Role of Sonographic Diagnosis in Managing Bethesda Class III Nodules

BACKGROUND AND PURPOSE: Bethesda class III cytology is an important limitation of the US-FNA in assessing thyroid nodules. This study aimed to assess the diagnostic efficacy of US in evaluating thyroid nodules with Bethesda class III cytology.

MATERIALS AND METHODS: From January 2008 to December 2009, 1036 patients with 1289 thyroid nodules diagnosed by US and subsequent US-FNA biopsy were enrolled in the study. On the basis of US features, each thyroid nodule was prospectively classified by a single radiologist into 1 of 5 diagnostic categories: benign, probably benign, borderline, possibly malignant, and malignant. Solid nodules were classified by using all 5 categories, whereas partially cystic nodules were classified by using 4 (borderline was omitted). We calculated the diagnostic efficacy of thyroid US by comparing the US diagnoses with the histopathology results of Bethesda class III nodules.

RESULTS: Of the 51 Bethesda class III nodules, 35 were surgically confirmed and 8 were histologically diagnosed, and a malignancy rate of 46.5% (20/43) was determined. From the 43 nodules, the sensitivity, specificity, positive and negative predictive values, and accuracy were calculated with 95% confidence intervals. Borderline nodules excluded (100%, 94.7%, 93.3%, 100%, and 97.0%, respectively) and with the 9 when reclassified as benign (63.6%, 95.2%, 93.3%, 71.4%, and 79.1%, respectively) and malignant (100%, 85.7%, 88.0%, 100%, and 93.0%, respectively). The values obtained with exclusion and malignancy reclassification of 9 borderline nodules were not significantly different (P = .250).

CONCLUSIONS: US diagnosis by using the present US classification system can be helpful for managing Bethesda class III nodules.

ABBREVIATIONS: CNB = core-needle biopsy; FNA = fine-needle aspiration; PCTN = partially cystic thyroid nodule; US = sonography

U S-FNA is an easy-to-use and accurate tool for evaluating thyroid nodules. Many studies have reported that US-FNA has high diagnostic adequacy and efficacy when used to assess thyroid nodules. However, indeterminate cytology, defined as cytologic results that do not provide a differential diagnosis between malignancy and benignancy, is an important limitation of the US-FNA in assessing thyroid nodules. In the Bethesda System, indeterminate cytology was subdivided into Bethesda classes III (atypia of undetermined significance or follicular lesion of undetermined significance in the Bethesda System) and IV (follicular neoplasm or suspicious for a follicular neoplasm). However, there are diverse reports citing the risk of malignancy, ranging from 5% to 45%, and the management for Bethesda class III nodules seems to be diverse, also depending on physicians or institutions, including clinical observation, repeat FNA, CNB, or surgery. Furthermore, Cibas and Ali emphasized that the use of Bethesda class III should be restricted because its necessity is debated.

Recently, a few studies have reported the feasibility of using thyroid US to predict malignancy of the nodules assigned as Bethesda class III cytology in the initial US-FNA (termed “Bethesda class III nodules”). However, these studies were either retrospective or did not include a categoric diagnostic classification. Compared with prospective studies, retrospective evaluations of thyroid images are limited so that they are restricted to a pre-existing set of images. In the present study, we aimed to assess the feasibility and role of thyroid US in predicting malignancy for Bethesda class III nodules by using a real-time US examination and a specific US classification system.

Materials and Methods

Patients

From January 2008 to December 2009, a single radiologist performed thyroid US to diagnose nodular thyroid disease in a consecutive series of patients at our hospital. Of these patients, 1036 (876 women and 160 men; mean age, 49.0 ± 12.0 years) who underwent US-FNA for ≥1 thyroid nodule with the largest diameter ≥5 mm were enrolled in this study. We obtained informed written consent from all patients before performing US-FNAs. Our institutional review board approved the study.

Thyroid US

Thyroid US was performed by a single radiologist (D.W.K.) with 8 years of relevant experience by using a high-resolution sonographic instrument (IU22; Philips Medical Systems, Bothell, Washington) equipped with a 12- to 5-MHz linear probe. We used 2 different categoric systems for classifying solid thyroid nodules or PCTNs. On the basis of the real-time thyroid US, solid thyroid nodules (defined as purely solid or as predominantly solid with a cystic component com-
posing <10% of the total volume) were prospectively classified into 1 of 5 categories: 1) benign, 2) probably benign, 3) borderline, 4) possibly malignant, and 5) malignant. PCTNs (defined as thyroid nodules with a cystic component composing ≥10% of the total volume) were grouped on the basis of the same real-time US, but the borderline class was excluded.

