Eur Radiol (2017) 27:801–811
DOI 10.1007/s00330-016-4424-1

HEAD AND NECK

Fine-needle aspiration versus core needle biopsy for diagnosis of thyroid malignancy and neoplasm: a matched cohort study

Soo-Yeon Kim[1] · Hye Sun Lee[2] · Jieun Moon[2] · Eun-Kyung Kim[1] · Hee Jung Moon[1] · Jung Hyun Yoon[1] · Jin Young Kwak[1]

Received: 13 October 2015 / Revised: 29 January 2016 / Accepted: 20 May 2016 / Published online: 3 June 2016

Abstract

Objectives To compare the diagnostic performances of fine-needle aspiration (FNA) and core needle biopsy (CNB) in the diagnosis of thyroid malignancy and neoplasm in patients who underwent surgery for thyroid nodules.

Methods This retrospective study was approved by the institutional review board, and the need to obtain informed consent was waived. 3192 patients who underwent FNA ($n = 3048$) or CNB ($n = 144$) for diagnosis of thyroid nodules and then proceeded with surgery were included. Surgical pathologic diagnosis was the reference standard. Diagnostic performances of FNA and CNB to predict malignancy and neoplasm were compared. Propensity score matching was used to match patients with FNA with those with CNB because there were significant differences in the number of nodules and nodule characteristics between the FNA and CNB groups.

Results Before matching, the sensitivity and accuracy of FNA were significantly higher or comparable with those of CNB, and the specificity, negative predictive value and positive predictive value were comparable. After matching, the diagnostic performances were similar, with the exception of specificity for predicting neoplasm being higher with CNB than with FNA.

Conclusion FNA showed comparable diagnostic performance to CNB; therefore, there may be no benefit in performing CNB to diagnose papillary thyroid carcinoma and neoplasm.

Key Points

- *Diagnostic performances of FNA and CNB for thyroid malignancy and neoplasm were compared.*
- *FNA showed comparable performances to CNB both before and after statistical matching.*
- *There may be no benefit in performing CNB, given the comparable performances.*

Keywords Thyroid · Fine needle aspiration · Core needle biopsy · Malignancy · Neoplasm

Introduction

Fine-needle aspiration (FNA) is widely accepted as the primary diagnostic tool for the evaluation of thyroid nodules owing to its simplicity, safety, cost-effectiveness, and its high sensitivity of 83–98 % and specificity of 70–92 % [1, 2]. However, a major limitation of FNA is inconclusive diagnosis including rates of non-diagnostic results reported to be 2–24 % and atypia of undetermined significance/follicular lesion of undetermined significance (AUS/FLUS) results reported to be 1–27 % [3–5]. Core needle biopsy (CNB) has been suggested as a complementary method to FNA, mainly to overcome possible inconclusive diagnosis [6]. However, inconclusive results have still been unavoidable in CNB with reported rates of inconclusive results of 6.0–31.8 % [6–17].

A few limited studies have compared the diagnostic accuracy of ultrasound (US)-guided CNB with that of US-guided FNA in the diagnosis of malignancy [6, 9, 18–21]. The results of these studies have not been consistent: CNB was more accurate and sensitive than FNA in some studies [6, 9, 21], but other studies, including one meta-analysis, showed that

✉ Jin Young Kwak
docjin@yuhs.ac

[1] Department of Radiology, Severance Hospital, Research Institute of Radiological Science, Yonsei University College of Medicine, 50 Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea 03722

[2] Department of Research Affairs, Biostatistics Collaboration Unit, Yonsei University College of Medicine, Seoul, Korea

CNB had similar or rather lower accuracy and sensitivity compared to FNA [18–20]. If the diagnostic accuracy does not significantly differ between the two methods, there may be no benefit in performing CNB in spite of accepting risk of complications and technical difficulty, which are more common with CNB than with FNA.

Thus, the objective of our study was to compare the diagnostic performances of FNA and CNB in the diagnosis of thyroid malignancy and neoplasm in patients who underwent surgery for thyroid nodules.

Materials and methods

Patients

The institutional review board approved this retrospective study, and patient approval and informed consent were not required for the retrospective review of US images, medical records or cytopathology reports. However, written informed consent for US-FNA, CNB and surgery was obtained from all patients prior to procedures.

From July 2013 to April 2015, 3382 consecutive patients underwent staging US examinations prior to surgery. Of 3382 patients, 184 patients who did not proceed with surgery, and six patients who underwent FNA at another hospital but who did not have a slide review performed were excluded. Finally, a total of 3192 patients were included. The mean ± standard deviation age of the patients was 44.4 ± 12.3 years (range 10–87 years); there were 2522 women (mean age 44.4 ± 12.3 years, range 10–87 years) and 670 men (mean age 44.4 ± 12.3 years, range 20–85 years). Among the total included patients, 3048 (95.5 %) underwent US-FNA and 144 (4.5 %) underwent CNB to diagnose thyroid nodules prior to surgery. Of 3048 patients with FNA, 2538 (83.3 %) underwent FNA at another hospital, and 510 (16.7 %) did so at our institution. Of 144 patients with CNB, 134 (93.1 %) underwent CNB at another hospital, and 10 (6.9 %) did so at our institution to evaluate thyroid nodules with non-diagnostic ($n = 2$) or AUS/FLUS ($n = 5$) cytology on prior FNA, benign but discordant FNA cytology ($n = 1$) or increased size on follow-up US ($n = 2$). All 2672 patients with CNB or FNA performed at other hospitals had their slides reviewed by our pathologists prior to surgery. There were no complications or technical failures among the CNBs performed at our institution, but data on the occurrence of complications due to CNBs performed at other hospitals were not available.

