Published in final edited form as:

Am Surg. 2014 August ; 80(8): 811–816.

False-negative Results with the Bethesda System of Reporting Thyroid Cytopathology: Predictors of Malignancy in Thyroid Nodules Classified as Benign by Cytopathologic Evaluation

Bryan K. Richmond, M.D., M.B.A.^{*}, Rudy Judhan, M.D.^{*}, Benny Chong, M.D.^{*}, Adam Ubert, M.D.^{*}, Zachary Aburahma, M.S.-IV[†], William Mangano, M.D.[‡], and Stephanie Thompson, PH.D.[§]

^{*}Department of Surgery, West Virginia University/Charleston Division, Charleston West Virginia

[†]West Virginia School of Osteopathic Medicine, Lewisburg, West Virginia

[‡]Department of Pathology, Charleston Area Medical Center, Charleston, West Virginia

[§]Charleston Area Medical Center Health Education and Research Institute, Charleston, West Virginia

Abstract

The benign category of the Bethesda System for reporting thyroid cytopathology (BSRTC) predicts an incidence of malignancy from zero to three per cent. However, recent series report higher rates of malignancy ranging from eight to 14 per cent. Surgery is often performed for reasons other than their fine needle aspiration biopsy (FNAB) such as symptoms, nodule enlargement, or worrisome imaging. We hypothesized that an analysis of patients who underwent thyroidectomy despite a benign FNAB would identify predictors of malignancy, an area not currently addressed by American Thyroid Association guidelines. We performed a retrospective analysis of patients with benign FNAB results who underwent thyroidectomy from October 2007 to October 2012. Data collected included symptoms, imaging findings, FNAB results, and operative and histopathology results, all of which were obtained by chart review. Findings were compared between patients with and without a diagnosis of malignancy. Statistical significance was set as $P < 0.05$. Of 3839 FNABs, 2838 were benign. Of these, 180 underwent surgery for indications other than the FNAB category. Twenty-four (13.4%) malignancies were identified: 12 (6.7%) incidental microcarcinomas and 12 (6.7%) significant cancers (papillary greater than 1.0 cm, any nonpapillary histology). No patient's symptoms or signs reached significance as a predictor of malignancy. Suspicious ultrasound appearance was significantly associated with an underlying carcinoma ($P = 0.004$). The false-negative result with benign FNAB is higher in surgical series than suggested by the BSRTC. Patients with tolerable symptoms may be observed in the face of a benign FNAB. Additionally, despite a benign FNAB, recommendations for closer follow-up or surgical intervention are warranted if the ultrasound appearance is suspicious.

Address correspondence and reprint requests to Bryan K. Richmond, M.D., M.B.A., Professor of Surgery and Section Chief—General Surgery, West Virginia University/Charleston Division, Room 3023, 3110 MacCorkle Avenue, Charleston, WV 25304. brichmond@hsc.wvu.edu.

Presented at the Annual Scientific Meeting and Postgraduate Course Program, Southeastern Surgical Congress, Savannah, GA, February 22–25, 2014.

The benign category of the Bethesda System for reporting thyroid cytopathology (BSRTC) predicts an incidence of malignancy from zero to three per cent.¹ However, recent series have reported higher rates of malignancy in surgical series with rates ranging from six to 14 per cent.²⁻⁷ Surgery is often performed in these patients for reasons other than their fine needle aspiration biopsy (FNAB) findings, and these reasons include associated symptoms, cosmesis, nodular enlargement, risk factors for thyroid carcinoma, or worrisome ultrasound imaging findings such as microcalcifications, hypoechogenicity, hypervascularity, irregularity, or a “taller than wide” appearance.⁸ Knowledge about the predictive value of these factors would be of value in determining which patients with a benign FNAB result warrant closer follow-up or surgical management of their thyroid nodules.

