Four Sensitive Thyrotropin Assays Critically Evaluated and Compared

Demetrios S. Sgoutas, Earl G. Barton, Margareta Hammarstrom, Philomena M. Pete, and Sandra A. Sgoutas

We compared four sensitive procedures for thyrotropin (TSH)—Corning's Magic-Lite, ElectroNucleonics' Delfia, Baxter's Stratus, Nichols' Allegro—for their ability to completely discriminate TSH concentrations in sera in euthyroidism, hyperthyroidism, and hypothyroidism. We evaluated the analytical and clinical performance of these procedures according to previously published criteria. All procedures we examined fulfilled the criterion stipulating <1% overlap between the assay variation at the lower normal range limit and the assay sensitivity. Both measurements were made with an interassay CV not exceeding 10% to 15%. All procedures produced results that correlated well for specimens with TSH concentrations of >0.4 milli-int. unit/L, and all four procedures clearly distinguished hypothyroid from euthyroid subjects. In a hyperthyroid—euthyroid comparison, three of the procedures, the Magic-Lite, Delfia, and Allegro, differentiated the two with 97% accuracy, the Stratus procedure with only 90% accuracy. The procedures appeared to differ even more in the measurement of TSH in serum of patients undergoing suppressive treatment with thyroid hormones and in hospitalized euthyroidally ill patients. The observed differences among procedures were thought to be related in part to a matrix serum effect, which is accentuated in samples from hospitalized patients.

Historically, measurement of thyrotropin (TSH) in serum has been generally accepted as the most valuable biochemical test for diagnosis of primary hypothyroidism. It has also been used for diagnosis of secondary (pituitary) and tertiary (hypothalamic) hypothyroidism. The recent availability of monoclonal antibodies has allowed the introduction of immunometric assays with improved sensitivity, which reportedly enabled laboratories to measure suppressive influences on TSH secretion so precisely as to allow discrimination between low euthyroid and hyperthyroid populations. Several methods are now available for measuring TSH based on radiometric (1, 2), chemiluminesmetric (3), fluorometric (4), radial partition immunometric (5, 6), and enzyme-amplified immunoassays (7, 8). However, results generated by these assays have shown that some but not all methods allow use of the TSH concentration in sera to completely discriminate between euthyroid and hyperthyroid patients. Furthermore, monitoring TSH in patients during suppression therapy with thyroid hormone (12), and other studies conducted in hospitalized populations, show remarkable differences among results from various methods (13–15). Here we have evaluated and compared the performance of four highly sensitive procedures relative to currently proposed criteria for sensitive TSH assays (16–18). Our aim was to provide information about these procedures in regard to their usefulness for measuring very low TSH concentration and in diagnosis of thyroid disorders.

Materials and Methods

Blood Specimens and Subject Groups

The patients' samples were taken from specimens received in the last six months. Clinical data were obtained from medical records or from the attending physician. Blood was sampled in the morning between 0800 and 1000 hours, and the serum was separated and stored at −40 °C until analysis.

TSH was measured in five groups of patients. Subjects in these groups had previously been classified on the basis of both clinical data and results of measurements of total triiodothyronine (T3), total thyroxin (T4), free thyroxin (FT4), and thyrotropin (TSH) in serum.

Group A consisted of 48 subjects with no clinical signs of thyroid dysfunction and with normal results for biochemical assessment, including assays for T3, T4, and FT4. Most of the subjects were blood-bank donors and hospital employees who were undergoing physical examination.

Group B consisted of 28 samples from hyperthyroid patients who were chosen on the basis of biochemical profiles of above-normal T4, above-normal T3, normal T4, and increased FT4. The diagnosis of hyperthyroidism was clinically confirmed in addition to biochemical signs other than TSH measurements. Some patients in this group, given the thyroliberin-stimulation test, showed no response, which can occur in patients with ophthalmic Grave's disease or nodular goiter.

Group C consisted of 14 patients with untreated primary hypothyroidism. Their concentrations of T4 and FT4 in serum were below normal, and their TSH values were high (>8 milli-int. units/L).

Group D consisted of 18 patients who were receiving chronic treatment with thyroid hormone to suppress non-toxic goiter or thyroid cancer.

Group E consisted of 36 hospitalized patients, some of them in an intensive-care unit with critical non-thyroidal illnesses for whom routine thyroid-function tests had been ordered.

