Original Article | Thyroid

eISSN 2005-8330
https://doi.org/10.3348/kjr.2023.0215
Korean J Radiol 2023;24(9):912-923

Korean Journal of Radiology
KJR

Validation of Ultrasound and Computed Tomography-Based Risk Stratification System and Biopsy Criteria for Cervical Lymph Nodes in Preoperative Patients With Thyroid Cancer

Young Hun Jeon[1], Ji Ye Lee[1,2], Roh-Eul Yoo[1,2], Jung Hyo Rhim[3], Kyung Hoon Lee[4], Kyu Sung Choi[1], Inpyeong Hwang[1,2], Koung Mi Kang[1,2], Ji-hoon Kim[2]

[1]Department of Radiology, Seoul National University Hospital, Seoul, Republic of Korea
[2]Department of Radiology, Seoul National University College of Medicine, Seoul, Republic of Korea
[3]Department of Radiology, Seoul Metropolitan Government Seoul National University Boramae Medical Center, Seoul, Republic of Korea
[4]Department of Radiology, Soonchunhyang University Seoul Hospital, Seoul, Republic of Korea

Objective: This study aimed to validate the risk stratification system (RSS) and biopsy criteria for cervical lymph nodes (LNs) proposed by the Korean Society of Thyroid Radiology (KSThR).
Materials and Methods: This retrospective study included a consecutive series of preoperative patients with thyroid cancer who underwent LN biopsy, ultrasound (US), and computed tomography (CT) between December 2006 and June 2015. LNs were categorized as probably benign, indeterminate, or suspicious according to the current US- and CT-based RSS and the size thresholds for cervical LN biopsy as suggested by the KSThR. The diagnostic performance and unnecessary biopsy rates were calculated.
Results: A total of 277 LNs (53.1% metastatic) in 228 patients (mean age ± standard deviation, 47.4 years ± 14) were analyzed. In US, the malignancy risks were significantly different among the three categories (all $P < 0.001$); however, CT-detected probably benign and indeterminate LNs showed similarly low malignancy risks ($P = 0.468$). The combined US + CT criteria stratified the malignancy risks among the three categories (all $P < 0.001$) and reduced the proportion of indeterminate LNs (from 20.6% to 14.4%) and the malignancy risk in the indeterminate LNs (from 31.6% to 12.5%) compared with US alone. In all image-based classifications, nodal size did not affect the malignancy risks (short diameter [SD] ≤ 5 mm LNs vs. SD > 5 mm LNs, $P \geq 0.177$). The criteria covering only suspicious LNs showed higher specificity and lower unnecessary biopsy rates than the current criteria, while maintaining sensitivity in all imaging modalities.
Conclusion: Integrative evaluation of US and CT helps in reducing the proportion of indeterminate LNs and the malignancy risk among them. Nodal size did not affect the malignancy risk of LNs, and the addition of indeterminate LNs to biopsy candidates did not have an advantage in detecting LN metastases in all imaging modalities.
Keywords: Thyroid neoplasm; Computed tomography; Neoplasm metastasis; Lymph nodes

INTRODUCTION

Differentiated thyroid carcinoma, particularly papillary thyroid carcinoma (PTC), is characterized by high nodal positivity on preoperative evaluation, with an incidence ranging from 30% to 80% depending on the detection method [1,2]. The local recurrence rate of PTC is reported to be up to 35% during postoperative surveillance, and lymph nodes (LNs) missed during the initial surgery are a well-recognized source of recurrent/persistent disease [3,4]. Therefore, accurate detection and localization of metastatic LN during preoperative evaluation are important

Received: January 4, 2023 **Revised:** June 26, 2023 **Accepted:** July 15, 2023
Corresponding author: Ji Ye Lee, MD, Department of Radiology, Seoul National University Hospital, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea
• E-mail: peachwh@naver.com

for precisely determining the extent of surgery and reducing the risk of repetitive surgery [4,5].

Although ultrasound (US) is the primary imaging modality for evaluating LNs in patients with thyroid cancer [5-8], it is limited by operator dependence and inability to visualize deep-seated or shadowed regions [9]. Recently, the added role of contrast-enhanced computed tomography (CT) has been used to detect additional LN metastasis [8,10-14]. Since 2016, the Korean Society of Thyroid Radiology (KSThR) has provided practice guidelines for imaging-based diagnosis of cervical LNs, including a US- and CT-based risk stratification system (RSS) and biopsy criteria for patients with suspected or proven thyroid cancer [6,15]. In this guideline, biopsy is recommended for suspicious and indeterminate LNs on US and CT, with size thresholds of 3–5 and 5 mm, respectively [6,15]. Although these categories are important in practice, the exact performance and unnecessary biopsy rates of the US-, CT-, and combined US + CT-based categories using the KSThR have not been investigated. Therefore, we aimed to validate the current US- and CT-based RSS and size thresholds for biopsy of cervical LNs as suggested by the KSThR.

MATERIALS AND METHODS

This retrospective study was approved by the Institutional Review Board of the Seoul National University Hospital (IRB No. H-1506-107-682). The requirement for patient consent was waived because of the retrospective nature of the study.

Study Population

A consecutive series of patients who underwent fine-needle aspiration (FNA) or core-needle biopsy (CNB) for neck lesions as part of the preoperative staging of thyroid cancer between December 2006 and June 2015 were included. The exclusion criteria were as follows: 1) history of other malignancies or previous surgery for thyroid cancer; 2) FNA or CNB for lesions other than LNs; and 3) no corresponding CT images. One hundred and sixty patients in this study population have been previously reported [14]. A previous study evaluated the added value of CT to US in distinguishing metastatic from benign LNs.

US- and CT- Image Acquisition

US images of the thyroid and neck were obtained by radiologists or radiology residents under the supervision of radiologists using linear transducers (7.5–15.0 MHz). Grayscale and color Doppler images were taken before biopsy. CT scans were obtained using multidetector CT scanners (16–256 channels). Pre-contrast CT images were obtained, and subsequently followed by contrast-enhanced CT scans within 40–43 s after a bolus injection of 90 mL of iodinated nonionic contrast agent (300–350 mg I/mL). Detailed CT scan parameters are described in Supplementary Material 1.

Image Analysis and Image-Based Classification of Cervical LNs

Two experienced radiologists (R–E. Y. and J–h. K., with 9 and 20 years of experience in thyroid US, respectively), who were blinded to all clinical information, independently analyzed all US and CT images. A consensus was obtained by the two reviewers for discordant cases.

LNs were classified as probably benign, indeterminate, or suspicious based on the KSThR guidelines [6]. Based on US findings, LNs were considered suspicious if they exhibited any of the following features: echogenic foci, cystic changes, cortical hyperechogenicity, or abnormal vascularity on Doppler imaging. When LN showed any cystic change, calcification, or strong or heterogeneous enhancement on CT, they were classified as suspicious. Probably benign LNs were defined as those that showed an echogenic hilum on US, hilar fat on CT, hilar vessel vascularity on US, or enhancement on CT without features of suspicious LNs. Indeterminate LNs were defined as LNs that did not have imaging features of suspicious or benign LNs. Detailed imaging criteria for US and CT are described in Supplementary Material 2.

LNs were further classified according to their final imaging classification, which was determined based on both CT and US features. For the integrative evaluation of US and CT, suspicious LNs were defined as those exhibiting suspicious features on either US or CT. Probably benign LNs were defined as LNs exhibiting probably benign features on either US or CT and no suspicious features on both US and CT, whereas indeterminate LNs were defined as LNs showing indeterminate features on both US and CT. The KSThR guidelines recommend biopsy of suspicious LNs with a short diameter (SD) > 3–5 mm and indeterminate LNs with an SD > 5 mm in preoperative patients with possible or proven thyroid cancer [6].

