The Clinical and Diagnostic Impact of Using Standard Criteria of Adequacy Assessment and Diagnostic Terminology on Thyroid Nodule Fine Needle Aspiration

Xin Jing, M.D., Claire W. Michael, M.D., and Robert T. Pu, M.D., Ph.D.*

The study was aimed to investigate the impact of using standard criteria for assessing specimen adequacy and diagnostic terminology (CAST) on fine-needle aspiration (FNA) diagnosis and clinical management of thyroid nodules. The study included similar numbers of FNAs performed in 2 year before (group A) and 1.5 year after (group B) implementing the standard CAST. In comparison to group A, group B showed a significantly lower rate of nondiagnostic specimens (RND) (16.1% vs. 21.6%, $P \leq 0.01$*) and rate of descriptive diagnoses (RDD) (3.8% vs. 14.5%,* $P \leq 0.001$*) and greater non-neoplastic (70.0% vs. 64.1%,* $P < 0.05$*) and follicular cell lesions (7.4% vs. 4.3%,* $P < 0.05$*) but a similar percentage of neoplastic diagnoses. The rate of surgical follow-up (RSF) was significantly higher in group B than in group A, overall (21.6% vs. 17.0%,* $P < 0.05$*), or in subgroups of non-neoplastic (12.6% vs. 5.4%,* $P < 0.01$*) and neoplastic categories (81.0% vs. 61.0%,* $P < 0.05$*). The rate of cytohistologic concordance was higher in group B although the difference was not statistically significant. We concluded that use of the standard CAST on FNA diagnosis of thyroid nodules significantly reduced RND and RDD, providing more consistent diagnoses among the pathologists as well as better and more uniform communication between the pathologists and the clinicians. Furthermore, the cytohistological concordance was slightly better after CAST implementation, indicating that the improvement of diagnostic consistency among pathologists did not sacrifice the diagnostic accuracy.* Diagn. Cytopathol. 2008;36:161–166.

Key Words: FNA; thyroid nodules; standard CAST

Fine-needle aspiration (FNA) is a widely used triage procedure in the management of thyroid nodules, which aims to distinguish nodular lesions that require surgical intervention from those that may be managed conservatively with clinical and imaging follow-up.[1–4] As previously reported, on-site assessment of specimen adequacy has played an important role with respect to a decreased nondiagnostic rate and an increased diagnostic yield of FNA of thyroid nodules.[5–9] Various criteria for adequacy assessment are used not only among institutions[10] but also among pathologists in the same institution.[6] In this regard, some require six or more cell groups comprising of 10–20 cells per group on two separate slides[5] and others require no less than six clusters containing at least 10–12 follicular cells each.[8,11] Furthermore, it is noted that different diagnostic categories and terminology are used for cytological diagnosis of thyroid nodules.[6,8,11,12]

In our institution, lack of uniform criteria for adequacy assessment and diagnostic terminology created diagnostic inconsistency between pathologists and difficult communications with clinicians. This prompted the development of standard criteria for assessment of specimen adequacy and terminology (CAST). The standard CAST was distributed and explained to all involved parties including endocrinologists, radiologists, and surgeons and has been implemented in our cytology practice since 2005. This study was carried out to investigate the impact of using the standard CAST on FNA diagnosis and clinical management of thyroid nodules.

Materials and Methods

The study included FNA specimens obtained from two individual periods, 2 years (January 2003–December

Department of Pathology, University of Michigan, Ann Arbor, Michigan
*Correspondence to: Robert T. Pu, M.D., Ph.D., Department of Pathology, University of Michigan, 1500 E. Medical Center Drive, 2G332 UH, Ann Arbor, MI 48109-0054. E-mail: robertpu@umich.edu
Received 24 May 2007; Accepted 6 October 2007
DOI 10.1002/dc.20762
Published online in Wiley InterScience (www.interscience.wiley.com).

