Interobserver Reproducibility of Thyroid Fine-Needle Aspiration Using the UK Royal College of Pathologists’ Classification System

To the Editor

In their article in the June 2011 issue of the Journal, Kocjan et al.1 raise very important questions about the terminology of thyroid cytopathology. Specifically, a concept has aroused our curiosity: interobserver reproducibility.

The United Kingdom Royal College of Pathologists Classification System (UKRCPCS) for reporting thyroid fine-needle aspiration (FNA) specimens2 undoubtedly represents another step toward standardization, improved clinical significance, and usefulness of thyroid FNA.

During the past decade, several classification schemes for thyroid gland FNA have been proposed by various professional organizations. Most of these schemes consist of 4 to 6 diagnostic categories,3-10 which are not always comparable with each other. This has led to confusion and differences in perceptions of diagnostic terminology in cytopathology reporting of thyroid FNA between cytopathologists and clinicians.11,12 This confusion is even more significant if it is considered that many pathology departments do not use standardized diagnostic categories for the reporting of cytologic diagnoses of thyroid nodular lesions, but rather describe the findings.13

Among several classification schemes, the 5-class system proposed by the British association, Royal College of Physicians, in 20027 was accepted with some changes by the Italian Society of Pathology and Cytopathology—Italian Section of the International Academy of Pathology (SIAPEC-IAP) in 2007.8 The reporting system in use in the United Kingdom, following publication of a 6-class system, namely, The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC),4 was updated by a working group of The Royal College of Pathologists2 and is now quite similar to TBSRTC. Particularly in category Thy1, the use of Thy1c, in which “c” indicates a cystic lesion, was introduced; also in category Thy2, the use of Thy2c, in which again c indicates a cystic lesion, was introduced; the Thy3 category was classified as Thy3f for follicular lesions and Thy3a for atypia insufficient to enable confident placing into any other category. Nevertheless, in our view, these changes have not been largely adopted. For many years, all classification systems have provided a category for nondiagnostic FNA, a category for benign lesions, and a category for malignant lesions. However, there are also notable differences.

The UKRCPCS introduces 2 categories for borderline lesions, namely, “neoplasm possible, atypia/nondiagnostic (Thy3a)” and “neoplasm possible, suggesting follicular neoplasm (Thy3f).” Conversely, the previous British System and the SIAPEC-IAP system provide a single category for all borderline lesions, namely, “follicular lesion” and “indeterminate (follicular proliferation),” respectively7,8

Table I.

Independent from the adopted system, the main difficulty is represented by borderline lesions characterized by atypia of undetermined significance and/or by a microfollicular pattern.14-17 The differences in reporting borderline lesions outline the well-recognized difficulties in lesions belonging to the gray zone; these are classified by some authors as low- and high-risk according to immunocytochemical findings.18

As an initial step in our discussion, we would like to ask the authors the same questions that we have already put to Cibas and Ali19 in our letter on The Bethesda System for Reporting Thyroid Cytopathology20: (1) Are there recognizable strict morphologic quantitative and qualitative criteria in cytologic preparations that allow us to divide borderline follicular lesions in 2 categories? (2) If so, are these criteria adequate to ensure a satisfactory interobserver and intraobserver diagnostic reproducibility? (3) Are they uniformly applicable? (4) Could they vary significantly depending on the operator performing the FNA procedure? Are these criteria affected by quantitative and qualitative issues, such as representative cellularity and adequate fixation?
Thyroid Classification Schemes

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Thyroid Classification Schemes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UK RCPath</strong></td>
<td><strong>Bethesda</strong></td>
</tr>
<tr>
<td>Thy1</td>
<td>Nondiagnostic</td>
</tr>
<tr>
<td>Thy1c</td>
<td>Nondiagnostic, cystic lesion</td>
</tr>
<tr>
<td>Thy2</td>
<td>Nonneoplastic</td>
</tr>
<tr>
<td>Thy3a</td>
<td>Neoplasm possible, atypia/follicular neoplasm</td>
</tr>
<tr>
<td>Thy4</td>
<td>Suspicious for malignancy</td>
</tr>
</tbody>
</table>

UK RCPath, UK The Royal College of Pathologists.

of the smears? (5) Finally, are they prototypes of a clear diagnostic terminology?

