



# Malignancy risk stratification and subcategorization of K-TIRADS intermediate suspicion thyroid nodules: a retrospective multicenter study

ULTRA  
SONO  
GRAPHY

Boeun Lee<sup>1</sup>, Dong Gyu Na<sup>2</sup>, Ji-hoon Kim<sup>3</sup>

<sup>1</sup>Department of Radiology, Ewha Womans University Seoul Hospital, Ewha Womans University College of Medicine, Seoul; <sup>2</sup>Department of Radiology, Gangneung Asan Hospital, University of Ulsan College of Medicine, Gangneung; <sup>3</sup>Department of Radiology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea

**Purpose:** This study aimed to develop the ultrasonography (US) criteria for risk stratification of the Korean Thyroid Imaging Reporting and Data System (K-TIRADS) 4 nodules, and to evaluate the diagnostic yield of a modified biopsy criterion in a multicenter cohort.

**Methods:** In total, 1,542 K-TIRADS 4 nodules ( $\geq 1$  cm) were included in the study. US criteria for the subcategorization of K-TIRADS 4 nodules were developed based on high-risk US features. The diagnostic yields and false referral rates of biopsy criterion 1 (size cut-off of 1 cm), biopsy criterion 2 (size cut-off of 1.5 cm), and modified biopsy criterion 3 (size cut-off of 1 cm for K-TIRADS 4B and 1.5 cm for K-TIRADS 4A) were evaluated.

**Results:** The five high-risk US features (solid composition, marked hypoechogenicity, macrocalcification, punctate echogenic foci, and irregular margin) independently increased the malignancy risk of the K-TIRADS 4 nodules ( $P < 0.001$ ). The K-TIRADS 4 nodules could be subcategorized into higher- and lower-risk subcategories according to the number of high-risk US features: K-TIRADS 4B ( $\geq 2$  US features) and K-TIRADS 4A ( $\leq 1$  US feature). The modified biopsy criterion increased the diagnostic yield by 7.8% compared with criterion 2 and reduced the false referral rate by 15.3% compared with criterion 1 ( $P < 0.001$ ).

**Conclusion:** The K-TIRADS 4 nodules were subcategorized as K-TIRADS 4B and K-TIRADS 4A based on high-risk US features. The modified biopsy criterion 3 showed a similar diagnostic yield and reduced false referral rate compared to criterion 1.

**Keywords:** Thyroid nodule; Thyroid neoplasms; Ultrasonography;

Thyroid Imaging Reporting and Data System; Fine-needle aspiration

**Key points:** The five high-risk ultrasonography (US) features independently increased the malignancy risk of Korean Thyroid Imaging Reporting and Data System (K-TIRADS) 4 nodules. K-TIRADS 4 nodules could be subcategorized according to the number of high-risk US features: K-TIRADS 4B ( $\geq 2$  US features) and K-TIRADS 4A ( $\leq 1$  US features). The modified biopsy criterion 3 showed a similar or higher diagnostic yield compared to criteria 1 or 2.

## ORIGINAL ARTICLE

<https://doi.org/10.14366/usg.23203>

eISSN: 2288-5943

Ultrasonography 2024;43:132-140

Received: October 30, 2023

Revised: December 11, 2023

Accepted: December 12, 2023

### Correspondence to:

Dong Gyu Na, MD, PhD, Department of Radiology, Gangneung Asan Hospital, University of Ulsan College of Medicine, 38 Bangdong-gil, Gangneung 25440, Korea

Tel. +82-33-610-4310

Fax. +82-33-610-3490

E-mail: nndgna@gmail.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2024 Korean Society of Ultrasound in Medicine (KSUM)



### How to cite this article:

Lee B, Na DG, Kim JH. Malignancy risk stratification and subcategorization of K-TIRADS intermediate suspicion thyroid nodules: a retrospective multicenter study. Ultrasonography. 2024 Mar;43(2):132-140.

## Introduction

Ultrasonography (US) is the primary imaging modality for evaluating thyroid nodules [1,2]. The US risk stratification system (RSS) and the Thyroid Imaging Reporting and Data System (TIRADS) are essential in diagnosing and managing these nodules. The RSS is particularly important in selecting patients who are eligible for biopsy and in ruling out thyroid cancer. As a triage tool, RSSs must demonstrate adequate sensitivity in detecting thyroid cancer within clinically relevant thyroid nodules (>1 cm) and should aid in minimizing unnecessary biopsies of benign nodules [3].

The recently updated 2021 Korean (K)-TIRADS increased the biopsy size cut-offs for low and intermediate suspicion nodules, significantly reducing the rate of unnecessary biopsies for small (1–2 cm) nodules while maintaining high sensitivity for detecting malignancy in larger (>2 cm) nodules [4,5]. K-TIRADS 4 nodules present an intermediate malignancy risk ranging from 10%–40%, and the 2021 K-TIRADS recommends biopsy for K-TIRADS 4 nodules measuring 1–1.5 cm, taking into account US features, nodule location, clinical risk factors, and patient considerations [5]. However, the US features that increase the risk of malignancy in K-TIRADS 4 nodules have been seldom explored. Two studies have investigated the risk stratification and subcategorization of K-TIRADS 4 nodules based on US characteristics, but their findings are inconsistent [6,7]. There is a need to establish US criteria for risk stratification of K-TIRADS 4 nodules to provide useful information for making optimized biopsy decisions. Additionally, the diagnostic yield of the modified biopsy criteria, based on the subcategorization of K-TIRADS 4 nodules, should be evaluated against the biopsy criteria of the 2021 K-TIRADS to determine their effectiveness in diagnosing malignancy.

This study aimed to develop US criteria for the risk stratification of K-TIRADS 4 nodules and to evaluate the diagnostic yield of a

modified biopsy criterion derived from subcategorizing K-TIRADS 4 nodules in a multicenter cohort.

