

The Bethesda System for Reporting Thyroid Cytopathology: A Meta-Analysis

Massimo Bongiovanni^a Alessandra Spitale^a William C. Faquin^b
Luca Mazzucchelli^a Zubair W. Baloch^c

^aInstitute of Pathology, Locarno, Switzerland; ^bMassachusetts General Hospital, Boston, Mass., and

^cUniversity of Pennsylvania Medical Center, Philadelphia, Pa., USA

Key Words

Bethesda · Thyroid fine-needle aspiration · Meta-analysis

Abstract

Objective: We aimed to investigate the validity of the Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) through meta-analysis. **Study Design:** All publications between January 1, 2008 and September 1, 2011 that studied TBSRTC and had available histological follow-up data were retrieved. To calculate the sensitivity, specificity and diagnostic accuracy, the cases diagnosed as follicular neoplasm, suspicious for malignancy and malignant which were histopathologically confirmed as malignant were defined as true-positive. True-negative included benign cases confirmed as benign on histopathology. The nondiagnostic category was excluded from the statistical calculation. The correlations between the 6 diagnostic categories were investigated. **Results:** The publications review resulted in a case cohort of 25,445 thyroid fine-needle aspirations, 6,362 (25%) of which underwent surgical excision; this group constituted the basis of the study. The sensitivity, specificity and diagnostic accuracy were 97, 50.7 and 68.8%, respectively. The positive predictive value and negative predictive value were 55.9 and 96.3%, respectively. The rates of false negatives and false positives were low: 3 and 0.5%, respectively. **Conclusions:**

The results of meta-analysis showed high overall accuracy, indicating that TBSRTC represents a reliable and valid reporting system for thyroid cytology.

Copyright © 2012 S. Karger AG, Basel

Introduction

Thyroid nodules are common and can be detected by ultrasound in up to 60% of the general population [1–3]. Data from the Surveillance Epidemiology and End Results (SEER) registry show an increasing prevalence of differentiated thyroid cancer worldwide [4, 5], most likely due to the increased detection of small papillary carcinomas [6]. For the initial evaluation of patients with such nodules, thyroid fine-needle aspiration (FNA) has proven to be a rapid, cost-effective, safe and reliable method of investigation. Moreover, this procedure is the most appropriate diagnostic tool to distinguish between patients that require clinical management or surgical excision. The increasing prevalence of thyroid cancer and improvements in the technology and resolution of ultrasound machines have led to an increasing number of cytological diagnostic procedures [7].

In 2007, a conference with one of the objectives being to standardize the diagnostic terminology for the report-

ing of thyroid cytopathology results was held in Bethesda, Md. The recommendations resulting from this conference led to the formation of The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC). This classification scheme has achieved its purpose of standardization of thyroid-reporting cytopathology, as evidenced by several publications [8]. By the same token, the category of follicular lesion/atypia of undetermined significance remains the most controversial category due to heterogeneity in its use among institutions and follow-up. This is due to the fact that it is nearly impossible to establish distinct morphologic criteria for diagnosing atypia among cytopathologists [9–16]. It has been shown that the best way to implement TBSRTC in a pathology laboratory and to convince the clinicians of its validity and robustness is to compare the outcome data prior to and after the implementation of TBSRTC. A comparison of the outcomes obtained by several institutions showed notable differences in the percentage of patients referred to surgery and the risk of malignancy for different diagnostic categories (DCs).

So the question one must ask now after 4 years of TBSRTC is whether this classification scheme has served its purpose of refining the cytopathologic interpretation of thyroid FNA and the management of thyroid nodules. In this investigation, we performed a comprehensive meta-analysis of thyroid FNA studies conducted using TBSRTC, to assess the variability in the use of the 6 DCs between different institutions with the aim of investigating the validity of this reporting system in view of histological outcomes.

