



Diagnostic Performance of Thyroid Core Needle Biopsy Using the Revised Reporting System: Comparison with Fine Needle Aspiration Cytology

Kwangsoon Kim¹, Ja Seong Bae¹, Jeong Soo Kim¹, So Lyung Jung², Chan Kwon Jung^{3,4}

Departments of ¹Surgery, ²Radiology, ³Hospital Pathology, ⁴Cancer Research Institute, College of Medicine, The Catholic University of Korea, Seoul, Korea

Background: We aim to validate the diagnostic performance of thyroid core needle biopsy (CNB) for diagnosing malignancy in clinical settings to align with the changes made in recently updated thyroid CNB guidelines.

Methods: We retrospectively analyzed 1,381 thyroid CNB and 2,223 fine needle aspiration (FNA) samples. The FNA and CNB slides were interpreted according to the Bethesda System for Reporting Thyroid Cytopathology and updated practice guidelines for thyroid CNB, respectively.

Results: Compared to FNA, CNB showed lower rates of inconclusive results: categories I (2.8% vs. 11.2%) and III (1.2% vs. 6.2%), and higher rates of categories II (60.9% vs. 50.4%) and IV (17.5% vs. 2.0%). The upper and lower bounds of the risk of malignancy (ROM) for category IV of CNB were 43.2% and 26.6%, respectively. The CNB subcategory IVb with nuclear atypia had a higher ROM than the subcategory without nuclear atypia (40%–62% vs. 23%–36%). In histologically confirmed cases, there was no significant difference in the diagnostic performance between CNB and FNA for malignancy. However, neoplastic diseases were more frequently detected by CNB than by FNA (88.8% vs. 77.6%, $P=0.046$). In category IV, there was no difference in unnecessary surgery rate between CNB and FNA (4.7% vs. 6.9%, $P=0.6361$).

Conclusion: Thyroid CNB decreased the rate of inconclusive results and showed a higher category IV diagnostic rate than FNA. The revised guidelines for thyroid CNB proved to be an excellent reporting system for assessing thyroid nodules.

Keywords: Biopsy, fine-needle; Biopsy, large-core needle; Retrospective studies; Unnecessary procedures; Thyroid neoplasms; Thyroid nodule; Thyroid cancer, papillary

INTRODUCTION

Ultrasonography-guided fine needle aspiration (FNA) is the safest and most cost-effective diagnostic method in the preoperative evaluation of thyroid nodules [1,2]. About 60% of all FNA samples are classified as benign and 10% as malignant [3]. The

remaining 30% of thyroid FNA include nondiagnostic, atypia of unknown significance/follicular lesions of unknown significance (AUS/FLUS), follicular neoplasm/suspicious of follicular neoplasm (FN/SFN), and suspicious for malignancy [3]. Nondiagnostic results occur in about 10% of all FNA cases [3]. Patients with inconclusive FNA results may undergo repeat FNA,

Received: 11 October 2021, Revised: 29 November 2021,
Accepted: 6 January 2022

Corresponding author: Chan Kwon Jung
Department of Hospital Pathology, College of Medicine, The Catholic University of Korea, 222 Banpo-daero, Seocho-gu, Seoul 06591, Korea
Tel: +82-2-2258-1622, Fax: +82-2-2258-1627, E-mail: ckjung@catholic.ac.kr

Copyright © 2022 Korean Endocrine Society

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://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.

molecular testing, or surgical interventions [1]. However, approximately 10% to 30% of thyroid nodules initially diagnosed as AUS/FLUS would be repeatedly classified as AUS/FLUS even by repeat FNA [4]. About 60% to 70% of resected nodules with a cytologic diagnosis of AUS/FLUS or FN/SFN turn out to be benign after surgery [3].

Core needle biopsy (CNB) has been introduced as an alternative to FNA that addresses some of the aforementioned problems. The CNB collects a larger tissue sample with potential information on architectural histological structures and is useful for establishing the diagnosis of thyroid nodules [5]. Although CNB uses a thicker needle of 18 to 21 gauge than FNA, the procedure is safe and well-tolerated when performed under the guidance of real-time ultrasound by experienced operators [5]. Previous studies demonstrated that CNB decreased nondiagnostic result rates and increased malignancy diagnostic accuracy compared to FNA cytology [6-8]. However, most studies were conducted before introducing a noninvasive follicular thyroid neoplasm with papillary-like nuclear feature (NIFTP). Given that the second edition of the thyroid CNB pathology reporting system was released in January 2020 [5], the diagnostic performance of CNB must be validated in clinical settings to align with the changes made in these updated guidelines and reflect the reclassification of a subset of papillary thyroid carcinoma (PTC) as NIFTP.

We retrospectively examined a consecutive series of FNA and CNB samples to assess the risk of malignancy (ROM) of each CNB and FNA category and their diagnostic performance in the era of NIFTP.

METHODS

Patients

We retrospectively analyzed 3,604 patients who underwent FNA or CNB to evaluate thyroid nodules from January 2018 to December 2019 at Seoul St. Mary's Hospital (Seoul, Korea). The medical charts and pathology reports of 3,604 patients were reviewed and analyzed. According to the diagnostic technique, patients were divided into two groups: 1,381 patients who underwent CNB from 2018 to 2019 and 2,223 patients who underwent FNA in 2018. This study was carried out following the Declaration of Helsinki (as revised in 2013) and was approved by the Institutional Review Board of Seoul St. Mary's Hospital, The Catholic University of Korea (IRB No: KC16RISI0038). The need for informed consent was waived due to the retrospective nature of this study.

