Ultrasound Sensitivity for Thyroid Malignancy Is Increased by Real-Time Elastography: A Prospective Multicenter Study

Pierpaolo Trimboli, Rinaldo Guglielmi, Salvatore Monti, Irene Misischi, Filomena Graziano, Naim Nasrollah, Stefano Amendola, Sara Nazzarena Morgante, Maria Grazia Deiana, Stefano Valabrega, Vincenzo Toscano, and Enrico Papini

Sections of Endocrinology and Diabetology (P.T., S.A.) and Surgery (N.N.), Ospedale Israelitico, 00148 Rome, Italy; Department of Endocrinology (R.G., I.M., F.G., E.P.), Ospedale Regina Apostolorum of Albano Laziale, 00041 Rome, Italy; and Departments of Endocrinology (S.M., S.N.M., M.G.D., V.T.) and Surgical Science (S.V.), Ospedale S. Andrea, Sapienza University, 00189 Rome, Italy

Context: Thyroid nodules are selected for biopsy on the basis of clinical and ultrasound (US) findings. Ultrasonography detects nodules at risk of malignancy, but its diagnostic accuracy does not rule out with certainty the possibility of cancer in lesions without suspicious findings.

Objective: The objective of the study was to evaluate the diagnostic accuracy of real-time elastography (RTE) in thyroid nodules and to assess the improvement provided by combination of RTE, B-mode US, and color flow Doppler (CFD).

Design: This was a prospective multicenter study.

Patients: A consecutive series of 498 thyroid nodules was blindly evaluated by US, CFD, and RTE before biopsy or surgery. Nodules were classified at RTE by four-class color scale. Patients with benign cytology underwent follow-up over 12 months, whereas patients with indeterminate, suspicious, or malignant cytology were surgically treated.

Results: At follow-up, 126 nodules were malignant and 372 benign. RTE classes III-IV showed 81% sensitivity and 62% specificity. The presence of at least one US risk factor (hypoechogenicity, microcalcifications, irregular margins, intranodular vascularity, and taller than wide shape) had 85% sensitivity and 91% negative predictive value. When RTE was combined with US, the presence of at least one of the six parameters had 97% sensitivity and 97% negative predictive value, with an odds ratio of 15.8 (95% confidence interval 5.7–43.8).

Conclusions: RTE is a valuable tool for detecting malignant thyroid lesions with a sensitivity similar to traditional US and CFD features. By adding RTE evaluation, the sensitivity for malignancy of US findings is markedly increased and the selection of nodules that do not need cytology is made more reliable. (J Clin Endocrinol Metab 97: 4524–4530, 2012)
atic nodule undergoes surgery or clinical follow-up mainly on the basis of the cytological outcome. Because of the high prevalence of nodular thyroid disease, it is not feasible or cost effective to perform a cytological assessment for all or even most thyroid nodules. US examination represents the main tool in risk stratification of palpable and nonpalpable lesions and their selection for FNA (5, 6). Several papers have reported that a few B-mode US and color flow Doppler (CFD) features (i.e., hypoechoigenicity, irregular margins, microcalcifications, more tall than wide shape, intralesional vascular signals) are suggestive for malignancy (7, 8). Thyroid nodules with suspicious US findings need a cytological assessment, but the diagnostic value of these features is hampered by their low sensitivity (9–12). Hence, devices improving US diagnostic accuracy are required to decrease the risk of overlooking malignant nodules not submitted to FNA because of the lack of suspicious US findings. In this context, elastography, first described on thyroid lesions in 2005 (13), has been reported as a tool that may potentially improve the management of thyroid lesions.

Elastography consists of real-time evaluation [real time elastography (RTE)] of an area of interest during the application of an external force by the echographic probe. A software tracks the tissue displacement that is displayed by a color scale in accordance with nodular and extranodular hardness. Thyroid cancer usually appears as a hard and firm lesion and several papers reported a high specificity and sensitivity of RTE in predicting thyroid cancer (14–22). Yet no prospective multicenter study has been carried out and the diagnostic accuracy of this imaging technique has been reported as quite variable. The inconsistent results are probably due to the presence of some limits in most of these studies: small series of patients, previous selection, retrospective design, absence of blind examination, single examiners, single-center trials, and lack of standardization of the procedure.