Materials and Methods

Selection Criteria

The primary source of the reviewed publications was the PubMed database (<http://preview.ncbi.nlm.nih.gov/pubmed>) and manual searches of the reference list included in the selected studies. The search consisted of literature published from January 2008 to September 2011. The inclusion criteria for the meta-analysis were publications written in the English language which classified thyroid FNA specimens according to TBSRTC and included surgical follow-up. Eight studies met the inclusion criteria and were utilized (table 1) [17–24]. The data collected from the University of Pennsylvania Medical Center, Philadelphia and the Massachusetts General Hospital, Boston, have already been reported in a different study [23] and have been split for this meta-analysis. The data points for all of the DCs and histopathologic follow-up were extracted and entered into an Excel datasheet for statistical analysis.

Cytological Classification

For the purpose of this report, the 6 DCs used in the TBSRTC were defined as: DC I = nondiagnostic, DC II = benign, DC III =

atypia/follicular lesion of undetermined significance (AUS/FLUS), DC IV = follicular neoplasm/suspicion for a follicular neoplasm (FN/SFN), DC V = suspicious for malignancy and DC VI = malignant (table 2) [8, 25, 26].

Follow-Up Data

The follow-up data included only reported histopathologic follow-up. Any incidentally detected lesions in the surgical resection specimens that were not the target of the thyroid FNA were excluded from the statistical analysis. One author included microcarcinomas in the histopathological follow-up [24].

Statistical Analysis

Statistical analysis was performed on the final Excel spreadsheet containing the data extracted from each publication employing both a log-linear model (likelihood ratio) and a χ^2 model with symmetric measures of association. Two-sided p values of <0.05 were considered statistically significant. The phi correlation was calculated to measure the strength of the association between categorical data; values ranged from +1 to -1, with +1 denoting a strong positive association, -1 indicating a strong negative association and 0 demonstrating no association.

The sensitivity, specificity and diagnostic accuracy were calculated considering thyroid FNA as a 'screening test'; FNA specimens interpreted as benign (DC II) were considered to be true-negative samples and the remaining categories (DC IV, DC V and DC VI) were considered to be true-positive samples because they led to a recommendation of surgery. The false-positive category included cases that were diagnosed as follicular neoplasm, suspicious for malignancy and malignant – that were confirmed as benign. The false-negative cases included those diagnosed as benign on FNA but confirmed as malignant upon surgical excision.

The use of the term 'positive' is for statistical purposes only and does not indicate 'malignant.' The DC I category was excluded from the statistical analysis because these diagnoses usually led to a repeat FNA rather than to surgical excision. Considering that almost 40% of cases in the AUS/FLUS category underwent surgery (see results, table 3), we performed an additional statistical analysis including the AUS/FLUS category as true-positive samples in the calculation of sensitivity, PPV and diagnostic accuracy.

SAS statistical software (version 9.1; SAS Institute Inc., Cary, N.C., USA) was used for the statistical analyses.

Results

The 8 published studies used for our meta-analysis together with the percentage of FNA cases placed into each DC from all publications included are listed in table 1. The 6 DCs used in the TBSRTC are described in table 2.

A total of 6,362 (25%) of the 25,445 thyroid FNAs included in the meta-analysis were followed by surgery, and the cytohistological correlations from these cases are presented in table 3. For each of the 6 individual DCs in TBSRTC, we observed the following:

DC I: FNAs placed in this category ranged from 1.8% [21] to 23.6% [18] with an overall value of 12.9%, when