FNA and CNB procedures

All FNA and CNB procedures were performed under ultrasound guidance using Philips HDI 5000 (Philips Medical Systems, Bothell, WA, USA) or Aplio 500 Platinum (Toshiba Medical Systems, Tokyo, Japan) device by radiologists with 1 to 20 years' experience in thyroid ultrasonography and interventional ultrasound. FNA was performed with the needle oriented parallel to the ultrasound probe using a 23 gauge needle connected to a 2 to 10 mL syringe. The thyroid samples were aspirated with to-and-fro needle movements for 5 to 10 seconds, preserved in ThinPrep PreservCyt solution (Hologic, Marlborough, MA, USA), and processed on a ThinPrep processor. CNB was performed using an 18 gauge needle with a biopsy notch (1.1 or 1.6 cm in length). After administration of local anesthetic (1% lidocaine) along the approach route, the CNB needle was inserted through the isthmus of the thyroid and 1 to 2 cores per nodule were obtained. The biopsy specimen included intranodular tissue, the interface around the nodule, and surrounding normal thyroid tissue. After visual assessment for specimen adequacy, the harvested tissue was immediately fixed in 10% neutral buffered formalin and sent for pathologic examination. If not adequate, an additional tissue core was obtained. Real-time color Doppler imaging was used to minimize bleeding during the procedure. After the procedure of FNA or CNB, patients were asked to compress the biopsy site and observed for 10 to 30 minutes to monitor for complications.

In our institution, we use CNB as a second-line tool for the diagnosis of thyroid nodules. But, CNB was performed in preference to FNA in the following cases: (1) thyroid nodules with macrocalcifications or hypervascularity; (2) thyroid nodules with scanty aspirates or prior nondiagnostic/inconclusive FNA; or (3) thyroid nodules with suspicious features of malignancy on ultrasound but prior nonmalignant FNA. An 18 gauge single action needle was used for the CNB procedure.

Categorical diagnosis of FNA and CNB

The FNA cytology and CNB pathology slides were examined according to the Bethesda System for Reporting Thyroid Cytopathology and recently updated practice guidelines for thyroid CNB, respectively [2,5]. The FNA and CNB results were classified into six categories. All cases were reviewed by a pathologist with extensive experience in thyroid disease. Category III was further classified into five subcategories: IIIa, with cytologic atypia; IIIb, with architectural atypia; IIIc, with both cytologic and architectural atypia; IIId, Hürthle cell type; and IIIe, not otherwise specified. Category IV was further classified into four

subcategories: IVa, conventional; IVb, with cytologic/nuclear atypia; IVc, Hürthle cell type; and IVd, not otherwise specified.

Statistical analysis

Continuous variables were presented as means with standard deviation and categorical variables as numbers with percentages. ROM was estimated as the lower bound and the upper bound because a subset of patients with FNA or CNB diagnosis had a surgical follow-up. The lower bound ROM was calculated using the total number of FNA or CNB cases as the denominator. The upper bound ROM was calculated by having the total number of resected cases as the denominator. Thus, the real ROM settles somewhere between the upper and lower values.

The surgical diseases refer to neoplastic lesions, including follicular adenoma, NIFTP, and thyroid cancers, that require surgery for diagnosis and treatment. Unnecessary surgery was defined as any surgical operation that is not needed as follows: (1) thyroid nodules with a preoperative diagnosis of malignancy (category V or VI) but benign pathology at the final diagnosis or (2) nodules with a preoperative diagnosis of FN/SFN (category IV) but non-neoplastic benign pathology such as nodular hyperplasia and chronic lymphocytic thyroiditis at the final diagnosis [9]. The surgery was justified when a nodule with category IV was diagnosed as a neoplasm regardless of whether the final pathology was benign or malignant.

The diagnostic performance of FNA and CNB was evaluated for the detection of surgical disease and thyroid cancer. In addition, subgroup analysis was performed according to tumor size. All statistical analyzes were performed using the SPSS software for Windows version 23.0 (IBM Corp., Armonk, NY, USA) and the MedCalc Statistical Software version 20.011 (MedCalc Software Ltd., Ostend, Belgium). A *P* value less than 0.05 was considered statistically significant.

RESULTS

There were no significant differences in age or sex between the two groups (*P*=0.534 and *P*=0.477, respectively), as shown in Table 1. Compared to FNA, CNB showed lower rates for diagnostic categories I (2.8% vs. 11.2%, *P*<0.001), III (1.2% vs. 6.2%, *P*<0.001), V (0.5% vs. 3.1%, *P*<0.001), and VI (17.1% vs. 27.0%, *P*<0.001), and higher rates for categories II (60.9% vs. 50.4%, *P*<0.001) and IV (17.5% vs. 2.0%, *P*<0.001).

Regarding the histologically confirmed diagnosis, the malignancy was confirmed in surgical specimens or by CNB diagnosis with specific malignant terms. Other lesions were histologi-

Table 1. Clinicopathologic Characteristics of Patients

Characteristic	CNB (n=1,381)	FNA (n=2,223)	<i>P</i> value
Age, yr	47.1±13.7	48.9±16.2	0.534
Sex			0.477
Female	1,045 (75.7)	1,658 (74.6)	
Male	336 (24.3)	565 (25.4)	
Diagnostic category			
I	39 (2.8)	250 (11.2)	<0.001
II	841 (60.9)	1,121 (50.4)	<0.001
III	17 (1.2)	138 (6.2)	<0.001
IV	241 (17.5)	45 (2.0)	<0.001
V	7 (0.5)	68 (3.1)	<0.001
VI	236 (17.1)	601 (27.0)	<0.001
Histologically confirmed cases	429 (31.1)	666 (30.0)	0.557
Histologic diagnosis			
FA/HA	70 (16.3)	27 (4.1)	<0.001
Other benign lesions	32 (7.5)	76 (11.4)	0.042
NIFTP	15 (3.5)	7 (1.1)	0.005
PTC	244 (56.9)	539 (80.9)	<0.001
FTC/HCC	32 (7.5)	9 (1.4)	<0.001
PDTC/ATC	9 (2.1)	2 (0.3)	0.004
MTC	3 (0.7)	3 (0.5)	0.587
Lymphoma	18 (4.2)	0	<0.001
Other malignancy	6 (1.4)	3 (0.5)	0.091
Nodule size of histologically confirmed cases, cm ^a			
Mean	2.0±1.5	1.1±0.8	<0.001
≤1.0	111 (25.9)	428 (64.3)	<0.001
1.1–2.0	143 (33.3)	177 (26.6)	0.045
2.1–3.0	101 (23.5)	38 (5.7)	<0.001
>3.0	74 (17.3)	23 (3.4)	<0.001

Values are expressed as mean±standard deviation or number (%).

