Performance Characteristics of Six Third-Generation Assays for Thyroid-Stimulating Hormone

Mindy L. Rawlins¹ and William L. Roberts²*

**Background:** Thyroid-stimulating hormone (TSH) is used to detect primary hypo- and hyperthyroidism. Current guidelines for TSH assays recommend a functional sensitivity of ≤0.02 mIU/L. The protocol for determining the functional sensitivity of TSH assays specifies analyses of serum samples with two reagent lots over a 6- to 8-week period.

**Methods:** We determined the functional sensitivities of the Access 2, ADVIA Centaur, ARCHITECT i2000, E170, IMMULITE 2000, and Vitros ECI automated methods, using seven serum pools and two reagent lots for each method.

**Results:** The observed functional sensitivities were as follows: Access 2, <0.020 mIU/L; ADVIA Centaur, 0.039 mIU/L; ARCHITECT i2000, <0.005 mIU/L; E170, 0.011 mIU/L; IMMULITE 2000, 0.014 mIU/L; Vitros ECI, 0.004 mIU/L. However, there were large differences between some method means for the seven serum pools. For the pool with the lowest TSH concentration, mean results were as follows: Access 2, 0.0203 mIU/L; ADVIA Centaur, 0.0085 mIU/L; ARCHITECT i2000, 0.0049 mIU/L; E170, 0.0098 mIU/L; IMMULITE 2000, 0.0077 mIU/L; Vitros ECI, 0.0014 mIU/L. Method-comparison studies using samples with TSH concentrations >0.2 mIU/L also showed method differences. The ARCHITECT i2000 method was the most precise at low TSH concentrations.

**Conclusions:** TSH methods do not provide comparable results for serum pools with TSH concentrations <0.2 mIU/L or for patient results across the analytic measurement range. Further investigation into the cause of these differences and additional harmonization efforts are required.

© 2004 American Association for Clinical Chemistry

Thyroid-stimulating hormone (TSH) measured by sensitive immunoassays can be used to detect primary hypo- and hyperthyroidism. Most current automated methods are based on nonisotopic immunometric assay principles. The National Academy of Clinical Biochemistry has recommended that the functional sensitivity of TSH assays be ≤0.02 mIU/L. This permits patients with nonthyroidal illnesses to be distinguished from those with primary hyperthyroidism. This is particularly important in patients hospitalized with nonthyroidal illnesses in whom TSH concentrations as low as 0.02 mIU/L may be encountered. Ill hyperthyroid patients should have profoundly low TSH concentrations (<0.02 mIU/L). In addition, high-risk patients with well-differentiated thyroid carcinoma who are receiving thyroxine for suppression therapy should have TSH <0.01 mIU/L.

It is recommended that the functional sensitivity of the TSH assay be verified independently of the manufacturer’s claim by use of human serum pools. At least 10 different runs spaced over 6–8 weeks with at least two different lots of reagents and two different instrument calibrations during the test period should be performed. A review of the literature failed to identify any studies conducted in the past 5 years in which the functional sensitivity of multiple TSH methods had been evaluated with these guidelines. We therefore conducted experiments according to these guidelines, using six automated TSH methods that were referred to as “third generation” in the manufacturers’ package inserts. We also compared the mean TSH results, using the same serum pools that were used to assess imprecision, to evaluate the agreement of these automated methods for low TSH concentrations.

---

¹ ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT.
² Department of Pathology, University of Utah Health Sciences Center, Salt Lake City, UT.
*Address correspondence to this author at: ARUP Laboratories, 500 Chipeta Way, Salt Lake City, UT 84108. Fax 801-584-5207; e-mail william.roberts@aruplab.com.

