'Suspicious for Papillary Thyroid Carcinoma' before and after The Bethesda System for Reporting Thyroid Cytopathology: Impact of Standardized Terminology

Matthew T. Olson^a Thiraphon Boonyarunnate^{a, c} Aysegul Aksoy Atlinboga^d
Syed Z. Ali^{a, b}

Departments of ^aPathology and ^bRadiology, The Johns Hopkins Hospital, Baltimore, Md., USA; ^cDepartment of Pathology, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok, Thailand; ^dDepartment of Pathology, Division of Clinical Cytology, Ankara University School of Medicine, Ankara, Turkey

Key Words

Fine-needle aspiration · Thyroid · The Bethesda System for Reporting Thyroid Cytopathology · Suspicious for malignancy · Suspicious for papillary thyroid carcinoma

Abstract

Background: The high-risk 'suspicious for papillary thyroid carcinoma' (SPTC) is a clinically relevant diagnosis in the cytological interpretation of thyroid aspirates. While The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) has provided invaluable terminology standardization, a performance comparison for this diagnostic category has not been performed. Therefore, this study evaluates the SPTC diagnosis before and after the introduction of TBSRTC in a large meta-analysis and at a single institution. **Materials and Methods:** The meta-analysis analyzed publications of SPTC or similar diagnoses before and after the introduction of TBSRTC. Similarly our own institutional experience was analyzed for the 8 years surrounding the introduction of TBSRTC. A correlation of the cytopathology and surgical pathology diagnoses was performed. **Results:** The introduction of TBSRTC coincided with a significant decrease in the fraction of cases called SPTC in the meta-analysis (4.5–3.1%, $p < 0.00001$) and in the institutional review (1.7–0.9%,

$p = 0.005$). Meanwhile, the malignancy risk for those cases increased significantly in the meta-analysis from 62.5 to 80.5% ($p < 0.00001$) and trended upwards in the institutional review from 69 to 79% ($p = 0.4$). The follow-up rate was similar in both time periods in the meta-analysis and the institutional review. **Conclusions:** The introduction of TBSRTC coincided with a decrease in the fraction of cases called SPTC and an increase in the malignancy risk associated with that diagnosis.

© 2013 S. Karger AG, Basel

Introduction

Thyroid nodules are very common and usually benign [1]. However, since thyroid nodules may contain a malignancy, and the incidence of thyroid cancer appears to be increasing [2], thyroid nodules are routinely biopsied with fine-needle aspiration (FNA) [1, 3]. When the cytological diagnosis is malignant, the positive predictive value is greater than 99%, and when it is benign, the negative predictive value is typically less than 5% [4, 5]. While these performance characteristics are desirable, indeterminate categories are unavoidable [6], and confusion over nomenclature necessitated The Bethesda System for

Reporting Thyroid Cytopathology (TBSRTC) [5, 7]. Before TBSRTC, terminologies for indeterminate diagnostic categories varied significantly among institutions [1, 3, 8] even though it was clear that the subset of indeterminate aspirates with features that were worrisome but not completely diagnostic for malignancy had a significantly higher risk than those with focal and less worrisome atypical features [9–11]. Thus, the ‘suspicious for malignancy’ (SFM) category is an important component of TBSRTC and is mainly comprised of cases that are suspicious for papillary thyroid carcinoma (SPTC) but lack the necessary features to make a definite malignant diagnosis. While the category is similar in many ways to ‘suspicious’ categories that had been used in thyroid cytopathology prior to TBSRTC, a comparison of patients and outcomes with these diagnoses before and after TBSRTC has not been performed. The objective of the present study is to evaluate if the introduction of TBSRTC coincided with any change in the fraction of thyroid FNA called SPTC, the rate at which patients with that diagnosis underwent surgery, or the malignancy risk either in the published literature or at our institution.

Methods

Literature Review

A literature search was performed on the PubMed database with a boolean term including suspicious, malignancy, papillary thyroid carcinoma, cytopathology, and thyroid. Each paper was reviewed for the following variables: number of total biopsies, incidence of SPTC, number of SPTC cases with surgical follow-up, and number of resected SPTC cases with a malignant histological diagnosis. If a publication subdivided the SFM category into entities such as suspicious for medullary thyroid cancer or suspicious for lymphoma, then those non-SPTC cases were excluded. Some studies only considered the categorical diagnosis of SFM, and these cases were counted as SPTC cases given the rarity of other suspicious subcategories. The publications were classified as pre-TBSRTC if the studies were performed prior to the introduction of a risk-tiered system at that institution and as post-TBSRTC if they were performed after the introduction of a risk-tiered system at that institution. The studies were additionally classified as prospective and retrospective. A publication was classified as prospective if it reported diagnoses that were made preoperatively according to a predetermined diagnostic definition. A publication was classified as retrospective if it reported diagnoses that were made in a batch review of cases that were classified with a different classification scheme than the one in use at the time of original clinical evaluation.