CNB, core needle biopsy; FNA, fine needle aspiration; FA, follicular adenoma; HA, Hürthle cell adenoma; NIFTP, noninvasive follicular thyroid neoplasm with papillary-like nuclear feature; PTC, papillary thyroid carcinoma; FTC, follicular thyroid carcinoma; HCC, Hürthle cell carcinoma; PDTC, poorly differentiated thyroid carcinoma; ATC, anaplastic thyroid carcinoma; MTC, medullary thyroid carcinoma.

^aThe nodule size was determined as the largest diameter measured in the surgical specimen, but if surgery was not performed, it was measured by ultrasound.

cally confirmed in surgical specimens. Histological diagnoses were available in 429 (31.1%) of 1,381 CNB samples and 666 (30.0%) of 2,223 FNA samples (Table 1). The prevalence of malignancies was higher in the FNA group than in the CNB

group (83.5% vs. 72.7%, $P<0.001$). CNB group showed higher rates of follicular adenoma/Hürthle cell adenoma, NIFTP, follicular carcinoma/Hürthle cell carcinoma, lymphoma than FNA (16.3% vs. 4.1%, $P<0.001$; 3.5% vs. 1.1%, $P=0.042$; 7.5% vs. 1.4%, $P<0.001$; 4.2% vs. 0%, $P<0.001$, respectively), whereas the rate of PTC was higher in the FNA group than in CNB (80.9% vs. 56.9%, $P<0.001$). The mean nodule size of the CNB group was larger than that of the FNA group (2.0 cm vs. 1.1 cm, $P<0.001$) in the histologically confirmed cases (Table 1). The proportion of subcentimeter nodules was higher in the FNA group than in the CNB group ($P<0.001$).

Malignancy risk for each category in CNB

The distribution of 1,381 CNB results was 39 (2.8%) in category I, 841 (60.9%) in category II, 17 (1.2%) in category III, 241 (17.5%) in category IV, seven (0.5%) in category V, and 236 (17.1%) in category VI (Table 2). Category III was further classified into category IIIa ($n=3$), category IIIb ($n=6$), category IIIc ($n=3$), and category IIIe ($n=5$). Likewise, 241 (17.5%) samples with category IV were assigned to the subclassification: category IVa ($n=144$), category IVb ($n=57$), category IVc

($n=39$), and category IVd ($n=1$).

The upper and lower ROM bounds in category III were 25% (1/4) and 5.9% (1/17), respectively. NIFTP was diagnosed predominantly in category IV. When NIFTP was not included in malignancy, the upper and lower ROM bounds in category IV were 43.2% (64/148) and 26.6% (64/241), respectively. When NIFTP was included in malignancy, the upper bound of ROM was calculated as 78 of 148 (52.7%), and the lower bound of ROM was speculated as 78 of 241 (32.4%) in patients with category IV. Thus, the IVb subcategory with nuclear atypia had a higher ROM than the other subcategories without nuclear atypia (upper ROM bounds, 78% [29/37] vs. 44% [49/111], $P=0.013$; lower ROM bounds, 51% [29/57] vs. 27% [49/184], $P=0.005$, respectively).

Malignancy risk of each category in FNA

The distribution of 2,223 FNA results was 250 (11.2%) in category I, 1,121 (50.4%) in category II, 138 (6.2%) in category III, 45 (2.0%) in category IV, 68 (3.1%) in category V, and 601 (27.1%) in category VI (Table 3). Categories III and IV were further classified into subcategories, as shown in Table 3.

Table 2. Thyroid Core Needle Biopsy Results and Histologically Confirmed Diagnosis

CNB category	No.	Histologic confirmation	Benign	NIFTP	Malignant	ROM without NIFTP, %	ROM with NIFTP, %
I	39	9 (23.1)	4	0	5	13–56	13–56
II	841	28 (3.3)	25	1	2	0.2–7.1	0.4–10.7
III	17	4 (24)	3	0	1	6–25	6–25
IIIa	3	0	0	0	0	NA	NA
IIIb	6	3 (50)	2	0	1	17–33	17–33
IIIc	0	NA	NA	NA	NA	NA	NA
IIId	3	1 (33)	1	0	0	0	0
IIIe	5	0	0	0	0	NA	NA
IV	241	148 (61.4)	70	14	64	26.6–43.2	32.4–52.7
IVa	144	91 (63.2)	51	7	33	22.9–36.3	27.8–44.0
IVb	57	37 (65)	8	6	23	40–62	51–78
IVc	39	19 (49)	11	1	7	18–37	21–42
IVd	1	1 (100)	0	0	1	100	100
V	7	4 (57)	0	0	4	57–100	57–100
VI	236	236 (100)	0	0	236	100.0	100.0
Total	1,381	429 (31.1)	102	15	312	22.6–72.7	23.7–76.2

Values are expressed as number (%). The upper bound of the ROM was calculated by having all resected cases as denominator. The malignant cases (category VI) diagnosed by CNB were considered as histologically confirmed cases. The lower bound of ROM was calculated by having the total number of CNB cases as denominator.

CNB, core needle biopsy; NIFTP, noninvasive follicular thyroid neoplasm with papillary-like nuclear feature; ROM, risk of malignancy; NA, not available.

NIFTP was observed in categories III, IV and IV of the FNA group after surgery (Table 3). In category III ($n=138$), the upper and lower bounds of ROM were calculated as 70.8% (34/48) and 24.6% (34/138), respectively. When NIFTPs were considered malignant, the upper and lower bounds of ROM were speculated as 75.0% (36/48) and 26.1% (36/138), respectively. Subcategory IIIa with nuclear atypia had a higher ROM than the other subcategories without nuclear atypia (upper ROM bounds, 94% [29/31] vs. 41% [7/17], $P=0.045$; lower ROM bounds, 36% [29/80] vs. 12% [7/58], $P=0.006$, respectively).

