Original Article | Thyroid

eISSN 2005-8330
https://doi.org/10.3348/kjr.2024.0292
Korean J Radiol 2024;25(10):924-933

Korean Journal of Radiology
KJR

Risk Stratification of Thyroid Nodules Diagnosed as Bethesda Category III by Ultrasound, Size, and Cytology

Hye Shin Ahn[1], Dong Gyu Na[2,3], Ji-Hoon Kim[4]

[1]Department of Radiology, Chung-Ang University Hospital, Chung-Ang University College of Medicine, Seoul, Republic of Korea
[2]Department of Radiology, Gangneung Asan Hospital, University of Ulsan College of Medicine, Gangneung, Republic of Korea
[3]Department of Radiology, Human Medical Imaging and Intervention Center, Seoul, Republic of Korea
[4]Department of Radiology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Republic of Korea

Objective: This study aimed to evaluate the performance of an integrated risk stratification system (RSS) based on ultrasound (US) RSSs, nodule size, and cytology subcategory for diagnosing malignancy in thyroid nodules initially identified as Bethesda category III on fine-needle aspiration.

Materials and Methods: This retrospective study was conducted at two institutions and included consecutive patients with Bethesda category III nodules, and final diagnoses confirmed by repeat biopsy or surgery. A total of 320 Bethesda category III nodules (≥1 cm) from 309 patients (223 female and 86 male; mean age, 50.9 ± 12.0 years) were included. The malignancy risk of Bethesda category III nodules and predictors of malignancy were assessed according to US RSSs, nodule size, and cytology subcategory. The diagnostic performances of US-size cytology (USC) RSS and US RSS alone for malignancy were compared.

Results: The intermediate or high suspicion US category independently increased the malignancy risk in all US RSSs ($P \leq 0.001$). Large nodule size (≥3 cm) independently increased the malignancy risk of low- or intermediate suspicion US category nodules. Additionally, the atypia of undetermined significance cytology subcategory independently increased the malignancy risk of low suspicion US category nodules in most US RSSs. The area under the receiver operating characteristic curve of the USC RSSs was greater than that of the US RSSs alone ($P < 0.048$). Malignancy was not found in the very low risk category of USC RSS.

Conclusion: The diagnostic performance of USC RSS for malignancy was superior to that of US RSS alone in Bethesda category III nodules. Malignancy can be ruled out in the very low-risk category of USC RSS.

Keywords: Thyroid gland; Thyroid neoplasms; Thyroid nodule; Ultrasonography; Fine-needle aspiration

INTRODUCTION

Ultrasound (US)-guided fine-needle aspiration (FNA) is the standard procedure for primary diagnosis of thyroid nodules. However, this method is limited by high rates of inconclusive results, including Bethesda category I (non-diagnostic) and Bethesda category III (atypia/follicular lesion of undetermined significance [AUS/FLUS]). The reported rate of Bethesda category III cases ranges from 3.0% to 20.5%, depending on the institution [1-7], and these cases show higher rates of inconclusive results (up to 65.4%) on repeat FNA [8-10]. As a result, a considerable number of patients diagnosed with Bethesda category III on initial FNA undergo unnecessary diagnostic surgery following inconclusive results on repeat FNA. The estimated malignancy risk for this category is 13% to 30% according to a recent Bethesda System report [11] and approximately 27% to 34% according to systematic reviews and meta-analyses [12,13].

The management of nodules initially diagnosed as Bethesda category III remains controversial and includes observation with US surveillance, repeat FNA or core

Received: March 25, 2024 **Revised:** June 30, 2024
Accepted: August 3, 2024
Corresponding author: Dong Gyu Na, MD, PhD, Department of Radiology, Gangneung Asan Hospital, 38 Bangdong-gil, Sacheon-myeon, Gangneung 25440, Republic of Korea
• E-mail: nndgna@gmail.com

Korean Journal of Radiology
KJR

needle biopsy (CNB), molecular tests, or diagnostic surgery, depending on the US features, clinical risk factors, and patient factors [14-19]. Therefore, researchers have investigated the estimation and stratification of individual malignancy risks for Bethesda category III nodules to optimize patient management. Previous studies have consistently reported that US features or US risk stratification systems (RSSs) can stratify the malignancy risk of Bethesda category III nodules [20-23] and that AUS nodules with nuclear atypia tend to have a higher malignancy rate than FLUS nodules with architectural atypia [11,24,25,26]. Although the predictive role of nodule size for malignancy in Bethesda category III nodules is unclear [26,27], nodule size may stratify malignancy risk in the mutation-negative subgroup [28]. However, the integrated RSS based on these three predictors of malignancy has rarely been investigated in Bethesda category III nodules. Therefore, this study aimed to evaluate the performance of an integrated RSS based on the US RSSs, nodule size, and cytology subcategory for diagnosing malignancy in Bethesda category III nodules on FNA.

MATERIALS AND METHODS

Study Design and Patient Characteristics

This retrospective observational cohort study was approved by the Institutional Review Boards of Seoul National University Hospital and the Human Medical Imaging and Intervention Center (IRB No. 2010-028-1162 and HI2020-01, respectively). Furthermore, the requirement for informed consent was waived owing to its retrospective nature. This study was performed in accordance with the Standards of Reporting of Diagnostic Accuracy Studies statement [29]. This study was conducted at two institutions from January 2010 to December 2014, and included consecutive patients with thyroid nodules initially identified as Bethesda category III by FNA, with final diagnoses confirmed by repeat biopsy or surgery. Thyroid nodules were excluded based on the following criteria: 1) nodules less than 1 cm, 2) nodules without a follow-up biopsy, or 3) nodules without a final diagnosis.