## Materials and Methods

### Compliance with Ethical Standards

This study was approved by the institutional review board of GangNeung Asan Hospital in Korea (2018-04-008-003), and informed consent was waived for this retrospective study. The methods and data reporting were performed in accordance with the Standards for Reporting Diagnostic Accuracy Studies [8].

### Study Population

This study included a subset of patient data retrospectively collected from 26 different hospitals in Korea (Thyroid Imaging Network of Korea registry), which included 5,081 consecutive patients with 5,708 thyroid nodules ( $\geq 1$  cm) [4,9]. After the exclusion of 3,581 patients with thyroid nodules classified as K-TIRADS 2, 3, or 5 by the 2021 K-TIRADS [5], the study included 1,542 thyroid nodules classified as K-TIRADS 4 in 1,500 patients (1,244 women and 256 men; median age, 55 years) (Fig. 1).

Malignant nodules were diagnosed based on histopathological results after surgery ( $n=304$ ) or malignant (Bethesda VI) fine needle aspiration (FNA) or core-needle biopsy (CNB) results ( $n=58$ ). Benign nodules were diagnosed based on histopathological results after surgery ( $n=94$ ), with at least two benign FNA or CNB results ( $n=155$ ) and one benign FNA or CNB result ( $n=931$ ).

### US Examination and Image Analysis

All US examinations were performed using 5–14 MHz linear probes and high-resolution US systems. The US images of thyroid nodules were independently reviewed by one of 17 experienced radiologists. These reviewers were blinded to both the biopsy results

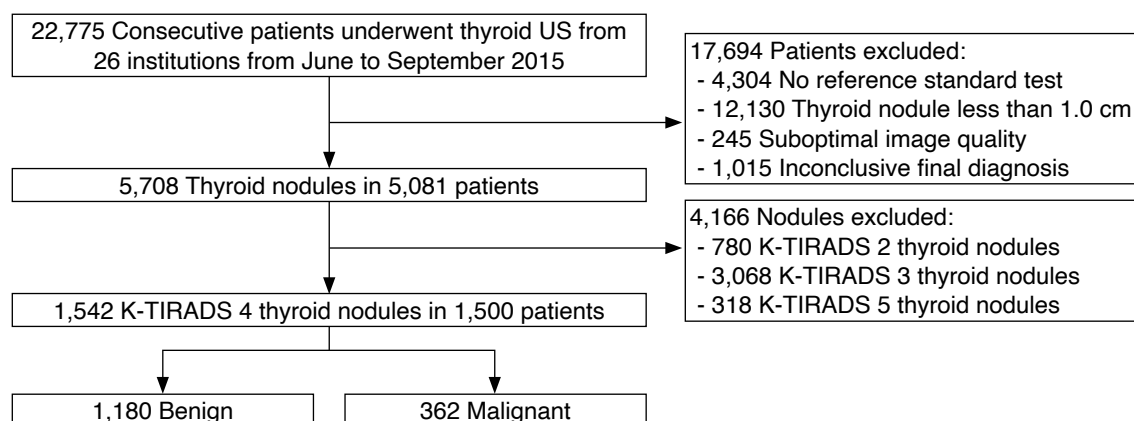


Fig. 1. Flow diagram of study patients. US, ultrasonography; K-TIRADS, Korean Thyroid Imaging Reporting and Data System.

and the final diagnoses. They evaluated the US characteristics of the nodules, focusing on composition, echogenicity, orientation, margin, and echogenic foci (calcification). The detailed methodology for US assessment of thyroid nodules has been previously described [4,9]. In our study, the US features of the nodules were defined according to the lexicon provided by the 2021 K-TIRADS [5]. A nonparallel orientation was identified when the anteroposterior diameter of a nodule exceeded its transverse diameter in the transverse plane. Marked hypoechogenicity was characterized by a nodule's echogenicity being hypoechoic or similar to that of the anterior neck muscle. A nodule that was entirely calcified (isolated macrocalcification) was classified as a K-TIRADS 4 nodule according to the 2021 K-TIRADS [5]. For the purposes of this study, such a nodule was considered to have macrocalcification and was placed into subgroup 1 for analysis.

### Development of US Criteria for the Subcategorization of K-TIRADS 4 Nodules

The K-TIRADS 4 nodules were subdivided into two subgroups based on US patterns: solid hypoechoic nodules without suspicious US features (subgroup 1) and partially cystic or iso-/hyperechoic nodules with suspicious US features (subgroup 2), as defined by the 2021 K-TIRADS [5]. The associations between US features—such as composition, echogenicity, echogenic foci (including punctate echogenic foci [PEF], macrocalcification, and rim calcification), irregular margins, nonparallel orientation, and hypoechoic halos—and malignancy were assessed within each subgroup and across all K-TIRADS 4 nodules. Criteria for the subcategorization of K-TIRADS 4 nodules were established based on high-risk US features that significantly increased the likelihood of malignancy within this group. Consequently, K-TIRADS 4 nodules were further classified into K-TIRADS 4B nodules, which have a higher risk of malignancy, and K-TIRADS 4A nodules, which have a lower risk, according to these criteria. This subcategorization was then applied to a modified biopsy criterion, wherein the size threshold for biopsy was differentiated: a 1 cm cut-off for K-TIRADS 4B nodules and a 1.5 cm cut-off for K-TIRADS 4A nodules.

