

Thyroid Fine-Needle Aspiration

Does Case Volume Affect Diagnostic Yield and Interpretation?

Jeffrey J. Houlton, MD; Gordon H. Sun, MD; Nella Fernandez, MD; Qihui "Jim" Zhai, MD; Fred Lucas, MD; David L. Steward, MD

Objective: To evaluate the effect of case volume on the diagnostic yield and interpretation of thyroid fine-needle aspiration (FNA).

Design: Retrospective case series.

Setting: An academic tertiary referral center and 2 community hospital centers.

Patients: Data were retrospectively reviewed for all consecutive patients undergoing thyroid FNA at these institutions during the 2009 calendar year.

Main Outcome Measures: Differences in diagnostic distribution and yield among pathologists and clinicians of differing case volume.

Results: A total of 790 patients underwent thyroid FNA, with the results interpreted as benign (479 [60%]), atypical (166 [22%]), malignant (9 [1%]), or nondiagnostic (136 [17%]). The FNAs were performed by 134 physicians and interpreted by 16 pathologists with varying case volumes. Low-volume pathologists (<50 FNAs inter-

preted) were more likely to report atypical FNAs (32% vs 13%; $P < .001$) and less likely to call FNAs benign (50% vs 70%; $P < .001$) compared with high-volume pathologists (≥ 50 FNAs interpreted), and compared with expected normative data (benign, $P < .001$; atypical, $P < .001$). Atypical FNA findings reported by low-volume pathologist were more likely to yield benign permanent results than those read by high-volume pathologists (64% vs 42%; $P < .02$). Low-volume clinicians (< 20 FNAs performed) were not more likely to perform nondiagnostic FNAs compared with high-volume clinicians (≥ 20 FNAs performed) (16% vs 15%; $P = .47$).

Conclusions: Case volume significantly influences the pathologic interpretation of thyroid FNA, as low-volume pathologists report more atypical and fewer benign FNA results. Case volume did not have a significant impact on diagnostic yield, because thyroid FNAs performed by low-volume clinicians did not result in more frequent nondiagnostic results compared with those performed by high-volume clinicians.

Arch Otolaryngol Head Neck Surg. 2011;137(11):1136-1139

THYROID FINE-NEEDLE ASPIRATION (FNA) remains a cornerstone in the diagnostic evaluation of thyroid nodules. Thyroid FNA is generally reported to yield sensitive and specific results for the detection of thyroid carcinoma, reducing the rate of thyroidectomy, and increasing the yield of cancer on permanent pathologic results.¹⁻⁵ Fine-needle as-

However, the accuracy of thyroid FNA interpretation is somewhat variable among differing series in the literature. In a recent report by Lewis et al,¹¹ which included a review of 19 large prospective series, highly varied sensitivities (38%-95%), specificities (46%-99%), and positive (28%-99%) and negative (66%-99%) predictive values were reported. The categories used to report results were also highly variable among differing series, varying from 3 to 8 category systems.¹¹ Even pathologists of the same institution are reported to interpret thyroid FNA results with significant interobserver variability, particularly for atypical lesions.¹²⁻¹⁴

This diagnostic variability is problematic as thyroid FNA biopsy is heavily relied on for decisions regarding the treatment of thyroid nodules. As such, the use

See Editor's Note at end of article

piration is recognized as the primary screening modality for thyroid carcinoma by the American Thyroid Association, National Cancer Institute, and the National Comprehensive Cancer Network.⁶⁻¹⁰

Author Affiliations:

Departments of Otolaryngology–Head and Neck Surgery (Drs Houlton, Sun, and Steward) and Pathology (Drs Fernandez, Zhai, and Lucas), University of Cincinnati Academic Health Center, Cincinnati, Ohio.

Table 1. Thyroid Fine-Needle Aspiration Results by Pathologist Case Volume

Source of Results	Diagnostic Category, %			
	Benign	Atypical	Malignant	Nondiagnostic
High-volume pathologists	68	13	1.50	17
Low-volume pathologists	50	32	0.60	17
Gharib and Papinil ⁴ and Gharib and Goellner ²⁰	70	10	5	15

of thyroid FNA as a screening modality necessitates that it be sufficiently sensitive to detect the low prevalence of malignancy among thyroid nodules (generally accepted as 5%-10%) while remaining specific enough to avoid a high rate of false-positive test results, which result in subsequent nontherapeutic thyroidectomy.¹⁵⁻¹⁷ Achieving a low incidence of nontherapeutic thyroidectomy is further challenged by the largely prevalent (20%-76% prevalence on thyroid ultrasonography) nature of thyroid nodules in the adult population.^{18,19}

In the current study, we examine the affect of differing case volume on the diagnostic yield and interpretation of thyroid FNA, in a regional setting, as we hypothesize that this variable may play a role in the accuracy of thyroid FNA results.

