

UCSF

UC San Francisco Previously Published Works

Title

A Large Thyroid Fine Needle Aspiration Biopsy Cohort with Long-Term Population-Based Follow-Up

Permalink

<https://escholarship.org/uc/item/99w209ch>

Journal

Thyroid, 31(7)

ISSN

1050-7256

Authors

Ng, Dianna L
van Zante, Annemieke
Griffin, Ann
[et al.](#)

Publication Date

2021-07-01

DOI

10.1089/thy.2020.0689

Peer reviewed

A Large Thyroid Fine Needle Aspiration Biopsy Cohort with Long-Term Population-Based Follow-Up

Dianna L. Ng,¹⁻³ Annemieke van Zante,¹ Ann Griffin,² Nancy K. Hills,^{4,5} and Britt-Marie Ljung^{1,2}

Background: Prior studies evaluating thyroid fine needle aspiration biopsies (FNABs) have limited the calculation of risk of malignancy (ROM) to cytologic specimens with corresponding histologic specimens, and clinical follow-up for those patients who do not undergo immediate surgery has been largely disregarded. Moreover, there is marked variability in how researchers have approached thyroid FNAB statistical analyses. This study addresses the urgent need for information from a large cohort of patients with long-term clinical follow-up to more accurately determine the performance of thyroid FNAB and ROM for each diagnostic category.

Methods: A retrospective review of the University of California, San Francisco (UCSF), pathology database for thyroid FNABs from January 1, 1997, to December 31, 2004, was performed. Diagnoses were coded using the 2017 The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC), and patients were matched to both the UCSF cancer registry and California Cancer Registry. Data were analyzed using the Kaplan–Meier method, and stratified by TBSRTC diagnostic category. Kaplan–Meier curves were used to estimate incidence rates of malignancy, stratified by FNAB category. Cox proportional hazards models were used to determine the instantaneous ROM.

Results: Initial FNABs from 2207 patients were included. Median follow-up period after the first thyroid FNAB was 13.9 years (range: 10.5–18.4 years). During follow-up, there were 279 confirmed diagnoses of thyroid malignancy. Estimates derived from Kaplan–Meier curves demonstrated that the risk of having a thyroid malignancy was low for nondiagnostic and benign categories, intermediate for atypia of undetermined significance (AUS), follicular lesion of undetermined significance (FLUS), AUS/FLUS combined, and follicular neoplasm, and high for suspicious and malignant categories. A total of 52/1575 false-negative cases (3.2%) were identified. Excluding papillary microcarcinomas, the false-negative rate was 1.5% (23/1575). No patients with a false-negative diagnosis died of thyroid cancer during the follow-up period.

Conclusions: Asymptomatic patients with low-risk clinical and radiologic features and initially benign or unsatisfactory biopsy are unlikely to develop thyroid malignancy and highly unlikely to die of thyroid cancer. FNAB is highly accurate in detecting malignancy. Additional studies evaluating similar large data sets after the adoption of TBSRTC and the integration of molecular testing are needed.

Keywords: adequacy, Bethesda, cytopathology, fine-needle aspiration biopsy, thyroid

Introduction

APPROXIMATELY 7–15% of thyroid nodules are malignant, of which more than 90% are papillary or follicular carcinomas (1). In the United States, there were an estimated 52,070 new cases of thyroid cancer in 2019 (2). Fine needle aspiration biopsy (FNAB) is a well-established and cost-effective diagnostic tool for evaluation of thyroid nodules

and is endorsed by the American Thyroid Association (1). The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) provides a standardized category-based reporting system that allows cytopathologists to communicate defined diagnoses concisely. It also provides estimated risk of malignancy (ROM) and recommendations on clinical management for each category. The TBSRTC was recently updated to incorporate new literature regarding ROM, the rapidly

¹Department of Pathology, University of California, San Francisco, San Francisco, California, USA.

²Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, California, USA.

³Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, New York, USA.

Departments of ⁴Epidemiology and Biostatistics, and ⁵Neurology, University of California, San Francisco, San Francisco, California, USA.

expanding role of molecular testing, and the reclassification of encapsulated follicular variant of papillary thyroid carcinoma to noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP) (3,4). The estimated ROM in TBSRTC is based on publications evaluating the performance of thyroid FNAB; these studies largely rely on the subset of patients with corresponding cytologic and histologic specimens, and exclude patients who underwent clinical follow-up. This selection bias results in the overestimation of the ROM, because nodules that are resected often have concerning radiographic and/or clinical features, abnormal repeat FNAB results, or abnormal molecular test results. Moreover, there is marked variability in how researchers have approached the statistical analysis of thyroid FNAB, precluding meaningful interinstitutional comparisons, and resulting in a wide range of ROMs (5–11). For example, studies assessing the performance of thyroid FNAB and determining ROM have inconsistently considered indeterminate categories negative or positive test results, or excluded them from analysis altogether (12). Studies have also applied sensitivity, specificity, and diagnostic accuracy calculations to indeterminate TBSRTC categories, when these metrics are only applicable to binary tests that produce definitive results (5,12). Given that published ROM data inform clinical decision-making, there is an urgent need for studies with long-term clinical follow-up and robust statistical methodology to more accurately determine the performance of thyroid FNAB.

To our knowledge, our study represents the first to evaluate a large cohort of patients who underwent thyroid FNAB with long-term clinical follow-up and utilizing a statewide population-based cancer registry. We review the experience of thyroid FNAB at the University of California, San Francisco (UCSF), and investigate the validity of the TBSRTC and the recently revised ROMs therein.

Materials and Methods

A retrospective review of the UCSF Pathology database was performed for thyroid FNABs from January 1, 1997, to