In the 45 FNA cases with category IV, the upper and lower bounds of ROM without NIFTP were calculated as 41.1% (12/29) and 26.7% (12/45), respectively. The upper and lower bounds of ROM when including NIFTP in the malignant group were speculated to be 55.2% (16/29) and 35.6% (16/45), respectively. There were no significant differences in ROMs between subcategory IVb with nuclear atypia and other subcategories without nuclear atypia (upper ROM bounds, 63% [5/8] vs. 52% [11/21], $P=0.743$; lower ROM bounds, 38% [5/13] vs. 34% [11/32], $P=0.835$, respectively).

Diagnostic performance of CNB and FNA

Diagnostic performance of CNB and FNA may vary depending on reference standard and cut-off value. The diagnostic performance was evaluated only for patients with available histological diagnoses after CNB or FNA. Histological diagnoses were available in 429 CNB and 666 FNA samples (Table 1). Regarding positive results by CNB and FNA, categories V and VI were considered positive for determination of malignancy, and categories IV, V, and VI were considered positive for determination of malignancy/NIFTP and neoplastic lesions including benign neoplasm, NIFTP, and malignancy.

Table 4 summarizes the observed performance of CNB and FNA according to three different approaches in histologically confirmed cases. The diagnostic sensitivity and specificity for the diagnosis of malignancy were as follows: 76.9% (95% confidence interval [CI], 71.8% to 81.5%) and 100% by CNB and 88.0% (95% CI, 85.0% to 90.5%) and 99.1% (95% CI, 95.0% to 100.0%) by FNA, respectively. Although patients with thyroid cancer more frequently underwent FNA than CNB ($P<0.001$), there was no significant difference in the diagnostic performance between CNB and FNA for malignancy.

Table 3. Thyroid Fine Needle Aspiration Cytology Results and Histologically Confirmed Diagnosis

FNA category	No.	Histologically confirmed cases	Benign	NIFTP	Malignant	ROM without NIFTP, %	ROM with NIFTP, %
I	250	27 (10.8)	11	0	16	6.4–59.3	6.4–59.3
II	1,121	72 (6.4)	67	0	5	0.4–6.9	0.4–6.9
III	138	48 (34.8)	12	2	34	24.6–70.8	26.1–75.0
IIIa	80	31 (35)	2	1	28	35–90	36–94
IIIb	22	10 (46)	8	0	2	9–20	9–20
IIIc	16	5 (31)	1	0	4	25–80	25–80
IIId	20	2 (10)	1	1	0	0	5–50
IIIe	0	NA	NA	NA	NA	NA	NA
IV	45	29 (64)	13	4	12	27–41	36–55
IVa	16	10 (63)	5	0	5	31–50	31–50
IVb	13	8 (62)	3	2	3	23–36	39–63
IVc	16	11 (69)	5	2	4	25–36	38–55
IVd	0	NA	NA	NA	NA	NA	NA
V	68	48 (71)	0	1	47	69–98	71–100
VI	601	442 (73.5)	0	0	442	73.5–100	73.5–100
Total	2,223	666 (30.0)	103	7	556	25.0–83.5	25.3–84.5

Values are expressed as number (%). The upper bound of the ROM was calculated by having all resected cases as denominator. The lower bound of ROM was calculated by having the total number of FNA cases as denominator.

FNA, fine needle aspiration; NIFTP, noninvasive follicular thyroid neoplasm with papillary-like nuclear feature; ROM, risk of malignancy; NA, not available.

In the analysis of diagnostic performance for the diagnosis of malignancy and NIFIP in histologically confirmed cases, FNA had higher specificity (87.4% vs. 31.4%, $P<0.001$) and positive predictive value (PPV) (97.5% vs. 82.0%, $P=0.015$) than CNB. There was no significant difference in sensitivity, negative pre-

dictive value (NPV), and accuracy between CNB and FNA for the diagnosis of malignancy/NIFTP (Table 4).

In histologically confirmed cases, surgical neoplastic diseases were more frequently detected by CNB than by FNA (88.8% vs. 77.6%, $P=0.046$). However, when categories IV, V, and VI of

Table 4. Diagnostic Performance of Thyroid Core Needle Biopsy and Fine Needle Aspiration Cytology for the Diagnosis of Malignancy, Malignancy/NIFTP, and Surgical Diseases

Outcomes	CNB (n=429)		FNA (n=666)		P value
	Malignant	Benign+NIFTP	Malignant	Benign+NIFTP	
Analysis 1^a					
Category V–VI	240	0	489	1	
Category I–IV	72	117	67	109	
Diagnostic performance	Incidence	95% CI	Incidence	95% CI	
Disease prevalence	55.9%	49.1–63.5	73.4%	67.1–80.0	<0.001
Sensitivity	76.9%	71.8–81.5	88.0%	85.0–90.5	0.089
Specificity	100%	96.9–100.0	99.1%	95.0–100.0	0.945
PPV	100%	100.0	99.8%	98.6–100.0	0.979
NPV	61.9%	57.0–66.6	61.9%	56.5–67.1	0.997
Accuracy	83.2%	79.3–86.6	89.8%	87.2–92.0	0.256
Analysis 2^b	Malignant+NIFTP	Benign	Malignant+NIFTP	Benign	P value
Category IV–VI	318	70	506	13	
Category I–III	9	32	57	90	
Diagnostic performance	Incidence	95% CI	Incidence	95% CI	
Disease prevalence	76.2%	68.2–85.0	84.5%	77.7–91.8	0.137
Sensitivity	97.2%	94.8–98.7	89.9%	87.1–92.2	0.271
Specificity	31.4%	22.5–41.3	87.4%	79.4–93.1	<0.001
PPV	82.0%	79.9–80.2	97.5%	95.9–98.5	0.015
NPV	78.0%	63.7–87.8	61.2	55.0–67.1	0.237
Accuracy	81.6%	77.6–85.1	89.5%	86.9–91.7	0.170
Analysis 3^c	Neoplastic (surgical disease)	Non-neoplastic	Neoplastic (surgical disease)	Non-neoplastic	P value
Category IV–VI	381	7	517	2	
Category I–III	16	25	78	69	
Diagnostic performance	Incidence	95% CI	Incidence	95% CI	
Disease prevalence	88.8%	80.1–98.2	77.6%	71.1–84.6	0.046
Sensitivity	96.0%	93.5–97.7	86.9%	83.9–89.5	0.141
Specificity	78.1%	60.0–90.7	97.2%	90.2–99.7	0.349
PPV	98.2%	96.6–99.1	99.6%	98.5–99.9	0.832
NPV	61.0%	48.3–72.3	46.9%	41.7–52.2	0.261
Accuracy	94.6%	92.1–96.6	88.0%	85.3–90.4	0.259

NIFTP, noninvasive follicular thyroid neoplasm with papillary-like nuclear feature; CNB, core needle biopsy; FNA, fine needle biopsy; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value.