Case Selection

After obtaining institutional review board approval, the pathology database was searched for the 6,998 thyroid FNA procedures performed at The Johns Hopkins Hospital over an 8-year period

from January 1, 2005, to December 31, 2012. This period of time included the 4 years prior to the adoption TBSRTC (January 1, 2005, to December 31, 2008) in which 3,364 procedures were performed and the 4 years after the adoption of TBSRTC (January 2, 2009, to December 31, 2012) in which 3,634 procedures were performed. These included some of the same patients already described in our preliminary report of the 26 months immediately after the institution of TBSRTC [4]. A small group ($n = 212$) of patients had biopsies both before and after January 1, 2009. Patients who were biopsied prior to January 1, 2009, and underwent thyroidectomy after this date were counted in the pre-TBSRTC group.

Cytopathology and Cytological-Histological Correlation

All cases were obtained at The Johns Hopkins Hospital under ultrasound guidance with on-site evaluation by either a cytopathologist or a cytotechnologist as has been described [12]. Paired direct smears were performed; one slide was stained with Diff-Quik® for on-site evaluation, and another slide was immediately immersed in 95% alcohol and stained with the Papanicolaou stain. Final diagnoses for all cases were made on both Diff-Quik-stained and Papanicolaou-stained direct smear slides by board-certified cytopathologists. In this study surgical follow-up was available for cytological-histological correlation in 89 cases. The preoperative size and site of the nodule were matched with the findings of the surgical pathology for each biopsy site. In a single case, a correlation could not be made due to unresolved documentation about where the biopsy was performed, and this case was excluded. Papillary microcarcinoma discovered incidentally away from the FNA site was not counted as malignant follow-up. Demographic data such as age, gender, and the size of the nodule were collected. Statistical analysis was performed using native functions in the statistical programming language R (<http://cran.r-project.org>). Continuous variables were analyzed with the Student t test, and binomial variables were analyzed with the χ^2 test or the Fisher exact test if any contingency contained a count of less than 30.

Results

Meta-Analysis of Cytopathological Data

Studies published before and after the adoption of TBSRTC were analyzed to investigate any differences in the fraction of cases diagnosed as SPTC, the fraction of cases diagnosed as SPTC that were followed up with surgery, and the malignancy rate of those cases. The articles and their findings are summarized in table 1. There were 13 studies that included 51,863 thyroid FNA in the pre-TBSRTC period and 12 studies that included 50,192 thyroid FNA in the post-TBSRTC period. These numbers are minimum estimates because, as shown in table 1, only some publications included the total number of patients seen during their study periods. All publications in both pre- and post-TBSRTC periods provided the number of SPTC cases and the malignant rate of resected SPTC cases. The fraction of cases diagnosed as SPTC that were ul-