Table I. Diagnostic Category for Thyroid FNA

Interpretations	*Descriptions*
1. Suggestive of colloid nodule	Abundant colloid with scant or without follicular cells
2. Consistent with cyst contents	Numerous macrophages without follicular cells
3. Hemorrhagic cyst, suggest biopsy	Hemorrhagic cyst, ≥4 cm, with numerous macrophages but without follicular cells
4. Thyroiditis (lymphocytic or Hashimoto's)	Rich with lymphoplasmacytic infiltrate without atypia
5. Nodular hyperplasia	Six or more follicular cell clusters, colloid and histocytes
6. Follicular cell lesions (Nodular hyperplasia vs. follicular neoplasm)	Numerous follicular/Hurthle cell clusters with little colloid and equivocal features
7. Follicular/Hurthle cell neoplasm	Follicular/Hurthle cell clusters with loss of honeycombing, presence of microfollicles and syncytial sheets, with or without colloid
8. Positive for malignancy	Features consistent with malignancy

2004) prior to and one and a half years (January 2005–June 2006) after implementing the standard CAST. Pathologists involved in cytopathology service included the same seven individuals for both periods, although two of them were replaced in late period B by a new pathologist, all had similar levels of experience and subspecialty training in evaluating thyroid FNA. Overall, the staffing is not considered to have changed significantly. In our institution, endocrinologists, radiologists, and surgeons performed FNAs with or without ultrasound guidance and with or without cytopathologist-assisted on-site assessment of specimen adequacy. For FNAs with on-site adequacy assessment, two direct smears were made for each pass and the needle was then rinsed in Cytolyt^{TM} solution for a ThinPrep and/or a cell block. One smear was air-dried, stained with Diff-Quik stain, and then evaluated immediately for specimen adequacy. The other smear was quickly fixed with Sprayfix^{TM} and later stained with Papanicolaou stain before examination. For FNAs without on-site adequacy assessment, each specimen was submitted in Cytolyt^{TM} solution from which a ThinPrep and a cell block slide were prepared.

Our standard CAST defined specimens as adequate by the presence of at least six groups of follicular cells with a minimum of 10 cells in each group. Eight FNA diagnostic categories (Table I) were established according to Guidelines of the Papanicolaou Society of Cytopathology for the Examination of Fine-Needle Aspiration Specimens from Thyroid Nodules[13] with minor modification. An educational grand round was given to the endocrinologist and radiologist performing FNA procedures to explain the terms and meaning of the categories before the implementation of the CAST. Diagnoses made for adequate specimens without assigned diagnostic categories were termed as descriptive diagnoses in this study. Exception was made for suboptimal specimens containing atypical cells with features suggestive of a neoplasm and a diagnosis was made as such noting the suboptimal nature of the specimen regardless of the number of groups of cells. Cytohistologic correlation was performed for adequate specimens with categorized diagnoses and available surgical follow-up.

The parameters including rate of nondiagnostic specimen (RND), rate of descriptive diagnoses (RDD), rate of surgical follow-up (RSF), and rate of concordance (RC) were calculated as follows:

RND = (inadequate specimens/total specimens) × 100%

RDD = (diagnostic specimens with descriptive diagnoses/total diagnostic specimens) × 100%

RSF = (diagnostic specimens with surgical follow-up/total diagnostic specimens) × 100%

RC = (histology confirmed specimens/specimens with categorized diagnosis and surgical follow-up) × 100%.

For the purpose of this study, cytologic diagnoses made before implementing the standard CAST were converted to the standard terminology based upon the description and the same parameters mentioned above were calculated accordingly. The data generated before (group A) and after (group B) implementing the standard CAST were compared by using the χ^2 test, Fisher exact test, and unpaired *t*-test with the help of the SigmaStat.3.5 program (Systat Software). The level of difference was considered statistically significant when $P \leq 0.05$.

Results

A total of 764 specimens from 641 patients in group A and 824 specimens from 649 patients in group B were examined. More female than male patients were included in both group A (female/male = 537/112) and group B (female/male = 536/105). The gender distributions between the two groups were not statistically different. However, the average age of patients in group B (52.6 ± 0.6 yr-old) was older ($P < 0.001$) than those in group A (49.6 ± 0.6 yr-old). A summary of the specimen's demographic data is presented in Table II. As noted, specimens in group B were more frequently prepared with on-site adequacy assessment compared to those in group A (64.0% vs. 41.2%, $P < 0.001$). Significantly lower RND was achieved in specimens prepared with on-site adequacy assessment than those without on-site adequacy assessment in both group A (6.7% vs. 32.1%, $P < 0.001$) and group B (4.7% vs. 36.3%, $P < 0.001$). Overall RND was significantly lower in group B than in group A (16.1% vs. 21.6%, $P < 0.01$).