December 31, 2004. During this period, most palpable nodules were sampled by pathologists, with a subset performed by endocrinologists and surgeons; all ultrasound-guided FNABs were performed by radiologists. The FNABs performed by both pathologists and radiologists all underwent rapid on-site evaluation. Because these FNABs predate TBSRTC, diagnoses were recoded using the 2017 TBSRTC diagnostic categories for the purpose of this study (3,4). Due to the absence of ultrasonographic imaging data, specimens that consisted of cyst fluid only were considered nondiagnostic. Patients who underwent thyroid FNAB at UCSF during the study time frame were matched to both the UCSF cancer registry and California Cancer Registry (CCR). CCR is a statewide population-based cancer registry administered by the California Department of Public Health. Matching allowed identification of patients diagnosed with thyroid malignancy during the follow-up period. The patients were divided into two categories: (a) absent from both UCSF and CCR records, or (b) present in the UCSF cancer registry and/or CCR. Patients absent from both CCR and UCSF records were assumed to be free of thyroid malignancy. Patients with a prior diagnosis of thyroid malignancy, prior thyroid surgery, or who developed a thyroid malignancy contralateral to the site of biopsy were excluded. Only each patient’s initial thyroid FNAB was included in the analysis. If patients had multiple nodules biopsied during the same procedure yielding different diagnoses, the TBSRTC category with the highest associated ROM was used for calculating the incidence of malignancy during follow-up. Whether or not a patient underwent repeat FNABs and the time interval to repeat biopsy was documented. The follow-up interval was defined from the time of FNAB to July 10, 2015, when the CCR match was performed, and which represents the date of the most recent clinical follow-up. The data were analyzed using Kaplan–Meier techniques, and stratified by TBSRTC diagnostic category. The date of the FNAB was considered to be time origin, and the patients were followed until they experienced an event (defined as the diagnosis of a thyroid malignancy); patients who did not develop a thyroid malignancy were censored as of July 10, 2015. Kaplan–Meier

TABLE 1. DISTRIBUTION OF CYTOLOGIC DIAGNOSES FOR INITIAL FINE NEEDLE ASPIRATION BIOPSY AND “RISK OF MALIGNANCY”

<i>Cytologic diagnosis</i>	<i>No. of specimens</i>	<i>Distribution of diagnoses (%)</i>	<i>Patients with cancer</i>	<i>“ROM” (%)</i>
Nondiagnostic	236	10.7*	16	6.78
Benign	1575	71.4	51	3.23**
AUS/FLUS	135	6.12	33	24.4
AUS	57	2.58	19	33.0
FLUS	78	3.53	14	17.9
Follicular neoplasm	107	4.85	31	29.0
SUS	20	0.91	14	70.0
Malignant	134	6.07	133	99.2
Total	2207	100	279	100

ROM is defined as number of thyroid cancers divided by the number of patients receiving each diagnosis over the follow-up period (mean = 13.9 years).

*The percentage of cases assigned the nondiagnostic category includes cyst fluid without thyroid epithelium. If these cases are considered benign, then the percentage is 7.6%.

**Of note, the ROM calculation includes any cancer occurring in the ipsilateral lobe and includes papillary microcarcinoma. If papillary microcarcinoma is removed from this calculation, the ROM for both nondiagnostic and benign categories is significantly lower (i.e., 1.4% for the benign category).

AUS, atypia of undetermined significance; FLUS, follicular lesion of undetermined significance; ROM, risk of malignancy; SUS, suspicious for malignancy.

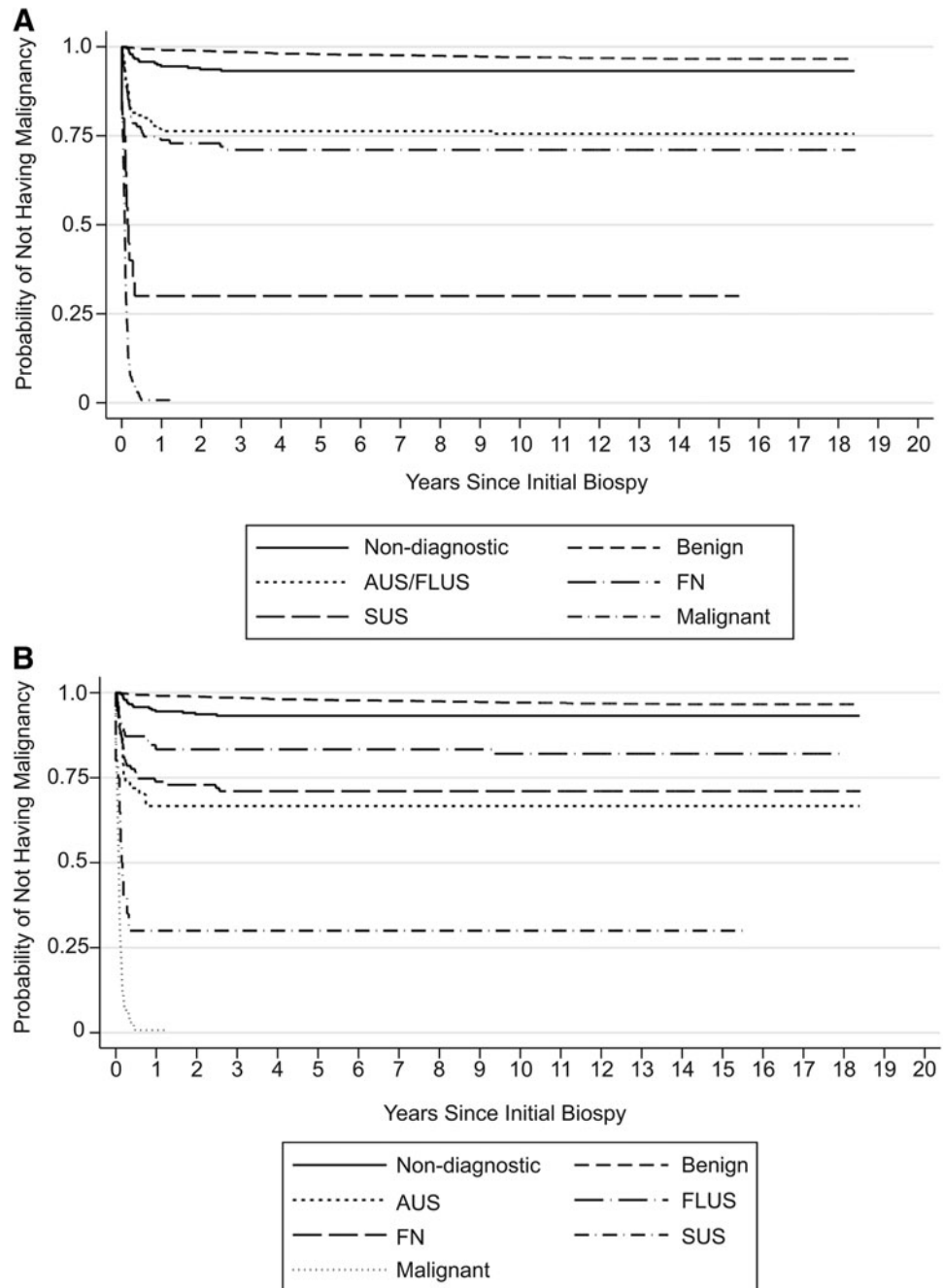
curves were used to estimate the incidence rates of malignancy, stratified by FNAB category. Cox proportional hazards models were used to determine the instantaneous ROM given: having a repeat biopsy, having a repeat biopsy within different intervals, having multiple biopsies, and having a given diagnosis (with a benign diagnosis as the reference).

