Thyroid fine-needle aspiration diagnostics in a real-life setting:

Experiences of the implementation of the Bethesda system in Finland

Published in Cytopathology 29(2018):2, doi: https://doi.org/10.1111/cyt.12513
First published: 18 December 2017

Authors: Paajanen Ilona¹, Metso Saara¹,², Jaatinen Pia¹,²,³, Kholová Ivana¹,⁴

¹ Faculty of Medicine and Life Sciences, University of Tampere, P.O.Box 100, Tampere, Finland
² Department of Internal Medicine, Tampere University Hospital, Tampere, Finland
³ Division of Internal Medicine, Seinäjoki Central Hospital, Seinäjoki, Finland
⁴ Pathology, Fimlab Laboratories, Tampere, Finland

Running title Thyroid FNA diagnostics

Key words Thyroid nodules, thyroid cytology, thyroid cancer, Bethesda System, thyroid ultrasonography (US)

Word count (excluding abstract, tables and references): 2952

Word count in abstract: 250

Number of references: 21 (+1 in text)

Conflict of interest: This work was supported by grants from the Competitive Research Funding of the Pirkanmaa Hospital District. The authors have nothing else to disclose.

Corresponding author:

Adjunct Professor Ivana Kholová, MD, PhD
Pathology, Fimlab Laboratories
P.O.Box 66
FIN 33101 Tampere, Finland
phone: +358 3 311 74851
fax: +358 3 311 75503
email: ivana.kholova@sll.fimnet.fi
Abstract

Objectives: The Bethesda system is widely accepted for thyroid fine-needle aspiration (FNA) diagnostics, but has scarcely been analyzed in relation to clinical background data. Our aim was to analyze the thyroid FNA diagnostic process in view of clinical data, and to assess the validity of the Bethesda system during the first year of implementation.

Methods: 415 thyroid FNAs were taken from 363 patients during October 2011 – September 2012 in the Pirkanmaa Hospital District, Finland. The median age of the patients was 59 years, and the female-to-male ratio 4:1. Clinical data were collected from patient registries, and thyroid FNA and histopathological data from the pathology registry.

Results: The Bethesda categories were represented as follows: non-diagnostic 94 cases (26%), benign 177 (49%), AUS/FLUS 32 (9%), follicular neoplasm 31 (9%), suspicious for malignancy 20 (5%), and malignant 9 cases (2%). Of the non-diagnostic samples only 23 (24%), and of the AUS/FLUS 18 (56%) led to repeat FNA. Thyroid cancer was histopathologically diagnosed in 28 cases (8%). When the categories requiring surgical treatment were considered true positive findings, the sensitivity of the Bethesda system was 90%, and specificity was 70%. Inter-observer accuracy was 86%.

Conclusions: Already during the first year of implementation the Bethesda system proved reliable in evaluating the risk of thyroid malignancy. Nevertheless, the clinical judgement of the indication of US/FNA, and management according to the FNA findings need improvement. The relatively high proportion of non-diagnostic FNAs could be diminished by obtaining the samples by radiologists experienced in US-guided FNA techniques.
Introduction

A palpable thyroid nodule is found in approximately 5% of the general population, more commonly in women than in men. In ultrasonography (US) examination thyroid nodules can be detected in up to 68% of the general population and the increasing use of computed tomography (CT) and magnetic resonance imaging (MRI) has also uncovered many nonpalpable thyroid incidentalomas [1]. The management of these asymptomatic findings is still controversial. The malignancy risk of a thyroid nodule is 5-15%. The incidence of thyroid cancer has increased worldwide in the past few decades [1-4]. Despite the generally indolent behavior of a well-differentiated thyroid cancer, some patients have a recurrent or a persistent disease. (http://seer.cancer.gov/statfacts/html/thyro.html, 16.10.2016)

The most commonly used diagnostic procedure for thyroid nodules is US-guided fine-needle aspiration (FNA). Recent guidelines, in general, recommend evaluation of nodules >1 cm in diameter [1,5,6]. A recent clinical practice survey showed that there is variation in the management of thyroid nodules among clinical endocrinologists [7]. The variation is probably even wider among general practitioners and other clinicians making decisions on thyroid diagnostics and treatment.

The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) was created in 2007 to standardize the reporting of thyroid cytopathology and to improve the clinical management of thyroid findings [8]. The TBSRTC was implemented and generally accepted worldwide [9,10]. The system has standardized the reporting and the management of thyroid FNA findings and provides better support to a clinician in treatment decisions than the former cytologic reporting system.[11] Its effectiveness and feasibility have been confirmed in several meta-analyses [11-15].