US-Guided Biopsy and Reference Standard

All US-guided biopsies were performed by board-certified radiologists who were aware of the size, location, and extent of the primary tumors. Grayscale and color Doppler images were obtained before the biopsy. The biopsies were targeted for LNs with indeterminate or suspicious features on imaging and for benign-looking LNs upon the attending physician's request (CT probably benign LNs, n = 19: two with large nodal size, seven with round shape, two with advanced primary tumor, four with equivocal suspicious features on US, and four with an indistinct hilum). Fine-needle aspiration and thyroglobulin (FNA-Tg) measurement was performed concurrently when the primary cancer was suspected to be a differentiated thyroid carcinoma. For CNB, spring-activated needles were used for Tg staining.

The final diagnosis of LNs was determined based on the biopsy results on a node-by-node basis because the resected and image-detected LNs could not be directly correlated in a level-by-level correlation. LN matching on US and CT was performed by a radiologist (J.Y.L., with nine years of experience in head and neck imaging). LNs were selected and labeled by matching the images with the corresponding US, CT, and final diagnosis reports. Efforts were made to reduce inaccuracies by matching the size measured on imaging with adjacent anatomical landmarks. LNs were diagnosed as malignant if they met at least one of the following criteria: 1) confirmation based on cytology or histopathology and 2) FNA-Tg cutoff > 8.3 ng/mL or greater than the serum Tg levels for cystic LNs [16]. The FNA-Tg cutoff results were only adapted for cystic LNs because of the possibility of false-positive results in ectopic benign thyroid tissue within the LN [17,18]. LNs were designated as benign when they showed benign cytology on FNA or when the biopsy specimen on CNB showed lymphoid tissue with no tumor and negative Tg immunohistochemistry results. Details of the US-guided biopsy methods and reference standards are presented in Supplementary Material 3.

Statistical Analysis

To compare the demographic characteristics of benign and metastatic LNs, student's *t*-test, chi-square test, or Fisher's exact test were used as appropriate. The frequency and malignancy risk, according to each category of the KSThR guidelines, were calculated as percentages. For each diagnostic category, malignancy risks of size stratified LNs (SD ≤ 5 mm and SD > 5 mm) were calculated and compared using the Fisher's exact test. The diagnostic performances of the US-, CT-, and US + CT-based categories were assessed by calculating their sensitivity, specificity, positive predictive value, negative predictive value, and diagnostic accuracy.

Subsequently, the LNs were dichotomized into two groups: biopsy-indicated and biopsy-not indicated, based on the criteria for biopsy from the 2021 KSThR guidelines. Seven different biopsy criteria were simulated according to the US-, CT-, and US + CT-based categories and size thresholds (Table 1): 1) any suspicious LNs; 2) suspicious LNs > 3 mm SD; 3) suspicious LNs > 5 mm SD; 4) any suspicious or indeterminate LNs; 5) suspicious LNs > 3 mm and indeterminate LNs > 5 mm; and 6) suspicious LNs > 5 mm and indeterminate LNs > 5 mm. Diagnostic performance for the diagnosis of metastasis and unnecessary biopsy rates (percentage of biopsied benign LNs among all benign LNs) were calculated. The diagnostic performance and unnecessary biopsy rates of different biopsy criteria were compared using McNemar's test.

Interobserver agreement of US- and CT-based classifications was calculated between the two radiologists using Cohen's kappa (κ) statistic. Values < 0.20 indicated slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, near-perfect agreement.

All statistical analyses were performed using MedCalc software version 20.112 (MedCalc) and SPSS version 23 for Windows (IBM). All tests were two-sided, and *P*-values < 0.05 were considered statistically significant.

RESULTS

Patient demographics and clinical characteristics of the LNs are listed in Table 2. Among 55276 patients who underwent FNA or CNB, those with a history of other malignancies (n = 6768) or previous surgery for thyroid

Table 1. Simulated different size thresholds for biopsy for cervical lymph nodes in thyroid cancer

Classification	Size stratified criteria	Description
Suspicious LNs only	(1)	Any suspicious LNs
	(2)	Suspicious LN > 3 mm in SD
	(3)	Suspicious LN > 5 mm in SD
Suspicious and indeterminate LNs	(4)	Any suspicious and indeterminate LNs
	(5)*	Suspicious > 3 mm and Indeterminate > 5 mm
	(6)*	Suspicious > 5 mm and Indeterminate > 5 mm

*Biopsy criteria of the Korean Society of Thyroid Radiology.
LN = lymph node, SD = short diameter

Korean Journal of Radiology
KJR

cancer (n = 793) were excluded. Furthermore, patients who underwent FNA or CNB for lesions other than LNs (n = 47427) and those who did not undergo corresponding CT imaging (n = 60) were excluded. Finally, 277 LNs from 228 patients (61 males and 167 females; mean age, 47.4 ± 14 [standard deviation] years; age range, 34–61 years) were included (Fig. 1).

Of the 277 LNs, 130 (46.9%) were benign and 147 (53.1%) were metastatic. The groups were balanced in terms of age (46.5 ± 11.7 years vs. 48.1 ± 15.2 [standard deviation] years, $P = 0.374$) and sex (% of female patients, 78.5% vs. 68.6%, $P = 0.384$). The mean long diameter (LD) of the primary tumors was 11.7 mm (range, 2.6–20.8 mm). The diagnoses of primary thyroid cancer included PTCs (n = 217), follicular carcinoma (n = 1), medullary carcinoma (n = 4), anaplastic carcinoma (ATC) (n = 2), poorly differentiated carcinoma (PDTC) (n = 1), metastasis (n = 1), and unspecified malignant tumors (n = 2). The mean LD of the primary tumor ($P = 0.009$), mean SD ($P < 0.001$), and LD of metastatic LNs ($P = 0.002$) were significantly larger than those of benign LNs. Regarding nodal locations, metastatic LNs were less frequently observed in level II and more frequently observed in level VI than benign LNs ($P = 0.007$).

Malignancy Risk of US-, CT-, and US + CT-Based Categories and Size

In US, the overall malignancy risks of the probably benign, indeterminate, and suspicious categories were 1.4% (95% CI: 0.04–8.2), 31.6% (95% CI: 18.7–49.9), and 84.2% (95% CI: 70.3–100.0), respectively (difference of malignancy risks, $P < 0.001$). In contrast, the malignancy risks of the CT probably benign, indeterminate, and suspicious categories were 12.5% (95% CI: 2.6–36.5), 7.5% (95% CI: 3.0–15.5), and 85.6% (95% CI: 71.9–101.2), respectively. Although the malignancy risk of suspicious LNs was significantly higher than that of the probably benign ($P < 0.001$) and indeterminate LNs ($P < 0.001$), the malignancy risk of indeterminate and probably benign LNs was not significantly different between the two groups ($P = 0.468$) (Table 3).

When LNs were categorized according to their final

Table 2. Demographic and clinical characteristics of patients and lymph nodes

Parameter	Benign	Metastasis	*P*
No. of patients	107	121	-
Sex, female:male	84:23	83:38	0.384
Age at diagnosis, yr	46.5 ± 11.7	48.1 ± 15.2	0.374
No. of LNs	130	147	-
Method of diagnosis			0.544
FNA	106 (81.5)	121 (82.4)	
CNB	19 (14.6)	17 (11.5)	
Both	5 (3.8)	9 (6.1)	
Mean maximal size of largest primary tumor, mm	10.5 ± 7.3	13.5 ± 9.7	0.009
Mean maximal size of LN, mm	9.4 ± 4.3	11.5 ± 6.7	0.002
Mean SD of LN, mm	5.8 ± 2.4	8.5 ± 5.2	< 0.001
Laterality in relation to the largest primary tumor, mm			0.270
Ipsilateral	106 (81.5)	128 (87.1)	
Contralateral	24 (18.5)	19 (12.9)	
Location			0.007
Level I	2 (1.5)	0 (0.0)	
Level II	25 (19.2)	13 (8.8)	
Level III	30 (23.1)	46 (31.3)	
Level IV	55 (42.3)	64 (43.5)	
Level V	7 (5.4)	2 (1.4)	
Level VI	6 (4.6)	18 (12.2)	
Supraclavicular fossa	5 (3.8)	4 (2.7)	