A false-negative diagnosis was defined as an FNAB interpreted as benign with subsequent identification of malignancy in the ipsilateral thyroid on histopathologic evaluation. A false-positive diagnosis was defined as an FNAB interpreted as malignant, resected, and found to be benign on histopathology. This study received Institutional Review Board approval by the UCSF Committee for Human Research (10-04747).

Results

A total of 2233 patients were identified for whom 2758 FNAB reports were available; 26 patients were excluded. The first FNAB from each of the remaining 2207 patients [median age: 48 years old, range 7–92 years; 1880 (85.2%) females and 327 (14.8%) males] was included in the analysis. The median follow-up period after the initial thyroid FNAB was 13.9 years (range: 10.5–18.4 years). The distribution of TBSRTC diagnostic categories was 236 (10.7%) nondiagnostic, 1575 (71.4%) benign, 57 (2.58%) atypia of undetermined significance (AUS), 78 (3.53%) follicular lesion of undetermined significance (FLUS), 107 (4.85%) follicular neoplasm or Hurthle cell neoplasm (FN),

FIG. 1. Estimated probability of diagnosis of thyroid malignancy for each diagnostic category. The date of the initial FNAB was considered to be time origin, and the patients were followed until they experienced an event (defined as the histologic diagnosis of a thyroid malignancy); mean follow-up period was 13.9 years. For (A), the Kaplan–Meier curves are categorized according to TBSRTC, while (B) separates AUS and FLUS into two distinct categories. AUS, atypia of undetermined significance; FLUS, follicular lesion of undetermined significance; FN, follicular neoplasm; FNAB, fine needle aspiration biopsy; SUS, suspicious for malignancy; TBSRTC, The Bethesda System for Reporting Thyroid Cytopathology.



20 (0.9%) suspicious for malignancy (SUS), and 134 (6.07%) malignant (Table 1). During follow-up, 279 patients were diagnosed with thyroid malignancy. Using the conventional method of calculating the ROM for thyroid FNAB, dividing the number of thyroid cancers by the total number of patients within a category, our data are concordant with the 2017 TBSRTC ROM (Table 1).

Using time from first biopsy to diagnosis of thyroid malignancy as the primary endpoint, estimates derived from Kaplan–Meier curves demonstrated that the risk of being diagnosed with a thyroid malignancy was low for nondiagnostic and benign FNAB categories, intermediate for AUS/FLUS and FN, and high for SUS and malignant biopsy results (Fig. 1A). When AUS and FLUS were analyzed as distinct groups, the Kaplan–Meier curves for both groups demonstrated interme-

diante risk of being diagnosed with a thyroid malignancy. Of note, the risk for AUS was similar to FN, while FLUS was in between AUS and benign (Fig. 1B). The distribution of risk was similar when analyses were repeated excluding patients who had repeat biopsies (Fig. 2A, B). Among the 262 patients who had repeat biopsies, the interval from first biopsy to diagnosis of malignancy was 40 days to 12.4 years. There were 10 patients with initially nondiagnostic FNA biopsies, 11 patients with benign FNA biopsies, and 2 patients with AUS FNA biopsies, who were ultimately diagnosed with thyroid malignancy. Of the 23 cases, on repeat, 2 were categorized to AUS, 2 to FN, and 9 to malignant from benign, one to FLUS from nondiagnostic, and one to papillary carcinoma from AUS, with the remainder unchanged. In comparison with the benign FNAB category, all other diagnostic categories were associated with significantly