We analyzed a population of patients who underwent thyroidectomy despite a benign FNAB with the goal being to identify predictors of malignancy in patients found to have a benign FNAB by BSRTC criteria. If predictive factors in this subpopulation could be identified, then they could be used to select patients for surgery because of their cancer risk despite their benign FNAB. This management decision is not currently addressed by American Thyroid Association (ATA) guidelines with respect to the treatment of thyroid nodules.⁸

Methods

This was a retrospective analysis of all individuals 18 years of age or older who 1) presented with a thyroid nodule and underwent a FNAB, which was classified as benign using the BSRTC; and 2) then subsequently underwent thyroid surgery at our institution for any reason other than the FNAB results. All aspects of the study were reviewed and approved by the Charleston Area Medical Center (CAMC)/West Virginia University, Charleston Division Institutional Review Board. Patients included in the study were identified by CAMC’s pathology department as having a thyroid FNAB classified as “benign” by the BSRTC during the study period. All included patients underwent thyroid surgery (lobectomy, near total thyroidectomy, or total thyroidectomy) at CAMC between October 1, 2008, and October 1, 2012. Both the FNAB results and the final pathology report on the resected surgical specimen must have been available for review for patient inclusion in the study.

For patients meeting inclusion criteria, basic demographics were collected, which included age, sex, and race. In addition, any family history of thyroid cancer or personal history of radiation exposure was noted. Other collected clinical data collected included symptoms present or signs present such as pain, hoarseness, or dysphagia. Any progressive enlargement of the nodule and the indication for and the results of the FNAB were recorded as well as whether the FNAB was obtained under ultrasound guidance. The results of any preoperative imaging studies (ultrasound, nuclear scanning, computerized tomographic [CT] scans) and the presence of any suspicious imaging findings such as microcalcifications, hypervascularity, irregularity, hypoechogenicity, and a “taller than wide appearance” were noted.⁸ The actual indication for surgery was also recorded. The final histopathology report was used to define the overall incidence of malignancy and the incidence of clinically significant cancers. We defined “clinically significant” thyroid carcinomas as papillary

carcinomas with tumor size 1.0 cm or larger, papillary carcinomas that were multifocal or high-risk histologic type, or cancers of any nonpapillary histologic type (such as follicular, medullary, or anaplastic carcinomas). Unifocal papillary microcarcinomas (less than 1.0 cm) were not considered within the “clinically significant” subgroup.

Mean and standard deviation for continuous variables and proportion and frequency for categorical variables were used to describe patient and procedure characteristics. Fisher’s exact (categorical data) and the Mann-Whitney *U* test (continuous variables) were used to evaluate associations between predictor variables and clinically significant thyroid malignancy. Statistical significance was set as a *P* value < 0.05. Statistical analysis was performed using IBM SPSS Statistics for Windows, Version 19.0 (IBM, Armonk, NY).

Results

Of 3839 thyroid FNABs obtained during the study period, 2838 (74%) aspirates were classified as meeting criteria for classification within the “benign” category of the BSRTC. In all, 180 (6.3% of all benign FNABs) patients subsequently underwent thyroid surgery for indications other than the FNAB findings. Twenty-four of the 180 patients were found to have a thyroid malignancy on final pathology (13.4%), of which 12 (6.7%) were incidental papillary microcarcinomas and 12 (6.7%) were clinically significant cancers. Of the 12 clinically significant carcinomas, six were papillary carcinomas (greater than 1.0 cm), four were follicular variants of papillary carcinoma, and two were medullary carcinomas. Final diagnoses present in patients not found to have thyroid malignancy included adenomatoid or hyperplastic nodule(s) (45%), multinodular goiter (39%), Hashimoto’s thyroiditis (19%), colloid nodule(s) (10%), and follicular adenoma (3%).

The median patient age was 55 years (range, 18 to 87 years) with 86 per cent of patients being female. Patient gender, age, or presence of pre-existing thyroid disease did not differ between patients with and without thyroid carcinomas ($P > 0.20$ for all comparisons; Table 1). The overall incidence of patients with a family history of thyroid cancer or previous radiation exposure was low (6 and 3%, respectively) and did not differ across cohorts ($P = 1.00$) (Table 1).

Ultrasound guidance for thyroid FNAB was used in 75 per cent of the patients with carcinomas *versus* 47 per cent in patients with benign pathological findings ($P = 0.07$), which trended toward, but ultimately failed to reach, statistical significance. The use of imaging modalities (ultrasound, nuclear scan, CT scan) did not differ among the two groups (benign *vs* malignant) with ultrasound imaging being the most common modality and used in 100 per cent of patients (Table 2).

