Measuring Serum Thyroglobulin in Patients with Follicular Thyroid Nodule: Its Diagnostic Implications

TAKAHIRO OKAMOTO, MASAKO KANBE, MASATOSHI IIHARA, KIYOMI YAMAZAKI, JOJI OKAMOTO, TOMOYUKI YAMASHITA, YUKIO ITO, MAKIO KAWAKAMI*, AND TAKAO OBARA

Departments of Endocrine Surgery and *Surgical Pathology, Tokyo Women's Medical College, Tokyo 162, Japan

Abstract. To determine the diagnostic implications of measuring the serum thyroglobulin level in patients with a solitary follicular thyroid tumor, a retrospective study was conducted on 122 consecutive patients with a solitary follicular thyroid nodule who underwent thyroidectomy. Data for eight variables were collected: the serum thyroglobulin (Tg) level (µg/l), age, maximum diameter of the nodule, gender, histopathologic type, presence or absence of metastases, macroscopic characteristics of the cut surface of the resected tumor, and smoking habit. Multiple regression analyses were used to investigate the relationships between the serum Tg level and the seven other variables. The diagnostic value of serum Tg was examined by means of receiver operating characteristic (ROC) curves. There were significant correlations between the serum Tg level and the maximum diameter of the nodule, the macroscopic characteristics, and the smoking habit. The sensitivity and specificity of the serum Tg level with a cut-off value of 1,000 µg/l were 57% and 86%, respectively. The likelihood ratio favouring follicular carcinoma associated with the serum Tg > 1,000 µg/l was 4.41. Measuring the level of serum Tg may be useful in discriminating follicular carcinoma from follicular adenoma, but since there may be some biases in this retrospective study, the results are less definitive. Further research activities are mandatory to obtain valid evidence.

Key words: Thyroid nodule, Serum thyroglobulin, Follicular thyroid tumor, Sensitivity and specificity, Likelihood ratio

PROPER management of thyroid nodules still remains controversial because of some uncertainties in the diagnostic process [1, 2]. One of the problems results from difficulty in differentiating between follicular carcinoma and follicular adenoma. For example, three diagnostic tests are widely used in evaluating thyroid nodules: palpation, ultrasonography, and fine needle aspiration cytology. These tests have demonstrated good success in making the diagnosis of papillary carcinoma, and we can reasonably guarantee reduced probability of papillary carcinoma if all the tests are negative; however, this is not the case for follicular carcinoma [3]. Many investigators have recommended operating on patients who are diagnosed as having follicular neoplasm based on cytology [4–7]. Nevertheless, some patients or attending physicians resist accepting surgical treatment because the risk of a life threatening event caused by the follicular tumor may be low, though the risk has not been sufficiently evaluated.

One of the possible tests for the differential diagnosis is the determination of serum thyroglobulin (Tg). Although many researchers have concluded that the measurement is not helpful in making the diagnosis for thyroid nodules [8–10], none of them has addressed the problem specifically focusing on the differential diagnosis.
of follicular thyroid tumors. In addition, as the
determination of serum Tg is affected by various
conditions [8], it may not be easy to carry out a
proper statistical analysis.

Our research questions in the present study are
twofold: in patients with solitary, follicular thyroid
nodule, (1) Is there any association between the
serum Tg levels and clinical information which
includes age, maximum diameter of the nodule,
gender, histopathologic type of the tumor, status
of metastases, macroscopic characteristics of the
cut surface of the resected tumor, and smoking
habit? (2) What is the test performance of serum
Tg levels in discriminating follicular carcinoma
from follicular adenoma? Physicians are naturally
expected to obtain reliable and valid answers to
such clinical questions for patients, but it is
impossible to do so unless we have a valid research
design. On the other hand, we very often confront
the fact that, in reality, we have only retrospective
data; in fact many studies have been carried out
on data collected retrospectively from consecutive
patients. Evidence from such studies may be less
definitive because of the designs. In the present
paper, we are going to try to answer the questions
by using our retrospective data, and will also
discuss methodologic issues in the study.

Patients and Methods

Study design and sampling method

A total of 1853 patients visited our institution
because of their thyroid nodules during the 5-year
period from 1987 through 1991. The study was
retrospective in design with a sampling frame of
773 patients who underwent thyroid surgery
among them. The main indications for surgical
treatment were clinical evidence or suspicion of
malignancy or symptoms caused by the nodules.
The distribution of histopathologic diagnoses in the
773 patients was as follows: malignant tumor, 451
(papillary carcinoma, 392; follicular carcinoma, 28;
other malignancies, 31); and benign tumor, 322 [3].
The patients’ records were reviewed and evaluated
for eligibility; therefore, the sampling was non-
random. The inclusion criteria for the study were
solitary thyroid nodule, and either follicular
adenoma or follicular carcinoma on histopathology,
and euthyroidism. The exclusion criteria were
multiple nodules on palpation and/or
ultrasonography, or concomitant thyroid diseases,
or previous history of thyroid diseases, or positive
thyroglobulin antibody, or medication with thyroid
hormone and/or oestrogen, or pregnancy, or
concomitant liver dysfunction. These inclusion and
exclusion criteria were applied to each patient
based on pre-operative information. After the
evaluation, 122 patients were eligible.

