PREOPERATIVE PREDICTION OF THE RISK OF MALIGNANCY IN THYROID NODULES

WALTER KUNZ1*, AYMAN MISMAR2*, GEORGE WILLE1, ROHANA AHMAD1, GABRIELE MATERAZZI3, PAOLO MICCOLI4

1Department of General Surgery/ABC Medical Centre, (Mexico City- Mexico) - 2Department of General Surgery/ Faculty of Medicine/The University of Jordan (Amman- Jordan) - 3Department of Thoracic and Visceral Surgery, Kantonsspital Winterthur. (Switzerland) - 4Department of Breast and Endocrine Surgery, Hospital Putrajaya. (Malysia) 52nd Department of General Surgery / Faculty of Medicine/University of Pisa (Pisa- Italy)

*W. Kunz and A. Mismar contributed equally to this work.

Introduction

There has been an increased frequency of diagnosis of the common and widespread thyroid nodular disease because of the use of ultrasound screening tests. The reported incidence ranges from 19% to 76%1. The importance of detection of thyroid nodules resides in the capability to differentiate benign from malignant nodules (<5%) and to treat them in a safer and less invasive manner2. The thyroid nodules are studied using different tools like needle aspiration cytology/histology, serum concentrations of TSH, T4, T3, calcitonin and imaging studies like ultrasonography (US), Magnetic Resonance Imaging (MRI), computed tomography (CT) and sometimes by using techniques such as spectroscopic analysis of the MRI3.

Thyroid US is the first line of imaging study to visualize the gland. Combined with color flow-doppler and elastographic analysis of thyroid nodules, this test has yielded promising results in pre-operatively determining the nature of the thyroid nodules. Although, US features like presence of incomplete halo sign, interrupted margins, micro-calculifications and hypoechogeticity of the nodule are reported as predictors of malignancy, they yield low positive and negative predictive values4. Similarly, the presence of intranodular (Type III)

ABSTRACT

Objective: This study aimed to propose a formula for predicting the biological nature of thyroid nodules, thus helping in clinical decision making and avoiding unnecessary surgery.

Patients and methods: Prospective, longitudinal and observational study was conducted on 400 patients submitted to primary thyroid surgery at the Department of Surgery, University of Pisa, Italy, between April 1st and June 30th 2008, whose clinical charts indicated gender, age, ultrasonographic findings, TSH concentrations, while not receiving thyroid hormones and whose definite histopathology report was available. Other variables recorded were: lymph node characteristics, estimated thyroid volume, thyroid functional tests, thyroglobulin, anti thyroid antibodies, calcitonin, fine needle aspiration cytology, frozen sections and vocal cord assessment through laryngoscopy. Statistical analysis was performed using Pearson correlation test and logistic regression backward elimination.

Results: We designed two formulas which included thyroid functional tests, calcitonin and fine needle aspiration with and without anti thyroid antibodies and demographic variables. These formulas could predict the biological nature of thyroid nodules with 92.3% sensitivity and 91.67% specificity, positive and negative predicting values of 88.5 % and 94.49%, respectively. Both formulas showed a better performance compared to fine-needle aspiration biopsy (FNA) alone.

Conclusion: We developed a formula that can predict the biological behavior of thyroid nodules in Italian patients, their applicability in other countries requires further study.

Key words: Thyroid nodule, Thyroid neoplasms, Thyroid carcinoma, Diagnosis, Prediction of malignancy risk.
flow seen at the color flow-doppler and non-compressible nodules at elastography has shown variable specificity and sensitivity\(^9\). Other drawbacks are the lack of a standardized technique, inter-observer variation and small series of patients\(^9\).

Fine needle aspiration (FNA) is still the best single test for discriminating malignant thyroid nodules. However, the diagnostic efficacy of FNA declines sharply in case of follicular patterned lesions of the thyroid. These lesions (adenomas, carcinomas and follicular variants of papillary carcinomas) are diagnosed as follicular neoplasms with malignancy rates ranging from 10 to 46%. Therefore, 54-90% of the patients presenting follicular lesions undergo unnecessary surgery because FNA cannot distinguish benign lesions from the malignant ones on the basis of cyto-morphology (7). Molecular markers (HBME1, Galectin, CK19, CITED1, Cyclin D1, and HMWCK) have yielded promising results, reaching a diagnostic accuracy of 95.3% with positive predictive values of 100%. But this method is expensive and requires large needle aspiration biopsies instead of fine needle aspiration\(^9\).

