Quantification of cancer risk of each clinical and ultrasonographic suspicious feature of thyroid nodules: a systematic review and meta-analysis

Paolo Campanella*, Francesca Ianni*, Carlo Antonio Rota, Salvatore Maria Corsello and Alfredo Pontecorvi

Unit of Endocrinology, Università Cattolica del Sacro Cuore, Largo Agostino Gemelli 8, 00168 Rome, Italy *(P Campanella and F Ianni contributed equally to this work)

Correspondence should be addressed to S M Corsello
Email corsello.sm@mclink.it

Abstract

Objective: In order to quantify the risk of malignancy of clinical and ultrasonographic features of thyroid nodules (TNs), we did a systematic review and meta-analysis of published studies.

Methods: We did a literature search in MEDLINE for studies published from 1st January 1989 until 31st December 2012. Studies were considered eligible if they investigated the association between at least one clinical/ultrasonographic feature and the risk of malignancy, did not have exclusion criteria for the detected nodules, had histologically confirmed the diagnoses of malignancy, and had a univariable analysis available. Two reviewers independently extracted data on study characteristics and outcomes.

Results: The meta-analysis included 41 studies, for a total of 29 678 TN. A higher risk of malignancy expressed in odds ratio (OR) was found for the following: nodule height greater than width (OR: 10.15), absent halo sign (OR: 7.14), microcalcifications (OR: 6.76), irregular margins (OR: 6.12), hypoechogenicity (OR: 5.07), solid nodule structure (OR: 4.69), intranodular vascularization (OR: 3.76), family history of thyroid carcinoma (OR: 2.29), nodule size ≥4 cm (OR: 1.63), single nodule (OR: 1.43), history of head/neck irradiation (OR: 1.29), and male gender (OR: 1.22).

Interestingly, meta-regression analysis showed a higher risk of malignancy for hypoechoic nodules in iodine-sufficient than in iodine-deficient geographical areas.

Conclusions: The current meta-analysis verified and weighed out each suspicious clinical and ultrasonographic TN feature. The highest risk was found for nodule height greater than width, absent halo sign, and microcalcifications for ultrasonographic features and family history of thyroid carcinoma for clinical features. A meta-analysis-derived grading system of TN malignancy risk, validated on a large prospective cohort, could be a useful tool in TN diagnostic work-up.

Introduction

The widespread use of ultrasonography (US) has resulted in a dramatic increase in the prevalence of clinically unapparent thyroid nodules (TN), with a prevalence of 19–67% by US (1), and with higher occurrence observed in iodine-deficient areas (2). However, approximately only 5% are malignant (3).

Fine-needle aspiration (FNA) is the most accurate diagnostic test for TN, but its value is limited when cytological findings are inconclusive or samples are inadequate, which is the case for ~10–25% of all aspirates (4).

More interestingly, recently Cibas et al. (5) have reported a not negligible inter- and intraobserver
variability of preoperative cytopathologic diagnosis, demonstrating a concordance in only 64.0% of interobserver diagnoses and in 74.7% of intraobserver diagnoses.

Therefore, the role of clinical and US TN features predictive of malignancy should be reconsidered, regarding them not only as risk factors requiring FNA, as stated by guidelines (6, 7), but also as important elements supporting the final decision in the TN preoperative management.

Hence, we intended to conduct a systematic review and meta-analysis of published studies on this topic to quantify the risk of each clinical and US suspicious feature of TN.

**Subjects and methods**

This study was conducted and reported in accord with MOOSE guidelines for meta-analyses and systematic reviews of observational studies (8).

**Search strategy and study selection**

We performed a literature search of MEDLINE for studies published between 1989 and December 2012 that evaluated clinical or US TN features associated with an increased risk of malignancy using the following algorithm: ‘thyroid nodule’ AND (‘risk of malignancy’ OR ‘cancer risk’ OR ‘risk factor’ OR predict*).

Our search was restricted to English language publications. Studies were considered eligible for inclusion if they investigated the association between one or more clinical and/or US TN feature and the risk of malignancy, did not have any restriction criteria for the inclusion of detected nodules in the study (e.g. nodule size or thyroid-stimulating hormone (TSH) levels), had histologically confirmed diagnoses of malignancy, and reported univariable analysis (or reported data sufficient to perform it).

Titles and abstracts of all identified citations were screened by one reviewer for inclusion, and checked for discrepancies by a second reviewer. We subsequently obtained the full text for all potential citations and screened them for inclusion. Additional relevant publications were identified from the references of the initially retrieved articles.

**Data extraction**

From each study, we extracted data on the first author’s last name, year of publication, study location, number of patients included, method used to diagnose malignant and benign nodules, clinical and US features evaluated for risk of malignancy, and odds ratios (ORs) with their CIs resulting from univariable analysis (either published or derived from reported data). All data extraction was conducted independently by two reviewers, and disagreements were adjudicated.

