not only is invasive and costly but also leads to repeated FNAB, core needle biopsy, or open surgical biopsy for diagnosis due to initial nondiagnostic FNAB [18]. Data of Hoang et al showed that ACR TI-RADS recommended less nodules for biopsy than other biopsy guidelines. Most of the nodules spared from biopsy with ACR TI-RADS were benign nodules [5]. We not only confirmed these results in the overall group but also reconfirmed these results in the subgroup analysis of different sizes of thyroid nodules. In the present study, the biopsy yield of malignancy of ACR TI-RADS (59.5%) was greater than that of ATA guidelines (38.5%) and unnecessary biopsy rate of ACR TI-RADS

Table 2 Comparison of malignancy rates with ACR TI-RADS and ATA guidelines

Guidelines and categories	Final diagnosis		Recommended malignancy risk (%)	Calculated malignancy rate (%)	p value
	Benign	Malignant			
ACR TI-RADS					
Highly Suspicious	32 (5.3)	307 (78.3)	> 20	90.6	< 0.001
Moderately Suspicious	106 (17.4)	72 (18.4)	5–20	40.4	
Mildly Suspicious	281 (46.1)	9 (2.3)	5	3.1	
Not Suspicious	183 (30.1)	4 (1.0)	< 2	2.1	
Benign	7 (1.2)	0 (0.0)	< 2	0.0	
ATA guidelines					
High suspicion	36 (5.9)	301 (76.8)	> 70–90	89.3	< 0.001
Intermediate suspicion	38 (6.2)	19 (4.9)	10–20	33.3	
Low suspicion	408 (67.0)	32 (8.2)	5–10	7.3	
Very low suspicion	91 (14.9)	3 (0.8)	< 3	3.2	
Benign	7 (1.2)	0 (0.0)	< 1	0.0	
Unclassified	29 (4.8)	37 (9.4)	Not mentioned	56.1	

(40.5%) was lower than that of ATA guidelines (61.5%). Those exact reasons are unclear. But we can analyze the reasons from these two aspects. On the one hand, management of nodules that are highly likely to benign was different between the two guidelines. ACR TI-RADS does not recommend these nodules for FNAB. However, ATA guidelines considers FNAB at a size threshold of 2 cm or larger. On the other hand, size threshold for biopsy of low suspicion nodules and intermediate suspicion nodules is inconsistent. Size threshold of ACR TI-RADS (≥ 2.5 cm and ≥ 1.5 cm, respectively) is higher than that of ATA guidelines (≥ 1.5 cm and ≥ 1.0 cm, respectively).

Biopsy rate of malignancy of ACR TI-RADS was lower than that of ATA guidelines. This result was unexpected but was acceptable. There were two main reasons for this result. For one thing, there were some nodules that had benign US features. For another thing, ACR TI-RADS recommended many highly suspicious nodules < 1 cm for follow-up. Because of indolent behavior of papillary microcarcinoma, active surveillance is performed in Japan and has been acknowledged in USA.

For nodules that do not meet the criteria for biopsy, there were comprehensive indications in ACR TI-RADS regarding follow-up. For a highly suspicious nodule, ACR TR-RADS recommend follow-up annually for up to 5 years. For a moderately suspicious nodule, follow-up should be done at 1, 2, 3, and 5 years. For a mildly suspicious nodule, follow-up may be performed at 1, 3, and 5 years. According to ATA guidelines, high suspicion or intermediate suspicion nodule that did not meet the criteria for biopsy should be followed up. For low suspicion nodule, follow-up may be warranted until the size is 1.5 cm. However, there was little consensus about follow-up intervals.

Result of Ha et al showed that ACR TI-RADS was more accurate for differentiating benign and malignant thyroid nodules. In their cohort, sensitivity, specificity, PPV, NPV, and accuracy were 74.7%, 67.3%, 40.2%, 90.1%, and 69.0%, respectively, for ACR TI-RADS and 89.6%, 33.2%, 28.3%, 91.6%, and 46.0%, respectively, for ATA guidelines [19]. In the present study, we also confirmed that ACR TI-RADS had significantly greater diagnostic performance over ATA guidelines. The greater diagnostic performance of ACR TI-RADS might attribute to

Table 3 Diagnostic performance of ACR TI-RADS and ATA guidelines

Parameter	ACR TI-RADS	ATA guidelines
Sensitivity (%)	96.7 (379/392) [94.9, 98.5]	99.2 (389/392) [98.4, 100.0]
Specificity (%)	77.3 (471/609) [74.0, 80.7]	16.1 (98/609) [13.2, 19.0]
PPV (%)	73.3 (379/517) [69.5, 77.1]	43.2 (389/900) [40.0, 46.5]
NPV (%)	97.3 (471/484) [95.9, 98.8]	97.0 (98/101) [93.7, 100.0]
Accuracy (%)	84.9 (850/1001) [82.7, 87.1]	48.7 (487/1001) [45.6, 51.8]
AUC	0.935 (0.918, 0.949)	0.884 (0.862, 0.903)
p value	< 0.001	