^aCategories V and VI were considered diagnostic-positive results for malignancy. Diagnostic performance was evaluated for the differentiating malignancy from nonmalignant lesions including NIFTP; ^bCategories IV, V and VI were considered diagnostic-positive results for malignancy and NIFTP. Diagnostic performance was evaluated for the differentiating malignancy and NIFTP from benign lesions; ^cCategories IV, V and VI were considered diagnostic-positive results for surgical diseases. Diagnostic performance was evaluated for the differentiating surgical diseases (neoplasms) from non-neoplastic lesions.

CNB and FNA were considered positive for the determination of neoplastic lesions including follicular adenoma, NIFTP, and malignancy, there was no significant difference in the diagnostic performance between CNB and FNA for the diagnosis of surgical diseases (Table 4).

The expected PPV and NPV in CNB and FNA based on the three different approaches and performance adjusted for the prevalence of malignancy, malignancy/NIFTP, and surgical diseases are shown in Fig. 1. For example, at the malignancy prevalence of 20%, the expected PPV and NPV are calculated as

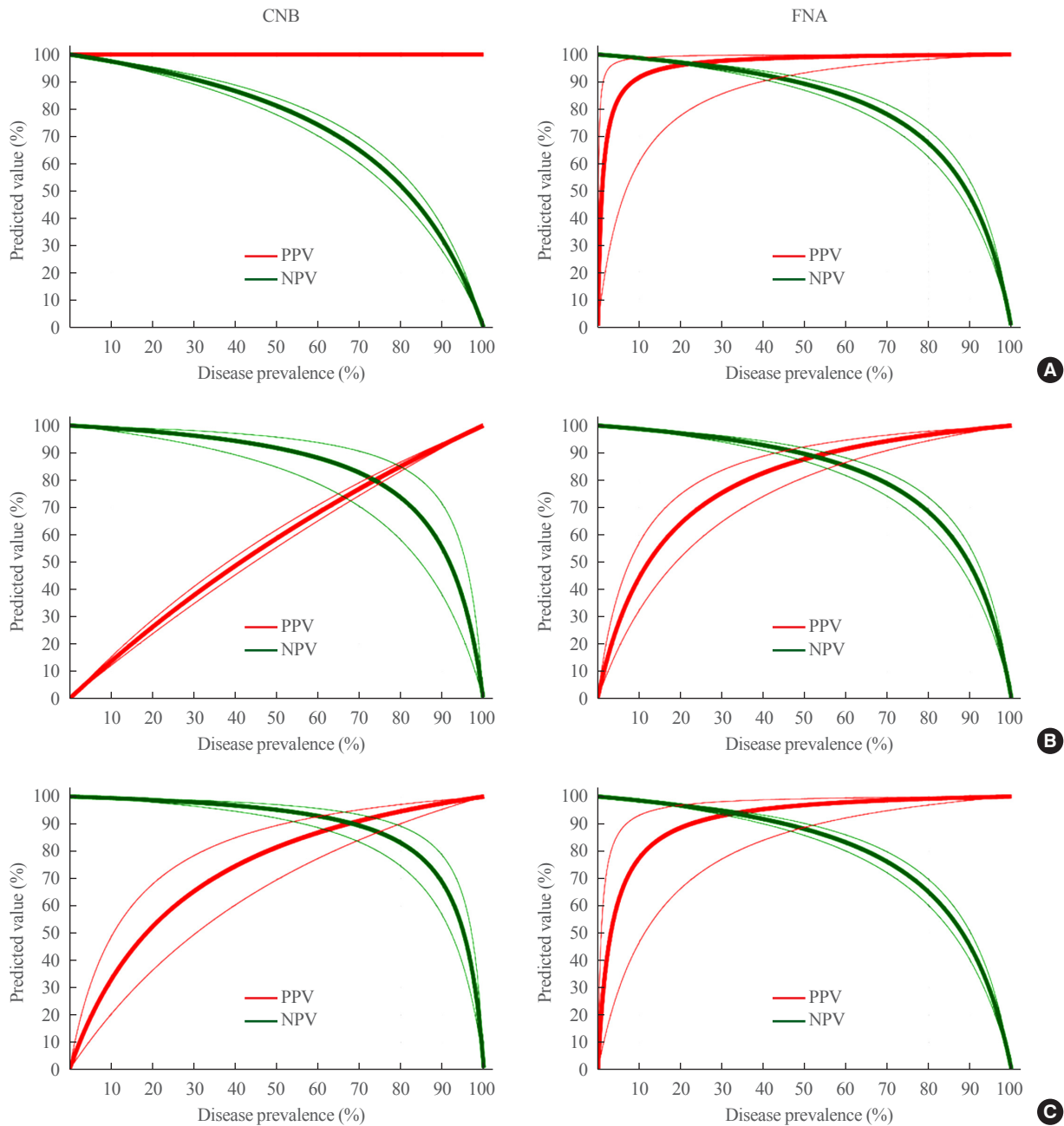


Fig. 1. Expected diagnostic performance of core needle biopsy (CNB) and fine needle aspiration (FNA). (A) The expected positive predictive value (PPV) and negative predictive value (NPV) for the diagnosis of thyroid malignancy with a variable prevalence of the disease. (B) The diagnosis of thyroid malignancy and noninvasive follicular thyroid neoplasm with papillary-like nuclear feature (NIFTP) with a variable prevalence of the disease. (C) The diagnosis of surgical diseases including malignancy, NIFTP, and benign neoplasm with a variable prevalence of the disease. The thin red and green lines represent the corresponding 95% confidence intervals for PPV and NPV, respectively.