Variables

Data were extracted from the patients’ records
for eight variables (capital letters in the following
parentheses denote each corresponding variable
name which will be used in the analyses); the three
numerical variables were the serum Tg level (μg/
l, TG), age (AGE), maximum diameter of the nodule
(cm, DIAMETER), and the five categorical variables
were the gender (GENDER), pathology (follicular
adenoma vs. follicular cancer, PATHO), presence
or absence of metastases (META), macroscopic
characteristics of the cut surface of the removed
tumor determined by a surgeon (“A type” versus
“B type”, MACRO) and smoking habit (non-smoker
vs. smoker, SMOKING).

Methods of the measurements

Serum Tg was measured by double antibody
radioimmunoassay with a commercial kit
(Thyroglobulin ‘Eiken’, Eiken, Japan). The diameter
of the thyroid nodule was evaluated by radiologists
using ultrasonography with a 7.5 MHz linear probe.
The maximum diameter for each patient was used
for the analysis.

Macroscopic characteristics of a cut surface of a
removed tumor were classified as either “A type”
or “B type” by a surgeon just after surgery. The
“A type” denotes a homogeneous, expansive and
grayish white cut surface, and the “B type”
indicates a less homogeneous, not expansive, tan
cut surface. It was assumed that microfollicular
tumors, particularly follicular carcinomas, were
likely to show A type features rather than B type
ones on macroscopic assessment. Pathologic
diagnosis was made on surgical materials which
were processed by routine histologic techniques,
and embedded in paraffin. Each section was
stained with hematoxylin-eosin, and subsequently, an Elastica-van Gieson stain was used to visualize the vascular wall, if needed. The diagnostic criteria for follicular carcinoma are follicular neoplasm with capsular invasion and/or vascular invasion.

**Reliability and validity of the measurements**

The reported accuracy of the radioimmunoassay kit for serum Tg was within ± 15% of the indicated value. The precision of the kit was described as the coefficient of variation which was less than 15%. The usual assay range of this method is 10–320 µg/l, and the expected normal value is equal to or less than 35 µg/l. The reproducibility of measuring the maximum diameter of a thyroid nodule by ultrasonography remains to be established. The reliability and validity of the pathologic diagnosis and macroscopic evaluation of the cut surface of the surgical materials have not been assessed.

**Statistical analyses**

Statistical analyses were carried out with the computer software package SYSTAT version 5.2 for the Macintosh [11]. Multiple regression analyses were used for the first research question. The dependent variable was the serum Tg level and covariates were the others. As for the selection of the covariates for the regression models, we did not employ any automatic selection procedures because the study was explorative. The probability of type I error for the test was considered significant at the level of 0.05. The appropriateness of using a multiple regression model was evaluated by means of analysis of variance. Regression diagnoses, which check assumptions of the multiple regression, were performed with residual analyses. In order to detect outliers among the data, leverage was also determined for each patient. Values of leverage less than 0.2 appear safe, between 0.2 and 0.5 risky, and above 0.5 to be avoided [11]. For the second research question, receiver operating characteristic (ROC) curves were generated to examine the test performance. The ROC curve is a plot of pairs of the true-positive rate (sensitivity) and the false-positive rate (1 – specificity) that correspond to each possible cut-off value for the diagnostic test result. Likelihood ratios for several cut-off points were also calculated. The likelihood ratio is an alternative way of describing the performance of a diagnostic test. There is one likelihood ratio for a positive test and another for a negative test.

The relevant equations are:

- Likelihood ratio for positive test result = sensitivity / (1 – specificity),
- Likelihood ratio for negative test result = (1 – sensitivity) / specificity.

These ratios express how many times more (or less) likely a test result is to be found in diseased, as compared to non-diseased, people. In other words, the likelihood ratios indicate by how much a given diagnostic result will raise or lower the pretest probability of the target disease [12, 13].

Since patients with follicular carcinoma who already have metastases at the time of diagnosis sometimes have a very high level of serum Tg, it may be unrealistic to include those patients in the sample when we evaluate the implications of the test in the differential diagnosis. In order to assess the effect of those patients on the results, we also examined the test performance excluding the patients with metastases from the sample.