Received June 25, 2004; accepted September 14, 2004.
Previously published online at DOI: 10.1373/clinchem.2004.039156
Materials and Methods
The following automated methods and instruments were evaluated: the Access 2 analyzer (Beckman Coulter), the ADVIA Centaur analyzer (Bayer Diagnostics), the ARCHITECT i2000 analyzer (Abbott Diagnostics), the Elecsys E170 analyzer (Roche Diagnostics), the IMMULITE 2000 analyzer (Diagnostics Products Corporation), and the Vitros ECI analyzer (Ortho-Clinical Diagnostics). All of the above immunoassays use either chemiluminescent or electrochemiluminescent detection. The ADVIA Centaur and IMMULITE 2000 have reagent sets available for both second- and third-generation assays. All assays used in this study were labeled third generation by their manufacturers. Third generation is defined as having a functional sensitivity, i.e., the lowest analyte concentration with an interassay imprecision ≤20%, of 0.01–0.02 mIU/L. Two different lots of reagents were used for each method. According to the manufacturers’ information, all assay calibrators are traceable to the WHO Second International Reference Preparation 80/558.

Assay imprecision was assessed by use of commercial quality-control materials, Anemia and Lyphocheck Levels 1–3 (Bio-Rad Laboratories). Each control material was analyzed by each method in duplicate per run. Daily runs were performed 2 days a week over a period of 3 weeks for each lot of reagent for a total of 24 replicates for each control. We estimated the functional sensitivity of each assay by preparing seven human serum pools from samples submitted for routine clinical testing. To prepare the pools, we assayed these samples on an E170 analyzer and combined samples with similar results. Aliquots of these pools were stored at −70°C until use. Each pool was analyzed as outlined above for the commercial control material.

The overall CV was calculated by use of each individual replicate for each concentration of the control material and serum pool for each method. The functional sensitivity of each method, except for the Access 2 and ARCHITECT i2000 methods, was estimated by fitting a power function to the imprecision data for pooled serum samples and calculating the TSH concentration that gave a CV of 20%. The Access 2 and ARCHITECT i2000 methods did not achieve a CV >20% with any pool that was tested. A method-comparison study was also conducted with one lot of reagent and using 104 patient samples with TSH concentrations that spanned the analytic measurement range of each assay. All studies using samples from humans were approved by the Institutional Review Board of the University of Utah. Passing–Bablok analysis of these data was performed with Analyze-it™ software version 1.71 (Analyze-it Software, Ltd.). The standard error of the estimate ($s_{y|x}$) and correlation coefficient ($r$) were obtained from linear regression analysis using Analyze-it software.

Results
Commercial quality-control materials were assayed in parallel with human serum pools to assess imprecision at higher TSH concentrations and to confirm that each assay was functioning as expected. A summary of the imprecision results obtained with these materials is shown in Table 1. The range of mean TSH concentrations observed for the lowest control (Anemia) was from 0.058 mIU/L for the ARCHITECT i2000 to 0.104 mIU/L for the E170 analyzer. The range of mean TSH concentrations observed for the high control (Lyphocheck Level 3) for all analyzers except the Vitros ECI was much narrower (23.4–27.7 mIU/L). The Vitros ECI, however, gave a mean concentration of 37.8 mIU/L for this control. These quality-control materials are prepared with artificial matrices and may not accurately reflect results seen with patient samples.

The results of imprecision studies conducted with human serum pools are summarized in Table 2. The estimated functional sensitivities for each method obtained with the mean pool concentration determined by each method are summarized in the legend to Fig. 1. It is noteworthy that for the Access 2 and ARCHITECT i2000 assays, the pool with the lowest TSH concentration did not have CVs >20%; therefore, the true functional sensitivities of these assays could not be adequately assessed. The CV of the Access 2 method was 11% at a TSH concentration of 0.0203 mIU/L, and the ARCHITECT i2000 produced a CV of 6.9% at a TSH concentration of 0.0049 mIU/L. The ADVIA Centaur method did not meet the definition of a third-generation TSH assay because the TSH concentration at which it had a CV of 20% was 0.039 mIU/L and not <0.02 mIU/L. It is also noteworthy that for the human serum pool with the lowest TSH concen-