Table 1. Pre- and post-TBSRTC studies of the SPTC category

First author (year)	Study type	Total cases, n	SPTC, n	Surgical follow-up, n	SPTC with surgical follow-up, n	Malignant rate, n
<i>Pre-TBSRTC studies</i>						
Lew [13], 2011	prospective	–	–	797	85	78 (91.8%)
Renshaw [14], 2001	prospective	1,085	27 (2.5%)	291 (26.8%)	16 (59.3%)	13 (81.3%)
Seningen [15], 2010	prospective	1,927	–	–	201	137 (68.1%)
Yassa [16], 2007	prospective	3,589	314 (8.7%)	1,242 (34.6%)	288 (91.7%)	173 (60.1%)
Punthakee [17], 2005	prospective	–	–	–	57	51 (89.5%)
Yang [18], 2007	prospective	4,703	124 (2.6%)	1,052 (22.4%)	105 (84.7%)	68 (64.8%)
Yoder [19], 2006	prospective	1,043	30 (2.9%)	200 (19.2%)	17 (56.7%)	11 (64.7%)
Goldstein [20], 2002	prospective	–	–	–	25	21 (84.0%)
Banks [21], 2008	prospective	–	–	–	150	91 (60.7%)
Raj [22], 2010	prospective	–	–	1,373	51	21 (41.2%)
Wu [23], 2006	prospective	1,621	42 (2.6%)	401 (24.7%)	19 (45.2%)	13 (68.4%)
Zhang [24], 2001	retrospective	–	–	57	25	17 (68.0%)
Ravetto [25], 2000	retrospective	37,895	–	4,069 (10.7%)	56	45 (80.4%)
Total		51,881	537	9,482	1,113	755 (67.8%)
<i>Post-TBSRTC studies</i>						
Theoharis [11], 2009	prospective	2,468	39 (1.6%)	378 (15.3%)	30 (76.9%)	26 (86.7%)
Castro [26], 2011	prospective	7,039	–	462 (6.6%)	136	101 (74.3%)
Deandrea [27], 2010	prospective	–	–	927	91	77 (84.6%)
Raparia [28], 2009	prospective	3,997	–	–	34	29 (85.3%)
Bongiovanni [9], 2012	prospective	7,686	266 (3.5%)	1,720 (22.4%)	222 (83.5%)	173 (77.9%)
Mastorakis [29], 2013	prospective	7,795	–	–	69	65 (94.2%)
Tepeoğlu [30], 2013	prospective	1,021	36 (3.5%)	219 (21.4%)	35 (97.2%)	32 (91.4%)
Ozluk [31], 2011	prospective	3,444	–	442 (12.8%)	42	34 (81.0%)
Nayar [32], 2009	retrospective	5,194	97 (1.9%)	986 (19.0%)	83 (85.6%)	44 (53.0%)
Jo [33], 2010	retrospective	3,080	71 (2.3%)	892 (29.0%)	56 (78.9%)	39 (69.6%)
Wu [34], 2011	retrospective	1,382	36 (2.6%)	221 (16.0%)	18 (50.0%)	12 (66.7%)
Renshaw [10], 2010	retrospective	7,089	123 (1.7%)	1,331 (18.8%)	73 (59.3%)	72 (98.6%)
Total		50,195	676	7,578	892	706 (79.1%)

timately malignant after resection was significantly higher in the post-TBSRTC period than in the pre-TBSRTC period. The malignancy rate for SPTC cases in the pre-TBSRTC period was 67.8% (739/1,095, range 41.2–91.8), and it was 79.1% (704/889, range 53–97.3) in the post-TBSRTC period ($p < 0.00001$).

Five [14, 16, 18, 19, 23] of the pre-TBSRTC studies and 6 [9–11, 32–34] of the post-TBSRTC studies reported the total number of thyroid FNA cases along with the fraction of them that were diagnosed as SPTC. The pre-TBSRTC publications involved 12,041 FNA biopsies of which 537 (4.5%, range 2.5–8.7) were diagnosed as SPTC. This is higher than the post-TBSRTC publications that involved 27,918 FNA of which 674 (2.4%, range 1.6–3.5) were diagnosed as SPTC. The difference was significant ($p <$

0.00001). Interestingly, these 2 time periods differed significantly in the rate of surgical follow-up. After the introduction of TBSRTC, the overall surgical follow-up rate for thyroid cases decreased significantly from 26.5% (3,186/12,041, range 19.2–34.6) to 20.6% (5,747/27,920, range 15.3–29.0, $p < 0.00001$), and the follow-up rate for cases diagnosed as SPTC similarly decreased from 82.9% (445/537, range 45.2–91.7) to 76.8% (519/676, range 50–97.2).

To correct for any bias that may have arisen from including studies in which cases were diagnosed retrospectively, the subset of studies that were diagnosed prospectively was considered separately. This group of publications included 5 pre-TBSRTC studies [14, 16, 18, 19, 23] that collectively involved at least 12,041 patients and 3

Table 2. Summary of prospective studies in the pre- and post-TBSRTC periods

	SPTC rate	Surgical follow-up rate of SPTC	Malignant rate
Pre-TBSRCT studies [14, 16, 18, 19, 23]	4.5% (537/12,041)	82.9% (445/537)	62.5% (278/445)
Post-TBSRCT studies [9, 11, 30]	3.1% (341/11,175)	84.2% (287/341)	80.5% (231/287)
p	<0.00001	0.61508	<0.00001

