Test Performances of Three Diagnostic Procedures in Evaluating Thyroid Nodules: Physical Examination, Ultrasonography and Fine Needle Aspiration Cytology


Departments of Endocrine Surgery, *Radiology, and **Surgical Pathology, Tokyo Women's Medical College, Tokyo 162, and ***Institute of Statistical Mathematics, Tokyo 106, Japan

Abstract. A retrospective study was performed to determine the reliability of physical examination (PE), ultrasonography (US) and fine-needle aspiration cytology (FNA) in the evaluation of thyroid nodules. Preoperative diagnoses of 252 euthyroid patients comprised 126 with benign lesions, 114 with papillary carcinoma and 12 with follicular carcinoma made by PE, US and FNA were reviewed. The specificity of PE, US and FNA for malignancy was 98%, 90% and 98%, respectively. The sensitivity of PE, US and FNA for malignancy was 63%, 78% and 80%. The sensitivity for papillary carcinoma of the three procedures was 68%, 83%, 88%, whereas that for follicular carcinoma was 25%, 25%, 8%, respectively. When all the test results were negative, the likelihood ratio favoring papillary carcinoma was 0.008 whereas that favoring follicular carcinoma was 0.6. The histological category of carcinoma should be considered when evaluating diagnostic procedures for thyroid nodules. No negative test result is conclusive for ruling out the possibility of follicular carcinoma.

Key words: Thyroid nodules, Sensitivity and specificity of diagnostic tests, Physical examination, Ultrasonography, Fine needle aspiration cytology.


THYROID NODULE is a common disorder throughout the world and proper management remains controversial. Because the prevalence of benign and malignant nodules of the thyroid and their clinical manifestations appear to be dissimilar in different areas according to the iodide intake, the problems which physicians have to face may differ from country to country. Furthermore, some uncertainty associated with the diagnostic procedures makes management difficult.

We have performed physical examination (PE), ultrasonography (US) and fine needle aspiration cytology (FNA) on patients with thyroid nodules as routine investigations. Although there are many authors reporting the sensitivity and specificity of FNA [1–3], corresponding figures for PE and US have rarely been published [2, 4]. This retrospective study was designed to disclose test performances and diagnostic implications of those three procedures carried out at our institution. As we have no test that always provides a perfect answer, a combination of the tests results seemed to be useful in the decision making.

Received: November 1, 1993
Accepted: January 21, 1994
Correspondence to: Dr. Takahiro OKAMOTO, Department of Endocrine Surgery, Tokyo Women's Medical College, 8–1 Kawada-cho, Shinjuku-ku, Tokyo 162, Japan
*This study was presented at the Eurosurgery 93 meeting in London, UK, September 15–17, 1993.
Materials and Methods

Patients

During the 5-year-period from 1987 through 1991, 1853 patients visited our institution because of thyroid nodules. The prevalence of malignant and benign tumors and benign inflammatory lesions was 24%, 73% and 3%, respectively (Table 1). Among them, we reviewed the cases of 252 euthyroid patients who had had all three diagnostic procedures preoperatively and underwent surgery for thyroid nodules. The patients consisted of 126 with benign tumors (adenoma or adenomatous goiter), 114 with papillary carcinoma and 12 with follicular carcinoma. None of the 126 patients with malignant neoplasm had evidence of metastasis to regional lymph nodes or a distant area at the time of diagnosis.

Diagnostic procedures

After the interview and PE, patients underwent US and FNA. US was performed with a 7.5-MHz liner probe. Smear specimens obtained by FNA were fixed in 95% alcohol and stained by a modified Papanicolaou technique. Test results from each procedure (PE, US, and FNA) were divided into three categories: 1) positive (malignant), 2) indeterminate (suspicious) and 3) negative (benign). Cases with unsatisfactory material for diagnosis by FNA were not included.

Evaluation

We checked the preoperative diagnoses made by PE, US and FNA in the patients’ medical records and calculated the sensitivity and specificity of each diagnostic procedure. Sensitivity was defined as True Positives / (True Positives + False Negatives) and specificity was defined as True Negatives / (True Negatives + False Positives). The likelihood ratios (the odds ratio form of Bayes’ Rule) associated with combinations of the test results were also determined. The relevant equations are [5]:

\[
\text{Likelihood ratio for positive test result} = \frac{\text{Sensitivity}}{1 - \text{Specificity}},
\]
\[
\text{Likelihood ratio for negative test result} = \frac{1 - \text{Sensitivity}}{\text{Specificity}},
\]
\[
\text{Post-test odds} = \text{Pre-test odds} \times \text{likelihood ratio}.
\]

The joint likelihood ratio for a combination of test results is the product of the separate ratios.

Statistical analysis

Fisher’s exact test was used to compare the sensitivity of each procedure according to the histological type of thyroid carcinoma.

Results

Overall test performances

The sensitivity and specificity of PE, US and FNA for diagnosing thyroid carcinoma are listed in Table 2. Each test performance was measured with and without the indeterminate results regarded as positive.

Sensitivity according to histological type of thyroid carcinoma

The sensitivity of each procedure in making the diagnosis of papillary carcinoma was higher than that for follicular carcinoma (PE, \( P=0.008 \); US, \( P=0.042 \); FNA, \( P=0.061 \); Table 3). Of 14 cases of papillary carcinoma which were false negative by
FNA, nine were positive by US.