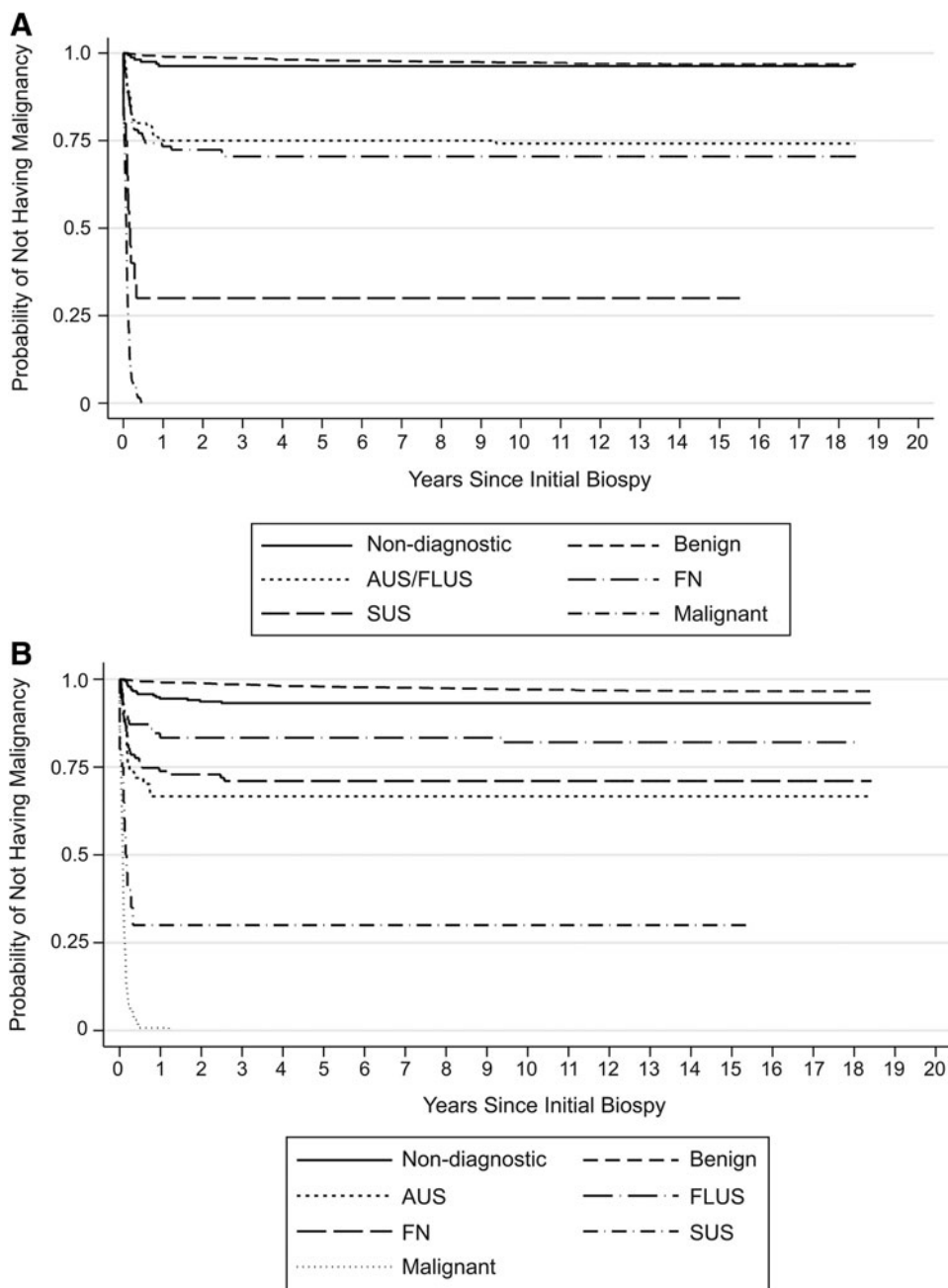


FIG. 2. Estimated probability of diagnosis of thyroid malignancy for each diagnostic category, excluding patients with repeat biopsies. The date of the initial FNAB was considered to be time origin, and the patients were followed until they experienced an event (defined as the histologic diagnosis of a thyroid malignancy); mean follow-up period was 13.9 years. For (A), the Kaplan–Meier curves are categorized according to TBSRTC, while (B) separates AUS and FLUS into two distinct categories.

higher risk of being diagnosed with a thyroid cancer during the follow-up period, with hazard ratios ranging from 2.09 (95% confidence interval [CI] 1.19–3.67, $p=0.01$) for nondiagnostic to 201 ([CI 138–293], $p\leq 0.001$) for malignant categories (Table 2). For patients with benign FNAB diagnoses, the estimated rate of malignancy was 2.42/1000 person-years (py) [CI 1.85–3.18]; only 3.2% of this group received a diagnosis of thyroid malignancy, of which 1.8% were microcarcinomas, during the entire time of observation; the median time to diagnosis was 3.44 years (Fig. 3). In comparison, all other diagnostic categories had a higher estimated rate of malignancy, including patients with nondiagnostic FNAB with a rate of 4.83/1000 py [CI 2.96–7.88] and median time to thyroid malignancy diagnosis of 4.4 months. AUS, FLUS, combined AUS/FLUS, and FN categories had similar and intermediate rates of malignancy at 34.5, 15.2, 22.4, and 29.1/1000 py, respectively, and the median time to thyroid malignancy was also similar at 2.17, 1.45, 2.1, and 1.94 months, respectively. Much higher incidence rates were seen for the SUS (183/1000 py) and malignant (980/1000 py) categories with median times to malignant diagnosis at 1.26 and 0.92 months, respectively. Patients who had multiple thyroid nodules biopsied during the first visit did not have an increased risk of being diagnosed with thyroid cancer (hazard ratio=0.88 [CI 0.55–1.42], $p=0.60$). Patients who had at least one repeat biopsy had a marginally reduced risk of being diagnosed with thyroid cancer (hazard ratio=0.63 [CI 0.41–0.97], $p=0.03$), while having a repeat biopsy within 6 months of the first FNAB suggests a minimally increased risk of thyroid cancer, but it was not statistically significant (hazard ratio=1.10 [CI 0.63–1.9], $p=0.73$).

Of the 134 FNABs diagnosed as malignant, 116 were papillary carcinomas (of which 15 measured <1 cm), 6 were medullary carcinomas, 2 were anaplastic carcinomas, 2 poorly differentiated carcinomas, 3 were lymphomas, 4 were metastatic cancers, and 1 was lost to follow-up. The distribution of malignant cases based on surgical resection or other clinical follow-up was 218 papillary carcinomas (of which 70 measured <1 cm), 27 follicular carcinomas, 6 Hurthle cell carcinomas, 11 medullary carcinomas, 4 poorly differentiated carcinomas, 2 anaplastic carcinomas, 4 lymphomas, and 4 metastatic cancers. Of the 20 FNABs diagnosed as SUS, 14 were confirmed to be malignant, with a distribution of 12 papillary carcinomas

(of which 3 were <1 cm) and 2 medullary carcinomas. The remaining 6 cases were composed of 2 follicular adenomas, and 4 received no follow-up confirmatory procedures.

One of the 134 malignant cases (0.75%) was not in the cancer registry, data suggesting a false-positive diagnosis. However, this patient was a visitor from outside the United States, received no health care at UCSF other than the FNAB visit, and was likely lost to follow-up. A total of 51 false-negative cases (3.2%) were identified: 28 papillary microcarcinomas (<1 cm), 11 follicular carcinomas (≥ 1 cm), 9 papillary carcinomas (≥ 1 cm), 1 medullary carcinoma (<1 cm), and 2 papillary carcinomas of unknown size. Excluding the papillary microcarcinomas, the false-negative rate was 1.5% (23/1575). The median time to thyroid cancer diagnosis for all false-negative FNABs was 3.44 years (0.04–13.6 years); no patients who had a false-negative diagnosis died of thyroid cancer during the follow-up period.

Overall, 15 patients in our cohort died of thyroid cancer during the follow-up period. One patient with a history of radiation exposure died of papillary thyroid carcinoma 16.8 years after an initial nondiagnostic biopsy. No patients who had a benign FNAB diagnosis died of thyroid cancer (Table 3). The remainder of the patients who died of thyroid cancer are detailed in Tables 3 and 4. For all patients, the overall cumulative hazard of dying 18 years after undergoing a thyroid FNAB was extremely low (Nelson–Aalen cumulative hazard ratio=0.01 [CI 0.005–0.02]), and was only marginally increased for patients who had a malignant FNAB (Nelson–Aalen cumulative hazard ratio=0.09 [CI 0.04–0.17])

Discussion

FNAB is the recommended first-line procedure for the diagnosis of thyroid nodules (1). However, to maintain thyroid FNAB as a clinically useful diagnostic tool, accuracy must be maintained. TBSRTC has successfully standardized diagnostic terminology and markedly improved communication between pathologists and referring physicians. However, the published ROM and recommendations for patient management are largely based on data sets including only those FNABs with corresponding histologic specimens. This selection bias leads to an overestimation of the ROM, particularly for low-risk diagnostic categories. Prior analyses of