The aims of this study were to prospectively assess the diagnostic accuracy of RTE in detecting thyroid cancer in a consecutive series of patients and the potential clinical improvement provided by the combined use of RTE with B-mode US and CFD examination.

**Materials and Methods**

In the period from October 1, 2010, to May 31, 2011, 602 thyroid nodules from 576 patients (438 females, 138 males, mean age 53 ± 12.8 yr) were evaluated by conventional US, CFD, and RTE at three thyroid referral centers in the greater Rome area, Ospedale Regina Apostolorum, Ospedale Israelitico, and Ospedale S. Andrea. Criteria of exclusion were the following: 1) presence of macrocalcifications (defined as hyperechoic spots >2 mm); 2) size smaller than 5 mm or larger than 30 mm; 3) anechoic or complex appearance at US; 4) multiple coalescent nodules; 5) Hashimoto’s thyroiditis; and 6) hot feature at thyroid radioisotope scan. All nodules had no previous cytology. Of the 576 patients, 532 underwent FNA, whereas the remaining 44 were addressed to our surgical departments from primary centers because of symptomatic multinodular goiter with suspicious nodules. All patients gave their informed consent to participate to the study.

According to the 2007 British Thyroid Association (23) and the 2010 American Association of Clinical Endocrinologists/Associazione Medici Endocrinologi/European Thyroid Association (5) guidelines, cytological samples were classified into five classes: Thy/class 1 (nondiagnostic), Thy/class 2 (benign), Thy/class 3 (indeterminate/follicular lesion), Thy/class 4 (suspicious for malignancy), and Thy/class 5 (malignant). Cytological evaluation was made by an expert cytopathologist at the Ospedale Regina Apostolorum and at the Ospedale Israelitico and by another experienced cytopathologist at the Ospedale S. Andrea. Cytology was combined with immunocytochemistry, when appropriate. Nodules reported as benign in the study had benign cytology (Thy/class 2) with clinical and US follow-up over a 12-month period. All cytologically benign nodules that showed a growth of their greater diameter greater than 20% or a suspicious change of their US features at follow-up were submitted to a second FNA for confirmation.

Thyroid US evaluation was performed using a Hitachi Logos High Vision E system equipped with a 6- to 14-MHz linear transducer (Hitachi Ltd., Tokyo, Japan). Ultrasound risk stratification, according to the Society of Radiologists in Ultrasound (10) and American Association of Clinical Endocrinologists/Associazione Medici Endocrinologi/European Thyroid Association criteria (5), was made by echostructure and echogenicity (solid hypoechoic, solid isoechoic, solid hyperechoic, mixed solid/cystic, spongiform, or purely cystic), the nodule’s margins (hypoechoic halo, regular or irregular), and the presence of microcalcifications (defined as hyperechoic spots <2 mm and with no posterior shadowing). The pattern of nodular vascular signal was evaluated by CFD and defined as follows: CFD-1 as absent signal, CFD-2 as perinodular signal, and CFD-3 as intranodular signal (9).

RTE was performed according to the technique previously described by Tranquart et al. (16). The RTE score was assessed based on a color scale, the blue color being correlated with hard tissue, red color with soft tissue, and green with intermediate hardness. Nodules were classified into four classes: class I, prevalence of red and green color; class II, green with prevalence in more than 50% of the nodule; class III, blue in at least 50% of nodule; class IV, blue with prevalence of at least 75% of nodule. Figure 1 illustrates the RTE assessment used herein.

In all cases US, CFD, and RTE examinations were performed before FNA or surgery by three experienced endocrinologists (E.P., P.T., S.M.). Before starting the trial, the US operators practiced together to standardize their technique of examination. The interobserver variability of RTE was assessed on 20 consecutive thyroid nodules. Thyroid lesions were analyzed by elastography, and the images were stored and separately evaluated by the three operators. Finally, the scores from the examiners were compared for each nodule.

