Ability of Two New Thyrotropin (TSH) Assays to Separate Hyperthyroid Patients from Euthyroid Patients with Low TSH

Eeva Taimela,1,4 Riitta Tähtelä,2 Pertti Koskinen,1 Pirjo Nuutila,3 Jari Forsström,3 Simo Taimela,1 Sirkka-Liisa Karonen,2 Matti Välimäki,2,4 and Kerttu Irjala1

We evaluated the ability of new thyrotropin (TSH) assays to separate hyperthyroid (n = 50) patients from clinically euthyroid subjects with low TSH values (nodular goiter, n = 20, and nonthyroidal illness, n = 22). Only patients whose serum TSH was <0.1 mIU/L by immunoradiometric assay were included. We used a new immunofluorometric method based on time-resolved fluorescence (TR-IFMA) and a new immunocholinuminometric assay (ICMA) to measure TSH in serum. Although the differences between the hyperthyroid patients and the euthyroid patients differed from each other by both methods (P = 0.0012 for TR-IFMA and P < 0.0001 for ICMA), there was no cutoff point that could definitely separate the groups. Thus, it is not possible to draw any definite conclusions on whether a patient is hyperthyroid or not, solely on the basis of TSH concentration measured with these new TSH assays.

Indexing Terms: thyroid function/immunofluorometric assay/immunocholinuminometric assay/goiter

Since their introduction in the 1990s, new thyrotropin (TSH) assays have been expected to facilitate diagnosis between hyperthyroid patients and euthyroid patients with low TSH values (<0.1 mIU/L) (1,2). Subnormal TSH values associated with euthyroidism are seen in patients with multinodular goiter, patients recovered from Graves disease (3), patients with nonthyroidal illness (NTI) (4–6), and patients taking medication (7). The variety of these groups confuses the ability of TSH immunoradiometric assays (IRMAs) to separate hyperthyroid patients from these euthyroid individuals.

New TSH assays with 10-fold greater assay functional sensitivity than TSH-IRMA are believed to aid the diagnosis of hyperthyroidism in the above groups (2,8). Recently, new TSH assays based on chemiluminescence and time-resolved fluorescence have been introduced (9–13). We studied the ability of a time-resolved immunofluorometric assay (TR-IFMA) and an immunocholinuminometric assay (ICMA) for TSH in serum to separate euthyroid individuals with low serum thyrotropin (<0.1 mIU/L) from hyperthyroid patients. We also studied the extent of agreement between these methods.

Materials and Methods

Subjects. The study population consisted of 92 patients at the University Central Hospital of Turku, Turku, Finland. The procedures followed were in accordance with the Helsinki Declaration of 1975, as revised in 1983. The serum specimens were preselected so that only those subjects who had serum TSH values <0.1 mIU/L by TSH-IRMA (Spectria®; Fannos Diagnostica, Turku, Finland) were included in the study. On the basis of clinical diagnosis, these patients formed two different groups: 50 hyperthyroid patients who had been diagnosed on the basis of characteristic clinical features and confirmed by laboratory tests, and 42 patients who were considered euthyroid on the basis of (a) the lack of typical clinical symptoms for hyperthyroidism, and (b) serum concentrations of free thyroxine (FT4) and triiodothyronine (FT3) between the confidence limits of the reference intervals (Table 1). The second group was further divided into two subgroups: 20 patients with nodular goiter and 22 patients with NTI. The clinical diagnosis of goiter was further confirmed by ultrasound and fine-needle biopsy in 16 of the 20 patients. The diseases of the NTI group concerned heart (n = 11), severe infection (4), lungs (2), kidneys (1), parkinsonism (1), psychosis (1), skin disorders (1), and fracture (1). No patient was taking medication that interfered with thyroid-function tests.

To determine reference intervals for TSH by ICMA, we measured serum TSH concentrations from 196 euthyroid subjects (70 men, 126 women; ages 22–64 years, mean 39.6, SD 9.1).

To assess the degree of agreement between TR-IFMA and ICMA methods for TSH, we analyzed 181 serum samples over a wide range of subnormal to normal values.

