Utility and Interobserver Agreement of Ultrasound Elastography in the Detection of Malignant Thyroid Nodules in Clinical Care

S. Merino, J. Arrazola, A. Cárdenas, M. Mendoza, P. De Miguel, C. Fernández and T. Ganado

AJNR Am J Neuroradiol 2011, 32 (11) 2142-2148
doi: https://doi.org/10.3174/ajnr.A2716
http://www.ajnr.org/content/32/11/2142

This information is current as of August 31, 2023.
Utility and Interobserver Agreement of Ultrasound Elastography in the Detection of Malignant Thyroid Nodules in Clinical Care

BACKGROUND AND PURPOSE: Malignancy correlates with hardness of tissues and US elastography can potentially analyze the stiffness of lesions. Our aim was to evaluate the utility of US elastography in the detection of malignant nodules and to investigate interobserver agreement with this technique.

MATERIALS AND METHODS: One-hundred three consecutive patients with 106 thyroid nodules were examined prospectively with conventional B-mode sonography and real-time US elastography. All patients were referred for FNAB. Conventional B-mode sonography and US elastographic examinations were performed, and images were separated and independently interpreted by 2 radiologists blinded to pathologic results. US elastogram evaluation was based on a simplified classification of stiffness based on gray-scale patterns, tumor size compared with B-mode, and margins. Interobserver agreement was studied. FNAB was used as the reference standard for the diagnosis of benign nodules, but histopathologic evaluations were performed when results suspicious for malignancy or malignant results were obtained on FNAB as well as in indeterminate lesions.

RESULTS: In our study, pattern of stiffness based on gray-scale and classification proposed were statistically significant and predicted malignancy with 100% sensitivity and 40.6% specificity. Tumor size when compared with B-mode images or margins was not statistically significant in our study. No false-negatives were found, and an NPV of 100% was seen. Interobserver agreement for US elastography was excellent in our study, with a k index of 0.82 (95% CI).

CONCLUSIONS: We believe that US elastography is a promising technique that can assist in the evaluation of thyroid nodules and can potentially diminish the number of FNAB procedures needed. We believe that it may be useful to introduce US elastography into routine clinical practice.

ABBREVIATIONS: CI = confidence interval; FNAB = fine-needle aspiration biopsy; NPV = negative predictive value; PPV = positive predictive value; OF = quality factor; US = ultrasound

Thyroid nodules are found in 4%–8% of adults by palpation, 41% by sonography, and 50% by pathologic postmortem examination. Although in the general population, most thyroid nodules are benign, the prevalence of thyroid cancer is as high as 5%–10%. Papillary carcinoma represents 75%–80% of the cases, and other histologic types include follicular carcinoma (10%–20%), medullar carcinoma (3%–5%), and anaplastic carcinoma (1%–2%). Lymphoma and metastases of the thyroid gland are less frequent.

The management of thyroid nodules is of clinical importance due to the relatively high incidence of malignant disease and the relative difficulty in distinguishing benign from malignant nodules by conventional sonographic imaging. Despite being extremely accurate, conventional sonography is not as useful for differentiating benign and malignant nodules. Conventional sonographic signs that correlate with malignant lesions include the following: hypoechogenicity, blurred or irregular margins, microcalcifications, an anteroposterior/transverse diameter of ≥1 cm, and intranodular predominantly centric vascularity. However the sensitivity, specificity, NPV, and PPV are considerably variable from study to study. FNAB provides the best method to differentiate benign from malignant nodules. In most of the proposed guidelines, current management of thyroid nodules includes the FNAB procedure in all nodules with a diameter >1 cm and in those with lesser diameters but with different classic sonographic signs suspicious for malignancy. US-guided FNAB has proved to be an efficient tool for thyroid cancer diagnosis, especially in the case of papillary carcinoma. In follicular carcinoma, histopathology is required because the diagnosis is based on the presence or absence of capsular or vascular invasion. However, US-guided FNAB is an invasive procedure, subject to sampling errors, and consumes time and economic resources, and it is associated with minor complications.