Serum TSH, free T4, antibodies for thyroperoxidase, and calcitonin were measured in all patients.
Statistical analysis

The US, CFD, and RTE characteristics were compared by a χ² test or a Fisher exact test, when appropriate. A predictivity test was obtained according to Galen and Gambino (24). Odds ratio and 95% confidence interval was used to evaluate the association with malignancy of several parameters. The statistical significance was set at P < 0.05.

Results

Of the initial series of 602 nodules, 104 (17.2%) were excluded from the study because of inadequate (Thy/class 1) cytology, diagnosis of indeterminate (Thy/class 3) cytology with refusal of surgery, or because of loss at follow-up. Then the final series of the study comprised 498 thyroid nodules from 446 patients (358 females and 88 males, mean age 53.4 ± 12.9 yr). Four hundred forty-two nodules from 397 subjects had a cytological report reading of benignancy, malignancy, or suspicion for malignancy. Five cases, even if with inadequate cytological sample, were included in the study because they were addressed to surgery due to their suspicious clinical parameters. As mentioned above, 51 nodules from 44 patients did not undergo cytology before thyroidectomy.

Histological and clinical follow-up

After follow-up, 126 malignant (25%) and 372 benign (75%) nodules were recorded. Of the cancers, 117 were papillary carcinomas (PTC) (101 classic, 14 follicular, two oncocytic variant) of which 46 microcarcinomas, four minimally invasive follicular carcinomas (FTCs), three medullary carcinomas, one mixed PTC-FTC, and one Hürthle cell cancer. Malignancy was confirmed in all 44 nodules cytologically classified as Thy/class 5, whereas 39 of 42 nodules (92.8%) assessed as Thy/class 4 were malignant with three negative cases. Sixteen of 42 nodules with indeterminate cytological report (Thy/class 3) (38.1%) were malignant (14 PTCs, one FTC, one medullary carcinoma) and 26 benign. Also, four of 314 nodules with Thy/class 2 (1.3%) and three of 5 (60%) nodules with Thy/class 1 operated upon because they were clinically suspicious were cancers at histology. In the series of 44 patients addressed to surgery without preoperative cytological assessment, 18 of 51 nodules (35.3%) were cancers after surgery. Finally, during follow-up of nodules with initial benign (Thy/class 2) cytology, no significant clinical or sonographic change was registered in 302 of 314 cases (96.2%). According to the study design, US-guided FNA was repeated in 12 lesions (3.8%) with suspicious modi-

Study design and main results

![FIG. 1. Schematic representation of the herein used RTE assessment based on a color scale. Nodules were classified into four classes. The RTE classes are detailed in Material and Methods (modified from Ref. 16).](https://academic.oup.com/jcem/article-abstract/97/12/4524/2536507/fig-1)

![FIG. 2. Flow chart of the study.](https://academic.oup.com/jcem/article-abstract/97/12/4524/2536507/fig-2)
Real time elastography

By RTE, 82 nodules (16.5%) were classified as class IV, 161 (32.3%) as class III, 222 (44.6%) as class II, and 33 (6.6%) as class I. Of the 126 cancers, 60 (47.6%) had class IV, 42 (33.3%) class III, 23 (18.2%) class II, and one (0.8%) class I. Table 1 shows the prevalence of cancer in the four RTE classes. The RTE class IV was significantly \( P < 0.0001 \) associated with a malignant outcome with a 14.4 odds ratio [OR; 95% confidence interval (CI) ranging from 8.3 to 25.1]. Also, the combined IV and III RTE scores were significantly \( P < 0.0001 \) more frequent in malignant (102 cases, 81%) than in benign (141 cases, 37.9%) nodules, with an OR of 6.9 (95% CI 4.2–11.3) (Table 2).

Regarding the predictivity test, RTE IV showed high accuracy in predicting a nodule’s malignancy. When classes IV and III were analyzed as a whole, they showed a high sensitivity and negative predictive value (NPV) (Table 2).