**Data analysis**

Meta-analysis was performed for each clinical and US feature evaluated in the studies. Because of the expected heterogeneity among the studies, the random effects model was employed using DerSimonian & Laird’s method (9). Heterogeneity was quantified using the Cochran Q test and I² statistics (10).

Sensitivity analyses were conducted by excluding one study at a time from the meta-analysis to determine whether results of the meta-analysis were influenced by individual studies and whether risk estimates and heterogeneity were substantially modified.

Random-effect meta-regression models (11) were fitted to explore whether the extracted study variables could explain the presence of heterogeneity. Variables included in the meta-regression were year of publication, study location (iodine-deficient areas vs iodine-sufficient areas according to WHO), and method used for diagnosis of benign TN (histologically confirmed or clinical observation in patients who did not undergo surgery).

The presence of publication bias was assessed using a visual funnel plot inspection and Egger’s test (12). If the funnel plot was asymmetric and the Egger test reported a P value < 0.05, we assumed publication bias was present. If publication bias was present, the Duval & Tweedie (13) non-parametric ‘trim and fill’ method was used to adjust for it. All statistical tests were performed with Comprehensive Meta-Analysis Software, version 2.2.064 (Biostat, Engelwood, NJ, USA).

**Results**

**Search results and study characteristics**

The initial literature search resulted in a total of 749 citations identified. Of these, 235 studies were retrieved for further investigation, and 41 studies (14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54) met the final selection criteria and were included in the analysis.
All 41 studies were observational studies published in peer-reviewed journals between 1989 and 2012. Sample size varied largely across studies, ranging from >1000 patients in seven studies to <100 patients in seven studies, and >100 to <1000 patients in 27 studies. Twenty-two studies were conducted in iodine-sufficient areas and 19 in iodine-deficient areas.

The sum of TNs included was 29,678 in more than 10,000 patients. All diagnoses of malignancy were histologically confirmed. In 23 studies where all patients underwent surgery, the diagnosis of benign TN was all histologically confirmed, and in 18 studies that included subjects who did not undergo surgery, the diagnosis of benign TN was histologically confirmed in patients who underwent surgery and by clinical observation in patients who did not undergo surgery.

Results of univariable analyses (or data sufficient to perform it) were reported in 21 studies for microcalcifications, in 17 studies for hypoechogenicity, in 16 studies for male gender, in 15 studies for irregular margins, in 15 studies for intranodular vascularization, in 12 studies for nodule size greater than width, in ten studies for solid nodule structure, in 17 studies for hypoechogenicity, in 16 studies for intranodular vascularization, in 15 studies for irregular margins, in 15 studies for intranodular vascularization, and in 12 studies for nodule height greater than width, absence of the halo sign, microcalcifications, irregular margins, hypoechogenicity, solid nodule structure, intranodular vascularization, nodule size varied largely across studies, ranging from 4 cm, and single nodule as US TN features (Fig. 2).

In contrast, meta-analysis did not show any statistically significant difference in the rate of malignancy for age <18 years, age >65 years, or TSH >2.5 mU/l. The summary OR was in a descending order (Table 1): 10.15 (95% CI=6.72–15.33) for nodule height greater than width, absence of the halo sign, microcalcifications, irregular margins, hypoechogenicity, solid nodule structure, intranodular vascularization, nodule size ≥4 cm, and single nodule as US TN features (Fig. 2).

The funnel plot asymmetry (Fig. 3) and the Egger's test for any other clinical or US features, the risk of malignancy was similar regardless of the year of publication, study location, and method used for diagnosis.

**Sensitivity and meta-regression analyses**

Sensitivity analysis has shown the stability of the overall ORs with the withdrawal of any of the study from the analysis without a significant improvement of the heterogeneity.

Meta-regression analyses showed a higher risk of malignancy for hypoechoic nodules in iodine-sufficient areas than in iodine-deficient areas (P=0.01) with a summary OR of 7.14 (95% CI=4.77–10.68) in iodine-sufficient areas and 3.11 (95% CI=2.13–4.52) in iodine-deficient areas. However, this difference did not explain the observed heterogeneity (residual $I^2=58\%$). There was no difference in the risk of malignancy for hypoechoic nodules, with respect to year of publication and method used for diagnosis.

In the meta-regression analyses for all the other clinical and US features, the risk of malignancy was similar regardless of the year of publication, study location, and method used for diagnosis.