METHODS

Approval was granted by the institutional review board of the University of Cincinnati Academic Health Center, Cincinnati, Ohio. Data were retrospectively reviewed for all consecutive patients who underwent thyroid FNA during the 2009 calendar year at 3 separate hospital centers, in a midsize American city. These 3 institutions included (1) an academic health center (including inpatient biopsies and associated outpatient clinics), (2) a privately employed community hospital and its surrounding clinics, and (3) a second community hospital and surrounding clinics that employ 6 community pathologists and 3 pathologists who also read FNA results at the academic center (and were responsible for 27% of biopsies read at this institution). All eligible patient data and related FNA data were included in the study. Data were obtained by searching pathologic accession numbers at each individual institution. Relevant data included the FNA accession number, interpreting pathologist, clinician performing FNA, institution of interpretation, and diagnostic result. The FNA findings were uniformly reported as 1 of 4 diagnostic categories: benign, atypical (indeterminate), malignant, or nondiagnostic results, as was customary at all institutions.

When obtainable, permanent pathologic results were reviewed for patients with atypical and malignant FNA findings. Follow-up permanent pathologic results were available only if thyroidectomy took place at the same institution as the initial FNA, and the results were unobtainable if surgery took place at a separate institution (eg, a surgical center or separate hospital). As such, these follow-up data were used to extrapolate accuracy data and not to determine the proportion of patients who ultimately underwent surgery. Permanent pathologic results were categorized as either malignant, adenoma, or benign (nonadenoma). Differences in permanent pathologic results were compared using χ^2 analysis.

To analyze the effect of case volume on FNA interpretation, pathologists were divided into 2 subgroups: those who read less than 50 FNA slides in the year (termed low-volume pathologists) and those who read at least 50 FNA slides in the

year (termed high-volume pathologists). We used χ^2 analysis to identify significant differences between low-volume pathologists and high-volume pathologists in their reporting of each of the 4 diagnostic results (benign, atypical, malignant, or nondiagnostic). Differences in permanent pathologic results for patients with FNA findings initially reported as atypical were also compared between these subgroups.

The diagnostic results of the low- and high-volume FNA subgroups were also compared with existing "near-normative" data. In the absence of validated normative data we used publications by Gharib and Papinil⁴ and Gharib and Goellner²⁰ to approximate an expected distribution of diagnoses. These studies were selected because they are large, well-respected publications and because they also approximated ranges of expected FNA diagnoses as published in the American Association of Clinical Endocrinologists and Associazione Medici Endocrinologi Medical Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules.²¹ While this was recognized as an imperfect statistical comparison, we believe it was important to provide context for our results. Comparisons were made using χ^2 analysis.

To evaluate the affect of case volume on diagnostic yield, the rates of nondiagnostic results were analyzed with regard to the clinician performing the FNA biopsy. The rate of nondiagnostic results submitted by low-volume clinicians (defined as clinicians performing < 20 FNA biopsies per year) was compared with the rate of high-volume clinicians (defined as clinicians who had performed $\geq$ 20 FNA biopsies per year) using χ^2 analysis. We also compared differences in each of the other 3 diagnostic categories (benign, atypical, and malignant) between clinicians of high and low volume. Data pertaining to whether biopsies were palpation guided or ultrasonography guided were unobtainable. Statistical significance was defined as $P < .05$.

RESULTS

A total of 790 thyroid FNAs were interpreted at the 3 hospital centers in the 2009 calendar year: 384 biopsies were interpreted by 8 pathologists at the Academic Health Center, 292 biopsies were interpreted by 9 pathologists (including 3 Academic pathologists) at the second community hospital, and 115 biopsies were interpreted by 4 pathologists at the first community hospital. All FNA results were reported and totaled uniformly as 1 of 4 categories: benign (479 [60%]), atypical (166 [22%]), malignant (9 [1%]), or nondiagnostic (136 [17%]). In total, FNAs were performed by 134 physicians and interpreted by 16 pathologists.