Table 3. Histological correlation for nodules diagnosed preoperatively as SPTC

First author (year)	Malignant	FC	HCC	PTC	FV-PTC	Other
<i>Histological correlation for malignant cases from studies before TBSRTC</i>						
Lew [13], 2011	78	–	–	58	18	2
Renshaw [14], 2001	13	–	–	13	–	–
Seningen [15], 2010	129	2	–	108	18	1
Yassa [16], 2007	173	–	–	171	–	2
Punthakee [17], 2005	51	–	–	51	–	–
Yang [18], 2007	68	5	–	63	–	–
Yoder [19], 2006	11	–	–	8	3	–
Goldstein [20], 2002	21	1	–	15	3	2
Banks [21], 2008	91	2	2	60	26	1
Raj [22], 2010	21	2	–	18	–	1
Wu [23], 2006	13	–	–	8	4	1
Zhang [24], 2001	17	1	–	16	–	–
Ravetto [25], 2000	45	21	–	19	–	5
Total	731	34	2	608	72	15
<i>Histological correlation for malignant cases from studies after TBSRTC</i>						
Theoharis [11], 2009	26	–	–	26	–	–
Castro [26], 2011	96	2	–	77	16	1
Deandrea [27], 2010	77	7	9	31	13	17
Raparia [28], 2009	29	1	2	26	–	–
Mastorakis [29], 2013	65	–	–	63	–	2
Tepeoglu [30], 2013	32	1	–	31	–	–
Nayar [32], 2009	44	–	1	40	–	3
Jo [33], 2010	39	3	–	25	8	3
Total	414	14	12	319	37	26
FC = Follicular carcinoma; HCC = Hürthle cell carcinoma.						

post-TBSRTC studies [9, 11, 30] that collectively involved at least 11,175 patients. The summary statistics for these studies are shown in table 2. Similar to the meta-analysis including all studies, the limited set of prospectively diagnosed studies showed that the rate of SPTC

diagnosis decreased significantly from 4.5% (537/12,041, range 2.5–8.7) to 3.1% (341/11,175, range 1.6–3.5, $p < 0.00001$). Also similar to the meta-analysis that included all studies, the malignancy rate for the publications that reported prospectively diagnosed cases increased from 62.5% (278/445, range 60.1–81.3) to 80.5% (231/287, range 77.9–91.4, $p < 0.00001$). Interestingly, after removing the retrospectively diagnosed cases, the surgery rate actually increased marginally from 82.9% (445/537, range 45.2–91.7) to 84.2% (287/341, range 76.9–97.2) despite an overall decrease in the percent of patients who underwent surgery – in the prospectively diagnosed studies, the overall surgical follow-up rate was 26.5 (3,186/12,041, range 19.2–34.6), and it was 20.7 (2,317/11,175, range 15.3–22.4) after TBSRTC. While the surgical follow-up rate for cases diagnosed as SPTC before and after TBSRTC was not significantly different ($p = 0.7$), it was substantially different from the dramatic decrease seen when all the publications were considered. Additionally, the marginal increase in the surgical follow-up rate for cases diagnosed as SPTC contrasts starkly with the clear decrease in the overall surgical follow-up rate observed among the same studies.

Meta-Analysis of Cytopathological-Histopathological Correlation Data

To better understand the malignancy rate of the SPTC diagnosis before and after the introduction of TBSRTC, the malignant histological diagnoses were examined. Thirteen studies in the pre-TBSRTC [13–25] and 8 studies in the post-TBSRTC [11, 26–30, 32, 33] provided the specific histopathological diagnoses for their cases with malignant follow-up, and these are detailed in table 3. From this data, the rate of the follicular variant of papillary thyroid carcinoma (FV-PTC) as a fraction of all malignancy was equivalent, i.e. 9.8% (72/731, range 0–30.8) before and 8.9% (37/414, range 0–20.5) after TBSRTC ($p = 0.73613$). This equivalence is meaningful because the overall malignancy rate increased, at a fraction significantly greater than the change in FV-PTC ($p = 0.00073$).

Table 4. Demographic data from the institutional study

Variable	Pre-TBSRTC	Post-TBSRTC	p value
Male sex, %	33.9 (19/56)	30.3 (10/33)	0.72447
Mean age (range), years	51.4 (19–85)	50 (15–87)	0.77644
Nodule size, cm	2.3 ± 1.7	2 ± 1.7	0.40685
SPTC rate, %	1.7 (56/3,362)	0.9 (33/3,629)	0.00477
Surgical follow-up, %	15.8 (530/3,362)	12.1 (441/3,629)	0.00001
Surgical follow-up of SPTC, %	75 (42/56)	78.8 (26/33)	0.7
Malignant surgical pathology, %	69.0 (29/42)	78.6 (22/28)	0.4

The increase in the malignancy rate was commensurate with the increase in cases that ultimately turned out to be classical PTC on resection ($p = 0.44$).