Conventional US and CFD

Five established US risk factors (hypoechogenicity, microcalcifications, irregular margins, more tall than wide shape, and intranodular vascular signals) were analyzed as predictors of nodule’s malignancy. When the presence of at least one of these five US-CFD risk factors was analyzed as a predictor test, US was significantly \( P < 0.0001 \) associated with malignancy (107 cases, 85%) with an OR of 6.6 OR (95% CI 3.9–11.2) (Table 2). These parameters showed similar predictivity with respect to RTE III-IV (Table 2).

Combined conventional US-CFD and RTE

When US and CFD were combined with RTE findings for the assessment of the risk of malignancy, the presence of at least one of these six risk features (hypoechogenicity, microcalcifications, irregular margins, more tall than wide shape, intranodular vascularization, and RTE class III-IV) was significantly \( P < 0.0001 \) associated with cancer with a 15.8 OR (95% CI 5.7–43.8). Moreover, sensitivity and NPV were increased by the use of a combination of US and RTE findings (Table 2). In particular, 220 thyroid nodules showed no conventional US or CFD risk factor. Of these, 89 nodules (40.4%) showed RTE III-IV scores and 15 of 89 (16.8%) were cancers at histological examination. Thus, RTE examination made it possible to identify as malignant 6.8% of the nodules that did not show any B-mode suspicious finding. On the other hand, 255 nodules showed RTE I-II scores, and 124 of these had at least one B-mode US suspicious finding, and 20 of 124 (16.1%) were cancers. Thus, B-mode and CFD examination made it possible to identify as malignant 7.8% of the nodules that did not show RTE suspicious scores. Of the 131 nodules with no US or RTE risk features, four were malignant (3.2% of the 126 cancers) (Fig. 3). Thus, in a series of lesions with a 25% prevalence of malignancy as a whole, the risk of thyroid cancer was decreased to 3.0% for nodules with a combined negative examination.

Interobserver variability

The percentage of agreement between the operators was 85% for class I and IV and 60% for class II and III of RTE. When RTE results were considered as a dichotomous variable (merging class I with class II and class III with class IV), the agreement rate between the investigators was 80%. Hence, we used RTE as a dichotomous variable because of the high percentage of agreement, similar to that reported for the established US risk factors for thyroid malignancy (25).

Discussion

Thyroid nodules are common and carry a low but not negligible risk of malignancy. The challenge of clinical management is to identify benign nodules and to accurately diagnose and treat malignant thyroid disease early in a large number of patients. The current treatment plan mainly relies on the sonographic appearance of the nodule (9, 11, 12, 26, 27). Hence, US tools with a high sensitivity for malignancy are needed to minimize the risk of missing malignant nodules not submitted to FNA because of the lack of suspicious US findings.

The use of RTE for improving the diagnostic accuracy of sonographic examination of the thyroid gland was first described by Lyshchik et al. (13). Fifty-two nodules were evaluated by a US scanner modified for research, and images were exported and used for off-line strain image re-

