Bethesda System in the evaluation of thyroid nodules: Review

Marta Wesoła A–D, Michał Jeleń E, F

Department of Pathomorphology and Oncological Cytology, Wroclaw Medical University, Poland

A – research concept and design; B – collection and/or assembly of data; C – data analysis and interpretation;
D – writing the article; E – critical revision of the article; F – final approval of article

Abstract

Fine needle aspiration (FNA) biopsy is an important and widely accepted method used in the diagnosis of patients with thyroid nodules. The lack of consistency in reporting FNA taken from the thyroid gland has led to divergences in the calculation of the sensitivity and specificity of the method. Discrepancies depend on what was considered to be true or false negative/positive. This resulted in confusion among doctors, who had to decide how to manage the patient with constantly changing positive or negative results of the FNA. The problem was solved in 2007, when “The Bethesda System for Reporting Thyroid Cytopathology” (TBSRTC) was introduced. Generally, TBSRTC, compared to previous systems, results in reducing the number of non-diagnostic/indeterminate cases and enables a better clinical usefulness of the results of the FNA of the thyroid gland. This is probably due to the introduction of more standard criteria for interpreting and reporting. TBSRTC improves communication between cytopathologists, reduces the number of unnecessary operations on benign lesions, and makes it possible to perform the operation on time in patients with malignant lesions and predict the risk of thyroid cancer. It provides a simple and reliable exchange of data not only between various laboratories but also between institutions throughout the world. Research shows that all indicators of malignancy calculated for all categories have similar value to the recommended rate of malignancy.

Key words: thyroid, Bethesda System, cytology, fine-needle aspiration, terminology

DOI
10.17219/acem/27319

Copyright
© 2017 by Wroclaw Medical University
This is an article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Fine needle aspiration (FNA) biopsy is an important and widely accepted method used in the diagnosis of patients with thyroid nodules. It is simple, thorough and safe for the patient as well as cost-effective.\(^1\) FNA is characterized by high sensitivity and accuracy.\(^2\) It reduces the number of thyroid surgeries on patients with benign lesions as well as appropriately qualifies patients with thyroid cancer for surgery.\(^3\) Despite the wide application of FNA in the diagnosis of thyroid nodules, there was confusion related to diagnostic terminology in the assessment of the preparations. The reasons for this were the multiplicity of the category names, descriptive reports without assigned categories and variable terminology of surgical pathology.\(^4\) The lack of consistency in the reporting of FNA taken from the thyroid gland has led to divergences in the calculation of the sensitivity and specificity of the method. The discrepancies depend on what was considered to be true or false negative/positive. This resulted in confusion among doctors, who had to decide how to manage the patient with a constantly changing positive or negative result of the FNA.\(^5\) The problem was solved in 2007, when it was discussed by the National Cancer Institute (NCI) during the NCI Thyroid Fine Needle Aspiration State-of-the-Science Conference, where terminology and morphological criteria for the evaluation of the preparations from the FNA of the thyroid gland were defined, thereby creating a framework for “The Bethesda System for Reporting Thyroid Cytopathology” (TBSRTC).\(^6,7\) The introduction of the new system of diagnostics of the thyroid gland was supported by many researchers.\(^8,9\)

Fig. 1 presents the recommended diagnostic categories for the Bethesda System and the risk of malignancy for a specific category.\(^10\)

Earlier studies show that individual diagnostic categories are interlinked.\(^11\) For example, Renshaw et al. suggest in their studies a negative link between ND (non-diagnostic) categories and MGT (malignant), implying that the cases qualified to ND are most likely to be with benign rather than malignant lesions. A recent survey by Krane et al.\(^13\) shows that there is a relationship between AUS/AFLUS categories (atypia of undetermined significance/ atypical follicular lesion of undetermined significance) and MGT, suggesting that during the AUS/AFLUS diagnosis, the malicious changes are more probable than benign changes. Therefore, it was proposed that the (AUS/AFLUS)/MGT ratio could be a useful tool for assessing quality in cytology.