TABLE 2. RELATIVE RISK OF DEVELOPING THYROID MALIGNANCY, ESTIMATED TIME TO DIAGNOSIS, AND ESTIMATED RATES

<i>Cytologic diagnosis</i>	<i>Category</i>	<i>Hazard ratio</i>	<i>[CI]</i>	<i>p</i>	<i>Estimated rate (per 1000 person-years)</i>	<i>[CI]</i>
Nondiagnostic	1	2.09	[1.19–3.67]	0.01	4.82	[2.96–7.88]
Benign	2	Reference group			2.42	[1.85–3.18]
AUS/FLUS	3	8.8	[5.69–13.6]	<0.002	22.4	[16–31.6]
AUS	3A	13.0	[7.69–22.0]	<0.001	34.5	[22.0–54.1]
FLUS	3F	6.12	[3.39–11.04]	<0.001	15.2	[9.0–25.7]
Follicular neoplasm	4	10.9	[7.0–17.0]	<0.003	29.1	[20.4–41.3]
SUS	5	49.1	[27.1–88.9]	<0.004	183	[108–310]
Malignant	6	201	[138–293]	<0.005	980	

The hazard ratio is the relative risk of developing thyroid cancer when compared with the reference group (benign diagnosis). All diagnostic categories show a significantly higher risk of cancer when compared with the benign category. The estimated rate of thyroid cancer diagnosis is given for each diagnostic category. Of note, these data include all thyroid malignancies occurring in the ipsilateral lobe over the follow-up period.

CI, 95% confidence interval.

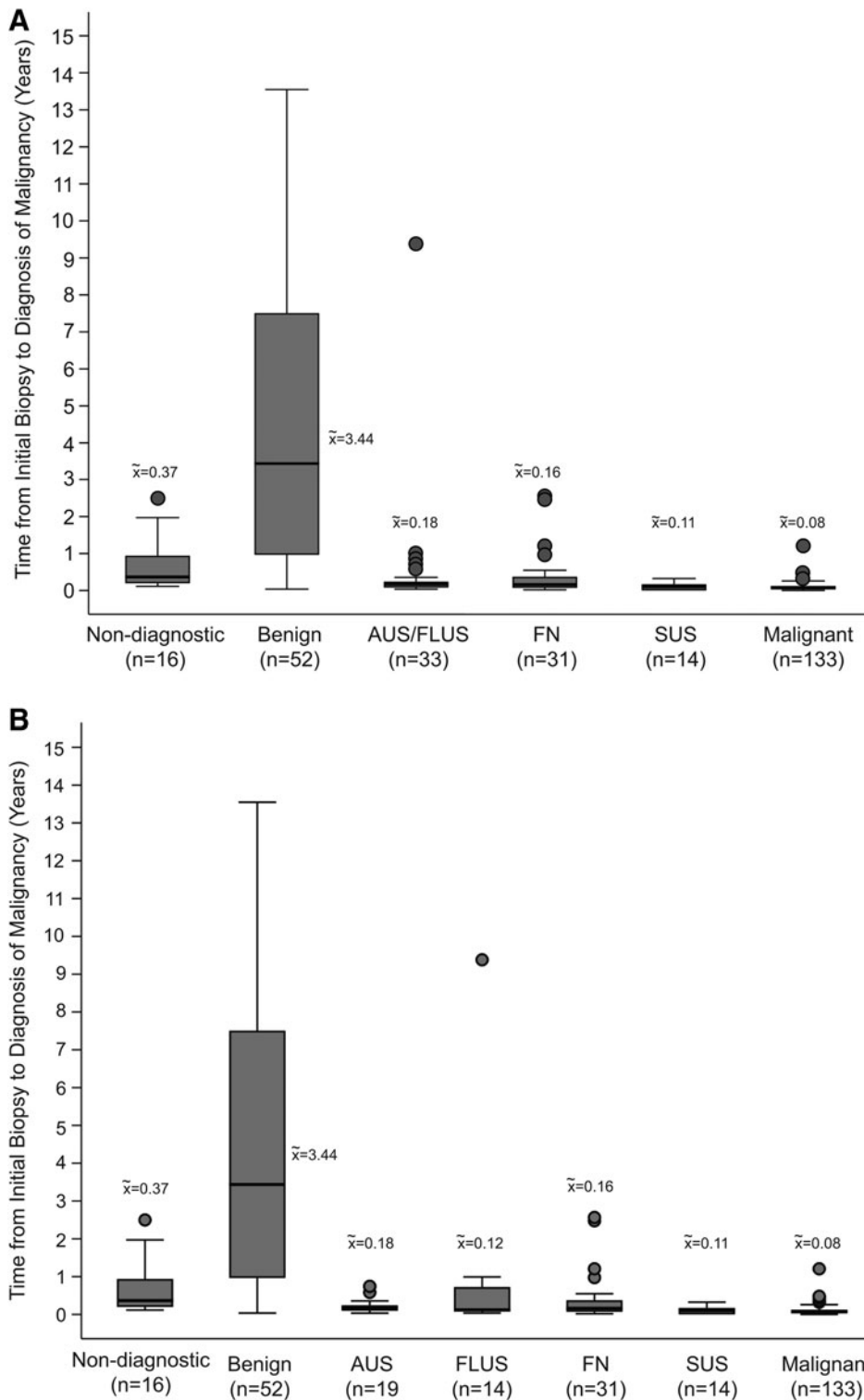


FIG. 3. Time from initial biopsy to diagnosis of malignancy (years). The date of the initial FNAB was considered to be time origin, and the patients were followed until they experienced an event (defined as the histologic diagnosis of a thyroid malignancy); mean follow-up period was 13.9 years. The box outlines denote the IQR, the solid line within the box denotes the median, the whiskers represent upper adjacent (75th percentile +1.5*IQR) and lower adjacent (25th percentile - 1.5*IQR) values, and the dots represent outliers. IQR, interquartile range.

thyroid FNABs also vary greatly in statistical methods. One study found four different approaches to indeterminate FNAB results (defined as any result that is not definitively benign, malignant, or nondiagnostic) when thyroid statistics was calculated; indeterminate FNABs were variously considered positive, negative, excluded from analysis, or considered a positive result when calculating sensitivity and negative when calculating specificity (12).