TSH Assays

We evaluated the following commercial immunometric procedures involving monoclonal antibodies for human TSH: (I) Magic-Lite (Ciba-Corning Diagnostics Corp., Medfield, MA), (II) Delfia (ElectroNucleonics, Fairfield, NJ), (III) Stratus (Dade Division, Baxter Inc., Miami, FL), and (IV) Allegro (Nichols Institute, San Juan, Capistrano, CA).

The Magic-Lite method is a two-site immunochemoluminescence assay involving two mouse monoclonal antibodies to TSH. An acridinium ester is used as chemiluminescence label, the final separation step is accomplished magnetically, and the chemiluminescence of the separated particles is measured in an automated luminometer. A detailed description of the method was given previously (3).
The nonisotopic Delfia is a dissociation-enhanced lanthanide fluoroimmunoassay. It incorporates a europium label and solid-phase separation (microwell trays). A detailed description of the method, including description of the Arcus (Model 1230) time-resolved fluorometer, appeared previously (4).

The Stratus radial partition fluorometric immunoassay, performed on the Stratus analyzer, was also described in detail elsewhere (5, 6). All pipetting, washes, incubation, and data reduction are performed automatically by the instrument.

The Allegro procedure is an immunoradiometric assay involving three monoclonal antibodies with specificity for three different sites of the TSH molecule. One of the monoclonal antibodies is labeled with $^{125}$I, and the other two are coupled to biotin. The addition of an avidin-coated plastic bead allows specific and quantitative binding and separation of the bound hormone. The procedure has been previously described in detail (9).

We counted radioactivity with a Crystal II (Packard Instrument Co., Downers Grove, IL), which reduces the data for immunoradiometric assays on-line by using a spline function with an automatic smoothing factor. Procedures I, III, and IV involve calibrators that are standardized with the World Health Organization's 2nd International Reference Preparation for Human TSH (IRP 80/688), whereas the calibrators in procedure II are standardized with the 1st Standard (IRP 68/38).

All assays were performed as recommended by the manufacturer. However, because we desired to measure very low TSH concentrations but the dose–response curve was not linear between the zero-point and the lowest standard point, we added an extra low standard to all procedures. We prepared low standards by serially diluting the lowest standard provided with the kit by the zero standard of the corresponding kit. We did not test for hormone cross-reactivities, but data supplied by the manufacturers demonstrated no significant cross-reaction with a wide variety of possible cross-reacting peptides. Human serum freed of $T_4$ and $T_3$ was from Chemicon International (Los Angeles, CA) and drug-free serum from Bio-Rad Laboratories (Anaheim, CA).

Other Methods

We measured $T_4$ concentrations in serum with an analog-based technique (Amersham International, Bucks., U.K.), total $T_4$ by using Abbott's TDX methodology (Abbott Diagnostics, North Chicago, IL), and total triiodothyronine ($T_3$) by using Magic $T_3$ immunoradiometric assay (Ciba Corning Diagnostics). Occasionally, and during thyroliberin testing, we measured TSH by a conventional RIA (Hybritech Inc., San Diego, CA). The between-batch precision of the $T_4$ assay was 5.2% at 16.5 pmol/L and 6% at 36 pmol/L; that of $T_3$ was 5.1% at 1 nmol/L and 43% at 2 nmol/L; and that of TSH was 4.8% at 1.5 milli-int. unit/L. Reference values were 9–26 pmol/L for $T_4$, 1.0–2.5 nmol/L for $T_3$, and 0.2–4/milli-int. units/L for TSH.

Statistical methods. Regression analysis, and Student's t-test were carried out with statistical programs for the Apple-McIntosh computer.

Results

Table 1 shows the sensitivities of the four procedures. We assessed analytical sensitivity (19, 20) by replicate (n = 10) measurements of the zero standard supplied with each procedure and derived a mean and standard deviation (SD) from which we calculated analytical sensitivity by interpolating the mean plus 2 SDs from the standard curve. The analytical sensitivities for procedures I, II, III, and IV were 0.06, 0.06, 0.12, and 0.1 milli-int. unit/L, respectively. These values were obtained with intra-assay CVs ranging from 10% to 12%. Assaying pooled sera from thyrotoxic patients who did not respond to thyroliberin stimulation yielded a more meaningful estimate of actual sensitivity: values were 0.1, 0.12, 0.22, and 0.12 milli-int. unit/L for procedures I, II, III, and IV, respectively, with intraassay CVs of 12% to 15% (21). In this study any TSH value of less than the actual sensitivity of a procedure was considered undetectable for that procedure. Following the example of other investigators, we make the distinction between analytical and actual sensitivity because human serum as matrix is different from the zero standard of commercial kits. In fact, the actual sensitivity with an intraassay CV of 12% to 15% is closer to the “functional” methodologic sensitivity limit, as previously defined (16, 17, 20, 21).