**Publication bias**

The funnel plot asymmetry (Fig. 3) and the Egger's test result (P=0.01) for intranodular vascularization suggest the presence of publication bias that may distort the meta-analysis. The Duval & Tweedie non-parametric ‘trim and fill’ method was employed and generally resulted in similar conclusions of the unadjusted random-effect model of 3.76 (95% CI=2.04–6.95); we calculated a summary adjusted OR of 3.31 (95% CI=1.81–6.06). Publication bias was not evident from the reviews of the funnel plot or Egger’s test for any other clinical or US feature.
Discussion

The current meta-analysis provided a comprehensive and systematic evaluation of the clinical and US features of TNs associated with an increased risk of thyroid cancer as listed in Table 1.

Clinical TN features

Family history of thyroid carcinoma, prior head or neck irradiation, and male gender are the clinical features that were associated with a higher risk of thyroid carcinoma in our analysis.

Not surprisingly, among the parameters related to clinical features, the highest OR was observed for family history of thyroid carcinoma. In fact, familial medullary thyroid carcinoma, multiple endocrine neoplasia type 2, familial papillary thyroid tumors, familial polyposis coli, Cowden disease, and Gardner syndrome are all familial syndromes associated with a higher incidence of thyroid carcinomas (7). Medullary thyroid carcinoma is the type of thyroid cancer most often present in the context of a familial syndrome, but the OR computed in our meta-analysis referred primarily to papillary and follicular thyroid carcinomas because they accounted for more than 90% of the thyroid cancer cases.

A history of head or neck irradiation during prior radiation therapy or exposure to atmospheric emissions of $^{131}$I following nuclear accidents was also associated with a higher risk of thyroid carcinoma. The OR computed in our meta-analysis shows the likelihood of TN malignancy in patients with a history of head or neck irradiation without thyroid carcinoma, multiple endocrine neoplasia type 2, familial papillary thyroid tumors, familial polyposis coli, Cowden disease, and Gardner syndrome are all familial syndromes associated with a higher incidence of thyroid carcinomas (7). Medullary thyroid carcinoma is the type of thyroid cancer most often present in the context of a familial syndrome, but the OR computed in our meta-analysis referred primarily to papillary and follicular thyroid carcinomas because they accounted for more than 90% of the thyroid cancer cases.

Figure 1

Forest plot for the meta-analysis of studies reporting on the association with the risk of thyroid cancer of (a) family history of thyroid carcinoma, (b) prior head or neck irradiation, (c) male gender, (d) age < 18 years, (e) age > 65 years, and (f) TSH > 2.5 mU/l. The overall estimate of the effect is represented by a diamond in each plot.
We must note that some studies have shown a significant linear dose–response relationship over a relatively broad range of doses and a higher risk after exposure in childhood (55).

Although some studies have reported a higher risk of thyroid carcinoma for TNs occurring in young people and older people with a bimodal distribution (31, 45, 56, 57), our analysis did not show these findings. However, limitations are present in our meta-analysis for age because it is a continuous variable, and most studies presented non-uniform, aggregate data on age.

**Figure 2**

Forest plot for the meta-analysis of studies reporting on the association with the risk of thyroid cancer of (a) nodule height greater than width, (b) absent halo sign, (c) microcalcifications, (d) irregular margins, (e) hypoechogenicity, (f) solid nodule structure, (g) intranodular vascularization, (h) nodule size \( \geq 4 \text{ cm} \), and (i) single nodule. The overall estimate of the effect is represented by a diamond in each plot.

**Ultrasonographic TN features**

We are progressively increasing our knowledge of benign and malignant TN US features. Although different terminology, qualities of US instruments, and experience of radiologists can produce variable results, thyroid US is currently the most accurate imaging modality for detecting TNs (58). In our meta-analysis, the highest risk of malignancy for US characteristics was seen for nodule height greater than width, absent halo sign, microcalcifications and irregular margins, and the lowest for nodule size \( \geq 4 \text{ cm} \) and single nodule.
Limitations are present in our meta-analysis for nodule size because it is a continuous variable and the included studies presented only non-uniform aggregated data on nodule size.

In iodine-sufficient areas, hypoechogenicity was also strongly associated with a higher risk of malignancy, but in iodine-deficient areas, this association was weaker. We could not find any consistent explanation for these findings in the literature.

Moreover, solid nodule structure was also associated with a higher risk of thyroid cancer. Intranodular vascularization was associated with a higher risk of malignancy when the unadjusted random-effect model was used, and this was confirmed when the ‘trim and fill’ method was used to adjust the meta-analysis for publication bias (Table 1).

Clinical perspectives of the meta-analysis results

The current guidelines recommend to biopsy TN >5 mm in diameter in presence of suspicious clinical and/or US features (6, 7), listing them without providing a singularly defined risk of malignant potential.

The present meta-analysis yields an overview of the TN features known to be linked to malignancy, quantifying their importance in the assessment of the TN cancer risk.