Institutional Experience

The review of our institutional experience covered 2 time periods: the 4 years before the adoption of TBSRTC in which 3,362 biopsies were performed and the 4 years after the adoption of the TBSRTC in which 3,629 were performed. There were 46 cases of SPTC in pre-TBSRTC periods and 29 cases in post-TBSRTC periods. These comprised 1.7% (46/3,362) and 0.9% (29/3,629) of the cases before and after TBSRTC, respectively. The decrease was statistically significant ($p = 0.00477$). As seen in table 4, there were no significant differences in the basic demographic features of patients before and after the introduction of TBSRTC. The introduction of TBSRTC coincided with an insignificant upward trend in the surgical follow-up rate for cases diagnosed as SPTC (78 vs. 86%, $p = 0.5$); the upward trend was remarkable given the significant decrease in overall surgical follow-up from 15.8% (530/3,362) to 12.1% (441/3,629, $p = 0.00001$).

Paradoxically, this decrease in the SPTC call rate coincided with a significant increase in the fraction of thyroid FNA cases that were diagnosed as malignant from 3.4% (115/3,362) to 4.6% (167/3,629, $p = 0.014$) and an increase in the fraction of cases that were diagnosed as atypia of undetermined significance (AUS) from 9.1% (305/3,364) to 11.5% (419/3,640, $p = 0.001$). The increase in the AUS rate was statistically equivalent to the increase in the malignancy call rate ($p = 0.8$). As seen in table 5, the call rate for non-SPTC categories within SFM was low and equivalent for both time periods. The malignancy rate for the SPTC category after the introduction of TBSRTC was similar to the malignancy rate before TBSRTC, i.e. 69% (29/42) and 68.6% (22/28, $p = 0.4$). The fraction of PTC-

Table 5. Institutional histologic diagnoses

Time period	Interpretation	Follow-up	Histology
Pre-TBSRTC (n = 56)	SFM: 2	no follow-up: 1 benign: 0 malignant: 1	– – PTC-FV 1
	SLYMPH: 6	no follow-up: 5 benign: 1 malignant: 0	– HT 1 –
	SMTC: 2	no follow-up: 0 benign: 0 malignant: 2	– – PTC-FV 1, MTC 1
	SPTC: 46	no follow-up: 8 benign: 12 malignant: 26	– AN 9, HT 2, FA 1 PTC 15, PTC-FV 11
Post-TBSRTC (n = 33)	SFM: 2	no follow-up: 1 benign: 0 malignant: 1	– – PD-FC 1
	SLYMPH: 1	no follow-up: 1 benign: 0 malignant: 0	– – –
	SMTC: 1	no follow-up: 0 benign: 0 malignant: 1	– – HCC 1
	SPTC: 29	no follow-up: 4 benign: 5 malignant: 20	no follow-up: 3 unable to correlate: 1 AN 2, HT 1, FA 1, HTA 1 PTC 13, PTC-FV 6, PTC-micro 1

SLYMPH = Suspicious for lymphoma; SMTC = suspicious for medullary thyroid carcinoma; HT = Hashimoto thyroiditis; MTC = medullary thyroid carcinoma; AN = adenomatous nodule; FA = follicular adenoma; PD-FC = poorly differentiated follicular carcinoma; HCC = Hürthle cell carcinoma; HTA = hyalinizing trabecular adenoma; PTC-micro = papillary thyroid microcarcinoma.

FV cases was subjectively similar, although our institutional study was not sufficiently large to detect a difference in the PTC-FV rate.

Discussion

The practical reason why SPTC is part of TBSRTC is to preserve the high positive predictive value of the malignant category by providing the cytopathologist with a less-binding diagnosis when some of the cytological criteria of malignancy are missing. On an intuitive level, SPTC is a necessary diagnostic category because of the morphological heterogeneous nature of thyroid tumors and the intrinsic incompleteness of sampling that occurs in FNA. The findings of both a decrease in the SPTC call rate and the increase in the SPTC malignancy rate parallel the consensus in TBSRTC that the malignant category should have a high positive predictive value and the consensus on the AUS category that provides a lower-risk category for compromised specimens. Given that many of the papers reviewed here did not report the incidences of all categories, it is not clear from the meta-analysis how much the existence of the AUS category led to the decrease in the SFM call rate and the increase in the predictive value.