Our data demonstrate a low false-negative rate of 1.5% when papillary microcarcinomas are excluded, which is commensurate with the predicted, and lower than the observed, ROM in published literature (4,5,7). Of note, our methods classify any cancer arising in the ipsilateral lobe to represent a false-negative biopsy; it is likely that many of the small cancers identified histologically did not represent the target of the preceding FNAB. Importantly, no patients with a

TABLE 3. FREQUENCY OF DEATH ATTRIBUTABLE TO THYROID MALIGNANCY AND TIME TO DEATH

<i>Cytologic diagnosis</i>	<i>Total deaths attributable to thyroid malignancy</i>	<i>Median time to death (years, range)</i>
Nondiagnostic	1	16.8
Benign	0	NA
AUS/FLUS	2	2.43 (1.29–3.58)
AUS	0	n/a
FLUS	2	2.43 (1.29–3.58)
Follicular neoplasm	2	9.25 (5.2–13.3)
SUS	1	9.01
Malignant	9	2.41 (0.01–15.02)
Total	15	

A total of 15 deaths attributable to thyroid cancer were recorded over the follow-up period. One patient who had an initially nondiagnostic biopsy succumbed to thyroid cancer. No deaths attributable to thyroid cancer were recorded for those patients who received a benign cytologic diagnosis.

n/a, not available.

benign FNAB died of thyroid cancer during the follow-up period, suggesting that a benign cytologic diagnosis essentially excludes an aggressive thyroid malignancy. False-negative diagnosis can result in a delay in treatment. However, our data suggest that those thyroid cancers identified at a later date are low risk and effectively treated without increased mortality. Our data suggest that the value of repeat biopsies may be limited in the appropriate clinical setting. Of 262 patients who had repeat biopsies, malignancy was diagnosed in only a small proportion. More definitive conclusions regarding the value of repeat biopsy are difficult to make without considering other clinical factors, including size and sonographic findings. To date there is one retrospective study with long-term follow-up, which analyzed the 2010 benign thyroid FNAB from 1369 patients with a mean follow-up of 8.5 years. Only 18 false-negative FNABs were identified in this cohort and, similar to our data set, none of the deaths that occurred was attributable to thyroid cancer (13). Unfortunately, other diagnostic categories were not analyzed in this retrospective study.

Although many studies and the most recent edition of TBSRTC suggest that the ROM for the nondiagnostic FNAB category is 5–10% higher compared with the benign category, our data suggest that while there is an increased ROM in

the short term, long-term outcomes for both nondiagnostic and benign categories are fairly similar. Of note, one high-risk patient with initial nondiagnostic FNAB in our cohort died of thyroid cancer. Therefore, an unsatisfactory diagnosis may still warrant repeat biopsy and/or close follow-up in certain clinical settings. There was one potential false-positive diagnosis in our cohort, but this patient was lost to follow-up and likely received care outside of California.

In other large studies, including one meta-analysis, the percentage of thyroid FNABs categorized as benign ranges from 21% to 89%, while the nondiagnostic category ranges from 2% to 43%. Indeterminate diagnoses, which encompass all diagnoses that were not definitely benign or malignant, range from 6% to 55%, and malignant from 3% to 32% (3,11–14). At our institution, the proportion of definitive diagnosis (either benign or malignant) makes up a majority of diagnoses rendered (77.5%), whereas equivocal diagnoses (i.e., AUS/FLUS, FN, and SUS) represent only 11.9% of diagnoses, suggesting that the proportion of specimens where adjunctive molecular testing is of value is relatively low in our practice. The long-term probability of developing malignancy for AUS/FLUS and FN categories was similar at our institution, raising the possibility of a five-tier diagnostic system, rather than the current six-tier system. Another suggestion might be to create distinct categories for AUS and FLUS, since our data and several previous studies have shown that FNABs diagnosed as AUS have an increased probability of developing malignancy over those diagnosed as FLUS (15–20). Nondiagnostic specimens represented 10.7% of FNABs, which is similar to published data from other large academic institutions, which range from 2% to 23.6%. We considered the implications of categorizing the 69 FNABs containing cyst fluid only as benign and found that 2 (2.9%) of these patients were later found to have a thyroid malignancy. The cysts measured 2 and 4 cm, respectively. Therefore, if cyst fluid cases are considered benign, the nondiagnostic rate falls to 7.6%, but the false-negative rate increases. Therefore, we adhered to TBSRTC criteria and characterized the cyst fluid-only cases as nondiagnostic, allowing for clinical correlation with imaging studies. Numerous factors can influence adequacy rates, including the type of clinician performing the biopsy, training and experience, and the use of ultrasound-guidance and rapid on-site evaluation (14,21–24). Our study period predates publication of TBSRTC and formal specimen adequacy criteria, and before widespread adoption of ultrasound guidance for thyroid FNAB by physicians other than radiologists (25). The

TABLE 4. RATE OF OVERALL SURVIVAL AND PROBABILITY THAT PATIENTS WOULD REMAIN FREE OF THYROID MALIGNANCY UP TO 15 YEARS BY DIAGNOSTIC CATEGORY

<i>Cytologic diagnosis</i>	<i>Patients (n)</i>	<i>Free of thyroid malignancy (%)</i>				<i>Overall survival (%)</i>	
		<i>1 Year</i>	<i>5 Years</i>	<i>10 Years</i>	<i>15 Years</i>	<i>Still in study</i>	<i>15 Years</i>
Benign	1575	99.1±0.2	97.9±0.4	97.1±0.4	96.6±0.4	512	1.00
Nondiagnostic	236	94.5±1.5	93.2±1.6	93.2±1.6	93.2±1.6	126	1.00
AUS/FLUS	135	76.3±3.7	76.3±3.7	75.6±3.7	75.6±3.7	38	98.5±1.0
AUS	57	66.7±6.2	66.7±6.2	66.7±6.2	66.7±6.2	21	1.00
FLUS	78	83.3±4.2	83.3±4.2	83.3±4.2	82.1±4.4	17	97.4±1.8
Follicular neoplasm	107	73.8±4.2	71.0±4.4	71.0±4.4	71.0±4.4	34	97.3±1.9
SUS	20	30.0±10.2	30.0±10.2	30.0±10.2	30.0±10.2	10	95.0±4.9
Malignant	134	0.7±0.07	n/a	n/a	n/a	41	93.1±2.2

nondiagnostic biopsies in this data set were primarily submitted by a small number of clinicians lacking formal training or experience in performing FNAB on palpable lesions and without the benefit of rapid on-site evaluation. In addition, we recognize that academic and large referral centers may receive a higher proportion of cases that may be more challenging to aspirate (23).

Commentaries by Sebo (26) and Ljung *et al.* (27) identify several factors necessary to maintain the quality of FNAB specimens and to optimize interpretation: (a) maintaining strong communication and multidisciplinary cooperation between referring providers, pathologists, and radiologists; (b) minimizing indeterminate diagnoses thereby maximizing specificity; (c) intradepartmental consultation among pathologists to reduce interobserver variability; and (d) limiting FNAB procurement to skilled operators to reduce problems related to poor sampling technique. Overall, the vast majority of patients at our institution received a definitive diagnosis, which provided clear clinical guidance and reduced the need for repeat biopsy, molecular testing, or diagnostic lobectomy.

