

# Ultrasound is helpful to differentiate Bethesda class III thyroid nodules

## A PRISMA-compliant systematic review and meta-analysis

Lu-Ying Gao, MD, Ying Wang, MD, Yu-Xin Jiang, MD, Xiao Yang, MD, Ru-Yu Liu, MB, Xue-Hua Xi, MB, Shen-Ling Zhu, MD, Rui-Na Zhao, MD, Xing-Jian Lai, MD, Xiao-Yan Zhang, MD, Bo Zhang, MD\*

### Abstract

**Background:** Fine-needle aspiration (FNA) is the most dependable tool to triage thyroid nodules for medical or surgical management. However, Bethesda class III cytology, namely “follicular lesion of undetermined significance” (FLUS) or “atypia of undetermined significance” (AUS), is a major limitation of the US-FNA in assessing thyroid nodules. As the most important imaging method, ultrasound (US) has a high efficacy in diagnosing thyroid nodules. This meta-analysis aimed to assess the role of US in evaluating Bethesda class III thyroid nodules.

**Methods:** With keywords “Undetermined Significance,” “Bethesda Category III,” “Bethesda system,” “Cytological Subcategory,” “AUS/FLUS,” “Atypia of Undetermined Significance,” and “Ultrasound/US,” papers in PubMed, Cochrane Library, Medline, Web of Science, Embase, and Google Scholar from inception to December 2016 were searched. A meta-analysis of these trials was then performed for evaluating the diagnostic value of thyroid ultrasound in Bethesda Category III thyroid nodules.

**Results:** Fourteen studies including 2405 nodules were analyzed. According to the criteria for US diagnosis of thyroid nodules in each article, with any one of suspicious features as indicators of malignancy, US had a pooled sensitivity of 0.75 (95% CI 0.72–0.78) and a pooled specificity of 0.48 (95% CI 0.45–0.50) in evaluating Bethesda Class III Nodules. The pooled diagnostic odds ratio was 10.92 (95% CI 6.04–19.74). The overall area under the curve was 0.84 and the Q\* index was 0.77. With any 2 or 3 of US suspicious features as indicators of malignancy, the sensitivity and specificity were 0.77 (95% CI 0.71–0.83) and 0.54 (95% CI 0.51–0.58), 0.66 (95% CI 0.59–0.73) and 0.71 (95% CI 0.68–0.74), respectively.

**Conclusions:** US was helpful for differentiating benign and malignant Bethesda class III thyroid nodules, with the more suspicious features, the more likely to be malignant.

**Abbreviations:** 95% CI = 95% confidence interval, ATA = American Thyroid Association, AUC = area under the curve, AUS = atypia of undetermined significance, DOR = diagnostic odds ratio, FLUS = follicular lesion of undetermined significance, FN = false-negative, FNA = fine-needle aspiration, FP = false-positive, GEC = gene expression classifier, LR = likelihood ratio, PCTNs = partially cystic thyroid nodules, PPV = positive predictive value, PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analysis, QUADAS = Quality Assessment of Diagnostic Accuracy Studies, SROC = summary receiver operating characteristic, TI-RADS = thyroid imaging reporting and data system, TN = true-negative, TP = true-positive, US = ultrasound.

**Keywords:** Bethesda class III, thyroid nodule, thyroid ultrasonography

Editor: Eleonore Fröhlich.

LG and YW contributed equally to this work and should be considered co-first authors.

Funding: This study was supported by a grant from the National Natural Science Foundation of China (No. 81541131).

The authors have no conflicts of interest to disclose.

Supplemental Digital Content is available for this article.

Department of Ultrasound, Chinese Academy of Medical Sciences & Peking Union Medical College Hospital, Beijing, China.

\* Correspondence: Bo Zhang, Department of Ultrasound, Peking Union Medical College Hospital, Dongdansantiao, Beijing, China (e-mail: zhangbop0417@pumch.cn).

Copyright © 2017 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the Creative Commons Attribution-No Derivatives License 4.0, which allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to the author.