The diagnostic categories of the TBSRTC include different risks of malignancy, which influence the recommended managements after the FNA.[8] However, the category AUS/FLUS is still rather problematic, because it is heterogeneous, overused and the cytohistological correlation is assessed in a minority of cases [14]. Approximately 15–20 percent of FNA biopsies are insufficient, non-diagnostic, or AUS/FLUS, and a repeat FNA is the procedure of choice in these cases [8].
The clinical background data (age, gender, referring health care unit, indication for US/FNA, thyroid function tests (TFTs)) of patients subjected to thyroid FNA have scarcely been analysed in relation to the Bethesda system. The aim of the present study was to find out the clinical background and US, cytological and histopathological findings, as well as the consequences of the thyroid FNAs performed in an unselected cohort from primary and specialized health care in a real-life setting. A quality control analysis of the cytological analysis was included, to ensure the quality of the TBSRTC during the first year of its implementation in the Pirkanmaa Hospital District, Finland.

**Materials and Methods**

Fimlab Laboratories began the use of the TBSRTC on October 1, 2011. The study material consists of data collected from patient and sample registries on the thyroid FNA samples taken in the Pirkanmaa Hospital District (ca. 520 000 residents), and analyzed in Fimlab Laboratories during a one-year period from October 1, 2011. The retrospective analysis was based on the clinical data collected at Tampere University Hospital, the regional hospitals, the primary healthcare centers of Pirkanmaa Hospital District, and the occupational health care of Tampere University Hospital. In cases with unavailable patient records, the information was collected from the referral to FNA and the pathology report.

The following clinical information was collected from the patient records: the clinical problem that led to the thyroid US and FNA, the results of the TFTs made less than three months before the FNA, the specific findings of the US examination, and the consequences of the FNA biopsy (complications and management after the FNA). Three to 5 passes are usually performed per nodule on US-guided FNA. All the FNA samples were alcohol-fixed, cytospined and Papanicolaou stained. All the FNAs were analyzed according to the TBSRTC. If more than one nodule was sampled by FNA in the same examination, the more severe Bethesda finding was included in the analyses.
An experienced thyroid pathologist (IK) re-evaluated 20 % (n=73) of the FNA samples, blinded to the results of the previous examination. The goal was to guarantee high quality categorization during the first year of implementation of the TBSRTC. If the management after the FNA was surgical, the final histopathological diagnosis was also reviewed. The FNA-targeted nodule was identified during the dissection. The accuracy of the Bethesda system was determined according to the final histopathological diagnosis of the surgically obtained samples. The sensitivity, specificity and the positive and negative predictive values were calculated, considering the ability of the TBSRTC to discover a neoplasia requiring surgical treatment, or a cancer.

Statistical analysis was performed using the IBM SPSS Statistics (version 20.0). The variables were mainly categorical, and the tests used were the Chi-square test and the Monte Carlo simulation test. A p value was considered significant when < 0.05. The thyroid FNA was considered a ‘screening test’ when the sensitivity and specificity were calculated. These analyses included all the cases where the final histopathological diagnosis was available. In the first analysis the true positive findings were all the Bethesda categories requiring surgical treatment (neoplasia), and in the second analysis only malignant and suspicious for malignancy categories were considered true positive findings.

The study was conducted according to the Declaration of Helsinki, and the protocol was approved by the Regional Ethics Committee of Tampere University Hospital (R14124).

Results

Altogether 415 thyroid FNAs were obtained from the study cohort of 363 patients. In 52 patients, two different foci were sampled by FNA in the same examination, and the more severe FNA finding was included in the analyses. The clinical characteristics of the patients are shown in Table 1. The median age of the patients was 59 years, and the female-to-male ratio was 4:1. Full patient record data were available for 298 (80 %) cases. Before the FNA, TFTs had been performed in only 211 (58 %) of the patients. Of these cases, 65 (18 %) were dysthyroid (hypothyroid 51 (14 %), hyperthyroid
14 (4 %)). In 74 cases (20%) the indication for thyroid US was other than a palpable or an incidentally found nodule or a goitre (e.g., hypothyroidism, complaints of compression, pain, or swelling in the neck).

The frequency of the cytological diagnoses is shown in Table 2. One third (126, 35%) of the samples belonged to a TBSRTC-group, for which repeat FNA is recommended; 94 (26%) samples were non-diagnostic, and 32 (9 %) were classified into the AUS/FLUS category. One third of the samples classified as non-diagnostic contained cyst fluid only.