The high incidence of nodules, the invasiveness of the procedure, and the possibility of sampling error require performing protocols for determining with relative confidence which nodules should be followed up and which should be aspirated.

Malignant nodules tend to be much harder than benign ones. However, the stiffness of thyroid nodules in palpation is a nonobjective indicator for the differential diagnosis. US elastography has been developed to obtain information about tissue stiffness noninvasively. This technique can evaluate the degree of distortion of a tissue under application of an external force and is based on the principle that the softer parts of...
tissues deform easier than the harder ones.\textsuperscript{17} US elastography has been successfully applied to breast lesions, the prostate, pancreas, lymph nodes, and thyroid gland, all easily assessable on elastographic examination.\textsuperscript{18–21}

The aim of this prospective study was to find out possible elastographic signs suspicious for malignancy in thyroid nodules, to evaluate interobserver agreement for US elastography, and to explore the sensitivity and specificity of US elastography for the differential diagnosis of thyroid cancer, with pathologic analysis as the reference standard.

**Materials and Methods**

**Research Design**

The study design was based in a series of consecutive thyroid nodules evaluated with conventional B-mode sonography and US elastography between October 2008 and November 2009, with prospective capture of information.

**Patients**

One-hundred six consecutive patients who were referred for FNAB of thyroid nodules were initially included in this prospective study. In 3 of these patients, it was not possible to obtain pathologic results, so they were excluded from the final statistical study. The sample at FNAB was insufficient in 1 patient who left our country and could not be followed up. The cytologic result in another patient was atypia of uncertain significance in subsequent explorations, and the patient declined new FNAB or surgery. The third patient also declined surgery even with FNAB results suspicious for follicular neoplasia.

Finally, 103 patients (89 women and 14 men) were included. The mean age was 58 years (range, 21–87 years). In each patient, 1 thyroid nodule was studied, except in 3 patients in whom 2 nodules were evaluated, so 106 nodules were definitively included in the study (Fig 1).

Before enrollment, each patient gave written informed consent.

**Equipment**

Conventional B-mode sonography and elastographic examinations were performed by using a real-time machine (Acuson S2000; Siemens, Erlangen, Germany) and a 13.5-MHz linear probe. Specific software was used to measure tissue distortion.

**Imaging Methods**

**Conventional Sonographic Examination.** Initially, conventional sonographic images were obtained with the same method used for standard clinical thyroid gland explorations, with the neck slightly extended. Morphologic aspects of the nodule such as size (<1 cm; 1–2 cm; and >2 cm), echogenicity (hypoechoic, hyperechoic, isoechoic, heterogeneous), margins (regular/well-defined or irregular/ill-defined), presence or absence of calcifications (microcalcifications and coarse calcifications), presence or absence of cysts, acoustic transmission (bad, good, or indeterminate), and absence or presence (continuous or interrupted) of halo were evaluated.

The images were acquired, reviewed, and interpreted by 2 radiologists with 6 and 12 years’ experience, respectively, in the evaluation of thyroid nodules. Both were blinded to the final pathologic diagnoses. When there was no agreement, consensus was obtained. It is not the purpose of this article to evaluate the utility of B-mode findings in the differential diagnosis of benign or malignant nodules, or even the interobserver agreement with this technique.

During B-mode sonography exploration, thyroid gland lesions were identified and the region of interest for elastography was selected.

**US Elastography.** After the B-mode sonographic examination, US elastography was performed. The same real-time instruments and probes were used. The probe was placed on the neck with light pressure, and the operator highlighted a box, which included the nodule and sufficient surrounding thyroid tissue when possible. The freehand compression applied on the neck region was slight, and an overlay box indicated too little/too much axial motion or excessive lateral motion. The application displayed a real-time QF score by using advanced signal-intensity analysis. A high QF value (>60) indicated minimal global motion artifacts, and a low QF value (<50) indicated global motion artifacts that may decrease diagnostic value.

The US-elastogram was displayed next to the B-mode image; grayscale images were registered and recorded for subsequent analysis.