Medicine (2017) 96:16(e6564)

Received: 11 January 2017 / Received in final form: 9 March 2017 / Accepted: 10 March 2017

<http://dx.doi.org/10.1097/MD.0000000000006564>

## 1. Introduction

Thyroid nodules are very common medical problem with the prevalence of 19% to 68% in general population.<sup>[1,2]</sup> About 7% to 15% of thyroid nodules are thyroid cancer, it was estimated that 96% of all new endocrine organ cancers originate from the thyroid gland in 2014.<sup>[3,4]</sup> According to American Thyroid Association (ATA) guidelines, FNA is the most accurate and cost-effective method for evaluating thyroid nodules.<sup>[5]</sup> The reports of FNA are dependable to triage thyroid nodules for medical or surgical management. From 2007 to now, the Bethesda classification system for reporting thyroid FNA has been used world widely. The most important contributions are the risk of malignancy and management recommendation for each category, and the creation of an atypical category for repeat FNA, that is Bethesda class III (atypia of undetermined significance or follicular lesion of undetermined significance in the Bethesda System), with the risk of malignancy 5% to 15%.<sup>[6,7]</sup> Bethesda class III, which do not provide a differential diagnosis between malignant and benign lesion, is a major limitation of the US-FNA in assessing thyroid nodules.<sup>[5]</sup> Management of Bethesda class III nodule seems to be

diverse, including clinical observation, ultrasound follow-up, repeat FNA, molecular test or surgery,<sup>[8]</sup> sometimes depending on the willing of patients or experiences of physicians. Recently, several studies have focused on the feasibility of using thyroid US to predict the malignancy of Bethesda class III thyroid nodules.<sup>[9–22]</sup> Although its clinical use was uncertain, these studies showed that promising results of US might provide useful information. We conducted this meta-analysis to evaluate the role of US in the diagnosis of Bethesda class III thyroid nodules.

## 2. Materials and methods

### 2.1. Search strategy and inclusion criteria

Our systematic review was designed and performed according to the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement.<sup>[23]</sup> The ethical approval was not necessary because our meta analysis was based on data from previously published studies. We searched PubMed, Cochrane Library, Medline, Web of Science, Embase, and Google Scholar from inception to December 2016 under the key words relating to “Undetermined Significance,” “Bethesda Category III,” “Bethesda system,” “Cytological Subcategory,” “AUS/FLUS,” “Atypia of Undetermined Significance” and “Ultrasound/US.” The inclusion criteria were studies of using thyroid US to predict malignancy of the Bethesda class III nodules in the initial US-FNA. Informed consent was necessary. References cited in original and review articles were cross checked. Reviews, abstracts, and duplicate data were removed. No language restrictions were applied.

Two authors (LYG and YW) performed the search job and data screening independently, and the discrepancies resolved by consensus (BZ).

### 2.2. Data extraction, quality assessment of articles, and statistical analysis

We attempted to determine the extent to which US is diagnostic that identifies Bethesda class III nodules as malignant or benign. However, US classification schemes for thyroid nodules are diverse, and different reports used various categories. So according to number of the suspicious malignant features of US, data were extracted by 2 independent authors. As the criteria for US diagnosis of thyroid nodules in each article, suspicious malignant features included hypoechoogenicity, irregular margins, microcalcifications, taller-than-wide shape or macrocalcifications, increasing size during follow-up, increased vascularization, and disrupted rim calcifications.<sup>[9–22]</sup> The absolute number of true-positive (TP), false-positive (FP), false-negative (FN), and true-negative (TN) were retrieved or calculated from the articles; other characteristics, including publication year, country or region, average age of patients, proportion of males, nodules number, retrospective or prospective set-up of the study and the reference standard that was used in the study were also recorded.

Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) criteria was applied in the methodological quality of the included studies.<sup>[24]</sup>

Sensitivity, specificity, diagnostic accuracy, and diagnostic odds ratio (DOR) were calculated using Meta-Disc version 1.4 statistical software (Meta-Disc, Unit of Clinical Biostatistics Team of the Romany Cajal Hospital, Madrid, Spain). All results were estimated as the pooled odds ratio with 95% confidence interval. To detect heterogeneity, the likelihood ratios and DORs

were graphically displayed using forest plots. A summary receiver operating characteristic (SROC) curve was constructed, the Q\*index was calculated. Heterogeneity among studies was assessed by the  $I^2$  statistic.

