Ultrasound-guided fine-needle aspiration of thyroid nodules: stratification of malignancy risk using follicular proliferation grading, clinical and ultrasonographic features

Sandrine Rorive, Nicky D’Haene, Caroline Fossion, Isabelle Delpierre, Nizar Abarguia, Freddy Avni, Christine Decaestecker and Isabelle Salmon

Department of Pathology, Department of Radiology, Erasme Hospital and Laboratory of Image Synthesis and Analysis, Faculty of Applied Sciences, Université Libre de Bruxelles (ULB), B-1050 Brussels, Belgium

(Correspondence should be addressed to I Salmon; Email: isabelle.salmon@erasme.ulb.ac.be)

Abstract

Objective: To evaluate the diagnostic value of fine-needle aspiration (FNA) cytology and the additive contribution brought by clinical and ultrasound (US) features.

Method: Cytological and histological diagnoses were compared in a series of 924 patients who underwent US-guided FNA before surgery. We additionally developed a grading system for follicular proliferation (FP) FNA diagnosis, and investigated its impact on the malignancy risk as well as the additive contribution of clinical and US features by means of decision tree analysis.

Results: Excluding FP cases (n = 395), our data demonstrated that strictly benign or malignant FNA diagnoses exhibit great concordance with benign or malignant histological diagnoses (97.8% accuracy). Our grading system that was applied to the 395 FP cases revealed that grades 1, 2 and 3 were associated with a 7.7, 17.7 and 45.7% incidence of malignancy respectively. Decision tree analysis resulted in a classification model which involved FP grade, patient’s age, serum thyroglobulin level, nodule size and nodule uniqueness. This model identified a subgroup of patients with grade 1 FP nodules who were older than 50 years, and who had a higher risk of malignancy (17.9%). In addition, high serum thyroglobulin levels were associated with a very high malignancy risk (75.0%) for patients with grade 3 FP nodules. Finally, among grade 2 FP patients, unique and large nodules were associated with a high malignancy risk of 36.1%.

Conclusions: The integration of FP grade, clinical and US features allows the stratification of patients with FP cytology according to their risk of malignancy.

European Journal of Endocrinology 162 1107–1115

Introduction

The assessment of thyroid nodules is a common clinical problem. The frequency of thyroid nodules detected by ultrasound (US) has strongly increased in recent years, reaching 67% of the adult population (1). In contrast, thyroid cancers are rare, accounting for only 1% of all cancers and occurring in ~5% of all thyroid nodules independent of their size (1, 2). Thus, the challenge that the physicians who manage patients with thyroid nodules face is the efficient stratification of patients according to their risk of malignancy in order to identify the best follow-up and therapeutic options.

Fine-needle aspiration (FNA) has become the predominant method used for the primary diagnosis of benign and malignant thyroid nodules, resulting in the categorisation of patients as operative or non-operative candidates (1–3). However, FNA has intrinsic limitations in distinguishing between benign and malignant follicular lesions (4). More particularly, evaluation and treatment of patients with follicular proliferation (FP) cytology still remain problematic. As this category is associated with a 20–30% incidence of malignancy, patients with this cytological diagnosis are referred for surgery (3). In addition to the development of biomarkers applied to FNA (5, 6), efforts to improve the management of these patients have focused on identifying additional clinical and US data that efficiently predict malignancy (7, 8).

In the present study, we investigated a series of 924 patients who underwent US-guided FNA before surgery, allowing the comparison of cytological and histological diagnoses. Our analysis aimed to evaluate the diagnostic value of FNA cytology, and to examine the risk factors associated with malignancy. In particular, we developed a grading system for FP FNA diagnosis, and
investigated its impact on the malignancy risk as well as the additive contribution of clinical and US features. Finally, we have proposed a refined decision aid tool for surgical indication integrating clinical, US and cytological features.

Materials and methods

Patient selection

The series of 924 patients analysed in the present study were identified by a retrospective study conducted in the Department of Pathology, Erasme Hospital (Brussels, Belgium). The selection criteria consisted of patients who underwent FNA followed by surgery (lobectomy or total thyroidectomy) from 1986 to 2007 at Erasme Hospital. The clinical data collected for each patient included age, sex, prior medical history, thyroid scan results, biochemical functional status, serum thyroglobulin level and thyroid antibodies (anti-thyroglobulin, anti-thyroid peroxidase or anti-TSH receptor antibodies). Their distributions are detailed in the Results section.