Staging US examinations

Preoperative staging US included an examination of the thyroid nodule (size, extrathyroidal extension and US features) and cervical lymph nodes. US features of each thyroid nodule were described and recorded by the radiologists who performed staging US examinations according to the following categories: (a) internal component (completely solid nodules, or mainly cystic (cystic portion ≥ 50 %) or mainly solid (cystic portion <50 %) nodules in mixed cystic and solid nodules), (b) echogenicity (hyper-, iso- and hypoechogenicity when compared to the echogenicity of the underlying thyroid parenchyma, or marked hypoechogenicity compared to the adjacent strap muscle), (c) margin (circumscribed or non-circumscribed i.e. microlobulated or irregular), (d) calcifications (no calcification, microcalcification, macrocalcification or mixed calcification with both micro- and macrocalcifications) and (e) shape (parallel, greater in the transverse dimension than the anteroposterior dimension or non-parallel, greater in the anteroposterior dimension than the transverse dimension). Suspicious US features included marked hypoechogenicity, non-circumscribed margin, microcalcification or mixed calcification, and non-parallel shape, based on published criteria [22]. Final assessment was given as 'probable benign' for nodules without suspicious US features and 'suspicious malignant' for those with one or more suspicious US features.

Cytopathology reports

During the study period, eight cytopathologists with 8–21 years of experience in thyroid cytopathology reviewed the FNA and CNB slides according to the review schedule. FNA cytology diagnoses were reported on the basis of the Bethesda System for Reporting Thyroid Cytopathology with the following six categories: non-diagnostic, benign, atypia of undetermined significance or follicular lesion of undetermined significance (AUS/FLUS), FN/SFN (follicular neoplasm or suspicious for follicular neoplasm), suspicious for malignancy, and malignancy [23]. The diagnostic criteria for CNB of thyroid nodules have not yet been standardized [6, 9, 15]; in this study, CNB readings were therefore classified according to the six categories of the Bethesda system, like with the prior studies [6, 9, 15, 23]. A non-diagnostic reading included the absence of any identifiable follicular thyroid tissue, the presence of only a normal thyroid gland, and tissue containing only a few follicular cells insufficient for diagnosis. A benign reading included colloid nodules, nodular hyperplasia and lymphocytic thyroiditis. An AUS/FLUS reading included nodules in which some atypical cells were present, but were not diagnostic of suspicious malignancy or malignancy, and cellular follicular nodules in which distinction between adenomatous hyperplasia (AH) and FN was not possible. An FN/SFN reading included nodules with histological features favouring follicular neoplasm. A suspicious for malignancy reading was assigned when a specimen exhibited some atypical cells but was without sufficient evidence for malignancy. A malignancy reading was assigned when a specimen exhibited unequivocal features of cancer.

Data and statistical analysis

Because there were significant differences in nodule characteristics between patients with CNB and FNA as well as the overall number of nodules, patients with CNB were matched with those with FNA using propensity score matching with greedy algorithms and logistic regression analysis [24]. Matched variables included sex, age, Bethesda category, size on US, US composition, and US final assessment. Clinical and US characteristics were compared between patients with CNB and those with FNA using the independent two-sample *t* test for continuous variables and the Chi-square or Fisher's exact test for categorical variables before matching, and the paired *t* test for continuous variables and McNemar's test for categorical variables after matching.

Surgical pathologic diagnosis was defined as the reference standard. Malignancy rates and neoplasm rates were compared between patients with CNB and those with FNA using the Chi-square test in the non-matched cohort and generalized estimating equations (GEE) in the matched cohort. Diagnostic performances (sensitivity, specificity, accuracy, positive predictive value (PPV) and negative predictive value (NPV)) to assess malignancy and neoplasm were calculated, and compared between the CNB and FNA groups using the Chi-square or Fisher's exact test in the non-matched cohort and GEE in the matched cohort with the following six criteria. Neoplasm was defined as malignancy plus follicular adenoma and Hürthle cell adenoma. For the diagnostic criteria to assess malignancy, test-positives were defined by the Bethesda categories of suspicious for malignancy and malignancy. Test-negatives were defined by the Bethesda categories according to the following four criteria to assess malignancy:

Criterion 1: Non-diagnostic, Benign, AUS/FLUS, FN/SFN
Criterion 2: Non-diagnostic, Benign, AUS/FLUS
Criterion 3: Non-diagnostic, Benign
Criterion 4: Benign

For the diagnostic criteria to assess neoplasm, test-positives were defined by the Bethesda categories of FN/SFN, suspicious for malignancy and malignancy. Test-negatives were defined by the Bethesda categories according to the following three criteria to assess neoplasm:

Criterion 1: Non-diagnostic, Benign, AUS/FLUS
Criterion 2: Non-diagnostic, Benign
Criterion 3: Benign

All analyses were performed with SAS, version 9.2 (SAS Institute, Cary, NC). Two-sided *P* values less than 0.05 were considered statistically significant.

Results

Clinical and US characteristics of patients with FNA versus CNB

Table 1 demonstrates the clinical and US characteristics of the two groups with FNA and CNB before and after matching. Before matching, sex and age were similar for both groups. In terms of the Bethesda category, no patients had non-diagnostic results from CNB, while 0.5 % (15 of 3048) of patients had non-diagnostic results from FNA. The proportions of the AUS/FLUS, FN/SFN and malignancy categories were higher in the CNB group than in the FNA group. The proportion of the suspicious for malignancy category was lower in the CNB group than in the FNA group. Patients with CNB had more nodules that were 1 cm or larger in size, with mixed cystic and solid composition, and with probable benign assessments on US than patients with FNA. Patients with FNA were statistically matched with patients with CNB. After matching, there were no differences between the two groups with respect to the matched variables.

Comparison of malignancy rate and neoplasm rate of FNA versus CNB

Table 2 compares the malignancy rates and the neoplasm rates between the two groups according to the Bethesda category, and Table 3 compares the FNA or CNB results to the final surgical diagnosis. Before matching, the overall malignancy rate of the FNA group was higher than that of the CNB group (Table 2, 97.4 % vs. 91.0 %, $P<0.001$), and the overall neoplasm rate was not different between the two groups (98.0 % vs. 96.5 %, $P=0.239$). Malignancy rates and neoplasm rates according to each Bethesda category were not significantly different between the two methods, with the exception that the malignancy rate of the FNA group was higher than that of the CNB group in the AUS/FLUS category (Table 2, 83.0 % vs. 60.9 %, $P=0.022$).

After matching, no differences were found between both groups for the overall malignancy rate (Table 2, 90.3 % for FNA, and 91.0 % for CNB, $P=0.796$), overall neoplasm rate (93.8 % for FNA, 96.5 % for CNB, $P=0.273$) and malignancy rates and neoplasm rates according to each Bethesda category.