The cytological findings differed between male and female patients, benign findings being more frequent in the female ones (Table 3). There was no statistically significant difference in the distribution of cytological findings between the euthyroid patients (TSH 0.27-4.2 mU/l) and the dysthyroid ones. Neither did the Bethesda findings differ between FNAs obtained from a solitary nodule and those from a multinodular goitre. If the indication of thyroid US was other than a nodule, the FNA finding was malignant or suspicious for malignancy in only 3 (4 %) of the cases, while 12 cases (11 %) of palpable nodules and 10 (14 %) of nodules found incidentally on previous imaging yielded malignant or suspicious-for-malignancy FNA findings (Table 3).

A repeat FNA was performed on 23 (24%) of the patients with a non-diagnostic finding and on 18 (56%) of those with an AUS/FLUS finding (Table 4). Of the 51 repeat FNAs, 41 (85%) were performed because of a non-diagnostic or an AUS/FLUS finding. Thirty (73 %) of the findings changed into another diagnostic category and 21 (27 %) remained non-diagnostic or AUS/FLUS. When the FNA was repeated, as recommended, 14 (77 %) of the AUS/FLUS and 14 (59%) of the non-diagnostic samples were diagnosed as benign (data not shown in the Tables).

The FNA finding led to surgery in 7 (78%) of the patients with a malignant FNA, and 18 (90 %) of those with a finding suspicious for malignancy (Table 4). Two malignant FNA biopsies (22%) leading to clinical follow-up were taken from inoperable tumors (anaplastic carcinoma, metastatic melanoma), treated with palliative radiation. When the FNA categories follicular neoplasia,
suspicious for malignancy and malignant were considered true-positive findings (requiring surgical
treatment), the sensitivity of the Bethesda system was 90 %, specificity was 70 %, PPV 83 % and
NPV 81 %. When only suspicious for malignancy and malignant were classified as true-positive
findings, sensitivity was 81 %, specificity 88 %, PPV 79 % and NPV 90 %.

In the quality control analysis (Table 5), 73 (20%) of the FNA samples were re-evaluated by an
experienced thyroid pathologist. The diagnostic category changed on re-evaluation in 10 (14 %) of
the samples. Most of these samples changed from the benign to the non-diagnostic category. The
inter-observer agreement was 86 %.

A total of 78 (22 %) patients in the study population underwent thyroid surgery (Table 6), either
a lobectomy or a total thyroidectomy. Of the operated cases, 27 (35 %) were malignant tumors, and
in 51 (65 %) the final histopathological diagnosis was benign (nodular goitre in 23 (18%) cases,
follicular adenoma in 21(16%), and Hashimoto’s thyroiditis in 7(5%)). In the suspicious for
malignancy category, 15 (71 %) of the histopathological diagnoses were malignant, and in the
malignant category 7 (100 %). When the TBSRTC category was follicular neoplasm, the malignancy
rate was 12 % (n=3). When the category was non-diagnostic, benign or AUS/FLUS, only 2 out of the
26 surgical samples (7 %) were malignant. The most common indications for surgery, when it was
not the management recommended for the TBSRTC category, were recurring cysts, cosmetic
disturbance, or compression symptoms in the neck.

Of the patients whose indication for thyroid US was other than a palpable or an incidentally found
nodule, 30 were operated on. The final histopathological diagnosis was papillary carcinoma in 8
patients (27 %). All these patients with a malignant histopathological finding had neck or throat
symptoms, and a visible nodule in the US examination. When the final histopathological diagnosis
was Hashimoto’s thyroiditis (n = 7), the indication for US was other than a nodule in every patient.

Of the patients with an AUS/FLUS finding, 7 (22 % ) were referred straight to a diagnostic
lobectomy (Table 4). The final histopathological diagnosis was goitre in 4, Hashimoto’s thyroiditis in
and follicular adenoma in one of the cases. Twenty-two (73%) of the patients with a follicular neoplasm cytology were referred to a diagnostic lobectomy (Table 4). Among these patients the histopathological diagnoses were goitre in 4, follicular adenoma in 15, follicular carcinoma in 2, and Hashimoto’s thyroiditis in one of the cases.

Discussion

The present study showed a high sensitivity and specificity of the TBSRTC, even during the first year of its implementation. The steps in thyroid US and FNA diagnostics that still need to be improved are the clinical judgement of the indication of US and FNA, and the proper management of the thyroid condition, according to the TBSRTC findings. The proportion of non-diagnostic FNA samples was also relatively high, which could be improved by obtaining the samples by radiologists experienced in US-guided thyroid FNA techniques.