Sera had been stored at −20°C before analysis. TR-IFMA. DELFIA® hTSH Ultra (Wallac, Turku, Finland) is a solid-phase, two-site immunofluorometric assay involving three monoclonal antibodies directed against human TSH. The solid phase of the assay, i.e., microtitration wells, is coated with the capture antibody. Two other monoclonal antibodies directed against different specific antigenic sites on the β-subunit of hTSH are labeled with europium. Enhancement solu-
tion dissociates europium ions from the labeled antibody into solution, where the ions form highly fluorescent chelates with components of the enhancement solution. Fluorescence is measured in a time-resolved fluorometer. We recently reported the assay performance of this method (13). The standards are calibrated against the World Health Organization Second International Reference Preparation (WHO 2nd IRP) 80/558. The curve-fitting algorithm was linear interpollation at concentrations <0.1 mIU/L.

ICMA. This one-step ICMA is a modification of the BeriLux® hTSH method (modified by United Laboratories, Helsinki, Finland). The method uses two monoclonal antibodies to hTSH: The solid phase of the assay, i.e., tubes, is coated with the catching antibody. The other antibody is labeled with a luminogenic label, an acridinium amide. We followed the modified procedure, as follows: Add 200 μL of serum and 100 μL of tracer to the tubes. After 2 h of incubation, remove the unbound labeled antibody by decanting and washing the tubes four times. Measure the amount of the label in the sandwich complex. We used a Magic® Lite II Analyzer (Ciba Corning, East Walpole, MA), which automatically injects the reagents necessary to initiate the luminescence reaction, and quantitates the photon output. To calculate the TSH concentration, we used a master curve that is corrected in each assay by two-point calibration. The master curve is produced for each reagent lot by using hTSH standards from the Nichols Institute (San Juan Capistrano, CA) third-generation TSH-kit; these standards are calibrated against the WHO 2nd IRP 80/558.

Other assays. FT₃ was analyzed by DELFIA TR-FIA (Wallac) (13); the reference interval was 4.2–7.6 pmol/L. FT₄ was analyzed by Amerlex MAB® (Amersham International, Amersham, UK); the reference interval was 11–24 pmol/L. TSH was primarily analyzed by Spectria, with a reference interval of 0.4–4.5 mIU/L.

Statistics. Analytical sensitivity for ICMA (i.e., the concentration corresponding to the number of counts 3 SD above mean for the zero standard) and functional sensitivity (10% between-assay CV) were measured according to Ekins (14). To assess the degree of agreement between TR-IFMA and ICMA for TSH, we used the methods described by Bland and Altman (15).

The within-assay precision for ICMA was calculated from 10 to 16 replicates of serum specimens, and the between-assay precision was calculated from the means of 10 to 40 duplicate measurements of serum specimens with low, intermediate, and high concentrations. Statistical treatment of reference values was performed as recommended by the International Federation of Clinical Chemistry to determine the reference interval (Program Refval 3.2) (16). Nonparametric estimates were used for TSH.

Nonparametric Kruskal–Wallis analysis of variance was calculated by using proc npar1way of the SAS/STAT® (SAS Institute, Cary, NC) program library to compare TSH concentrations determined by TR-IFMA and ICMA between the three groups: overtly hyperthyroid, goiter, and NTI patients. Post hoc comparisons were calculated by Mann–Whitney U-test with an a priori P value of 0.017.

The areas under relative operating characteristic (ROC) curves were compared between the two methods, TR-IFMA and ICMA. The areas under ROC curves, standard errors (SE), and the average correlation were calculated according to Forström (17) on the basis of reports by Beck and Shultz (18) and Hanley and McNeil (19). The level of significance between the two methods was read from a z-score table. Calculation of diagnostic efficiency was based on the assumption that sensitivity and specificity were equally important in all clinical situations.

Results and Discussion

Assay Performance

Sensitivity. The analytical sensitivity of the TR-IFMA was 0.001 mIU/L (13); that of the ICMA was 0.002 mIU/L. The functional sensitivity of the TR-IFMA (10% between-assay CV) was 0.016 mIU/L (13); that of the ICMA was 0.011 mIU/L—in both cases better than in other new TSH assays (1, 11, 12), except for the in-house method previously reported (20). In the other studies, the accepted CV limit for functional sensitivity varied from 10% to 30%, so it is difficult to intercompare the sensitivities reported (21).