### Comparison of previous system and the Bethesda System for reporting thyroid cytopathology

Pathak et al.\(^14\) conducted a study in which 39 cases specified as not conclusive using the 4-phase system were rated again using TBSRTC. The reassessment was carried out by a person with the same diagnostic experience. Of the 39, only 2 cases were left in the category of ND, which constitutes 5.13%. The others were qualified...
for other categories. In the AUS/AFLUS category, there were 15 cases (38.46%). The authors note that the introduction of the 6-Phase Bethesda System has resulted in reducing the number of non-diagnostic/indeterminate cases as compared to the previously-used system, enabling better clinical usefulness of the results of the FNA of the thyroid gland.14 Song et al.15 also believe that the introduction of TBSRTC reduces the number of unspecified cases of thyroid gland FNA. They conducted a study in which, among 154 indeterminate cases, only 2 were in the ND group (and 77 cases in the AUS/AFLUS group).15 Richmond et al.16 analyzed 100 cases before and after the introduction of TBSRTC. Of the 68 cases previously referred to as indeterminate, 32 (41.7%) were assigned to the BN category (benign). The authors believe that the introduction of TBSRTC resulted in the classification of a significant proportion of the previously unspecified cases as benign lesions. This is probably due to the introduction of more standard criteria for interpreting and reporting. The authors conclude that, while maintaining the same diagnostic accuracy in detecting cancers of the thyroid, it is possible to reduce the number of surgeries to remove part or all of the thyroid gland (thyroidectomy).16 Other authors (Kiernan et al.)17 on the basis of their studies, conclude that the introduction of TBSRTC in a single institution has resulted in an increase in the percentage of indeterminate FNA cases, but the diagnostic accuracy of the SM category (suspicous for malignancy) increased.17 McElroy et al.18 compared the earlier system and TBSRTC. On this basis, they stated that the introduction of TBSRTC led to improved predictions of malignancy and more accurate diagnosis. After the introduction of TBSRTC, the clinical management of patients with indeterminate diagnosis has not changed. Interobserver concordance was rated moderate, what could have been caused by a different level of experience in cytology.18 Rabaglia et al.19 said, after comparing the system before and after the introduction of TBSRTC, that after the introduction of TBSRTC, the number removals of the thyroid gland has not changed for the indeterminate cases, but the risk of malignancy has been reduced.

Impact of the Bethesda System for reporting thyroid cytopathology on quality of reporting

This system improves communication between cytopathologists, reduces the number of unnecessary operations on benign lesions and makes it possible to perform the operation on time in patients with malignant lesions and predict the risk of thyroid cancer. It provides a simple and reliable exchange of data not only between various laboratories but also between institutions throughout the world.10 Bongiovanni et al. carried out a meta-analysis, on the basis of which they claimed that TBSRTC has high diagnostic accuracy in general and provides a reliable and valid reporting system for cytology of the thyroid gland.20 Hershman21 in his meta-analysis compared the results of the reports from 8 outlets. The MGT category accounted for 5.4% of all thyroid FNAs, SM for 2.6%, SFN (suspicious for follicular neoplasm): 10.1%, AUS/AFLUS: 9.6%, BN: 59% and ND: 13%. On the basis of these results, Hershman found that TBSRTC is an effective classification system of FNA for the treatment of patients with thyroid nodules.21 Crowe et al. claim22 TBSRTC affects the quality of reporting. By comparing the cases of thyroid FNA assessed by TBSRTC and the previous system, the authors proved that TBSRTC has reduced the number of subsequent surgical resection and reduced the number of ambiguous diagnoses. The overall number of operations carried out has been reduced, especially for benign lesions. However, the diagnostic accuracy has not changed significantly following the introduction of TBSRTC.23 According to Olson et al.23 the introduction of TBSRTC coincided with a reduction in the number of cases of „suspicous for papillary thyroid carcinoma” (SPTC) and at the same time caused an increase in the indicator of malignancy for that category, called in the Bethesda System, „suspicous for malignancy” (SM).

Comparison of interobserver concordance

As already mentioned, the introduction of TBSRTC led to the unification of diagnostic terminology and reporting of the results of the thyroid gland in order to improve the clinical management and reduce the number of indeterminate cases.18 However, morphological interpretation remains subjective and there are discrepancies between observers.24 Pathak et al.14 conducted a study in which 415 FNA thyroid gland preparations were assessed. All preparations were assessed by three observers with varying degrees of experience in cytopathology. In addition, for comparison of the results, all preparations were reassessed by an experienced cytopathologist. The results of this test show that the Bethesda System in FNA of the thyroid gland has good repeatability between observers at different levels of experience in cytopathology with greater repeatability for those with growing experience.14 In other studies, Prabath et al.25 compared the interobservers’ compatibility between the system used in Sri Lanka and the Bethesda System. Studies have shown a significant interobserver concordance for both systems. TBSRTC showed a slightly higher percentage of compliance results, but not statistically significant. Jing et al.26 did a reassessment of cases from AUS/AFLUS. Part of the cases remained in this Bethesda System diagnostic cat-
category and the rest were allocated to one of the remaining other categories. Compatibility between observers for these 5 categories ranges from 71.6% to 100%, which gives it an average of 88.8%.26

Comparison of the malignancy indicator between the recommended and calculated rate on the basis of the studies