The recommended indication of thyroid US and FNA is a palpable or an incidentally found thyroid nodule [1,5,6]. In 45% of this cohort, the indication of the US scan was against the guideline recommendations. Our results also showed that malignant findings were infrequent, if the indication of the US was other than a thyroid nodule. Therefore, avoiding FNAs in the absence of a remarkable thyroid nodule (by size or US appearance criteria) would decrease the number of unnecessary invasive procedures (FNAs and surgery).

The first year of using the TBSRTC showed that there are challenges in the clinical interpretation of the FNA findings. The clinical management after the FNA grossly deviated from the prevailing recommendations [8], especially in the non-diagnostic and benign categories. The latest ATA guidelines recommend repeating the US within 12 months, if the US pattern is highly suspicious. If the nodules are of a very low suspicion pattern, the interval can be over 24 months [1]. In our study these recommendations were fulfilled for only 37 (27%) of the patients with a benign FNA finding.
Only 23 (24%) of the non-diagnostic samples, and only 18 (56%) of AUS/FLUS findings led to repeat FNA. However, if FNA was repeated as suggested, 14 (77%) of the AUS/FLUS samples were diagnosed as benign. This correlates well with a recent meta-analysis, where 62% of the AUS/FLUS lesions where benign in histological analyses [14].

The non-diagnostic category was more frequent (25%) in this study than in a recent meta-analysis [11]. One third of the samples classified as non-diagnostic contained cyst fluid only, underlining the proper management of cystic lesions. The cystic lesions should be emptied prior to the FNA biopsy. If there is a visible solid component after emptying the cyst, a US-guided FNA should be performed.[17] The high malignancy rate (33%) among the non-diagnostic samples of this cohort is likely to result from a random effect, as only three non-diagnostic samples were referred to surgery and histopathological analysis.

Approximately a half of the cytological samples fell into the benign Bethesda category, in line with previous studies. The risk of malignancy in the FNA samples classified as benign was somewhat higher in the present study than in the meta-analysis of previous reports (9% vs. 3.7%) [11]. This finding may be partly due to a relatively large number of radiologists performing a small number of FNA biopsies each, increasing the risk of false negative or non-diagnostic samples. It is also possible that the pathologists tended to approve more insufficient samples during the early phase of the TBSRTC implementation. In our quality control analysis, 19% of the samples in the benign category changed into another category, mostly to the non-diagnostic one.

The proportion of the AUS/FLUS category samples in the present study (9%) is in agreement with the recent meta-analyses. The risk of malignancy in the AUS/FLUS category, however, was lower in the present study (9%) than the reported risk of 16-34% in previous meta-analyses. [11,14]

In the present cohort, 2% of the FNA samples fell into the malignant category. All the tumors classified as malignant on FNA and subsequently operated on, were papillary carcinomas in the histological analysis. In addition, the histological analysis verified 5 follicular carcinomas, the
TBSRTC categories of which were follicular neoplasia in 3, suspicious for malignancy in one and non-diagnostic in one of the cases. Of the cases classified as suspicious for malignancy, 90% were referred for thyroid surgery, and their malignancy rate was 71%. These results correlate well with the meta-analysis of Bongiovanni et al. [11], and confirm the high reliability of the TBSRTC right from the start of its application at the pathology unit analyzing all the samples of the present study.

The aim of a FNA biopsy is to verify the nature of a thyroid nodule, and to help selecting the ones that need operative treatment. In this study, the overall malignancy rate of the operated cases (35%) was similar to the average rate in previous studies (33.8%) [11], and the rate of neoplasia (either benign or malignant) was 62%. In 33% of the operated cases, the recommended management according to the TBSRTC category was not surgical. The most common reasons for surgery in these cases were a refilling cyst or a large goitre, causing discomfort in the neck.

Currently, follicular neoplasms are treated operatively, as benign and malignant follicular lesions cannot be differentiated on cytological examination. The risk of neoplasia has been included in the latest analyses of thyroid cytology, to highlight the fact that also adenomas should often be surgically removed, but the rate of neoplasia is still rarely reported in literature. In the study of Dincer et al, the rate of neoplasia in the AUS/FLUS category was 34%, which is comparable with our results (27%). [14,18] The development of molecular testing might help in avoiding unnecessary thyroidectomies in patients with indeterminate thyroid nodules [6,19,20,21]. More data are needed to evaluate the clinical applicability of the evolving molecular diagnostics [6,16].