Data are mean ± standard deviation or number of lesions with percentage among benign or metastatic lymph nodes in parentheses unless otherwise indicated.
LN = lymph node, FNA = fine-needle aspiration, CNB = core needle biopsy, SD = short diameter

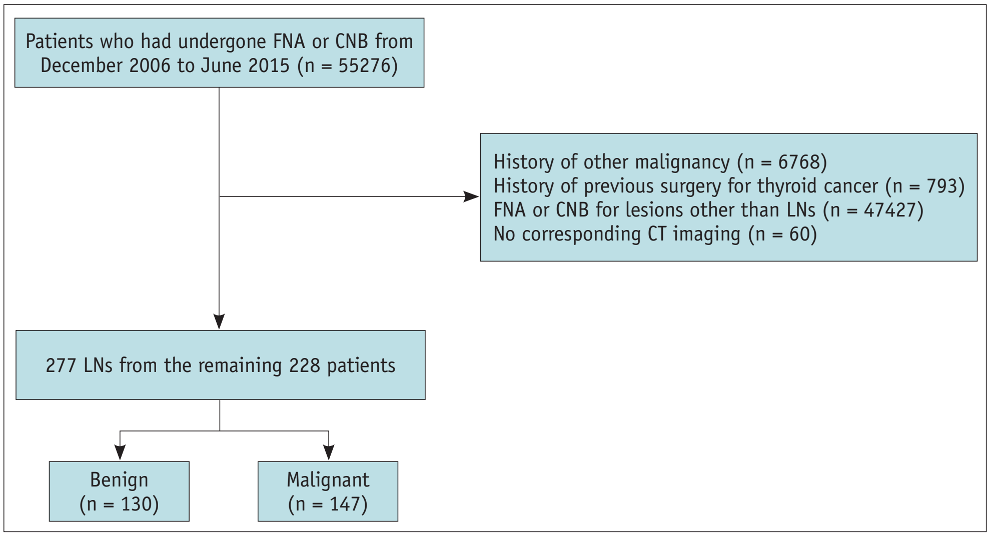

Fig. 1. Flow chart of the study population. A total of 55276 patients with screened from thyroid cancer patients who underwent US, CT and biopsy for cervical LNs. After excluding patients with past history of other malignancy, previous thyroid cancer surgery, and no CT images, 277 LNs from 228 patients were included in this study. US = ultrasound, FNA = fine-needle aspiration, CNB = core-needle biopsy, CT = computed tomography, LNs = lymph nodes

Table 3. Malignancy risk of US-, CT-, US + CT-classified LNs

Classification of LNs	All LNs*	Malignant LNs*	Malignancy risk†
US			
Suspicious	152 (54.9)	128 (87.1)	84.2 (70.3, 100.0)
Indeterminate	57 (20.6)	18 (12.2)	31.6 (18.7, 49.9)
Probably benign	68 (24.5)	1 (0.7)	1.4 (0.04, 8.2)
CT			
Suspicious	160 (57.8)	137 (93.2)	85.6 (71.9, 100.0)
Indeterminate	93 (33.6)	7 (4.8)	7.5 (3.0, 15.5)
Probably benign	24 (8.7)	3 (2.0)	12.5 (2.6, 36.5)
US + CT			
Suspicious	176 (63.5)	141 (95.9)	80.1 (67.4, 94.5)
Indeterminate	40 (14.4)	5 (3.4)	12.5 (4.1, 29.2)
Probably benign	61 (22.0)	1 (0.7)	1.6 (0.04, 9.1)

Suspicious LNs: Suspicious on either US or CT; indeterminate LNs: Indeterminate on both US and CT; and probably benign LNs: Benign features on either US or CT. Comparison of malignancy risks: US probably benign vs. indeterminate, $P < 0.001$; US suspicious vs. probably benign, $P < 0.001$; US indeterminate vs. probably benign, $P < 0.001$. CT suspicious vs. indeterminate, $P < 0.001$; CT suspicious vs. probably benign, $P < 0.001$; CT indeterminate vs. probably benign, $P = 0.437$. US + CT suspicious vs. indeterminate, $P < 0.001$, US + CT suspicious vs. probably benign, $P < 0.001$, US + CT indeterminate vs. probably benign, $P = 0.024$.
*Data represent the number of lesions, with percentages in parentheses, †Malignancy risk is calculated as the number of malignant LNs divided by the total number of LNs and is presented as a percentage (95% CI).
US = ultrasound, CT = computed tomography, LN = lymph node, CI = confidence interval

imaging categories based on combined US + CT features, suspicious LNs (having suspicious features on either US or CT), indeterminate LNs (indeterminate on both US and CT), and probably benign LNs (benign features either on US or CT) had malignancy risks of 80.1%, 12.5%, and 1.6%, respectively. Malignancy risk was significantly different between each risk category (suspicious vs. indeterminate, $P < 0.001$; indeterminate vs. probably benign, $P < 0.001$). When compared to the US criteria, the combined diagnosis of US and CT resulted in a decrease in the proportion of indeterminate LNs (from 20.6% to 14.4%) and malignancy risk (from 31.6% to 12.5%).

When LNs were further categorized by size (Table 4), indeterminate LNs (0%–22.7%) and probably benign LNs (0%) > 5 mm showed a low malignancy risk (Fig. 2). However, tiny indeterminate LNs (< 3 mm SD) showed a relatively high malignancy risk based on US (66.7%) and combined US + CT diagnosis (50%). In contrast, the malignancy risk of suspicious LNs was high overall (65.4%–100.0%), even when their sizes were ≤ 5 mm (Fig. 3). The malignancy risks of small (≤ 5 mm SD) and large (> 5 mm SD) LNs did not show significant differences in suspicious (US, $P = 0.274$; CT, $P = 0.673$; US + CT, $P = 0.422$) or indeterminate LNs (US, $P = 0.177$; CT, $P = 0.588$; US + CT, $P = 0.767$).

Interestingly, high incidence of metastatic LNs were observed in groups with tiny US indeterminate LNs observed (SD ≤ 5 mm, 37.9%–66.7%), rather than those with large US indeterminate LNs (SD > 5 mm, 0%–22.7%). On CT, the

Korean Journal of Radiology KJR

Table 4. Malignancy risk of US-, CT- and US + CT-classified LNs according to size thresholds

Size	US			CT			US + CT		
	Malignant LNs	All LNs	Malignancy risk (malignancy/all), %	Malignant LNs	All LNs	Malignancy risk (malignancy/all), %	Malignant LNs	All LNs	Malignancy risk (malignancy/all), %
Suspicious									
SD < 3 mm	1 (0.8)	1 (0.7)	100.0	1 (0.7)	1 (0.6)	100.0	1 (0.7)	1 (0.6)	100.0
3 ≤ SD < 5 mm	17 (13.3)	26 (17.1)	65.4	26 (19.1)	33 (20.6)	78.8	27 (19.1)	39 (22.2)	69.2
5 ≤ SD < 8 mm	53 (41.4)	64 (42.1)	82.8	54 (39.7)	64 (40.0)	84.4	56 (39.7)	71 (40.3)	78.9
8 ≤ SD < 10 mm	19 (14.8)	20 (13.2)	95.0	19 (14.0)	20 (12.5)	95.0	19 (13.5)	21 (11.9)	90.5
SD ≥ 10 mm	38 (29.7)	41 (27.0)	92.7	37 (27.2)	42 (26.3)	88.1	38 (27.0)	44 (25.0)	86.4
All	128 (100.0)	152 (100.0)	84.2	136 (100.0)	160 (100.0)	85.6	141 (100.0)	176 (100.0)	80.1
Indeterminate									
SD < 3 mm	2 (11.1)	3 (5.3)	66.7	1 (14.3)	5 (5.4)	5.4	1 (20.0)	2 (5.0)	50.0
3 ≤ SD < 5 mm	11 (61.1)	29 (50.9)	37.9	3 (42.9)	38 (40.9)	7.9	2 (40.0)	19 (47.5)	10.5
5 ≤ SD < 8 mm	5 (27.8)	22 (38.6)	22.7	3 (42.9)	44 (47.3)	6.8	2 (40.0)	17 (42.5)	11.8
8 ≤ SD < 10 mm	0 (0.0)	2 (3.5)	0.0	0 (0.0)	3 (3.2)	3.2	0 (0.0)	2 (5.0)	0.0
SD ≥ 10 mm	0 (0.0)	1 (1.8)	0.0	0 (0.0)	3 (3.2)	3.2	0 (0.0)	0 (0.0)	0.0
All	18 (100.0)	57 (100.0)	31.6	7 (100.0)	93 (100.0)	7.5	5 (100.0)	40 (100.0)	12.5
Probably benign									
SD < 3 mm	0 (0.0)	6 (8.8)	0.0	1 (33.3)	4 (16.7)	16.7	1 (100.0)	7 (11.5)	14.3
3 ≤ SD < 5 mm	1 (100.0)	22 (32.4)	4.5	0 (0.0)	6 (25.0)	25.0	0 (0.0)	19 (31.1)	0.0
5 ≤ SD < 8 mm	0 (0.0)	30 (44.1)	0.0	1 (33.3)	8 (33.3)	33.3	0 (0.0)	28 (45.9)	0.0
8 ≤ SD < 10 mm	0 (0.0)	4 (5.9)	0.0	1 (0.0)	3 (12.5)	12.5	0 (0.0)	3 (4.9)	0.0
SD ≥ 10 mm	0(0.0)	6 (8.8)	0.0	1 (33.3)	3 (12.5)	12.5	0 (0.0)	4 (6.6)	0.0
All	1 (100.0)	68 (100.0)	1.5	3 (100.0)	24 (100.0)	100.0	1 (100.0)	61 (100.0)	1.6