As with conventional sonography, the 2 radiologists independently evaluated the different variables defined for US elastography.
analysis for each nodule. Moreover, when discrepancies occurred, consensus was obtained.

In terms of elastography, radiologists evaluated elasticity (predominantly rigid, indeterminate, and predominantly soft), size (larger diameter on US elastography than on B-mode sonography, similar size or smaller diameter on US elastography), and margins (ill-defined or irregular, indeterminate, or well-defined or regular). Elasticity was evaluated according to gray-scale elastograms: Lesions were considered predominantly rigid when >90% of the area of the lesion was black and predominantly soft when >90% area of the lesion was white in gray-scale; an indeterminate category was considered when lesions could not be included in any of these categories.

The best B-mode US elastographic image pairs were acquired and stored for posterior evaluation. Both images were physically separated to avoid the influence of one exploration over another.

According to variables described above and before consensus was reached, we established some criteria to let each radiologist classify nodules into 3 groups, first separately and then by consensus, depending on the suspicion for malignancy or benignity of the lesions with US elastography: type I (suspicious for benignity), type II (indeterminate), and type III (suspicious for malignancy). These criteria are as follows: Type I nodules (suspicious for benignity or predominantly soft); predominantly white in gray-scale elastograms (>90% of the area of the nodule); Type II nodules (indeterminate): classification in type I or III was not possible; Type III nodules (suspicious for malignancy or predominantly rigid): predominantly black in gray-scale elastograms (>90% of the area of the nodule) or greater size with US elastography than on the B-mode image.

**Histopathologic Diagnosis**

Cytologic or histologic diagnosis served as the reference standard for comparison between conventional sonographic imaging and US elastography.

According to cytologic results obtained with FNAB and the classification proposed by the Bethesda System for Reporting Thyroid Cytopathology, we classified thyroid nodules into 1 of the following groups: 1) malignant, 2) suspicious for malignancy, 3) suspicious for follicular or Hürthle cell neoplasia, 4) atypia of uncertain significance, 5) benign, and 6) insufficient or not valid material.20,21 In our hospital, the pathologist is always present when FNAB samples are obtained to reduce the amount of insufficient material or samples with atypia of uncertain significance; an insufficient amount or atypia requires repetition of the procedure.

A definite diagnosis was obtained with FNAB and cytology only in patients included in category 5 (benign). Patients with nodules included by FNAB in groups 1 (malignant) and 2 (suspicious for malignancy) were scheduled for surgery, and histologic results were analyzed. The same procedure occurred with nodules considered in category 3 (suspicious for neoplasia), in which surgery was needed to distinguish carcinoma and adenoma because cytology cannot differentiate these 2 entities. Nodules included in category 4 (atypia of uncertain significance) or 6 (insufficient or not valid material) were re-evaluated to reclassify them into other groups mentioned before.

Definite nodule pathologic classification from cytologic or histologic analyses established 2 groups: 1) malignant nodules, and 2) benign nodules.

**Statistical Analysis**

Qualitative variables were studied as percentages, and quantitative variables were analyzed as means and SDs. Parametric tests were used for statistical evaluation. Results obtained were compared by using the χ² test or the Fisher exact test and McNemar test for paired data and logistic regression analysis. The sensitivity and specificity, PPV, NPV, false-positive rates, and false-negative rates were calculated.

To evaluate interobserver agreement, we calculated the weighted κ index with a CI of 95%.

For all tests, a P value < .05 was considered statistically significant.

**Results**

As mentioned above, 3 patients with their respective 3 nodules were initially excluded from the prospective study because it was not possible to know definite pathologic results. Therefore, 103 patients with a total of 106 thyroid nodules were studied with conventional B-mode sonography and US elastography. Eighty-nine patients were women (86.4%) and 14 were men (13.6%).

Fifty-two of the 106 nodules (49.1%) were located in the right lobe of the gland; 49 nodules (46.2%), in the left lobe; 5 nodules were sited at the isthmus of the thyroid gland (4.7%).