Table 4 Reliability for recommending FNAB of ACR TI-RADS and ATA guidelines

Parameters	No. of biopsy nodules	Biopsy yield of malignancy (%)	Biopsy rate of malignancy (%)	Unnecessary biopsy rate (%)
All				
ACR	351	59.5 (209/351)	91.3 (209/229)	40.5 (142/351)
ATA	579	38.5 (223/579)	97.4 (223/229)	61.5 (356/579)
1.0–1.9 cm				
ACR	164	81.1 (133/164)	88.7 (133/150)	18.9 (31/164)
ATA	251	57.4 (144/251)	96.0 (144/150)	42.6 (107/251)
2.0–2.9 cm				
ACR	85	56.5 (48/85)	96.0 (48/50)	43.5 (37/85)
ATA	184	27.2 (50/184)	100.0 (50/50)	72.8 (134/184)
≥3.0 cm				
ACR	102	27.5 (28/102)	96.6 (28/29)	72.5 (74/102)
ATA	144	20.1 (29/144)	100.0 (29/29)	79.9 (115/144)

allocation of points for each US characteristic which reflected the likelihood of malignancy. ACR TI-RADS was a “point-based” stratification system. First, points are given for five US features, including composition, echogenicity, shape, margins, and echogenic foci. Second, more suspicious feature was awarded additional points. Finally, total points determined the nodule’s ACR TI-RADS level.

Previous data of Middleton et al and Ha et al showed that ACR TI-RADS was able to classify all the included 3422 nodules and 2000 nodules, respectively [19, 20]. We confirmed this result again by using a Chinese cohort with higher rate of malignant nodules in a university hospital. However, our data showed that 6.6% (66/1001) nodules could not be classified using ATA guidelines. The malignancy rate of these nodules was 56.1%. This result was consistent with previous studies. Yoon et al and Ha et al reported that ATA guidelines were unable to classify 3.4% (44/1293) nodules and 5.0% (100/2000) nodules. The malignancy rates of these nodules were 18.2% and 19.0%, respectively [19, 21]. The reason why ATA guidelines could not classify every nodule contained two aspects. For one thing, although the important US features could be assessed, but the combination of US features was not included in ATA guidelines. For another thing, the important US features could not be determined by using ATA guidelines, for instance, macro-calcification with resulting shadowing. However, ACR TI-RADS solved this problem by specific instruction advised how to allocate points if an US characteristic could not be determined. For instance, assign 2 points if composition cannot be determined because of calcification, assign 1 point if echogenicity cannot be determined, and assign 0 point if margin cannot be determined.

We validated risk in each category of both guidelines. Basically, calculated malignancy rates according to each category of both guidelines were within the recommended malignancy risk interval. However, calculated malignancy rate of moderately suspicious nodules of ACR TI-

RADS and intermediate suspicion nodules of ATA guidelines were relatively higher than the recommended malignancy risk. These unexpected results might due to our database included a higher number of malignant nodules that likely had a higher frequency of suspicious US features.

Both of ACR TI-RADS and ATA guidelines not only provided uniform standards for the interpretation of thyroid nodule sonograms but also provide terminology to communicate more effectively with pathologists, thyroid surgeon, and endocrinologists. However, ACR TI-RADS was relatively easy to apply across a wide gamut of ultrasound practice. First, ACR TI-RADS facilitated comparison of examinations from different US practitioners and institutions because it forced the US practitioner who was obtaining and interpreting the thyroid images to concentrate on each of the five important US features. Second, it could be readily integrated into reporting templates. By implementing an ACR TI-RADS-structured reporting template, Griffin et al improved the quality of thyroid ultrasound reports that led to better description of US features and improved the number of reports with definitive management recommendations [22].

Our study has several limitations. First of all, our data set included only thyroid nodules that had undergone FNAB. These nodules might have a disproportionately higher frequency of suspicious US features. Therefore, selection bias may exist. Second, a relatively high malignant rate of included thyroid nodules was present in our study, because the present study took place in a university hospital. The malignant risk of some categories may be affected by the high malignancy rate of the study population. Third, we did not evaluate the inter-observer agreement in the imaging analysis, because there was inconsistency in subclassification of five important US features between the two guidelines. For instance, the important US feature of calcification was subdivided into rim calcification and micro-calcification by ATA guidelines instead of

macro-calcification, rim calcification, and punctate calcification by ACR TI-RADS. Fourth, only 222 of 609 benign thyroid nodules were confirmed by surgical histologic finding. The rest of benign thyroid nodules were confirmed by cytopathologic finding. Although benign thyroid nodules on cytology required no further immediate diagnostic studies, initial or repeated FNAB may miss some malignancy [9, 23]. This may cause false negative and false positive.

Conclusions

In summary, our results indicate that ACR TI-RADS was more accurate in differentiating malignant thyroid nodules from benign nodules and more reliable in recommending thyroid nodules for FNAB. We believe that constant improvement of guidelines will promote the standardization of diagnosis and treatment of thyroid nodules and reduces unnecessary excessive FNAB.

