Sonographic Features of Benign Thyroid Nodules
Interobserver Reliability and Overlap With Malignancy

Jeffrey R. Wienke, MD, Wui K. Chong, MD, Julia R. Fielding, MD, Kelly H. Zou, PhD, Carol A. Mittelstaedt, MD

Objective. To prospectively determine the sonographic findings of nodular hyperplasia of the thyroid, to compare these with reported findings associated with malignancy, and to assess interobserver reliability. Methods. Seventy thyroid nodules were scanned, and then biopsies of the nodules were performed under sonographic guidance with fine-needle cytologic analysis; in all cases images were reviewed by 2 experienced radiologists without knowledge of clinical outcome. Findings reported associated with malignancy were specifically assessed. Interobserver agreement between the expert and secondary readers for each finding was calculated by the $\kappa$ or weighted $\kappa$ statistic and the Fisher exact test of independence. Results. There were 68 benign and 2 malignant nodules in a population of 63 female and 7 male patients. The mean benign nodule size was 2.9 cm; 60% were solid; 54% were hypoechoic; 59% were microlobulated or macrolobulated; 47% had central vascularity; 24% contained calcifications; and 82% were elliptical in shape. There was very good interobserver reliability for the presence of calcium ($\kappa = 0.91$) and good agreement for the presence and location of vascularity ($\kappa = 0.75$) and the amount of cystic components ($\kappa = 0.62$; all $P < .01$). Conclusions. Sixty-nine percent of benign nodules had at least 1 finding reported previously as associated with malignancy. The interobserver reliability of the sonographic findings was good to very good for 3 of the 5 findings assessed. Key words: interobserver reliability; thyroid nodules; thyroid sonography.

Received March 10, 2003, from the Department of Radiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina USA (J.R.W., W.K.C., J.R.F., C.A.M.); and Department of Radiology, Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts USA (K.H.Z.). Revision requested April 1, 2003. Revised manuscript accepted for publication May 21, 2003.

Address correspondence and reprint requests to Julia R. Fielding, MD, Department of Radiology, University of North Carolina at Chapel Hill, Campus Box 7510, 101 Manning Dr, Chapel Hill, NC 27599 USA.
E-mail: julia_fielding@med.unc.edu.

Fine-needle aspiration is the established diagnostic procedure for palpable and nonpalpable thyroid nodules. Recently, Kim et al$^1$ reported sonographic criteria predictive of thyroid cancer and indicating the need for fine-needle aspiration. These sonographic features included microcalcifications, irregular or macrolobulated borders, marked hypoechogenicity, and a shape that was more tall than wide. Additionally, in separate articles, Papini et al$^2$ and Frates et al$^3$ reported central or intranodular vascularity as a sonographic finding associated with thyroid cancer, although the finding did not obviate the need for biopsy.

Our clinical experience, like that of others, has been that no sonographic finding is adequately sensitive or specific for diagnosis of cancer and that solid thyroid nodules always require biopsy. The goal of our study was to deter-
mine the frequency of the above-described sono-
graphic criteria in our patient population, which
consists primarily of patients with benign nodular
hyperplasia or goiters. In addition, we evaluated the
interobserver reliability of review of the sonograph-
ic findings.

Materials and Methods

Clinical Procedures
Between March 2001 and June 2002, 82 thyroid
nodules were assessed by a standardized sono-
graphic examination of the thyroid, including color
Doppler flow imaging. The examination was imme-
diately followed by fine-needle aspiration. Thyroid
nodules were scanned in the Department of
Radiology with either an Acuson Sequoia system
(Siemens Medical Solutions, Mountain View, CA)
using a 10-MHz linear transducer or an HDI 5000
systems (Philips Medical Systems, Bothell, WA)
using a 12-MHz linear transducer. Biopsies of the
nodules were performed with a 23-gauge hepar-
arinized needle. The first pass was obtained without
aspiration; subsequent samples were obtained with
aspiration. A cytopathologist immediately checked
the samples for adequacy. Three to 7 passes were
performed on each nodule. All fine-needle aspira-
tion specimens with the diagnosis of benign
multinodular goiter were considered benign for this
study. For those nodules with follicular cells, surgi-
cal results after partial thyroidectomy were used to
determine benign versus malignant etiology.