The final diagnosis of malignancy was determined by the histopathology obtained from surgical resection or a malignant result (category VI) on FNA or CNB. Final diagnoses of benign nodules were determined by histopathology obtained from surgical resection or at least one benign diagnosis (category II) on repeat FNA or CNB without a diagnosis of follicular neoplasm, suspicious for malignancy, or malignancy on additional repeat FNA or CNB.

From January 2010 to December 2014, 798 thyroid nodules (10.4%) were initially diagnosed as Bethesda category III among 7657 consecutive nodules that underwent FNA at the two institutions [30]. Of the 798 nodules, sub-centimeter nodules (n = 297), nodules with no follow-up repeat FNA or CNB (n = 95), and nodules with no final diagnosis (n = 86) were excluded. Therefore, 320 consecutive thyroid nodules (≥1 cm) with final diagnoses, from 309 patients (223 female and 86 male; mean age, 50.9 ± 12.0 years) were finally included in this study (Fig. 1).

US Examination and Image Analysis

A high-resolution color Doppler US with a 10–12 MHz or 5–14 MHz linear array transducer (Aplio XG, Toshiba, Otawara, Japan; iU22, Philips Medical Systems, Bothell, WA, USA) was used. All US images were retrospectively assessed by two radiologists (D.G.N. and H.S.A.) who had 22 and 7 years of experience, respectively, in performing thyroid US, who were unaware of the FNA results or final diagnoses. The radiologists independently evaluated the US features of thyroid nodules: composition, echogenicity, margin, orientation (taller-than-wide), echogenic foci (calcification) and the US features of discrepant cases were determined by a consensus of the two reviewers. Thyroid nodules were assessed and classified according to the US RSSs of four societies: the American College of Radiology (ACR) Thyroid Imaging Reporting and Data System (TI-RADS), American Thyroid Association (ATA) system, European

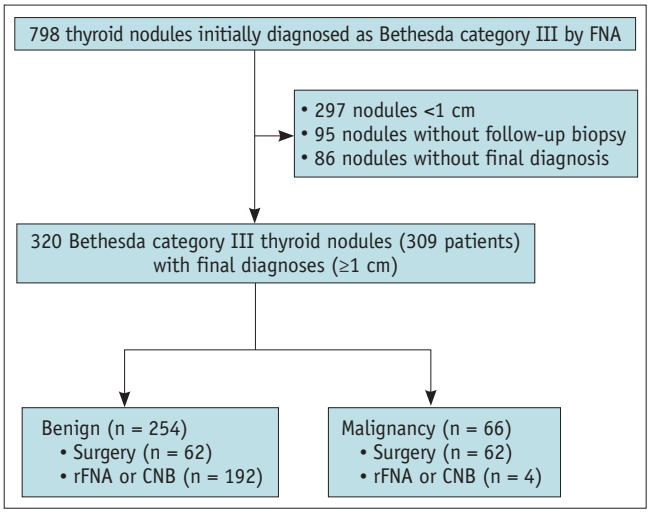

Fig. 1. Flow diagram of patient enrollment. FNA = fine-needle aspiration, rFNA = repeat FNA, CNB = core needle biopsy

https://doi.org/10.3348/kjr.2024.0292

(EU)-TIRADS, and Korean (K)-TIRADS [14,16,31,32]. The nodules classified as benign, very low suspicion, or low suspicion according to the US RSSs were categorized into the same low-suspicion US category to develop the RSS for the Bethesda category III nodules. The 38 unclassified nodules (12.6%), including three entirely calcified nodules and 35 isoechoic nodules with suspicious features, were categorized as intermediate suspicion nodules according to the ATA system. Moreover, the three unclassified, entirely calcified nodules (0.9%) were categorized as intermediate risk nodules for this study, considering the estimated intermediate malignancy risk of these nodules [33,34].

US-Guided FNA Procedure and Cytology Analysis

During the study period, US-guided FNA was routinely performed as a first-line assessment for suspicious or indeterminate thyroid nodules measuring >1 cm [35]. A conventional freehand method was used and at least two samples were obtained from each nodule [30]. FNA interpretation was based on the Bethesda System for Reporting Thyroid Cytopathology [36]. Two endocrine pathologists retrospectively subcategorized the Bethesda category III cytology into two subcategories: AUS and FLUS. The AUS subcategory included nodules with nuclear atypia, but not enough atypia to be considered suspicious for malignancy. The FLUS subcategory included nodules with architectural atypia but not enough to be diagnosed as a follicular neoplasm or as suspicious for a follicular neoplasm [36,37].

Development of Integrated RSS for Bethesda Category III Nodules Based on US Category, Nodule Size, and Cytology Subcategory

The malignancy rates of Bethesda category III nodules were assessed according to the US category, nodule size, and cytology subcategory in all nodules, and subgroup analysis was performed according to the classified US categories. The associations of US categories, nodule size, and cytology subcategories with malignancy were assessed in Bethesda category III nodules and subgroups of the classified US categories. The integrated RSS for Bethesda category III nodules was developed based on the estimated malignancy rates according to the classified US categories, nodule size, and cytology subcategory, and the predictors for malignancy were identified according to the classified US categories.

