Progress Report on a National Quality-Control Survey of Triiodothyronine and Thyroxin Assay

G. C. Zucchelli, A. Pilo, M. R. Chiesa, and M. A. Piro

In January 1980, a national external quality-control survey was organized to evaluate assays for triiodothyronine (T₃) and thyroxin (T₄). Currently, about 150 laboratories are involved. Each participant has received and assayed 100 quality-control samples during four periods of about six months each. The average analytical performance achieved by the participants in each six-month period was estimated by computing the average between-laboratory agreement (CVₒ), the overall average bias, and the average laboratory imprecision. During the 2.5 years of the survey, analytical performance has improved for both assays (CVₒ decreased from 17.0 to 15.7% for T₃ and from 13.1 to 12.7% for T₄). Analysis of survey results according to the method-kit used (mean kit bias and kit imprecision for the nine kits most used by participants) showed that the analytical reliability of the T₄ assay is generally better than that observed for T₃, mainly because of the larger systematic differences among T₃ kits.

Interlaboratory surveys, widely used in clinical chemistry, have been extended to radioimmunoassay procedures in the last few years (1-5). The major goal of external quality-control surveys (EQCS) is to improve the reliability of analyses done by the participating laboratories. By comparing their own performance with that of other laboratories and assessing the periodic EQCS reports, laboratories obtain useful information on the analytical reliability of a method or a kit. Update of the analytical performances of various kits, which is easily generated from the large amount of data collected during the interlaboratory survey, is a major tool for improving the quality of the assays for which they are used. The laboratories gain some quantitative basis for choosing among the available kits, and kit manufacturers are prompted to provide more reliable products.

Starting in January 1980, we organized a national EQCS for assays for T₃ and T₄ in which more than 150 laboratories are now involved (4, 5). During the first two-and-a-half years of the survey 100 samples have been sent to each participant, in 22 monthly dispatches.

From the results we have prepared monthly reports summarizing the statistics associated with each sample and four end-of-period reports containing estimates of the mean bias and imprecision achieved by each laboratory. Here we report a comparison of the analytical reliability achieved by the participants during the four periods of the EQCS, and evaluate the performances of the T₃ and T₄ kits the laboratories used in the fourth six-month period.

Materials and Methods

Outline of the Scheme

Every month, each participating laboratory receives three to six samples of human serum (nonreactive for hepatitis B surface antigen), pooled from the excess after routine analysis for T₃ and T₄. Such pooled serum is filtered through 0.45-μm pore-size filters (catalog designation HA; Millipore Corp., Bedford, MA), then is mixed with 10 mmol of sodium azide per liter, as preservative. EQCS samples are sent by mail, at room temperature. The laboratories are instructed to measure T₃ and T₄ in the samples by their routine procedures and to return the results thus obtained together with the name of the method (or kit) they used. Results are computer processed and a monthly report is printed containing mean, median, SD, CV, and range—both for all results and for results subdivided according to the method or kit—and sent back to each participant (6). All data accumulated during a six-month period are used to prepare an end-of-period report containing data on bias and imprecision of each participating laboratory; a plot of imprecision—bias is also included, to facilitate evaluation of each laboratory's performance (4, 6).

Data Analysis

From all results reported to the survey, the following parameters have been computed:

Laboratory bias: This is the mean of the percent deviations—in comparison with the consensus mean—of the results the laboratory obtained for all the EQCS samples assayed in the considered period. The bias bᵢ of the ith laboratory was computed as:

\[ bᵢ = \frac{\sum_{j} xᵢ - mᵢ}{mᵢ} \times 100/M \]

where \( xᵢ \) is the result of the ith laboratory for the jth EQCS sample; \( mᵢ = \sum xᵢ/N \) is the consensus mean of the jth EQCS sample, where \( N \) is the number of results returned (or the number of participants); and \( M \) is the number of EQCS samples mailed out during the six-month period.

Laboratory imprecision: The mean imprecision estimated from results reported by the laboratory for unidentified replicate samples. A CV was computed from results of EQCS samples prepared from the same pool and assayed by the laboratory on two or more occasions during the six-month period; the mean imprecision was obtained by pooling the CVs for different replicate pools.

\[ CV = \sqrt{\frac{(n₁ - 1)CV₁ + (n₂ - 1)CV₂}{n₁ + n₂ - 2}} \]

C.N.R. Institute of Clinical Physiology, Via Savi, 8-56100 Pisa, Italy.

Nonstandard abbreviations: EQCS, external quality-control survey; T₃, triiodothyronine; T₄, thyroxin; CVₒ, average between-laboratory agreement; CVᵢ, average laboratory imprecision; and BIAS, overall average bias.

2 Work performed as a part of the Special Project "Tecnicole Biomediche e Sanitarie," supported by C.N.R. (National Research Council).

Received July 11, 1983; accepted November 14, 1983.
Imprecision-bias plot: This is a plot illustrating the laboratory imprecision and the laboratory bias achieved by each participant during the considered period. The distance from the origin on this plot, which is numerically equal to \( \sqrt{\text{lab. imprecision}^2 + \text{lab. bias}^2} \), is referred to as "analytical reliability," and it corresponds to the percent deviation from the target value attributable to both random and systematic errors.

\[
\text{CV}_T \%, \text{ average between-laboratory agreement} \quad \text{The pooled between-laboratory CV} \quad \text{for all the EQCS samples mailed out in the considered period, CV}_T, \text{ is computed as:}
\]

\[
\text{CV}_T = \sqrt{\frac{\sum CV_j^2}{MN}}
\]

where \( CV_j \) is the coefficient of variation computed from results of all laboratories for the \( j \)-th-EQCS samples. When the values for \( CV_j \) were derived from different numbers of results, a weighted mean was computed by using a formula analogous to that reported in footnote 3. \( \text{CV}_T \) reflects both the laboratory imprecisions and the systematic differences among laboratories.