### Diagnostic Yield of Subcategories and Biopsy Criteria in K-TIRADS 4 Nodules

The diagnostic yields, false referral rates, and positive predictive values (PPVs) of two US subcategories (K-TIRADS 4A and 4B) and three biopsy criteria (criterion 1: biopsy size cut-off of 1 cm for all nodules, criterion 2: biopsy size cut-off of 1.5 cm for all nodules, and criterion 3: biopsy size cut-off of 1 cm for K-TIRADS 4B and 1.5 cm for K-TIRADS 4A nodules) were evaluated with 95% confidence intervals for K-TIRADS 4 nodules. Nodules were dichotomized into

those for which a biopsy was indicated (test positivity) and those for which it was not (test negativity) according to each biopsy criterion. The diagnostic yield, or detection rate, was defined as the proportion of true-positive thyroid malignant tumors among all nodules, calculated by dividing the number of true-positive results by the total number of nodules [10]. The false referral rate, or unnecessary biopsy rate, was defined as the proportion of false-positive thyroid malignant tumors among all patients, determined by dividing the number of false-positive results by the total number of nodules [10]. The positive predictive value was defined as the proportion of true-positive malignant tumors among the nodules indicated for biopsy by each biopsy criterion.

### Statistical Analysis

Due to their nonparametric distribution, continuous variables such as age and nodule size are presented using the median and interquartile range. The Mann-Whitney U test was utilized to compare age and nodule size between the two subgroups within K-TIRADS 4 nodules. Categorical variables are expressed as frequencies and percentages for each category. For categorical variables, the chi-square test or the Fisher exact test was employed to compare the clinicopathological features of patients between the two subgroups of K-TIRADS 4 nodules, as well as to assess the difference in malignancy risk between the subcategorized nodules (K-TIRADS 4A and 4B). These tests were also applied to evaluate the malignancy risk of nodules based on the number of high-risk US features independently associated with malignancy within the overall K-TIRADS 4 nodules.

Univariable and multivariable logistic regression analyses were conducted to identify independent US predictors of malignancy within each subgroup and across all K-TIRADS 4 nodules. We compared the diagnostic yield, false referral rate, and PPV among the biopsy criteria using the McNemar test or weighted generalized score tests. The statistical analyses were carried out using IBM SPSS version 25 for Windows (IBM Corp., Armonk, NY, USA) and R 3.6.3 for Windows (R Foundation for Statistical Computing, Vienna, Austria). A P-value of less than 0.05 was considered statistically significant.

## Results

### Demographic Data and Clinical Characteristics of the Patients

The demographic data and clinical characteristics of the patients are summarized in Table 1. The K-TIRADS 4 category nodules were divided into two subgroups based on the US patterns of thyroid nodules. There was no significant difference in the malignancy rates

**Table 1.** Demographic and clinical characteristics of patients with intermediate suspicion (K-TIRADS 4) thyroid nodules

Parameter	Total	Subgroup 1 (solid hypoechoic nodules without suspicious US feature <sup>a)</sup> )	Subgroup 2 (partially cystic or Iso-/hyperechoic nodules with suspicious US features <sup>a)</sup> )	P-value
Patients				
No. of patients	1,500 <sup>b)</sup>	817	697	
No. of women	1,244 (82.9) <sup>c)</sup>	698 (85.4)	559 (80.2)	0.021 <sup>d)</sup>
Age (year), median (IQR)	55 (45–63)	54 (46–63)	55 (45–63)	0.720 <sup>b)</sup>
No. of nodules	1,542	839	703	
Maximal diameter of nodule (cm), median (IQR)	1.6 (1.2–2.3)	1.4 (1.2–2.0)	1.8 (1.3–2.6)	<0.001
≤2 cm	1,054 (68.3)	635 (75.7)	419 (59.6)	<0.001
>2 cm	488 (31.7)	204 (24.3)	284 (40.4)	
No. of malignant tumors	362 (23.5)	189 (22.5)	173 (24.6)	0.361
Papillary thyroid carcinomas <sup>d)</sup>	320 (88.4)	161 (85.1)	159 (91.9)	0.244
Follicular thyroid carcinomas	28 (7.8)	17 (9.0)	11 (6.4)	–
Other malignant tumors	14 (3.9)	11 (5.9)	3 (1.7)	–
Medullary thyroid carcinoma	5 (1.4)	5 (2.6)	0	
Anaplastic carcinoma	4 (1.1)	3 (1.6)	1 (0.6)	
Poorly differentiated carcinoma	2 (0.6)	1 (0.5)	1 (0.6)	
Lymphoma	1 (0.3)	1 (0.5)	0	
Metastasis	1 (0.3)	0	1 (0.5)	
Unspecified malignancy	1 (0.3)	1 (0.5)	0	

Values are presented as number of patients or nodules (%), unless otherwise indicated.

K-TIRADS, Korean Thyroid Imaging Reporting and Data System; US, ultrasound; IQR, interquartile range.

<sup>a)</sup>Punctate echogenic foci, irregular margin, nonparallel orientation. <sup>b)</sup>Fourteen patients had nodules in both subgroups, and these patients were excluded from the statistical comparison of age between the two subgroups. <sup>c)</sup>Thirteen patients had nodules in both subgroups, and these patients were excluded from the statistical comparison of sex between the two subgroups. <sup>d)</sup>Includes 43 follicular variant papillary thyroid carcinomas (31 in subgroup 1 and 12 in subgroup 2).

between subgroups 1 and 2 (22.5% vs. 24.6%,  $P=0.361$ ). The proportion of women was slightly higher in subgroup 1 compared to subgroup 2 ( $P=0.021$ ); however, the median ages of the two subgroups were not significantly different ( $P=0.720$ ). The median size (maximum diameter) of the nodules was larger in subgroup 2 than in subgroup 1 (1.8 cm vs. 1.4 cm,  $P<0.001$ ), and the incidence of large nodules (>2 cm) was greater in subgroup 2 than in subgroup 1 (40.4% vs. 24.3%,  $P<0.001$ ). Out of the 1,542 nodules examined, 362 (23.5%) were malignant, and 1,180 (76.5%) were benign. There was no significant difference in the distribution of histological types of malignant tumors between the two subgroups ( $P=0.244$ ).