Diagnostic performances of FNA versus CNB

Before matching, when predicting malignancy, the sensitivity and accuracy of FNA were significantly higher than those of CNB using criterion 1 (Table 4, 93.8 % vs. 84.7 %, $P<0.001$ for sensitivity, 93.7 % vs. 86.1 %, $P<0.001$ for accuracy) and criterion 2 (94.0 % vs. 88.1 %, $P=0.008$ for sensitivity, 93.9 % vs. 89.1 %, $P=0.023$ for accuracy). No differences were found when using criteria 3 and 4. When predicting

Table 1 Clinical and ultrasonographic characteristics of patients with fine-needle aspiration (FNA) or core needle biopsy (CNB) before and after propensity score matching

	Before matching			After matching		
	FNA ($n=3048$)	CNB ($n=144$)	*P* value	FNA ($n=144$)	CNB ($n=144$)	*P* value
Sex			0.499			0.758
Female	2405 (78.9)	117 (81.25)		115 (79.9)	117 (81.3)	
Male	643 (21.1)	27 (18.75)		29 (20.1)	27 (18.8)	
Age	44.44 ± 12.33	43.69 ± 11.34	0.475	45.72 ± 13.69	43.69 ± 11.34	0.150
Bethesda category			<0.001			0.849
Non-diagnostic	15 (0.5)	0 (0.0)	0.826	0 (0.0)	0 (0.0)	
Benign	32 (1.1)	3 (2.1)	0.451	2 (1.4)	3 (2.1)	
AUS/FLUS	194 (6.4)	23[a] (16.0)	<0.001	22 (15.3)	23[a] (16.0)	
FN/SFN	18 (0.6)	7 (4.9)	<0.001	8 (5.6)	7 (4.9)	
Suspicious for malignancy	1082 (35.5)	11 (7.6)	<0.001	11 (7.6)	11 (7.6)	
Malignancy	1707 (56.0)	100 (69.4)	0.001	101 (70.1)	100 (69.4)	
Size on US			0.020			>0.999
< 1 cm	1698 (55.7)	66 (45.8)		66 (45.8)	66 (45.8)	
≥ 1 cm	1350 (44.3)	78 (54.2)		78 (54.2)	78 (54.2)	
Internal composition on US			0.001			0.217
Mainly cystic	14 (0.5)	4 (2.8)		1 (0.7)	4 (2.8)	
Mainly solid	167 (5.5)	9 (6.3)		11 (7.6)	9 (6.3)	
Completely solid	2867 (94.1)	131 (91.0)		132 (91.7)	131 (91.0)	
Final US assessment			<0.001			>0.999
Probable benign	186 (6.1)	26 (18.1)		26 (18.1)	26 (18.1)	
Suspicious	2862 (93.9)	118 (81.9)		118 (81.9)	118 (81.9)	

Mean ± standard deviation for continuous variables, or number (percentage) for categorical variables

AUS/FLUS atypia of undetermined significance or follicular lesion of undetermined significance, *FN/SFN* follicular neoplasm or suspicious for follicular neoplasm, *US* ultrasound

[a] On CNB, 23 nodules were classified as AUS/FLUS category. Of them, 11 were nodules in which some atypical cells were present, but were not diagnostic of suspicious malignancy or malignancy; and 12 were nodules in which distinction between adenomatous hyperplasia and FN was not possible according to the CNB pathology reports

neoplasm, the sensitivity and accuracy of FNA were significantly higher than those of CNB using criterion 1 (Table 5, 93.6 % vs. 84.9 %, $P<0.001$ for sensitivity, 93.3 % vs. 85.4 %, $P<0.001$ for accuracy). No differences were found when using criteria 2 and 3. The specificity, negative predictive value and positive predictive value were comparable between FNA and CNB.

After matching, diagnostic performances to predict malignancy or neoplasm were not different between the FNA and CNB groups, except that the specificity of CNB was significantly higher than that of FNA when using criterion 2 or 3 to predict neoplasm (Table 5, 100.0 % vs. 50.0 %, $P=0.046$), because of the two false-positive diagnoses (Table 3, two AH) for the FN/SFN category of FNA.

Table 2 Comparison of the malignancy rate and neoplasm rate of fine-needle aspiration (FNA) versus core needle biopsy (CNB) according to the Bethesda category