In conclusion, the clinical judgement of the indications of thyroid US and FNA, and the proper management according to the FNA findings could be further improved, to target the invasive investigations and treatment to patients likely to benefit from them. When properly used, the Bethesda system is sensitive and specific, and useful in avoiding unnecessary surgery.
Acknowledgements: We thank Dr. Ritva Kannisto for her help in the acquisition of the clinical data, and Esko Väyrynen, MA, for revising the language of the manuscript. This work was supported by grants from the Competitive Research Funding of the Pirkanmaa Hospital District.
References


Table 1. Description of the study cohort (n=363) subjected to thyroid fine-needle aspiration (FNA)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>288</td>
<td>79</td>
</tr>
<tr>
<td>Male</td>
<td>75</td>
<td>21</td>
</tr>
<tr>
<td><strong>Referral to FNA from</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary health care</td>
<td>247</td>
<td>68</td>
</tr>
<tr>
<td>Specialized health care</td>
<td>116</td>
<td>32</td>
</tr>
<tr>
<td><strong>Patient data extracted from</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full patient records</td>
<td>298</td>
<td>82</td>
</tr>
<tr>
<td>Biopsy referral data only</td>
<td>65</td>
<td>18</td>
</tr>
<tr>
<td><strong>Was the patient euthyroid?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>173</td>
<td>48</td>
</tr>
<tr>
<td>No</td>
<td>36</td>
<td>10</td>
</tr>
<tr>
<td>No data available</td>
<td>154</td>
<td>42</td>
</tr>
<tr>
<td><strong>Indication for thyroid US</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nodule</td>
<td>112</td>
<td>31</td>
</tr>
<tr>
<td>Goitre</td>
<td>109</td>
<td>30</td>
</tr>
<tr>
<td>Incidentally found nodule*</td>
<td>68</td>
<td>19</td>
</tr>
<tr>
<td>Other**</td>
<td>55</td>
<td>15</td>
</tr>
<tr>
<td>No data available</td>
<td>19</td>
<td>5</td>
</tr>
<tr>
<td><strong>Nodule size on US</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 1 cm</td>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td>1 cm or more</td>
<td>261</td>
<td>72</td>
</tr>
<tr>
<td>No data available</td>
<td>72</td>
<td>20</td>
</tr>
<tr>
<td><strong>Indication for FNA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nodule confirmed on US</td>
<td>310</td>
<td>85</td>
</tr>
<tr>
<td>Goitre</td>
<td>34</td>
<td>9</td>
</tr>
<tr>
<td>Other</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>No data available</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

*Found incidentally on carotid US, MRI, CT, PET-CT, scintigraphy or SPET scan

** E.g. hypothyroidism, hyperparathyroidism, dysphagia, pain or swelling in the neck region
Table 2. Frequency of cytological diagnoses in the present cohort, compared to a recent meta-analysis by Bongiovanni et al. [11]

<table>
<thead>
<tr>
<th>Diagnostic category of the Bethesda system</th>
<th>Present study n (%)</th>
<th>Meta-analysis [11] n (%)</th>
<th>Risk of nonplasia(^a), this study %</th>
<th>Risk of malignancy, this study %</th>
<th>Risk of malignancy, meta-analysis [11] %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-diagnostic</td>
<td>94 (26)</td>
<td>3271 (13)</td>
<td>33</td>
<td>33</td>
<td>17</td>
</tr>
<tr>
<td>Benign</td>
<td>177 (49)</td>
<td>15 104 (59)</td>
<td>18</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>AUS/FLUS</td>
<td>32 (9)</td>
<td>2441 (10)</td>
<td>27</td>
<td>9</td>
<td>16</td>
</tr>
<tr>
<td>Follicular neoplasm</td>
<td>31 (9)</td>
<td>2571 (10)</td>
<td>65</td>
<td>13</td>
<td>26</td>
</tr>
<tr>
<td>Suspicious for malignancy</td>
<td>20 (5)</td>
<td>680 (3)</td>
<td>88</td>
<td>71</td>
<td>75</td>
</tr>
<tr>
<td>Malignant</td>
<td>9 (2)</td>
<td>1378 (5)</td>
<td>100</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>Total</td>
<td>363 (100)</td>
<td>25 445 (100)</td>
<td>62 %(^a)</td>
<td>35 %(^a)</td>
<td>34 %(^a)</td>
</tr>
</tbody>
</table>

\(^a\) Percentage of the operated patients; \(^b\) Percentage of all the patients; \(^c\) Neoplasia refers to both benign and malignant tumors