<table>
<thead>
<tr>
<th>Control material</th>
<th>Anemia</th>
<th>Lyphocheck Level 1</th>
<th>Lyphocheck Level 2</th>
<th>Lyphocheck Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean TSH, mIU/L</td>
<td>CV, %</td>
<td>Mean TSH, mIU/L</td>
<td>CV, %</td>
</tr>
<tr>
<td>Access 2</td>
<td>0.060</td>
<td>6.7</td>
<td>0.408</td>
<td>6.3</td>
</tr>
<tr>
<td>ADVIA Centaur</td>
<td>0.070</td>
<td>19</td>
<td>0.459</td>
<td>5.7</td>
</tr>
<tr>
<td>ARCHITECT i2000</td>
<td>0.058</td>
<td>3.6</td>
<td>0.362</td>
<td>4.3</td>
</tr>
<tr>
<td>E170</td>
<td>0.104</td>
<td>2.9</td>
<td>0.530</td>
<td>1.6</td>
</tr>
<tr>
<td>IMMULITE 2000</td>
<td>0.095</td>
<td>5.9</td>
<td>0.423</td>
<td>5.2</td>
</tr>
<tr>
<td>Vitros ECI</td>
<td>0.086</td>
<td>6.3</td>
<td>0.396</td>
<td>3.7</td>
</tr>
</tbody>
</table>
When the mean values for each method for pool 1 were compared by use of the $t$-test, the ADVIA Centaur value was not significantly different from either the E170 or the IMMULITE 2000 mean values ($P = 0.28$ and 0.70, respectively). All other combinations of methods including the E170 and IMMULITE 2000 were statistically different ($P < 0.01$).

Because all methods ranked the TSH concentrations in the seven human serum pools in the same order and the TSH concentration was lowest in pool 1, we also judged imprecision at low TSH concentrations based on the CV calculated for pool 1. In this case, the ARCHITECT i2000 was the most precise, followed by the Access 2, the E170, the IMMULITE 2000, the Vitros ECI, and the ADVIA Centaur in that order.

We assessed the comparability of methods, using patient samples with a wide range of TSH concentrations (Fig. 3). We used the IMMULITE 2000 as the comparison method because it gave “middle” values for the seven human serum pools and had a functional sensitivity $<0.02$ mIU/L. The Access 2, ARCHITECT i2000, and IMMULITE 2000 methods agreed well with one another and had slopes within 10% of 1.0, whereas the ADVIA Centaur, E170, and Vitros ECI yielded consistently higher results with slopes of 1.23–1.33. The scatter around the regression line was noticeably increased for TSH concentrations $>25$ mIU/L.

Inspection of results from the method-comparison study performed with individual patient samples revealed eight samples with TSH concentrations $<0.1$ mIU/L by the IMMULITE 2000 method. The sample with the lowest TSH concentration had an IMMULITE 2000 result of 0.01 mIU/L. The results obtained with the ARCHITECT i2000, E170, and Vitros ECI were 0.0173, 0.035, and 0.006 mIU/L, respectively. It is noteworthy that the ADVIA Centaur and Access 2 results for this sample were 0.133 and 0.20 mIU/L. The other seven samples in this group generally showed much closer agreement than was seen for the first sample. The method means for the seven samples were 0.091, 0.104, 0.054, 0.084, 0.053, and 0.040 mIU/L for the Access 2, ADVIA Centaur, ARCHITECT i2000, E170, IMMULITE 2000, and Vitros ECI methods, respectively.