In our series, the most common symptoms serving as surgical indications included dysphagia (43.9%), enlarging nodule (38.3%), patient request (27.2%), hoarseness (16.1%), and cosmesis (2.2%). Multiple symptoms were present in 51 per cent of patients. All symptoms identified as indications for surgery failed to reach significance as predictors of malignancy, regardless of how many individual symptoms were present ($P = 0.11$ for all symptoms collectively) (Table 3).

Suspicious ultrasound imaging findings was the indication for surgery in 21.1 per cent of patients. Although other imaging studies were obtained in a minority of the patients (nuclear scan 30%, CT scan 11%) (Table 2), only a suspicious ultrasound appearance was used as an indication for surgery in our series. Notably, nine of the 12 patients (75%) subsequently determined to have clinically significant thyroid carcinomas had suspicious findings on ultrasound imaging compared with 56 of 168 (33.3%) patients not found to have an underlying thyroid malignancy ($P = 0.01$) (Table 2). On univariate analysis, suspicious ultrasound appearance as an indication for surgery was significantly associated with the presence of an underlying carcinoma ($P = 0.004$) (Table 3).

Discussion

The diagnostic workup and subsequent treatment of thyroid nodules is one of the most common clinical scenarios encountered in the practice of endocrine surgery. FNAB remains the diagnostic test of choice for stratification of risk and selection of treatment modality, be it surgery, repeat biopsy, or observation with serial examinations and ultrasound examinations.^{1, 8} The BSRTC is a six-tiered reporting scheme that has gained favor among clinicians as a result of its standardized reporting and reproducibility of results in the clinical setting. In addition, the BSRTC includes a treatment algorithm that is based on the risk of malignancy in a given reporting category (Fig. 1).¹

The “benign” category of the BSRTC is reported to carry a malignancy risk (false-negative rate) of zero to three per cent.¹ Multiple recent surgical series however have reported substantially higher false-negative (FN) rates, ranging from eight to 14 per cent.^{2–7} It is logical to conclude that other factors were present that led the surgeons to operate on what was determined to be a cytologically benign nodule such as the presence of concerning symptoms (such as hoarseness or dysphagia), risk factors for thyroid carcinoma, enlargement of the nodule, or suspicious imaging findings. The 2009 Revised ATA Management Guidelines for Patients with Thyroid Nodules and Differentiated Thyroid Cancer provide specific guidelines as to which patients should undergo FNAB, which include, among others, the previously mentioned symptoms and nodule size.⁸ However, all treatment recommendations in the ATA guidelines subsequent to the FNAB are based solely on the FNAB results with two notable exceptions. The first of these is that of an enlarging nodule, in which a rebiopsy or resection is recommended, and the second being that of a recurrent cyst after repeated aspiration, in which a resection or ablation is recommended. The ATA guidelines call for observation of cytologically benign nodules in all other clinical circumstances.⁸ The higher FN rate in surgical series would suggest that other factors may be present in these patients that, if identified, could predict a higher risk of malignancy when present and be used to select which patients should undergo surgery despite a benign FNAB.

In expert hands, thyroid surgery is extremely safe with a major complication rate of approximately one per cent.⁹ Subsequently, in modern endocrine surgery practice, patients with benign FNAB results are often offered surgery for their thyroid nodules for a variety of other indications. These may include the presence of symptoms, risk factors for thyroid carcinoma (such as family history or radiation exposure), or suspicious imaging findings. In our series, the most common surgical indications included the presence of dysphagia