Unless otherwise indicated, data are numbers of lesions with percentage in parentheses. Comparison of malignancy risks between LNs with SD ≤ 5 mm vs. > 5 mm, US suspicious LNs, P = 0.274, US indeterminate LNs, P = 0.177, CT suspicious LNs, P = 0.673, CT indeterminate LNs, P = 0.588, US + CT suspicious LNs, P = 0.422, US + CT indeterminate LNs, P = 0.767.
US = ultrasound, CT = computed tomography, LN = lymph node, SD = short diameter

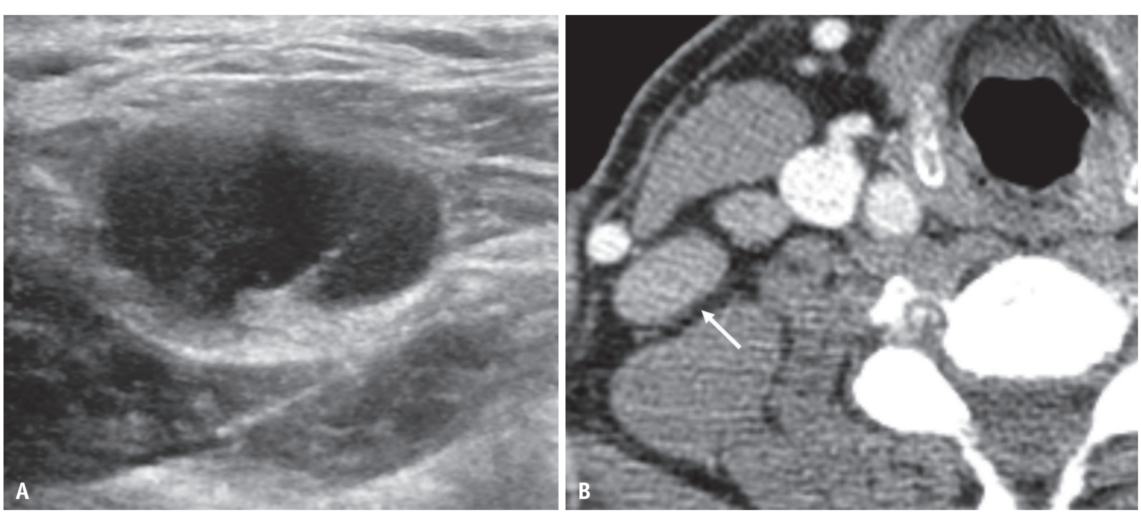

Fig. 2. A 79-year-old female with right papillary thyroid carcinoma. An enlarged lymph node (10 mm in short diameter) at the right level II with right papillary thyroid carcinoma showing a LN with a preserved echogenic hilum, which was classified as probably benign on US **(A)**. This LN shows loss of the hilum (arrow) on axial contrast-enhanced CT **(B)** and is classified as indeterminate. Core needle biopsy confirmed a benign hyperplastic lymph node. CT = computed tomography, US = ultrasound, LN = lymph node

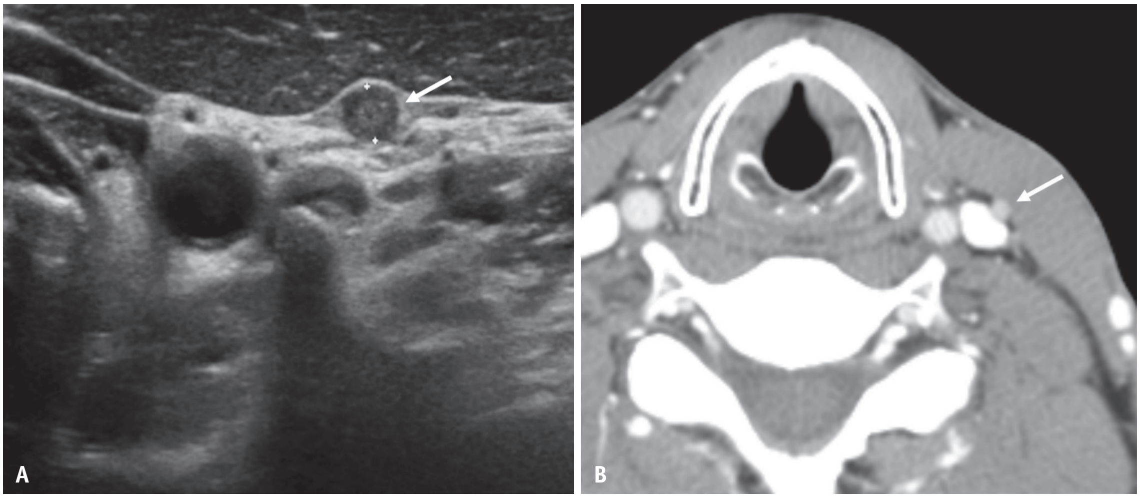

Fig. 3. A 65-year-old female with left papillary thyroid carcinoma. A small lymph node (3 mm in short diameter) at the left level III showing suspicious features of hyperechogenicity on US (arrow) **(A)**. Strong enhancement is noted on axial contrast-enhanced CT (arrow in **B**) and is classified as suspicious LN. Fine-needle aspiration confirmed a metastatic papillary thyroid carcinoma. US = ultrasound, CT = computed tomography, LN = lymph node

malignancy risk of indeterminate LNs was low, regardless of tumor size (3.2%–7.9%).

Diagnostic Performance and Rate of Unnecessary Biopsy according to KSThR LN Size Cutoff Classifications

In all imaging modalities, criteria (4), (5), and (6), as shown in Table 1, with a combination of suspicious and indeterminate LNs showed a relatively low specificity (16%–69%) and high unnecessary biopsy rate (27%–84%) (Table 5). Criteria covering suspicious LNs resulted in a significant improvement in specificity (13–30 percentage points [p.p] in criteria (1) vs. (4), all, $P < 0.001$, 13–36 p.p in (2) vs. (5), $P \leq 0.026$; and 13–39 p.p in (3) vs. (6), $P \leq 0.014$) in all imaging modalities. A significant reduction in unnecessary biopsy rate was observed on US and CT (27–66 p.p in criteria (1) vs. (4), $P \leq 0.003$; 13–36 p.p in (2) vs. (5), $P \leq 0.043$ and 13–15 p.p in (3) vs. (6), $P \leq 0.015$) while preserving sensitivity.