## 3. Results

### 3.1. Study characteristics

The flow chart is shown in Fig. 1. According to the searching strategy, the initial search retrieved 188 articles, of those, 146 were excluded based on the title and abstract. Full-text assessment was performed on 42 articles. Of these, 13 articles only analyzed the ultrasonic features which predicted the risk of malignancy, but did not examine US as a diagnostic tool in Bethesda class III nodules; 5 articles were written in Chinese or French; and 10 articles were duplicated articles. Finally, 14 articles including 2405 thyroid nodules were included in the analysis<sup>[9–22]</sup> (Fig. 1).

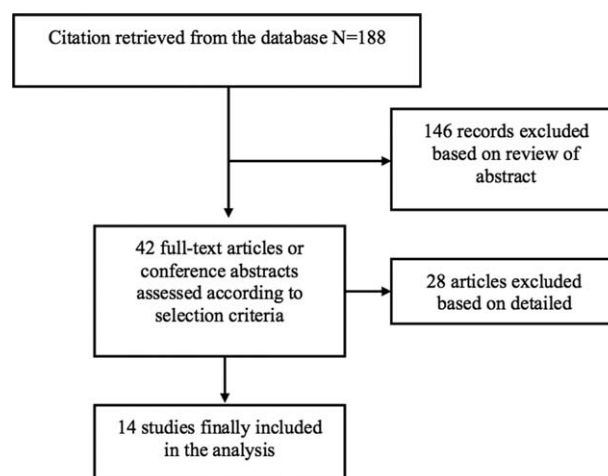
The 14 articles were published from 2011 to 2016. The study designs were retrospective (N=11) and prospective (N=3). The number of analyzed nodules ranged from 29 to 548. The average age of participants in the study was 48.4 years. Males constituted 19.3% of all patients. In 6 studies, surgery and repeat FNA was carried out for histopathological and cytological results. Four studies carried out surgery for histological results. In addition, 3 studies included follow-up US, and 1 study included a genetic test used as a tool to assess the final diagnosis (Table 1).

The included studies generally had a low risk of bias. Details about the quality of trials were according to the QUADAS-2.

The funnel plot indicated no likely publication bias, showing a symmetric shape when the log DORs of individual studies, was plotted against their sample sizes.

### 3.2. Summary of sensitivity, specificity, positivity, and SROC curves

When thyroid nodules with any one of US feature of malignancy were considered malignant, all 14 articles were included. The pooled sensitivity of US diagnostic accuracy for differentiating malignant and benign nodules was 0.75 (95% CI 0.72–0.78) (Fig. 2), and the pooled specificity of US was 0.48 (95% CI 0.45–0.50) (Fig. 3). The summary positive LR (likelihood ratio)



**Figure 1.** The procedure of study selection in our meta-analysis. A total of 14 studies were included in this systematic review, which fulfilled all the inclusion criteria.

**Table 1**  
**Basic characteristic of 14 studies.**

No	Authors	Publication year	Country/region	Study design	Number of nodules	Male%	Average age	Reference standard
1	Kim	2011	South Korea	Prospectively	388	15.4	49	Surgery
2	Gweon	2013	South Korea	Prospectively	155	14.6	48.4	Surgery and repeat FNA
3	Carr	2013	USA	Retrospective	95	20	N	Surgery
4	Jung	2014	South Korea	Retrospective	192	21.7	46	Surgery
5	Kwang	2014	South Korea	Retrospective	152	23	47.9	Surgery, CNB and repeat FNA
6	Pedro	2014	Brazil	Prospectively	150	17.3	N	Surgery and repeat FNA
7	Young	2014	South Korea	Retrospective	116	17.1	49.1	Surgery, gene, follow-up US, repeat FNA
8	Vivian	2015	South Korea	Retrospective	29	19.4	50.6	Surgery, follow-up US, repeat FNA
9	Aya	2015	USA	Retrospective	41	19.6	49.4	Surgery and repeat FNA
10	Mi	2015	South Korea	Retrospective	67	20	47.5	Surgery and repeat FNA
11	Kim	2016	South Korea	Retrospective	43	19.6	46.7	Surgery and repeat FNA
12	B. Kuru	2016	Turkey	Retrospective	153	20.9	N	Surgery
13	Jung	2016	South Korea	Retrospective	275	22.9	50.2	Surgery and repeat FNA
14	Hee	2016	South Korea	Retrospective	548	18.6	52.6	Surgery, follow-up US, repeat FNA

CNB = core needle biopsy, FNA = fine-needle aspiration, US = ultrasound.