Precision. The between-assay precision profiles for each of the three TSH assays—IRMA, TR-IFMA, and ICMA—are presented in Fig. 1. Between-assay and within-assay CVs were <4% in the euthyroid range in the TR-IFMA and ICMA methods. The between-assay CVs by TR-IFMA at 0.144 and 12.27 mIU/L were 7.9% and 2.1%, respectively. The between-assay CVs by

---

Table 1. Description of the study patients and TSH measurements.*

<table>
<thead>
<tr>
<th>Group</th>
<th>n Ref</th>
<th>FT₃ (range)</th>
<th>FT₄ (range)</th>
<th>Age, years</th>
<th>TSH (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothyroid</td>
<td>50</td>
<td>49.6 (28–100)</td>
<td>25 (7.9–85)</td>
<td>48.1 (18–87)</td>
<td>&lt;0.016 (&lt;0.016–0.203)</td>
</tr>
<tr>
<td>Goiter</td>
<td>20</td>
<td>20.4 (15–23)</td>
<td>5.7 (3.4–7.7)</td>
<td>62 (44–79)</td>
<td>&lt;0.011 (&lt;0.011–0.164)</td>
</tr>
<tr>
<td>NTI</td>
<td>22</td>
<td>19.9 (13–25)</td>
<td>4.9 (&lt;2.7–7.8)</td>
<td>70.6 (38–87)</td>
<td>0.034 (&lt;0.016–0.105)</td>
</tr>
</tbody>
</table>

* The serum samples were preselected so that only patients with serum TSH <0.1 mIU/L were included.

† Male/female ratio is in parentheses.

‡ Functional sensitivity limit of the assay.
ICMA at 0.169 and 12.82 mIU/L were 7.8% and 4.2%, respectively. These precisions were well acceptable in clinical practice.

Reference interval. The nonparametric central 0.95 fraction of the reference distribution for euthyroid subjects by TR-IFMA was 0.53–4.7 mIU/L (n = 200). The 0.90 confidence intervals for the 0.025 and 0.975 fractiles were 0.48–0.64 and 4.0–6.6 mIU/L, respectively (13). The nonparametric central 0.95 fraction of the reference distribution for euthyroid subjects by ICMA was 0.44–4.7 mIU/L (n = 196). The 0.90 confidence intervals for the 0.025 and 0.975 fractiles were 0.40–0.63 and 3.9–6.1 mIU/L, respectively.

Comparison study. To assess the degree of agreement between TR-IFMA and ICMA, we measured 136 serum samples in a wide range around the normal range (Fig. 2). The mean difference of the two methods was 0.056 (SD 0.621) mIU/L, ICMA giving lower concentrations. We also measured 41 serum samples between TSH concentrations of 0.01 and 0.1 mIU/L (Fig. 3). At those concentrations, the mean difference of the two methods was 0.021 (SD 0.028) mIU/L, with the ICMA giving the lower TSH concentrations. The differences were normally distributed. The bias around the reference interval was minor, but became more marked at TSH concentrations <0.1 mIU/L.

The bias can be explained partly by antibody specificities differences. TR-IFMA uses 3 antibodies against the hTSH molecule: The capture antibody recognizes parts of both the α- and β-subunits at their connection site, and the two labeled antibodies are specific to different antigenic sites of the β-subunit. In contrast, ICMA uses two antibodies against the hTSH molecule: One is specific to the β-subunit, and the other recognizes parts of both the α- and β-subunits at their connection site. Differences in standardization of the two assays is an unlikely source of bias. Both assays are calibrated against the same international reference preparation (WHO 2nd IRP 80/558). Had the bias been due to different calibration of the two methods, it should have been proportionally consistent in all samples. However, no such systematic difference was found between the two methods, as is apparent in Figs. 2 and 3. Heterophile antibodies may interfere with the immune reactions in immunoassays, and neutral mouse immunoglobulin is included in the TR-IFMA assay buffer to neutralize the interference by such antibodies. Because the TSH values of concern in our study are all in the subnormal concentration range, whereas TSH concentrations are spuriously elevated with interfering heterophile antibodies (22), we do not believe that heterophile antibodies are involved in the bias seen. Considering the acceptable performance of the TR-IFMA in terms of linearity and recovery (13), perhaps matrix effects influence results in the TSH concentration range close to the functional sensitivity of the assay.
TSH Measurements in Serum from Hyper- and Euthyroid Subjects