All patients underwent FNAB, and the cytologic results of the nodules based on the classification proposed by the Bethesda conference were as follows: 6 (5.7%) nodules corresponded to category 1 (malignant), 4 (3.8%) nodules were included in category 2 (suspicious for malignancy), 7 nodules (6.6%) were classified as category 3 (suspicious for follicular or Hürthle cell neoplasia), 88 nodules (83%) were benign (category 5), and 1 nodule (0.9%) was initially classified as category 6 (insufficient material) and required surgery. No patient from the final series was included in category 4 (atypia of uncertain significance).

On definite pathologic results of the 106 nodules, 10 (9.4%) nodules corresponded to malignant lesions (8 papillary carcinomas, 1 medullar carcinoma, and 1 anaplastic carcinoma) and 96 nodules (90.6%) were benign.

Table 1 shows different variables studied with US elastography in absolute numbers and percentages obtained with consensus from both radiologists and nodule location on the thyroid gland.

In reference to strain, 16 of the total of 106 nodules (15.1%) were predominantly rigid, 44 nodules (41.5%) were considered predominantly soft, and 46 (43.4%) lesions were classified as indeterminate if it was not possible to classify them in either of the 2 other groups. Fifteen (14.2%) of the nodules had a larger size on elastography than on B-mode sonography, 62 nodules were similar in size in both techniques, and 29 lesions (27.4%) showed a smaller size with US elastography. Ill-defined or irregular margins were defined in 41 nodules (38.7%), whereas the other 65 nodules showed well-defined or regular margins.

Although the number of cases classified as type I (suspicious for benignity) was similar for the 2 readers, reader 1 classified 6 more nodules in type III (suspicious for malignancy) than reader 2. This last radiologist classified 7 more cases in type II or the indeterminate category. When consensus was obtained, more cases were included in the indeterminate group, and the total number of nodules classified type I or suspicious for benignity was reduced (Fig 2).

Interobserver agreement in the classification of thyroid nodules into the previously defined US elastography groups I,
II, or III with \( \frac{P}{H11021} \) 0.001 was excellent in our study, with a weighted index of 0.82 (95% IC, 0.74–0.89).

As a verification of the utility of US elastography, different variables studied with US elastography and defined by consensus between both radiologists correlated with definite pathologic results, considered the reference standard in this study (Table 2).

Six nodules from the group of 10 malignant nodules confirmed by cytology or histology (60%) were predominantly rigid on US elastography, and the remaining 4 nodules (40%) showed an indeterminate aspect, which did not allow classifying them as rigid or soft (Fig 3). None of the nodules were described as predominantly soft with US elastography in the group of malignant nodules, and 100% of the nodules described as soft were benign in the definite pathologic classification (Fig 4).

Regarding nodule size, most studied nodules (62 nodules, 58.5%) were similar in size on US elastography and conventional sonography. Fifteen nodules (14.2%) had a greater diameter on elastography than on B-mode sonography. Among these last ones, 4 (26.7%) and 11 (73.3%) corresponded to malignant and benign lesions, respectively.

Twenty-nine nodules were smaller on US elastography than with conventional sonography; 96% of these lesions were benign and 4% were malignant.

In the evaluation of the utility of the classification of thyroid nodules in groups I (suspicious of benignity), II (indeterminate), and III (suspicious of malignancy) with US elastography, classifications proposed by consensus correlated with pathologic results (Table 3).

Thirty-nine nodules were classified by consensus as type I or suspicious for benignity, and none corresponded to a malignant lesion. On the other hand, among 16 nodules considered by consensus as type III or suspicious for malignancy, 5 nodules corresponded to malignant lesions and 11 nodules were benign. Fifty-one nodules were considered indeterminate in our study, 5 corresponding to malignant lesions and 46 corresponding to benign nodules.

Discussion

Some pathologic conditions induce considerable changes in the soft-tissue structure, modifying its elastic properties and leading to increased stiffness of the evolved tissue. Because malignancy has been correlated with hardness of tissues, palpation of the thyroid gland constitutes a way of discriminating nodules suspicious for malignancy from those that are benign. This method is subjective and depends on the depth of the lesion.

