How Sensitive Are Immunoassays for Thyrotropin?


The usual method for calculation of the "sensitivity" of thyrotropin immunoassays is multireplicate analysis of the zero analyte standard. Although this is a statistically valid estimate of the scatter likely to be found in the response variable, it is unrelated to normal analytical practice (usually analysis in duplicate) and estimates intra-assay errors only. This study was designed to assess the analytical performance of 10 immunoassays used routinely for measurement of thyrotropin in human serum. Response data from each assay were accumulated to provide (a) an estimate of "sensitivity" from multireplicate analysis and (b) an estimate of "minimum detection limit," relating directly to errors associated with routine performance and derived from a minimum of 500 duplicate analyses. We conclude that the "minimum detection limit" should be promoted as a more meaningful measure of assay performance at low analyte concentrations than the "sensitivity" derived from multireplicate analysis of the zero-analyte standard.

Additional Keyphrases: definition of "minimum detection limit" • precision–dose profile • assay sensitivity

Advances in immunoassay technology leading to the development of immunoassay technique, and the availability of non-isotopic labeled compounds with high specific activity, have enabled the accurate measurement of low concentrations of human thyrotropin (thyroid-stimulating hormone, TSH) in serum (1). This has led to the development of several commercially available assays, with all of which it is claimed that untreated hyperthyroid and euthyroid subjects may be differentiated by measuring serum TSH concentration alone. The feature of immunoassay performance that defines the ability to measure low concentrations of analyte and, in the case of TSH measurement, to distinguish hyperthyroidism from euthyroidism, is the "sensitivity." Currently, there are problems with this term and its numerical definition. "Sensitivity" may be defined as the precision at zero analyte concentration, whereas the term "minimum detection limit" may be used to refer to the ability to distinguish a low analyte concentration as being statistically different from zero (2).

The most commonly quoted method of determination of "sensitivity" is by multireplicate analysis (usually n = 20) of the zero standard, followed by determination of the TSH concentration at 2, 2.5, or 3 standard deviations above the mean response for zero analyte concentration. This is a statistically valid estimate of the scatter likely to be found in response variable at that level, but it is unrelated to normal analytical practice (usually samples are measured in duplicate) and estimates intra-assay errors only. It also assumes that the matrix of the zero standard and that of the serum specimens is identical. An approach more likely to define the minimum detection limit encountered in a clinical chemistry laboratory would be that recommended by Ekins (3), involving the accumulation of error data from several assays before the minimum detection limit is calculated. This has the virtue that the minimum detection limit so defined relates to data derived from routinely performed assays by use of duplicate estimations and avoids the artificial practice of multireplicate analysis.

Because of the different claims being made (by manufacturers and in the scientific literature) for minimum detection limits of immunoassays for TSH, we undertook a study that had three aims:

1. To accumulate sufficient data from any one immunoassay for an estimate of the minimum detection limit to be made that relates directly to the errors associated with routine performance.

2. To determine whether estimates of sensitivity by multireplicate analysis within assay give an overoptimistic estimate of assay performance.

3. To suggest a method of deriving the minimum detection limit that may be used to compare the relative performance of different immunoassays.

Materials and Methods

The kits we used for the immunoassay of TSH are described in Table 1. The details of an "in-house" immunoradiometric assay (IRMA) included in the study are published elsewhere (4). Each assay was conducted in duplicate in a single laboratory, with rigid adherence to the manufacturer's recommended procedure (except for the Mauclaire assay, in which we used half the recommended volumes for sample and reagents). Individual standard deviations for at least 500 duplicate analyses, including standards, controls, and samples collected from several consecutive assays by each method, were calculated and pooled to derive a response–error relationship (3, 5). This relationship, which predicts the expected standard deviation for any given level of response, was used with typical standard curve response variables to calculate the range of coefficients of variation (CVs) encountered for that standard curve—the precision–dose profile (3, 5). The minimum detection limit was interpolated from that precision–dose profile at a CV of 22% (see Appendix).