	Before matching													
	FNA ($n=3048$)						CNB ($n=144$)							
	Total	Benign	Malignancy	Malignancy rate (%)	Neoplasm	Neoplasm rate (%)	Total	Benign	Malignancy	Malignancy rate (%)	Neoplasm	Neoplasm rate (%)	*P* value for the malignancy rate	*P* value for the neoplasm rate
Non-diagnostic	15 (0.5)	2 (2.5)	13 (0.4)	86.7	13 (0.4)	86.7	0 (0.0)	0 (0.0)	0 (0.0)	0.0	0 (0.0)	–	–	–
Benign	32 (1.1)	27 (33.8)	5 (0.2)	15.6	9 (0.3)	28.1	3 (2.1)	2 (15.4)	1 (0.8)	33.3	1 (0.7)	33.3	0.442	>0.999
AUS/FLUS	194 (6.4)	33 (41.3)	161 (5.4)	83.0	169 (5.7)	87.1	23[a] (15.9)	9 (69.2)	14 (10.7)	60.9	20 (14.4)	87.0	**0.022**	>0.999
FN/SFN	18 (0.6)	12 (15.0)	6 (0.2)	33.3	12 (0.4)	66.7	7 (4.9)	2 (15.4)	5 (3.8)	71.4	7 (5.0)	100.0	0.178	0.137
Suspicious for malignancy	1082 (35.5)	4 (5.0)	1078 (36.3)	99.6	1078 (36.1)	99.6	11 (7.6)	0 (0.0)	11 (8.4)	100.0	11 (8.0)	100.0	>0.999	>0.999
Malignancy	1707 (56.0)	2 (2.5)	1705 (57.5)	99.9	1705 (57.1)	99.9	100 (69.5)	0 (0.0)	100 (76.4)	100.0	100 (71.9)	100.0	>0.999	>0.999
Total	3048	80	2968	97.4	2986	98.0	144	13	131	91.0	139	96.5	<0.001	0.239
	After matching													
	FNA ($n=144$)						CNB ($n=144$)							
	Total	Benign	Malignancy	Malignancy rate (%)	Neoplasm	Neoplasm rate (%)	Total	Benign	Malignancy	Malignancy rate (%)	Neoplasm	Neoplasm rate (%)	*P* value for the malignancy rate	*P* value for the neoplasm rate
Non-diagnostic	0 (0.0)	0 (0.0)	0 (0.0)	–	0 (0.0)	–	0 (0.0)	0 (0.0)	0 (0)	–	0 (0)	–	–	–
Benign	2 (1.4)	2 (14.3)	0 (0.0)	0.0	0 (0.0)	0	3 (2.0)	2 (15.4)	1 (0.8)	33.3	1 (0.7)	33.3	0.221	>0.999
AUS/FLUS	22 (15.3)	6 (42.9)	16 (12.3)	72.7	17 (12.6)	77.3	23[a] (16.0)	9 (69.2)	14 (10.7)	60.9	20 (14.4)	86.6	0.395	0.456
FN/SFN	8 (5.6)	6 (42.9)	2 (1.5)	25.0	6 (4.4)	75	7 (4.9)	2 (15.4)	5 (3.8)	71.4	7 (5.0)	100	0.075	0.467
Suspicious for malignancy	11 (7.6)	0 (0.0)	11 (8.5)	100.0	11 (8.2)	100	11 (7.6)	0 (0.0)	11 (8.4)	100.0	11 (7.9)	100	>0.999	>0.999
Malignancy	101 (70.1)	0 (0.0)	101 (77.7)	100.0	101 (74.8)	100	100 (69.4)	0 (0.0)	100 (76.3)	100.0	100 (71.9)	100	>0.999	>0.999
Total	144	14	130	90.3	135	93.8	144	13	131	91.0	139	96.5	0.796	0.273

Values given are numbers (percentages) or percentages. Neoplasm is defined as malignancy plus follicular adenoma and Hürthle cell adenoma. Matched variables included sex, age, Bethesda category, nodule size on ultrasound (US), internal composition and final US assessment

AUS/FLUS atypia of undetermined significance or follicular lesion of undetermined significance, *FN/SFN* follicular neoplasm or suspicious for follicular neoplasm, *PTC* papillary thyroid carcinoma

P values with statistical significance ($P<0.05$)

[a] Of 23 nodules with AUS/FLUS result on CNB, 11 were nodules in which some atypical cells were present, but were not diagnostic of suspicious malignancy or malignancy; and 12 were nodules in which distinction between adenomatous hyperplasia and FN was not possible according to the CNB pathology reports. Of the 11 nodules, malignancy rate and neoplasm rate was 90.9 % (10/11). Of the 12 nodules, malignancy rate was 33.3 % (4/12) and neoplasm rate was 83.3 % (10/12)

Table 3 Comparison of the fine-needle aspiration (FNA) or core needle biopsy (CNB) diagnosis with the final surgical diagnosis

FNA before matching ($n = 3408$)		
Diagnosis	No. of nodules	Surgical diagnosis
Non-diagnostic	15	2 Benign (AH)
		13 Malignancy (12 PTC, 1 MTC)
Benign	32	27 Benign (21 AH, 4 FA, 2 LT)
		5 Malignancy (4 PTC, 1 FTC)
AUS/FLUS	194	33 Benign (20 AH, 8 FA, 4 LT, 1 fibrosis)
		161 Malignancy (151 PTC, 9 FTC, 1 MTC)
FN/SFN	18	12 Benign (6 FA, 5 AH,1 LT)
		6 Malignancy (5 FTC, 1 PTC)
Suspicious for malignancy	1082	4 Benign (2 AH, 2 LT)
		1078 Malignancy (1075 PTC, 1 metastasis from lung cancer, 1 poorly differentiated carcinoma, 1 carcinoma showing thymus-like differentiation)
Malignancy	1707	2 Benign (2 HTT)
		1705 Malignancy (1696 PTC, 5 MTC, 2 FTC, 2 anaplastic carcinoma)
FNA after matching ($n = 144$)		
Diagnosis	No. of nodules	Surgical diagnosis
Non-diagnostic	0	
Benign	2	2 Benign (1 AH, 1 LT)
AUS/FLUS	22	6 Benign (3 AH, 1 FA, 1 fibrosis, 1 LT)
		16 Malignancy (15 PTC, 1 FTC)
FN/SFN	8	6 Benign (4 FA, 2 AH)
		2 Malignancy (1 PTC, 1 FTC)
Suspicious for malignancy	11	11 Malignancy (11 PTC)
Malignancy	101	101 Malignancy (101 PTC)
CNB ($n = 144$)		
Diagnosis	No. of nodules	Surgical diagnosis
Non-diagnostic	0	
Benign	3	2 Benign (2 AH)
		1 Malignancy (1 FTC)
AUS/FLUS	23[a]	9 Benign (4 FA, 3 AH, 2 HCA)
		14 Malignancy (14 PTC)
FN/SFN	7	2 Benign (2 FA)
		5 Malignancy (4 PTC, 1 FTC)
Suspicious for malignancy	11	11 Malignancy (9 PTC, 2 FTC)
Malignancy	100	100 Malignancy (99 PTC, 1 MTC)

AUS/FLUS atypia of undetermined significance or follicular lesion of undetermined significance, *FN/SFN* follicular neoplasm or suspicious for follicular neoplasm, *AH* adenomatous hyperplasia, *FA* follicular adenoma, *FTC* follicular thyroid carcinoma, *HCA* Hürthle cell adenoma, *HTT* hyaline trabecular tumour, *LT* lymphocytic thyroiditis, *MTC* medullary thyroid carcinoma, *PTC* papillary thyroid carcinoma