Likelihood ratios for combinations of the test results

The joint likelihood ratios for combinations of the test results were calculated from the test performances, with indeterminate results regarded as negative, and are shown in Table 4. When all the test results were negative, the joint likelihood ratio favoring papillary carcinoma was 0.008, whereas that favoring follicular carcinoma was 0.6. If we adopt the prevalence of each thyroid carcinoma as the pre-test probability from Table 1, the post-test probability can be determined by the joint likelihood ratio. Pre-test and post-test probability percentages for thyroid cancer were 21.2% and 0.2% for papillary carcinoma, and 1.5% and 0.9% for follicular carcinoma, respectively, when all the tests were negative.

Discussion

The prevalence of thyroid nodules among the general population is 2.9%–10% in Western countries [6,7] whereas it is 0.8%–1.7% in Japan [8,9]. The frequency of malignancy of all the nodules is 3%–7.6% in Western countries [1, 6, 10–12] and 14.2%–24.3% in Japan, including our own series. Thus in the iodide-rich area Japan, thyroid nodules are less frequent, but the prevalence of cancer among thyroid nodules is much higher than in Western countries. In addition, there are differences in the frequencies of histological types of

| Table 2. Overall test performances of physical examination, ultrasonography and fine needle aspiration cytology for making a diagnosis of thyroid carcinoma, with or without regarding the “indeterminate” category as a positive test result |
|-----------------|-----------------|
|                  | Sensitivity %   | Specificity % |
| Physical Examination excluding “indeterminate” | 63 (80/126) | 98 (123/126) |
| including “indeterminate” | 70 (88/126) | 91 (115/126) |
| Ultrasonography excluding “indeterminate” | 78 (98/126) | 90 (113/126) |
| including “indeterminate” | 82 (103/126) | 71 (89/126) |
| Fine Needle Aspiration Cytology excluding “indeterminate” | 80 (101/126) | 98 (124/126) |
| including “indeterminate” | 87 (109/126) | 95 (120/126) |

| Table 3. Sensitivity of physical examination, ultrasonography and fine needle aspiration cytology for making a diagnosis of papillary and follicular thyroid carcinoma, with or without regarding the “indeterminate” category as a positive test result |
|-----------------|-----------------|-----------------|
|                  | Sensitivity for papillary carcinoma % | Sensitivity for follicular carcinoma % |
| Physical Examination excluding “indeterminate” | 68 (77/114) | 25 (3/12) |
| including “indeterminate” | 73 (83/114) | 42 (5/12) |
| Ultrasonography excluding “indeterminate” | 83 (95/114) | 25 (3/12) |
| including “indeterminate” | 87 (99/114) | 33 (4/12) |
| Fine Needle Aspiration Cytology excluding “indeterminate” | 88 (100/114) | 8 (1/12) |
| including “indeterminate” | 93 (106/114) | 25 (3/12) |
thyroid cancer; namely the prevalence of follicular carcinoma is relatively high in European iodide-deficient areas [3, 13, 14] but as high as 87% of thyroid malignancies are papillary carcinomas in Japan. Characteristics which favor the suspicion of malignancy on PE are a hard nodule, irregular shape, irregular surface and fixation to adjacent structures. Though some authors claim that PE is a poor predictor of malignancy [10, 14], these findings are specific and relatively sensitive in making a diagnosis of papillary carcinoma.

Many investigators believe that US has no role as a primary diagnostic test for thyroid nodules because it cannot differentiate between benign and malignant tumors [10, 14–16]. But most papillary carcinomas can easily be diagnosed by US, which frequently shows an irregularly shaped, hypoechoic area with an unclear verge in the thyroid. The margin of the sternothyroid muscle overlying the nodule is sometimes blurred due to adhesion or invasion. There are often tiny calcifications within the nodule, representing psammoma bodies. It is very important to describe the ultrasonographic appearance of the nodule in detail, instead of simply classifying it as a solid and/or a cystic lesion, to make a diagnosis of malignancy. This approach enables us to reach a sensitivity of 83% and a specificity of 90% in diagnosing papillary carcinoma. Furthermore US has several additional roles to play in the management of patients with thyroid nodules [12]: to ascertain that there is no lesion in the other lobe which will be left behind at surgery, and to observe changes in the size and structure of the nodules during a conservative follow-up period. Therefore US is recommendable as a useful method to use in the evaluation of thyroid nodules.

FNA provides the most reliable information for the evaluation of thyroid nodules and it plays a decisive role in their management [10, 15]. Several authors reported sensitivity from 73% to 98.5% and specificity from 74% to 100% [1–3, 6]. Because of its high specificity, the introduction of FNA has reduced the number of operations and increased the yield of malignancy at thyroid surgery [11, 17].

The frequency of indeterminate results on FNA ranged from 13% to 30% and the prevalence of malignancy in this category proved to be 10% to 50% [1, 3, 6, 14, 15]. As the major causes of an indeterminate result are the difficulty in differentiating follicular adenoma from follicular carcinoma and the relatively high incidence of malignancy in this category, some authors advocate surgical excision of all such lesions [14] or follicular neoplasms on FNA [16, 18, 19].

A negative report of the tests should not be conclusive for ruling out follicular carcinoma. In fact, probabilistic change in doubting follicular carcinoma when all the tests were negative was very small in contrast to the degree of papillary carcinoma.

Test performances in all the diagnostic procedures are dependent upon the examiner’s experience and skill. They also can be affected by the prevalence of follicular carcinoma in the study.
population. Therefore, the figures for the sensitivity and specificity of those tests reported in the literature should be interpreted carefully. The optimal method for evaluating thyroid nodules in each institution should be determined by weighing the available clinical expertise against the cost of the diagnostic procedures.

References