For solid thyroid nodules, the US features indicating benignancy included an ovoid or flat shape, isoechoegenicity, a smooth margin, and peripheral vascularity. The US features of thyroid nodules that were still indeterminate for benignancy/malignancy (classified as having borderline features) included hypoechogenicity, centrally predominant vascularity, and macrocalcifications (including eggshell calcification and intranodular macrocalcifications). Solid thyroid nodules diagnosed as malignant were characterized by marked hypoechogenicity, a spiculated margin, microcalcifications, a taller-than-wide shape, and the existence of lymphadenopathy with intranodular cystic components or microcalcifications in the perithyroidal region. For PCTNs, the US characteristics of benignancy included a configuration that was either concentric or eccentric with a blunt angle, a smooth free-margin, peripheral or no vascularity, a spongiform appearance or daughter cysts in the solid component, intranodular comet-tail artifacts formed by colloidal crystals, and isoechoegenicity of the solid component. The US features of a malignant PCTN included an eccentric configuration with an acute angle, microcalcifications, macrolobulation or irregularity of the free-margin, perinodular infiltration, a centripetal vascularity in the pedicle, and the existence of lymphadenopathy with intranodular cystic components or microcalcifications in the perithyroidal region.

The criteria for US diagnosis of thyroid nodules differed on the basis of the type of nodule (solid thyroid nodule or PCTN) (Fig 1). For solid thyroid nodules, those with ≥3 US features of benignancy and no malignant or borderline US features were considered “benign.” Solid thyroid nodules with 1 or 2 US features of benignancy and no malignant or borderline US features were considered “probably benign.” Those with ≥1 borderline US feature and no US features of malignancy, regardless of benign US features, were considered “borderline.” Those with 1 US feature of malignancy, regardless of borderline or benign US features, were considered “possibly malignant.” Solid thyroid nodules with ≥2 US features of malignancy, regardless of borderline or benign US features, were considered “malignant.” The criteria underlying the US diagnosis of PCTNs were as follows: PCTNs with ≥3 US features of benignancy and no features of malignancy were considered “benign.” PCTNs with 1 or 2 US features of malignancy and no features of malignancy were considered “possibly malignant.” PCTNs with 1 US feature of malignancy, regardless of other benign features, were considered “possibly malignant.” PCTNs with ≥2 US features of malignancy, regardless of other benign features, were considered “malignant.”

**US-FNA and Cytologic Analysis**

US-FNA was performed immediately after thyroid US by the same radiologist. All 1036 patients underwent US-FNA, and they had a total of 1289 nodules (nodule size range, 0.5–9.8 cm; mean size, 1.5 cm). For each sample, a smear was prepared on 4–6 slides, fixed in 95% ethanol, and sent to the department of pathology for Papanicolaou staining. In case of PCTN, the remaining aspirate within the syringe was sent for cell analysis.

Cytologic diagnoses were made as follows: 1) inadequate (nondiagnostic or unsatisfactory in the Bethesda System), 2) benign (benign in the Bethesda System), 3) indeterminate (atypia of undetermined...
significance or follicular lesion of undetermined significance in the Bethesda System), 4) follicular neoplasm (follicular neoplasm or suspicious for a follicular neoplasm in the Bethesda System), 5) suspicious for malignancy (suspicious for malignancy in the Bethesda System), and 6) positive for malignancy (malignant in the Bethesda System). The cytologic results were unsatisfactory when <6 clusters of thyroid follicular cells containing no identifiable colloid were observed in a preparation. Benign cytology included nodular goiter, nodular goiter with hyperplastic nodules, colloid nodules, cyst contents with or without benign follicular cells, and lymphocytic thyroiditis. Indeterminate cytology was equivalent to the specimens with atypical cells or follicular cells of undetermined significance. Cellular specimens with abundant follicular cells arranged in a microfollicular pattern with little or no colloid or cellular specimens with a predominant population of Hurthle cells were reported as follicular neoplasms. Specimens were considered suspicious for malignancy if they demonstrated features of a malignant neoplasm that were quantitatively or qualitatively insufficient to make a definite diagnosis of malignancy. Specimens showing abundant cells with malignant cytologic features had a positive-for-malignancy cytology.