Table 1. Summary of the 8 articles used in the meta-analysis

First author	FNA cases	ND	Benign	AUS/FLUS	FN/SFN	SM	Malignant	Cases with follow-up	Benign histology	Malignant histology
Jo [17]	2,987	509 (17.0)	1,792 (60.0)	101 (3.4)	298 (10.0)	71 (2.4)	216 (7.2)	1,022 (34.2)	746 (73.0)	276 (27.0)
Renshaw [18]	7,086	1,671 (23.6)	3,829 (54.0)	548 (7.7)	606 (8.6)	131 (1.9)	301 (4.2)	1,331 (18.8)	906 (68.1)	425 (31.9)
Nayar [19]	5,194	274 (5.3)	3,337 (64.2)	924 (17.8)	307 (5.9)	97 (1.9)	255 (4.9)	1,413 (27.2)	1,079 (76.4)	334 (23.6)
Theoharis [20]	3,207	357 (11.1)	2,368 (73.8)	95 (3.0)	176 (5.5)	43 (1.4)	168 (5.2)	378 (11.8)	176 (46.6)	202 (53.4)
Kim [21]	865	16 (1.8)	504 (58.3)	141 (16.3)	10 (1.2)	54 (6.2)	140 (16.2)	204 (23.6)	22 (10.8)	182 (89.2)
Her-Juing Wu [22]	1,382	278 (20.1)	539 (39.0)	376 (27.2)	116 (8.4)	36 (2.6)	37 (2.7)	205 (14.8)	157 (76.6)	48 (23.4)
Bongiovanni [23]	3,474 ¹ 250 ²	70 (2.0) 40 (16.0)	1,898 (54.7) 166 (66.4)	220 (6.3) 28 (11.2)	880 (25.3) 6 (2.4)	219 (6.3) 5 (2.0)	187 (5.4) 5 (2.0)	1,305 (37.6) 53 (21.2)	763 (58.5) 42 (79.2)	542 (41.5) 11 (20.8)
Bohacek [24]	1,000	56 (5.6)	671 (67.1)	8 (0.8)	172 (17.2)	24 (2.4)	69 (6.9)	451 (45.1)	321 (71.2)	130 (28.8)
Total	25,445	3,271	15,104	2,441	2,571	680	1,378	6,362	4,212	2,150

Numbers in parentheses are percentages. ND = Nondiagnostic; SM = suspicious for malignancy.

¹ Cases provided by Z.W.B., University of Pennsylvania Medical Center, Philadelphia, Pa., USA. ² Cases provided by W.C.F., Massachusetts General Hospital, Boston, Mass., USA.

Table 2. Diagnostic categories, associated risk of malignancy and clinical management in TBSRTC (modified by Ali and Cibas [8])

Diagnostic category	Cytological diagnosis	Risk of malignancy, %	Usual management
I	nondiagnostic or unsatisfactory	1–4	repeat FNA with ultrasound guidance
II	benign	0–3	clinical follow-up
III	AUS/FLUS	5–15	repeat FNA
IV	FNS/SFN	15–30	surgical lobectomy
V	suspicious for malignancy	60–75	near-total thyroidectomy or surgical lobectomy
VI	malignant	97–99	near-total thyroidectomy

Table 3. Cytohistological correlations in the 6-tiered Bethesda System

Cytological diagnosis	All FNAs		All FNAs with histological follow-up			Benign histology		Malignant histology	
	n	% total	n	% total ^a	% category ^b	n	% ^c	n	% ^c
Nondiagnostic	3,271	12.9	530	8.3	16.2	441	83.2	89	16.8
Benign	15,104	59.3	1,563	24.6	10.4	1,505	96.3	58	3.7
AUS/FLUS	2,441	9.6	957	15.0	39.2	805	84.1	152	15.9
FN/SFN	2,571	10.1	1,791	28.2	69.7	1,323	73.9	468	26.1
Suspicious for malignancy	680	2.7	501	7.9	73.7	124	24.8	377	75.2
Malignant	1,378	5.4	1,020	16.0	74.0	14	1.4	1,006	98.6
Total	25,445	100	6,362	100	25.0	4,212	66.2	2,150	33.8

^a Percentage of the 6,362 cases with follow-up. ^b Percentage of cases operated in each DC. ^c Percentage of cases calculated of the total number of operated cases in each category.