FNA indications

In our hospital, we carried out FNA of nodules with a diameter larger or equal to 1 cm and exhibiting hypofunction (radionuclide scan), except if they are cystic and without solidity. In addition, each nodule associated with suspicious clinical features (rapid growth or adenopathy) was also considered for FNA without taking into account the above-mentioned criteria.

US-guided FNA procedure

FNAs were performed by an experienced pathologist (I S or S R) following a well-defined procedure involving US guidance by a radiologist. As detailed in the Results section, the US examination allowed the assessment of a series of features characterising multinodularity, nodule size, solid or cystic nodule (with three levels of cystic changes: +, ++ and +++), endocystic proliferation, calcification, echogenicity, echogenicity pattern and significant regional lymphadenopathy (lymph node diameter > 1 cm). A US probe was placed on the neck perpendicular to the thyroid during all the procedures, allowing clear visualisation of the needle movements in the nodule. The aspiration was carried out using a 21-gauge needle attached to a 10-ml syringe. The aspirated material was smeared on slides (around three slides per aspiration), air-dried and subjected to a difficult stain. The number of passes depended on the nodule size and the smear cellularity (enabling two passes for the majority of the patients).

Cytological diagnosis

Based on our personal experience, our cytological diagnoses were classified into four major categories as detailed below, and in the case of repeated FNAs with different diagnoses, the most ‘malignant’ diagnosis was kept for the patient:

i) Unsatisfactory or non-diagnostic samples: this involved specimens that could not be diagnosed because of poor fixation, poor cell preservation or hypocellularity (i.e. less than eight clusters of well-preserved cells at least on each of two slides).

ii) Benign diagnosis: this involved specimens categorised as benign, which included i) colloid nodules with abundant colloid and benign follicular cells. ii) Colloid nodules with cystic changes consisting of abundant colloid-containing macrophages and benign follicular cells and iii) lymphocytic thyroiditis presenting lymphocytes, plasma cells and oncocytic follicular cells.

iii) FP diagnosis: this category was subclassified into three grades as follows:
   a) Grade 1 FP: this is applicable to specimens with variable amounts of colloid, including some sheets of follicular cells presenting with low architectural atypia and without cellular atypia. In these specimens, the pathologist favoured a follicular neoplasm, but a benign lesion could not be excluded.
   b) Grade 2 FP: this is applicable to hypercellular specimens with scant colloid and sheets of follicular cells and with numerous atypical architectural features. In these specimens, the pathologist favoured a follicular neoplasm.
   c) Grade 3 FP: this is applicable to specimens with grade 2 FP criteria without a papillary pattern. In addition, these samples presented cytological features suggestive of papillary carcinoma (PC), such as nuclear grooves or ground glass nuclei, but not in sufficient quantity or quality for a definitive diagnosis of the follicular variant of PC.

iv) Malignant diagnosis: this involved specimens that had cytological features of malignant neoplasms including PC, medullary carcinoma (MC) or anaplastic carcinoma (AC). Lymphoma and metastasis are also included.

Surgical indications

Cytological diagnoses of FP and malignancy were considered to be a surgical indication. Certain patients with benign RNA diagnoses were referred for surgery when their clinical or US follow-up suggested a high suspicion of carcinoma, or when the symptoms were caused by nodules and/or patient anxiety.
**Histological diagnosis**

The histological diagnoses were assessed according to the criteria of WHO classification (9). The differential diagnosis of thyroid tumours with follicular patterns is an important problem for the pathologist (10). For most of the cases, WHO criteria can be easily applied to distinguish follicular adenoma, follicular carcinoma and the follicular variant of PC. Nevertheless, the diagnosis remains difficult for a few cases that do not perfectly satisfy the criteria. For instance, the diagnosis of minimally invasive follicular carcinoma continues to be controversial, and the poor reproducibility of this diagnosis is well documented (11). For such cases, we therefore decided to use the terminology of 'follicular tumour with uncertain malignant potential' (FT-UMP). We believe that such a terminology better reflects the clinical and morphological realities. The distribution of the different benign and malignant categories in our series are detailed in the Results section.