### Discussion

Our results indicated that all third-generation TSH assays had functional sensitivities $<0.02$ mIU/L except for the ADVIA Centaur method, which had a functional sensitivity of 0.039 mIU/L. The imprecision of the ADVIA Centaur with the Bio-Rad anemia control was 19% at a TSH concentration of 0.07 mIU/L, and the imprecision of the ADVIA Centaur with pool 2 was 32% at a TSH concentration of 0.021 mIU/L. One previous report on the ACS:180 TSH-3 assay found a functional sensitivity of 0.015 mIU/L (2). This assay is quite similar to the ADVIA Centaur assay that we used. A second report found functional sensitivities of 0.033 mIU/L for the ADVIA Centaur and 0.039 mIU/L for the ACS:180 when testing was conducted over a period of 6 weeks with a single lot of reagents (3). This report agrees well with our estimate of the functional sensitivity of the ADVIA Centaur. The reason for the poor precision of the current ADVIA Centaur is unclear. The instrument we used was being used for routine TSH measurements and was maintained according to the manufacturer’s recommendations. One other study examined the functional sensitivity of several automated methods (4). This study found a functional sensitivity of 0.028 mIU/L for the Access method, 0.015 mIU/L for the IMMULITE method, and 0.026 for the ACS:180 method. Our results for the IMMULITE 2000 method are comparable, whereas we found a lower functional sensitivity for the Access 2. Our functional sensitivity estimate of $<0.0049$ mIU/L for the ARCHITECT i2000 method is comparable to an earlier study (5).

One limitation of our study design was the manner in which samples were incorporated with routine testing. The instrument operator was not blinded to the presence of test pools in the run. The technologist responsible for

<table>
<thead>
<tr>
<th>Pool</th>
<th>ACCESS 2</th>
<th>ADVIA Centaur</th>
<th>ARCHITECT i2000</th>
<th>E170</th>
<th>IMMULITE 2000</th>
<th>Vitros ECI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0203 (11)</td>
<td>0.0085 (73)</td>
<td>0.0049 (6.9)</td>
<td>0.0098 (27)</td>
<td>0.0077 (30)</td>
<td>0.0014 (41)</td>
</tr>
<tr>
<td>2</td>
<td>0.0290 (6.3)</td>
<td>0.0213 (32)</td>
<td>0.0138 (5.7)</td>
<td>0.0213 (16)</td>
<td>0.0176 (18)</td>
<td>0.0033 (23)</td>
</tr>
<tr>
<td>3</td>
<td>0.0502 (7.3)</td>
<td>0.0300 (19)</td>
<td>0.0204 (4.6)</td>
<td>0.0309 (8.3)</td>
<td>0.0273 (16)</td>
<td>0.0056 (15)</td>
</tr>
<tr>
<td>4</td>
<td>0.0505 (5.8)</td>
<td>0.0460 (17)</td>
<td>0.0302 (5.4)</td>
<td>0.0466 (6.3)</td>
<td>0.0360 (9.1)</td>
<td>0.0112 (12)</td>
</tr>
<tr>
<td>5</td>
<td>0.0927 (4.8)</td>
<td>0.0852 (11)</td>
<td>0.0613 (5.6)</td>
<td>0.0875 (3.6)</td>
<td>0.0696 (6.4)</td>
<td>0.0355 (6.2)</td>
</tr>
<tr>
<td>6</td>
<td>0.1465 (6.5)</td>
<td>0.1498 (6.9)</td>
<td>0.1037 (5.9)</td>
<td>0.1478 (2.4)</td>
<td>0.1156 (5.4)</td>
<td>0.0838 (3.2)</td>
</tr>
<tr>
<td>7</td>
<td>0.1931 (6.3)</td>
<td>0.2116 (7.0)</td>
<td>0.1459 (5.5)</td>
<td>0.1955 (2.8)</td>
<td>0.1693 (5.6)</td>
<td>0.1377 (3.3)</td>
</tr>
</tbody>
</table>
Fig. 1. Plots of functional sensitivity data.