As in the meta-analysis, in our institutional experience the rate of SPTC decreased and the predictive rate for malignancy increased. Paradoxically, this coincided with a proportionate increase in both the fraction of cases that were cytologically malignant and the fraction of cases that were diagnosed as AUS. If the decrease in the SPTC call rate were solely attributable to the consensus acceptance of AUS, then the increase in the AUS rate would be expected to be greater than the increase in the malignancy rate. As this is not the case, it is unlikely that the decrease in the SFM call rate is solely attributable to cytopathologist factors. An alternative explanation could be that the rate of BRAF mutation-associated PTC is increasing [35], and the classical morphological findings in PTC are more pronounced in these tumors [36]. While there is no evidence in this retrospective review to test such a hypothesis, the influence of such environmental and epidemiological factors on the morphology and the performance of FNA in the thyroid cannot be eliminated and may have potential indications for the future of ancillary test development and use.

Before TBSRTC, various diagnostic terminologies were used to describe what is now called SFM. These included 'atypical follicular proliferation, highly suspicious

for papillary thyroid carcinoma', 'thyroid neoplasm with features likely consistent with papillary thyroid carcinoma', and descriptive diagnoses. While the standardization of terminology within TBSRTC is obvious, the relationship between that standardization and performance metrics in published clinical or pathological practice is much less obvious. Interestingly, consolidating the different terminologies into the SPTC category led to a decrease in the use of that category. Our meta-analysis shows that the introduction of TBSRTC coincided with a decrease in the rate at which thyroid FNA biopsies were diagnosed as SFM but an increase in the fraction of those cases that were malignant. The impact on clinical management appears unchanged – most of the patients with this diagnosis underwent surgery regardless of the terminology with which they were diagnosed. There are two possible explanations for the decreased rate of SFM and the increased malignancy rate that we have shown in the meta-analysis. The first is that standardized terminology, especially the establishment of AUS, allowed for a seemingly less risky way to report specimens that lack the full complement of malignant findings. The second possible explanation is that SFM is a surrogate for non-classical genetic variants of PTC that are decreasing in abundance. The exact contribution of either of these two possibilities, if any, remains unclear.

The meta-analysis presented here intentionally separates studies reporting prospectively and retrospectively diagnosed material. This is done because studies with retrospectively assigned diagnoses introduce bias into the some of the most important features of the data – overall and cytological diagnosis-specific surgical follow-up rates. This is not surprising from a practical perspective since this bias is the product of other basic sources of bias. First, as the studies involve reviewing slides from distant history, more material is available in the study than at the time of the original read – the resection diagnosis may even be known to the person reviewing the case. Second, retrospective reviews of multiple years of cases within a limited period of study time inherently lead to intrastudy training and fatiguing. This also occurs in prospectively diagnosed cases but to a much smaller extent due to the necessarily higher attention given in the first-time legally binding diagnostic process and the fact that multiple years worth of cases are interpreted over a period of multiple years rather than concentrated. Finally, retrospective reviews that are limited to enrich a particular diagnostic category intrinsically contain selection bias, and that is significant enough to lead to artificial estimates of follow-up disease. Despite these shortcomings, retrospec-

tively reviewed case studies are a chief source of the cytopathology knowledge base since they enable a first and safe validation of diagnostic criteria and categories. Also, when they are performed in an exhaustive or sizeable randomly selected set, they provide a first approximation of the expected incidences and problems in new categories. When performed methodically, they can even lead to the formulation of diagnostic criteria [37]. As such, series involving retrospectively reviewed material were included in the meta-analysis for purposes of estimating how diagnostic categories are distributed. However, the biases involved in retrospectively reviewed studies led to some significantly misleading results – especially those relating to the surgical follow-up. Accordingly, all of our conclusions are derived from the prospectively reviewed studies.

The meta-analysis and institutional study have the usual unavoidable biases of retrospective reviews: selection bias, referral bias, and the contribution of unknown confounding variables. The greatest limitation in this is study the lack of universal surgical follow-up. As a result

of this limitation, the natural incidence of disease within this patient population remains unknown, and it is unclear if the changes described here are due to the consensus terminology in TBSRTC or underlying shifts in the molecular epidemiology of PTC. As an equivalent fraction of patients with a diagnosis of SFM underwent surgery in both time periods and in both the meta-analysis and the institutional study, it is likely that selection biases possibly introduced by a less-than-perfect surgical follow-up rate are minimal.