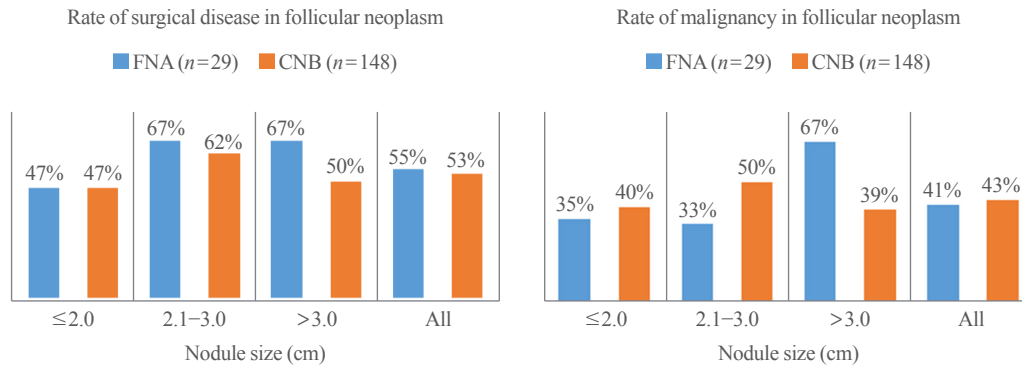


Fig. 2. Rates of (A) surgical disease and (B) malignancy according to the size of the nodule in patients who underwent diagnostic surgery for thyroid nodules with a preoperative diagnosis of follicular neoplasm by core needle biopsy (CNB) or fine needle aspiration (FNA) cytology.

follows: 100% and 94.5% (95% CI, 93.4% to 95.5%) by CNB and 96.0% (95% CI, 77.5% to 99.4%) and 97.0% (95% CI, 96.3% to 97.6%) by FNA, respectively.

The unnecessary surgery rates for CNB and FNA were 4.7% (7/148) and 6.9% (2/29) in category IV, respectively (Table 4). The difference was not statistically significant ($P=0.6361$). In categories V and VI, there was no unnecessary surgery in both groups.

Outcomes of CNB and FNA in follicular neoplasm according to tumor size

Histologically confirmed diagnoses of CNB and FNA results with FN/SFN were further analyzed according to tumor size (Fig. 2). There were no significant differences in the rates of surgical disease and malignancy between CNB and FNA according to tumor size.

DISCUSSION

Ultrasound-guided FNA is the first-line diagnostic tool for thyroid nodules [1]. However, the main drawback of FNA is inconclusive results, including nondiagnostic and AUS/FLUS categories, which can lead to repeat FNA or diagnostic surgery [10–12]. In this present study, we found that the rate of inconclusive results was lower in CNB than in FNA. Nondiagnostic and AUS/FLUS results were found in 11.2% and 6.2% of all FNA specimens, respectively. The ROMs of FNA fell within the range of 6.4% to 59.3% for nondiagnostic results and 24.6% to 70.8% for AUS/FLUS results. On the other hand, the nondiagnostic and AUS/FLUS results for CNB were 2.8% and 1.2%, respectively. The ROM of the nondiagnostic CNB samples was within the range of 13% to 56%. Therefore, the low rate of inconclusive results from CNB decreased the need for additional

biopsies or unnecessary diagnostic surgery.

The Bethesda System for Reporting Thyroid Cytopathology recommends that patients with nondiagnostic conditions undergo a repeat FNA [2]. However, repeated FNA often has unsatisfactory results. Choi et al. [13] reported that repeat FNA was not useful to reduce the number of nondiagnostic categories. Lee et al. [14] showed that one-third of repeat FNAs were nondiagnostic on repeat examination. Previous studies have demonstrated the feasibility of CNB as an alternative to unsatisfactory FNA results. Na et al. [15] reported that the nondiagnostic rate was 45.3% with repeat FNA and 12.5% with CNB. Furthermore, Park et al. [16] reported that the rate of inconclusive results was 48.6% with repeat FNA and 1.8% with CNB in thyroid nodules that were diagnosed with initially inconclusive results by FNA. Our data are in line with those of previous studies.

Sampling errors and inadequate specimens in FNA can occur under variable conditions, such as small nodule size, predominantly cystic nodule, heavy calcification, fibrosis, and hypervascularity [17]. CNB has been suggested as an alternative first-line diagnostic tool to FNA to diagnose thyroid nodules [18–22]. CNB preserves the cellular architecture, facilitates histopathological evaluation, and enables ancillary immunohistochemical tests if necessary [23,24]. Trimboli et al. [25] reported that CNB had a higher diagnostic accuracy for malignancy than FNA in solid thyroid nodules at high risk for malignancy. In our institution, we performed CNB as a second tool for the diagnosis of thyroid nodules with previous nondiagnostic or inconclusive FNA results. CNB was performed as a first-line tool in thyroid nodules with marked calcification or hypervascularity. Most of the patients who were transferred to our hospital for thyroid cancer treatment were diagnosed with FNA rather than CNB. CNB was often performed when it was difficult to diagnose thyroid cancer. Therefore, the prevalence of malignancy was

higher in patients who underwent FNA than in those who underwent CNB. The selection bias for patients who underwent CNB or FNA could have influenced the outcome of diagnostic performance.