---

**Table 1.** Prevalence of malignancy in thyroid nodules according to the elastographic score (I-IV)

<table>
<thead>
<tr>
<th>RTE score</th>
<th>Prevalence of cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I</td>
<td>1/33 (3.0%)</td>
</tr>
<tr>
<td>Class II</td>
<td>23/222 (10.4%)</td>
</tr>
<tr>
<td>Class III</td>
<td>42/161 (26.1%)</td>
</tr>
<tr>
<td>Class IV</td>
<td>60/82 (73.2%)</td>
</tr>
</tbody>
</table>
construction. A strain index greater than 4 was the strongest independent factor in predicting malignancy of nodules, with 82% sensitivity and 96% specificity. Afterward, a retrospective postsurgery study classified 92 nodules into five classes by RTE, and the classes 4 and 5 as a whole were an independent predictor of malignancy with a very high sensitivity (97%) and specificity (100%) (14). Following this finding, 81% sensitivity and 94% specificity were recorded in a prospective series of 86 nodules. Therefore, to analyze the receiver-operating characteristic curve, a presumptive malignant or benign diagnosis was assigned to RTE classes 3 and 4 and 1 and 2, respectively (15). Eight studies (13–20), comprising the above-mentioned three, were included in a meta-analysis by Bojunga et al. (28). The results showed that RTE approached to 88–96% sensitivity and 85–95% specificity in detecting thyroid cancer. This wide interobserver variability was analyzed by Park et al. (29) for both US and RTE; fifty-two thyroid lesions scheduled to surgery because of malignant cytology were included, and significant concordance between examiners was found for all features but echogenicity and elastography (29). In all, papers analyzing the qualitative use of RTE on the basis of colors scale showed controversial results and reported a wide range of diagnostic accuracy (13–22). The main methodological limits of most of these trials were the small series of patients usually enrolled in only one center, their previous selection among candidates for surgery, the retrospective study design, and the absence of blinded examination. Thus, the actual role of RTE for the selection of thyroid nodules to biopsy is still not well clarified in clinical practice (30).

To date, RTE has been analyzed as a single factor in comparison to B-mode and color Doppler US. In the present prospective series, RTE and conventional US-CFD features of 126 cancers and 372 benign nodules were blindly evaluated as predictors of malignancy. Microcalcifications, irregular margins, RTE class IV, and tall shape of the nodule, when considered as distinct risk factors, confirmed a strong association with thyroid cancer with a high OR. The RTE analysis demonstrated that a relevant stiffness (scores IV and III) was present in less than half of the nodules under evaluation and that this subgroup of thyroid nodules harbored the vast majority of thyroid cancers. When considered as a whole, the combined RTE III and IV scores had an OR of 6.9 with a high NPV (91%) and sensitivity (81%). Their sensitivity was higher but their specificities was lower than the conventional US features (Table 2). The specificity of RTE recorded herein was in agreement with previous papers, but its sensitivity appeared lower than in prior reports.

The present study showed that the predictivity of malignancy of conventional US and CFD is improved by adding elastography. In particular, using one of the five US-CFD features, it was possible to correctly identify 85% of the malignant lesions (OR 6.6), whereas the presence of at least one of six US, CFD and RTE features was able to detect 97% of thyroid cancers (OR 15.8). On the basis of these data, the finding of a RTE III-IV score should be considered as predictive for malignancy as the conventional US-CFD parameters. It is remarkable that 15 cancers (12%) lacked the conventional B-mode US risk factors but were classified as III-IV score at RTE, whereas 20 cancers (15.8%) were benign at RTE and had at least one

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Se</th>
<th>Sp</th>
<th>PPV</th>
<th>NPV</th>
<th>Acc</th>
<th>P</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypoechogenicity</td>
<td>72</td>
<td>60</td>
<td>38</td>
<td>87</td>
<td>63</td>
<td>&lt;0.0001</td>
<td>3.9 (2.5–6.2)</td>
</tr>
<tr>
<td>Microcalcifications</td>
<td>31</td>
<td>98</td>
<td>87</td>
<td>81</td>
<td>81</td>
<td>&lt;0.0001</td>
<td>27.3 (11.2–66.6)</td>
</tr>
<tr>
<td>Irregular margins</td>
<td>25</td>
<td>99</td>
<td>86</td>
<td>80</td>
<td>80</td>
<td>&lt;0.0001</td>
<td>24.9 (9.4–65.8)</td>
</tr>
<tr>
<td>Intranodular vascularization</td>
<td>37</td>
<td>88</td>
<td>52</td>
<td>81</td>
<td>75</td>
<td>&lt;0.0001</td>
<td>4.4 (2.7–7.1)</td>
</tr>
<tr>
<td>More tall than wide</td>
<td>14</td>
<td>99</td>
<td>82</td>
<td>77</td>
<td>78</td>
<td>&lt;0.0001</td>
<td>15.3 (5.1–46.2)</td>
</tr>
<tr>
<td>At least one of five US-CFD</td>
<td>85</td>
<td>54</td>
<td>38</td>
<td>91</td>
<td>62</td>
<td>&lt;0.0001</td>
<td>6.6 (3.9–11.2)</td>
</tr>
<tr>
<td>RTE III or IV</td>
<td>81</td>
<td>62</td>
<td>42</td>
<td>91</td>
<td>67</td>
<td>&lt;0.0001</td>
<td>6.9 (4.2–11.3)</td>
</tr>
<tr>
<td>At least one of six US-CFD-RTE</td>
<td>97</td>
<td>34</td>
<td>33</td>
<td>97</td>
<td>50</td>
<td>&lt;0.0001</td>
<td>15.8 (5.7–43.8)</td>
</tr>
<tr>
<td>Hypoechogenicity feature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microcalcifications feature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irregular margins feature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intranodular vascularization feature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>More tall than wide feature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At least one of five US-CFD feature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RTE III or IV feature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At least one of six US-CFD-RTE feature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FIG. 3. US, RTE, and CFD risk factors in 126 histologically proved thyroid cancers.