Some reports described the use of epigenetic and peripheral blood markers in patients with equivocal FNA\(^9, 10\). Though useful, these markers have limited sensitivity and a limited negative predictive value; therefore, they fail to detect more than 33% of cancers\(^11-13\).

Studies have described the development of gene-expression classifiers that better distinguish benign from malignant thyroid nodules\(^12\).

Gene expression profiling using DNA microarrays or serial analysis of gene expression (SAGE) has revealed several hundreds of genes that are differentially expressed between malignant and benign thyroid nodules\(^15, 16\). However, several studies have demonstrated that none of these genes individually has sufficient sensitivity and specificity to be exploited as an independent diagnostic biomarker. On the other hand, using very large panels of genes may be expensive and not practical\(^17, 18\).

Until now, none of the preoperative investigations have satisfactorily identified the subset of patients to be electively investigated by fine needle aspiration cytology (FNAC) or able to accurately predict the presence of malignant disease in thyroid nodules.

Boelaert et al. designed a formula to predict the preoperative risk of malignancy in a thyroid nodule for each individual patient; this formula considered clinical variables (age, gender, goiter type) and TSH concentration but it does not include FNA, which is one of the most informative preoperative tests\(^19\).

We set out to find a formula that can predict the risk of malignancy of thyroid nodules combining all available preoperative variables. The benefits of this formula include a significant reduction in surgical morbidity, reduction of patient stress, and facilitating definitive management of patients with surgical requirements.

**Material and methods**

The study was prospective, observational, longitudinal, held at the Department of Surgery, University of Pisa, Italy, which houses the WHO Collaborating Center for the Diagnosis and Treatment of Thyroid Cancer and Other Thyroid Diseases. All patients who underwent primary thyroid surgery between the 1st April - 30th June 2008, not treated with thyroid hormones and whose final histopathological diagnosis was available were included in the study. During the period of study, 694 patients were operated in the second department of General Surgery. Out of them, 182 operations done in the day surgery facility were eliminated to reduce variables. Of the remaining 512 patients, only 417 met the inclusion criteria, 17 were eliminated due to inability to get the final results of pathology. In order to maintain the initial protocol guidelines approved by the Ethics Committee of Santa Chiara Hospital, a series of 400 patients were accepted and analyzed.

The variables collected were: goiter type as assessed by ultrasonography, characteristics of regional nodal groups to ultrasound, estimated thyroid volume (VTE), thyroid functional tests including thyroglobulin (TG), thyroglobulin antibodies (ABTG), anti-peroxidase (AbTPO), anti-TSH receptor (TRAb), calcitonin, FNAC, intraoperative studies and evaluation of mobility of vocal cords by laryngoscopy. Data was collected by one of three investigators (WK, AM and RA) within the operating room prior to surgery.

The final histopathological result was collected by an investigator blinded to the predicted probability of malignancy.

**Statistical analysis**

The diagnosis result was defined as presence or absence of malignancy. The two groups were
analyzed to describe their characteristics, and correlation was made between the predicted probabilities of malignancy (expressed as a percentage) and the histopathology results. The type of goiter was classified into diffuse, multinodular, solitary and dominant nodules. Because of the disparity between physical examination and ultrasonography, a solitary nodule was defined when the rest of the gland was normal or had micro nodules gland less than 1 cm (not palpable on physical examination). A dominant nodule was defined when it had a diameter of more than twice the largest of the non-dominant nodules. The presence of goiter was classified dichotomously (1 = present, 0 = absent) and was included in the analysis. Similarly, the existence of suspicious cervical lymph nodes was defined irrespective of the level in which it was found. A numerical value was assigned to the gender of the patients (1 for women and 2 for men). Serum thyroid hormones and antibodies, and thyroid volume estimate were included with the corresponding measurement units. The result of the FNA included in accordance with the classification of the British Thyroid Association.