### US Features Associated with Thyroid Malignancy in K-TIRADS 4 Nodules

Table 2 shows the US features associated with malignancy in subgroups 1 and 2 of K-TIRADS 4 nodules. In subgroup 1, marked hypoechoogenicity and macrocalcification emerged as significant independent predictors of thyroid malignancy in the multivariable

analysis ( $P=0.002$  and  $P<0.001$ , respectively). For subgroup 2, solid composition, marked hypoechoogenicity, PEF, irregular margins, and macrocalcification were identified as significant independent predictors of malignancy, with all  $P$ -values being  $\leq 0.009$ . The presence of two or three suspicious US features in subgroup 2 nodules significantly increased the risk of malignancy compared to nodules with only one suspicious feature (51.1% vs. 20.7%,  $P<0.001$ ). However, the US feature of nonparallel orientation did not significantly affect the risk of malignancy in subgroup 2 nodules ( $P=0.899$ ). Table 3 details the US features associated with malignancy across all K-TIRADS 4 nodules. Multivariable binary logistic regression analysis revealed that five US features—solid composition, marked hypoechoogenicity, PEF, irregular margins, and macrocalcification—were independently associated with an increased risk of malignancy (all  $P<0.001$ ).

### US Criteria for Subcategorization of K-TIRADS 4 Nodules

The risk of malignancy varied significantly with the number of high-risk US features present ( $P<0.001$ ), showing a tendency to

**Table 2.** US features associated with thyroid malignancy in two subgroups of intermediate suspicion (K-TIRADS 4) nodules

US feature	No. of nodules	No. of benign nodules, n (%)	No. of malignant nodules, n (%)	Malignancy risk (%)	Univariable analysis <sup>a)</sup>		Multivariable analysis <sup>a)</sup>	
					OR (95% CI)	P-value	OR (95% CI)	P-value
Subgroup 1 (solid hypoechoic nodules without suspicious features <sup>b)</sup> )								
All	839	650	189	22.5				
Marked hypoechoogenicity	343	250 (38.5)	93 (49.2)	28.1	1.55 (1.12–2.15)	0.008	1.70 (1.22–2.38)	0.002
Mixed echogenicity	208	168 (25.8)	40 (21.2)	19.2	0.77 (0.52–1.14)	0.190	–	–
Macrocalcification	141	89 (13.7)	52 (27.5)	36.4	2.39 (1.62–3.53)	<0.001	2.59 (1.74–3.86)	<0.001
Rim calcification	56	41 (6.3)	15 (7.9)	26.8	1.28 (0.69–2.37)	0.431	–	–
Hypoechoic halo	142	108 (16.6)	34 (18.0)	24.1	1.10 (0.72–1.68)	0.658	–	–
Subgroup 2 (partially cystic or Iso-/hyperechoic nodules with suspicious features <sup>b)</sup> )								
All	703	530	173	24.6				
Solid	300	207 (39.1)	93 (53.8)	31.0	1.81 (1.28–2.56)	0.001	2.09 (1.42–3.10)	<0.001
Marked hypoechoogenicity	41	22 (4.2)	19 (11.0)	46.3	2.85 (1.50–5.40)	0.001	3.53 (1.70–7.31)	0.001
Mixed echogenicity	277	192 (36.2)	85 (49.1)	30.7	1.70 (1.20–2.40)	0.003	1.38 (0.95–2.00)	0.087
Punctate echogenic foci	567	414 (78.1)	153 (88.4)	27.0	2.14 (1.29–3.57)	0.003	3.30 (1.87–5.82)	<0.001
Irregular margin	88	46 (8.7)	42 (24.3)	47.7	3.37 (2.13–5.35)	<0.001	3.60 (2.12–6.12)	<0.001
Nonparallel	152	114 (21.5)	38 (22.0)	25.0	1.03 (0.68–1.56)	0.899	–	–
Macrocalcification	124	80 (15.1)	44 (25.4)	35.5	1.92 (1.26–2.91)	0.002	1.82 (1.16–2.85)	0.009
Rim calcification	38	25 (4.7)	13 (7.5)	34.2	1.64 (0.82–3.28)	0.161	–	–
Hypoechoic halo	150	109 (20.6)	41 (23.7)	27.3	1.20 (0.80–1.81)	0.383	–	–
1 suspicious feature <sup>b)</sup>	613	486 (91.7)	127 (73.4)	20.7	0.25 (0.16–0.40)	<0.001	–	–
2 or 3 suspicious features <sup>b)</sup>	90	44 (8.3)	46 (26.6)	51.1	4.00 (2.53–6.32)	<0.001	–	–

K-TIRADS, Korean Thyroid Imaging Reporting and Data System; US, ultrasonography; OR, odds ratio; CI, confidence interval.

<sup>a)</sup>Binary logistic regression analysis. <sup>b)</sup>Punctate echogenic foci, irregular margin, nonparallel orientation.

increase as the number of high-risk US features within a nodule rose ( $P<0.001$ ) (Supplementary Table 1). Nodules exhibiting two high-risk US features had a malignancy risk of 27.0%, which was significantly greater than the risk associated with nodules having none or one high-risk US feature (10%,  $P=0.039$  and 15.4%,  $P<0.001$ , respectively). Furthermore, the malignancy risk for nodules with three high-risk US features was 50.7%, markedly higher than the risk for nodules with two high-risk US features ( $P<0.001$ ).