Mufti et al.27 conducted a study to determine the indicator of malignancy and compare it to the recommended rate. The indicator of malignancy for specific categories was as follows: ND/UNS: 20%, BN: 3.1%, AUS/AFLUS: 50%, SFN: 20%, SM: 80% and MGT: 100%. Similar studies conducted by Lee et al.28 showed the malignancy indicator valued: ND/UNS: 4.2%, BN: 0%, AUS/AFLUS: 41.2%, SFN: 46.4%, SM: 48.4% and MGT: 78.4%. Vickie et al.29 noted the following malignancy indicator: ND/UNS: 8.9%, BN: 1.1%, AUS/AFLUS: 17%, SFN: 25.4%, SM: 70% and MGT: 98.1%. Wu et al.30 noted, in their studies, the malignancy indicator as follows: BN: 3%, AUS/AFLUS: 6%, SFN: 22%, SM: 56% and MGT: 100%. Yassa et al.31 noted the malignancy indicator for specific categories to be: ND: 10%, BN: 0.3%, AUS/AFLUS: 24%, SFN: 28%, SM: 60% and MGT: 97%. Similar studies conducted by Yang et al.32 recorded the indicators at the following level: ND: 0%, BN: 4.5%, AUS/AFLUS: 20%, SFN: 36%, SM: 75% and MGT: 97.8%. Tepeoğlu et al. conducted similar studies.34 On their basis, the malignancy indicators were calculated as: BN: 0%, AUS/AFLUS: 12.7%, SFN: 35%, SM: 91.4% and MGT: 100%. Sheffield et al.35 conducted a review on the subject of malignancy level between TB-SRTC categories: ND: 11–26%, BN: 4–9%, AUS/AFLUS: 19–38%, SFN: 27–40%, SM: 50–79% and MGT: 98–100%

Table 1 shows all the results which are similar to the Sheffield et al. results.

The collected data shows that the values for the AUS/AFLUS and SFN categories overlap each other.35 On the basis of the results, it can be seen that the diagnostic category AUS/AFLUS is multi-faceted. Ohori and Schoedel36 did a review of the AUS/AFLUS category which showed variation in the frequency of its occurrence ranging from 0.7 to 18% and the malignancy level wavering from 6 to 48%. Other authors (Broome and Solorzano),37 on the basis of studies in which samples were compared before and after the introduction of TBSRTC, said that the AUS/AFLUS category after the introduction of the system was used more often than recommended (14%) and the degree of malignancy was higher than expected (20%). On the basis of the study results,29 Vickie et al. state that classifying the changes in thyroid cells to the AUS/AFLUS category plays an important role in triaging of the patients with thyroid nodules because those patients happen to have a lower malignancy risk after surgical observation than those with preliminary diagnosis of SNF after cytology. It should be considered that the AFLUS/AUS category is subjective and non-homogeneous, and that there are and will always be differences between observers.38 Other authors39 feel that the AUS/AFLUS category, because of its heterogeneity and the fact that it consists of the samples with borderline cellularity (the quality of sample is compromised) as well as those with true atypia, should be eliminated. Studies conducted by them show that cases assigned to the AUS/AFLUS category can ultimately be assigned to other Bethesda System groups.39 Jing et al.26 conducted a study in which they reassessed

<table>
<thead>
<tr>
<th>Malignancy Rates among Published Studies</th>
<th>Diagnostic category</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ND [%]</td>
</tr>
<tr>
<td>Mufti et al.27</td>
<td>20</td>
</tr>
<tr>
<td>Lee et al.28</td>
<td>4.2</td>
</tr>
<tr>
<td>Vickie et al.29</td>
<td>8.9</td>
</tr>
<tr>
<td>Wu et al.30</td>
<td>–</td>
</tr>
<tr>
<td>Yassa et al.31</td>
<td>10</td>
</tr>
<tr>
<td>Yang et al.32</td>
<td>10.7</td>
</tr>
<tr>
<td>Nayar and Ivanovic5</td>
<td>9</td>
</tr>
<tr>
<td>Mondal et al.33</td>
<td>0</td>
</tr>
<tr>
<td>Tepeoğlu et al.34</td>
<td>–</td>
</tr>
<tr>
<td>Sheffield et al.37 (review)</td>
<td>11–26</td>
</tr>
</tbody>
</table>
the cases from the category of AUS/AFLUS. The analysis shows that from 50 preparations in the AUS/AFLUS group, ultimately 11 (22%) remained. In the ND category there were 2 (4%), BN: 26 (52%), SM: 10 (20%) and MGT: 1 (2%). Walls et al. conducted a study, on the basis of which they wanted to determine whether the AUS/AFLUS category should be divided into further subcategories according to the malignancy rate. Studies suggest that cases of AUS/AFLUS showing the presence of focal nuclear atypia carry a higher risk of cancer than other cases assigned to AUS/AFLUS and fit better into the category of SM, while those cases that do not show atypical nuclear characteristics show the risk of malignancy recommended for TBSRTC AUS/AFLUS categories. 40

Generally, compared to previous systems, TBSRTC results in reducing the number of non-diagnostic/indeterminate cases and enables better clinical usefulness of the results of a FNA of the thyroid gland. This is probably due to the introduction of more standard criteria for interpreting and reporting. TBSRTC improves communication between cytopathologists, reduces the number of unnecessary operations on benign lesions and makes it possible to perform the operation on time in patients with malignant lesions and predict the risk of thyroid cancer. It provides a simple and reliable exchange of data not only between various laboratories but also between institutions throughout the world. Research shows that all indicators of malignancy calculated for all categories in TBSRTC have similar values to their recommended rates.

References