Diagnostic category IV (FN/SFN) is more easily diagnosed in CNB specimens than FNA because CNB provides better architectural histologic features than FNA [5]. In the present study, FN/SFN was diagnosed more frequently in the CNB group than in the FNA group (17.5% vs. 2.0%), while the category III rate was lower in the CNB than in the FNA (1.2% vs. 6.2%). Although FN/SFN was preoperatively diagnosed with a higher frequency on CNB samples, the ROM for CNB was similar as a result for FNA (26.6%–43.2% vs. 26.7%–41.1%) even when including NIFTP in the malignant group (32.4%–52.7% vs. 35.6%–55.2%). When we further analyzed the surgical results of thyroid nodules with a preoperative diagnosis of FN/SFN according to tumor size, the rates of surgical disease and malignancy were similar regardless of tumor size or diagnostic methods used (Fig. 2). Therefore, FN/SFN can be underdiagnosed or misclassified as other categories on FNA samples with a follicular architecture. These findings are in line with other previous reports in the literature. For example, Lee et al. [26] reported in their validation study that CNB was a good method for the diagnosis of FN/SFN in thyroid nodule screening. Yoon et al. [27] presented that CNB showed a significantly lower false-positive and unnecessary surgery rate than FNA in the diagnosis of FN/SFN. Since CNB can provide sufficient tissue for molecular testing, it can be possible to improve diagnostic performance by applying molecular tests on CNB specimens with indeterminate results.

In diagnostic category III, the subcategories with nuclear atypia showed higher ROM than other subcategories in both the CNB and FNA groups. Reclassification of a subtype of the follicular variant PTC to NIFTP impacted ROM predominantly in category IV of FNA and CNB in our study. Most NIFTPs were diagnosed as a diagnostic subtype with nuclear atypia: 71% (5/7) in FNA with category IIIa/IVb/IVc and 47% (7/15) in CNB with category IVb/IVc. Therefore, NIFTP should be considered for the differential diagnosis when thyroid nodules with category III or IV show nuclear atypia in CNB and FNA specimens.

Previous studies have reported that the rates of CNB categories vary significantly among different institutions. Supplemental Table S1 shows the preoperative and postoperative results of the CNB performed in three Korean institutions [28,29]. The rate of inconclusive results (category I and III) was lower in our

institution than in the other two institutions. The rates of categories II and IV were higher in our institution than in the other two institutions. However, the ROMs of CNB category IV after surgery were similar between the three institutions. The strength of our study is that most of the CNB samples were initially diagnosed by a pathologist with extensive experience who has dedicated more than 20 years to thyroid disease. Therefore, the diagnosis of thyroid nodules could maintain consistency because all cases were reviewed for this study by the same pathologist. However, in other institutions, multiple pathologists diagnosed the CNB specimens. Therefore, we suggest that the diagnostic performance of CNB can vary between institutions and pathologists depending on their experience and diagnostic threshold.

This study has several limitations. First, this study was designed as a retrospective study in nature. There may have been selection bias because all patients were enrolled from a tertiary referral hospital and do not reflect the entire patient population. We did not perform a head-to-head comparison of the diagnostic performance of CNB and FNA. Therefore, the rates of malignant CNB and FNA results were relatively higher in our study than those reported in other similar studies [1,2]. There was a significant difference in rates and types of cancers between CNB and FNA because we preferred FNA over CNB as the first-line diagnostic tool for thyroid nodules with suspicious findings on ultrasonography. Second, this study did not evaluate radiologists' variability and ultrasonographic characteristics in the performance of FNA and CNB. Third, we did not compare the frequency and degree of complications after procedures such as pain, hematoma, transient parenchymal edema, etc. However, complications after CNB have been reported to be similar compared to FNA [5]. Fourth, we did not compare the diagnostic performance between the first and second edition of the CNB pathology reporting system because the main change in the second edition, the subclassification of categories III and IV, has already been conceptually introduced in the first edition, and NIFTP was newly introduced in the second edition. Finally, we demonstrated that CNB could reduce the need for unnecessary repeat FNA or diagnostic operations performed to diagnose thyroid nodules with initially inconclusive results. However, there is no consensus on how to apply CNB for the evaluation of inconclusive thyroid nodules. These limitations could be overcome by conducting more prospective or multicenter studies in the future.

In conclusion, the present study demonstrated that CNB showed a lower rate of inconclusive results and a higher category IV diagnostic rate than FNA in the diagnosis of thyroid nod-

ules. Diagnostic categories III and IV should be further subdivided into subcategories with or without nuclear atypia because a subcategory with nuclear atypia has a higher ROM. The diagnostic performance of CNB depends on the diagnostic threshold and the experience of the pathologists, specifically in relation to the diagnosis of category III.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENTS

This research was supported by a grant (NRF-2020R1F1-A1070028) from the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Science and ICT.

AUTHOR CONTRIBUTIONS

Conception or design: C.K.J. Acquisition, analysis, or interpretation of data: K.K., J.S.B., J.S.K., S.L.J., C.K.J. Drafting the work or revising: K.K., J.S.B., J.S.K., S.L.J., C.K.J. Final approval of the manuscript: K.K., J.S.B., J.S.K., S.L.J., C.K.J.