Based on these findings, K-TIRADS 4 nodules can be divided into two subcategories (K-TIRADS 4A: lower risk and K-TIRADS 4B: higher risk). The K-TIRADS 4B subcategory included K-TIRADS 4 nodules that showed at least two of the five following US features: (1) solid composition, (2) marked hypoechoogenicity, (3) macrocalcification, (4) PEF, and (5) irregular margin (Table 4). The K-TIRADS 4A subcategory included K-TIRADS 4 nodules with absent or one high-risk US feature. The malignancy risk in the K-TIRADS 4B category was significantly higher than that in the K-TIRADS 4A category (31.7% vs. 15.2%,  $P<0.001$ ) (Table 4). The malignancy risk in the K-TIRADS 4B category was also significantly higher than that in the K-TIRADS 4A category in subgroup 1 (28.3% vs. 16.1%,

$P<0.001$ ) and subgroup 2 (36.0% vs. 14.2%,  $P<0.001$ ).

The diagnostic yield of K-TIRADS 4B was significantly higher than that of K-TIRADS 4A (16.0% and 7.5%,  $P<0.001$ ) and the false referral rate of K-TIRADS 4B was significantly lower than that of K-TIRADS 4A (34.4% and 42.1%,  $P<0.001$ ) (Table 4).

### Diagnostic Yield of the Modified Biopsy Criterion Based on the Subcategorization of K-TIRADS 4 Nodules

Table 5 presents the diagnostic yields, false referral rates, and PPVs for two biopsy criteria from the 2021 K-TIRADS (criteria 1 and 2) alongside a modified biopsy criterion (criterion 3) based on the subcategorization of K-TIRADS 4 nodules within the overall K-TIRADS 4 category. Criterion 3 resulted in a 2.9% decrease in diagnostic yield and a 15.3% reduction in the false referral rate compared to criterion 1. In contrast, criterion 3 led to a 7.8% increase in diagnostic yield and a 16.0% rise in the false referral rate when compared to criterion 2. In the subgroup of small ( $\leq 2$  cm) K-TIRADS 4 nodules, criterion 3 reduced the diagnostic yield by 4.3% and the false referral rate by 22.5% relative to criterion 1. Meanwhile, compared to criterion 2, criterion 3 increased the

**Table 3.** US features associated with thyroid malignancy in intermediate suspicion (K-TIRADS 4) nodules

US feature	No. of nodules	No. of benign nodules, n (%)	No. of malignant nodules, n (%)	Malignancy risk (%)	Univariable analysis <sup>a)</sup>		Multivariable analysis <sup>a)</sup>	
					OR (95% CI)	P-value	OR (95% CI)	P-value
Total	1,542	1,180	362	23.5				
Solid	1,139	857 (72.6)	282 (77.9)	24.7	1.33 (1.00–1.76)	0.046	2.06 (1.45–2.94)	<0.001
Marked hypoechogenicity	384	272 (23.1)	112(30.9)	29.8	1.50 (1.15–1.94)	0.002	2.04 (1.52–2.73)	<0.001
Mixed echogenicity	485	360 (30.5)	125 (34.5)	25.8	1.20 (0.94–1.54)	0.15	–	–
Punctate echogenic foci	567	414 (35.1)	153 (42.3)	27.0	1.35 (1.06–1.72)	0.013	2.24 (1.62–3.10)	<0.001
Irregular margin	88	46 (3.9)	42 (11.6)	47.7	3.24 (2.09–5.00)	<0.001	3.58 (2.26–5.69)	<0.001
Nonparallel	152	114 (9.7)	38 (10.5)	25.0	1.10 (0.74–1.62)	0.641	–	–
Macrocalcification	265	169 (14.3)	96 (26.5)	36.0	2.16 (1.62–2.87)	<0.001	2.23 (1.66–3.00)	<0.001
Rim calcification	94	66 (5.6)	28 (7.7)	29.8	1.42 (0.89–2.24)	0.138	–	–
Hypoechoic halo	295	219 (18.6)	76 (21.0)	25.8	1.16 (0.86–1.56)	0.323	–	–

K-TIRADS, Korean Thyroid Imaging Reporting and Data System; US, ultrasonography; OR, odds ratio; CI, confidence interval.

<sup>a)</sup>Binary logistic regression analysis.

**Table 4.** Malignancy risk, diagnostic yield, and false referral rate according to the subcategories of 2021 K-TIRADS 4 nodules

Category	US criteria	No. of nodules	No. of benign nodules, n (%)	No. of malignant nodules, n (%)	Malignancy risk (%) [95% CI]	Diagnostic yield (%) [95% CI]	False referral rate (%) [95% CI]
K-TIRADS 4	1) Solid hypoechoic nodule without any of 3 suspicious US features <sup>a)</sup> 2) Partially cystic or isoechoic and hyperechoic nodule with any of 3 suspicious US features <sup>a)</sup>	1,542	1,180	362	23.5 [21.4–25.6]	–	–
K-TIRADS 4B	Presence of at least two of five high-risk US features: 1) Solid 2) Marked hypoechogenicity 3) Macrocalcification 4) Punctate echogenic foci 5) Irregular margin	777	531 (45.0)	246 (68.0)	31.7 [28.4–34.9]	16.0 (246/1,542) [14.1–17.8]	34.4 (531/1,542) [32.6–36.8]
K-TIRADS 4A	Absence or one of five high-risk US features	765	649 (55.0)	116 (32.0)	15.2 [12.6–17.7]	7.5 (116/1,542) [6.2–8.8]	42.1 (649/1,542) [39.6–44.6]

K-TIRADS, Korean Thyroid Imaging Reporting and Data System; US, ultrasonography; CI, confidence interval.