Table 4. Statistical analysis of the 6-tiered Bethesda System

Cytological comparisons	χ^2 ^a	phi ^b	LR ^c	DF	p value
DC I vs. DC II vs. DC III vs. DC IV vs. DC V vs. DC VI	3,186.02	0.71	3,559.46	5	<0.0001
DC II vs. DC III vs. DC IV vs. DC V vs. DC VI	3,046.39	0.72	3,476.17	4	<0.0001
DC II vs. DC VI	2,295.45	0.94	2,856.43	1	<0.0001
DC II vs. DC V	1,167.29	0.75	1,069.13	1	<0.0001
DC II vs. DC IV	317.26	0.31	360.28	1	<0.0001
DC II vs. DC III	115.13	0.21	111.93	1	<0.0001
DC II vs. DC I	103.72	-0.23	88.62	1	<0.0001
DC III vs. DC VI	1,393.17	0.84	1,696.60	1	<0.0001
DC III vs. DC V	501.32	0.59	511.56	1	<0.0001
DC III vs. DC IV	37.49	0.12	39.09	1	<0.0001
DC III vs. DC I	0.21	-0.01	0.21	1	0.6485
DC IV vs. DC VI	1,369.51	0.70	1,684.74	1	<0.0001
DC IV vs. DC V	405.81	0.42	399.13	1	<0.0001
DC IV vs. DC I	19.55	0.09	20.72	1	<0.0001
DC V vs. DC VI	222.58	0.38	216.86	1	<0.0001
DC V vs. DC I	355.31	0.59	379.32	1	<0.0001
DC VI vs. DC I	1,126.32	0.85	1,248.81	1	<0.0001

LR = Likelihood ratio; DF = degrees of freedom.

^a χ^2 model for associations between categorical data. ^b phi correlation coefficient for the strength of association between categorical data. ^c Likelihood ratio (log-linear model) for association between categorical data.

considering all the studies in the meta-analysis (table 3). Surgical resection was performed in 16.2% of cases, and the risk of malignancy was 16.8%.

DC II: The cases in this category ranged from 39% [22] to 73.8% [20] with an overall value of 59.3%, and a cumulative malignancy rate of 3.7%.

DC III: FNA cases in this category ranged from 3% [20] to 27.2% [22] with an overall value of 9.6% and an overall rate of malignancy of 15.9%.

DC IV: Cases in this category ranged from 1.2% [21] to 25.3% [23] with an overall value of 10.1%. More than two-thirds of these (approx. 70%) underwent surgery, with a risk of malignancy of 26.1%.

DC V: FNA cases in this category ranged from 1.4% [20] to 6.3% [23] with an overall value of 2.7%. The mean risk of malignancy in this category was 75.2%.

DC VI: The reported rate of malignancy in all publications included in this analysis ranged from 2% [23] to 16.2% [21] % with an overall value of 5.4% and a risk of malignancy of 98.6%.

There was a robust correlation between the DCs and the histological follow-up, except between the DCs AUS/FLUS and nondiagnostic (table 4).

The main diagnostic indicators in the TBRSTC are presented in table 5. The negative predictive value (NPV) and PPV for the benign and malignant categories were

96.3 and 98.6%, respectively. The sensitivity, specificity and diagnostic accuracy were 97, 50.7 and 68.8%, respectively with a false-negative rate of 3% and false-positive rate of 0.5%. The sensitivity and diagnostic accuracy were 97.2 and 60.2%, respectively, when considering the AUS/FLUS category as true-positive samples.

Discussion

As evidenced by the extensive body of literature, a tiered diagnostic classification scheme, such as the 6-tiered one proposed by TBSRTC or those proposed by other associations, is an effective approach for the diagnosis and management of thyroid nodules [8, 27, 28]. The advantage of TBSRTC is the standardization of the reporting of thyroid cytology, which, prior to 2007, consisted of nonreproducible classification schemes that in some cases included either too few or too many DCs. The 6 DCs of TBSRTC arose from a probabilistic approach: the probability that a thyroid lesion placed into a specific category would show histological evidence of malignancy [29]. The advantage of this approach is that each of the 6 DCs can be associated with an implied risk of malignancy that translates into a recommendation for clinical management. Since its inception, many authors have