Table 2 shows the within- and between-assay precision of analyses for pooled patients' sera containing a low concentration of TSH (0.2, 0.5, and 1.0 milli-int. unit/L) and of "Lyphochek" controls (1.6 and 9.1 milli-int. units/L; Bio-Rad Labs). Each pooled serum and commercial control was

---

**Table 1. TSH Assay Sensitivity (Intra-Assay Analysis of 10 Replicates)**

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>Magic-Lite</th>
<th>Delfia</th>
<th>Stratus</th>
<th>Allegro</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analyticala</td>
<td>0.06</td>
<td>0.06</td>
<td>0.12</td>
<td>0.10</td>
</tr>
<tr>
<td>Reportedb</td>
<td>0.04</td>
<td>0.04</td>
<td>0.08</td>
<td>0.04</td>
</tr>
<tr>
<td>Actualc</td>
<td>0.10</td>
<td>0.12</td>
<td>0.22</td>
<td>0.12</td>
</tr>
</tbody>
</table>

*a Analysis of the zero standard of each kit, mean + 2 SDs. 
b Statistical values reported by the manufacturer. c Analysis of serum with suppressed TSH, mean + 2 SDs.

---

**Table 2. Precision of the Four TSH Assays**

<table>
<thead>
<tr>
<th>TSH concn, milli-int. units/La</th>
<th>Magic-Lite</th>
<th>Delfia</th>
<th>Stratus</th>
<th>Allegro</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Within</td>
<td>Between</td>
<td>Within</td>
<td>Between</td>
</tr>
<tr>
<td>0.2</td>
<td>8.8</td>
<td>10.2</td>
<td>9.2</td>
<td>10.9</td>
</tr>
<tr>
<td>0.5</td>
<td>5.5</td>
<td>6.2</td>
<td>6.5</td>
<td>7.4</td>
</tr>
<tr>
<td>1.0</td>
<td>5.1</td>
<td>5.8</td>
<td>5.8</td>
<td>6.2</td>
</tr>
<tr>
<td>1.6</td>
<td>3.1</td>
<td>4.2</td>
<td>4.6</td>
<td>4.8</td>
</tr>
<tr>
<td>9.1</td>
<td>2.6</td>
<td>3.4</td>
<td>2.5</td>
<td>3.1</td>
</tr>
</tbody>
</table>

*a Pools and their assays are described in the text.*
assayed 10 times with each procedure for the calculation of the within-assay CV and seven times in duplicate over a period of 10 days for the calculation of the between-assays CV.

Table 3 shows the TSH values for 48 euthyroid subjects measured by the four procedures. Procedure II had the lowest and procedure III the highest TSH values. The normal ranges were practically identical and all had a normal distribution. Statistical analysis of the natural-log-transformed values gave a normal distribution with calculated cutoff values for normal of 0.31, 0.28, 0.45, and 0.36 for procedures I, II, III, and IV, respectively. Serum pools with TSH concentration near the lower limit of normal values (0.4 milli-int. unit/L) gave CV of 10–12% and the mean lower limit value (log-transformed) minus 2.6 SD gave values of 0.31, 0.26, 0.44, and 0.36 milli-int. unit/L for procedures I, II, III, and IV, respectively. These values were comparable with the cutoff values (Table 3) and represent the expression of variability at the low normal limit (16).

The correlation among results by all four methods was assessed by analysis by all methods of 88 patients’ samples covering a range of results from 0.4 to 45 milli-int. units/L. Regression of these samples yielded the following: Magic-Lite = (0.921 × Delfia) + 0.328, r = 0.973; Magic-Lite = (0.805 × Stratus) + 0.120, r = 0.879; Magic-Lite = (1.07 × Allegro) + 0.12, r = 0.923; Delfia = (1.29 × Stratus) + 0.530, r = 0.963; Delfia = (1.08 × Allegro) − 0.1, r = 0.910; and Stratus = (1.087 × Allegro) + 0.12, r = 0.946.