Participants in the study performed, in one assay, multireplicate analysis (n = 20) of both the analyte-free standard from their own method and a sample of TSH-free serum provided. The raw data from the analysis were used to derive a mean and standard deviation from which the sensitivity was calculated by interpolating mean +2, 2.5, and 3 SD from the standard curve. An estimate of any statistically significant difference between TSH-free standard and TSH-free serum was provided by an unpaired t-test.

Received August 31, 1988; accepted November 1, 1988.
Table 1. Immunometric TSH Assay Sensitivity (Intra-Assay Analysis of 20 Replicates of Zero Standard or TSH-Free Serum)

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Sensitivity, milli-int. unit/L</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zero standard</td>
</tr>
<tr>
<td></td>
<td>+3.0 SD</td>
</tr>
<tr>
<td>RIAs</td>
<td></td>
</tr>
<tr>
<td>Behring Diagnostics</td>
<td>0.007</td>
</tr>
<tr>
<td>Celltech Diagnostics Ltd</td>
<td>0.025</td>
</tr>
<tr>
<td>Amersham International plc</td>
<td>0.109</td>
</tr>
<tr>
<td>Becton-Dickinson UK (Ltd)</td>
<td>0.118</td>
</tr>
<tr>
<td>Serono Diagnostics Ltd</td>
<td>0.087</td>
</tr>
<tr>
<td>Pharmacia</td>
<td>0.134</td>
</tr>
<tr>
<td>Non-isotopic assay</td>
<td></td>
</tr>
<tr>
<td>Delfia (GB Ltd)</td>
<td>0.007</td>
</tr>
<tr>
<td>Amersham International plc</td>
<td>0.011</td>
</tr>
<tr>
<td>Celltech Diagnostics Ltd</td>
<td>0.060</td>
</tr>
<tr>
<td>Magic-Lite (CIBA-Corning Diagnostics)</td>
<td>0.016</td>
</tr>
</tbody>
</table>

*Significant difference (P < 0.001) between zero standard (column 3) and TSH-free serum (column 4).

TSH-free serum was obtained from three euthyroid volunteers who each took, on consecutive weeks, 0.1, 0.2, and 0.3 mg of thyroxin per day. At the end of the three-week period, each volunteer donated a pint of blood. Serum from these donations was screened by multiplicate analysis in the in-house IRMA, and the serum pool with the lowest signal was selected for use.

Results

Assay sensitivities calculated from multiplicate analysis of the analyte-free standard are summarized in Table 1. There is no consensus regarding the number of standard deviations to be used for this calculation. The most commonly occurring increments are 2.0, 2.5, and 3.0 SD, and data for all three appear in Table 1. In all cases the sensitivities obtained are consistent with those claimed by the manufacturers in the literature accompanying each kit.

For half the immunometric assays studied, there was a small but statistically very significant difference (P < 0.001) in sensitivity derived by using TSH-free serum in place of TSH-free standard (Table 1). In all cases a lower value was achieved with TSH-free serum, although this difference was negligible in terms of clinical significance.

Summarized in column 1 of Table 2 are the minimum detection limit values for the immunometric assays derived from the precision–dose profile techniques of Ekins (3). A direct comparison of minimum detection limit (column 1) and 2 SD sensitivity (column 2) illustrate that in all but two assays, the parameter derived from the precision–dose profile was three- to 24-fold less favorable than that derived from multiplicate analysis (column 3). Representative examples of precision–dose profiles are illustrated in Figure 1.

Discussion

An appreciation of the limitations of any assay requires that the errors associated with that assay be fully defined. For the new generation of TSH assays, where a low subnormal concentration of TSH in serum is of great significance in the strategy of investigation of thyroid disease (1), it is especially important to define the errors of measurement close to zero analyte concentration and to understand the basis on which assay sensitivity or minimum detection limit is determined.

The most commonly quoted definition of TSH assay sensitivity is based on multiplicate analysis of analyte-free standard, but this takes no account of inter-assay errors or those arising from the varying matrices found in different sera (6). Therefore, this definition of sensitivity probably gives an overly optimistic estimate of the performance of a TSH assay at low analyte concentrations.