<sup>a)</sup>Punctate echogenic foci, irregular margin, nonparallel orientation.

diagnostic yield by 11.3% and the false referral rate by 23.3% (Table 5). Both criteria 2 and 3 demonstrated significantly lower diagnostic yields and false referral rates than criterion 1 ( $P<0.001$  for both). Criterion 3 also showed significantly higher diagnostic yield, false referral rate, and PPV compared to criterion 2 (all  $P<0.001$ ). Additionally, criterion 3 exhibited significantly greater sensitivity for detecting malignancy in both the overall and small ( $\leq 2$  cm) K-TIRADS 4 nodules than biopsy criterion 2 (87.6% vs. 54.4%,  $P<0.001$  and 82.0% vs. 34.0%,  $P<0.001$ , respectively).

## Discussion

This study demonstrated that the presence of five high-risk US features—solid composition, marked hypoechogenicity, macrocalcification, PEF, and irregular margins—independently increased the malignancy risk of K-TIRADS 4 nodules. Furthermore, we found that K-TIRADS 4 nodules could be stratified into higher and lower risk subcategories based on the number of these high-risk US features. Specifically, K-TIRADS 4B is characterized by the presence of two or more high-risk US features, while K-TIRADS 4A is defined by the absence or presence of only one high-risk US feature. The modified biopsy criterion (criterion 3) based on the

**Table 5.** Diagnostic yield and false referral rate of modified biopsy criteria of the 2021 K-TIRADS in 2021 K-TIRADS 4 nodules

Parameter	Criterion 1 (>1.0 cm)	Criterion 2 (>1.5 cm)	Criterion 3 <sup>a)</sup> (4B >1 cm, 4A >1.5 cm)
All nodules (n=1,542)			
Diagnostic yield	23.5 (362/1,542) [21.4–25.6]	12.8 (197/1,542) [11.1–14.4]	20.6 (317/1,542) [18.5–22.6]
False referral rate	76.5 (1,180/1,542) [74.4–78.6]	45.2 (697/1,542) [42.7–47.7]	61.2 (943/1,542) [58.7–63.6]
Positive predictive value	23.5 (362/1,542) [21.4–25.6]	22.0 (197/894) [19.3–24.8]	25.2 (317/1,260) [22.8–27.6]
Small (≤2 cm) nodules (n=1,054)			
Diagnostic yield	23.7 (250/1,054) [21.2–26.3]	8.1 (85/1,054) [6.4–9.7]	19.4 (205/1,054) [17.1–21.9]
False referral rate	76.3 (804/1,054) [73.7–78.9]	30.5 (321/1,054) [27.7–33.2]	53.8 (567/1,054) [50.8–56.8]
Positive predictive value	23.7 (250/1,054) [21.1–26.3]	20.9 (85/406) [17.0–24.9]	26.6 (205/772) [23.4–29.7]

Values are presented as percentages (number of nodules, numerator/denominator). Values in square brackets are 95% confidence intervals.

K-TIRADS, Korean Thyroid Imaging Reporting and Data System.

<sup>a)</sup>Modified biopsy criterion by the subcategorization of K-TIRADS 4 nodules: 4B includes nodules with two or more of the five high-risk ultrasound features including solid composition, marked hypoechogenicity, macrocalcification, punctate echogenic foci, and irregular margin. 4A indicates nodules with absent or one high-risk ultrasound features.

subcategorization of K-TIRADS 4 nodules reduced the false referral rate (unnecessary biopsy rate) by 15.3% and 22.5% compared with criterion 1 (biopsy size cut-off of 1 cm) and increased the diagnostic yield (detection rate) for malignancy by 7.8% and 11.3% compared with criterion 2 (biopsy size cut-off of 1.5 cm) in the overall and small (≤2 cm) K-TIRADS 4 nodules, respectively.

Marked hypoechogenicity and macrocalcification were identified as independent US features that increased the risk of malignancy in all K-TIRADS 4 nodules, as well as within each subgroup of nodules. These findings align with those of previous studies [11,12]. Lee et al. [12] found that nodules with moderately or markedly hypoechoic characteristics (similar in echogenicity or hypoechoic compared to neck muscle) had a significantly higher risk of malignancy than those that were mildly hypoechoic or iso-/hyperechoic across all subgroups, based on their composition and suspicious US features. Similarly, Shin et al. [11] found that macrocalcification was independently associated with malignancy in all nodules and significantly increased the malignancy risk in both subgroups of K-TIRADS 4 nodules. These consistent results strongly suggest that marked hypoechogenicity and macrocalcification are US features that independently elevate the risk of malignancy in K-TIRADS 4 nodules. A trend was also observed for an increased risk of malignancy with a higher number of high-risk US features in K-TIRADS 4 nodules, which is in line with the outcomes of previous research [13,14]. The US patterns of higher-risk K-TIRADS 4 nodules (K-TIRADS 4B) can be further categorized based on the subgroups of K-TIRADS 4 nodules. These subcategories include: (1) solid hypoechoic K-TIRADS 4 nodules with at least one feature of marked hypoechogenicity or macrocalcification, (2) solid isoechoic K-TIRADS 4 nodules with at least one feature of macrocalcification, punctate echogenic foci, or irregular margin, and (3) partially cystic K-TIRADS

4 nodules with at least two features of marked hypoechogenicity, macrocalcification, punctate echogenic foci, or irregular margin.

The sensitivity and unnecessary biopsy rate for the 2021 K-TIRADS using biopsy criterion 1 (a size cut-off of 1 cm for K-TIRADS 4 nodules) were the highest. In contrast, those for the 2021 K-TIRADS with biopsy criterion 2 (a size cut-off of 1.5 cm for K-TIRADS 4 nodules) were the lowest when compared to the American College of Radiology (ACR)-TIRADS and European-TIRADS in small (≤2 cm) thyroid nodules [9]. The diagnostic performance of current RSSs or TIRADS in detecting malignant tumors is primarily influenced by the nodule size cut-off for biopsy [15–17]. Utilizing higher size cut-offs for biopsy, such as those in ACR-TIRADS, results in relatively higher specificity, a lower rate of unnecessary biopsies, and greater diagnostic accuracy, but this approach inevitably reduces sensitivity in the diagnosis of malignancy [18–20]. The approach of lowering the rate of unnecessary biopsies at the expense of reduced sensitivity may be suitable for small (≤2 cm) malignant tumors, which generally have a better prognosis than larger (>2 cm) malignant tumors. Ideally, the most effective RSS should minimize unnecessary biopsies while preserving adequate sensitivity for detecting malignancy [3]. However, there is currently no consensus on the optimal sensitivity level for identifying malignancies in small (≤2 cm) thyroid nodules.