Fine-needle aspiration specimens that were
deemed inadequate for diagnosis were excluded
from our study (n = 6). Nodules that underwent
biopsies but that did not have diagnostic-quality
gray scale images were excluded (n = 6). Two
patients had inadequate color Doppler images but
were included in the study because the other find-
ings were identifiable. These exclusions resulted in
70 nodules for further study.

Institutional Review Board approval was obtained
for review of medical records. Images were elec-
tronically stripped of identifying data. Two experi-
enced radiologists without knowledge of clinical
outcome independently reviewed each set of
images. The findings of the most experienced read-
er, designated as the expert reader, were compared
with pathologic outcome. Sonographic findings
were divided into 7 categories designed to assess
those reported as associated with malignancy.
These included calcifications, echogenicity, shape,
borders, vascularity, size, and composition.

Calcifications were divided into those located
centrally within the nodule, peripherally, and
none. Posterior shadowing of at least 1 of the
suspected calcifications was required to con-
sider the finding present. Echogenicity cate-
gories included markedly hypoechoic (the
thyroid nodule was hypoechoic in relation to
the adjacent strap musculature), hypoechoic,
isoechoic, and hyperechoic in relation to the
adjacent normal thyroid. Dimensions were
measured in the superior-to-inferior and right-
to-left dimensions to assess an elliptical versus
a circular shape. Borders were defined as
smooth, microlobulated, or macrolobulated.
Vascularity as determined by color Doppler
imaging was defined as absent, peripheral, or
central. If both peripheral vascularity and cen-
tral vascularity were identified, the lesions were
characterized as having central vascularity. Size
was defined as the maximal dimension of the
nodule and was divided into 3 categories:
smaller than 1, 1 through 1.5, and larger than
1.5 cm. Composition was defined as solid, one-
third cystic, two-thirds cystic, or greater than
two-thirds cystic.

Statistical Analysis
Findings of the expert reader were used to deter-
mine the percentage of benign nodules with each
feature based on summary statistics. Interobserver
agreement between the expert and secondary read-
ers for each finding was calculated by the $\kappa$
or, in the case of nominal data, weighted $\kappa$
statistic and the Fisher exact test of independence.

Results
There were 68 benign nodules and 2 malignant
nodules in a population of 63 female and 7 male
patients. Both malignant nodules occurred in male
patients. Benign nodules ranged in size from 0.8 to
8.0 cm with a mean of 2.9 cm. Forty-one (60%)
of the 68 benign nodules were solid; 11 (16%) were
one-third cystic; an equal percentage were two-
thirds cystic; and 5 (7%) were greater than two-
thirds cystic. Thirty-five (51%) of 68 nodules were
hypoechoic to the thyroid, and 2 (3%) were hypo-
echoic to the adjacent strap muscles. Twenty-eight
(41%) of 68 nodules had smooth borders; 21 (31%)
were microlobulated; and 19 (28%) were macrolob-
ulated. Sixteen (24%) of 66 nodules contained calci-
fications, and 52 (76%) had none (Fig. 1). Of these
16, 15 were within the thyroid parenchyma, and 1
was peripheral and curvilinear. Sixteen (24%) of 66 nodules had no vascularity; 19 (29%) had peripheral vascularity only; and 31 (47%) had central vascularity (Fig. 2). Most benign nodules were elliptical in shape (56 [82%] of 68), with a median ratio of 1.28. Sonographic features are summarized in Table 1.

There was very good interobserver reliability for the presence of calcium ($\kappa = 0.91$) and good agreement for the presence and location of central vascularity ($\kappa = 0.75$) and the amount of cystic components ($\kappa = 0.62$; all $P < .01$ by Fisher exact test). There was only fair agreement for echogenicity ($\kappa = 0.37$) and poor agreement for border assessment ($\kappa = 0.13$).

**Discussion**

We evaluated previously reported sonographic findings associated with thyroid malignancy in 68 benign lesions. We found that 69% of histologically proved benign thyroid nodules had at least 1 of the positive predictors of malignancy reported by Kim et al.\(^1\) This corroborates the general consensus that sonographic findings are not adequate for determining the need for biopsy. Regarding the qualitative assessment of sonographic findings, experienced readers produce highly reproducible reports of thyroid calcification, the presence and location of vas-