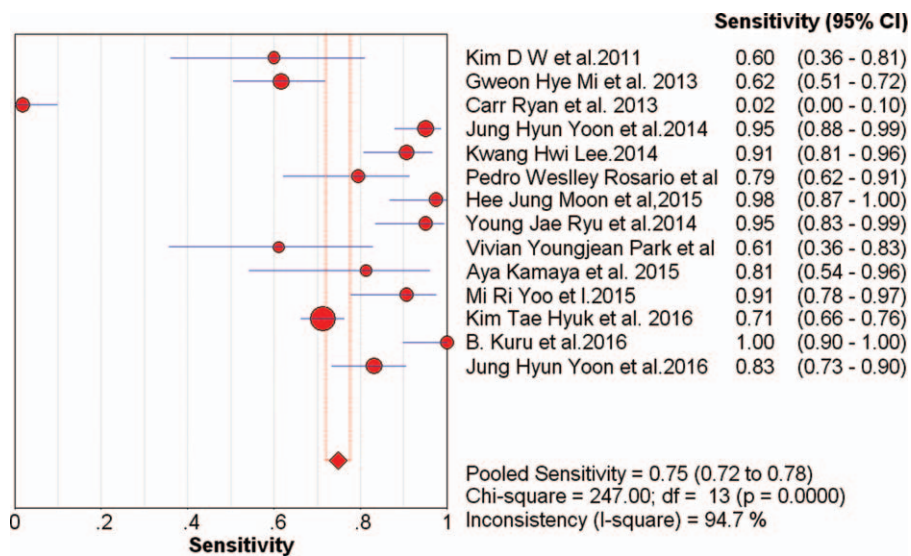


Figure 2. Forest plot showed pooled sensitivity of US in the differentiated diagnosis of Bethesda class III thyroid nodules. US = ultrasound.

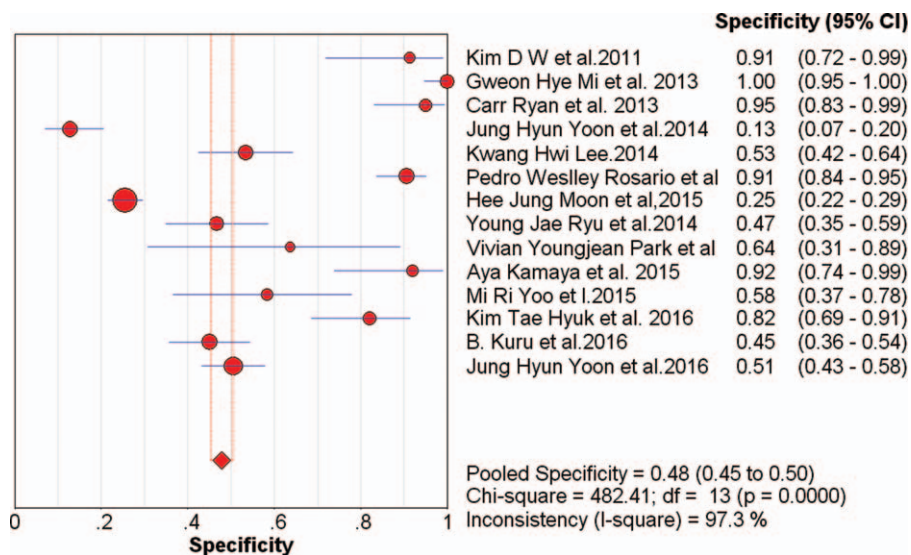
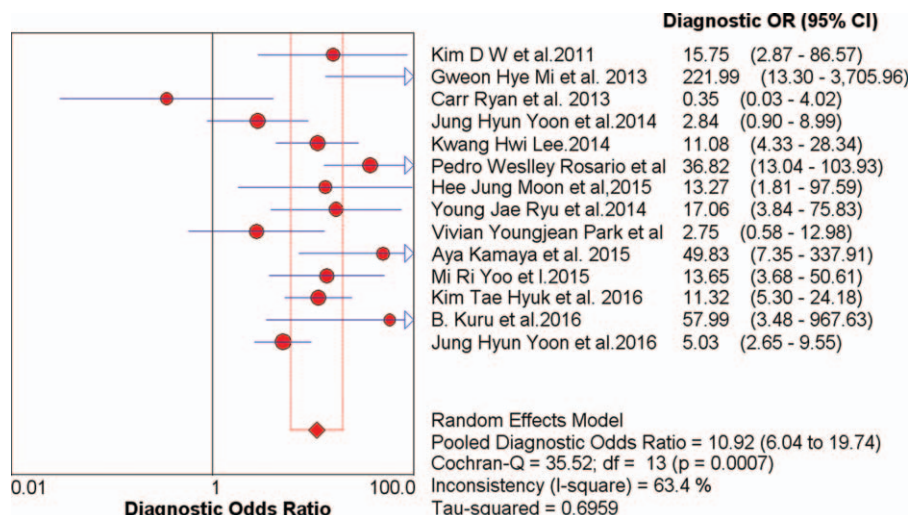