Although in other settings, the introduction of NIFTP has significantly altered the rates of malignancy, a prior study by our group found that the vast majority of tumors diagnosed as NIFTP on histology were classified as FN or AUS/FLUS on preceding FNAB (28–32). None was diagnosed as SUS and only 1 of 22 NIFTPs was diagnosed as papillary thyroid carcinoma on cytology. In our experience using direct smears, cytologic evaluation can reliably triage NIFTP for conservative management and accurately diagnose classic papillary thyroid carcinoma (and often infiltrative follicular variant of papillary carcinoma) as SUS or malignant (32). Given that the vast majority of NIFTPs were appropriately classified as AUS/FLUS or FN on FNAB, the adoption of the NIFTP diagnostic category will likely have little impact on the rates of malignancy in our practice.

There are a few limitations in our study. This study was retrospective in design and focused on thyroid FNAB performed at a single institution. Although using a statewide population-based cancer registry minimizes the number of patients who are lost to follow-up, we were unable to determine whether individuals not represented in the registry had moved outside of California and received health care out of state. Therefore, there may be some underestimation of the ROM. While we were able to evaluate the development of thyroid malignancy and related mortality from the CCR, detailed diagnostic and clinical data were not available for patients who were not treated at UCSF. Our practice also depends primarily on direct smears, with ThinPrep as an adjunct preparation in a minority of cases. Therefore, our findings may not be applicable to settings that rely on liquid-based cytology (LBC) preparations. The performance of direct smear preparations may explain the difference in the distribution of diagnoses and the high percentage of FNABs with definitive diagnosis at our institution (77.5%) when compared with those where LBC is used exclusively (60–65%) (9,33,34). Studies have shown conflicting results when conventional smears and LBC are compared, with some studies showing increased rates of malignant, nondiagnostic, and/or benign diagnoses with LBC, while others showed decreased rates (33,35–38).

In conclusion, our study is the first to assess the long-term ROM after thyroid FNAB, including mortality data. Most of

the false-negative cancers were papillary microcarcinomas, followed by follicular carcinomas. There were no deaths attributable to thyroid cancer for patients who had a benign diagnosis on FNAB, with a median of 10.5 years of follow-up. Among patients with nondiagnostic specimens, one high-risk patient died more than 15 years after the initial biopsy. Thus, asymptomatic patients without worrisome clinical and radiographic findings and an initially benign or unsatisfactory biopsy are highly unlikely to develop thyroid malignancy and even less likely to die of thyroid cancer. In our practice, FNAB is highly accurate in detecting not only papillary carcinoma but also malignancy overall. We recognize that additional large studies with long-term follow-up and rigorous statistical methodology are needed since the impacts of TBSRTC, increasing use of ultrasound, and molecular testing are unknown.

Authors' Contributions

D.L.N. collected and analyzed the data, designed the figures, and wrote and edited the article. A.v.Z. designed and supervised the study, analyzed the data, and wrote and edited the article. A.G. coordinated the study, collected and analyzed the data, and performed the cancer registry matches. N.K.H. performed the biostatistical analysis and review. B.M.L. conceived, designed, and supervised the study, analyzed the data, and wrote and edited the article. All authors read the article and agreed to its contents.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

No funding was received for this article.

References

1. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, *et al.* 2016 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* **26**:1–133.
2. Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds) SEER Cancer Statistics Review, 1975–2016, Bethesda, MD, based on November 2018 SEER data submission, posted to the SEER web site, April 2019. Available at https://seer.cancer.gov/csr/1975_2016 (accessed November 20, 2020).
3. Cibas ES, Ali SZ, NCI Thyroid FNA State of the Science Conference 2009 The Bethesda System For Reporting Thyroid Cytopathology. *Am J Clin Pathol* **132**:658–665.
4. Cibas ES, Ali SZ 2017 The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid* **27**:1341–1346.
5. Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW 2012 The Bethesda system for reporting thyroid cytopathology: a meta-analysis. *Acta Cytol* **56**:333–339.
6. Yang J, Schnadig V, Logrono R, Wasserman PG 2007 Fine-needle aspiration of thyroid nodules: a study of 4703 patients with histologic and clinical correlations. *Cancer* **111**:306–315.