A US-CNB was performed with an 18-ga automatic biopsy gun (Accut; TSK Laboratory, Tochigi, Japan) by the same radiologist as mentioned previously. The interval between US-FNA and US-CNB was 17.3 days (range, 6–35 days). For each biopsy, 2 samples were obtained after the administration of local anesthesia.

**Statistical Analysis**

The Bethesda class III nodules were prospectively classified on the basis of real-time US. Thyroid nodules that had been diagnosed with US as benign or probably benign were classified as negative (benign category), and those diagnosed as possibly malignant and malignant were classified as positive (malignant category). We calculated the sensitivity, specificity, positive and negative predictive values, and accuracy of US diagnosis in comparison with US diagnoses and histopathology results for Bethesda class III nodules. We used the McNemar test to calculate the sensitivity, specificity, positive and negative predictive values, and accuracy associated with the US diagnoses of Bethesda class III nodules. A P value <.05 was statistically significant. Data analyses were performed by using the Statistical Package for the Social Sciences for Windows (Version 17.0.1, SPSS, Chicago, Illinois).

**Results**

For the 1289 thyroid nodules across all 1036 patients (mean number of nodules obtained by US-FNA, 1.24 per patient), the incidence of adequate sampling was 92.6% (1193/1289). Of the 1289 nodules, cytologic results led to 51 classified as indeterminate (atypia of undetermined significance or follicular lesion of undetermined significance in the Bethesda System) (51/1289, 4.0%).

Of the 51 Bethesda class III nodules (45 from women and 6 from men; range of nodule size, 0.5–5.4 cm; mean size, 1.6 cm), there were 49 solid nodules (Fig 2) and 2 PCTNs. Thirty-five of these nodules (33 from women and 2 from men; range of nodule size, 0.5–4.2 cm; mean size, 1.43 cm) were surgically removed for reasons beyond Bethesda class III, including an US diagnosis of suspected malignancy (n = 12), malignant cytology on repeat US-FNA (n = 4, Fig 3), Bethesda class III in repeat US-FNA (n = 2), malignant histology in US-CNB (n = 4), the presence of associated thyroid malignancy (n = 10), and a patient request (n = 3). Of the 35 surgically removed nodules, 34 were solid and 1 was a PCTN.
The nodule was confirmed to be a papillary thyroid microcarcinoma (follicular variant) that arose from nodular hyperplasia. A small hypoechoic solid component of the nodule (arrow) was the focus of the sampling, and the cytologic result was suspicious for malignancy (Bethesda class V). After thyroid surgery, the nodule was confirmed to be a papillary thyroid microcarcinoma (follicular variant) that arose from nodular hyperplasia.

Of the 16 nonsurgical Bethesda class III nodules, 8 were nonsurgically confirmed by US-CNB with the histology results indicating 8 nodular hyperplasias. On the basis of histopathologic results, a malignancy rate of 46.5% (20/43) was determined. The US diagnoses and histopathologic results for the 43 Bethesda class III nodules are summarized in Table 1. Of the 16 nonsurgical nodules, 8 were not followed up due to patient loss and were excluded in the calculation of diagnostic indices for thyroid US.

When 9 nodules assigned a borderline US diagnosis were excluded, the sensitivity, specificity, positive and negative predictive values, and accuracy of US diagnoses for differentiating malignancy and benignancy were 100%, 94.7%, 93.3%, 100%, and 97.0%, respectively. If the same 9 nodules were reclassified as malignant, the sensitivity, specificity, positive and negative predictive values, and accuracy of US diagnoses were 63.6%, 95.2%, 93.3%, 71.4%, and 79.1%, respectively. Excluding nodules with a borderline US diagnosis yielded a high diagnostic efficacy of thyroid US, but it was not significantly different (P = .250) from that obtained when the same nodules were reclassified as malignant. However, compared with these approaches, the diagnostic efficacy of thyroid US was significantly lower (P = .039) when the nodules with a borderline US diagnosis were reclassified as benign (Table 2).