Figure 3. Forest plot showed pooled positivity of US in the differentiated diagnosis of Bethesda class III thyroid nodules. US = ultrasound.



**Figure 4.** Forest plot showed diagnostic odds ratio (DOR) of US in the differentiated diagnosis of Bethesda class III thyroid nodules. DOR = diagnostic odds ratio, US = ultrasound.

were 2.29 (95% CI 1.65–3.18), and the summary negative LR<sub>s</sub> were 0.27 (95% CI 0.12–0.20), respectively. The pooled DOR was 10.92 (95% CI 6.04–19.74) (Fig. 4). The SROC curve was symmetric. The overall AUC (area under the curve) was 0.8358, and the Q\* index was 0.7679 (Fig. 5).

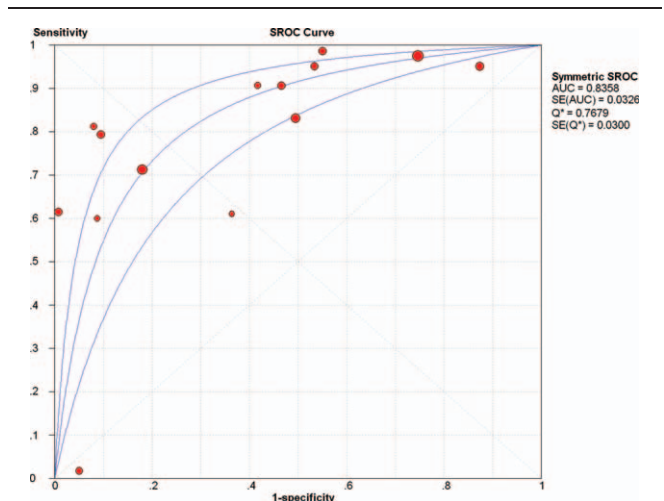
When thyroid nodules with any 2 of suspicious features were considered malignant nodules, 5 articles were included. The pooled sensitivity was 0.77 (95% CI 0.71–0.83), and the pooled specificity was 0.54 (95% CI 0.51–0.58). The pooled DOR was 7.61 (95% CI 3.04–19.07) (see Supplemental Digital Content Fig. 1s, <http://links.lww.com/MD/B635>, which shows the DOR of US). The overall AUC was 0.8116, and the Q\* index was 0.7461.

When thyroid nodules with any 3 of suspicious features were considered malignant nodules, 4 articles were included. The pooled sensitivity was 0.66 (95% CI 0.59–0.73), and the pooled specificity was 0.71 (95% CI 0.68–0.74). The pooled DOR was

5.74 (95% CI 3.02–10.93) (see Supplemental Digital Content Fig. 2s, <http://links.lww.com/MD/B635> which shows the DOR of US). The overall AUC was 0.7425, and the Q\* index was 0.6873.

**3.3. Subgroup analysis**

Because noticeable heterogeneity was observed in the tests of sensitivity and specificity, the random effects model was used. In the analysis, only 14 studies were available, which prevented the use of a meta-regression. Therefore, we performed a subgroup analysis of 3 subgroups (increased vascularization as one of the US malignant features, number of nodules, and thyroid imaging reporting and data system (TI-RADS) criteria) to determine the potential sources of heterogeneity. Among them, only 1 main factor (increased vascularization as one of the US malignant features) contributed to the heterogeneity of our systematic-review (Table 2). The result indicated that studies including increased vascularization as one of US malignant features had important influence on the overall specificity and sensitivity ( $P < 0.05$ ).