Table II. Data of Specimens Before and After Implementation of the Standard CAST

	Number of specimens (%)		*Nondiagnostic rate*	
Types of specimen	*A*	*B*	*A*	*B*
With on-site adequacy assessment	315 (41.2)	527 (64.0)[a]	21 (6.7)[b]	25 (4.7)[b]
Without on-site adequacy assessment	449 (58.8)	297 (36.0)[a]	144 (32.1)	108 (36.3)
Overall	764 (100)	824 (100)	165 (21.6)	133 (16.1)[c]

Before (A) and after (B) implementing the standard CAST.
[a]$P < 0.001$ compared to group A.
[b]$P < 0.001$ compared to those without on-site adequacy assessment.
[c]$P < 0.01$ compared to group A.

Table III. Cytologic Diagnoses and Surgical Follow-Up

	Number of specimens (%)		*Number of specimens with surgical follow-up (%)*	
Cytologic diagnosis	*A*	*B*	*A*	*B*
Lymphocytic thyroiditis	25 (4.2)	51 (7.4)[a]	4 (16.0)	4 (7.8)
Non-neoplastic	384 (64.1)	484 (70.0)[b]	22 (5.4)	61 (12.6)[a]
Neoplastic	77 (12.9)	79 (11.4)	47 (61.0)	64 (81.0)[b]
FCL (indeterminate)	26 (4.3)	51 (7.4)[b]	14 (53.8)	18 (35.2)
Descriptive diagnoses	87 (14.5)	26 (3.8)[c]	15 (17.2)	2 (7.7)
Overall	599 (100)	691 (100)	102 (17.0)	149 (21.6)[b]

Before (A) and after (B) implementing the standard CAST.
FCL, follicular cell lesions.
[a]$P < 0.01$ compared to group A.
[b]$P < 0.05$ compared to group A.
[c]$P < 0.001$ compared to group A.

Table IV. Cytohistologic Correlation in Cases with Surgical Follow-Up

	Histologic diagnosis					
Cytologic diagnosis	*Lymphocytic thyroiditis*	*Nodular hyperplasia*	*Follicular adenoma*	*Follicular carcinoma*	*Papillary carcinoma*	*Concordant rate*
Group A						
Lymphocytic thyroiditis	4	–	–	–	–	4/4 (100)
Non-neoplastic	–	17	4	1	–	17/22 (77.3)
Neoplastic	3	8	–	2	34	36/47 (76.6)
FCL (indeterminate)	–	11	–	1	2	12/14 (85.7)
Overall	7	36	4	4	36	69/87 (79.3)
Group B						
Lymphocytic thyroiditis	1	2	–	–	1	1/4 (25)
Non-neoplastic	1	46	1	–	3	46/5 (190.2)
Neoplastic	–	13	3	1	47	51/64 (79.7)
FCL (indeterminate)	–	10	4	1	3	15/18 (83.3)
Overall	2	71	8	2	54	113/137 (82.5)

Before (Group A) and after (Group B) the implementation of standard CAST.
FCL, follicular cell lesions.
Data are presented as number (percentage).

Table III illustrates distribution of adequate (diagnostic) specimens among the designated categories, which included lymphocytic thyroiditis (category 4), non-neoplastic (categories 1, 2, 3, 5), neoplastic (categories 7, 8), and follicular cell lesions (category 6). RDD and RSF are also presented hereby. Group B had a significantly lower RDD than group A (3.8% vs. 14.5%, $P < 0.001$) and all individuals made fewer numbers of descriptive diagnoses in group B compared to group A. Although both groups showed similar percentages in the neoplastic category, group B revealed a greater percentage than group A with regard to lymphocytic thyroiditis (7.4% vs. 4.2%, $P < 0.01$), non-neoplastic categories (70.0% vs. 64.1%, $P < 0.05$), and follicular cell lesions (7.4% vs. 4.3%, $P < 0.05$). Surgical follow-up was performed in a total of 102/599 specimens in group A and 149/691 specimens in group B. The overall RSF was significantly higher in Group B than in Group A (21.6% vs. 17.0%, $P < 0.05$). The same trend was observed in both non-neoplastic (12.6% vs. 5.4%, $P < 0.01$) and in neoplastic categories (81.0% vs. 61.0%, $P < 0.05$).