7. Renshaw A 2010 An estimate of risk of malignancy for a benign diagnosis in thyroid fine-needle aspirates. *Cancer Cytopathol* **118**:190–195.
8. Oertel YC, Miyahara-Felipe L, Mendoza MG, Yu K 2007 Value of repeated fine needle aspirations of the thyroid: an analysis of over ten thousand FNAs. *Thyroid* **17**:1061–1066.
9. Yassa L, Cibas ES, Benson CB, Frates MC, Doubilet PM, Gawande AA, *et al.* 2007 Long-term assessment of a multidisciplinary approach to thyroid nodule diagnostic evaluation. *Cancer* **111**:508–516.
10. Jo VY, Stelow EB, Dustin SM, Hanley KZ 2010 Malignancy risk for fine-needle aspiration of thyroid lesions according to the Bethesda System for Reporting Thyroid Cytopathology. *Am J Clin Pathol* **134**:450–456.
11. Iskandar ME, Bonomo G, Avadhani V, Persky M, Lucido D, Wang B, *et al.* 2015 Evidence for overestimation of the prevalence of malignancy in indeterminate thyroid nodules classified as Bethesda category III. *Surgery* **157**:510–517.
12. Lewis CM, Chang KP, Pitman M, Faquin WC, Randolph GW 2009 Thyroid fine-needle aspiration biopsy: variability in reporting. *Thyroid* **19**:717–723.
13. Nou E, Kwong N, Alexander LK, Cibas ES, Marqusee E, Alexander EK 2014 Determination of the optimal time interval for repeat evaluation after a benign thyroid nodule aspiration. *J Clin Endocrinol Metab* **99**:510–516.
14. Redman R, Zalaznick H, Mazzaferri EL, Massoll NA 2006 The impact of assessing specimen adequacy and number of needle passes for fine-needle aspiration biopsy of thyroid nodules. *Thyroid* **16**:55–60.
15. Rosario PW, Calsolari MR 2017 Importance of cytological subclassification of thyroid nodules with Bethesda category III cytology (AUS/FLUS) into architectural atypia only and nuclear atypia: a prospective study. *Diagn Cytopathol* **45**:604–607.
16. Lim JXY, Nga ME, Chan DKH, Tan WB, Parameswaran R, Ngiam KY 2018 Subclassification of Bethesda atypical and follicular neoplasm categories according to nuclear and architectural atypia improves discrimination of thyroid malignancy risk. *Thyroid* **28**:511–521.
17. Kim SJ, Roh J, Baek JH, Hong SJ, Shong YK, Kim WB, *et al.* 2017 Risk of malignancy according to sub-classification of the atypia of undetermined significance or follicular lesion of undetermined significance (AUS/FLUS) category in the Bethesda system for reporting thyroid cytopathology. *Cytopathology* **28**:65–73.
18. Gan TRX, Nga ME, Lum JHY, Wong WM, Tan WB, Parameswaran R, *et al.* 2017. Thyroid cytology-nuclear versus architectural atypia within the “Atypia of undetermined significance/follicular lesion of undetermined significance” Bethesda category have significantly different rates of malignancy. *Cancer Cytopathol* **125**:245–256.
19. Shrestha RT, Hennessey JV 2016 Cytologic subclassification of atypia of undetermined significance may predict thyroid nodules more likely to be malignant at surgery. *Diagn Cytopathol* **44**:492–498.
20. Johnson DN, Cavallo AB, Uraizee I, Tanager K, Lastra RR, Antic T, *et al.* 2019 A proposal for separation of nuclear atypia and architectural atypia in Bethesda category III (AUS/FLUS) based on differing rates of thyroid malignancy. *Am J Clin Pathol* **151**:86–94.
21. Amrikachi M, Ramzy I, Rubinfeld S, Wheeler TM 2001 Accuracy of fine-needle aspiration of thyroid. *Arch Pathol Lab Med* **125**:484–488.
22. Burch HB, Burman KD, Reed HL, Buckner L, Raber T, Ownbey JL 1996 Fine needle aspiration of thyroid nodules. Determinants of insufficiency rate and malignancy yield at thyroidectomy. *Acta Cytol* **40**:1176–1183.
23. Renshaw AA 2011 Non-diagnostic rates for thyroid fine needle aspiration are negatively correlated with positive for malignancy rates. *Acta Cytol* **55**:38–41.
24. Ljung BM, Drejet A, Chiampi N, Jeffrey J, Goodson WH 3rd, Chew K, *et al.* 2001 Diagnostic accuracy of fine-needle aspiration biopsy is determined by physician training in sampling technique. *Cancer* **93**:263–268.
25. Wu M, Choi Y, Zhang Z, Si Q, Salem F, Szporn A, *et al.* 2016 Ultrasound guided FNA of thyroid performed by cytopathologists enhances Bethesda diagnostic value. *Diagn Cytopathol* **44**:787–791.
26. Sebo TJ 2012 What are the keys to successful thyroid FNA interpretation? *Clin Endocrinol (Oxf)* **77**:13–17.
27. Ljung B-M 2008 Thyroid fine-needle aspiration: smears versus liquid-based preparations. *Cancer* **114**:144–148.
28. Lau RP, Paulsen JD, Brandler TC, Liu CZ, Simsir A, Zhou F 2017 Impact of the reclassification of “noninvasive encapsulated follicular variant of papillary thyroid carcinoma” to “noninvasive follicular thyroid neoplasm with papillary-like nuclear features” on the Bethesda system for reporting thyroid cytopathology: a large academic institution’s experience. *Am J Clin Pathol* **149**:50–54.
29. Strickland KC, Howitt BE, Marqusee E, Alexander EK, Cibas ES, Krane JF, *et al.* 2015 The impact of noninvasive follicular variant of papillary thyroid carcinoma on rates of malignancy for fine-needle aspiration diagnostic categories. *Thyroid* **25**:987–992.
30. Faquin WC, Wong LQ, Afrogheh AH, Ali SZ, Bishop JA, Bongiovanni M, *et al.* 2016 Impact of reclassifying non-invasive follicular variant of papillary thyroid carcinoma on the risk of malignancy in The Bethesda System for Reporting Thyroid Cytopathology. *Cancer Cytopathol* **124**:181–187.
31. Zhou H, Baloch ZW, Nayar R, Bizzarro T, Fadda G, Adhikari-Guragain D, *et al.* 2018 Noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP): implications for the risk of malignancy (ROM) in the Bethesda System for Reporting Thyroid Cytopathology (TBSRTC). *Cancer Cytopathol* **126**:20–26.
32. Ng DL, Can NT, Ma ZV, van Zante A, Ljung BM, Khanafshar E 2017 Cytomorphologic features of noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP): a comparison with infiltrative follicular variant of papillary thyroid carcinoma. *J Bas Clin Med* **1**:51.
33. Siddiqui MA, Griffith KA, Michael CW, Pu RT 2008 Nodule heterogeneity as shown by size differences between the targeted nodule and the tumor in thyroidectomy specimen: a cause for a false-negative diagnosis of papillary thyroid carcinoma on fine-needle aspiration. *Cancer* **114**:27–33.
34. VanderLaan PA, Marqusee E, Krane JF 2011 Clinical outcome for atypia of undetermined significance in thyroid fine-needle aspirations: should repeated FNA be the preferred initial approach? *Am J Clin Pathol* **135**:770–775.
35. Fischer AH, Clayton AC, Bentz JS, Wasserman PG, Henry MR, Souers RJ, *et al.* 2013 Performance differences between conventional smears and liquid-based preparations of thyroid fine-needle aspiration samples: analysis of 47,076

- responses in the College of American Pathologists Inter-laboratory Comparison Program in Non-Gynecologic Cytology. *Arch Pathol Lab Med* **137**:26–31.
36. Witt RL 2004 A comparison of the results of monolayer versus smear cytopreparatory techniques for fine-needle aspiration in 100 consecutive patients undergoing thyroidectomy: a surgeon's perspective. *Otolaryngol Head Neck Surg* **131**:964–967.
 37. Frost AR, Sidawy MK, Ferfelli M, Tabbara SO, Bronner NA, Brosky KR, *et al.* 1998 Utility of thin-layer preparations in thyroid fine-needle aspiration: diagnostic accuracy, cytomorphology, and optimal sample preparation. *Cancer* **84**:17–25.
 38. Michael CW, Hunter B 2000 Interpretation of fine-needle aspirates processed by the ThinPrep technique: cytologic artifacts and diagnostic pitfalls. *Diagn Cytopathol* **23**:6–13.

Address correspondence to:

Dianna L. Ng, MD
Department of Pathology
Memorial Sloan Kettering Cancer Center
1275 York Avenue
New York, NY 10065
USA

E-mail: ngd2@mskcc.org