[a] Of 23 nodules with AUS/FLUS result on CNB, 11 were nodules in which some atypical cells were present, but were not diagnostic of suspicious malignancy or malignancy; and 12 were nodules in which distinction between AH and FN was not possible according to the CNB pathology reports. After surgery, of the 11 nodules, one was benign (AH), and 10 were malignancy (PTC). Of the 12 nodules, 8 were benign (4 FA, 2 AH, 2 HCA) and 4 were malignancy (PTC)

Table 4 Diagnostic performances of fine-needle aspiration (FNA) versus core needle biopsy (CNB) to predict malignancy

	Before matching											
	Criterion 1			Criterion 2			Criterion 3			Criterion 4		
	FNA	CNB	*P* value	FNA	CNB	*P* value	FNA	CNB	*P* value	FNA	CNB	*P* value
TP	2783	111		2783	111		2783	111		2783	111	
TN	74	13		62	11		29	2		27	2	
FP	6	0		6	0		6	0		6	0	
FN	185	20		179	15		18	1		5	1	
Sensitivity	93.8 (92.9–94.6)	84.7 (78.6–90.9)	**<0.001**	94.0 (93.1–94.8)	88.1 (82.4–93.7)	**0.008**	99.4 (99.1–100.0)	99.1 (97.4–100.0)	0.526	99.8 (99.7–99.9)	99.1 (97.4–100.0)	0.211
Specificity	92.5 (86.7–98.3)	100.0 (100.0–100.0)	0.590	91.2 (84.4–97.9)	100.0 (100.0–100.0)	0.588	82.9 (70.4–95.3)	100.0 (100.0–100.0)	>0.999	81.8 (68.7–95.0)	100.0 (100.0–100.0)	>0.999
Accuracy	93.7 (92.9–94.6)	86.1 (80.5–91.8)	**<0.001**	93.9 (93.0–94.7)	89.1 (83.8–94.3)	**0.023**	99.2 (98.8–99.5)	99.1 (97.4–100.0)	>0.999	99.6 (83.8–94.3)	99.1 (97.4–100.0)	0.379
Positive predictive value	99.8 (99.6–99.9)	100.0 (100.0–100.0)	>0.999	99.8 (99.6–100)	100.0 (100.0–100.0)	>0.999	99.8 (99.6–100)	100.0 (100.0–100.0)	>0.999	99.8 (99.6–99.9)	100.0 (100.0–100.0)	>0.999
Negative predictive value	28.6 (23.1–34.1)	39.4 (22.7–56.1)	0.201	25.7 (20.2–31.2)	42.3 (23.3–61.3)	0.072	61.7 (47.8–75.6)	66.7 (13.3–100.0)	>0.999	84.4 (71.8–97.0)	66.7 (13.3–100.0)	0.442
	After matching											
	Criterion 1			Criterion 2			Criterion 3			Criterion 4		
	FNA	CNB	*P* value	FNA	CNB	*P* value	FNA	CNB	*P* value	FNA	CNB	*P* value
TP	112	111		112	111		112	111		112	111	
TN	14	13		8	11		2	2		2	2	
FP	0	0		0	0		0	0		0	0	
FN	18	20		16	15		0	1		0	1	
Sensitivity	86.2 (80.2–92.1)	84.7 (78.6–90.9)	0.596	87.5 (81.8–93.2)	88.1 (82.4–93.7)	0.800	100.0 (100.0–100.0)	99.1 (97.4–100)	0.315	100.0 (100.0–100.0)	99.1 (97.4–100.0)	0.315
Specificity	100.0 (100.0–100.0)	100.0 (100.0–100.0)	>0.999	100.0 (100.0–100.0)	100.0 (100.0–100.0)	>0.999	100.0 (100.0–100.0)	100.0 (100.0–100.0)	>0.999	100.0 (100.0–100.0)	100.0 (100.0–100.0)	>0.999
Accuracy	87.5. (82.1–92.9)	86.1 (80.5–91.8)	0.617	88.2 (82.8–93.7)	89.1 (83.8–94.3)	0.737	100.0 (100.0–100.0)	99.1 (97.4–100.0)	0.315	100.0 (100.0–100.0)	99.1 (97.4–100.0)	0.315
Positive predictive value	100.0 (100.0–100.0)	100.0 (100.0–100.0)	>0.999	100.0 (100.0–100.0)	100.0 (100.0–100.0)	>0.999	100.0 (100.0–100.0)	100.0 (100.0–100.0)	>0.999	100.0 (100.0–100.0)	100.0 (100.0–100.0)	>0.999
Negative predictive value	43.8 (26.6–60.9)	39.4 (22.7–56.1)	0.711	33.3 (14.5–52.2)	42.3 (23.3–61.3)	0.477	100.0 (100.0–100.0)	66.7 (13.3–100.0)	0.221	100.0 (100.0–100.0)	66.7 (13.3–100.0)	0.221

Values are numbers or percentages (95 % confidence interval)

TP true positive, *TN* true negative, *FP* false positive, *FN* false negative

Test-negatives are defined according to criteria 1 to 4 by the following Bethesda categories:

Criterion 1: Non-diagnostic, Benign, AUS/FLUS, FN/SFN

Criterion 2: Non-diagnostic, Benign, AUS/FLUS

Criterion 3: Non-diagnostic, Benign

Criterion 4: Benign

Test-positives are defined by the following Bethesda categories: suspicious for malignancy, malignancy

P values with statistical significance ($P<0.05$)

Table 5 Diagnostic performances of fine-needle aspiration (FNA) versus core needle biopsy (CNB) to predict neoplasm