Conclusion

Standardized terminology in TBSRTC coincided with a decrease in the use of the SFM category and an increase in the fraction of cases with that diagnosis that were ultimately found to be malignant. These findings were consistent in both the meta-analysis and the institutional review.

References

- Cooper DS, Doherty GM, Haugen BR, Hauger 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.
- Cooper JS, Porter K, Mallin K, Hoffman HT, Weber RS, Ang KK, Gay EG, Langer CJ: National Cancer Database report on cancer of the head and neck: 10-year update. *Head Neck* 2009;31:748–758.
- Gharib H, Goellner JR: Fine-needle aspiration biopsy of the thyroid: an appraisal. *Ann Intern Med* 1993;118:282–289.
- Olson MT, Clark DP, Erozan YS, Ali SZ: Spectrum of risk of malignancy in subcategories of 'atypia of undetermined significance'. *Acta Cytol* 2011;55:518–525.
- Cibas ES, Ali SZ: The Bethesda System for Reporting Thyroid Cytopathology. *Thyroid* 2009;19:1159–1165.
- Olson MT, Boonyaarunnate T, Aragon Han P, Umbricht CB, Ali SZ, Zeiger MA: A tertiary center's experience with second review of 3,885 thyroid cytopathology specimens. *J Clin Endocrinol Metab* 2013;98:1450–1457.
- Ali SZ: Thyroid cytopathology: Bethesda and beyond. *Acta Cytol* 2011;55:4–12.
- Guidelines of the Papanicolaou Society of Cytopathology for fine-needle aspiration procedure and reporting – the Papanicolaou Society of Cytopathology Task Force on Standards of Practice. *Diagn Cytopathol* 1997;17:239–247.
- Bongiovanni M, Krane JF, Cibas ES, Faquin WC: The atypical thyroid fine-needle aspiration: past, present, and future. *Cancer Cytopathol* 2012;120:73–86.
- Renshaw AA: Should 'atypical follicular cells' in thyroid fine-needle aspirates be subclassified? *Cancer Cytopathol* 2010;118:186–189.
- Theoharis CGA, 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.
- Olson MT, Tatsas AD, Ali SZ: Cytotechnologist attended on-site adequacy evaluation of thyroid fine needle aspiration: comparison with cytopathologists and correlation with the final interpretation. *Am J Clin Pathol* 2011;138:90–95.
- Lew JI, Snyder RA, Sanchez YM, Solorzano CC: Fine needle aspiration of the thyroid: correlation with final histopathology in a surgical series of 797 patients. *J Am Coll Surg* 2011;213:188–194, discussion 194–195.
- Renshaw AA: Accuracy of thyroid fine-needle aspiration using receiver operator characteristic curves. *Am J Clin Pathol* 2001;116:477–482.
- Senigen JL, Nassar A, Henry MR: Correlation of thyroid nodule fine-needle aspiration cytology with corresponding histology at Mayo Clinic, 2001–2007: an institutional experience of 1,945 cases. *Diagn Cytopathol* 2012;40(suppl 1):E27–E32.
- Yassa L, Cibas ES, Benson CB, Frates MC, Doubilet PM, Gawande AA, Moore FD, Kim BW, Nosé V, Marqusee E, Larsen PR, Alexander EK: Long-term assessment of a multidisciplinary approach to thyroid nodule diagnostic evaluation. *Cancer* 2007;111:508–516.
- Punthakee X, Palme CE, Franklin JH, Zhang I, Freeman JL, Bedard YC: Fine-needle aspiration biopsy findings suspicious for papillary thyroid carcinoma: a review of cytopathological criteria. *Laryngoscope* 2005;115:433–436.
- Yang J, Schnadig V, Logrono R, Wasserman PG: Fine-needle aspiration of thyroid nodules: a study of 4703 patients with histologic and clinical correlations. *Cancer* 2007;111:306–315.
- 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.
- Goldstein RE, Nettekville JL, Burkey B, Johnson JE: Implications of follicular neoplasms, atypia, and lesions suspicious for malignancy diagnosed by fine-needle aspiration of thyroid nodules. *Ann Surg* 2002;235:656–662, discussion 662–664.