The diagnostic yields of biopsy criteria were analyzed in a subgroup of small (≤2 cm) K-TIRADS 4 nodules to discern the differences in diagnostic yield among the three biopsy criteria specific to small nodules. Our study revealed that the diagnostic yield and sensitivity of biopsy criterion 2 for detecting malignancy were notably low in small (≤2 cm) K-TIRADS 4 nodules. In contrast, the modified biopsy criterion 3, which is based on the subcategorization of K-TIRADS 4, demonstrated a significant improvement in diagnostic yield and sensitivity for malignancy, albeit

with an increased false referral rate compared to biopsy criterion 2. It may be prudent to employ biopsy criterion 2 (size cut-off of 1.5 cm) for most K-TIRADS 4 nodules that lack high-risk clinical or ultrasonographic features of metastasis or gross extrathyroidal extension, despite the relatively low diagnostic yield, because the majority of malignant tumors missed are small (<1.5 cm) low-risk tumors. However, our findings suggest that the modified biopsy criterion 3 could be selectively utilized in place of biopsy criterion 2 for patients who prioritize a higher detection rate for malignancy and aim to minimize the oversight of small malignant tumors.

Subcategorizing K-TIRADS 4 nodules by risk stratification could provide valuable insights for the management of thyroid nodules. First, a K-TIRADS 4B nodule may be considered for biopsy if a patient presents with multiple K-TIRADS 4 nodules of comparable size. Second, modified biopsy criterion 3 might be applied in patients with clinical risk factors, particularly when the K-TIRADS 4 nodule is in proximity to vital structures such as the trachea or recurrent laryngeal nerves. Third, employing modified biopsy criterion 3 instead of criterion 2 could enhance the detection of malignancy in children, who may require more aggressive management [21–24]. However, the effectiveness of these modified biopsy criteria, based on subcategorization, will need future validation in pediatric populations.

A recent study from a single institution [7] reported a higher malignancy rate in K-TIRADS 4 nodules with suspicious findings (subgroup 2 nodules) compared to those without suspicious findings (subgroup 1 nodules). This finding contrasts with the results of the present study and previous research, which indicated similar malignancy rates between the two subgroups of K-TIRADS 4 nodules [11,25,26]. The discrepancy may be partly due to an overestimation of the malignant risk associated with PEF in retrospective US assessments [27]. Notably, the malignancy risk for K-TIRADS 4 nodules with PEF in their study was significantly higher (47.9%) than the risks estimated in other studies, which ranged from 9.2% to 25.9% [25–27].

The present study has several limitations. First, this was a retrospective study and included only nodules that had undergone a biopsy. Therefore, selection bias was unavoidable. Second, the final diagnosis of many benign nodules was based on one FNA or CNB benign result, which may have resulted in false-negative results.

In conclusion, K-TIRADS 4 nodules were subcategorized into K-TIRADS 4B (higher risk) and K-TIRADS 4A (lower risk), based on the presence of five high-risk US features that increase the risk of malignancy in K-TIRADS 4 nodules. The modified biopsy criterion 3 significantly improved the diagnostic yield for malignancy compared to biopsy criterion 2, which has a size cut-off of 1.5 cm. It also demonstrated a similar diagnostic yield and a reduced false referral

rate compared to biopsy criterion 1, with a size cut-off of 1 cm. Modified biopsy criterion 3 could be used complementarily with biopsy criterion 2 in patients who require a higher detection rate for malignancy.

ORCID: Boeun Lee: <https://orcid.org/0000-0001-5471-8825>; Dong Gyu Na: <https://orcid.org/0000-0001-6422-1652>; Ji-hoon Kim: <https://orcid.org/0000-0002-6349-6950>

### Author Contributions

Conceptualization: Lee B, Na DG, Kim JH. Data acquisition: Lee B, Na DG. Data analysis or interpretation: Lee B, Na DG. Drafting of the manuscript: Lee B, Na DG. Critical revision of the manuscript: Lee B, Na DG, Kim JH. Approval of the final version of the manuscript: all authors.

### Conflict of Interest

Ji-hoon Kim serves as Editor for the *Ultrasonography*, but has no role in the decision to publish this article. All remaining authors have declared no conflicts of interest.

### Acknowledgments

This research was supported by the Medical Research Promotion Program through the GangNeung Asan Hospital funded by the Asan Foundation (2018-C03).

### Supplementary Material

Supplementary Table 1. Malignancy risk of intermediate suspicion (K-TIRADS 4) nodules according to the number of high risk US features (<https://doi.org/10.14366/usg.23203>).

## References

1. Ha EJ, Lim HK, Yoon JH, Baek JH, Do KH, Choi M, et al. Primary imaging test and appropriate biopsy methods for thyroid nodules: guidelines by Korean Society of Radiology and National Evidence-Based Healthcare Collaborating Agency. *Korean J Radiol* 2018;19:623-631.
2. 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.
3. Ha EJ, Na DG, Baek JH. Korean Thyroid Imaging Reporting and Data System: current status, challenges, and future perspectives. *Korean J Radiol* 2021;22:1569-1578.
4. Chung SR, Ahn HS, Choi YJ, Lee JY, Yoo RE, Lee YJ, et al. Diagnostic performance of the modified Korean Thyroid Imaging Reporting and Data System for thyroid malignancy: a multicenter validation study.