Elastography is a technique that uses sonography to analyze

<table>
<thead>
<tr>
<th>Total of No. of Nodules</th>
<th>No. of Nodules in Left Lobe</th>
<th>No. of Nodules in Right Lobe</th>
<th>No. of Nodules in the Isthmus</th>
<th>( P ) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predominantly rigid</td>
<td>16 (15.1%)</td>
<td>9 (56.3%)</td>
<td>7 (43.7%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>46 (43.4%)</td>
<td>18 (39.1%)</td>
<td>26 (56.5%)</td>
<td>2 (4.4%)</td>
</tr>
<tr>
<td>Predominantly soft</td>
<td>44 (41.5%)</td>
<td>22 (50%)</td>
<td>19 (43.2%)</td>
<td>3 (6.8%)</td>
</tr>
<tr>
<td>Size</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greater diameter in US elastography than B-mode US</td>
<td>15 (14.2%)</td>
<td>7 (46.7%)</td>
<td>8 (53.3%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Same size</td>
<td>62 (58.5%)</td>
<td>29 (46.8%)</td>
<td>32 (51.6%)</td>
<td>1 (1.6%)</td>
</tr>
<tr>
<td>Lesser diameter in US elastography than B-mode US</td>
<td>29 (27.3%)</td>
<td>13 (44.8%)</td>
<td>12 (41.4%)</td>
<td>4 (13.8%)</td>
</tr>
<tr>
<td>Margins</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ill-defined or irregular</td>
<td>41 (38.7%)</td>
<td>20 (48.8%)</td>
<td>19 (46.3%)</td>
<td>2 (4.9%)</td>
</tr>
<tr>
<td>Well-defined or regular</td>
<td>65 (61.3%)</td>
<td>29 (44.6%)</td>
<td>33 (50.8%)</td>
<td>3 (4.6%)</td>
</tr>
</tbody>
</table>
the stiffness of a nodule by measuring the amount of distortion that occurs when the nodule is subjected to external pressure. This technique has been extensively studied in breast lesions and the prostate, pancreas, and lymph nodes. Recent studies have evaluated the use of US elastography for detecting malignant thyroid nodules. These uniformly suggest that US elastography increases the ability to discriminate benign and malignant nodules. Recently, Bojunga et al published a meta-analysis of elastography for the differentiation of benign and malignant thyroid nodules, concluding that US elastography can be used with high sensitivity in the work-up of thyroid nodules and might be a useful method in addition to or even instead of FNAB to select patients for surgery.

We present a large series of 103 patients who have been studied with elastography for the evaluation of thyroid nodules. All patients included had been referred for the FNAB procedure as occurs in real clinical practice, so we did not study any nodule that was not previously indicated for aspiration. We believe that this method is more representative of the population that will undergo the test in clinical care.

We proposed a simplified classification of stiffness based on gray-scale patterns instead of color because we believe that this method of interpretation is more reliable and reproducible. We also separated images from conventional sonography and US elastography for interpretation to avoid the influence of one reader over another; to our knowledge, only Lyshchik et al have used this procedure, which we believe is essential for blinded analysis. In addition, we evaluated and compared tumor size with B-mode sonography and US elastography; to our knowledge, only Lyshchik et al have included this variable in the analysis.

In our study, there was a significant statistical association between elasticity as an independent variable and the classification proposed (combining elasticity, size, and margins) and definite pathologic results. Malignant nodules could be excluded by elastography (44 nodules appeared soft in US elastography; all were histologically benign) in this cohort of patients. US elastograms predicted malignancy with 100% sensitivity and 40.6% specificity. Sensitivity was very high in this study and no false-negatives were found, so all nodules interpreted as suspicious for benignity in our study, mostly due to extensive white areas in gray-scales in elastograms, were benign on pathologic results. This result is concordant with a high NPV of the technique and suggests that US elastography could be helpful in the discrimination of nodules that should be left alone and for which conventional sonographic follow-up would be sufficient. All malignant nodules on definite