	Before matching								
	Criterion 1			Criterion 2			Criterion 3		
	FNA	CNB	*P* value	FNA	CNB	*P* value	FNA	CNB	*P* value
TP	2795	118		2795	118		2795	118	
TN	50	5		25	2		23	2	
FP	12	0		12	0		12	0	
FN	191	21		22	1		9	1	
Sensitivity	93.6 (92.7–94.5)	84.9 (78.9–90.9)	**<0.001**	99.2 (98.9–99.5)	99.2 (97.6–100)	0.615	99.7 (99.5–99.9)	99.2 (97.5–100.0)	0.341
Specificity	80.7 (70.8–90.5)	100.0 (100.0–100.0)	0.576	67.6 (52.5–82.7)	100.0 (100.0–100.0)	>0.999	65.7 (50.0–81.4)	100.0 (100.0–100.0)	>0.999
Accuracy	93.3 (92.5–94.2)	85.4 (79.7–91.2)	**<0.001**	98.8 (98.4–99.2)	99.2 (97.6–100)	>0.999	99.3 (99.0–99.6)	99.2 (97.6–100.0)	0.602
Positive predictive value	99.6 (99.3–99.8)	100.0 (100.0–100.0)	>0.999	99.6 (99.3–99.8)	100.0 (100.0–100.0)	>0.999	99.6 (99.3–99.8)	100.0 (100.0–100.0)	>0.999
Negative predictive value	20.8 (15.6–25.9)	19.2 (4.1–34.4)	0.856	53.2 (38.9–67.5)	66.7 (13.3–100.0)	>0.999	71.9 (56.3–87.5)	66.7 (13.3–100.0)	>0.999
	After matching								
	Criterion 1			Criterion 2			Criterion 3		
	FNA	CNB	*P* value	FNA	CNB	*P* value	FNA	CNB	*P* value
TP	118	118		118	118		118	118	
TN	7	5		2	2		2	2	
FP	2	0		2	0		2	0	
FN	17	21		0	1		0	1	
Sensitivity	87.4 (81.8–93.0)	84.9 (78.9–90.9)	0.220	100.0 (100.0–100.0)	99.2 (97.5–100.0)	0.315	100.0 (100.0–100.0)	99.2 (97.5–100.0)	0.315
Specificity	77.8 (50.6–100.0)	100.0 (100.0–100.0)	0.109	50.0 (1.0–99.0)	100.0 (100.0–100.0)	**0.046**	50.0 (1.0–99.0)	100.0 (100.0–100.0)	**0.046**
Accuracy	86.8 (81.3–92.3)	85.4 (79.7–91.2)	0.527	98.4 (96.1–100.0)	99.2 (97.6–100.0)	0.566	98.4 (96.1–100.0)	99.2 (97.6–100.0)	0.566
Positive predictive value	98.3 (96.0–100.0)	100.0 (100.0–100.0)	0.154	98.3 (96.0–100.0)	100.0 (100.0–100.0)	0.154	98.3 (96.0–100.0)	100.0 (100.0–100.0)	0.154
Negative predictive value	29.2 (11.0–47.4)	19.2 (4.1–34.4)	0.366	100.0 (100.0–100.0)	66.7 (13.3–100.0)	0.221	100.0 (100.0–100.0)	66.7 (13.3–100.0)	0.221

Values are numbers or percentages (95 % confidence interval)

TP true positive, *TN* true negative, *FP* false positive, *FN* false negative

Test-negatives are defined according to criteria 1 to 3 by the following Bethesda categories:

Criterion 1: Non-diagnostic, Benign, AUS/FLUS

Criterion 2: Non-diagnostic, Benign

Criterion 3: Benign

Test-positives are defined by the following Bethesda categories: FN/SFN, suspicious for malignancy, malignancy

P values with statistical significance ($P<0.05$)

Discussion

We compared the diagnostic performances of FNA and CNB to diagnose malignancy and neoplasm in patients who underwent surgery for thyroid nodules. Since there were significant differences in terms of the Bethesda category, nodule size, internal composition on US, and final US assessment, as well as the overall number of nodules between the FNA and CNB groups, the results found after matching these variables were considered to be more appropriate for the comparison of FNA and CNB. Before matching, FNA showed similar or significantly higher sensitivity and accuracy than CNB. After matching, the diagnostic performances of the two methods were similar, with the exception of criteria 2 and 3 showing a higher specificity of CNB in predicting neoplasm. Therefore, our results found that there may be no benefit in performing CNB over FNA, given the two methods show comparable diagnostic performances.

Until now, only a few studies have compared the diagnostic performances of CNB with those of FNA in the diagnosis of thyroid malignancy [6, 9, 18–21]. Of these prior studies, the study by Sung et al. evaluated 555 thyroid nodules, the largest number with regard to sample size, and used the same diagnostic criteria as our criterion 1 to calculate diagnostic performances [9]. In that study, the sensitivity and accuracy of CNB were significantly higher than those of FNA to diagnose malignancy (86.8 % vs. 68.6 % for sensitivity, 92.1 % vs. 82.0 % for accuracy) and neoplasm (84.7 % vs. 65.5 % for sensitivity, 90.3 % vs. 78.9 % for accuracy) [9], contrary to our results. These conflicting results may be due to how the final diagnoses were determined as they were defined differently in the two studies. In our study, all final diagnoses were determined by histopathological results after surgical resection, while in the study by Sung et al. the final diagnoses for benignity were mostly determined by the combination of follow-up US and FNA or CNB results without surgery (82.7 % of benign nodules) [9]. Also, inter-observer variability among the different pathologists might have affected the interpretations of FNA and CNB results [25].

Several recent studies have reported the usefulness of CNB for nodules with initial non-diagnostic FNA results, showing significantly lower non-diagnostic rates in CNB compared to repeat US-FNA [6, 8, 10, 26, 27]. The lower non-diagnostic rate of CNB may be a natural result because CNB can obtain larger tissue samples, and thus CNB may help clinicians appropriately manage nodules with non-diagnostic FNA results [16, 28]. However, the full clinical impact of this approach remains uncertain [16, 28]. In recent studies which included non-diagnostic FNA samples with repeat FNA or follow-up US, as well as those with surgery, the frequency of the malignancy rate was 1.6–3.0 % for nodules lacking suspicious US features. Given this low malignancy rate, a more conservative approach with follow-up US examinations instead of repeat FNA or CNB has been proposed for nodules with non-diagnostic FNA results, particularly for those lacking suspicious US features [29, 30].