- Korean J Radiol 2021;22:1579-1586.
5. 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.
  6. Na DG, Paik W, Cha J, Gwon HY, Kim SY, Yoo RE. Diagnostic performance of the modified Korean Thyroid Imaging Reporting and Data System for thyroid malignancy according to nodule size: a comparison with five society guidelines. *Ultrasonography* 2021;40:474-485.
  7. Kim H, Shin JH, Kim KE, Kim MK, Oh J, Hahn SY. Subcategorization of intermediate suspicion thyroid nodules based on suspicious ultrasonographic findings. *Ultrasonography* 2023;42:307-313.
  8. 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.
  9. Ha EJ, Shin JH, Na DG, Jung SL, Lee YH, Paik W, et al. Comparison of the diagnostic performance of the modified Korean Thyroid Imaging Reporting and Data System for thyroid malignancy with three international guidelines. *Ultrasonography* 2021;40:594-601.
  10. Park HY, Suh CH, Kim SO. Use of "diagnostic yield" in imaging research reports: results from articles published in two general radiology journals. *Korean J Radiol* 2022;23:1290-1300.
  11. Shin HS, Na DG, Paik W, Yoon SJ, Gwon HY, Noh BJ, et al. Malignancy risk stratification of thyroid nodules with macrocalcification and rim calcification based on ultrasound patterns. *Korean J Radiol* 2021;22:663-671.
  12. Lee JY, Na DG, Yoon SJ, Gwon HY, Paik W, Kim T, et al. Ultrasound malignancy risk stratification of thyroid nodules based on the degree of hypoechoogenicity and echotexture. *Eur Radiol* 2020;30:1653-1663.
  13. Middleton WD, Teefey SA, Reading CC, Langer JE, Beland MD, Szabunio MM, et al. Multiinstitutional analysis of thyroid nodule risk stratification using the American College of Radiology Thyroid Imaging Reporting and Data System. *AJR Am J Roentgenol* 2017;208:1331-1341.
  14. Kwak JY, Han KH, Yoon JH, Moon HJ, Son EJ, Park SH, et al. Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk. *Radiology* 2011;260:892-899.
  15. Ha SM, Baek JH, Na DG, Suh CH, Chung SR, Choi YJ, et al. Diagnostic performance of practice guidelines for thyroid nodules: thyroid nodule size versus biopsy rates. *Radiology* 2019;291:92-99.
  16. Yoon SJ, Na DG, Gwon HY, Paik W, Kim WJ, Song JS, et al. Similarities and differences between Thyroid Imaging Reporting and Data Systems. *AJR Am J Roentgenol* 2019;213:W76-W84.
  17. Yim Y, Na DG, Ha EJ, Baek JH, Sung JY, Kim JH, et al. Concordance of three international guidelines for thyroid nodules classified by ultrasonography and diagnostic performance of biopsy criteria. *Korean J Radiol* 2020;21:108-116.
  18. Ha EJ, Na DG, Baek JH, Sung JY, Kim JH, Kang SY. US fine-needle aspiration biopsy for thyroid malignancy: diagnostic performance of seven society guidelines applied to 2000 thyroid nodules. *Radiology* 2018;287:893-900.
  19. Middleton WD, Teefey SA, Reading CC, Langer JE, Beland MD, Szabunio MM, et al. Comparison of performance characteristics of American College of Radiology TI-RADS, Korean Society of Thyroid Radiology TIRADS, and American Thyroid Association guidelines. *AJR Am J Roentgenol* 2018;210:1148-1154.
  20. Grani G, Lamartina L, Ascoli V, Bosco D, Biffoni M, Giacomelli L, et al. Reducing the number of unnecessary thyroid biopsies while improving diagnostic accuracy: toward the "right" TIRADS. *J Clin Endocrinol Metab* 2019;104:95-102.
  21. Richman DM, Benson CB, Doubilet PM, Wassner AJ, Asch E, Cherella CE, et al. Assessment of American College of Radiology Thyroid Imaging Reporting and Data System (TI-RADS) for pediatric thyroid nodules. *Radiology* 2020;294:415-420.
  22. Lee SB, Cho YJ, Lee S, Choi YH, Cheon JE, Kim WS. Korean Society of Thyroid Radiology guidelines for the management of pediatric thyroid nodules: suitability and risk factors. *Thyroid* 2021;31:1472-1480.
  23. Kim PH, Yoon HM, Baek JH, Chung SR, Choi YJ, Lee JH, et al. Diagnostic performance of five adult-based US risk stratification systems in pediatric thyroid nodules. *Radiology* 2022;305:190-198.
  24. Kim PH, Yoon HM, Baek JH, Chung SR, Choi YJ, Lee JH, et al. Diagnostic performance of the 2021 Korean thyroid imaging reporting and data system in pediatric thyroid nodules. *Eur Radiol* 2023;33:172-180.
  25. Na DG, Baek JH, Sung JY, Kim JH, Kim JK, Choi YJ, et al. Thyroid Imaging Reporting and Data System risk stratification of thyroid nodules: categorization based on solidity and echogenicity. *Thyroid* 2016;26:562-572.
  26. Ha EJ, Moon WJ, Na DG, Lee YH, Choi N, Kim SJ, et al. A multicenter prospective validation study for the Korean Thyroid Imaging Reporting and Data System in patients with thyroid nodules. *Korean J Radiol* 2016;17:811-821.
  27. Yang GE, Na DG. Impact of the ultrasonography assessment method on the malignancy risk and diagnostic performance of five risk stratification systems in thyroid nodules. *Endocrine* 2022;75:137-148.