Figure 1 shows results for patients in groups A, B, and C. For group A, and in procedure I, 20 patients’ samples had undetectable TSH concentrations; seven patients had detectable but below normal TSH, and one sample was in the normal range. In procedure II, TSH in 19 samples was undetectable, detectable but below normal in eight, and in one sample was in the normal range. Procedure III gave undetectable TSH concentrations in 12 samples, detectable but below normal TSH in 13, and three were in the normal range. With procedure IV, 20 samples had undetectable TSH concentrations, seven were detectable but below normal, and one was in the normal range. The same sample appeared to be unexpectedly normal in all four procedures. Figure 1 also shows that all hypothyroid patients had TSH concentrations above the high reference limit (>7.0 milli-int. units/L) and none in the normal range.

Figure 2 shows the distribution of TSH concentration among patients taking thyroid hormones for suppression therapy (group D). In procedure I, 15 of 18 samples had undetectable TSH concentrations; the other three were detectable but below the normal range. In procedure II, 16 had TSH concentrations that were undetectable and two were detectable but below normal. In procedure III, 12 were undetectable; the remaining six were detectable and clearly in the normal range. In procedure IV, 16 were undetectable and two were detectable. With a few exceptions, the same samples had undetectable TSH concentrations by all procedures.

Figure 3 shows the distribution of TSH concentrations among patients with major nonthyroid illness (group E). In procedure I, eight of 36 patients had undetectable TSH concentrations, eight had detectable TSH, and 20 were in the normal range. The respective distribution of TSH values in procedure II was 10, 8, and 18; in procedure III, 5, 11, and 20; and in procedure IV, 6, 12, and 18. When serum TSH concentrations were grouped according to serum T₄ and T₃ concentrations [subgroup I (n = 22) low serum T₄; normal serum T₃; subgroup II (n = 14) low serum T₃ and low serum T₄], the differences were not statistically significant. For subgroup I, the mean serum TSH in the Delfia assay was 1.4 (SD 1.4) milli-int. units/L, and for subgroup II, it was 1.5 (SD 1.6) milli-int. units/L. Most patients in this group did not undergo thyroliberin testing.

---

**Table 3. Means, SDs, Normal Range, and Cutoff Values with the Four Assays**

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Mean Untransformed TSH, milli-int. units/L</th>
<th>Mean Ln-transformed TSH</th>
<th>Cutoff value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magic-Lite</td>
<td>2.09</td>
<td>0.57</td>
<td>0.31</td>
</tr>
<tr>
<td>Delfia</td>
<td>1.88</td>
<td>0.46</td>
<td>0.28</td>
</tr>
<tr>
<td>Stratus</td>
<td>2.75</td>
<td>0.85</td>
<td>0.45</td>
</tr>
<tr>
<td>Allegro</td>
<td>2.20</td>
<td>0.64</td>
<td>0.36</td>
</tr>
</tbody>
</table>

*The corresponding antilog (e) to 3 natural log value for the healthy population.

---

![Graph showing TSH levels for different procedures](image1)

**Fig. 1.** TSH values for hypothyroid, hyperthyroid, and normal subjects, as measured by all four procedures. Each value is the mean of duplicate determinations.

![Graph showing TSH results for treated patients](image2)

**Fig. 2.** TSH results from treated patients compared with those from controls. Each point is the average of two determinations.
Finally, to evaluate the matrix effect, we tested horse, rabbit, bovine, and mouse sera; serum free of T₃ and T₄; and commercial drug-free serum. The results, in apparent human TSH milli-int. units/L, were respectively: Magic-Lite 0.11, 0.33, 0.04, 0.09, 0.16, and 0.05; Delta 0.01, 0.01, 0.01, 0.05, 0, and 0.02; Stratus 0.04, 3.12, 0.28, 2.76, 1.34, and 1.85; and Allegro 0.21, 0.03, 0.02, 0.30, 0.05, and 0.1.

Discussion

We now discuss our results relative to these criteria as recently described by Klee and Hay (16) for analytical performance and clinical usefulness of sensitive TSH assays.

The most important criterion with regard to analytical performance requires that the assay gives values for serum TSH that do not overlap between the limits of variability of the lower value for normal and the assay sensitivity limit (16, 17). Results in Tables 1 and 3 clearly show no overlap between lower normal values and assay sensitivities. This suggests that all four procedures meet the first requirement, that they can justifiably be called "sensitive," and that they merit further clinical trials with hyperthyroid patients.