**Data analysis**

Taking the histological diagnosis as the gold standard, we first compared benign and malignant cases in terms of clinical and US features by means of univariate analyses (Mann–Whitney test and $\chi^2$ or Fisher’s exact test in the case of quantitative and qualitative features respectively). We then evaluated the sensitivity, specificity, positive and negative predictive values and global accuracy of the cytological diagnosis for identifying the malignancy.

To refine the predictive value in the case of FP, we investigated the contribution of the clinical and US features in discriminating between benign and malignant diagnoses by using a decision tree approach (12). To construct a decision tree, the algorithm recursively splits the training dataset into subsets to enrich each subset in one of the classes as much as possible. At each step, the algorithm generates univariate tests based on the observed feature values, and selects the best one by evaluating how well the resulting data split can differentiate observations from the different classes. The algorithm then recurses on each resulting subset until no further gain in class homogeneity can be achieved. To be more accurate on unseen cases, the decision tree is then usually pruned by discarding one or more sub-trees and replacing them with terminal nodes labelled by the majority class (12). In the present study, we tested three different criteria (i.e. Gini index, entropy and misclassification error) to construct trees, and their respective performances were evaluated by means of cross-validation to select the best approach (12). We also used cross-validation for selecting stable features in order to extract robust information from our data, as already detailed elsewhere (13, 14).

The statistical analyses were carried out using Statistica (Statsoft, Tulsa, OK, USA) and R, an open-source software environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria).

**Results**

**Clinical series analysis**

Between 1986 and 2007, 5283 patients were subjected to FNA at Erasme Hospital. Among them, 924 patients underwent surgery at Erasme Hospital. We focused our analysis on the latter series of 924 patients whose histological diagnoses were available. They included 766 cases with a histological diagnosis of benign disease, 15 well-differentiated FT-UMP and 143 cancers. A large majority of the patients were females (77.8%). The mean age was 47 years (range: 6–93 years). Table 1 details the breakdown between patients with benign and malignant histological lesions in terms of clinical and US features, and illustrates significant differences between these two groups ($P$ values <0.05). In particular, we found significant associations between malignancy and higher patient’s age, increased thyroglobulin levels and cold nodules. In addition, malignancy was significantly more often observed in unique, solid or hypoechoic nodules. Finally, lymphadenopathy detected by US was significantly associated with malignancy. Among the 26 patients with malignant nodules and regional lymphadenopathy, 21 had metastatic lymphadenopathy. 1 had benign lymphadenopathy and 4 were not evaluated. Among the six patients with benign nodules and regional lymphadenopathy, five had benign lymphadenopathy and one was not evaluated. In contrast, benign and malignant nodules did not significantly differ with regard to patient’s gender, thyroid antibodies, nodule size, calcification, endocystic proliferation or echogenicity pattern (Table 1).

**FNA diagnostic value**

In the 5283 patients subjected to FNA, cytological diagnoses were categorised as follows: 222 unsatisfactory cases (4.2%), 4091 benign diagnosis (77.4%), 827 FP (15.7%), and 143 malignant diagnosis (2.7%). In the subseries of 924 patients who underwent surgery in our hospital, cytological diagnoses were categorised as follows: 24 unsatisfactory cases (2.6%), 118 colloid nodules (12.8%), 295 colloid nodules with cystic changes (31.9%), 13 lymphocytic thyroiditis (1.4%), 78 grade 1 FP (8.4%), 271 grade 2 FP (29.3%), 46 grade 3 FP (5.0%), 50 PC (5.4%), 9 MC (1.0%), 6 AC (0.6%), 5 metastases (0.5%), 5 lymphomas (0.5%) and 4 cases indeterminate between AC and metastasis (0.4%). The correspondence between cytological and histological diagnoses is detailed in Table 2.
Regarding the four major cytological diagnosis categories, we observed a 4.2% risk of malignancy for unsatisfactory specimens, 1.6% for benign diagnosis, 19.0% for FP and 94.9% for cancers.

Among the 474 patients with a cytological diagnosis who were referred for surgery (FP or cancer), 390 (82.3%) had a histological diagnosis of adenoma or cancer (including FT-UMP). Regarding the surgical indication, the sensitivity and the specificity of FNA were 80.4 and 79.8% respectively for the histological diagnoses of adenoma and cancer. The positive and negative predictive values were 82.3 and 77.7% respectively, with a global accuracy of 80.1%.