Discussion

US-FNA is the first-line method for the diagnostic evaluation of nodular thyroid disease because it is simple, safe, accurate, and cost-effective. The use of US-FNA to evaluate thyroid nodules has improved the detection rates for thyroid cancer, increased cancer detection rates, and decreased the number of benign thyroid surgeries performed.1-4 However, one of the significant limitations of US-FNA is Bethesda class III cytologic criteria. To overcome this limitation, many physicians have investigated potential risk factors for malignancy associated with Bethesda class III nodules, such as sex, age, and nodule size.10-15 Several studies have suggested that being male, being older than 40 years of age, and having a nodule with the largest diameter of ≥2 cm are significantly related to the risk of a Bethesda class III nodule being malignant.11,12 Nevertheless, these risk factors do not help to prevent or decrease the rate of unnecessary surgery for Bethesda class III nodules.

The incidence of Bethesda class III thyroid cytology varies greatly among studies,7-15 but limited use of Bethesda class III interpretation is recommended for approximately 7% or fewer of all thyroid FNAs.6 In our study, the incidence of Bethesda class III thyroid cytology of all thyroid FNAs was 4.0% (51/1289). The management for Bethesda class III nodules seems to be differ-

---

Table 1: US diagnoses and histopathologic results in 43 thyroid nodules with Bethesda class III

<table>
<thead>
<tr>
<th>US Diagnoses (No.)</th>
<th>Histopathologic Results (No.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benign (5)</td>
<td>NH (5)</td>
</tr>
<tr>
<td>Probably benign (14)</td>
<td>FA (1), NH (13)</td>
</tr>
<tr>
<td>Borderline (10)</td>
<td>PTC (7), FTC (1), NH (2)</td>
</tr>
<tr>
<td>Possibly malignant (8)</td>
<td>PTC (6), MTC (1), NH (2)</td>
</tr>
<tr>
<td>Malignant (6)</td>
<td>PTC (6)</td>
</tr>
<tr>
<td>Total (43)</td>
<td></td>
</tr>
</tbody>
</table>

Note:—NH indicates nodular hyperplasia; FA, follicular adenoma; PTC, papillary thyroid carcinoma; FTC, follicular thyroid carcinoma; MTC, medullary thyroid carcinoma.

Table 2: Diagnostic indices of US diagnoses for Bethesda class III nodules

<table>
<thead>
<tr>
<th>Measure</th>
<th>Exclusion of Borderline US Diagnosis (%)</th>
<th>Inclusion of Borderline US Diagnosis into Benign Category (%)</th>
<th>Inclusion of Borderline US Diagnosis into Malignant Category (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>14/14 (100)</td>
<td>14/22 (63.6)</td>
<td>22/22 (100)</td>
</tr>
<tr>
<td>Specificity</td>
<td>18/19 (94.7)</td>
<td>20/21 (95.2)</td>
<td>18/21 (85.7)</td>
</tr>
<tr>
<td>PPV</td>
<td>14/15 (93.3)</td>
<td>14/15 (93.3)</td>
<td>22/25 (88.0)</td>
</tr>
<tr>
<td>NPV</td>
<td>18/18 (100)</td>
<td>20/28 (71.4)</td>
<td>18/18 (100)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>32/33 (97.0)</td>
<td>34/43 (79.1)</td>
<td>40/43 (93.0)</td>
</tr>
<tr>
<td>False-negative rate</td>
<td>0/14 (0)</td>
<td>8/22 (36.4)</td>
<td>0/22 (0)</td>
</tr>
<tr>
<td>False-positive rate</td>
<td>1/19 (5.3)</td>
<td>1/21 (4.8)</td>
<td>3/21 (14.3)</td>
</tr>
</tbody>
</table>

Note:—PPV indicates positive predictive value; NPV, negative predictive value.
ent depending on the physician or institution, including US follow-ups, repeat US-FNAs, US-CNBs, and thyroid surgery.7-16 For the management of Bethesda class III nodules, 1 cytopathologist group recommended a repeat FNA at an appropriate interval.5 However, Lee et al7 insisted on limiting the use of repeat US-FNAs because a discrepancy might be unavoidable in the cytologic interpretation of the nodules classified as benign or as indeterminate aspirates because of overlapping cytologic criteria. There was limited use of repeat US-FNA for Bethesda class III nodules in our study (8/51, 15.7%), and we consequently believe that the selection bias based on US diagnosis by using the present classification system can be helpful for managing Bethesda class III nodules. In our study, US-CNB of Bethesda class III nodules was performed in 12 cases, and the histologic results were helpful in their management. In particular, Park et al16 suggested that US-CNB can be a better complementary tool for evaluating thyroid nodules with indeterminate cytology in the initial FNA compared with repeat US-FNA. We concur with prior studies that suggest US-CNB may become an alternative to repeat US-FNA or surgery for Bethesda class III nodules.16-18