Data Analysis and Statistics

Continuous variables are presented as mean ± standard deviation or median (interquartile range) according to parametric or nonparametric distribution, respectively. Categorical variables are reported as frequencies and percentages. The unpaired *t*-test or Mann-Whitney U test was used to compare continuous variables between benign and malignant nodules. The chi-square test or Fisher's exact test was used to compare categorical variables among the categories of nodule size and cytology subcategories in all categories and subgroups of the three US categories. The Mantel-Haenszel chi-square trend test was used to investigate the trend of the malignancy rate as the scores of the US RSSs and the developed integrated US-size-cytology (USC) RSS increase. Multivariable logistic regression analyses were performed to determine the independent predictors of malignancy among the US categories, nodule size, and cytology subcategories. The performance of US RSSs and USC RSS in diagnosing malignancy was compared using receiver operating characteristic (ROC) analysis. Test positivity was defined as intermediate or high risk in the USC RSS and intermediate or high suspicion in the US RSSs. Sensitivity and specificity were compared between the USC RSS and US RSSs using the McNemar test. In addition, interobserver agreement for the classified US categories of the four US RSSs were assessed using Cohen's κ statistics (0.81–1.00, almost perfect agreement; 0.61–0.80, substantial agreement; 0.41–0.60, moderate agreement; 0.21–0.40, fair agreement; and 0.00–0.20, slight agreement) [38]. Statistical analyses were performed using SPSS version 28.0 for Windows (IBM Corp., Armonk, NY, USA) and MedCalc 19.3.1 software (Ostend, Belgium). Statistical significance was defined as a *P*-value <0.05.

RESULTS

Demographic Data and the Characteristics of Thyroid Nodules

Table 1 shows the patients' demographic data and characteristics of the Bethesda category III nodules. The malignancy rate of the Bethesda category III nodules was 20.6%. The maximum nodule diameter did not significantly differ between benign and malignant nodules ($P = 0.123$), however, the number of large nodules (≥3 cm) was higher among malignant nodules than among benign nodules ($P = 0.013$). The number of AUS subcategory nodules was higher among malignant than among benign nodules; however,

https://doi.org/10.3348/kjr.2024.0292

Korean Journal of Radiology
KJR

Table 1. Demographic data and the characteristics of Bethesda category III thyroid nodules

Parameter	Benign	Malignant	All	*P*
Patient data				
No. of patients	246	63	309	
No. of female patients	177 (72.0)	46 (73.0)	223 (72.2)	0.866
Age, yrs, mean ± SD	51.3 ± 11.9	49.3 ± 12.0	50.9 ± 12.0	0.238
Nodule data				
No. of nodules	254 (79.4)	66 (20.6)	320	
Maximal nodule diameter, cm, median (IQR)	1.4 (1.1–1.9)	1.6 (1.2–2.5)	1.4 (1.1–2.0)	0.123
Nodule size category				0.013
<3 cm	234 (81.2)	54 (18.8)	288	
≥3 cm	20 (62.5)	12 (37.5)	32	
FNA cytology subcategory				0.145
AUS	190 (74.8)	55 (83.3)	245 (76.6)	
FLUS	64 (25.2)	11 (16.7)	75 (23.4)	
Final diagnosis				
By surgery	62 (24.4)	62 (93.9)		
	Nodular hyperplasia: 42 Follicular adenoma: 18 Benign fibrotic nodule: 1 Hyalinizing trabecular tumor: 1	Papillary thyroid carcinoma: 51 (28 conventional and 23 follicular variants) Follicular thyroid carcinoma: 10 Medullary thyroid carcinoma: 1		
By FNA or CNB	192 (75.6)	4 (6.1)		
	At least two benign diagnoses: 87 One benign diagnosis: 105	Papillary thyroid carcinoma: 3 Lymphoma: 1		

Data show the number of patients or nodules, with percentages in parentheses, unless otherwise indicated.
SD = standard deviation, IQR = interquartile range, FNA = fine needle aspiration, AUS = atypia of undetermined significance, FLUS = follicular lesion of undetermined significance, CNB = core needle biopsy

the difference was not statistically significant (P = 0.145). The proportion of papillary thyroid carcinoma (PTC) among malignant tumors was lower in larger (≥3 cm) than in smaller (<3 cm) malignant tumors (50.0% vs. 88.9%, P = 0.005); however, the proportion of non-PTC malignant tumors was higher in larger (≥3 cm) than in smaller (<3 cm) malignant tumors (50.0% vs. 11.1%, P = 0.005). While the risk of PTC was not associated with nodule size (P = 0.598), the risk of non-PTC malignant tumors increased with increasing nodule size (P < 0.001), and this risk was significantly higher in larger nodules (≥3 cm) than in smaller (<3 cm) nodules (18.8% vs. 2.1%, P = 0.001). Out of 12 large malignant tumors (≥3 cm), 9 to 11 (75.0% to 91.7%) were categorized as low or intermediate suspicion according to the US RSSs.

Interobserver agreement for classified US categories of nodules was substantial for all four RSSs, as shown by Cohen's kappa value (ACR TI-RADS, 0.72, 95% confidence interval [CI], 0.66, 0.78; ATA system, 0.77, 95% CI, 0.71, 0.82; EU-TIRADS, 0.78, 95% CI, 0.72, 0.84; and K-TIRADS, 0.73, 95% CI, 0.67, 0.80).