Among these criteria, the biopsy criteria for any suspicious LNs showed an overall high diagnostic sensitivity (87%–93%), specificity (73%–82%), and accuracy (85%–88%), with a low unnecessary biopsy rate (18%–27%) in all imaging modalities. Criteria (1) and (2) showed a significantly higher sensitivity than criterion (3) for all imaging modalities ($P < 0.012$). The diagnostic specificities and accuracies of Criteria (1), (2), and (3) did not show significant overall differences (Figs. 4-6).

Interobserver Agreement of CT Categories

In the US, the two readers showed substantial agreement for the three-tiered classification (κ = 0.801; 95% CI: 0.739–0.863) and the binary classification of suspicious LNs (κ = 0.88; 95% CI: 0.826–0.938). For CT, the two readers showed moderate agreement (κ = 0.773; 95% CI: 0.684–0.862) for determining the three-tiered classification and nearly perfect agreement (κ = 0.957; 95% CI: 0.909–1.000) for determining the binary classification (Supplementary Table 1).

DISCUSSION

Our study demonstrated that, unlike the US-based classification, CT-identified indeterminate and probably benign LN categories based on the KSThR guidelines have similarly low malignancy risks (7.5% vs. 12.5%). In all image-based classifications, the malignancy risks of LNs were not significantly affected by size. In addition, biopsy criteria including only suspicious LNs showed higher specificity (82.3%–87.7% vs. 16.2%–51.5%) and lower unnecessary biopsy rate (12.7%–14.5% vs. 33.5%–43.1%) without compromising sensitivity (74.8%–93.2% vs. 76.9%–97.9%) compared to the current criteria, including suspicious and indeterminate LNs in US, CT, and combined US + CT criteria.

Although CT is widely used as a complementary imaging modality to US for evaluating LN in thyroid cancer [5], the malignancy risk and biopsy criteria for US, CT, and combined

Korean Journal of Radiology KJR

Table 5. Diagnostic performance of US- and CT-based risk stratification criteria according to various size thresholds for biopsy

Modality	Classification	Size stratified criteria*	Sensitivity	Specificity	Accuracy	Unnecessary biopsy rate
US†	Suspicious	(1)	87 (81, 92)	82 (74, 88)	85 (80, 89)	19 (12, 28)
		(2)	86 (80, 92)	82 (74, 88)	84 (79, 88)	19 (12, 28)
		(3)	75 (67, 82)	89 (82, 93)	81 (76, 86)	12 (7, 19)
	Suspicious and indeterminate	(4)	99 (96, 100)	52 (43, 60)	77 (72, 82)	49 (37, 62)
		(5)	89 (84, 94)	66 (57, 74)	79 (73, 83)	34 (25, 45)
		(6)	78 (70, 85)	50 (43, 57)	62 (57, 68)	27 (19, 37)
CT‡	Suspicious	(1)	93 (88, 97)	82 (75, 88)	88 (84, 92)	18 (11, 27)
		(2)	92 (87, 96)	82 (75, 88)	88 (83, 91)	18 (11, 27)
		(3)	75 (67, 82)	88 (81, 93)	81 (76, 85)	12 (7, 20)
	Suspicious and indeterminate	(4)	98 (94, 100)	16 (10, 24)	88 (68, 96)	84 (69, 100)
		(5)	95 (90, 98)	46 (37, 55)	72 (66, 77)	54 (42, 68)
		(6)	77 (69, 83)	52 (43, 60)	65 (59, 71)	49 (37, 62)
US + CT§	Suspicious	(1)	96 (91, 99)	73 (65, 81)	85 (81, 89)	27 (19, 37)
		(2)	95 (90, 98)	73 (65, 81)	85 (80, 89)	27 (19, 37)
		(3)	77 (69, 83)	82 (75, 88)	79 (74, 84)	18 (11, 27)
	Suspicious and indeterminate	(4)	99 (96, 100)	46 (37, 55)	74 (69, 49)	54 (42, 68)
		(5)	97 (92, 99)	60 (51, 69)	79 (74, 84)	40 (30, 53)
		(6)	78 (71, 85)	69 (61, 77)	74 (68, 79)	31 (22, 42)

Data are percentage with 95% confidence interval in parentheses.
*Criteria (1) through (6) refer to any suspicious LNs; suspicious LNs > 3 mm in short diameter (SD); suspicious LNs > 5 mm in SD; any suspicious and indeterminate LNs; suspicious LNs > 3 mm and indeterminate LNs > 5 mm in SD; and suspicious LNs > 5 mm and indeterminate LNs > 5 mm, respectively, as shown in Table 1, †Comparison of sensitivity: US criteria (1) vs. (4), $P = 0.278$, criteria (2) vs. (5), $P = 0.368$, criteria (3) vs. (6), $P = 0.491$. Comparison of specificity: criteria (1) vs. (4), $P < 0.001$, criteria (2) vs. (5), $P = 0.005$, criteria (3) vs. (6), $P = 0.002$. Comparison of accuracy: criteria (1) vs. (4), $P = 0.491$, criteria (2) vs. (5), $P = 0.002$, criteria (3) vs. (6), $P = 0.121$. Comparison of unnecessary biopsy rate: criteria (1) vs. (4), $P = 0.003$, criteria (2) vs. (5), $P = 0.043$, criteria (3) vs. (6), $P = 0.015$, ‡Comparison of sensitivity: US criteria (1) vs. (4), $P = 0.045$, criteria (2) vs. (5), $P = 0.464$, criteria (3) vs. (6), $P = 0.689$. Comparison of specificity: criteria (1) vs. (4), $P < 0.001$, criteria (2) vs. (5), $P < 0.001$, criteria (3) vs. (6), $P < 0.001$. Comparison of accuracy: criteria (1) vs. (4), $P < 0.001$, criteria (2) vs. (5), $P < 0.001$, criteria (3) vs. (6), $P < 0.001$. Comparison of unnecessary biopsy rate: criteria (1) vs. (4), $P < 0.001$, criteria (2) vs. (5), $P < 0.001$, criteria (3) vs. (6), $P < 0.001$, §Comparison of sensitivity: US criteria (1) vs. (4), $P = 0.560$, criteria (2) vs. (5), $P = 0.556$, criteria (3) vs. (6), $P = 0.780$. Comparison of specificity: criteria (1) vs. (4), $P < 0.001$, criteria (2) vs. (5), $P = 0.026$, criteria (3) vs. (6), $P = 0.014$. Comparison of accuracy: criteria (1) vs. (4), $P = 0.002$, criteria (2) vs. (5), $P = 0.096$, criteria (3) vs. (6), $P = 0.987$. Comparison of unnecessary biopsy rate: criteria (1) vs. (4), $P = 0.048$ criteria (2) vs. (5), $P = 0.124$, criteria (3) vs. (6), $P = 0.066$.
US = ultrasound, CT = computed tomography, LNs = lymph nodes

US and CT based on KSThR have not yet been validated. Our study has additional value because previous studies have mainly focused on the diagnostic performance of US and CT features based on level-by-level analyses [10,12,19-21]. This is the first study to validate the US-, CT-, and combined US + CT-based KSThR RSS and size thresholds for the biopsy of cervical LN on a node-by-node basis.