The distribution of diagnoses in the low-volume pathologist subgroup, the high-volume pathologist subgroup, and the near-normative group (Gharib and Papinil⁴ and Gharib and Goellner²⁰) is displayed in **Table 1**. Low-volume pathologists were more likely to report atypi-

Table 2. Nondiagnostic Results by Clinician Case Volume

Source of Results	Clinicians, No.	FNAs	
		Performed, Mean, No.	Nondiagnostic Results, %
High-volume clinicians	9	45	16
Low-volume clinicians	125	3.1	15

Abbreviation: FNA, fine-needle aspiration.

cal FNAs (32% vs 13%; $P < .001$) and less likely to call FNAs benign (50% vs 68%; $P < .001$) compared with high-volume pathologists. There was no statistically significant difference in the reporting of nondiagnostic (17% vs 17%) or malignant (1% vs 1%) results. When compared with the near-normative data, low-volume pathologists again reported more atypical FNAs (32% vs 10%; $P < .001$) and fewer benign FNAs (50% vs 70%; $P < .001$) but also reported fewer malignant FNAs (1% vs 5%; $P = .002$). High-volume pathologists were not significantly different in their reporting of atypical (13% vs 10%) and benign (68% vs 70%) FNA results compared with the data reported by Gharib and Papinil⁴ and Gharib and Goellner,²⁰ but, like the low-volume pathologists, were less likely to report malignant FNAs (1% vs 5%; $P = .003$).

The 790 total FNA biopsies were performed by 134 clinicians, as displayed in **Table 2**. A total of 125 low-volume clinicians performed a mean of 3.1 biopsies in the year, totaling 382 FNAs. The high-volume clinicians performed a mean of 45 biopsies in the year, totaling 408 FNAs. Low-volume clinicians were not more likely to perform nondiagnostic FNA compared with high-volume clinicians (16% vs 15%; $P = .42$). There were also no significant differences in the other 3 diagnostic categories (atypical, $P = .58$; malignant, $P = .42$; benign, $P = .38$).

Of the patients whose FNA findings were reported as atypical, 55% (91 of 166) had permanent pathologic results available for review; this included 66% (31 of 47) at the Academic Health Center, 41% (16 of 39) at the first community hospital, and 55% (44 of 80) at the second community hospital. Of the patients whose FNA findings were reported as malignant, 44% (4 of 9) had permanent pathologic results available; this included 50% (3 of 6) at the Academic Health Center, 0% (0 of 2) at the first community hospital, and 100% (1 of 1) at the second community hospital. Atypical FNA findings read by low-volume pathologists were more likely to have benign permanent pathologic results than those read by high-volume pathologists (64% vs 42%; $P = .02$) (**Table 3**).

COMMENT

Thyroid FNA is an invaluable diagnostic tool, yet its use in the workup of thyroid nodules is not without limitation. In our series of almost 800 patients, these limitations are most clearly illustrated by the 39% rate of FNA findings read as either atypical (22%) or nondiagnostic (17%). These 2 categorical results fail to convey a definitive diagnosis and, as such, are typically frustrating for both surgeon and patient alike. In our study, we sought to analyze the affects of both pathologic case volume and

Table 3. Permanent Pathologic Results for Atypical Fine-Needle Aspiration Findings

Source of Results	Results, %		
	Benign	Adenoma	Malignant
High-volume pathologists	42	28	31
Low-volume pathologists	64	11	25

clinical case volume on the diagnostic yield and interpretation of thyroid FNA. In addition, we looked particularly closely at these correlations as related to atypical and nondiagnostic FNA results, given the clinical uncertainty these results present.

In our series, low-volume pathologists were found to report atypical FNA findings almost 2½ times as often as high-volume pathologists (32% vs 13%; $P < .001$). This increase in atypical FNA findings seemed to correlate with a significant decrease in the reporting of benign FNA findings (50% vs 68%; $P < .001$). A similar result was produced by our imperfect comparison with “near-normative” data reported by Gharib and Papinil⁴ and Gharib and Goellner,²⁰ because low-volume pathologists again reported atypical FNA findings more often, and benign less often, than this data set. However, high-volume pathologists did not.