Table 5. Main diagnostic indicators for the 6-tiered Bethesda System

Indicator	6-tiered system	
Sensitivity (DC VI + DC V + DC IV) ^a	97.0%	(1,851/1,909)
Sensitivity (DC VI + DC V + DC IV + DC III) ^b	97.2%	(2,003/2,061)
Specificity	50.7%	(1,505/2,966)
PPV for DC III	15.9%	(152/957)
PPV for DC VI	98.6%	(1,006/1,020)
PPV for DC V	75.2%	(377/501)
PPV for DC IV	26.1%	(468/1,791)
PPV for DC VI + DC V + DC IV	55.9%	(1,851/3,312)
PPV for DC VI + DC V + DC IV + DC III	46.9%	(2,003/4,269)
NPV for DC II	96.3%	(1,505/1,563)
Rate of false-negative	3.0%	(58/1,909)
Rate of false-positive	0.5%	(14/2,966)
Diagnostic accuracy (DC VI, DC V, DC IV, DC II) ^c	68.8%	(3,356/4,875)
Diagnostic accuracy (DC VI, DC V, DC IV, DC III, DC II) ^d	60.2%	(3,508/5,832)
Rate of DC I	12.9%	(3,271/25,445)
Rate of DC II	59.4%	(15,104/25,445)
Rate of DC III	9.6%	(2,441/25,445)
Rate of DC IV	10.1%	(2,571/25,445)
Rate of DC V	2.7%	(680/25,445)
Rate of DC VI	5.4%	(1,378/25,445)

^a Complete sensitivity considers DC VI + DC V + DC IV categories true-positive cases. ^b Complete sensitivity considers DC VI + DC V + DC IV + DC III categories true-positive cases. ^c We considered DC VI + DC V + DC IV categories to be true-positive. ^d We considered DC VI + DC V + DC IV + DC III categories to be true-positive.

published their experience with TBSRTC; therefore, we believe it is time to take a closer look at the combined clinical applicability of this classification scheme.

We approached this meta-analysis with the following questions. How has TBSRTC impacted the clinical management of thyroid nodules, and how has the variability in reporting among different institutions as well as the average risk of malignancy for each category changed since the implementation of TBSRTC? We acknowledge that the comparison between pre- and post-TBSRTC has significant limitations; however, the utility of the information gained from the analysis as presented here has a practical value in reflecting the current state of TBSRTC and its validity/applicability in various practices.

Surgical Follow-Up

The percentage of cases that underwent surgery varied greatly between different institutions, ranging from 11.8% [20] to 45.1% [24] with an average rate of 25% (6,362 out of 25,445). These variable rates could be explained by the different follow-up times in the studies we used, rather than being considered as true differences in the surgical follow-up. More than two-thirds of the cases for which

the TBSRTC indicated surgical management (a DC \geq FN/SFN) were resected. Interestingly, the rate of surgery was almost identical (i.e. 69.7, 73.7 and 74.0%) in the FN/SFN, suspicious for malignancy and malignant categories, respectively.

Diagnostic Categories

The validity of using 6 DCs is justified by the strong correlation between each DC and the histological outcomes in predicting benignity versus malignancy. The only nonstatistically significant difference was observed between the nondiagnostic specimens and the AUS/FLUS DC. Interestingly, both AUS/FLUS and non-diagnostic cases had similar risks of malignancy, 16.8 and 15.9%, respectively.

The average percentage of cases in the non-DC was 12.9%, ranging from 1.8 to 23.6%. This shows that even when the criteria for a satisfactory specimen are so-called 'well established' for solid or complex nodules, there exists a noticeable variability among institutions in classifying thyroid FNA cases as nondiagnostic. Thus, the range of nondiagnostic cases among different institutions is still as high as it was in the pre-TBSRTC era [30]. One explana-

tion for this high nondiagnostic rate, previously suggested by Renshaw [30], is that published studies such as those used in this meta-analysis are primarily based upon the experience of academic/large reference centers, and individual thyroid FNA cases sent to these centers for aspiration may represent cases that are technically challenging and where it is difficult to procure an adequate specimen.

If we calculate the rate of positive cases as the number of cases diagnosed as malignant divided by all of the cases aspirated (i.e. 1,378/25,445), we obtain a malignancy rate of 5.4% that falls exactly in line with the inverse correlation of the nondiagnostic rate and positive rate described by Renshaw [30], thus supporting the findings of our study. Up to 16.2% of cases classified as nondiagnostic underwent surgical excision with a malignancy rate of approximately 17%. This is much higher than proposed by the TBSRTC (1–4%). Interestingly, the malignancy rate of the non-DC is even slightly more elevated than the one observed for the AUS/FLUS category, even though more than twice the number of cases classified as AUS/FLUS underwent surgical excision. It is not known whether all of the thyroid FNA cases that underwent surgery were classified as nondiagnostic on the initial or the repeat FNA. This finding from the meta-analysis suggests that a nondiagnostic aspirate obtained by an experienced operator from a sonographically suspicious nodule should be managed cautiously due to the appreciable risk of malignancy [31, 32].