Table IV shows cytohistologic correlation in categorized specimens. The overall RC was slightly improved in group B although the difference did not achieve statistical

Table V. Distribution of Discrepant Cases

	Group A		*Group B*	
Cytologic diagnosis	*ThinPrep*	*Conventional smear*	*ThinPrep*	*Conventional smear*
Lymphocytic thyroiditis	0	0	1	2
Non-neoplastic	4	1	0	4
FCL (indeterminate)	0	2	1	2
Neoplastic				
Follicular neoplasm	5	4	5	2
Papillary carcinoma	1	1	1	5
Total	10	8	8	15

Before (Group A) and after (Group B) the implementation of standard CAST.
FCL, follicular cell lesion.
Data presented as number of cases.

significance between group A (79.3%) and group B (82.5%). RC was greater in group B than in group A for both non-neoplastic (other than lymphocytic thyroiditis) and neoplastic lesions while it appeared to be greater in group A than group B for lymphocytic thyroiditis, although the differences were below the level of statistical significance. For follicular cell lesion cytology diagnoses, although it can encompass from benign to malignant lesions, we only considered papillary carcinoma, only two cases were follicular variant of papillary carcinoma, on surgical resection as discordant cases with this group. The RC was similar between group A and group B (85.7% vs. 83.3%). The difference on overall diagnostic sensitivity of FNA cytology between group A (58/87 = 66.7%) and group B (104/137 = 76.4%) was not significantly different.

As shown in Table V, group A had 18 and group B had 23 discrepant cases. Ten (group A) and 8 (group B) cases were prepared by ThinPrep, and the remaining 8 (group A) and 15 (group B) cases were prepared by conventional smear. False positive diagnoses of follicular neoplasm attributed to most of these discrepancies regardless the methods of specimen preparation. Histology-confirmed neoplasms (adenoma or carcinoma) in a background of nodular hyperplasia/thyroiditis were seen in 7 cases in group A and 10 cases in group B in which cytologic diagnoses were mostly nodular hyperplasia or thyroiditis. The cytohistologic discrepancy was due to sampling error since even a second review of the cytology slides at the time of this study failed to identify cytological features diagnostic for a neoplasm with the knowledge that the follow-up diagnosis was neoplasm. Cytologic diagnosis of follicular neoplasm or papillary carcinoma was made in 11 cases in group A and 13 cases in group B, which was found to be non-neoplastic on surgical resection. This discrepant phenomenon resulted from errors in overinterpretation of some cytological features in those benign lesions that could be suggestive but not diagnostic of neoplasm.

Discussion

This study was carried out to evaluate the impact of using the standard CAST on FNA diagnosis and clinical management of thyroid nodules. Our data demonstrates that use of standard CAST on FNA diagnosis of thyroid nodules has significantly reduced RND and RDD without influencing the diagnostic concordance rate significantly. Similar to previously published data,[5–9] our results prove the benefit of performing on-site assessment of specimen adequacy in FNA diagnosis of thyroid nodules in terms of a decrease in nondiagnostic rate and an increase in diagnostic yield, regardless of the implementation of the standard CAST.

As previously mentioned, different criteria for specimen adequacy and diagnostic categories were used for cytological diagnosis of thyroid nodules. While assessing specimen adequacy we acknowledge that the nature of thyroid lesions should be considered. In this regard, follicular cells may be absent or rare in some benign thyroid nodules mainly composed of colloid material or undergoing extensive cystic degeneration with or without hemorrhage. Large ($\geq$3cm) and cystic/solid lesions may predict high prevalence of malignancy.[14] We thus used number of follicular cells as a general rule but not a sole measurement, and provided suggestive diagnoses (categories 1–3) while interpreting these specimens which might be classified as nondiagnostic due to the insufficient number of follicular cells. This approach proved to be useful in the appropriate clinical context. Hurthle cell neoplasm and follicular neoplasm were grouped into one category in the current study as a previous study conducted in our institution has indicated that Hurthle cell neoplasm did not predict more malignant potential compared to follicular neoplasm.[15]