One possible conclusion is that there is a significant proportion of FNA findings being reported as atypical by low-volume pathologists, which would otherwise be reported as benign by high-volume pathologists. This conclusion is supported by our examination of the available permanent pathologic results for atypical FNA findings in which there was a significantly higher rate of benign permanent pathologic results for atypical FNAs reported by low-volume pathologists compared with those reported by high-volume pathologists (64% vs 42%; $P = .02$). These follow-up data were somewhat limited because only 55% of permanent pathologic results were available, which is a significantly lower rate than thyroidectomy rates for atypical aspirates in other series.²² However, this result is likely the effect of patients undergoing thyroidectomy at institutions other than the 3 study locations, as is common in our region, rather than foregoing surgery all together. Nonetheless, these missing data are a limitation of the current study.

Several other limitations were presented by the retrospective and regional nature of our study design. Demographic data were not obtainable and therefore not controlled for in the study groups, a particular limitation because the study included both academic and community centers. The inclusion of 3 separate hospital centers also introduced other variables, such as differences in slide preparation and differences in quality control among institutions. Information regarding whether biopsies were ultrasonography or palpation guided was not easily obtainable and may have played a role in diagnostic yield. In addition, our study addressed a 4-category system of FNA interpretation, and conclusions may not extrapolate to other systems being used, such as the 6-category Bethesda system.

That being said, we do feel that our data strongly support the conclusion that case volume does significantly affect diagnostic interpretation, particularly for the atypical subset. Implications for improving accuracy for this atypical subset include the implementation of departmental quality improvement measures, further workup for atypical/indeterminate FNA results (either by obtaining a second opinion from a dedicated cytopathologist or in the future through cytomolecular testing), or by encouraging interpretation by higher-volume pathologists.

Finally, despite data suggesting that nondiagnostic results are higher among inexperienced clinicians performing fine-needle biopsies, there was surprisingly no significant difference between high-volume clinicians and low-volume clinicians in regard to nondiagnostic results in our series.²³ This suggests that other factors, such as slide preparation, pathologic interpretation, and the nature of the thyroid lesion itself, likely play a more important role in determining the overall rate of nondiagnosis than does clinician case volume. Clinician errors that do not result in nondiagnosis are obviously quite possible, however.

In conclusion, case volume significantly influences the pathologic interpretation of thyroid FNA, because low-volume pathologists report more atypical and fewer benign FNA results. Case volume did not have a significant impact on diagnostic yield, because thyroid FNAs performed by low-volume clinicians did not result in more frequent nondiagnostic result compared with those performed by high-volume clinicians.

Submitted for Publication: March 10, 2011; final revision received May 23, 2011; accepted September 14, 2011.

Corresponding Author: Jeffrey J. Houlton, MD, Department of Otolaryngology–Head and Neck Surgery, University of Cincinnati Academic Health Center, 231 Albert Sabin Way, Room 6407 MSB, Cincinnati, OH 45267-0528 (houltojj@ucmail.uc.edu).

Author Contributions: All authors had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. *Study concept and design:* Houlton, Sun, Lucas, and Steward. *Acquisition of data:* Houlton, Fernandez, and Lucas. *Analysis and interpretation of data:* Houlton, Sun, and Zhai. *Drafting of the manuscript:* Houlton and Fernandez. *Critical revision of the manuscript for important intellectual content:* Sun, Zhai, Lucas, and Steward. *Statistical analysis:* Houlton. *Administrative, technical, and material support:* Houlton, Zhai, and Lucas. *Study supervision:* Lucas and Steward.

Financial Disclosure: None reported.

Previous Presentation: This study was presented at the American Head and Neck Society 2011 Annual Meeting; April 27, 2011; Chicago, Illinois.