(43.9%), an enlarging nodule (38.3%), and patient request (27.2%). Other indications included abnormal (suspicious) imaging findings (21.1%), hoarseness (16.1%), and cosmesis (2.2%). No clear indication for surgery could be ascertained from the operative report in 5.6 per cent, which is a limitation of this retrospective analysis. Analysis of our data revealed that the presence of symptoms had no predictive value with respect to the presence of an underlying thyroid carcinoma. This was the case for all individual symptoms and regardless of the number of symptoms present (Table 3). These data would suggest that symptomatology is an unreliable indicator of an underlying thyroid carcinoma and that patients with symptoms may be safely observed if the symptoms are tolerable. This is consistent with other published literature indicating a poor correlation between patient symptoms and the presence of an underlying thyroid malignancy.¹⁰ It is important to note, however, that this specifically refers to the presence of symptoms (such as hoarseness or dysphagia) rather than objective signs such as vocal cord paralysis or a fixed neck mass, which more are strongly correlated with the presence of an underlying carcinoma.^{11, 12} Additionally, the absolute size of a thyroid nodule has also been correlated with both FN FNAB results and increased risk of malignancy with larger nodules displaying both higher FN FNAB results and higher rates of underlying malignancy.^{13, 14}

Abnormal imaging studies are referenced in the ATA guidelines as an indication to proceed with FNAB. It is recommended that all solid nodules measuring greater than 1 cm undergo FNAB. It is further recommended that patients with smaller nodules (5 mm) undergo biopsy as well if suspicious ultrasound findings are present such as microcalcifications, hypoechogenicity, an irregular border, increased vascularity, or a “taller than wide” appearance to name a few.^{8, 15} Although these findings may be used to justify FNAB of thyroid nodules, the subsequent treatment recommendations provided by the ATA guidelines are based on the results obtained from the FNAB.⁸ Our data found that the presence of an abnormal ultrasound appearance was significantly associated with the presence of an underlying thyroid carcinoma despite a benign FNAB result. Consequently, these data would suggest that it may be appropriate to consider closer follow-up or surgery in these patients based on the sonographic appearance of the nodule even if the results of the FNAB are benign.

Limitations of this study include its retrospective nature and small sample size. Despite review of 3839 FNABs, of which 2838 were benign, only 180 patients subsequently underwent surgery. Of these, 24 carcinomas were identified (12 incidental, 12 clinically significant). Clearly, a larger study group would serve to strengthen the subsequent conclusions drawn. Although prospectively collected data would certainly be of greater use, the logistics of obtaining prospective data on the natural history of thyroid nodules is daunting owing to the indolent nature of differentiated thyroid cancers. In addition, this study did not include patients who did not undergo FNAB but rather were taken directly to surgery as a result of severe symptoms, physical findings suggestive of malignancy, risk factors for thyroid carcinoma, or the previously specified abnormal imaging characteristics. It is also our practice to offer surgery to patients presenting with nodules of size greater than or equal to 4.0 cm based on the increased risk of carcinoma in these patients.¹⁴ We felt it inappropriate to include these patients in this study, because the focus of our project was assess the true FN rate of the BSRTC and to identify predictors of malignancy specifically in

a population of patients with a benign FNAB. Also, because of the limitations encountered in the retrospective extraction of data from the final pathology report, it was not possible to determine whether the carcinomas detected were identified in the biopsied nodules or elsewhere within the specimen. In addition, it could be argued that any study evaluating the accuracy of FNAB reporting could be affected by inherent interobserver variability between the interpreting cytopathologists. Although this may be true (and perhaps inherently unavoidable to some extent), we feel that the standardization of reporting criteria within the BSRTC makes reproducibility of results much less of a concern than it was with older, less standardized reporting schemes.¹ Also, the percentage of cases reported as benign on FNAB was 74 per cent, which is consistent with the percentage reported by other academic centers.¹⁶ Finally, because our focus was solely on the accuracy of the BSRTC, we did not consider any data that were available from molecular testing of obtained thyroid FNAB specimens. We acknowledge, however, the expanding role of molecular diagnostics in the evaluation and treatment planning of thyroid nodules and anticipate the integration of molecular diagnostics into subsequent treatment guidelines.¹⁷

Conclusions

The BSRTC continues to emerge as a highly useful tool for the evaluation and management of thyroid nodules. Despite this, surgical series continue to report FN rates for the “benign” category significantly higher than the zero to three per cent proposed by the BSRTC.¹ This suggests that patients taken to surgery for clinical indications other than the FNAB have a higher risk of an underlying carcinoma and that identification of these risk factors is useful for selecting patients for surgery despite a benign FNAB. Our study suggests that symptoms are not useful in predicting malignancy risk and, consequently, patients with symptoms (if tolerable) may be safely observed. Conversely, if suspicious ultrasound imaging findings are present, our study suggests that closer observation, a lower threshold for rebiopsy, or surgery may be warranted. Further investigation and larger-scale validation of this hypothesis will be useful to practicing endocrine surgeons.