**Figure 5.** Summary receiver operating characteristic (SROC) curves of US in the evaluation of Bethesda class III thyroid nodules. SROC = summary receiver operating characteristic, US = ultrasound.

**4. Discussion**

It is necessary to improve the management of AUS/FLUS nodules based on FNA biopsy diagnosis. Our meta-analysis showed that ultrasound classification scheme did facilitate malignancy prediction and helped guide the therapeutic plan for AUS/FLUS nodules.

The systematic review focused on the diagnostic value of US in Bethesda class III nodules. According to our meta-analysis, with 1 US suspicious feature as indicators of malignancy, the pooled sensitivity was 0.75, and the pooled specificity was 0.48. The overall AUC was 0.84. These values indicated that US has very good diagnostic accuracy for the differentiation of Bethesda class III nodules. Moreover, the DOR in our review was 10.92 (95% CI 6.04–19.74), which demonstrated that the US was a good diagnosis test for the differentiation of Bethesda class III nodules. In most studies, thyroid nodules with 1 US feature of malignancy were considered as malignant nodules. For nodules with 2 US features of malignancy, the pooled sensitivity was 0.77 and the

**Table 2**  
**Diagnostic accuracy of subgroups studies.**

Subgroups	Number of studies	Pooled sensitivity (95% CI)	Pooled specificity (95% CI)	DOR	P
Total	14	0.75 (0.72–0.78)	0.48 (0.45–0.50)	10.92	0.001
US malignant criteria including increased vascularization	3	0.92 (0.85–0.96)	0.62 (0.57–0.68)	30.4	0.63
US malignant criteria excluding increased vascularization	11	0.73 (0.70–0.76)	0.44 (0.41–0.47)	25.5	0.005
No. of nodules ( $\geq 100$ )	7	0.81 (0.78–0.84)	0.41 (0.38–0.43)	10.9	0.009
No. of nodules ( $< 100$ )	7	0.62 (0.56–0.67)	0.78 (0.73–0.83)	10.8	0.005
US Criteria (Tirads criteria)	6	0.82 (0.78–0.86)	0.36 (0.33–0.39)	7.35	0.022
US Criteria (non-Tirads criteria)	8	0.71 (0.67–0.74)	0.68 (0.64–0.72)	14.94	0.032

CI = confidence interval, DOR = diagnostic odds ratio, TI-RADS = thyroid imaging reporting and data system, US = ultrasound.

pooled specificity was 0.54. The AUC was 0.81, and the DOR in our review was 7.61 (95% CI 3.04–19.07). For thyroid nodules with 3 US features of malignancy, the pooled sensitivity was 0.66 and the pooled specificity was 0.71. The overall AUC was 0.74, and the DOR in our review was 5.74 (95% CI 3.02–10.93). Therefore, it is reasonable that when Bethesda class III nodules showed more US features of malignancy, the specificity of diagnostic malignancy were higher.

The diversity of US classification schemes for nodules may contribute to the heterogeneity of this meta-analysis. Many studies have indicated the US features of thyroid nodal diseases and some malignant features have been generally accepted.<sup>[25–27]</sup> However, the criteria for the ultrasound features of malignant nodules are still controversial. In our meta-analysis, we found a large range of sensitivity and specificity (0.02–1.00 and 0.13–1.00) with high heterogeneity ( $P < 0.01$ ). The various criteria for the US malignant features in these publications may contribute to the different sensitivities and specificities. In most studies, nodules diagnosed as malignant were characterized by hypoechogenicity, a taller than wide shape, microcalcifications, and a spiculated margin. However, macro-calcifications, increasing size during follow-up, increasing vascularization and disrupted rim calcifications with extrusive soft tissue components were considered to be malignant US features in several studies, but not in others. Kim et al<sup>[9]</sup> emphasized different US categories for solid thyroid nodules and partially cystic thyroid nodules (PCTNs), but the other studies used the same US classification schemes for solid nodules and PCTNs.