One practical alternative to the derivation of sensitivity...
from intra-assay multiplicate analysis is that proposed by Bayer (6), in which five serum pools with TSH concentrations corresponding to the subnormal (at least two pools) and normal range are included in eight to 10 consecutive routine assays to assess the precision actually achieved by a laboratory over the normal and subnormal range. We feel that the information derived from this analysis, which is restricted to determination of precision at low analyte concentration, can be extended to provide precision data at the full range of analyte concentrations by the calculation of a precision–dose profile for each assay. The ready availability of basic microcomputer facilities and the choice of appropriate software mean that all laboratories performing TSH assays should be able to derive a precision–dose profile from their raw data. It has been our experience (5) that at least 500 duplicate analyses are required to provide a database that accurately reflects the errors encountered in an assay. This database should be derived from several assays and should reflect the range of values encountered in routine practice. The software used should be capable of continuous updating of this database so that it always consists of at least the most recent 500 duplicate analyses. The software used in this study (adapted to the Apple IIe microcomputer, courtesy of Prof. R. P. Ekins and Dr. P. R. Edwards, Department of Molecular Endocrinology, the Middlesex Hospital Medical School, London, U.K.) has this facility, as do a number of alternative software packages (7).

To clarify the existing confusion over terminology we recommend that the term “minimum detection limit” be reserved for that measure of an analyte concentration that is significantly different from zero as determined by the precision–dose profile, which takes into account the whole range of assay errors. Furthermore, we recommend that the minimum detection limit be the measure of choice when comparing the performance of different methods for the same analyte at low analyte concentrations. In this study, therefore, Table 2, column 1, is our preferred basis for assay comparison.

No firm conclusion can be reached about the detailed ranking of performance of the TSH assays included in this study, because each assay was used by one laboratory only and the data obtained reflected the performance of the individual analyst as well as the assay itself. Nevertheless, Table 2, column 1, suggests that the non-isotopic immuno- metric TSH assays generally have lower minimum detection limits than do isotopic assays. Although this may have been anticipated on the basis of assay theory (8), the factor displayed in Table 2, column 3, reveals that IMMAES are in general less precise than non-isotopic assays. There could be two reasons for this: first, the higher specific activity of non-isotopic label, and second, the deterioration in measured response as a function of age with respect to isotopic labels.

In conclusion, we recommend that each laboratory should derive the minimum detection limit for each TSH assay that they perform and use this when interpreting patients’ results.

Appendix

Statistical analysis of the probability of a certain analyte concentration being significantly different from zero is required, and this is provided by the t-test.

\[
t = \frac{X_1 - X_2}{\text{standard error}} \]

where \(X_1\) is the response at minimum detection limit and \(X_2\) is the response at zero analyte concentration.

\[
\text{Standard error} = \frac{\text{standard deviation (SD)}}{\sqrt{n}}
\]

Re-arranging equation 1 gives:

\[
t \cdot \frac{\text{SD}}{\sqrt{n}} = X_1 - X_2
\]

Hence, the minimum detection limit for the assay is represented by an analyte concentration whose response is greater than that encountered at zero analyte concentration by the factor, \(t \cdot (\text{SD}/\sqrt{n})\).

For duplicates, \(n = 2\), so that:

\[
\text{minimum detection limit} = t \cdot \frac{\text{SD}}{\sqrt{2}}
\]

However, the coefficient of variation (CV), is calculated from:

\[
\frac{\text{SD}}{\text{mean}} \times 100\%
\]

which, in this context, becomes:

\[
\frac{\text{SD}}{\text{minimum detection limit}} \times 100\% = \frac{100 \times \sqrt{2}}{t}\%
\]

Then, substituting the equation for minimum detection limit (no. 3) into the above (no. 4) gives:

\[
\text{CVDL} = \frac{100 \times \sqrt{2}}{t}\% = \frac{6.31 \times \sqrt{2}}{t}\% = 22\%
\]

References
7. RIACALC is a trademark of Pharmacia (GB) Ltd, Milton Keynes, U.K.