Malignancy Rate of Bethesda Category III Nodules According to US Category, Nodule Size, and Cytology Subcategory

Table 2 shows the malignancy rates of Bethesda category III nodules according to the US categories classified by US RSSs. In all US RSSs, there was a trend toward an increasing malignancy rate with increasing US scores for overall Bethesda category III nodules (P < 0.001) and for each AUS and FLUS subcategory nodule (P < 0.026). Larger nodules (≥3 cm) showed a higher malignancy rate than smaller (<3 cm) nodules in all Bethesda category III nodules (37.5% vs. 18.8%, P = 0.013), and in each AUS and FLUS subcategory nodule (P = 0.057 and 0.050, respectively). However, there was no difference in the malignancy rate between the 1–1.9-cm and the 2–2.9-cm nodules (18.9% and 18.0%, respectively; P = 0.881) in all Bethesda category III nodules. The larger (≥3 cm) nodules showed significantly higher malignancy rates than smaller (<3 cm) nodules in the low suspicion US categories of all US RSSs (P ≤ 0.035) and intermediate suspicion US categories of the ATA system (P =

 https://doi.org/10.3348/kjr.2024.0292

Table 2. Malignancy rates of Bethesda category III nodules according to US categories classified by four US RSS

US RSS (score)	Final diagnoses		All (n = 320)	Malignancy rate, %	*P**
	Benign (n = 254)	Malignancy (n = 66)			
ACR TI-RADS					<0.001
1	2	0	2	0	
2	42	5	47	10.6	
3	103	8	111	7.2	
4	87	32	119	26.9	
5	20	21	41	51.2	
ATA system					<0.001
2	56	7	63	11.1	
3	100	10	110	9.1	
4	79	30	109	27.5	
5	19	19	38	50.0	
EU-TIRADS					<0.001
3	142	15	157	9.6	
4	58	24	82	29.3	
5	54	27	81	33.3	
K-TIRADS					<0.001
2	2	0	2	0	
3	154	17	171	9.9	
4	87	30	117	25.6	
5	11	19	30	63.3	

**P*-values represent the trend toward an increasing malignancy rate with increasing RSS score.
US = ultrasound, RSS = risk stratification system, ACR TI-RADS = American College of Radiology Thyroid Imaging Reporting and Data System, ATA = American Thyroid Association, EU-TIRADS = European-Thyroid Imaging Reporting and Data System, K-TIRADS = Korean-Thyroid Imaging Reporting and Data System

0.025) and K-TIRADS (P = 0.017) (Supplementary Table 1). However, there were no significant differences in malignancy risk between larger (≥3 cm) and smaller (<3 cm) nodules in the high suspicion US categories of all US RSSs ($P \geq 0.395$). The AUS subcategory nodules showed higher malignancy rates than the FLUS subcategory nodules only in the low suspicion US categories; however, the observed differences did not reach statistical significance in all US RSSs (10.2%–12.4% vs. 2.3%–2.4%, respectively; P = 0.084–0.146).

Association of US Categories, Nodule Size, and Cytology Subcategory With Malignancy in Bethesda Category III Nodules

Multivariable logistic regression analyses showed that the intermediate- and high-suspicion US categories of all RSSs and larger nodule size (≥3 cm) independently increased malignancy risk in all Bethesda category III nodules ($P \leq 0.001$ for all) (Supplementary Table 2). The AUS subcategory was not significantly associated with malignancy in all US RSSs ($P \geq 0.062$) except the EU-TIRADS (P = 0.049). Subgroup analysis showed that nodule size (≥3 cm) was independently associated with malignancy in the subgroup of low suspicion US categories of all US RSSs ($P \leq 0.010$) and in the subgroup of intermediate suspicion US categories of the ATA system (P = 0.030) and K-TIRADS (P = 0.020). However, there was no association between large nodule size (≥3 cm) and malignancy in the high suspicion US categories of all US RSSs ($P \geq 0.128$) in multivariable analyses. Subgroup analysis of cytology subcategory showed that the AUS subcategory was independently associated with malignancy only in the low suspicion US categories of all RSSs ($P \leq 0.049$) except the ACR TI-RADS (P = 0.089). However, the AUS subcategory was not predictive of malignancy in intermediate- and high suspicion categories of all US RSSs, according to multivariable analyses ($P \geq 0.069$).

Development of Integrated Risk Stratification System for Bethesda Category III Nodules Based on US Category, Nodule Size, and Cytology Subcategory

The four-tiered USC RSS for Bethesda category III nodules was developed based on the US category, nodule size, and

https://doi.org/10.3348/kjr.2024.0292

Korean Journal of Radiology
KJR

Table 3. Integrated risk stratification system for Bethesda category III nodules based on US category, nodule size, and cytology subcategory

US category (score)	Nodule size, cm	Cytology subcategory	Calculated malignancy risk, %				Integrated risk stratification system for Bethesda category III nodules		Proposed management strategy
			ACR TI-RADS	ATA system	EU-TIRADS	K-TIRADS	Category	Range of malignancy risk, %	
Low suspicion (1/2/3)	1–2.9	FLUS	0.0 (0/35)	0.0 (0/35)	0.0 (0/32)	0.0 (0/35)	Very low risk	0	US surveillance
		AUS	8.3 (9/108)	10.2 (12/118)	9.5 (10/105)	10.2 (12/118)	Low risk	8.3–10.2	Repeat biopsy or molecular study
	≥3	Any subcategory	23.5 (4/17)	25.0 (5/20)	25.0 (5/20)	25.0 (5/20)	Intermediate risk	22.8–26.0	Repeat biopsy or molecular study
Intermediate suspicion (4)	1–2.9	Any subcategory	23.4 (25/107)	24.2 (24/99)	22.2 (20/76)	22.4 (24/107)			
	≥3	Any subcategory	50.0 (6/12)	60.0 (6/10)	66.7 (4/6)	60.0 (6/10)	High risk	35.6–62.5	Repeat biopsy or molecular study
High suspicion (5)	Any size	Any subcategory	53.7 (22/41)	50.0 (19/38)	33.3 (27/81)	63.3 (19/30)			