ORCID

Kwangsoon Kim <https://orcid.org/0000-0001-6403-6035>

Chan Kwon Jung <https://orcid.org/0000-0001-6843-3708>

REFERENCES

- 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.
- Cibas ES, Ali SZ. The 2017 Bethesda System for Reporting Thyroid Cytopathology. *Thyroid* 2017;27:1341-6.
- Vuong HG, Chung DG, Ngo LM, Bui TQ, Hassell L, Jung CK, et al. The use of the Bethesda System for Reporting Thyroid Cytopathology in pediatric thyroid nodules: a meta-analysis. *Thyroid* 2021;31:1203-11.
- Rossi ED, Pantanowitz L, Faquin WC. The role of molecular testing for the indeterminate thyroid FNA. *Genes (Basel)* 2019;10:736.
- Jung CK, Baek JH, Na DG, Oh YL, Yi KH, Kang HC. 2019 Practice guidelines for thyroid core needle biopsy: a report of the Clinical Practice Guidelines Development Committee of the Korean Thyroid Association. *J Pathol Transl Med* 2020;54:64-86.
- Yeon JS, Baek JH, Lim HK, Ha EJ, Kim JK, Song DE, et al. Thyroid nodules with initially nondiagnostic cytologic results: the role of core-needle biopsy. *Radiology* 2013;268:274-80.
- Chung SR, Suh CH, Baek JH, Choi YJ, Lee JH. The role of core needle biopsy in the diagnosis of initially detected thyroid nodules: a systematic review and meta-analysis. *Eur Radiol* 2018;28:4909-18.
- Na DG, Baek JH, Jung SL, Kim JH, Sung JY, Kim KS, et al. Core needle biopsy of the thyroid: 2016 consensus statement and recommendations from Korean Society of Thyroid Radiology. *Korean J Radiol* 2017;18:217-37.
- Chung SR, Baek JH, Choi YJ, Sung TY, Song DE, Kim TY, et al. The role of core needle biopsy for the evaluation of thyroid nodules with suspicious ultrasound features. *Korean J Radiol* 2019;20:158-65.
- Wolinski K, Stangierski A, Ruchala M. Comparison of diagnostic yield of core-needle and fine-needle aspiration biopsies of thyroid lesions: systematic review and meta-analysis. *Eur Radiol* 2017;27:431-6.
- Choi YJ, Baek JH, Suh CH, Shim WH, Jeong B, Kim JK, et al. Core-needle biopsy versus repeat fine-needle aspiration for thyroid nodules initially read as atypia/follicular lesion of undetermined significance. *Head Neck* 2017;39:361-9.
- Sung JY, Na DG, Kim KS, Yoo H, Lee H, Kim JH, et al. Diagnostic accuracy of fine-needle aspiration versus core-needle biopsy for the diagnosis of thyroid malignancy in a clinical cohort. *Eur Radiol* 2012;22:1564-72.
- Choi SH, Baek JH, Lee JH, Choi YJ, Hong MJ, Song DE, et al. Thyroid nodules with initially non-diagnostic, fine-needle aspiration results: comparison of core-needle biopsy and repeated fine-needle aspiration. *Eur Radiol* 2014;24:2819-26.
- Lee SH, Kim MH, Bae JS, Lim DJ, Jung SL, Jung CK. Clinical outcomes in patients with non-diagnostic thyroid fine needle aspiration cytology: usefulness of the thyroid core needle biopsy. *Ann Surg Oncol* 2014;21:1870-7.
- Na DG, Kim JH, Sung JY, Baek JH, Jung KC, Lee H, et al. Core-needle biopsy is more useful than repeat fine-needle aspiration in thyroid nodules read as nondiagnostic or atypia of undetermined significance by the Bethesda system for re-

- porting thyroid cytopathology. *Thyroid* 2012;22:468-75.
16. Park KT, Ahn SH, Mo JH, Park YJ, Park DJ, Choi SI, et al. Role of core needle biopsy and ultrasonographic finding in management of indeterminate thyroid nodules. *Head Neck* 2011;33:160-5.
 17. Raab SS, Vrbin CM, Grzybicki DM, Sudilovsky D, Balasanian R, Zarbo RJ, et al. Errors in thyroid gland fine-needle aspiration. *Am J Clin Pathol* 2006;125:873-82.
 18. Ahn HS, Youn I, Na DG, Kim SJ, Lee MY. Diagnostic performance of core needle biopsy as a first-line diagnostic tool for thyroid nodules according to ultrasound patterns: comparison with fine needle aspiration using propensity score matching analysis. *Clin Endocrinol (Oxf)* 2021;94:494-503.
 19. Ha EJ, Baek JH, Lee JH, Song DE, Kim JK, Shong YK, et al. Sonographically suspicious thyroid nodules with initially benign cytologic results: the role of a core needle biopsy. *Thyroid* 2013;23:703-8.
 20. Choi YJ, Baek JH, Ha EJ, Lim HK, Lee JH, Kim JK, et al. Differences in risk of malignancy and management recommendations in subcategories of thyroid nodules with atypia of undetermined significance or follicular lesion of undetermined significance: the role of ultrasound-guided core-needle biopsy. *Thyroid* 2014;24:494-501.
 21. Renshaw AA, Pinnar N. Comparison of thyroid fine-needle aspiration and core needle biopsy. *Am J Clin Pathol* 2007;128:370-4.
 22. Hong MJ, Na DG, Lee H. Diagnostic efficacy and safety of core needle biopsy as a first-line diagnostic method for thyroid nodules: a prospective cohort study. *Thyroid* 2020;30:1141-9.
 23. Jung CK, Min HS, Park HJ, Song DE, Kim JH, Park SY, et al. Pathology reporting of thyroid core needle biopsy: a proposal of the Korean Endocrine Pathology Thyroid Core Needle Biopsy Study Group. *J Pathol Transl Med* 2015;49:288-99.
 24. Min HS, Kim JH, Ryoo I, Jung SL, Jung CK. The role of core needle biopsy in the preoperative diagnosis of follicular neoplasm of the thyroid. *APMIS* 2014;122:993-1000.
 25. Trimboli P, Nasrollah N, Guidobaldi L, Taccogna S, Ciccarella Modica DD, Amendola S, et al. The use of core needle biopsy as first-line in diagnosis of thyroid nodules reduces false negative and inconclusive data reported by fine-needle aspiration. *World J Surg Oncol* 2014;12:61.
 26. Lee SH, Park GS, Jung SL, Kim MH, Bae JS, Lim DJ, et al. Core-needle biopsy for the preoperative diagnosis of follicular neoplasm in thyroid nodule screening: a validation study. *Pathol Res Pract* 2016;212:44-50.
 27. Yoon RG, Baek JH, Lee JH, Choi YJ, Hong MJ, Song DE, et al. Diagnosis of thyroid follicular neoplasm: fine-needle aspiration versus core-needle biopsy. *Thyroid* 2014;24:1612-7.
 28. Na HY, Woo JW, Moon JH, Choi JY, Jeong WJ, Kim YK, et al. Preoperative diagnostic categories of noninvasive follicular thyroid neoplasm with papillary-like nuclear features in thyroid core needle biopsy and its impact on risk of malignancy. *Endocr Pathol* 2019;30:329-39.
 29. Chung SR, Baek JH, Lee JH, Lee YM, Sung TY, Chung KW, et al. Risk of malignancy according to the sub-classification of atypia of undetermined significance and suspicious follicular neoplasm categories in thyroid core needle biopsies. *Endocr Pathol* 2019;30:146-54.