The current KSThR guidelines adopt a size range for cervical LNs in patients with thyroid cancer, which recommends biopsy for suspicious LNs > 3–5 mm and for indeterminate LNs > 5 mm in SD [6]. However, our results demonstrated that the current biopsy criteria covering both indeterminate and suspicious LNs had relatively low specificity and high unnecessary biopsy rate. Meanwhile, biopsy criteria covering only suspicious LNs showed high specificity and reduced unnecessary biopsy rates, while maintaining high sensitivity for detecting metastatic LNs. Additionally, nodal size did not influence the malignancy risk in suspicious and indeterminate LNs. This is consistent with recent observations revealing no associations between size and malignancy risk in indeterminate LNs [22,23]. This result supports the idea that the size threshold based on SD may not effectively detect metastatic LNs during the preoperative evaluation of thyroid cancer. Recent observations have suggested that the characteristics of the primary tumor, such as multiplicity, extrathyroidal extension, or nonparallel orientation, rather than their nodal sizes, could be risk factors for metastatic LNs in indeterminate LNs [22,23]. Interestingly, small indeterminate LNs were associated with high risk of

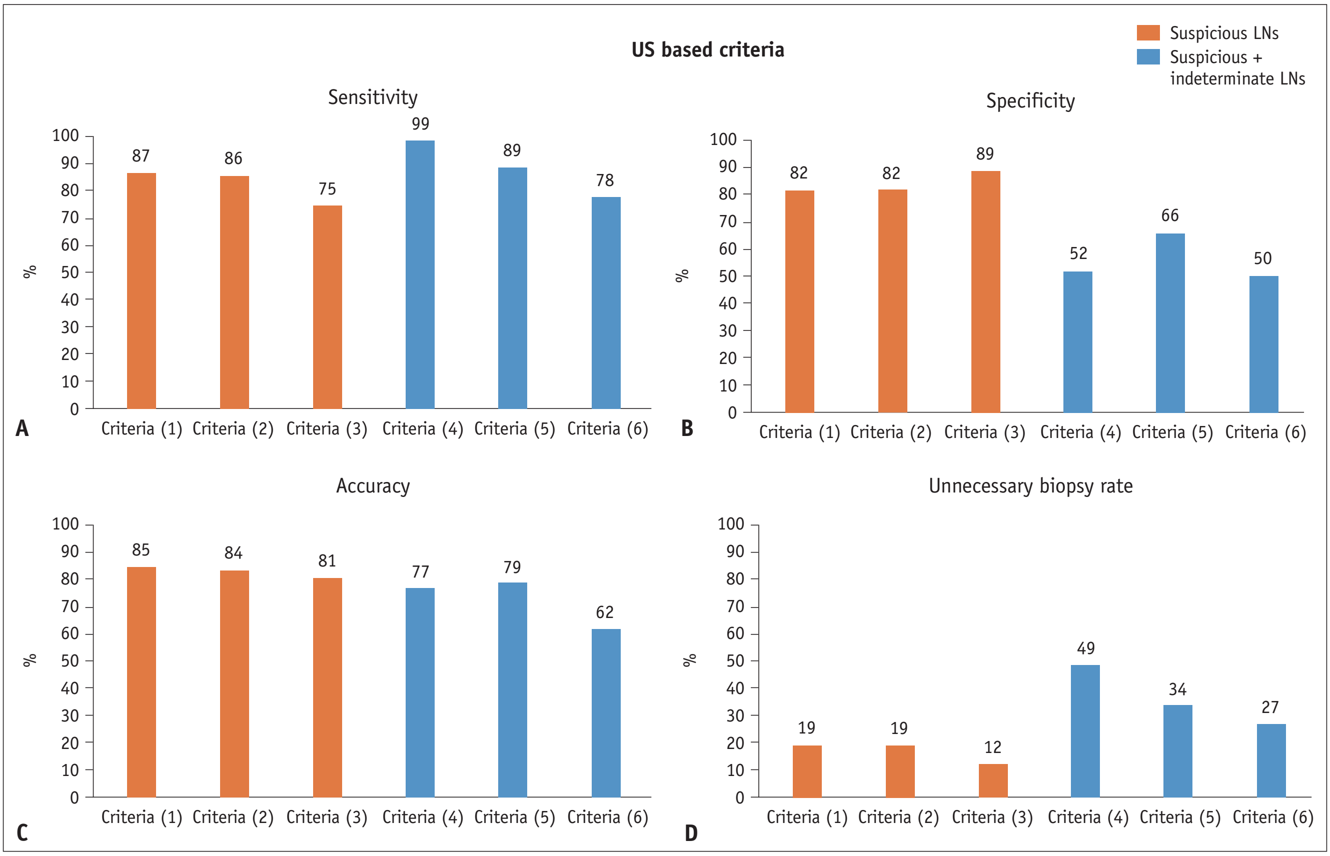

Fig. 4. Comparison of diagnostic performance and unnecessary biopsy rates according to US based classifications and size threshold simulations in Korean Society of Thyroid Radiology (KSThR) guideline for cervical LNs. Graphs show sensitivity **(A)**, specificity **(B)**, accuracy **(C)**, and unnecessary biopsy rate **(D)** of LN classifications according to the size cutoffs in KSThR guideline. Criteria (1): Any suspicious LNs. Criteria (2): Suspicious LNs > 3 mm in SD. Criteria (3): Suspicious LNs > 5 mm in SD. Criteria (4): Any suspicious and indeterminate LNs. Criteria (5): Suspicious LNs > 3 mm and indeterminate LNs > 5 mm in SD. Criteria (6): Suspicious LNs > 5 mm and indeterminate LNs > 5 mm. US = ultrasound, LN = lymph node, SD = short diameter

malignancy in this study. In this retrospective cohort, small indeterminate LNs may have selective biopsy, which may paradoxically resulted in high malignancy risk. In addition, it is also worth considering the possibility that suspicious features in tiny malignant LNs might have been undetected because of their small size and mistakenly classified as indeterminate. Given that the accurate detection of cervical LN metastasis is becoming increasingly important for determining the extent of surgery and whether to enroll in active surveillance (AS), understanding the impact of size cutoffs for each LN category is important. An appropriate biopsy indication can be selected to determine the most important factors for optimizing patient management. Omitting biopsy in indeterminate LNs spared a substantial proportion of patients (approximately 19.0%–28.7%, according to our study) from unnecessary, painful biopsy and additional medical costs. In addition, psychological stress due to the presumptive diagnosis of cervical LN metastasis is avoided, which could influence the decision to perform AS. For suspicious LNs, biopsies should be considered regardless of their size, although they may not be routinely required for large indeterminate LNs. In the case of small indeterminate LNs with high malignancy risk, it is necessary to carefully evaluate their imaging features along with the characteristics of the primary tumor to determine the indications for biopsy.

This study had several limitations. First, in addition to its retrospective nature, all patients were recruited from a single tertiary referral center; therefore, selection bias might exist. Second, the number of probably benign or indeterminate LNs was small for subcategorization, because most probably benign and small indeterminate LNs are not candidates for biopsy in practice. Considering that biopsy was performed in only a few cases for probably benign and indeterminate LN categories, the overall malignancy risk could have been overestimated. Future studies with larger sample sizes from