<table>
<thead>
<tr>
<th>Classification of Nodules with US Elastography by Consensus</th>
<th>Malignant Nodules</th>
<th>Benign Nodules</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I: suspicious for benignity (n = 39)</td>
<td>0</td>
<td>39</td>
<td>.001</td>
</tr>
<tr>
<td>Type II: indeterminate (n = 51)</td>
<td>5</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Type III: suspicious for malignancy (n = 16)</td>
<td>5</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>
diagnosis were suspicious for malignancy or indeterminate on elastograms in the present study; all those nodules needed to be studied by FNAB to avoid missed malignant lesions. However, in the group of nodules that appeared predominantly rigid, 62.5% were false-positive. The specificity of the US elastography in our study was a very low, probably secondary to a high number of indeterminately classified nodules, in which determining malignancy or benignity was not possible with elastography because only minor information was provided.

The κ index of concordance for elastographic interpretation in our study was 0.82 (95% CI), reflecting an excellent interobserver agreement. In their published article, Park et al.²⁹ did not find a statistically significant concordance.

Most publications referring to US elastography in the evaluation of thyroid nodules propose a scale of tissue stiffness based on the color of the elastograms: 4-point scales,²³–²⁶ 5-point scales,²⁵ and 6-point scales.²⁴ Lyshchik et al.²⁸ evaluated 52 thyroid nodules in 31 patients and found a specificity of 96% and a sensitivity of 82%. Asteria et al.²⁵ examined 86 nodules in 67 patients and found a sensitivity of 94% and a specificity of 81%. The PPVs and NPVs were 55% and 98%, respectively. Friedrich-Rust et al.²³ investigated 53 nodules in 50 patients and found an 86% sensitivity and an 87% specificity. Rago et al.²⁸ found scales 4–5 highly predictive of malignancy with a sensitivity of 97% and a specificity of 100%. An off-line processing of strain images, a different method from ours, was also used in the study of Lyshchik et al. Hong et al.²⁴ evaluated the diagnostic utility of real-time US elastography with a 6-point scale of color in elastograms of 90 patients who were referred for surgical treatment, so the final diagnosis was based on the results of the histopathologic examination of resected thyroid gland tissue of the 145 thyroid nodules studied. High sensitivity (88%) and specificity (90%) were found, with a PPV of 81% and a NPV of 93%.

Recently, Rago et al.²⁵ studied 176 patients with 195 nodules with specifically indeterminate or non-diagnostic cytology on FNAB for whom histology was available after thyroidectomy, and proposed a 3-level elasticity score, similar to the method we used. The score of 1, describing high elasticity, was strongly predictive of benignity. By combining scores 2 and 3, they found that US elastography had a sensitivity of 96.8% and a specificity of 91.8%. They also studied interobserver agreement (95%) but exclusively based on color elastograms. However they did not evaluate the size of the nodule with respect to B-mode conventional sonography or either margin of the nodule.

Some limitations in this study need to be addressed. First, some publications speculate that US-elastograms of nodules with cystic components or coarse calcifications are not reliable, and we have not excluded these nodules from the study. Second, as in other publications, it was not possible in all cases in our study to establish the rigidity of the lesion compared with the surrounding parenchyma, which is the classic proposed optimal interpretation. Previous reports concluded that US elastography and even FNAB are not suitable for the diagnosis of follicular carcinoma, and 4 of 9 follicular carcinomas in the meta-analysis published by Bojunga et al.²¹ were missed on US elastography. We had no case of follicular carcinoma in our series that may have altered the results observed. Further series including more nodules, specifically follicular carcinomas, are required.

Conclusions
We believe that US elastography is a promising technique that can assist in the differential diagnosis of thyroid cancer and in the selection of nodules suitable for FNAB. To date, US elastography is a relatively easy and quick-to-perform supplemental technique that may be practical for routine clinical care, mostly helping to identify probable benign nodules that could be followed up. US elastography does not require off-line strain imaging reconstruction, and performing it with other techniques (ie, by using shear-waves) will probably increase its overall accuracy. Consequently, we believe that it may be useful to introduce US elastography into routine clinical practice.

References