REFERENCES

- Amrikachi M, Ramzy I, Rubinfeld S, Wheeler TM. Accuracy of fine-needle aspiration of thyroid. *Arch Pathol Lab Med*. 2001;125(4):484-488.
- de Vos tot Nederveen Cappel RJ, Bouvy ND, Bonjer HJ, van Muiswinkel JM, Chadha S. Fine needle aspiration cytology of thyroid nodules: how accurate is it and what are the causes of discrepant cases? *Cytopathology*. 2001;12(6):399-405.
- Ravetto C, Colombo L, Dottorini ME. Usefulness of fine-needle aspiration in the diagnosis of thyroid carcinoma: a retrospective study in 37,895 patients. *Cancer*. 2000;90(6):357-363.
- Gharib H, Papini E. Thyroid nodules: clinical importance, assessment, and treatment. *Endocrinol Metab Clin North Am*. 2007;36(3):707-735, vi.
- Hamberger B, Gharib H, Melton LJ III, Goellner JR, Zinsmeister AR. Fine-needle aspiration biopsy of thyroid nodules. Impact on thyroid practice and cost of care. *Am J Med*. 1982;73(3):381-384.
- Cooper DS, Doherty GM, Haugen BR, et al; American Thyroid Association (ATA) Guidelines Taskforce on Thyroid Nodules and Differentiated Thyroid Cancer. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2009;19(11):1167-1214.
- Baloch ZW, LiVolsi VA, Asa SL, et al. Diagnostic terminology and morphologic criteria for cytologic diagnosis of thyroid lesions: a synopsis of the National Cancer Institute Thyroid Fine-Needle Aspiration State of the Science Conference. *Diagn Cytopathol*. 2008;36(6):425-437.
- Baloch ZW, Cibas ES, Clark DP, et al. The National Cancer Institute Thyroid fine needle aspiration state of the science conference: a summation. *Cytojournal*. 2008;5:6.
- Sherman SI, Angelos P, Ball DW, et al; National Comprehensive Cancer Network. Thyroid carcinoma. *J Natl Compr Canc Netw*. 2005;3(3):404-457.
- NCCN Clinical Practice Guidelines in Oncology: Thyroid Carcinoma. National Comprehensive Cancer Network Web site. <http://www.nccn.org>. Accessed October 30, 2010.
- Lewis CM, Chang KP, Pitman M, Faquin WC, Randolph GW. Thyroid fine-needle aspiration biopsy: variability in reporting. *Thyroid*. 2009;19(7):717-723.
- Hirokawa M, Carney JA, Goellner JR, et al. Observer variation of encapsulated follicular lesions of the thyroid gland. *Am J Surg Pathol*. 2002;26(11):1508-1514.
- Franc B. Observer variation of lesions of the thyroid. *Am J Surg Pathol*. 2003;27(8):1177-1179.
- Lloyd RV, Erickson LA, Casey MB, et al. Observer variation in the diagnosis of follicular variant of papillary thyroid carcinoma. *Am J Surg Pathol*. 2004;28(10):1336-1340.
- Filetti S, Durante C, Tortolano M. Nonsurgical approaches to the management of thyroid nodules. *Nat Clin Pract Endocrinol Metab*. 2006;2(7):384-394.
- Caruso DR, Mazzaferri EL. Fine needle aspiration biopsy in the management of thyroid nodules. *Endocrinologist*. 1991;1:194-202.
- Yassa L, Cibas ES, Benson CB, et al. Long-term assessment of a multidisciplinary approach to thyroid nodule diagnostic evaluation. *Cancer*. 2007;111(6):508-516.
- Tan GH, Gharib H. Thyroid incidentalomas: management approaches to nonpalpable nodules discovered incidentally on thyroid imaging. *Ann Intern Med*. 1997;126(3):226-231.
- Ezzat S, Sarti DA, Cain DR, Braunstein GD. Thyroid incidentalomas: prevalence by palpation and ultrasonography. *Arch Intern Med*. 1994;154(16):1838-1840.
- Gharib H, Goellner JR. Fine-needle aspiration biopsy of the thyroid: an appraisal. *Ann Intern Med*. 1993;118(4):282-289.
- Gharib H, Papini E, Valcavi R, et al; AACE/AME Task Force on Thyroid Nodules. American Association of Clinical Endocrinologists and Associazione Medici Endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules. *Endocr Pract*. 2006;12(1):63-102.
- Wang CC, Friedman L, Kennedy GC, et al. A large multicenter correlation study of thyroid nodule cytopathology and histopathology. *Thyroid*. 2011;21(3):243-251.
- Ljung BM, Drejet A, Chiampi N, et al. Diagnostic accuracy of fine-needle aspiration biopsy is determined by physician training in sampling technique. *Cancer*. 2001;93(4):263-268.