US = ultrasound, ACR TI-RADS = American College of Radiology Thyroid Imaging Reporting and Data System, ATA = American Thyroid Association, EU-TIRADS = European-Thyroid Imaging Reporting and Data System, K-TIRADS = Korean-Thyroid Imaging Reporting and Data System, FLUS = follicular lesion of undetermined significance, AUS = atypia of undetermined significance

cytology subcategory (Table 3). The malignancy risk of low suspicion US category nodules was further stratified into three risk categories: very low risk (size <3 cm and FLUS), low risk (size <3 cm and AUS), and intermediate risk (size ≥3 cm and AUS or FLUS). The malignancy risk of intermediate suspicion US category nodules was stratified into two categories based on nodule size: intermediate risk, size <3 cm, and high risk, size ≥3 cm. The observed malignancy risk increased as the USC RSS score increased in all US RSSs ($P < 0.001$) (Table 4): very low risk, 0%; low risk, 8.3%–10.2%; intermediate risk, 22.8%–26.0%; and high risk, 35.6%–62.5%. For nodules categorized as very low risk, US surveillance was proposed without additional diagnostic tests, considering their very low malignancy risk and relatively small size (<3 cm).

Performance of Ultrasound-Size-Cytology Risk Stratification System for Diagnosing Malignancy in Bethesda Category III Nodules: Comparison With US Risk Stratification Systems

The area under the ROC curve of the USC RSS was significantly greater than that of all US RSSs (0.708–0.763 and 0.674–0.731, $P < 0.048$) (Table 5). The diagnostic odds ratios of the USC RSS were also higher than those of the US RSSs (USC RSS, 5.60–7.07; US RSS, 4.31–5.60). In all Bethesda category III nodules, the sensitivity of the USC RSS for malignancy was 81.8%–86.4%, and the specificity was 50.0%–55.5%. The USC RSS increased the sensitivity by 6.1%–7.6% compared with the US RSSs. In the subgroup of large (≥2 cm) Bethesda category III nodules, the sensitivity of the USC RSS for malignancy was 90.5%–95.2%, and the specificity was 37.7%–42.6%. The USC RSS increased the sensitivity by 19.0%–23.8% compared with the US RSSs.

DISCUSSION

Our study demonstrated that the intermediate or high suspicion US categories independently increased the malignancy risk of Bethesda category III nodules in all US RSSs. A large nodule size (≥3 cm) independently increased the malignancy risk of low and intermediate suspicion US category nodules, while the AUS cytology subcategory independently increased the malignancy risk of low suspicion US category nodules in most US RSSs.

Our study validated that the US RSSs could stratify the malignancy risk of Bethesda category III nodules in the ACR TIRADS [22,23,39,40], ATA system [21,23,39,40], EU-TIRADS

https://doi.org/10.3348/kjr.2024.0292

Table 4. Estimated malignancy risks of categories classified by USC risk stratification system for Bethesda category III nodules

USC risk stratification system	Final diagnosis		All (n = 320)	Malignancy risk, % (95% CI)	*P**
	Benign (n = 254)	Malignant (n = 66)			
ACR TI-RADS					<0.001
1 (very low)	35	0	35	0.0 (0, 0)	
2 (low)	99	9	108	8.3 (3.1, 13.6)	
3 (intermediate)	95	29	124	23.4 (15.9, 30.8)	
4 (high)	25	28	53	52.8 (39.3, 66.2)	
ATA system					<0.001
1 (very low)	35	0	35	0.0 (0, 0)	
2 (low)	106	12	118	10.2 (4.7, 15.6)	
3 (intermediate)	90	29	119	24.4 (16.7, 32.1)	
4 (high)	23	25	48	52.1 (37.9, 66.2)	
EU-TIRADS					<0.001
1 (very low)	32	0	32	0.0 (0, 0)	
2 (low)	95	10	105	9.5 (3.9, 15.1)	
3 (intermediate)	71	25	96	26.0 (17.3, 34.8)	
4 (high)	56	31	87	35.6 (25.5, 45.7)	
K-TIRADS					<0.001
1 (very low)	35	0	35	0.0 (0, 0)	
2 (low)	106	12	118	10.2 (4.7, 15.6)	
3 (intermediate)	98	29	127	22.8 (15.5, 30.1)	
4 (high)	15	25	40	62.5 (47.5, 77.5)	

**P*-values represent the trend toward an increasing malignancy rate with an increasing classified risk score.
USC = ultrasound-size-cytology, CI = confidence interval, ACR TI-RADS = American College of Radiology Thyroid Imaging Reporting and Data System, ATA = American Thyroid Association, EU-TIRADS = European-Thyroid Imaging Reporting and Data System, K-TIRADS = Korean-Thyroid Imaging Reporting and Data System

[39,40], and K-TIRADS [20,23,39,40]. Previous studies have reported that the malignancy risk of Bethesda category III nodules can be stratified by a combination of US features and cytological subcategories [27,40,41]. However, our study showed that the AUS subcategory independently predicted malignancy only in low suspicion US category nodules and the impact of AUS subcategory on malignancy risk may differ according to US category. The impact of nodule size on malignancy risk may depend on the histology type of malignant tumors and US pattern of nodules. The large nodule size (≥3 cm) increased malignancy risk in low or intermediate suspicion US category nodules, in which the proportion of non-PTC malignant tumors among malignant tumors was higher compared with the high suspicion US category [42]. This finding may explain why nodule size predicted malignancy in Bethesda category III nodules with the low or intermediate suspicion US categories according to our results.