We performed a Pearson correlation test to select useful variables. Logistic regression was performed gathering clinical and investigation variables. Variables with p <0.05 were considered significant and retained in the model. The calculated percentages were plotted as graphs of the receiver operating characteristic (ROC). Then we compared the independent ROCs to determine the significant differences. The formula with the best overall performance, which incorporated only clinical variables, was compared with the FNA ROC. Analyses were performed using Microsoft Office Excel ® and MedCalc ® for Windows Version 11.2 (Mariarke, Belgium).

Results

In our series, 230 patients had benign pathologies while 170 were malignant; the distribution is shown in Table 1.

Through the Pearson correlation analysis, age, the presence of lymph node enlargement, the FNA diagnostic classification and TSH concentrations were correlated with the diagnosis of carcinoma as shown in Table 2.

We obtained the formula for Probability = 1 / (1-e-z), where e is the logarithm base value equals 2.71828182845904 and Z = 0.8606 - 0.01431 (Age) + 0.5033 (if there is a solitary nodule) + 1.57775 (if there are swollen lymph nodes). This formula, which included only clinical data (age, sex, type of goiter, Abnormal Vocal Cord Exploration) had an area under the curve (AUC) of 0.646 with standard error (SE) of 0.0285 with a confidence interval of 95% from 0.597-0.694.

<table>
<thead>
<tr>
<th>Benign n=230 (%)</th>
<th>Malignant n=170(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>145 (63) Multinodular Goiter</td>
<td>152* (89) Papillary ca.</td>
</tr>
<tr>
<td>62 (27) Follicular Adenoma</td>
<td>8* (4.7) Medullary ca.</td>
</tr>
<tr>
<td>17 (7.4) Graves disease</td>
<td>8* (4.7) Follicular ca.</td>
</tr>
<tr>
<td>1 (0.4) Schwannoma</td>
<td>2 (1.1) Poorly diff. ca</td>
</tr>
<tr>
<td>1 (0.4) Fibrosclerotic nodule</td>
<td>2 (1.1) Anaplásico ca</td>
</tr>
<tr>
<td>1 (0.4) Hashimoto thyroiditis</td>
<td></td>
</tr>
<tr>
<td>1 (0.4) Normal (Ret +ve)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Distribution of the cases presented in the series. *There were two papillary carcinomas coexisting with medullary carcinoma and follicular carcinoma.

The same inclusion criteria were employed in selecting paraclinical variables (thyroid profile, complete thyroid volume, thyroid antibody profile). The formula presented AUC of 0.788 with SE 0.0237 and 95% confidence interval (CI) from 0.745-0.827. By incorporating the findings of the FNA, a simple formula was obtained with AUC of 0.778 with SE 0.0242 and 95% CI from 0.734-0.818, where Z = -2.2289 + 0.7472 (FNA according to BTA) + 1.2323 (if lymphadenopathy was present).

The formula with the best performance was achieved by modifying the inclusion criteria of p <0.1, with backward elimination. It incorporated a large number of clinical and laboratory factors that were only available for analysis in 25 patients, where Z = -58.2268 -0.05319 ABTG + 0.02524 AbTPO – 66.5378 (if abnormal vocal cord exploration) + 2.5154 (Age) + 0.1284 calcitonin + 132.3299 (for diffuse goiter) –52.3352 (for dominant nodule) –70.3719 (for multinodular goiter) – 74.2454 (for a solitary nodule) + 23.1853 (FT3) –10.0076 (FT4) + 22.7265 (If there are cervical lymph nodes enlargement) - 5.7215 (gender: 1 for female, 2 for male). This formula presented a perfect performance with AUC of 1, 0.000 SE and 95% CI= 0.863-1, however, when applied to the entire sample, AUC fell to 0.788 because not all
patients had all necessary variables available as shown in Figure 1,2.

Additionally, a formula was obtained with similar performance, excluding thyroglobulin antibodies, peroxidase antibodies and evaluation of vocal cord motility while modifying the exclusion criteria if p > 0.9. This formula defined Z = -2.1724 + 0.0083844 (calcitonin) + 4.3706 (FT3) −2.0701 (FT4) + 2.4874 (Tir according to the BTA) −1.2673 (TSH). This formula presented with AUC of 0.949, SE 0.0465 and 95% CI 0780-0997. If 0.48 is used as the cut-off value, sensitivity is 92.3% (95% CI 64-99.8) and specificity of 91.67% (95% CI 61.5-99.8) with a positive and negative predictive value 88.50 to 94.49, respectively, as shown in Figure 3.