**Results**

**Descriptive analyses of the data**

The distribution of serum Tg was highly skewed, whereas those of age and diameter of the nodule looked roughly symmetric. The data for serum Tg were transformed with a logarithmic function for a dependent variable of regression models (LOGTg). About 70% of the patients were female. The two categories of macroscopic findings (A type and B type) were almost evenly distributed. There are a few patients with follicular cancer, and a small number of smokers in the sample (Table 1).

All the follicular carcinomas (14/14) and 44% (48/108) of the adenomas were classified as A type, and 56% (60/108) of the adenomas were judged as B type. Proportions of smokers among the patients were 33% (12/36) in men and 7% (6/86) in women.

**Multiple regression analyses**

To examine the relationships between serum Tg levels and the other variables, we put all 7
independent variables into the regression model. Among them, DIAMETER, MACRO and SMOKING were significant variables (Table 2). None of the independent variables were confounders, and there was no interaction effect among the three significant variables (Data are not shown, but available from the first author). The squared multiple correlation coefficient of the model was 0.296, and the regression diagnoses for the model showed no major violations of the assumptions for multiple regression analysis. The largest leverage was 0.128.

Test performance of serum Tg in the differential diagnosis of follicular tumors

The serum Tg level adjusted for the size of the nodule may be useful in discriminating follicular carcinoma from follicular adenoma because the multivariable analyses suggested that the size of the tumor may be one of the determinants of the serum Tg level. We, therefore, examined the test performances of two indices in discriminating follicular carcinoma from follicular adenoma; serum Tg (TG) and serum Tg divided by DIAMETER (TG/D). ROC curves of these indices showed little difference in their diagnostic abilities (Fig. 1). From a practical point of view, the serum Tg level of 1,000 μg/l seems to be a reasonable cut-off value in the differential diagnosis with a sensitivity of 57% (8/14) and specificity of 87% (94/108). The likelihood ratio favouring follicular carcinoma with the determination of serum Tg>1,000 μg/l, was 4.41 (Table 3). When we exclude patients with follicular carcinoma who already had metastases at the time of diagnosis, sensitivity and the likelihood ratio associated with the serum Tg>1,000 μg/l, were 55% (6/11) and 4.21, respectively.

Discussion

There have been many clinical studies on serum Tg in patients with thyroid diseases. Some of them examined factors which affect the level of serum Tg [14–17], and others investigated the diagnostic value of the test in the management of thyroid nodules [18–22]. Experts reported that the measurement of serum Tg is of no value in the differential diagnosis of thyroid tumor [8–10], but we should be critical of their findings in two respects; an analysis and a study population. First, it may be inappropriate to compare the mean values of the crude measurements of serum Tg among specified groups. As we can see in our sample, the crude determinations of serum Tg may not be normally distributed. In addition, there must be other factors affecting the determination. A simple comparison may, therefore, be
misleading. Second, as for the management of thyroid nodule, we rarely need the measurement when we make a diagnosis of papillary, medullary or anaplastic carcinoma because other diagnostic information, such as the clinical history, physical examination, fine needle aspiration cytology, or measurement of plasma calcitonin or of serum carcinoembryonic antigen (CEA), are very helpful. We need a further test when we want to discrim-

### Table 2. Multiple regression analysis with seven covariates

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std Error</th>
<th>Std Coef</th>
<th>t</th>
<th>P (2-tail)</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>3.300</td>
<td>0.650</td>
<td>0.000</td>
<td>5.080</td>
<td>0.000</td>
</tr>
<tr>
<td>DIAMETER</td>
<td>0.260</td>
<td>0.082</td>
<td>0.259</td>
<td>3.152</td>
<td>0.002</td>
</tr>
<tr>
<td>MACRO</td>
<td>1.143</td>
<td>0.277</td>
<td>0.352</td>
<td>4.132</td>
<td>0.000</td>
</tr>
<tr>
<td>SMOKING</td>
<td>−1.165</td>
<td>0.383</td>
<td>−0.255</td>
<td>−3.040</td>
<td>0.003</td>
</tr>
<tr>
<td>AGE</td>
<td>0.010</td>
<td>0.010</td>
<td>0.087</td>
<td>1.052</td>
<td>0.295</td>
</tr>
<tr>
<td>GENDER</td>
<td>−0.299</td>
<td>0.311</td>
<td>−0.084</td>
<td>−0.963</td>
<td>0.338</td>
</tr>
<tr>
<td>PATHO</td>
<td>0.395</td>
<td>0.481</td>
<td>0.078</td>
<td>0.822</td>
<td>0.413</td>
</tr>
<tr>
<td>META</td>
<td>0.502</td>
<td>0.935</td>
<td>0.048</td>
<td>0.537</td>
<td>0.592</td>
</tr>
</tbody>
</table>

DIAMETER, maximum diameter of thyroid nodule; FOLLICLE, macroscopic finding of the cut surface of the removed tumor; SMOKING, smoking habit; AGE, age; GENDER, gender; PATHO, histopathology; META, presence or absence of metastases. Std Error, standard error; Std Coef, stanardized coefficient. Squared multiple R=0.296.

### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F-ratio</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>95.290</td>
<td>7</td>
<td>13.613</td>
<td>6.860</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual</td>
<td>226.212</td>
<td>114</td>
<td>1.984</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SS, sum of squares; df, degree of freedom; MS, mean square.

### Table 3. Likelihood ratios

<table>
<thead>
<tr>
<th>Serum Tg level (μg/L)</th>
<th>Follicular carcinoma</th>
<th>Follicular adenoma</th>
<th>Likelihood ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tg &gt; 1,000</td>
<td>8</td>
<td>14</td>
<td>4.41</td>
</tr>
<tr>
<td>1,000 ≥ Tg &gt; 35</td>
<td>5</td>
<td>80</td>
<td>0.48</td>
</tr>
<tr>
<td>35 ≥ Tg</td>
<td>1</td>
<td>14</td>
<td>0.55</td>
</tr>
<tr>
<td>total</td>
<td>14</td>
<td>108</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. ROC (receiver operating characteristics) curves showing test results for the serum thyroglobulin level (TG, solid line) and serum thyroglobulin divided by tumor diameter (TG/D, broken line) in discriminating follicular carcinoma from follicular adenoma of the thyroid.
inate follicular carcinoma from benign tumors. A study of the measurement, specifically focusing on the study population with follicular thyroid tumors, will be more clinically relevant than those on a mixed population with various thyroid disorders.

The multiple regression analysis showed that there were correlations between the serum Tg level and the size of the nodule, the macroscopic characteristics and the smoking habit. This suggests that patients whose nodules are large, or A type on macroscopic assessment, or who are non-smokers may have higher determinations of serum Tg than those who do not have these characteristics, although the negative effect of smoking was an opposite to findings in another report [17]. These results have two clinical implications. First, if we want to know the diagnostic value of the serum Tg concentration in discriminating follicular carcinoma from follicular adenoma, the index involving both the serum Tg level and the size of the nodule may be more useful than the crude serum Tg level (for example, serum Tg/diameter), though this was not proved in the ROC analysis in this study. Second, the measurement may help to detect tumors with macroscopic features of A type which are associated with follicular carcinoma.

The sensitivity (57%) of crude serum Tg level with a cut-off value of 1,000 µg/l should be considered superior to figures obtained with other diagnostic procedures such as palpation (25%), ultrasonography (25%) and fine needle aspiration cytology (8%) in the differential diagnosis of follicular thyroid tumors [3]. Measuring serum Tg may, therefore, be useful in the management.

We should, however, be aware that this study has several limitations because the design was retrospective. First, some of the patients might have been referred to us and indicated for surgery because their nodules were large and their determinations of serum Tg were high. This process may produce a spurious correlation between DIAMETER and LOGTG in the regression analysis; hence, the finding from the study may be artificial. Second, because patients with a high serum Tg level were more likely to be indicated for surgery, we might not have included some patients with follicular thyroid nodules whose determinations of serum Tg were within the normal range in the study. Therefore, any cut-off value for the serum Tg level to discriminate follicular carcinoma from adenoma will show spuriously high sensitivity and low specificity. Third, the determination of serum Tg can be affected by other, uncontrolled factors (e.g., the value may be elevated right after the palpation or the aspiration for cytology). This may be the reason why the contribution of the regression model to the variation in the Tg was small. Fourth, the measurements of the variables may not be reliable. Lack of studies examining reliability and validity of those measurements make the study less credible in terms of its conclusions.

Retrospective designs have been used for some clinical research activities, but we should note that the designs may have limitations, and therefore that conclusions drawn from such studies may be least definitive. We should rather use the data to construct hypotheses in an explorative way with appropriate statistical analyses. We must then test such hypotheses in studies with more appropriate designs to avoid systematic or random errors. A well-designed study on patients with thyroid nodule, specifically focusing on follicular tumors, is therefore mandatory to provide reliable and valid answers.

Acknowledgements

We are grateful to Mrs. Lois Wright for her support in preparing this manuscript. A part of this work was supported by Sasakawa Health Science Foundation.

References

and fine needle aspiration cytology. Endocr J 41: 243–247 [Corrigendum, Endocr J 42(4)].