In agreement with other series (15, 16), our data demonstrated that strictly benign or malignant FNA results (i.e. excluding FP) show a great concordance with benign or malignant histological diagnoses (including FT-UMP). This concordance is highlighted by the grey areas in Table 2. FNA sensitivity and specificity for the diagnosis of malignancy reached 91.5 and 99.1% respectively. The positive and negative predictive values are 94.9 and 98.4% respectively, with a global accuracy of 97.8%. These results indicate that false-positive FNA results occurred for 4/79 (5.1%) patients, and that false-negative FNA results occurred for 7/426 patients (1.6%).

Our data also showed that the accuracy of strictly benign/malignant cytological diagnoses slightly decreased to 96.9% in the case of the 65 patients (including ten cancers) who benefited from only one FNA. In contrast, repeated FNAs (at least one time) increased the global accuracy to 98.8%.

**Multivariate risk model of malignancy for patients with graded FP FNA diagnosis**

Of the 395 patients with a cytological diagnosis of FP, the prevalence of malignancy (including FT-UMP and all thyroid carcinomas) increased in direct relation with the three FP grades that we defined. Table 2 illustrates the increased risk of malignancy from grade 1 FP (7.7%) to grade 2 FP (17.7%), and finally grade 3 FP (45.7%) \( (P=10^{-6}) \).

In order to improve malignancy prediction in each FP grade, we used the decision tree approach to discriminate between benign and malignant lesions on the basis

---

Table 1 Clinical and ultrasonographic data.

<table>
<thead>
<tr>
<th>Patient characteristics</th>
<th>Benign</th>
<th>Malignant&lt;sup&gt;a&lt;/sup&gt;</th>
<th>( P ) value&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years; mean ± S.E.M.)</td>
<td>46.1 ± 0.5</td>
<td>50.6 ± 1.3</td>
<td>0.002</td>
</tr>
<tr>
<td>Sex ratio (F/M)</td>
<td>599/167</td>
<td>120/38</td>
<td></td>
</tr>
<tr>
<td>Thyroid functional status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td>5</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Euthyroidism</td>
<td>689</td>
<td>131</td>
<td></td>
</tr>
<tr>
<td>Hyperthyroidism</td>
<td>27</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Thyroglobulin (ng/ml; mean ± S.E.M.)</td>
<td>204 ± 19</td>
<td>602 ± 174</td>
<td>( &lt;10^{-6} )</td>
</tr>
<tr>
<td>Thyroid antibodies (( -/+ ))</td>
<td>509/118</td>
<td>90/31</td>
<td>NS</td>
</tr>
<tr>
<td>Nodule characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radionuclide scan</td>
<td></td>
<td></td>
<td>0.02</td>
</tr>
<tr>
<td>Hypofunctioning</td>
<td>581</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>Isofunctioning</td>
<td>80</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Hyperfunctioning</td>
<td>41</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Ultrasonographic features</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unique/multinodular</td>
<td>217/538</td>
<td>81/72</td>
<td>( &lt;10^{-6} )</td>
</tr>
<tr>
<td>Mean size (mm; mean ± S.E.M.)</td>
<td>34.1 ± 0.5</td>
<td>32.7 ± 1.3</td>
<td>NS</td>
</tr>
<tr>
<td>Cystic + + +</td>
<td>124</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cystic + +</td>
<td>208</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Cystic +</td>
<td>249</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Solid</td>
<td>234</td>
<td>106</td>
<td></td>
</tr>
<tr>
<td>Endocystic proliferation (( -/+ ))</td>
<td>723/16</td>
<td>157/1</td>
<td>NS</td>
</tr>
<tr>
<td>Calcification (( -/+ ))</td>
<td>288/447</td>
<td>66/84</td>
<td>NS</td>
</tr>
<tr>
<td>Echogenicity</td>
<td></td>
<td></td>
<td>( &lt;10^{-6} )</td>
</tr>
<tr>
<td>Hypoechoic</td>
<td>271</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>Isoechoic</td>
<td>317</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>Hyperechoic</td>
<td>92</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Homogeneous/heterogeneous pattern</td>
<td>185/543</td>
<td>49/97</td>
<td>NS</td>
</tr>
<tr>
<td>Lymphadenopathy (( -/+ ))</td>
<td>732/6</td>
<td>128/26</td>
<td>( &lt;10^{-6} )</td>
</tr>
<tr>
<td>Cases</td>
<td>766</td>
<td>158</td>
<td></td>
</tr>
</tbody>
</table>

NS, non-significant \( (P>0.05) \). Cystic + + +, totally cystic; cystic + + , moderate cystic change; cystic + , minimal cystic change.