When comparing the performance of these two formulas we obtained a −0.051 difference with SE 0.047, z−1.081 and p= 0.279.

FNA alone was studied as a predictor of malignancy and it demonstrated an AUC of 0.789 with SE of 0.0258.

### Table 2: Studied variables and P value.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Benign N=230</th>
<th>Malignant N=170</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>54</td>
<td>45</td>
<td>0.5581</td>
</tr>
<tr>
<td>Age (years)</td>
<td>50.45</td>
<td>46.51</td>
<td>0.8711</td>
</tr>
<tr>
<td>Volume (cc)</td>
<td>46.52</td>
<td>30.87</td>
<td>0.2712</td>
</tr>
<tr>
<td>FT4 (pg/ml)**</td>
<td>11.79</td>
<td>11.16</td>
<td>0.0118***</td>
</tr>
<tr>
<td>FT3(pp/ml)**</td>
<td>4.02</td>
<td>3.82</td>
<td>0.0033***</td>
</tr>
<tr>
<td>TSH (U/ml)**</td>
<td>1.04</td>
<td>1.47</td>
<td>0.0288 95% CI 0.81a-0.04 p=0.0014** p&lt;0.0001***</td>
</tr>
<tr>
<td>TG (U/ml)*</td>
<td>63</td>
<td>50</td>
<td>0.0007 T test p=0.4115</td>
</tr>
<tr>
<td>ABTG (U/ml) *</td>
<td>206.33</td>
<td>148.78</td>
<td>0.1959**</td>
</tr>
<tr>
<td>ABTPO(U/ml)*</td>
<td>117.91</td>
<td>154.68</td>
<td>0.2037**</td>
</tr>
<tr>
<td>TRAB/U/L*</td>
<td>20.85</td>
<td>0.597</td>
<td>0.0893 p=0.0269**</td>
</tr>
<tr>
<td>Calcitonina (pg/ml)*</td>
<td>8.53</td>
<td>63.01</td>
<td>0.2501</td>
</tr>
<tr>
<td>VCA</td>
<td>225</td>
<td>163</td>
<td>0.3747</td>
</tr>
<tr>
<td>DG</td>
<td>10</td>
<td>7</td>
<td>0.001</td>
</tr>
<tr>
<td>MNG</td>
<td>113</td>
<td>64</td>
<td>0.0252</td>
</tr>
<tr>
<td>SN</td>
<td>60</td>
<td>65</td>
<td>0.0120</td>
</tr>
<tr>
<td>DN</td>
<td>45</td>
<td>34</td>
<td>0.001</td>
</tr>
<tr>
<td>CLN</td>
<td>10</td>
<td>37</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

* Fisher’s exact test (fp) was performed in these variables to identify statistical differences in the levels, wherever there was a statistically significant difference, a subgroup analysis was done.
** Indicates P value for the subgroup above the normal levels, and
*** Indicates P value for the subgroup below the normal level.

T test was used to evaluate the averages of two samples. VTE estimated thyroid volume in cubic centimeters, FT4 (7-17pg/ml), FT3 (2.7-5.7pg/ml), TSH, thyrotropin (0.4-3.4 uUI / ml), TG thyroglobulin (0-75 ng / ml), ABTG thyroglobulin antibodies (0-5 IU / ml), anti ABTPO peroxidase (0-10U/ml), TRAB TSH receptor antibodies (1 <IU / L), VCA assessment vocal cords, DG diffuse goiter, MNG multinodular goiter, SN solitary nodule, DN dominant nodule, CLN cervical lymphadenopathy.

Additionally, a formula was obtained with similar performance, excluding thyroglobulin antibodies, peroxidase antibodies and evaluation of vocal cord motility while modifying the exclusion criteria if p > 0.9. This formula defined Z = -2.1724 + 0.0083844 (calcitonin) + 4.3706 (FT3) −2.0701 (FT4) + 2.4874 (Tir according to the BTA) −1.2673 (TSH). This formula presented with AUC of 0.949, SE 0.0465 and 95% CI 0780-0997. If 0.48 is used as the cut-off value, sensitivity is 92.3% (95% CI 64-99.8) and specificity of 91.67% (95% CI 61.5-99.8) with a positive and negative predictive value 88.50 to 94.49, respectively, as shown in Figure 3.