Another criterion refers to the clinical usefulness of the assay, such that at least 95% of hyperthyroid patients should have values below the normal range. This criterion initially required that 95% of hyperthyroid patients should give undetectable TSH values. However, as the technology improved, undetectable values became detectable; consequently, this criterion has been changed to allow a decision level not based on assay detection limit. Hence, the American Thyroid Association's Nomenclature Committee has defined a sensitive TSH assay as one in which serum from 95% of the clinically hyperthyroid patients gives results >3 SD below the mean value (after logarithmic transformation) for serum from normal subjects (22). Our results indicate that in procedures I, II, III, and IV, 96%, 97%, 90%, and 97%, respectively, of hyperthyroid patients gave TSH concentrations >3 SD below the mean normal value (after logarithmic transformation). Therefore, all procedures except procedure III met this requirement; procedure III also gave more TSH values overlapping with the normal range.

Careyon et al. (10), evaluating the ability of five sensitive TSH kits to distinguish euthyroid subjects from hyperthyroid patients, reported undetectable TSH concentrations in euthyroid subjects and detectable TSH concentrations in 1.3% to 8.4% of the hyperthyroid subjects. Similarly, Rodriguez-Espinosa et al. (11) reported that four of the six kits they evaluated distinguished clearly between hyperthyroid and euthyroid subjects. In contrast, Hershman et al. (12) found that all five commercial immunometric kits they evaluated clearly distinguished hyperthyroid from euthyroid patients.

A third criterion requires that at least 95% of clinically euthyroid subjects should have basal TSH concentrations within the established normal range. All four procedures met this requirement.

Two other criteria require that at least 95% of patients who have a normal TSH should respond to thyroliberin testing and that at least 95% of patients who have an absent response to thyroliberin should have basal TSH values below the lower normal limits. TSH responses to thyroliberin testing were not studied here because appropriate documentation has been presented in the medical literature for all four procedures (4–9).

In addition to the comparison of the four assays by the proposed criteria by Klee and Hay (16), we also included a group of patients undergoing suppressive treatment with thyroid hormone. These patients had either undetectable or low TSH concentrations by all four procedures. The majority of these patients had high FT₃ values, and any who were tested showed no response to thyroliberin. Four patients, however, had detectable TSH concentrations by procedures I and IV and normal values by procedure III, suggesting some method dependency of the results.

Spencer (13) noted that some patients on Synthroid (T₄) suppression for thyroid carcinoma may display a measurable thyroliberin response when basal TSH concentrations are undetectable. Klee and Hay, on the other hand, used a different assay, suggested that this response is rarely seen (16). Nevertheless, and despite method dependency, sensitive TSH becomes the test of choice for monitoring thyroid therapy (22).

Another possible application of any sensitive TSH method is in the evaluation of thyroid function in hospitalized patients with nonthyroidal illness. Several previous studies have addressed the problem in detail (26–28). To further study the clinical usefulness of the procedures under evaluation, we analyzed sera from 36 hospitalized patients. The wide range of TSH values and the lack of correlation of TSH with T₄ and T₃ concentrations cast doubts on the clinical usefulness of sensitive TSH procedures for differentiating sick euthyroid patients from those with true thyroid disorders. Possible explanations for this include individual abnormalities of the hypothalamic–pituitary–thyroid axis, concomitant drug therapy such as dopamine or glucocorticoids (23–28), or methodological problems associated with assays. For patients with nonthyroidal illness, procedures I, II, and IV showed a higher number of equivocal undetectable TSH results than did procedure III. This difference clearly suggests a loss of adequate specificity of one or more of the procedures under study.

Some of the factors limiting the specificity of a procedure are specificity of antibodies or matrix effects (23). Other investigators have extensively studied specificity of antibodies and explained discrepancies in TSH concentrations on the basis of antibody differences (23). Matrix effects have received little attention, perhaps because of the technical difficulties involved (23, 29, 30). In our evaluation, tests of some animal and other sera showed that all procedures were affected to a certain extent (with the Stratus procedure showing the greatest effect). This might explain how
the same procedures could give undetectable TSH concentrations in procedures I, II, and IV and detectable or even normal in the Stratus procedure. We believe that further investigation is warranted to evaluate more fully the matrix effect on these and other sensitive assays.

References