Several studies have demonstrated that thyroid US is a feasible method for predicting malignancy in Bethesda class III nodules.10,11 Based on a retrospective evaluation of US images, Yoon et al10 emphasized that an irregular margin, microcalcifications, and taller-than-wide shape showed a significant correlation with malignancy. The results of a prospective study, which did not include an US classification scheme, led Mendez et al15 to suggest that an irregular margin, taller-than-wide shape, hypoechoogenicity, and microcalcifications were significantly associated with malignancy. We attempted to determine the extent to which US is diagnostic with a classification system that identifies Bethesda class III nodules as malignant or benign, though we did not focus on individual US features of Bethesda class III nodules. Our study found a relatively high accuracy for US diagnosis of Bethesda class III nodules when patients with borderline US diagnosis were either excluded or reclassified as malignant. The results of this study showed that our US classification scheme did help in predicting malignancy and determining the therapeutic plan for Bethesda class III nodules.

High-resolution thyroid US is regarded as the most useful diagnostic tool for evaluating nodular thyroid disease. Many studies have reported the US features of nodular thyroid disease, and several benign and malignant features have been generally accepted.19-23 The present criteria for the US features of solid thyroid nodules were determined on the basis of these studies. However, hypoechoogenicity, macrocalcifications, and centrally predominant vascularity were considered to be borderline US features in this study, which suggests that there is limited evidence that these features predict malignancy; hence their usefulness is a matter of debate. In particular, 9 surgical borderline nodules showed a high malignancy rate (8/9, 88.9%), which significantly influenced the diagnostic indices of thyroid US according to their reclassification into the benign or malignant category. Of the 9 surgical borderline nodules, 6 showed an eggshell calcification with interrupted eggshell or a thick hypoechoic outer rim on a thyroid US, which corresponds to interrupted calcification or a thick hypoechoic outer rim predicting malignancy in eggshell-calciﬁed nodules.24,25 Nevertheless, large-scale studies are needed to prove statistically that borderline US features predict the risk of malignancy.

In the present study, we made US diagnoses of thyroid nodules in accordance with a 5- and 4-category scheme for solid nodules and PCTNs, respectively. US classification schemes for thyroid nodules are diverse, but most reports include only 2 or 3 categories.10,23,26-28 Yoon et al10 emphasized the usefulness of thyroid US for predicting malignancy in Bethesda class III nodules, though they used only 2 categories (probably benign and suspicious) in their retrospective study of thyroid images. One of the authors (D.W.K.) recently reported that prospective studies using 5 US categories for solid thyroid nodules and 4 US categories for PCTNs showed high diagnostic efficacy.30,31 The results of these studies showed that our US classification schemes for solid nodules and PCTNs are useful because of their accuracy and the relative ease with which nodules can be assigned to particular categories.

There are several limitations to the present study. First, the sample size was relatively small. Second, a high malignancy rate of Bethesda class III nodules (46.5%, 20/43) was determined in our study, though the risk-of-malignancy reporting should be from 5% to 15%.6 Third, there were 16 nonsurgical Bethesda class III nodules; of these, 8 were histologically diagnosed and did not show suspicious US features on follow-up US during 12 months. The other nodules (n = 8) were not followed up by thyroid US, repeat US-FNA, US-CNB, or surgery; this lack of follow-up may represent a bias. Fourth, the differences in experience levels of the 3 cytopathologists in interpreting FNA slides (approximately 8, 9, and 15 years, respectively) might have resulted in variable cytologic diagnoses for individual cases; however, we did not evaluate the interobserver variations of the 3 cytopathologists in this study. Finally, only 1 radiologist performed the real-time thyroid US and made US diagnoses in all cases.

Conclusions

US diagnosis by using the present US classification system can be helpful for managing Bethesda class III nodules. Therefore, our US-based recommendations for Bethesda III nodules in the initial US-FNA are as follows: 1) when the US diagnosis for a thyroid nodule is benign or probably benign, repeat US-FNA or CNB may be considered; 2) when the US diagnosis for a thyroid nodule is borderline, repeat US-FNA or CNB should be considered; and 3) when the US diagnosis for a thyroid nodule is possibly malignant or malignant, repeat US-FNA may be unnecessary and thyroid surgery should be considered.

References

5. Layfield I, Cochran-Priollet B, LaVolli V, et al. Post thyroid FNA testing and