During the period in which specimens in group A were collected, the majority of ultrasound-guided FNAs were performed by radiologists with only 50% having on-site assessments performed. In early 2005, the endodocrinologists became more actively involved in ultrasound-guided FNAs and were more willing to have a cytopathologist present for on-site adequacy assessment. This higher percentage contributed to the higher on-site adequacy assessment (smear plus ThinPrep/CB) in group B. Despite on-site adequacy assessment, more diagnostic inconsistency was encountered in group A due to the lack of the standard CAST. In this regard, some cases that appeared to be adequate to one individual at on-site evaluation may be changed to inadequate by a different individual on final examination. Further, after implementing the standard CAST, suggestive diagnoses were generated for specimens from lesions mainly comprised of colloid or cyst,

which might be treated as nondiagnostic previously due to too few or absence of follicular cells. These factors contributed to a significant reduction of RND in group B in which the specimen were interpreted using the standard CAST.

Although the cost-effectiveness for on-site assessment of specimen adequacy in FNA of thyroid lesions remains a controversial issue,[6,16] its effect on reducing the nondiagnostic rate of FNA of thyroid nodules has been proven.[5–9] In this study, data generated from both group A and B consistently demonstrated a dramatic drop in the nondiagnostic rate with on-site assessment of specimen adequacy. Use of uniform criteria was previously recommended,[17] however, there is no published data demonstrating the effect of using a standardized system for cytologic examination of the thyroid to the best of our knowledge.

It may be postulated that use of the standard CAST would reduce the surgical procedure rate as better-defined criteria for adequacy and diagnosis would reduce the rate of unnecessary operative procedures while still capturing all patients needing surgery. On the contrary, the rates of surgical follow-up for both non-neoplastic and neoplastic categories, as well as overall RSF were significantly higher after implementing the standard CAST. This phenomenon may be related to the older age of the patients in this group or other clinical factors rather than use of the CAST itself. One reasonable explanation is that less ambiguous diagnostic languages (lower RDD) coupled with fewer nondiagnostic specimens in group B prompt more definitive treatment (surgical) procedure instead of clinical follow-up with imaging study or repeat FNA. The specimens evaluated with the standard CAST showed slightly better cytohistologic concordance compared to those in which the standard CAST was not used for interpretation. We considered this as no significant impact on the concordance rate since the difference was not statistically significant. Furthermore, the level of correlation between cytology and histology demonstrated in the current study is similar to those previously described.[12,18] False positive diagnoses for follicular neoplasm represented most of the discrepant cases in the current study. A considerable challenge in the diagnosis of Hashimoto's thyroiditis on specimens prepared by ThinPrep has been previously reported.[19,20] In our study, all five histology-proven lymphocytic thyroiditis cases and three out of four false negative cases were prepared by the ThinPrep method.

After implementation of the standard CAST, all individuals made much fewer numbers of descriptive diagnoses. It is noted that the individuals reviewed different cases with various levels of complexity and distribution of case load varied among the individuals. Thus, it is not optimal to compare RDD between individuals. The dramatic drop in RDD coupled with significantly fewer nondiagnostic cases corresponded to the increased rates of benign and follicular cell lesion diagnoses while the neoplastic diagnosis rate remained relatively constant. As the RC was slightly improved after the implementation of the CAST, the reduction of RDD and RND thus did not sacrifice the diagnostic accuracy. This reflects the effort by the pathologists to adhere to the standard criteria and to use more consistent terminology to provide a more definitive diagnosis with less uncertainty whenever possible.

In conclusion, the reduction in RND and RDD following implementing the standard CAST in FNA of thyroid nodules provides more diagnostic consistency among the pathologists thus a better and more uniform communication between the pathologists and the clinicians without negatively impacting the diagnostic accuracy.

Acknowledgment

We are grateful for Brian Smola, C.T. for proof reading our manuscript.