<sup>a</sup>Including 15 follicular tumours with uncertain malignant potential (FT-UMP).

<sup>b</sup>Computed by means of a Mann–Whitney test in the case of quantitative features and a \( \chi^2 \) or Fisher’s exact test in the case of qualitative features.
Table 2 Cytological–histological diagnosis correspondence.

<table>
<thead>
<tr>
<th>Cytological diagnosis</th>
<th>n</th>
<th>Benign (%)</th>
<th>Adenoma (%)</th>
<th>Cancer (%)</th>
<th>MNG</th>
<th>Benign$^b$</th>
<th>AD</th>
<th>FT-UMP</th>
<th>PC</th>
<th>FV PC</th>
<th>FC</th>
<th>MC</th>
<th>PDC</th>
<th>AC</th>
<th>Metastasis</th>
<th>Lymphoma</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unsatisfactory</strong></td>
<td>24</td>
<td>62.5</td>
<td>33.3</td>
<td>4.2</td>
<td>14</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Benign lesions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colloid nodule</td>
<td>118</td>
<td>71.2</td>
<td>27.1</td>
<td>1.7</td>
<td>83</td>
<td>1</td>
<td>32</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cystic colloid nodule</td>
<td>295</td>
<td>80.7</td>
<td>18</td>
<td>1.3</td>
<td>235</td>
<td>3</td>
<td>53</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lymphocytic thyroiditis</td>
<td>13</td>
<td>69.2</td>
<td>23.1</td>
<td>7.7</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Follicular proliferation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 1</td>
<td>78</td>
<td>38.5</td>
<td>53.8</td>
<td>7.7</td>
<td>29</td>
<td>1</td>
<td>42</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Grade 2</td>
<td>271</td>
<td>18.5</td>
<td>63.8</td>
<td>17.7</td>
<td>40</td>
<td>10</td>
<td>173</td>
<td>8</td>
<td>7</td>
<td>22</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Grade 3</td>
<td>46</td>
<td>4.3</td>
<td>50</td>
<td>45.7</td>
<td>1</td>
<td>1</td>
<td>23</td>
<td>1</td>
<td>8</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Cancer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papillary C.</td>
<td>50</td>
<td>2</td>
<td>4</td>
<td>94</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>35</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Medullary C.</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Anaplastic C.</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Metastasis</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Anaplastic C. vs metastasis</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>5</td>
<td>20</td>
<td>0</td>
<td>80</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>924</td>
<td>62.5</td>
<td>33.3</td>
<td>4.2</td>
<td>405</td>
<td>25</td>
<td>336</td>
<td>15</td>
<td>52</td>
<td>44</td>
<td>11</td>
<td>11</td>
<td>3</td>
<td>10</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

The grey areas indicate the cases concordantly classified as strictly benign or malignant by both cytological and histological diagnoses. The percentages in bold are computed per histological category (i.e. benign, adenoma or cancer). The numbers in bold are the marginal absolute frequencies (i.e. total numbers of cases computed per line or per column). CN, colloid nodule; MNG, multinodular goitre; AD, adenoma; FT-UMP, follicular tumours with uncertain malignant potential; PC, papillary carcinoma; FVPC, follicular variant papillary carcinoma; FC, follicular carcinoma; MC, medullary carcinoma; PDC, poorly differentiated carcinoma; AC, anaplastic carcinoma.

$^a$Cancer % include FT-UMP cases.

$^b$Benign lesions included lymphocytic thyroiditis, amyloidosis and Graves’ disease.
of clinical and US features. The decision tree-based analyses provided the best results (evaluated by means of cross-validation, see Materials and methods) when the three FP grades were submitted to separate analyses. Figure 1 details the resulting patient stratification based on features robustly selected by the decision tree approach (i.e. stable features, see Materials and methods). This result shows that the extracted risk profiles differed between the three FP grades, and enabled us to identify high-risk groups in each grade (see grey areas in Fig. 1). All the threshold values that appeared in this decision tree were automatically selected by the algorithm to differentiate at best benign and malignant histological diagnoses.

**Grade 1 FP risk profile** For patients with grade 1 FP, the patient’s age discriminates between a very low-risk group (2.0% of malignancy) of younger patients (<50 years) and an older group (>50 years) with a higher malignancy risk, i.e. 17.9%, compared with the global risk associated with grade 1 FP (7.7%).