The CV for each human serum pool was plotted against its mean concentration for each method, and the data for each method were fitted with a power curve. The functional sensitivity was the TSH concentration that gave a CV of 20%. The functional sensitivities were as follows: (A), Access 2, <0.020 mIU/L; (B), ADVIA Centaur, 0.039 mIU/L; (C), ARCHITECT i2000, <0.005 mIU/L; (D), E170, 0.011 mIU/L; (E), IMMULITE 2000, 0.014 mIU/L; (F), Vitros ECI, 0.004 mIU/L.
Figure 2. Mean TSH concentrations for each human serum pool and method.

The mean concentration of each human serum pool measured by each method was plotted against the pool number. □, Access 2; ◦, ADVIA Centaur; ○, ARCHITECT i2000; ∆, E170; ×, IMMULITE 2000; +, Vitros ECI.

This project loaded all samples on all analyzers and was the primary operator of the ARCHITECT i2000 analyzer. We doubt that this factor had a significant effect on our results because current analyzers are "black boxes" on which samples are loaded and results are produced automatically. However, we cannot exclude the possibility that functional sensitivity estimates performed in the routine clinical laboratory might actually be higher than our estimates.

There was good agreement of all methods for pool 7, which had a TSH concentration of 0.2 mIU/L. Below this concentration the Vitros ECI gave lower results than the other five methods. At TSH concentrations <0.05 mIU/L, the Access 2 method gave substantially higher concentrations than the other methods. For human serum pool 1, the Vitros ECI measured a mean concentration of 0.0014 mIU/L, and the mean for the Access 2 was 0.0203 mIU/L. The other four methods gave a mean of 0.0077 mIU/L for this pool. The Access 2 result was more than twice the next highest result.

Overrecovery with the Access 2 method has not been reported previously. The lower than expected results for the Vitros ECI have been described previously (6–8). An earlier study comparing four TSH methods with samples having TSH values <0.8 mIU/L found that only two of the four showed good agreement (9). Possible explanations include differences in calibration, although all assays appeared to have been calibrated according to the same international standard, and differences in the specificities of the anti-TSH monoclonal antibodies used. These authors recommended that an international collaborative study be conducted to clarify reasons for assay differences and to standardize TSH measurements. To our knowledge, such a study has not been performed, and it is still clearly needed. An opinion on the general topic of immunoassay standardization, including the issues of analyte heterogeneity and the lack of reference methods, has been published previously in this journal (10).

When we reviewed the data from the method-comparison study for the nine results that were <0.1 mIU/L on the IMMULITE 2000, most of the results from the Access 2 were comparable to the other methods. However, one of the nine samples gave a result that was 20-fold higher than the IMMULITE 2000 result. It is possible that a few patient samples that gave substantially higher results on the Access 2 method than on the E170 and IMMULITE 2000 methods, which were used to screen samples for pool preparation, could have caused an increase in the Access 2 results compared with the other methods. It would be interesting to prepare patient pools using the Access 2 method to screen samples and repeat the functional sensitivity study with new pools. An alternative approach that might be useful to address the ability of TSH assays to measure low TSH concentrations is to estimate the minimum detectable concentration (11). This approach incorporates both linearity and imprecision considerations and has the potential to overcome some of the limitations of the concept of functional sensitivity that we have highlighted.

The intermethod discrepancies at low TSH concentrations suggest that the clinical usefulness of some TSH assays may be impaired. Diagnosis of hyperthyroidism in hospitalized patients and administration of appropriate thyroxine suppression therapy for thyroid carcinoma require accurate TSH measurements at TSH concentrations of 0.01–0.02 mIU/L. It is difficult to find published studies on the clinical usefulness of commercial TSH assays for these indications. In light of our results, additional information on the clinical usefulness of current commercial TSH assays is needed.