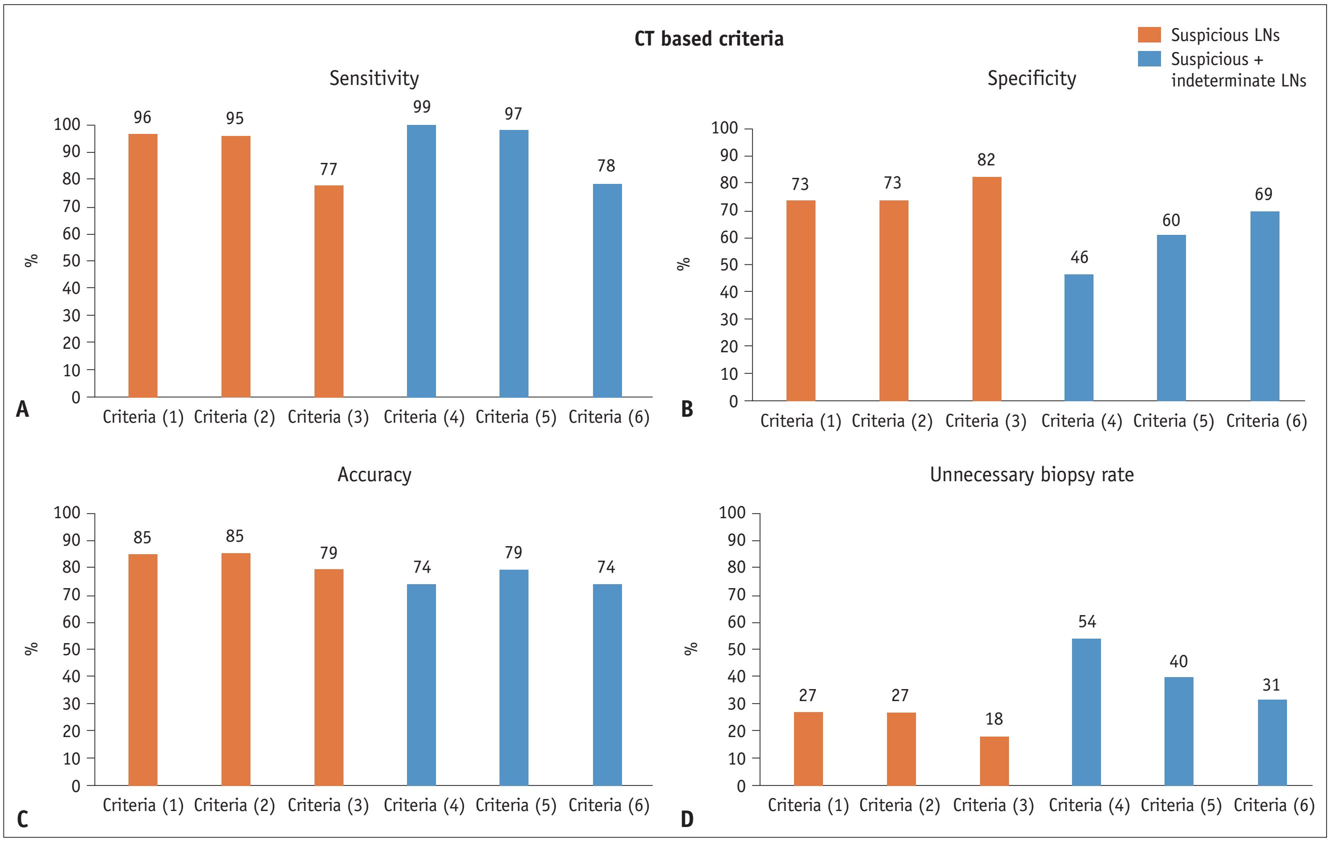

Fig. 5. Comparison of diagnostic performance and unnecessary biopsy rates according to CT based classifications and size threshold simulations in Korean Society of Thyroid Radiology (KSThR) guideline for cervical LNs. Graphs show sensitivity **(A)**, specificity **(B)**, accuracy **(C)**, and unnecessary biopsy rate **(D)** of LN classifications according to the size cutoffs in KSThR guideline. Criteria (1): Any suspicious LNs. Criteria (2): Suspicious LNs > 3 mm in SD. Criteria (3): Suspicious LNs > 5 mm in SD. Criteria (4): Any suspicious and indeterminate LNs. Criteria (5): Suspicious LNs > 3 mm and indeterminate LNs > 5 mm in SD. Criteria (6): Suspicious LNs > 5 mm and indeterminate LNs > 5 mm. CT = computed tomography, LN = lymph node, SD = short diameter

multiple centers may mitigate these limitations. Third, not all the thyroid cancer subtypes were included in this study. Most were conventional PTC, and very small cases of PDTC and ATC were included; thus, the importance of CT features suggestive of metastasis from undifferentiated thyroid cancer might have been underestimated. Finally, given that this study was based on a retrospective node-by-node analysis, the clinical relevance of these CT features in surgical management could not be determined.

In conclusion, integrative evaluation of US and CT was helpful in reducing the proportion of indeterminate LNs and malignancy risk. Nodal size did not affect the malignancy risk of LNs on either US or CT. The addition of indeterminate LNs to biopsy candidates did not provide an advantage in detecting LN metastasis using any imaging modality. These results could potentially refine the current RSS for cervical LNs in patients with thyroid cancer. Future studies with prospective data from larger populations are required to fully elucidate the relevance of these criteria.

Supplement

The Supplement is available with this article at https://doi.org/10.3348/kjr.2023.0215.

Availability of Data and Material

The datasets generated or analyzed during the study are not publicly available due the patient's privacy but are available from the corresponding author on reasonable request.

Conflicts of Interest

Ji-hoon Kim, a contributing editor of the *Korean Journal of Radiology*, was not involved in the editorial evaluation or decision to publish this article. All remaining authors have declared no conflicts of interest.

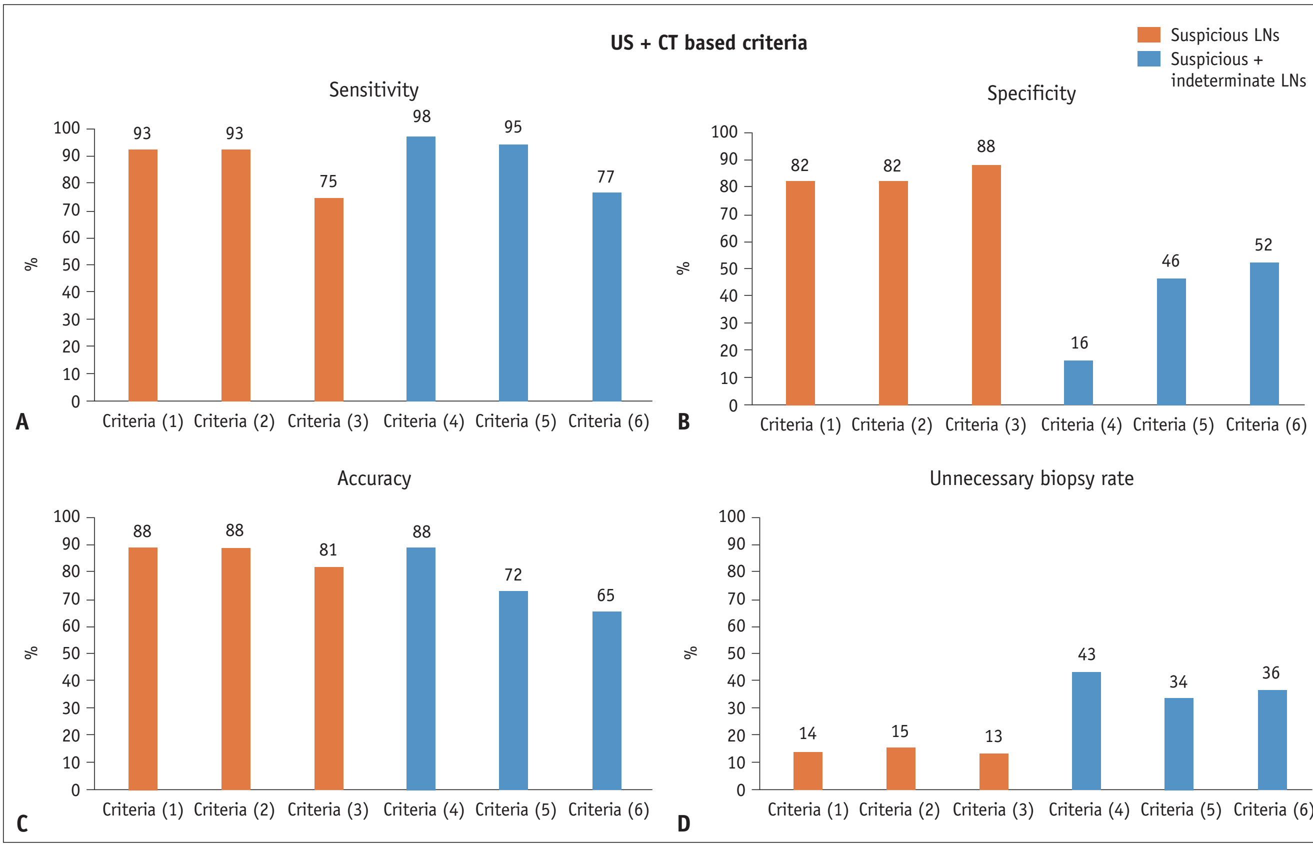

Fig. 6. Comparison of diagnostic performance and unnecessary biopsy rates according to US + CT based classifications and size threshold simulations in Korean Society of Thyroid Radiology (KSThR) guideline for cervical LNs. Graphs show sensitivity **(A)**, specificity **(B)**, accuracy **(C)**, and unnecessary biopsy rate **(D)** of LN classifications according to the size cutoffs in KSThR guideline. Criteria (1): Any suspicious LNs. Criteria (2): Suspicious LNs > 3 mm in SD. Criteria (3): Suspicious LNs > 5 mm in SD. Criteria (4): Any suspicious and indeterminate LNs. Criteria (5): Suspicious LNs > 3 mm and indeterminate LNs > 5 mm in SD. Criteria (6): Suspicious LNs > 5 mm and indeterminate LNs > 5 mm. US = ultrasound, CT = computed tomography, LN = lymph node, SD = short diameter