**Grade 2 FP risk profile** For patients with grade 2 FP, the decision criteria was more complex, but they still allowed us to identify a high-risk subgroup associated with a 36.1% malignancy risk compared with a global risk of 17.7% in this FP grade. This risk subgroup is characterised by a unique nodule with a diameter >35 mm. Additionally, we observed a subgroup with a low risk of malignancy (5.6%) consisting of patients younger than 46 years with a small (≤35 mm) and unique nodule.

**Grade 3 FP risk profile** Finally, a patient group with grade 3 FP and high thyroglobulin level (>400 ng/ml) presents a very high risk of malignancy (75.0%) compared with a global risk of 45.7% in this FP grade.

It should be noted that all the features included in the decision tree characterising FP were also identified in Table 1 as very significant for distinguishing between strictly benign and malignant histological diagnoses.

**Discussion**

In the present study, we investigated the clinical and US features together with FNA-based cytological diagnosis as malignancy risk factors for thyroid nodules. Our univariate analyses identified increased patient’s age, high serum thyroglobulin levels and the presence of a cold nodule as factors that are significantly associated with malignancy. While the literature agrees with our results on age and cold nodules (2, 15), the predictive value of thyroglobulin is not so clearly established (17). Generally, thyroglobulin assessment is not recommended as a routine initial evaluation of thyroid nodules (1, 18). However, measurement of serum thyroglobulin is primarily used as a tumour marker in
the postoperative management of patients with differentiated thyroid cancer (18). Additionally, in the literature, a number of US characteristics are associated with an increased risk of malignancy; these characteristics include large nodule size, hypoechogenicity, microcalcifications, solid nodules and the presence of lymphadenopathy (1, 2, 19). In agreement with the previous studies, we identified lymphadenopathy, solid nodules and hypoechogenicity as factors that are significantly associated with malignancy.

US-guided FNA remains the most sensitive investigational procedure in the evaluation of patients with thyroid nodules. However, as emphasised by various authors, the FNA technique and interpretation should be performed by an experienced team to increase the quality of FNA results (20, 21). The US guidance for FNA is certainly essential for decreasing the occurrence of unsatisfactory specimens. In our study, all FNAs were performed using US guidance, resulting in a very low percentage of unsatisfactory specimens (2.6%) compared with the range reported in the literature (10–25%) (22).

The major limitation of thyroid FNA remains in the management of patients with an FP FNA diagnosis because it includes both benign and malignant histological diagnoses. In the complete series of 5283 FNAs, this diagnosis represented 15.7%, which was in agreement with the 15–30% prevalence rate that has been reported previously (2, 20). The selection of patients who underwent surgery inevitably increased this prevalence (to 42.7% in our series of 924 patients), because the cytological FP diagnoses constituted a surgery indication. Another increase was observed for the cancer FNA diagnoses, which increased from 2.7% in the complete series to 8.5% in the subseries. Meanwhile, benign FNA diagnoses decreased from 77.4 to 46.1%. While affecting global accuracy measurements (combining the different diagnostic groups), these modifications in group prevalence theoretically do not impact specificity, sensitivity and predictive values, except if patient selection biases modify population characteristics (23). In the present study, both FP and malignant FNA diagnoses constituted surgery indications. In these diagnostic groups, the subsets of patients who underwent surgery at our hospital could be reasonably considered as being representative and unbiased samples of the complete FP and malignant FNA groups. The possible bias in our series essentially involves patients with benign FNA diagnoses. In fact, some of them could be referred for surgery when their clinical or US follow-ups indicated a high suspicion of malignancy. Consequently, the false-negative rates reported in the present study could be pessimistically biased, i.e. higher than their actual values.

In the literature, the reported incidences of false-positive FNA and false-negative FNA range from 0 to 9% and from 1 to 7% respectively (1, 15, 24–29). As highlighted by Yang et al. (16), the large variations of the reported FNA sensitivity and specificity were attributable mainly to the data analysis methods. Most studies exclude the indeterminate cytological categories (i.e. FP) from the computation. This is the reason why two accuracy analyses were carried out in the present study: one studying the accuracy of surgical indication because FNA had been used by clinicians as a screening test to determine whether surgery is required, and the other studying the accuracy of malignancy detection. The first analysis was carried out for the FNA diagnosis of neoplasm (including FP), and the false-positive and false-negative rates were 17.7 and 22.3% respectively; the second analysis was carried out for distinguishing between benign and malignant FNA diagnoses (excluding FP), and the false-positive and false-negative rates were 5.1 and 1.6% respectively.