References

1. Gharib H. Fine needle aspiration biopsy of thyroid nodules: Advantages, limitations, and effect. Mayo Clin Proc 1994;69:44–49.
2. Holleman F, Hoekstra JB, Ruitenberg HM. Evaluation of fine needle aspiration (FNA) cytology in the diagnosis of thyroid nodules. Cytopathology 1995;6:168–175.
3. Mandreker SR, Nadkarni NS, Pinto RG, Menezes S. Role of fine needle aspiration cytology as the initial modality in the investigation of thyroid lesions. Acta Cytol 1995;39:898–904.
4. Zagorianakou P, Malamou-Mitsi V, Zagorianakou N, Stefanou D, Tsatsoulis A, Agnantis NJ. The role of fine-needle aspiration biopsy in the management of patients with thyroid nodules. In Vivo 2005; 19:605–609.
5. Baloch ZW, Tam D, Langer J, Mandel S, LiVolsi VA, Gupta P. Ultrasound-guided fine needle aspiration biopsy of the thyroid: Role of on-site assessment and multiple cytologic preparations. Diagn Cytopathol 2000;23:425–429.
6. Eedes CR, Wang HH. Cost-effectiveness of immediate specimen adequacy assessment of thyroid fine-needle aspirations. Am J Clin Pathol 2004;121:64–89.
7. Ghofrani M, Beckman, Rimm D. The value of onsite adequacy assessment of thyroid fine-needle aspiration is a function of operator experience. Cancer (Cancer Cytopathol) 2006;108:110–113.
8. Kelly NP, Lim JC, DeJong S, Harmath C, Dudiak C, Wojcik E. Specimen adequacy and diagnostic specificity of ultrasound-guided fine needle aspirations of nonpalpable thyroid nodules. Diagn Cytopathol 2006;34:188–190.
9. Zhu W, Michael CW. How important is on site adequacy assessment for thyroid FNA? An evaluation of 883 cases. Diagn Cytopathol 2007;35:183–186.
10. AACE/AME Task Force on Thyroid Nodules. American Association of clinical endocrinologists and associazione medici endocrinology medical guidelines for clinical practice for the diagnosis and management of thyroid nodules. Endocr Pract 2006;12:63–102.
11. Methrotra P, Wiswanathan H, Johnson S, Wadehra V, Richardson DL, Lennard TWJ. Ultrasound guidance improves the adequacy of our preoperative thyroid cytology but not its accuracy. Cytopathology 2006;17:137–144.

12. Ylagan LR, Farkas T, Dehner LP. Fine needle aspiration of the thyroid: A cytohistologic correlation and study of discrepant cases. Thyroid 2004;14:35–41.
13. The Papanicolaou Society of Cytopathology Task Force on Standards of Practice. Guidelines of the Papanicolaou Society of Cytopathology for the examination of fine-needle aspiration specimens from thyroid nodules. Diagn Cytopathol 1996;15:84–89.
14. Meko JB, Norton JA. Large cystic/solid thyroid nodules: A potential false-negative fine-needle aspiration. Surgery 1995;118:996–1003.
15. Pu RT, Yang J, Wasserman PG, Bhuiya T, Griffith KA, Michael CW. Does Hurthle cell lesion/neoplasm predict malignancy more than follicular lesion/neoplasm on thyroid fine-needle aspiration? Diagn Cytopathol 2006;34:330–334.
16. Rimm DL, Stastny J, Rimm EB, Ayer S, Frable W. Comparison of the cost of fine-needle aspiration and open surgical biopsy as methods for obtaining a pathologic diagnosis. Cancer (Cancer Cytopathol) 1997;81:51–56.
17. Burch HB, Burman KD, Reed HL, Buckner L, Raber T, Ownbey JL. Fine needle aspiration of thyroid nodules. Determinants of insufficiency rate and malignancy yield at thyroidectomy. Acta Cytol 1996;40:176–183.
18. Sidway MK, Del Vecchio DM, Knoll SM. Fine-needle aspiration of thyroid nodules: Correlation between cytology and histology and evaluation of discrepant cases. Cancer (Cancer Cytopathol) 1997; 81:253–259.
19. Biscotti CV, Hollow JA, Toddy SM, Easley KA. ThinPrep versus conventional smears cytologic preparation in the analysis of thyroid fine-needle aspiration specimens. Am J Clin Pathol 1995;104:150–153.
20. Cochand-Priollet B, Prat JJ, Polivka M, Thienpont L, Dahan H, Wassef M, Guillausseau PJ. Thyroid fine needle aspiration: the morphological features of ThinPrep slide preparations. Eighty cases with histological control. Cytopathology 2003;14: 343–349.