Our results provide several insights into the management of Bethesda category III nodules. First, malignancy can be ruled out in the very low risk category of the USC RSS, which was categorized by stratifying the malignancy risk of nodules with a low suspicion US category and FLUS subcategory according to nodule size (cutoff, 3 cm). The estimated malignancy rates (3.9%–20.9%) of nodules in the low suspicion US category and FLUS subcategory [27,40,41] are not deemed suitable for US follow-up, especially in cases of large nodules. This is due to the potential risk of aggressive local invasion or distant metastasis [43,44]. Therefore, US surveillance can be confidently allowed without repeat biopsy or molecular study for very low risk category nodules stratified by the USC RSS. This avoids additional unnecessary diagnostic intervention in 12.6%–13.8% of benign nodules and 10.0%–10.9% of all Bethesda category III nodules. Second, large (≥3 cm) nodule size was a predictor of malignancy in the subgroup of Bethesda category III nodules in the low or intermediate suspicion US category, and it was associated with the increased risk of non-PTC malignant tumors in Bethesda category III nodules. Third, our study showed that the estimated diagnostic performance of the USC RSS was similar across the four widely used US RSSs.

https://doi.org/10.3348/kjr.2024.0292

Korean Journal of Radiology
KJR

Table 5. Performance of USC RSS for diagnosing malignancy in Bethesda category III nodules: comparison with US RSSs

US/USC RSS	Sensitivity*	Specificity*	PPV*	NPV*	DOR	AUROC
ACR TI-RADS	80.3 (53/66) [70.7, 89.9]	57.9 (147/254) [51.4, 64.0]	33.1 (53/160) [25.9, 41.0]	91.9 (147/160) [86.5, 95.6]	5.60 [2.90, 10.79]	0.718 [0.665, 0.766]
ACR USC RSS	86.4 (57/66) [75.7, 93.6]	52.8 (134/254) [46.4, 59.0]	32.2 (57/177) [25.4, 39.6]	93.7 (134/143) [88.4, 97.1]	7.07 [3.36, 14.89]	0.759 [0.709, 0.805]
ATA RSS	74.2 (49/66) [62.0, 84.2]	61.4 (156/254) [55.1, 67.4]	33.3 (49/147) [25.8, 41.6]	90.2 (156/173) [84.7, 94.2]	4.59 [2.50, 8.42]	0.706 [0.653, 0.755]
ATA USC RSS	81.8 (54/66) [70.4, 90.2]	55.5 (141/254) [49.2, 61.7]	32.3 (54/167) [25.3, 40.0]	92.2 (141/153) [86.7, 95.9]	5.62 [2.87, 11.00]	0.746 [0.695, 0.793]
EU-TIRADS	77.3 (51/66) [65.3, 86.7]	55.9 (142/254) [49.6, 62.1]	31.3 (51/163) [24.3, 39.0]	90.4 (142/157) [84.7, 94,6]	4.31 [2.30, 8.07]	0.674 [0.620, 0.725]
EU-USC RSS	84.8 (56/66) [73.9, 92.5]	50.0 (127/254) [43.7, 56.3]	30.6 (56/183) [24.0, 37.8]	92.7 (127/137) [87.0, 96.4]	5.60 [2.74, 11.46]	0.708 [0.655, 0.757]
K-TIRADS	74.2 (49/66) [62.0, 84.2]	61.4 (156/254) [55.1, 67.4]	33.3 (49/147) [25.8, 41.6]	90.2 (156/173) [84.7, 94.2]	4.59 [2.50, 8.42]	0.731 [0.679, 0.779]
K-USC RSS	81.8 (54/66) [70.4, 90.2]	55.5 (141/254) [49.2, 61.7]	32.3 (54/167) [25.3, 40.0]	92.2 (141/153) [86.7, 95.9]	5.62 [2.87, 11.00]	0.763 [0.712, 0.808]

Data in square brackets are 95% confidence intervals. The cutoff for the diagnosis of malignant tumors were the intermediate risk category in the USC RSS and the intermediate suspicion category in the US RSSs.
*Data are percentage with the raw data in parentheses.
USC = ultrasound-size-cytology, RSS = risk stratification system, US = ultrasound, PPV = positive predictive value, NPV = negative predictive value, DOR = diagnostic odds ratio, AUROC = area under the receiver operating characteristic, ACR TI-RADS = American College of Radiology Thyroid Imaging Reporting and Data System, ATA = American Thyroid Association, EU-TIRADS = European-Thyroid Imaging Reporting and Data System, K-TIRADS = Korean-Thyroid Imaging Reporting and Data System

This study had several limitations. First, it was retrospective, and the exclusion of patients who did not undergo repeat biopsy or those that had no final diagnosis might have introduced a selection bias. Second, the reference standard for benign diagnosis was based on one benign FNA or CNB in 41.3% of benign nodules to reduce selection bias, which may have resulted in false-negative results in rare cases. However, this possibility may not have significantly affected the results because the malignancy risk was very low in Bethesda category III nodules with one benign result on repeat FNA [45]. Third, the diagnostic performance of the USC RSS needs to be validated in further studies.

In conclusion, the malignancy risk of Bethesda category III nodules can be stratified by the USC RSS based on the US category classified by the US RSS, nodule size, and cytology subcategory. The performance of the USC RSS for diagnosing malignancy was superior to that of the US RSS alone in Bethesda category III nodules, and very low risk category nodules classified by the USC RSS can be conservatively managed with US surveillance, avoiding unnecessary diagnostic intervention.

Supplement

The Supplement is available with this article at https://doi.org/10.3348/kjr.2024.0292.