We estimated reference intervals (Table 3) for each of the methods, using the slope and intercept from Passing–Bablok analysis for each method compared with the IMMULITE 2000 and a reference interval of 0.4–4.0 for the IMMULITE 2000, which was obtained from its package insert. This approach yielded lower reference limits ranging from 0.39 to 0.55 mIU/L and upper reference limits from 4.0 to 5.3 mIU/L. According to each manufacturer, all methods have been standardized to the same reference material. Our results with patient samples call into question how well this standardization process is working for TSH. There seem to be two groups of methods. One group, which includes the Access 2, ARCHITECT i2000, and IMMULITE 2000, agree reasonably well for patient samples containing >0.1 mIU/L TSH. The other group, which includes the ADVIA Centaur, E170, and Vitros ECI, also agree reasonably well, but the two groups differ by 15–35%. TSH from the human pituitary
Fig. 3. Method comparison of TSH methods with individual patient samples spanning the analytic measurement range. The IMMULITE 2000 was chosen as the comparison method because it yielded the mean value for each of the human serum pools. The dotted line represents the ideal comparison ($x = y$), and the solid line is from Passing–Bablok analysis. The regression statistics are summarized in Table 3.
has several different glycoforms (12). Differences in how TSH glycoforms are recognized by different immunoassays may contribute to difficulties in assay standardization and lead to discordant results with individual patient samples (13). Additional investigations into the role TSH glycoforms may be useful.

In conclusion, additional studies are needed to understand why there are differences among methods for samples with TSH concentrations <0.2 mIU/L. Once these differences are better understood, additional harmonization efforts can be implemented so that TSH methods show better agreement for samples with low TSH concentrations. All of the methods that we examined had a functional sensitivity <0.02 mIU/L except for the ADVIA Centaur. The ARCHITECT i2000 and Vitros ECi demonstrated the lowest functional sensitivities, but the ARCHITECT i2000 was the most precise of the methods evaluated in this study, with a CV of 6.9% at a TSH concentration of 0.0049 mIU/L.

This work was supported by the Diagnostics Division of Abbott Laboratories and the ARUP Institute for Clinical and Experimental Pathology.

References

### Table 3. Summary of regression statistics for each method compared with the IMMULITE 2000 method.

<table>
<thead>
<tr>
<th>Method</th>
<th>Slope* (95% CI)</th>
<th>Intercept* (95% CI), mIU/L</th>
<th>Slope** (95% CI), mIU/L</th>
<th>Intercept** (95% CI), mIU/L</th>
<th>Mean value at 0.4 mIU/L</th>
<th>Mean value at 4.0 mIU/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access 2</td>
<td>1.07 (1.04–1.11)</td>
<td>0.02 (−0.02 to 0.05)</td>
<td>3.4</td>
<td>0.98</td>
<td>0.45</td>
<td>4.3</td>
</tr>
<tr>
<td>ADVIA Centaur</td>
<td>1.30 (1.25–1.34)</td>
<td>0.03 (−0.03 to 0.10)</td>
<td>4.2</td>
<td>0.97</td>
<td>0.55</td>
<td>5.2</td>
</tr>
<tr>
<td>ARCHITECT i2000</td>
<td>0.99 (0.97–1.02)</td>
<td>−0.01 (−0.02 to 0.01)</td>
<td>2.9</td>
<td>0.98</td>
<td>0.39</td>
<td>4.0</td>
</tr>
<tr>
<td>E170</td>
<td>1.33 (1.30–1.38)</td>
<td>0.01 (−0.04 to 0.06)</td>
<td>4.2</td>
<td>0.98</td>
<td>0.54</td>
<td>5.3</td>
</tr>
<tr>
<td>Vitros ECi</td>
<td>1.23 (1.20–1.27)</td>
<td>−0.03 (−0.09 to −0.01)</td>
<td>3.8</td>
<td>0.98</td>
<td>0.46</td>
<td>4.9</td>
</tr>
</tbody>
</table>

*The slope and intercept were determined by Passing–Bablok analysis.*

**CI, confidence interval.

*The mean values at the lower and upper limits of the IMMULITE 2000 reference interval were calculated for each method by use of the slope and intercept estimated from Passing–Bablok analysis.*