At least one of five US-CFD means that the presence of at least one of the five US-CFD parameters (hypoechogeticity, microcalcifications, irregular margins, intranodular vascularization, more tall than wide shape) was considered as a single test in predicting malignancy. At least one of six US-CFD-RTE means that one of the six US-CFD-RTE features (hypoechogeticity, microcalcifications, irregular margins, intranodular vascularization, more tall than wide shape, RTE class III-IV) was analyzed as a single test. Se, Sensitivity; Sp, specificity, PPV, positive predictive value, Acc, accuracy.
US risk feature (Fig. 3). Moreover, in the present series of lesions characterized by a 25% prevalence of malignancy, the risk of thyroid cancer was decreased to 3.0% for nodules with a combined negative examination. Hence, the combination of US-CFD-RTE risk features can be used for the selection of nodules to be biopsied because it effectively reduces the risk of overlooking a malignant nodule with otherwise benign B-mode or RTE sonographic findings.

For practical purposes, on the basis of our findings, the presence of a relevant region of stiffness, with a blue color revealed in at least 50% of the nodule area as in scores IV and III of the classification by Tranquart et al. (16), may be considered as a suspicious RTE pattern with no need of further elastographic categorization (13–22).

A partial limit of the present study is the high prevalence of malignant nodules in our series of cases. This is due to the enrollment of patients referred by primary centers for undergoing surgery and to the entry criteria (exclusion of cystic or complex lesions, calcified nodules, multiple coalescent nodules). Nodules containing fluid collections, indeed, cannot be reliably evaluated by RTE because their anechoic areas show a red appearance at elastography. Macrocalcifications (either coarse or shell calcifications) have a hard blue appearance at RTE. Thus, cystic or calcified lesions and nodular goiters with large or merging nodules were excluded from the study to avoid possible bias due to the technical limits of RTE technique. Hence, in clinical practice, when evaluating an unselected series of nodules with a lower (~5%) prevalence of malignant lesions, the cost-effectiveness of RTE should be probably less than in our study.

In conclusion, the present study confirms that elastographic evaluation of thyroid nodules is an effective tool for detecting malignant lesions. The presence of hardness in at least 50% of the nodule’s area has a sensitivity that is similar to the conventional US and CFD findings. The sensitivity of B-mode US findings traditionally used in risk stratification of thyroid lesions is increased by adding RTE evaluation. Thus, the selection of nodules that do not need cytology is made significantly more reliable by combining RTE with B-mode US and CFD evaluation.

Acknowledgments

Address all correspondence and requests for reprints to: Dr. Enrico Papini, Ospedale Regina Apostolorum, Via San Francesco 50, Albano Laziale, 00041 Rome, Italy. E-mail: enrico.papini@fastwebnet.it.

Disclosure Summary: The authors have nothing to disclose.

References


22. Unlüturk U, Erdogan MF, Demir O, Gülüü S, Başkal N 9 August 2012 Ultrasound-Elastography is not superior to gray-scale ultrasound in predicting malignancy in thyroid nodules. Thyroid 10.1089/thy.2011-0502