When comparing the performance of these two formulas we obtained a −0.051 difference with SE 0.047, z−1.081 and p= 0.279.

FNA alone was studied as a predictor of malignancy and it demonstrated an AUC of 0.789 with SE of 0.0259.

### Fig. 1: The best performance formula presented with a AUC of 1 but its predictive strength was reduced when applied to the entire sample (AUC 0.788) due to missing information. Probability=1/(1-e−Z) where e is the logarithm base value and equals 2.71828182845904 and Z= -58.2268 - 0.05319 AbTG-0.02524 AbTPO -66.5378 (if abnormal vocal cord exploration) + 2.5154 Age + 0.1284 Calcitonin + 132.3299 (for diffuse goiter) - 52.3352 (for dominant nodule) - 70.3719 (for multinodular goiter) - 74.2454 (for solitary nodule) + 23.1853 (FT3) – 10.0076 (FT4) + 22.7265 (if there were enlarged cervical lymph nodes) – 5.7215 (Gender: 1 for female, 2 for male).

### Fig. 2: If a 0.48 cutoff value is used, the Kunz-Mismar proposal has a 92.3% sensitivity (95% CI 64-99.8) and 91.67% sensitivity (95% CI 61.5-99.8) with positive and negative predictive values of 88.5 and 94.49%, respectively.
Discussion

The fear of malignancy is one of the main concerning issues for the patients with thyroid nodules; the presence of a precise malignancy risk calculator will help them and their physicians to reach a wise decision whether to undergo surgery or observation. From the results, we reached a formula that might predict the risk of malignancy in any individual with a thyroid nodule depending on clinical and investigation variables. This equation included patients with all differentiated cancer originating either from follicular or neuroectodermal tissue and its studied performance covering papillary, follicular and medullary thyroid carcinoma.

We reached two equations to predict the individual risk of malignancy through the equation: Probability = 1 / (1 - e^-z). The first one (Kunz-Mismar formula) presented the best performance in our series in which z value is calculated as:

\[ Z = -58.2268 - 0.05319 \ ABTG - 0.02524 \ AbTPO - 2.1724 + 0.008384 \text{Calcitonin} + 4.3706 \text{FT3} - 2.0701 \text{FT4} + 2.4874 \text{(BTA class for FNAC)} - 1.2623 \text{(TSH)}. \]

A second compact formula was obtained; it is applicable in cases where some variables are not available:

\[ Z = -2.1724 + 0.008384 \text{Calcitonin} + 4.3706 \text{FT3} - 2.0701 \text{FT4} + 2.4874 \text{(BNA according to BTA)} - 1.2623 \text{(TSH)}. \]

Our formulas, present a better performance than those designed by Boelaert et al. (19) (AUC 0.949 vs. 0.541 p <0.05). In comparison to FNA alone, our formulas presented a better diagnostic performance with AUC of 0.789 with SE of 0.0259. In contrary to the McGill Thyroid Nodule Score and the recent experimental and commercial gene panels our formula includes the most frequently used clinical, serologic and imaging studies performed while evaluating thyroid nodules worldwide. It avoided the additional expense of genetic profiling that represents a major cost for developing nations with.

We believe that FNA is still the best single test for the evaluation of thyroid nodules, but combining it with other variables related to increased risk of malignancy will augment its diagnostic yield. Our formula gathers clinical, biochemical and cytological tests, and thus the personal specifications for each individual with a thyroid nodule are taken into consideration when calculating his or her risk of having a malignant disease.

A major drawback of our study is that it was developed within a homogeneous Italian society. Therefore, it needs validation in other centers with ethnic variations before being applicable in clinical practice. Additionally, it integrates data from specialized last generation serological tests that might not be available in all centers worldwide, which undoubtedly limits its applicability in low socioeconomic societies.

In conclusion, we provide a formula that can predict the biological behavior of thyroid nodules in Italian patients; however, their applicability in other countries requires further study.

References