The percentage of cases classified as benign showed a malignancy rate of 3.7% which is slightly higher than that recommended by TBSRTC guidelines (0–3%); however, it is within the 0–5% range reported by the American Thyroid Association guidelines [3]. The studies by Theoharis et al. [20] and Her-Juing Wu et al. [22] showed the widest variation in the percentage of benign cases (73.8 vs. 39%). Interestingly, a similar wide variation was also observed for the AUS/FLUS category in these reports (3 vs. 27.2%). This variability could be explained by differences in patient population, nodule selection criteria or subjective classification issues.

The overall percentage of AUS/FLUS cases in this analysis was 9.6%, with a 15.9% risk of malignancy. As evidenced by many pre- and post-TBSRTC publications on this topic, the cases classified as AUS/FLUS represent the 'gray zone' in thyroid cytopathology for many reasons. Some authors have recommended a further subdivision of the AUS/FLUS category to better reflect the malignancy risk [18]. In our meta-analysis, 39.2% of cases in this category underwent surgical resection, although we were unable to determine whether the surgery was performed after

the first or repeat FNA diagnosis of AUS/FLUS. Because of this high percentage of AUS/FLUS cases undergoing surgery, we might ask if this category is perceived as a 'positive test', more than 'indeterminate results' that require repeated FNA, according to TBSRTC. Considering the AUS/FLUS category as a true positive sample in the statistical analysis, the sensitivity of thyroid FNA has not changed much, from 97 to 97.2% (table 5). In contrast, the PPV value and diagnostic accuracy have fallen down slightly from 55.9 to 46.9% and from 68.8 to 60.2%, respectively (table 5). The question arises how one can prevent the overuse of the AUS/FLUS category in a clinical practice. Krane et al. [33] proposed using the AUS/FLUS:malignant ratio as a performance measure for the reporting of thyroid FNA in TBSRTC, similar to the ASCUS:LSIL (atypical squamous cells of undetermined significance:low-grade squamous intraepithelial lesion) ratio used in gynecological cytology. These authors also recommended that this ratio should be between 1 and 3. In our analysis the AUS/FLUS:malignant ratio was 1.8%. It has been shown that adjunct molecular studies might be beneficial in further triaging of these cases. At present, these molecular tests, which consist of gene-mutation panels or RNA analysis, are being offered at both institutional and commercial levels with either a high NPV or PPV [34, 35].

The FN/SFN DC also showed a very wide range of 1.2 to 25.3% (mean 10.1%) with a risk of malignancy of 26.1%. Interestingly, this finding is compatible with rates given for thyroid FNAs prior to TBSRTC. Approximately 70% of patients in this category underwent surgical excision of the thyroid nodule, attesting to the implications of the use of the term 'neoplasm'. The mean risk of malignancy for cases classified as suspicious for malignancy and malignant was 75.2 and 98.6%, respectively. These rates are also consistent with those presented in pre-TBSRTC publications.

Conclusions

As evidenced by its high sensitivity and high NPV, TBSRTC has proven to be an effective and robust thyroid FNA classification scheme to guide the clinical management of patients with thyroid nodules. The findings of this analysis show a growing trend i.e. that many institutions both nationally and worldwide are adopting this reporting system in order to provide their clinicians with cytopathology reports that are clear and comprehensible and permitting comparisons and performance evaluations on a larger scale.