Acknowledgments

We thank Ms. Chess Stover for her invaluable assistance with this project.

REFERENCES

1. Cibas E, Ali S. The Bethesda system for reporting thyroid cytopathology. *Thyroid*. 2009; 19:1159–1165. [PubMed: 19888858]
2. Broome J, Solorzano C. The impact of atypia/follicular lesion of undetermined significance on the rate of malignancy in thyroid fine-needle aspiration: evaluation of the Bethesda system for reporting thyroid cytopathology. *Surgery*. 2011; 150:1234–1241. [PubMed: 22136846]
3. Richmond B, O’Brien B, Mangano W, et al. The impact of the Bethesda system for reporting thyroid cytopathology on the surgical treatment of thyroid nodules. *Am Surg*. 2012; 78:706–710. [PubMed: 22643269]
4. Negro R, Piana S, Ferrari M, et al. Assessing the risk of false-negative fine-needle aspiration cytology and of incidental cancer. *Endocr Pract*. 2013; 19:444–450. [PubMed: 23337148]
5. Renshaw A. An estimate of risk of malignancy for a benign diagnosis in thyroid fine-needle aspirates. *Cancer Cytopathol*. 2010; 118:190–195. [PubMed: 20586117]

6. Yang J, Schnadig V, Logrono R, et al. Fine-needle aspiration of thyroid nodules: a study of 4703 patients with histologic and clinical correlations. *Cancer Cytopathol.* 2007; 111:306–315.
7. Chernyavsky V, Shanker B, Davidov T, et al. Is one benign fine needle aspiration enough? *Ann Surg Oncol.* 2012; 19:1472–1476. [PubMed: 21969084]
8. Cooper D, Doherty G, Haugen B, et al. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid.* 2009; 19:1167–1214. [PubMed: 19860577]
9. Harness J, Fung L, Thompson N, et al. Total thyroidectomy: complications and technique. *World J Surg.* 1986; 10:781–786. [PubMed: 3776215]
10. Castro M, Gharib H. Continuing controversies in the management of thyroid nodules. *Ann Intern Med.* 2005; 142:926–931. [PubMed: 15941700]
11. Hamburger J, Husain M. Contribution of intraoperative pathology evaluation to surgical management of thyroid nodules. *Endocrinol Metab Clin North Am.* 1990; 19:509–522. [PubMed: 2261905]
12. Gharib H, Papini E, Paschke R, et al. American association of clinical endocrinologists, associazione medici endocrinologi, and European Thyroid Association medical guidelines for clinical practice for the diagnosis and treatment of thyroid nodules. *Endocr Pract.* 2006; 12:63–102. [PubMed: 16596732]
13. Meko J, Norton J. Large cystic/solid thyroid nodules: a potential false-negative fine-needle aspiration. *Surgery.* 1995; 118:996–1004. [PubMed: 7491545]
14. McCoy K, Jabbour N, Ogilvie J, et al. The incidence of cancer and rate of false-negative cytology in thyroid nodules greater than or equal to 4cm in size. *Surgery.* 2007; 142:837–844. [PubMed: 18063065]
15. Cappelli C, Castellano M, Pirola I, et al. The predictive value of ultrasound findings in the management of thyroid nodules. *QJM.* 2007; 100:29–35. [PubMed: 17178736]
16. Theorharis C, Adeniran AJ, Roman S, et al. The impact of implementing the Bethesda system for reporting of thyroid FNA at an academic center. *Diagn Cytopathol.* 2013; 41:858–863. [PubMed: 23512999]
17. Yip L, Kebebew E, Milas M, et al. Summary statement: utility of molecular marker testing in thyroid cancer. *Surgery.* 2010; 148:1313–1315. [PubMed: 21134567]

FNAB Category	Risk of malignancy	Recommendation
Benign	0-3%	Observation, follow up
Unsatisfactory	0-4%	Repeat biopsy (US guidance)
Atypia/follicular lesion of uncertain significance (AUS/FLUS)	5-15%	Repeat biopsy
Suspicious for follicular neoplasm	15-30%	Thyroidectomy (lobectomy)
Suspicious for malignancy	60-75%	Thyroidectomy
Malignant	97-99%	Thyroidectomy

Fig. 1.
The Bethesda System for reporting thyroid cytopathology.