To our knowledge, it represents the first effort to provide a complete TN cancer risk assessment not based on a single sample of patients, but on the review of all published studies from 1989 to 2012, which matched our stringent inclusion criteria.

As a limitation of the meta-analysis, high heterogeneity among the studies was evident and we explored the year of publication, study location, and method used for diagnosis as potential sources of heterogeneity; none of these factors emerged as significant contributors to combined risk estimates. Nevertheless, it is possible that

<table>
<thead>
<tr>
<th>Clinical features</th>
<th>OR (95% CI)</th>
<th>P value</th>
<th>Heterogeneity</th>
<th>Pub bias (Egger test)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Q test</td>
<td>I² (%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical features</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family history of thyroid carcinoma</td>
<td>2.29 (1.45–3.64)</td>
<td>&lt;0.001</td>
<td>0.30</td>
<td>18%</td>
</tr>
<tr>
<td>Prior head or neck irradiation</td>
<td>1.29 (1.02–1.64)</td>
<td>0.03</td>
<td>0.36</td>
<td>9%</td>
</tr>
<tr>
<td>Gender (male)</td>
<td>1.22 (1.01–1.47)</td>
<td>0.04</td>
<td>0.01</td>
<td>68%</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;18</td>
<td>1.33 (0.70–2.50)</td>
<td>0.37</td>
<td>0.25</td>
<td>28%</td>
</tr>
<tr>
<td>&gt;65</td>
<td>1.15 (0.70–1.89)</td>
<td>0.58</td>
<td>0.06</td>
<td>49%</td>
</tr>
<tr>
<td>TSH &gt;2.5 mU/l</td>
<td>1.05 (0.45–2.45)</td>
<td>0.90</td>
<td>0.01</td>
<td>73%</td>
</tr>
<tr>
<td>Ultrasonographic features</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nodule height greater than width</td>
<td>10.15 (6.72–15.33)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>86%</td>
</tr>
<tr>
<td>Absent halo sign</td>
<td>7.14 (3.71–13.71)</td>
<td>0.004</td>
<td>0.27</td>
<td>24%</td>
</tr>
<tr>
<td>Microcalcifications</td>
<td>6.76 (4.72–9.69)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>89%</td>
</tr>
<tr>
<td>Irregular margins</td>
<td>6.12 (3.12–12.02)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>95%</td>
</tr>
<tr>
<td>Hypoechogeticity</td>
<td>5.07 (3.47–7.43)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>74%</td>
</tr>
<tr>
<td>Solid nodule structure</td>
<td>4.69 (2.63–8.36)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>86%</td>
</tr>
<tr>
<td>Intranodular vascularization</td>
<td>3.76 (2.04–6.95)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>92%</td>
</tr>
<tr>
<td>‘trim and fill’ adjusted</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nodule size ≥4 cm</td>
<td>3.31 (1.81–6.06)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>86%</td>
</tr>
<tr>
<td>Single nodule</td>
<td>1.63 (1.04–2.55)</td>
<td>0.03</td>
<td>0.03</td>
<td>61%</td>
</tr>
<tr>
<td></td>
<td>1.43 (1.09–1.88)</td>
<td>0.01</td>
<td>0.006</td>
<td>70%</td>
</tr>
</tbody>
</table>

Age <18 and > 65 years and TSH levels >2.5 mU/l did not show a significantly increased risk.

Figure 3

Funnel plot assessing a negative influence of publication bias only for studies reporting on the association between intranodular vascularization and the risk of thyroid cancer. Open circles represent individual studies, while Open diamond represents the overall effect estimated.
other factors caused by high inter-observer variations in TN US assessment (55), which have effects that cannot be evaluated in meta-analysis, may have significantly contributed to the heterogeneity.

Another limitation of this meta-analysis is represented by having conducted a univariable analysis for each different characteristic. However, a multivariable analysis that would consider them all at once is not feasible because for almost all published studies only aggregated data are available.

The OR values derived from the meta-analysis aid to quantify the risk of thyroid cancer in the TN assessment, especially when other examinations such as elastography and thyroid molecular markers are not available and in case of preoperative not conclusive cytologic results.

In fact, in most cases FNA clearly distinguishes benign from malignant lesions, but the largest series show that about 5% of FNA results are false positive or false negative and that 20–25% of FNA are non-diagnostic or indeterminate (59, 60).

In order to improve management of such situations and to ameliorate TN diagnostic workup, it would be possible and desirable to create a meta-analysis-derived grading system of TN malignancy risk, which would need to validated in a large prospective cohort.

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

Funding
This review did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.

References
Review


Received 6 December 2013
Revised version received 12 February 2014
Accepted 17 February 2014