- 21 Banks ND, Kowalski J, Tsai H-L, Somervell H, Tufano R, Dackiw APB, Marohn MR, Clark DP, Umbricht CB, Zeiger MA: A diagnostic predictor model for indeterminate or suspicious thyroid FNA samples. *Thyroid* 2008;18: 933–941.
- 22 Raj MD, Grodski S, Martin SA, Yeung M, Serpell JW: The role of fine-needle aspiration cytology in the surgical management of thyroid cancer. *ANZ J Surg* 2010;80:827–830.
- 23 Wu HH, Jones JN, Osman J: Fine-needle aspiration cytology of the thyroid: ten years experience in a community teaching hospital. *Diagn Cytopathol* 2006;34:93–96.
- 24 Zhang Y, Fraser JL, Wang HH: Morphologic predictors of papillary carcinoma on fine-needle aspiration of thyroid with ThinPrep preparations. *Diagn Cytopathol* 2001;24: 378–383.
- 25 Ravetto C, Colombo L, Dottorini ME: Usefulness of fine-needle aspiration in the diagnosis of thyroid carcinoma. *Cancer* 2001;90:357–363.
- 26 Castro MR, Espiritu RP, Bahn RS, Henry MR, Gharib H, Caraballo PJ, Morris JC: Predictors of malignancy in patients with cytologically suspicious thyroid nodules. *Thyroid* 2011;21: 1191.
- 27 Deandrea M, Ragazzoni F, Motta M, Torchio B, Mormile A, Garino F, Magliona G, Gamarra E, Ramunni MJ, Garberoglio R, Limone PP: Diagnostic value of a cytomorphological subclassification of follicular patterned thyroid lesions: a study of 927 consecutive cases with histological correlation. *Thyroid* 2010;20: 1077–1083.
- 28 Raparia K, Min SK, Mody DR, Anton R, Amrikachi M: Clinical outcomes for ‘suspicious’ category in thyroid fine-needle aspiration biopsy: patient’s sex and nodule size are possible predictors of malignancy. *Arch Pathol Lab Med* 2009;133:787–790.
- 29 Mastorakis E, Meristoudis C, Margari N, Pouliakis A, Leventakos K, Chroniaris N, Panayiotides I, Karakitsos P: Fine needle aspiration cytology of nodular thyroid lesions: a 2-year experience of the Bethesda system for reporting thyroid cytopathology in a large regional and a university hospital, with histological correlation. *Cytopathology* 2013, DOI: 10.1111/cyt.12062.
- 30 Tepeoğlu M, Bilezikçi B, Bayraktar SG: A histological assessment of the Bethesda system for reporting thyroid cytopathology (2010) abnormal categories: a series of 219 consecutive cases. *Cytopathology* 2013, DOI: 10.1111/cyt.12051.
- 31 Ozluk Y, Pehlivan E, Gulluoglu MG, Poyanli A, Salmaslioglu A, Colak N, Kapran Y, Yilmazbayhan D: The use of the Bethesda terminology in thyroid fine-needle aspiration results in a lower rate of surgery for nonmalignant nodules: a report from a reference center in Turkey. *Int J Surg Pathol* 2011;19: 761–771.
- 32 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.
- 33 Jo VY, Renshaw AA, Krane JF: Relative sensitivity of thyroid fine-needle aspiration by tumor type and size. *Diagn Cytopathol* 2013, DOI: 10.1002/dc.22976.
- 34 Wu HH, 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.
- 35 Mathur A, Moses W, Rahbari R, Khanafshar E, Duh Q-Y, Clark O, Kebebew E: Higher rate of BRAF mutation in papillary thyroid cancer over time: a single-institution study. *Cancer* 2011;117:4390–4395.
- 36 Kleiman DA, Sporn MJ, Beninato T, Crowley MJ, Nguyen A, Uccelli A, Scognamiglio T, Zarnegar R, Fahey I, Thomas J: Preoperative BRAF(V600E) mutation screening is unlikely to alter initial surgical treatment of patients with indeterminate thyroid nodules: a prospective case series of 960 patients. *Cancer* 2013;119:1495–1502.
- 37 Vandenbussche CJ, Sathiyamoorthy S, Owens CL, Burroughs FH, Rosenthal DL, Guan H: The Johns Hopkins Hospital template for urologic cytology samples. 2, 3. Improving the predictability of indeterminate results in urinary cytologic samples: an outcomes and cytomorphologic study. *Cancer Cytopathol* 2013;121:21–28.