Funding This study has received funding by Natural Science Foundation of Guangdong Province (2016A030313306) and National Natural Science Foundation of China (81671704).

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Prof. Bao-ming Luo.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products of services may be related to the subject matter of the article.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology

- retrospective
- diagnostic or prognostic
- performed at one institution

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Davies L, Welch HG (2014) Current thyroid cancer trends in the United States. *JAMA Otolaryngol Head Neck Surg* 140:317–322
2. Ahn HS, Kim HJ, Welch HG (2014) Korea's thyroid-cancer "epidemic"—screening and overdiagnosis. *N Engl J Med* 371:1765–1767
3. Oda H, Miyauchi A, Ito Y et al (2017) Comparison of the costs of active surveillance and immediate surgery in the management of low-risk papillary microcarcinoma of the thyroid. *Endocr J* 64:59–64
4. Griffin A, Brito JP, Bahl M, Hoang JK (2017) Applying criteria of active surveillance to low-risk papillary thyroid cancer over a decade: how many surgeries and complications can be avoided? *Thyroid* 27:518–523
5. Hoang JK, Middleton WD, Farjat AE et al (2018) Reduction in thyroid nodule biopsies and improved accuracy with American College of Radiology Thyroid Imaging Reporting and Data System. *Radiology* 287:185–193
6. Shin JH, Baek JH, Chung J et al (2016) Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean Society of Thyroid Radiology consensus statement and recommendations. *Korean J Radiol* 17:370–395
7. Tessler FN, Middleton WD, Grant EG et al (2017) ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. *J Am Coll Radiol* 14:587–595
8. Frates MC, Benson CB, Charboneau JW et al (2005) Management of thyroid nodules detected at US: Society of Radiologists in Ultrasound consensus conference statement. *Radiology* 237:794–800
9. Haugen BR, Alexander EK, Bible KC et al (2016) 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* 26:1–133
10. Ha EJ, Moon WJ, Na DG et al (2016) A multicenter prospective validation study for the Korean Thyroid Imaging Reporting and Data System in patients with thyroid nodules. *Korean J Radiol* 17:811–821
11. Tang AL, Falciglia M, Yang H, Mark JR, Steward DL (2017) Validation of American Thyroid Association ultrasound risk assessment of thyroid nodules selected for ultrasound fine-needle aspiration. *Thyroid* 27:1077–1082
12. Horvath E, Silva CF, Majlis S et al (2017) Prospective validation of the ultrasound based TIRADS (Thyroid Imaging Reporting and Data System) classification: results in surgically resected thyroid nodules. *Eur Radiol* 27:2619–2628
13. Ha SM, Ahn HS, Baek JH et al (2017) Validation of three scoring risk-stratification models for thyroid nodules. *Thyroid* 27:1550–1557
14. Lubitz CC, Kong CY, McMahon PM et al (2014) Annual financial impact of well-differentiated thyroid cancer care in the United States. *Cancer* 120:1345–1352
15. Kwak JY, Han KH, Yoon JH et al (2011) Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk. *Radiology* 260:892–899
16. Leenhardt L, Erdogan MF, Hegedus L et al (2013) 2013 European thyroid association guidelines for cervical ultrasound scan and ultrasound-guided techniques in the postoperative management of patients with thyroid cancer. *Eur Thyroid J* 2:147–159
17. Seo H, Na DG, Kim JH, Kim KW, Yoon JW (2015) Ultrasound-based risk stratification for malignancy in thyroid nodules: a four-tier categorization system. *Eur Radiol* 25:2153–2162
18. Choi YJ, Baek JH, Suh CH et al (2017) Core-needle biopsy versus repeat fine-needle aspiration for thyroid nodules initially read as atypia/follicular lesion of undetermined significance. *Head Neck* 39:361–369
19. Ha EJ, Na DG, Baek JH, Sung JY, Kim JH, Kang SY (2018) US fine-needle aspiration biopsy for thyroid malignancy: diagnostic performance of seven society guidelines applied to 2000 thyroid nodules. *Radiology* 287:893–900
20. Middleton WD, Teefey SA, Reading CC et al (2017) Multiinstitutional analysis of thyroid nodule risk stratification using the American College of Radiology Thyroid Imaging

- Reporting and Data System. *AJR Am J Roentgenol* 208:1324–1334
21. Yoon JH, Lee HS, Kim EK, Moon HJ, Kwak JY (2016) Malignancy risk stratification of thyroid nodules: comparison between the Thyroid Imaging Reporting and Data System and the 2014 American Thyroid Association management guidelines. *Radiology* 278:917–924
 22. Griffin AS, Mitsky J, Rawal U, Bronner AJ, Tessler FN, Hoang JK (2018) Improved quality of thyroid ultrasound reports after implementation of the ACR thyroid imaging reporting and data system nodule lexicon and risk stratification system. *J Am Coll Radiol* 15: 743–748
 23. Tee YY, Lowe AJ, Brand CA, Judson RT (2007) Fine-needle aspiration may miss a third of all malignancy in palpable thyroid nodules: a comprehensive literature review. *Ann Surg* 246:714–720