The multiclassifications of US features in different medical institutions may confuse ultrasonologist who perform US examinations. Moreover, the experience differences of cytopathologists in interpreting FNA slides might have resulted in variable cytological diagnoses for cases, and the operator-dependent character of US might contribute to this divergence. In addition, through the subgroup regression analyses, we identify only 1 main factor (increased vascularization) had a significant influence on the specificity and sensitivity, and the factors may be responsible for the heterogeneity. The data demonstrated that only.

The management for AUS/FLUS nodules varies widely among institutions, including US follow-ups, repeat FNA, molecular test, and surgery.<sup>[28,29]</sup> Layfield et al<sup>[30]</sup> recommended a repeat US-FNA at a specified interval. However, 1 group recommended the limited use of repeat FNA because a discrepancy might be unavoidable in the interpretation of the overlapping cytological criteria.<sup>[31]</sup> The clinical application of molecular test for Bethesda class III nodules is fairly common practice; however, according to ATA, the utility of molecular testing is applicable only when combined with clinical and

ultrasonic risk factors for malignancy.<sup>[25]</sup> Although it has been reported that the Veracyte Afirma gene expression classifier (GEC) test had a positive predictive value (PPV) of 38% and an FN rate of 5%, the cost effectiveness of molecular test is still controversial.<sup>[32]</sup>

There are several limitations to our meta-analysis. First, only a limited number of studies have reported the accuracy of US diagnosis of AUS/FLUS, which limits the generalizability of our results. Second, the lack of uniform criteria of the US diagnosis scheme may result in evaluation bias of nodules. The diversity of US classification schemes of differentiation of thyroid nodules limits the use of US as an effective diagnostic test. Third, not all cases of included studies were confirmed by pathology. Some patients experienced US follow-ups.

US provides different risks of malignancy for nodules initially classified as Bethesda category III, with the more suspicious features, the more likely to be malignant.

## References

- [1] Tan GH, Gharib H. Thyroid incidentalomas: management approaches to nonpalpable nodules discovered incidentally on thyroid imaging. *Ann Intern Med* 1997;126:226–31.
- [2] Guth S, Theune U, Aberle J, et al. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. *Eur J Clin Invest* 2009;39:699–706.
- [3] Hegedüs L. Clinical practice. The thyroid nodule. *N Engl J Med* 2004;351:1764–71.
- [4] Mandel SJ. A 64-year-old woman with a thyroid nodule. *JAMA* 2004;292:2632–42.
- [5] Md HB, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer. *Thyroid* 2016;26:1–33.
- [6] Cibas ES, Ali SZ. NCI Thyroid FNA State of the Science Conference. The Bethesda system for reporting thyroid cytopathology. *Am J Clin Pathol* 2009;132:658–65.
- [7] Alexander EK, Heering JP, Benson CB, et al. Assessment of non-diagnostic ultrasound-guided fine needle aspirations of thyroid nodules. *J Clin Endocrinol Metab* 2002;87:4924–7.
- [8] Melo M, da Rocha AG, Vinagre J, et al. TERT promoter mutations are a major indicator of poor outcome in differentiated thyroid carcinomas. *J Clin Endocrinol Metab* 2014;99:E754–65.
- [9] Kim DW, Lee EJ, Jung SJ, et al. Role of sonographic diagnosis in managing Bethesda class III nodules. *AJNR Am J Neuroradiol* 2011;32:2136–41.
- [10] Carr R, Ustun B, Chhieng D, et al. Radiologic and clinical predictors of malignancy in the follicular lesion of undetermined significance of the thyroid. *Endocr Pathol* 2013;24:62–8.
- [11] Yoo MR, Gweon HM, Park AY, et al. Repeat diagnoses of Bethesda category III thyroid nodules: what to do next? *PLoS One* 2015;10:e0130138.
- [12] Park VY, Kim E, Kwak JY, et al. Malignancy risk and characteristics of thyroid nodules with two consecutive results of atypia of undetermined significance or follicular lesion of undetermined significance on cytology. *Eur Radiol* 2015;25:2601–7.