A recent meta-analysis by Li et al. [20] found that FNA and CNB do not differ significantly in sensitivity and specificity for diagnosis of thyroid malignancy, concordant with our results. The areas under the ROC curves of FNA were even higher than those of CNB without statistical significance (0.905 ± 0.030 for FNA vs. 0.745 ± 0.095, for CNB, $P = 0.053$) [20]. This meta-analysis included the aforementioned study by Sung et al. along with four other studies [6, 9, 18, 19, 21]. Of note, despite the large sample size of the study by Sung et al. (64.7 %, 555 of the total 858 thyroid nodules included in the meta-analysis), the results of the meta-analysis showed comparable sensitivity and accuracy of FNA and CNB in the diagnosis of malignancy [20]. A previous study by Yoon et al. reported that CNB can provide more accurate information in differentiating follicular neoplasms from non-neoplasms, because CNB can obtain tissue samples containing nodular tissue, the fibrous capsule of the nodule, and the extranodular tissue [31]. However, in our study, the diagnostic performance of CNB to diagnose neoplasm did not differ from that of FNA. Also, our results showed that it is still difficult to differentiate AH (which lacks a well-defined, complete fibrous capsule) and FN (completely encapsulated lesion by capsule) using CNB, concordant with several previous studies [32, 33]. These CNB readings showing the difficult differentiation between AH and FN accounted for 8.3 % (footnote of Table 1; 12/144) of total readings, and the neoplasm rate and malignancy rate of the readings were 83.3 % (footnote of Table 2; 10/12) and 33.3 % (4/12), respectively. CNB cannot be used to distinguish between follicular carcinoma and follicular adenoma, because the pathologist might not be able to review the whole nodule for invasion through the capsule [16, 33].

The main concerns with performing CNB are safety problems and complications [16, 32, 34]. The most common complications are bleeding and haematoma [32]. Although the reported complication rates of CNB are low, ranging from 0.5 % to 1.0 % [32], and although reported patient discomfort levels and tolerability are similar between FNA and CNB [35], CNB can lead to severe and critical complications such as injury to the carotid artery, trachea or adjacent nerves. An iatrogenic arteriovenous fistula can occur after CNB, causing tinnitus [34]. With the use of a larger needle size, CNB has a greater potential for more serious and permanent complications compared to the transient complications by FNA. Also, CNB may be technically unfeasible or difficult to perform in certain cases, especially in nodules in close proximity to the carotid artery or trachea or in nodules located at the posterior margin of the thyroid [9, 16]. While FNA is relatively safe and feasible even when performed by less-experienced performers, CNB should be performed by experienced

radiologists who are familiar with radiologic features of important anatomic structures of the cervical region to avoid major complications [16].

We acknowledge several limitations. First, there may be a selection bias. From the beginning of the study, thyroid nodules without surgery were not included because they did not have gold standard results. This may affect the high malignancy rates of the non-diagnostic, benign and AUS/FLUS category. Also, a comparison of the inconclusive rates (i.e. non-diagnostic plus AUS/FLUS) between FNA and CNB was inappropriate with our study population. Second, we did not perform simultaneous FNA and CNB on thyroid nodules, unlike the previously mentioned studies [6, 9, 18–21]. Instead, we chose to compare the diagnostic performances of the two methods by establishing a matched cohort. Third, another limitation is the lack of diagnostic category standardization in the pathologic interpretation of CNB. Recently, the Korean Endocrine Pathology Thyroid Core Needle Biopsy Study Group suggested a guideline for the standard pathology reporting system of CNB [36]. Using a standard pathology reporting system would help clinicians properly manage patients with CNB, and help them accurately evaluate the diagnostic performances of CNB. Fourth, the majority of FNA (83.3 %) and CNB (93.1 %) were performed at other hospitals, although all other hospital slides were re-reviewed by our cytopathologists. Inter-hospital and inter-performer variability might have occurred. Fifth, CNB has been reported to be helpful in the specific diagnosis for malignancy other than papillary thyroid carcinoma such as lymphoma or anaplastic carcinoma [37], but an analysis was not performed on this issue because of the small number of malignancies other than PTC. Our results cannot be generalized to all thyroid malignancies other than papillary thyroid carcinomas, because most of the malignancies included in our study were papillary thyroid carcinomas (98.9 %, 3065 of 3099).

In conclusion, FNA showed comparable diagnostic performance to CNB; therefore, there may be no benefit in performing CNB to diagnose papillary thyroid carcinoma and neoplasm.

Acknowledgments The scientific guarantor of this publication is Eun-Kyung Kim. The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article. We thank Hye Sun Lee and Jieun Moon, statistician of the Biostatistics Collaboration Unit, Medical Research Center, Yonsei University, College of Medicine, Seoul, Korea for her help with the statistical analysis. Institutional review board approval (Yonsei University Health System, 4-2015-0574) was obtained. Informed consent was waived by the institutional review board. No study subjects or cohorts have been previously reported. Methodology: retrospective, diagnostic or prognostic study, performed at one institution.