The distribution of serum TSH concentrations in the three groups of patients displayed the bias between the TSH results of the two TSH methods (Figs. 4 and 5). The descriptions of the distributions for TSH are presented in Table 1. Analysis of variance revealed significant differences between results for the three study groups by both TR-IFMA and ICMA (P = 0.0012 and P < 0.0001, respectively). For both methods, significant differences were found between hyperthyroid patients and patients with goiter, and between hyperthyroid and NTI patients in the post hoc comparisons. Areas under ROC curves for TR-IFMA and ICMA were 0.723 (SE 0.0536) and 0.759 (SE 0.0511), respectively, which were not significantly different from each other (P = 0.60). In other words, both methods could separate hyperthyroid patients from euthyroid subjects with low TSH values, and their discriminatory power was equal according to the ROC analysis. The validity of a test, i.e., the fundamental discriminatory power of any test to classify samples into alternative groups, can be summarized by the area under the ROC curve (23). The area under the ROC curve for an ideal test is 1.0; for a test of no value, the area is 0.5. However, a selection based solely on the area value is not appropriate. If two ROC curves differ markedly in their overall slopes, they still may intersect at some decision point. Therefore, a test performance has to be translated into optimal diagnostic efficacy by selecting the operating point of maximal clinical yield (23). The selection of the cutoff value depends on the task, and there is not just one suitable method for calculating the optimal cutoff limits (17).

The optimal diagnostic efficacy, based on the assumption that sensitivity and specificity were equally important, was achieved at cutoff points of 0.018–0.020 and 0.011 mIU/L for TR-IFMA and ICMA, respectively. At the cutoff point of 0.018–0.020 mIU/L, the sensitivity and specificity of the TR-IFMA for hyperthyroidism were 78% and 64%, respectively. Correspondingly, at the cutoff point of 0.011 mIU/L, the sensitivity and specificity of the ICMA for diagnosing hyperthyroidism were 97% and 57%, respectively. On the other hand, had our main interest been to find all hyperthyroid patients (test sensitivity of 100%), the specificities would have been inferior, only 0–2%. Because half the euthyroid patients had a TSH concentration below the functional sensitivity limit of the ICMA, we could not determine the exact cutoff point for the ICMA.

One must bear in mind that these numbers are valid only in this preselected material of patients whose TSH concentration was <0.1 mIU/L by a second-generation TSH assay. Furthermore, the incidence of patients with a goiter or NTI whose serum TSH is <0.1 mIU/L is less than the incidence of hyperthyroidism. Only 0–4% of patients with NTI have serum TSH <0.1 mIU/L (5, 24–26). Because of the number of patients with goiter or NTI whose serum TSH concentrations are >0.1 mIU/L, in clinical practice the sensitivities and especially the specificities of these tests to diagnose hyperthyroidism will be better. Given the lack of a gold standard for diagnosing hyperthyroidism, we combined
four components: serum TSH, FT₃, FT₄, and clinical signs and symptoms. All of these variables supported that the patients in the hyperthyroid group were hyperthyroid. Consequently, the fact that the range of FT₃ concentrations in the hyperthyroid group extended down to near the normal reference interval did not exclude the possibility that a patient is overtly hyperthyroid.

This study further confirmed the recent experience with new TSH assays, indicating that most of the depressed TSH values in NTI patients were only marginally below the 0.1 mIU/L assay limit of TSH-IRMA methods and far above those observed with fully developed thyrotoxicosis (27). Spencer et al. (1) suggested that a serum TSH value <0.01 mIU/L was highly suspicious of hyperthyroidism, regardless of whether the patient was ambulatory or hospitalized. We conclude that most hyperthyroid patients have undetectable serum TSH concentration by several new TSH assays, and that the degree of TSH suppression in goiter or NTI is not as drastic as in hyperthyroidism.

Although the differences between the hyperthyroid group and the euthyroid groups were statistically significant, there was no definite cutoff point that could separate the groups. Thus, when dealing with an unselected patient population in clinical practice, one cannot draw any absolute conclusion about whether a patient is hyperthyroid or not on the basis of TSH concentration alone, without regard to serum free thyroid hormone measurements and clinical examination. It is not surprising that patients with thyroid disease cannot be accurately categorized with only one laboratory test. Nonetheless, the new TSH assays are a big improvement over the previous methodologies.

This study was financially supported by Turku University Foundation.

References