References

- Mazzaferri EL: Management of a solitary thyroid nodule. *N Engl J Med* 1993;328:553–559.
- Gharib H, Papini E, Valcavi R, Baskin HJ, Crescenzi A, Dottorini ME, Duick DS, Guglielmi R, Hamilton CR Jr, Zeiger MA, Zini M: 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;12:63–102.
- Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, Mandel SJ, Mazzaferri EL, McIver B, Pacini F, Schlumberger M, Sherman SI, Steward DL, Tuttle RM: Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid* 2009;19:1167–1214.
- Ries LAG, Melbert D, Krapcho M, et al: SEER Cancer Statistics Review, 1975–2005. Bethesda, National Cancer Institute, 2007. http://seer.cancer.gov/csr/1975_2005/
- Chen AY, Jemal A, Ward EM: Increasing incidence of differentiated thyroid cancer in the United States, 1988–2005. *Cancer* 2009;115:3801–3807.
- Hassell LA, Gillies EM, Dunn ST: Cytologic and molecular diagnosis of thyroid cancers: Is it time for routine reflex testing? *Cancer Cytopathol* 2011;120:7–17.
- Ross DS: Predicting thyroid malignancy (editorial). *J Clin Endocrinol Metab* 2006;91:4253–4255.
- Ali SZ, Cibas ES: The Bethesda System for Reporting Thyroid Cytopathology. Definitions, criteria and explanatory notes. New York, Springer, 2010.
- Abele JS, Levine RA: Diagnostic criteria and risk-adapted approach to indeterminate thyroid cytodiagnosis. *Cancer Cytopathol* 2010;118:415–422.
- Marhefka GD, McDivitt JD, Shakir KM, Drake AJ 3rd: Diagnosis of follicular neoplasm in thyroid nodules by fine needle aspiration cytology: does the result, benign vs. suspicious for a malignant process, in these nodules make a difference? *Acta Cytol* 2009;53:517–523.
- Miller B, Burkey S, Lindberg G, Snyder WH 3rd, Nwariaku FE: Prevalence of malignancy within cytologically indeterminate thyroid nodules. *Am J Surg* 2004;188:459–462.
- Pang T, Ihre-Lundgren C, Gill A, McMullen T, Sywak M, Sidhu S, Delbridge L: Correlation between indeterminate aspiration cytology and final histopathology of thyroid neoplasms. *Surgery* 2010;148:532–537.
- Poller DN, Ibrahim AK, Cummings MH, Mikel JJ, Boote D, Perry M: Fine-needle aspiration of the thyroid. *Cancer* 2000;90:239–244.
- Rabaglia JL, Kabbani W, Wallace L, Holt S, Watumull L, Pruitt J, Snyder WH, Nwariaku FE: Effect of the Bethesda system for reporting thyroid cytopathology on thyroidectomy rates and malignancy risk in cytologically indeterminate lesions. *Surgery* 2010;148:1267–1272.
- Redman R, Yoder BJ, Massoll NA: Perceptions of diagnostic terminology and cytopathologic reporting of fine-needle aspiration biopsies of thyroid nodules: a survey of clinicians and pathologists. *Thyroid* 2006;16:1003–1008.
- Yoder BJ, Redman R, Massoll NA: Validation of a five-tier cytodiagnostic system for thyroid fine needle aspiration biopsies using cytohistologic correlation. *Thyroid* 2006;16:781–786.
- Jo VY, Stelow EB, Dustin SM, Hanley KZ: Malignancy risk for fine-needle aspiration of thyroid lesions according to the Bethesda system for reporting thyroid cytopathology. *Am J Clin Pathol* 2010;134:450–456.
- Renshaw AA: Should ‘atypical follicular cells’ in thyroid fine-needle aspirates be subclassified? *Cancer Cytopathol* 2010;118:186–189.
- Nayar R, Ivanovic M: The indeterminate thyroid fine-needle aspiration: experience from an academic center using terminology similar to that proposed in the 2007 National Cancer Institute Thyroid Fine Needle Aspiration State of the Science Conference. *Cancer Cytopathol* 2009;117:195–202.