References

1. Cooper DS, Doherty GM, Haugen BR et al (2009) Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association (ATA) guidelines taskforce on thyroid nodules and differentiated thyroid cancer. Thyroid 19:1167–1214
2. Gharib H, Papini E, Paschke R et al (2010) AACE/AME/ETA Task Force on Thyroid Nodules. American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and European Thyroid Association medical guidelines for clinical practice for the diagnosis and management of thyroid nodules: executive summary of recommendations. Endocr Pract 16:1–43
3. Yoon JH, Moon HJ, Kim E-K, Kwak JY (2011) Inadequate cytology in thyroid nodules: should we repeat aspiration or follow-up? Ann Surg Oncol 18:1282–1289
4. Yoon JH, Kwak JY, Kim E-K et al (2010) How to approach thyroid nodules with indeterminate cytology. Ann Surg Oncol 17:2147–2155
5. Ho AS, Sarti EE, Jain KS et al (2014) Malignancy rate in thyroid nodules classified as Bethesda category III (AUS/FLUS). Thyroid 24:832–839
6. Na DG, Kim J-H, Sung JY et al (2012) 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 reporting thyroid cytopathology. Thyroid 22:468–475
7. Ha EJ, Baek JH, Lee JH et al (2013) Sonographically suspicious thyroid nodules with initially benign cytologic results: the role of a core needle biopsy. Thyroid 23:703–708
8. Choi SH, Baek JH, Lee JH et al (2014) Thyroid nodules with initially non-diagnostic, fine-needle aspiration results: comparison of core-needle biopsy and repeated fine-needle aspiration. Eur Radiol 24:2819–2826
9. Sung JY, Na DG, Kim KS et al (2012) Diagnostic accuracy of fine-needle aspiration versus core-needle biopsy for the diagnosis of thyroid malignancy in a clinical cohort. Eur Radiol 22:1564–1572
10. Yeon JS, Baek JH, Lim HK et al (2013) Thyroid nodules with initially nondiagnostic cytologic results: the role of core-needle biopsy. Radiology 268:274–280
11. Park KT, Ahn SH, Mo JH et al (2011) Role of core needle biopsy and ultrasonographic finding in management of indeterminate thyroid nodules. Head Neck 33:160–165
12. Choi YJ, Baek JH, Ha EJ et al (2014) 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 24:494–501
13. Zhang M, Zhang Y, Fu S, Lv F, Tang J (2014) Thyroid nodules with suspicious ultrasound findings: the role of ultrasound-guided core needle biopsy. Clin Imaging 38:434–438
14. Khoo T-K, Baker C, Hallanger-Johnson J et al (2008) Comparison of ultrasound-guided fine-needle aspiration biopsy with core-needle biopsy in the evaluation of thyroid nodules. Endocr Pract 14:426–431
15. Ha EJ, Baek JH, Lee JH et al (2014) Core needle biopsy can minimise the non-diagnostic results and need for diagnostic surgery in patients with calcified thyroid nodules. Eur Radiol 24:1403–1409
16. Yoon JH, Kim E-K, Kwak JY, Moon HJ (2015) Effectiveness and limitations of core needle biopsy in the diagnosis of thyroid nodules: review of current literature. J Pathol Transl Med 49:230
17. Kim YH, Kwon HJ, Kim E-K, Kwak JY, Moon HJ, Yoon JH (2015) Applying ultrasound-guided core needle biopsy for

diagnosis of thyroid masses: preliminary results from a single institution. J Ultrasound Med 34:1801–1808
18. Karstrup S, Balslev E, Juul N, Eskildsen PC, Baumbach L (2001) US-guided fine needle aspiration versus coarse needle biopsy of thyroid nodules. Eur J Ultrasound 13:1–5
19. Renshaw AA, Pinnar N (2007) Comparison of thyroid fine-needle aspiration and core needle biopsy. Am J Clin Pathol 128:370–374
20. Li L, Chen B-D, Zhu H-F et al (2013) Comparison of pre-operation diagnosis of thyroid cancer with fine needle aspiration and core-needle biopsy: a meta-analysis. Asian Pac J Cancer Prev 15:7187–7193
21. Hakala T, Kholová I, Sand J, Saaristo R, Kellokumpu-Lehtinen P (2013) A core needle biopsy provides more malignancy-specific results than fine-needle aspiration biopsy in thyroid nodules suspicious for malignancy. J Clin Pathol 66:1046–1050
22. Kim EK, Park CS, Chung WY et al (2002) New sonographic criteria for recommending fine-needle aspiration biopsy of nonpalpable solid nodules of the thyroid. AJR Am J Roentgenol 178:687–691
23. Cibas ES, Ali SZ (2009) The Bethesda system for reporting thyroid cytopathology. Am J Clin Pathol 132:658–665
24. Austin PC, Grootendorst P, Anderson GM (2007) A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. Stat Med 26:734–753
25. Clary KM, Condel JL, Liu Y, Johnson DR, Grzybicki DM, Raab SS (2005) Interobserver variability in the fine needle aspiration biopsy diagnosis of follicular lesions of the thyroid gland. Acta Cytol 49:378–382
26. Lee SH, Kim MH, Bae JS, Lim DJ, Jung SL, Jung CK (2014) Clinical outcomes in patients with non-diagnostic thyroid fine needle aspiration cytology: usefulness of the thyroid core needle biopsy. Ann Surg Oncol 21:1870–1877
27. Samir AE, Vij A, Seale MK et al (2012) Ultrasound-guided percutaneous thyroid nodule core biopsy: clinical utility in patients with prior nondiagnostic fine-needle aspirate. Thyroid 22:461–467
28. Haugen B, Alexander E, Bible K et al (2016) 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: The American Thyroid Association Guidelines Taskforce on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid 26:1–133
29. Anderson TJ, Atalay MK, Grand DJ, Baird GL, Cronan JJ, Beland MD (2014) Management of nodules with initially nondiagnostic results of thyroid fine-needle aspiration: can we avoid repeat biopsy? Radiology 272:777–784
30. Yoon JH, Lee HS, Kim E-K, Moon HJ, Kwak JY (2015) Thyroid nodules: nondiagnostic cytologic results according to thyroid imaging reporting and data system before and after application of the Bethesda system. Radiology 276:579–587
31. Yoon RG, Baek JH, Lee JH et al (2014) Diagnosis of thyroid follicular neoplasm: fine-needle aspiration versus core-needle biopsy. Thyroid 24:1612–1617
32. Novoa E, Gürtler N, Arnoux A, Kraft M (2012) Role of ultrasound-guided core-needle biopsy in the assessment of head and neck lesions: a meta-analysis and systematic review of the literature. Head Neck 34:1497–1503
33. Min HS, Kim JH, Ryoo I, Jung SL, Jung CK (2014) The role of core needle biopsy in the preoperative diagnosis of follicular neoplasm of the thyroid. APMIS 122:993–1000
34. Bergeron M, Beaudoin D (2014) Simple core-needle biopsy for thyroid nodule, complicated tinnitus. Eur Thyroid J 3:130–133
35. Nasrollah N, Trimboli P, Rossi F et al (2014) Patient's comfort with and tolerability of thyroid core needle biopsy. Endocrine 45:79–83
36. Jung CK, Min HS, Park HJ et al (2015) Pathology reporting of thyroid core needle biopsy: a proposal of the Korean Endocrine Pathology Thyroid Core Needle Biopsy Study Group. J Pathol Transl Med 49:288
37. Hahn S, Shin J, Han B, Ko E, Ko E (2013) Ultrasonography-guided core needle biopsy for the thyroid nodule: does the procedure hold any benefit for the diagnosis when fine-needle aspiration cytology analysis shows inconclusive results? Br J Radiol 86:20130007