Availability of Data and Material

The datasets generated or analyzed during the study are available from the corresponding author on reasonable request

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: Dong Gyu Na. Data curation: Hye Shin Ahn, Ji-Hoon Kim. Formal analysis: Hye Shin Ahn, Dong Gyu Na. Investigation: Dong Gyu Na. Methodology: Dong Gyu Na. Software: Hye Shin Ahn, Dong Gyu Na. Supervision: Dong Gyu Na. Visualization: Hye Shin Ahn, Dong Gyu Na. Writing—original draft: Hye Shin Ahn. Writing—review & editing: Dong Gyu Na, Ji-Hoon Kim.

ORCID IDs
Hye Shin Ahn
https://orcid.org/0000-0001-7260-7467
Dong Gyu Na
https://orcid.org/0000-0001-6422-1652
Ji-hoon Kim
https://orcid.org/0000-0002-6349-6950

Funding Statement
None

REFERENCES

1. Nayar R, Ivanovic M. The indeterminate thyroid fine-needle aspiration: experience from an academic center using terminology similar to that proposed in the 2007 National Cancer Institute thyroid fine needle aspiration state of the science conference. *Cancer* 2009;117:195-202
2. Theoharis CG, Schofield KM, Hammers L, Udelsman R, Chhieng DC. The Bethesda thyroid fine-needle aspiration classification system: year 1 at an academic institution. *Thyroid* 2009;19:1215-1223
3. Ohori NP, Nikiforova MN, Schoedel KE, LeBeau SO, Hodak SP, Seethala RR, et al. Contribution of molecular testing to thyroid fine-needle aspiration cytology of "follicular lesion of undetermined significance/atypia of undetermined significance." *Cancer Cytopathol* 2010;118:17-23
4. Renshaw AA. Should "atypical follicular cells" in thyroid fine-needle aspirates be subclassified? *Cancer Cytopathol* 2010;118:186-189
5. Jo VY, Stelow EB, Dustin SM, Hanley KZ. Malignancy risk for fine-needle aspiration of thyroid lesions according to the Bethesda system for reporting thyroid cytopathology. *Am J Clin Pathol* 2010;134:450-456
6. Rabaglia JL, Kabbani W, Wallace L, Holt S, Watumull L, Pruitt J, et al. Effect of the Bethesda system for reporting thyroid cytopathology on thyroidectomy rates and malignancy risk in cytologically indeterminate lesions. *Surgery* 2010;148:1267-1272; discussion 1272-1273
7. Vanderlaan PA, Krane JF, Cibas ES. The frequency of 'atypia of undetermined significance' interpretations for thyroid fine-needle aspirations is negatively correlated with histologically proven malignant outcomes. *Acta Cytol* 2011;55:512-517
8. Sullivan PS, Hirschowitz SL, Fung PC, Apple SK. The impact of atypia/follicular lesion of undetermined significance and repeat fine-needle aspiration: 5 years before and after implementation of the Bethesda system. *Cancer Cytopathol* 2014;122:866-872
9. Allen L, Al Afif A, Rigby MH, Bullock MJ, Trites J, Taylor SM, et al. The role of repeat fine needle aspiration in managing indeterminate thyroid nodules. *J Otolaryngol Head Neck Surg* 2019;48:16
10. Evranos Ogmen B, Aydin C, Kilinc I, Aksoy Altinboga A, Ersoy R, Cakir B. Can repeat biopsies change the prognoses of AUS/FLUS nodule? *Eur Thyroid J* 2020;9:92-98
11. Ali SZ, Baloch ZW, Cochand-Priollet B, Schmitt FC, Vielh P, VanderLaan PA. The 2023 Bethesda system for reporting thyroid cytopathology. *Thyroid* 2023;33:1039-1044
12. Kholová I, Ludvíková M. Thyroid atypia of undetermined significance or follicular lesion of undetermined significance: an indispensable Bethesda 2010 diagnostic category or waste garbage? *Acta Cytol* 2014;58:319-329
13. Straccia P, Rossi ED, Bizzarro T, Brunelli C, Cianfrini F, Damiani D, et al. A meta-analytic review of the Bethesda system for reporting thyroid cytopathology: has the rate of malignancy in indeterminate lesions been underestimated? *Cancer Cytopathol* 2015;123:713-722
14. 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
15. Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedüs L, et al. American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules--2016 update. *Endocr Pract* 2016;22:622-639
16. Ha EJ, Chung SR, Na DG, Ahn HS, Chung J, Lee JY, et al. 2021 Korean thyroid imaging reporting and data system and imaging-based management of thyroid nodules: Korean Society of Thyroid Radiology consensus statement and recommendations. *Korean J Radiol* 2021;22:2094-2123
17. Haddad RI, Bischoff L, Ball D, Bernet V, Blomain E, Busaidy NL, et al. Thyroid carcinoma, version 2.2022, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw* 2022;20:925-951
18. Durante C, Hegedüs L, Czarniecka A, Paschke R, Russ G, Schmitt F, et al. 2023 European Thyroid Association clinical practice guidelines for thyroid nodule management. *Eur Thyroid J* 2023;12:e230067
19. Borson-Chazot F, Buffet C, Decaussin-Petrucci M, Cao CD, Drui D, Leboulleux S, et al. SFE-AFCE-SFMN 2022 consensus on the management of thyroid nodules: synthesis and algorithms. *Ann Endocrinol (Paris)* 2022;83:440-453
20. Hong MJ, Na DG, Baek JH, Sung JY, Kim JH. Cytology-ultrasonography risk-stratification scoring system based on fine-needle aspiration cytology and the Korean-thyroid imaging reporting and data system. *Thyroid* 2017;27:953-959
21. Valderrabano P, McGettigan MJ, Lam CA, Khazai L, Thompson ZJ, Chung CH, et al. Thyroid nodules with indeterminate cytology: utility of the American Thyroid Association sonographic patterns for cancer risk stratification. *Thyroid* 2018;28:1004-1012
22. Dickey MV, Nguyen A, Wiseman SM. Cancer risk estimation