- Theoharis CG, Schofield KM, Hammers L, Udelsman R, Chhieng DC: The Bethesda thyroid fine-needle aspiration classification system: year 1 at an academic institution. *Thyroid* 2009;19:1215–1223.
- Kim SK, Hwang TS, Yoo YB, Han HS, Kim DL, Song KH, Lim SD, Kim WS, Paik NS: Surgical results of thyroid nodules according to a management guideline based on the braf(v600e) mutation status. *J Clin Endocrinol Metab* 2011;96:658–664.
- Her-Juing Wu H, Rose C, Elsheikh TM: The Bethesda system for reporting thyroid cytopathology: an experience of 1,382 cases in a community practice setting with the implication for risk of neoplasm and risk of malignancy. *Diagn Cytopathol* 2012;40:399–403.
- Bongiovanni M, Crippa S, Baloch Z, Piana S, Spitale A, Pagni F, Mazzucchelli L, Di Bella C, Faquin W: Comparison of 5-tiered and 6-tiered diagnostic systems for the reporting of thyroid cytopathology: a multi-institutional study. *Cancer Cytopathol* 2012;120:117–125.
- Bohacek L, Milas M, Mitchell J, Siperstein A, Berber E: Diagnostic accuracy of surgeon-performed ultrasound-guided fine-needle aspiration of thyroid nodules. *Ann Surg Oncol* 2012;19:45–51.
- Cibas ES, Ali SZ: The Bethesda system for reporting thyroid cytopathology. *Thyroid* 2009;19:1159–1165.
- Cibas ES, Ali SZ: The Bethesda system for reporting thyroid cytopathology. *Am J Clin Pathol* 2009;132:658–665.
- British Thyroid Association, Royal College of Physicians: Guidelines for Management of Thyroid Cancer. Report of the thyroid cancer guidelines update group, ed 2. London, Royal College of Physicians, 2007.
- Fadda G, Basolo F, Bondi A, Bussolati G, Crescenzi A, Nappi O, Nardi F, Papotti M, Taddei G, Palombini L: Cytological classification of thyroid nodules. Proposal of the SIAPEC-IAP Italian consensus working group. *Pathologica* 2010;102:405–408.
- Ali SZ: Thyroid cytopathology: Bethesda and beyond. *Acta Cytol* 2011;55:4–12.
- Renshaw AA: Non-diagnostic rates for thyroid fine needle aspiration are negatively correlated with positive for malignancy rates. *Acta Cytol* 2011;55:38–41.
- Frates MC, Benson CB, Charboneau JW, Cibas ES, Clark OH, Coleman BG, Cronan JJ, Doubilet PM, Evans DB, Goellner JR, Hay ID, Hertzberg BS, Intenzo CM, Jeffrey RB, Langer JE, Larsen PR, Mandel SJ, Middleton WD, Reading CC, Sherman SI, Tessler FN: Management of thyroid nodules detected at US: Society of Radiologists in Ultrasound consensus conference statement. *Ultrasound Q* 2006;22:231–238.
- Baloch Z, LiVolsi VA, Jain P, Jain R, Aljada I, Mandel S, Langer JE, Gupta PK: Role of repeat fine-needle aspiration biopsy (FNAB) in the management of thyroid nodules. *Diagn Cytopathol* 2003;29:203–206.
- Krane JF, Vanderlaan PA, Faquin WC, Renshaw AA: The atypia of undetermined significance/follicular lesion of undetermined significance:malignant ratio: a proposed performance measure for reporting in the Bethesda system for thyroid cytopathology. *Cancer Cytopathol* 2012;120:111–116.
- Nikiforov YE, Ohori NP, Hodak SP, Carty SE, LeBeau SO, Ferris RL, Yip L, Seethala RR, Tublin ME, Stang MT, Coyne C, Johnson JT, Stewart AF, Nikiforova MN: Impact of mutational testing on the diagnosis and management of patients with cytologically indeterminate thyroid nodules: a prospective analysis of 1,056 FNA samples. *J Clin Endocrinol Metab* 2011;96:3390–3397.
- Chudova D, Wilde JI, Wang ET, Wang H, Rabbee N, Egidio CM, Reynolds J, Tom E, Pagan M, Rigl CT, Friedman L, Wang CC, Lanman RB, Zeiger M, Kebebew E, Rosai J, Fellegara G, LiVolsi VA, Kennedy GC: Molecular classification of thyroid nodules using high-dimensionality genomic data. *J Clin Endocrinol Metab* 2010;95:5296–5304.