- [13] Gweon HM, Son EJ, Youk JH, et al. Thyroid nodules with Bethesda system III cytology: can ultrasonography guide the next step? *Ann Surg Oncol* 2013;20:3083–8.
- [14] Ryu YJ, Jung YS, Yoon HC, et al. Atypia of undetermined significance on thyroid fine needle aspiration: surgical outcome and risk factors for malignancy. *Ann Surg Treat Res* 2014;86: 109–14.7.
- [15] Kim TH, Jeong DJ, Hahn SY, et al. Triage of patients with AUS/FLUS on thyroid cytopathology: effectiveness of the multimodal diagnostic techniques. *Cancer Med* 2016;5:769–77.
- [16] Kamaya A, Lewis GH, Liu Y, et al. Atypia of undetermined significance and follicular lesions of undetermined significance: sonographic assessment for prediction of the final diagnosis. *J Ultrasound Med* 2015;34:767–74.
- [17] Kuru B, Atmaca A, Tarim IA, et al. Risk factors associated with malignancy and with triage to surgery in thyroid nodules classified as Bethesda category III (AUS/FLUS). *Eur J Surg Oncol* 2016;42:87–93.
- [18] Lee KH, Shin JH, Oh YL, et al. Atypia of undetermined significance in thyroid fine-needle aspiration cytology: prediction of malignancy by US and comparison of methods for further management. *Ann Surg Oncol* 2014;21:2326–31.
- [19] Moon HJ, Kim EK, Yoon JH, et al. Malignancy risk stratification in thyroid nodules with nondiagnostic results at cytologic examination: combination of thyroid imaging reporting and data system and the Bethesda System. *Radiology* 2015;274:287–95.
- [20] Rosario PW. Thyroid nodules with atypia or follicular lesions of undetermined significance (Bethesda Category III): importance of ultrasonography and cytological subcategory. *Thyroid* 2014;24: 1115–20.
- [21] Yoon JH, Lee HS, Kim EK, et al. A nomogram for predicting malignancy in thyroid nodules diagnosed as atypia of undetermined significance/follicular lesions of undetermined significance on fine needle aspiration. *Surgery* 2014;155:1006–13.
- [22] Yoon JH, Kwon HJ, Kim EK, et al. Subcategorization of atypia of undetermined significance/follicular lesion of undetermined significance (AUS/FLUS): a study applying Thyroid Imaging Reporting and Data System (TIRADS). *Clin Endocrinol (Oxf)* 2016;85:275–82.
- [23] Knobloch K, Yoon U, Vogt PM. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement and publication bias. *J Craniomaxillofac Surg* 2011;39:91–2.
- [24] Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
- [25] Papini E, Guglielmi R, Bianchini A, et al. Risk of malignancy in nonpalpable thyroid nodules: predictive value of ultrasound and color-Doppler features. *J Clin Endocrinol Metab* 2002;87:1941–6.
- [26] Shimura H, Haraguchi K, Hiejima Y, et al. Distinct diagnostic criteria for ultrasonographic examination of papillary thyroid carcinoma: a multicenter study. *Thyroid* 2005;15:251–8.
- [27] Moon WJ, Jung SL, Lee JH, et al. Benign and malignant thyroid nodules: US differentiation—multicenter retrospective study. *Radiology* 2008;247:762–70.
- [28] Pang T, Ihre-Lundgren C, Gill A, et al. Correlation between indeterminate aspiration cytology and final histopathology of thyroid neoplasms. *Surgery* 2010;148:532–7.
- [29] Yoon JH, Kwak JY, Kim EK, et al. How to approach thyroid nodules with indeterminate cytology. *Ann Surg Oncol* 2010;17:2147–55.
- [30] Layfield LJ, Abrams J, Cochand-Priollet B, et al. Post-thyroid FNA testing and treatment options: a synopsis of the National Cancer Institute Thyroid Fine Needle Aspiration State of the Science Conference. *Diagn Cytopathol* 2008;36:442–8.
- [31] Lee YH, Kim BH, Suh SI, et al. Comparison of cytological results obtained by repeated US-guided fine-needle aspiration biopsies of thyroid nodules: focus on the rate of malignancy and diagnostic concordance. *Diagn Cytopathol* 2009;37:492–7.
- [32] Ferris RL, Baloch Z, Bernet V, et al. American Thyroid Association statement on surgical application of molecular profiling for thyroid nodules: current impact on perioperative decision making. *Thyroid* 2015;25:760–8.