using American College of Radiology thyroid imaging reporting and data system for cytologically indeterminate thyroid nodules. *Am J Surg* 2022;224:653-656
23. Staibano P, Forner D, Noel CW, Zhang H, Gupta M, Monteiro E, et al. Ultrasonography and fine-needle aspiration in indeterminate thyroid nodules: a systematic review of diagnostic test accuracy. *Laryngoscope* 2022;132:242-251
24. Rosario PW, Calsolari MR. Importance of cytological subclassification of thyroid nodules with Bethesda category III cytology (AUS/FLUS) into architectural atypia only and nuclear atypia: a prospective study. *Diagn Cytopathol* 2017;45:604-607
25. Gan TR, Nga ME, Lum JH, Wong WM, Tan WB, Parameswaran R, et al. Thyroid cytology-nuclear versus architectural atypia within the "atypia of undetermined significance/follicular lesion of undetermined significance" Bethesda category have significantly different rates of malignancy. *Cancer Cytopathol* 2017;125:245-256
26. Eisa N, Khan A, Akhter M, Fensterwald M, Saleem S, Fananapazir G, et al. Both ultrasound features and nuclear atypia are associated with malignancy in thyroid nodules with atypia of undetermined significancence. *Ann Surg Oncol* 2018;25:3913-3918
27. Rosario PW. Thyroid nodules with atypia or follicular lesions of undetermined significance (Bethesda category III): importance of ultrasonography and cytological subcategory. *Thyroid* 2014;24:1115-1120
28. Mehta RS, Carty SE, Ohori NP, Hodak SP, Coyne C, LeBeau SO, et al. Nodule size is an independent predictor of malignancy in mutation-negative nodules with follicular lesion of undetermined significance cytology. *Surgery* 2013;154:730-736; discussion 736-738
29. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;6:e012799
30. Joo L, Na DG, Kim JH, Seo H. Comparison of core needle biopsy and repeat fine-needle aspiration in avoiding diagnostic surgery for thyroid nodules initially diagnosed as atypia/follicular lesion of undetermined significance. *Korean J Radiol* 2022;23:280-288
31. 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;6:225-237
32. 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
33. Kwon D, Kulich M, Mack WJ, Monedero RM, Joyo E, Angell TE. Malignancy risk of thyroid nodules that are not classifiable by the American Thyroid Association ultrasound risk stratification system: a systematic review and meta-analysis. *Thyroid* 2023;33:593-602
34. Gwon HY, Na DG, Noh BJ, Paik W, Yoon SJ, Choi SJ, et al. Thyroid nodules with isolated macrocalcifications: malignancy risk of isolated macrocalcifications and postoperative risk stratification of malignant tumors manifesting as isolated macrocalcifications. *Korean J Radiol* 2020;21:605-613
35. Moon WJ, Baek JH, Jung SL, Kim DW, Kim EK, Kim JY, et al. Ultrasonography and the ultrasound-based management of thyroid nodules: consensus statement and recommendations. *Korean J Radiol* 2011;12:1-14
36. Cibas ES, Ali SZ. The Bethesda system for reporting thyroid cytopathology. *Thyroid* 2009;19:1159-1165
37. Na DG, Min HS, Lee H, Won JK, Seo HB, Kim JH. Role of core needle biopsy in the management of atypia/follicular lesion of undetermined significance thyroid nodules: comparison with repeat fine-needle aspiration in subcategory nodules. *Eur Thyroid J* 2015;4:189-196
38. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174
39. Słowińska-Klencka D, Wysocka-Konieczna K, Klencki M, Popowicz B. Diagnostic value of six thyroid imaging reporting and data systems (TIRADS) in cytologically equivocal thyroid nodules. *J Clin Med* 2020;9:2281
40. Yoo WS, Ahn HY, Ahn HS, Chung YJ, Kim HS, Cho BY, et al. Malignancy rate of Bethesda category III thyroid nodules according to ultrasound risk stratification system and cytological subtype. *Medicine (Baltimore)* 2020;99:e18780
41. Larcher de Almeida AM, Delfim RLC, Vidal APA, Chaves MCDCM, Santiago ACL, Gianotti MF, et al. Combining the American Thyroid Association's ultrasound classification with cytological subcategorization improves the assessment of malignancy risk in indeterminate thyroid nodules. *Thyroid* 2021;31:922-932
42. Hong MJ, Na DG, Baek JH, Sung JY, Kim JH. Impact of nodule size on malignancy risk differs according to the ultrasonography pattern of thyroid nodules. *Korean J Radiol* 2018;19:534-541
43. Machens A, Holzhausen HJ, Dralle H. The prognostic value of primary tumor size in papillary and follicular thyroid carcinoma. *Cancer* 2005;103:2269-2273
44. Nguyen XV, Roy Choudhury K, Tessler FN, Hoang JK. Effect of tumor size on risk of metastatic disease and survival for thyroid cancer: implications for biopsy guidelines. *Thyroid* 2018;28:295-300
45. Kim GR, Yoon JH, Kim EK, Moon HJ, Kwak JY. Benign aspirates on follow-up FNA may be enough in patients with initial Atypia of undetermined significance/follicular lesion of undetermined significance. *Int J Endocrinol* 2014;2014:354612

 https://doi.org/10.3348/kjr.2024.0292