As for the question: “Can these criteria, by themselves, provide sufficient information on which to decide the management of patients with nodular thyroid lesions?” In their article, Kocjan et al1 had explained that “UK RCPath [The Royal College of Pathologists] terminology assumes that all Thy3a, Thy3f, Thy4, and Thy5 aspirates (equivalent to TBSRTC III, IV, V, and VI, respectively) will be reviewed in a multidisciplinary clinical team meeting before a clinical decision is made about the required therapeutic action.”

We agree and think that FNA is the most important, but not the only important, deciding factor, and, therefore, it must be integrated with other diagnostic procedures, such as ultrasonographic and scintigraphic examination of the thyroid gland. In our experience 99m-Tc-sestamibi scanning can be useful in selecting what nodule to aspirate or to excise after an FNA borderline diagnosis.21-23 Kocjan et al1 explained the criteria to classify the cases into single categories, but our question remains: “Are there strict morphologic quantitative and qualitative criteria…”?

Morphologic criteria are not only hindered by poor reproducibility,15 but they are also affected by quantitative and qualitative variables that are strongly dependent on the expertise of the operator performing the FNA procedure. If we accept the intrinsic and extrinsic limitations of thyroid FNA, the validity of splitting borderline lesions into 2 categories seems dubious. We should not forget that the recipients of cytopathology reports are patients and clinicians and that the latter are not always specialized in thyroid diseases. Do we really think that everyone understands the differences between Thy1 and Thy1c or between Thy1c and Thy2c or between Thy3a and Thy3f? We think that this subclassification can generate confusion. For example, clinicians can face a lot of difficulty when reading in our reports “Nondiagnostic…(Thy1)” and “…/nondiagnostic (Thy3).” Kocjan et al1 explain that “samples that exhibit cytologic atypia or other features that raise the possibility of neoplasia, but are insufficient to enable confident placing into any other category, should be classed as Thy3a.” This includes “samples in which there is architectural atypia…,” “only sparse colloid is evident…,” “sparsely cellular samples containing predominantly microfollicles, focal cytologic changes that are most probably benign but papillary carcinoma cannot be confidently excluded…,” etc. For us, this is simply an indeterminate class, as recognized by the SIAPEC-IAP system8 (Table 1). Kocjan et al1 specify that “these samples should form only a small minority of Thy3 cases.” Then we can say, even if we manage to obtain good reproducibility among cytopathologists evaluating thyroid FNA and to correctly address terminology issues related to reporting, that the clinical impact of the category Thy3a remains poor. The corresponding category, according to the TBSRTC, should be used as a last resort and be limited to approximately 7% or fewer of all thyroid FNAs.

Cibas and Ali19 wrote “The AUS/FLUS category may never have good interobserver reproducibility, even after pathologists familiarize themselves with the criteria in the atlas. But the lack of good reproducibility does not necessarily disqualify a diagnostic category.” They agreed with Downs-Kelly et al24 that pathologists often report diagnostic categories with limited reproducibility because of a belief that certain distinctions are clinically important.

We think that it is better for pathologists to agree among themselves and to remember that, if the number of classes is increased, the classification system could become less useful.9 Besides, the issue is how one can use the classes. Most problems in classifying the difficult cases are correlated with poor samples or specimens. If we think that a case is nondiagnostic or unsatisfactory, it should be classified as Thy1. If we think that a case is atypical, it should be classified as Thy3. Finally, if a case is negative for malignant cells but with some atypia, we can place it in the Thy2 class with a recommendation for repeat FNA for confirmation. In the last case, we think that a Thy3a report meets more the needs of pathologists rather than the needs of patients.
Ravetto et al.\textsuperscript{25} and Piana et al.\textsuperscript{26} showed the accuracy of the 5-class reporting scheme for the reporting of thyroid cytopathology. The report of a multi-institutional study comparing 5- and 6-class diagnostic systems is in press; this study concluded that both systems have excellent sensitivity but different specificity and accuracy.\textsuperscript{27}

We favor a 5-category classification system that links all the borderline lesions, and we prefer the term indeterminate/follicular proliferation. The report of the cytopathologic evaluation, particularly of borderline lesions, should always be accompanied by a microscopic description and, ideally, by brief management advice. It is critical to recognize that experience and technical procedures in performing thyroid

References