Table 1

Patient Characteristics as a Predictor of Malignancy

	Total	Benign/Incidental Findings	Clinically Significant Malignancies (>1.0 cm papillary, or cancers of other types)	P Value (Fisher's exact test unless otherwise noted)
	n = 180	n = 168	n = 12	
Patient demographics				
Female	154 (85.6%)	144 (5.7%)	10 (83.3%)	0.69
Age, median (range), (years)	55 (18–87)	55 (18–87)	53 (30–83)	0.95
Time between FNA and surgery, median (range) (months)	1 (0–6)	1 (0–6)	2 (0–5)	0.95
Pre-existing thyroid disease				
None identified, other than index nodule	28 (15.6%)	25 (14.9%)	3 (25%)	0.40
Thyroid nodule(s) (pre-existing)	71 (39.4%)	65 (38.7%)	6 (50%)	0.54
Multinodular goiter	78 (43.3%)	75 (44.6%)	3 (25%)	0.24
Hashimoto's disease	11 (6.1%)	10 (6.0%)	1 (8.3%)	0.54
Graves disease	1 (0.6%)	1 (0.6%)	0 (0%)	1.00
Hypothyroidism	20 (11.1%)	20 (12.0%)	0 (0%)	0.37
Family history of thyroid cancer	10 (5.6%)	10 (6.0%)	0 (0%)	1.00
Previous radiation exposure	5 (2.8%)	5 (3.0%)	0 (0%)	1.00

* Mann-Whitney *U* test.

FNA, fine needle aspiration.

Table 2

Imaging Results as a Predictor of Malignancy

	Total	Benign/Incidental Findings	Clinically Significant Carcinomas (>1.0 cm papillary, or cancers of other types)	P Value (Fisher's exact test)
	n = 180	n = 168	n = 12	
Ultrasound				
Frequency of use	180 (100%)			
Suspicious findings	65 (3 6%)	56 (33%)	9 (75%)	0.01
Nuclear scan				
Frequency of use	54 (30%)			
Suspicious findings	31 (17%)	31 (18%)	0 (0%)	0.23
Computed tomography scan				
Frequency of use	20 (11%)			
Suspicious findings	12 (7%)	12 (7%)	0 (0%)	1.00

Table 3

Indications for Surgery/Symptoms as a Predictor of Malignancy

	Total	Benign/Incidental Findings	Clinical Significant Malignancies (>1.0 cm papillary, or cancers of other types)	P Value (Fisher's exact test)
	n = 180	n = 168	n = 12	
No recorded reason	10 (5.6%)	10 (6.0%)	0 (0%)	1.00
Hoarseness	29 (16.1%)	25 (14.9%)	4 (33.3%)	0.11
Dysphagia	79 (43.9%)	74 (44.0%)	5 (41.7%)	1.00
Patient request	49 (27.2%)	46 (27.4%)	3 (25%)	1.00
Enlarging nodule/nodule size	69 (38.3%)	66 (39.3%)	3 (25%)	0.38
Cosmesis	4 (2.2%)	4 (2.4%)	0 (0%)	1.00
Imaging results (ultrasound)	38 (21.1%)	31 (18.5%)	7 (58.3%)	0.004
Number of symptoms present				
No recorded symptom	10 (5.6%)	10 (6.0%)	0 (0%)	0.11
1 symptom	88 (48.8%)	82 (48.8%)	6 (50%)	
2 symptoms	68 (37.7%)	65 (38.6%)	3 (25%)	
3 symptoms	12 (6.7%)	10 (6.0%)	2 (16.7%)	
4 symptoms	2 (1.1%)	1 (0.6%)	1 (8.3%)	