---

**Table 1. Gray Scale and Doppler Sonographic Findings in 68 Benign Thyroid Nodules**

<table>
<thead>
<tr>
<th>Finding</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Borders</td>
<td></td>
</tr>
<tr>
<td>Smooth</td>
<td>28</td>
</tr>
<tr>
<td>Microlobulated</td>
<td>21</td>
</tr>
<tr>
<td>Macrolobulated</td>
<td>19</td>
</tr>
<tr>
<td>Calcifications</td>
<td></td>
</tr>
<tr>
<td>Central</td>
<td>15</td>
</tr>
<tr>
<td>Peripheral</td>
<td>1</td>
</tr>
<tr>
<td>None</td>
<td>52</td>
</tr>
<tr>
<td>Composition</td>
<td></td>
</tr>
<tr>
<td>Solid</td>
<td>41</td>
</tr>
<tr>
<td>$\frac{1}{2}$ cystic</td>
<td>11</td>
</tr>
<tr>
<td>$\frac{2}{3}$ cystic</td>
<td>11</td>
</tr>
<tr>
<td>$&gt;\frac{2}{3}$ cystic</td>
<td>5</td>
</tr>
<tr>
<td>Echogenicity*</td>
<td></td>
</tr>
<tr>
<td>Markedly hypoechoic</td>
<td>2</td>
</tr>
<tr>
<td>Hypoechoic</td>
<td>35</td>
</tr>
<tr>
<td>Isoechoic</td>
<td>18</td>
</tr>
<tr>
<td>Hyperechoic</td>
<td>13</td>
</tr>
<tr>
<td>Vascularity†</td>
<td></td>
</tr>
<tr>
<td>Peripheral</td>
<td>19</td>
</tr>
<tr>
<td>Central</td>
<td>31</td>
</tr>
<tr>
<td>None</td>
<td>16</td>
</tr>
</tbody>
</table>

*Markedly hypoechoic indicates hypoechoic in relation to the adjacent strap musculature; hypoechoic, hypoechoic to adjacent thyroid tissue; isoechoic, isoechoic to adjacent thyroid tissue; and hyperechoic, hyperechoic to adjacent thyroid tissue.

†Two patients had inadequate color Doppler images.
cularity, and the amount of cystic components. Border and echogenicity assessment will likely vary considerably with each reader.

Barring local invasion or lymphadenopathy, no sonographic finding can reliably discriminate between benign and malignant solitary thyroid nodules. Kim et al\(^1\) attempted to provide new sonographic criteria for fine-needle aspiration of nonpalpable solid thyroid nodules. They performed biopsies on 155 solid nonpalpable nodules. Eighty-two (53%) of these had at least 1 of the above-described sonographic findings the authors considered suggestive of malignancy, and 46 (56%) of 82 specimens were malignant. Although our study evaluated the sonographic findings of benign, rather than malignant, nodules, we think that this condition more closely approaches the patient population sent for evaluation in most US hospitals.

In our study, we found that 42 (60%) of 70 nodules were solid, 35 (83%) of which had at least 1 malignant characteristic. Of these 35, 1 (3%) was malignant. In a similar study, Papini et al\(^2\) found that 18 (9.2%) of 195 nonpalpable, solid, hypofunctioning nodules on technetium Tc 99m thyroid scans were malignant. Additionally, both Frates et al\(^3\) and Papini et al\(^2\) found central or intranodular vascularity to be associated with malignancy. In our population, approximately half of the benign nodules had intranodular vascularity, similar to the proportion of Frates et al\(^3\).

Given the inclusion criteria of a hypofunctioning nodule established by Papini et al\(^2\), one would expect a higher percentage of malignancy than in the study by Kim et al\(^1\). The discrepancy in the percentage of actual malignancies from those indicated by sonographic findings is intriguing. Perhaps the use of clinical criteria to select a portion of the patient population led to an increase in the percentage of malignancies. Kim et al\(^1\) had patients with hoarseness, swallowing difficulties, or neck discomfort as well as those who underwent breast cancer screening. There is a controversial association between thyroid and breast cancer.\(^4\)

The role of sonography in the thyroid nodule workup is detection of solid nodules and biopsy guidance for those nodules not easily palpated. These usually include nonpalpable nodules of 1 cm or larger, predominantly cystic lesions with small solid components, and lesions in positions that are prone to complications. Additionally, nodules smaller than 1 cm with a clinical or laboratory suggestion of malignancy should be considered for sonographically guided fine-needle aspiration.\(^5\)–\(^11\) In our patient population, which had nearly 100% benign lesions, sonographic findings reported as associated with thyroid cancer were present in most; therefore, these findings should not be used to exclude biopsy.

References