To refine the FP FNA diagnosis, we introduced a three-level grading system. This grading system was implemented before the publication of the NCI FNA guidelines (3, 30) (which was introduced in 2007–2008), and was improved over time. By including unsatisfactory, benign and malignant FNA diagnoses, we currently use a six-category system as favoured by the NCI FNA guidelines (unsatisfactory, benign, atypia of undetermined significance, follicular neoplasm, suspicious for malignancy and malignant). However, there are two differences pertaining to the NCI FNA guidelines: i) in grade 1 FP, the pathologist favours a follicular neoplasm with a very low probability of cancer. For clinicians, the choice of close clinical follow-up is justified, and a novel FNA can be performed after 6 or 12 months. In the NCI guidelines, the terminology ‘atypia of undetermined significance’ is defined as cytological findings not convincingly benign, and is not sufficiently informative to allow the clinicians to make an adequate therapeutic decision. It should be noted that some members of Committee IV suggested that this category should be optional (3); ii) as opposed to the NCI classification, we used the diagnosis of ‘suspicious for malignancy’ only for the follicular variant of PC (grade 3 FP). Our subclassification of FP presents a gradation in malignancy incidence (i.e. 8, 18 and 46% in our series) similar to the recently reported figures (5–10, 20–30 and 50–75%) (3). The new guidelines refer patients with a nodule diagnosed with an FP level of 2 or 3 for surgery (i.e. follicular neoplasm or suspicious for malignancy), whereas close clinical follow-ups with repeat US-guided FNA are recommended for the FP level 1 diagnosis (i.e. atypia of undetermined significance) (3).

We observed that our three FP grades included very different benign and malignant profiles in terms of clinical and US features. For this reason, we combined all the available information by means of a decision tree approach to propose a refined decision aid tool for surgical indication in the problematic cases of FP diagnosis. The obtained model stratifies patients into
different levels of malignancy risk on the basis of the FP grade combined with one or more features (which vary with the FP grade) including the patient’s age, thyroglobulin level, nodule size and uni/multinodularity (Fig. 1). More specifically, this stratification identified two patient subgroups in grade 2 FP (referred for surgery by the recent recommendations) with a very low risk of malignancy (5–8%), and who could thus be managed conservatively, i.e. close clinical follow-up with repeat US-guided FNA (31). In contrast, a subgroup of grade 1 FP cases (referred for clinical follow-up by the recent recommendations) exhibited a malignancy risk of 18%, and should thus be referred for surgery.

Our results agree with those obtained from a recent cohort of 468 patients with indeterminate or suspicious FNA for which age, nodule size and cytology were identified as risk factors for thyroid malignancy by a multivariate analysis (7). However, it should be noted that the indeterminate FNA category used in the latter study was more heterogeneous than ours due to its inclusion of Hurthle cell neoplasm, atypia, suspicious for PTC, follicular neoplasm and thyroid neoplasm. In addition, other studies analysing follicular neoplasms or indeterminate FNA results identified nodule size and age as malignancy factors (15, 32).

The present study proposes decision aid tools to refine patient stratification in FP cytological diagnoses. To further validate this proposal, we are planning a prospective study in which additional biochemical (such as serum TSH) and ultrasonographic features (such as Doppler studies, elastography and calcification evaluation) will be included so that data from recent literatures can be taken into account (33–35). In addition to clinical and ultrasonographic features, other nodule characteristics, such as those provided by biomarkers, could be useful for preoperative risk stratification, such as investigations for CD57 immuno-nostaining (5) and, more recently, BRAF mutation status (6) in FNA specimens.

In conclusion, this study confirms that US-guided FNA is a sensitive investigational procedure for the evaluation of patients with thyroid nodules. Moreover, for the indeterminate class of FP cytology, integration of FP grade together with clinical and US features allows the stratification of patients according to their risk of malignancy, enabling a possible refinement of their treatment options.

Declaration of interest
The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding
This work was supported by grants awarded by the Fonds Yvonne Boël (Brussels, Belgium).

References

www.eje-online.org