Author Contributions

Conceptualization: Young Hun Jeon, Ji Ye Lee, Roh-Eul Yoo, Jung Hyo Rhim, Ji-hoon Kim. Data curation: Young Hun Jeon, Ji Ye Lee, Roh-Eul Yoo, Jung Hyo Rhim, Ji-hoon Kim. Formal analysis: Young Hun Jeon, Ji Ye Lee, Roh-Eul Yoo, Ji-hoon Kim. Investigation: Young Hun Jeon, Ji Ye Lee, Roh-Eul Yoo, Jung Hyo Rhim, Ji-hoon Kim. Methodology: Young Hun Jeon, Ji Ye Lee, Roh-Eul Yoo, Ji-hoon Kim. Project administration: Young Hun Jeon, Ji Ye Lee. Resources: Young Hun Jeon, Ji Ye Lee. Supervision: all authors. Validation: Young Hun Jeon, Ji Ye Lee, Roh-Eul Yoo, Ji-hoon Kim. Visualization: Young Hun Jeon, Ji Ye Lee. Writing—original draft: Young Hun Jeon, Ji Ye Lee. Writing—review & editing: all authors.

ORCID IDs

Young Hun Jeon
https://orcid.org/0000-0001-7672-3932
Ji Ye Lee
https://orcid.org/0000-0002-3929-6254
Roh-Eul Yoo
https://orcid.org/0000-0002-5625-5921
Jung Hyo Rhim
https://orcid.org/0000-0001-5822-9770
Kyung Hoon Lee
https://orcid.org/0000-0002-4818-2937
Kyu Sung Choi
https://orcid.org/0000-0002-5175-3307
Inpyeong Hwang
https://orcid.org/0000-0002-1291-8973
Koung Mi Kang
https://orcid.org/0000-0001-9643-2008
Ji-hoon Kim
https://orcid.org/0000-0002-6349-6950

Korean Journal of Radiology
KJR

Funding Statement

None

REFERENCES

1. Stack BC Jr, Ferris RL, Goldenberg D, Haymart M, Shaha A, Sheth S, et al. American Thyroid Association consensus review and statement regarding the anatomy, terminology, and rationale for lateral neck dissection in differentiated thyroid cancer. *Thyroid* 2012;22:501-508
2. Mulla MG, Knoefel WT, Gilbert J, McGregor A, Schulte KM. Lateral cervical lymph node metastases in papillary thyroid cancer: a systematic review of imaging-guided and prophylactic removal of the lateral compartment. *Clin Endocrinol (Oxf)* 2012;77:126-131
3. Mazzaferri EL, Kloos RT. Clinical review 128: Current approaches to primary therapy for papillary and follicular thyroid cancer. *J Clin Endocrinol Metab* 2001;86:1447-1463
4. Tufano RP, Clayman G, Heller KS, Inabnet WB, Kebebew E, Shaha A, et al. Management of recurrent/persistent nodal disease in patients with differentiated thyroid cancer: a critical review of the risks and benefits of surgical intervention versus active surveillance. *Thyroid* 2015;25:15-27
5. 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
6. 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
7. Leenhardt L, Erdogan MF, Hegedus L, Mandel SJ, Paschke R, Rago T, et al. 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* 2013;2:147-159
8. Lee JY, Baek JH, Ha EJ, Sung JY, Shin JH, Kim JH, et al. 2020 Imaging guidelines for thyroid nodules and differentiated thyroid cancer: Korean Society of Thyroid Radiology. *Korean J Radiol* 2021;22:840-860
9. Hwang HS, Orloff LA. Efficacy of preoperative neck ultrasound in the detection of cervical lymph node metastasis from thyroid cancer. *Laryngoscope* 2011;121:487-491
10. Kim SK, Woo JW, Park I, Lee JH, Choe JH, Kim JH, et al. Computed tomography-detected central lymph node metastasis in ultrasonography node-negative papillary thyroid carcinoma: is it really significant? *Ann Surg Oncol* 2017;24:442-449
11. Suh CH, Baek JH, Choi YJ, Lee JH. Performance of CT in the preoperative diagnosis of cervical lymph node metastasis in patients with papillary thyroid cancer: a systematic review and meta-analysis. *AJNR Am J Neuroradiol* 2017;38:154-161
12. Lee Y, Kim JH, Baek JH, Jung SL, Park SW, Kim J, et al. Value of CT added to ultrasonography for the diagnosis of lymph node metastasis in patients with thyroid cancer. *Head Neck* 2018;40:2137-2148
13. Bongers PJ, Verzijl R, Dzingala M, Vriens MR, Yu E, Pasternak JD, et al. Preoperative computed tomography changes surgical management for clinically low-risk well-differentiated thyroid cancer. *Ann Surg Oncol* 2019;26:4439-4444
14. Yoo RE, Kim JH, Hwang I, Kang KM, Yun TJ, Choi SH, et al. Added value of computed tomography to ultrasonography for assessing LN metastasis in preoperative patients with thyroid cancer: node-by-node correlation. *Cancers (Basel)* 2020;12:1190
15. Shin JH, Baek JH, Chung J, Ha EJ, Kim JH, Lee YH, et al. Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean Society of Thyroid Radiology consensus statement and recommendations. *Korean J Radiol* 2016;17:370-395
16. Chung SR, Baek JH, Choi YJ, Sung TY, Song DE, Kim TY, et al. Diagnostic algorithm for metastatic lymph nodes of differentiated thyroid carcinoma. *Cancers (Basel)* 2021;13:1338
17. Moon JH, Kim YI, Lim JA, Choi HS, Cho SW, Kim KW, et al. Thyroglobulin in washout fluid from lymph node fine-needle aspiration biopsy in papillary thyroid cancer: large-scale validation of the cutoff value to determine malignancy and evaluation of discrepant results. *J Clin Endocrinol Metab* 2013;98:1061-1068
18. Ibrahim NB, Milewski PJ, Gillett R, Temple JG. Benign thyroid inclusions within cervical lymph nodes: an alarming incidental finding. *Aust N Z J Surg* 1981;51:188-189
19. Choi JS, Kim J, Kwak JY, Kim MJ, Chang HS, Kim EK. Preoperative staging of papillary thyroid carcinoma: comparison of ultrasound imaging and CT. *AJR Am J Roentgenol* 2009;193:871-878
20. Kim E, Park JS, Son KR, Kim JH, Jeon SJ, Na DG. Preoperative diagnosis of cervical metastatic lymph nodes in papillary thyroid carcinoma: comparison of ultrasound, computed tomography, and combined ultrasound with computed tomography. *Thyroid* 2008;18:411-418
21. Ahn JE, Lee JH, Yi JS, Shong YK, Hong SJ, Lee DH, et al. Diagnostic accuracy of CT and ultrasonography for evaluating metastatic cervical lymph nodes in patients with thyroid cancer. *World J Surg* 2008;32:1552-1558
22. Chung SR, Baek JH, Choi YJ, Sung TY, Song DE, Kim TY, et al. Risk factors for metastasis in indeterminate lymph nodes in preoperative patients with thyroid cancer. *Eur Radiol* 2022;32:3863-3868
23. Lee JY, Yoo RE, Rhim JH, Lee KH, Choi KS, Hwang I, et al. Validation of ultrasound risk stratification systems for cervical lymph node metastasis in patients with thyroid cancer. *Cancers (Basel)* 2022;14:2106