\( \text{BIAS} \%, \text{ average bias} \quad \text{This is the root mean square of all the laboratory biases, computed as:}
\]

\[
\text{BIAS} = \sqrt{\frac{\sum b_j^2}{MN}}
\]

\( \text{BIAS} \) reflects systematic differences among laboratories mainly deriving from the use of different methods/kits.

\( \text{CV}_L \%, \text{ average imprecision} \quad \text{This is the median of all the laboratory imprecisions. CV}_L \text{ accounts for the dispersion of the results of the laboratories in respect to their own means.}

\( \text{Kit bias} \quad \text{This is the mean of the percent deviations from the consensus mean of all the results reported by the users of the considered kit.}

\( \text{Kit imprecision} \quad \text{This is the pooled CV computed from all results reported by the users of the considered kit for unidentified replicate EQCS samples sent in different dispatches.}

In the computation of both the laboratory bias and the kit bias, the consensus mean was taken as the target value, because no authoritative reference method has as yet been developed for these assays. The validity of the consensus mean was checked by sending to the participants nine EQCS samples prepared from sera stripped of analyte and then supplemented with spectrophotometrically measured amounts of \( T_3 \) and \( T_4 \) (purchased from Henning, Berlin, F.R.G.) (7). The consensus means found for these samples were very close to the expected values [mean recovery 101.3% (SD 4.3%) for \( T_3 \) and 101.8% (SD 5.2%) for \( T_4 \)].

### Table 1. \( T_3 \) and \( T_4 \) Assay Variability during 2.5 Years of EGCS

<table>
<thead>
<tr>
<th>Data on triiodothyronine</th>
<th>Data on thyroxin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jan 80–Jun 80</td>
</tr>
<tr>
<td></td>
<td>Sep 80–Feb 81</td>
</tr>
<tr>
<td></td>
<td>Mar 81–Nov 81</td>
</tr>
<tr>
<td></td>
<td>Dec 81–Jun 82</td>
</tr>
<tr>
<td>No. labs</td>
<td>42</td>
</tr>
<tr>
<td>No. EQCS samples</td>
<td>17</td>
</tr>
<tr>
<td>Av between-lab.</td>
<td>17.0</td>
</tr>
<tr>
<td>agreement (CV)</td>
<td>11.6</td>
</tr>
<tr>
<td>(BIAS %)</td>
<td>11.9 (11.9)</td>
</tr>
<tr>
<td>Average imprecision</td>
<td>11.1 (10.4)</td>
</tr>
<tr>
<td>(CVL %)</td>
<td>11.1 (13.2)</td>
</tr>
<tr>
<td></td>
<td>8.8 (8.8)</td>
</tr>
<tr>
<td></td>
<td>9.7 (9.3)</td>
</tr>
<tr>
<td></td>
<td>8.6 (8.6)</td>
</tr>
<tr>
<td>*Values in parentheses are those for the subgroup of laboratories that participated in all the last three EQCS periods.</td>
<td></td>
</tr>
</tbody>
</table>

### Results and Comments

The results reported here were those obtained on 100 EQCS samples sent during the period January 1980–June 1982 in 22 monthly dispatches. Four end-of-period reports were compiled during this period. The concentration of the EQCS samples ranged from 60 to 450 ng/L for \( T_3 \) and from 20 to 170 \( \mu \)g/L for \( T_4 \); the between-laboratory agreement for both assays was found to be approximately independent of concentration except for the low-concentration samples (<90 ng/L for \( T_3 \), <40 \( \mu \)g/L for \( T_4 \)), for which the between-laboratory CVs were higher (5). Results on these latter samples were therefore excluded from the computations. The number of participant laboratories and of EQCS samples in each of the four six-month periods are reported in Table 1 for \( T_3 \) and \( T_4 \), respectively. This same table reports the values for \( \text{CV}_T \), BIAS, and \( \text{CV}_L \), which also are computed, for comparison, for the subgroup of laboratories that participated in all of the last three six-month periods. The values for \( \text{CV}_T \) that we found are similar to those reported by others. In fact, from the data of Wood et al. (2) we computed—for results on samples in the same concentration range as ours—a mean between-laboratory agreement of 19.9% for \( T_3 \) and 13.8% for \( T_4 \). More recently, Shishiba et al. (9) reported data from which a between-laboratory agreement of 12.9% can be derived for a sample containing \( T_4 \) in a concentration of 101 \( \mu \)g/L. During the two-and-a-half years of the EQCS, the average laboratory performance appears to improve slightly for \( T_3 \) and more appreciably for \( T_4 \). The data in Table 1 indicate that, as far as assay for \( T_3 \) is concerned, there was both an improvement in \( \text{CV}_L \) and a decrease in BIAS. For \( T_4 \) assay, the essential improvement was a decrease in the average bias. The statistics obtained only for the subgroup—those participating in all the last three periods—show a similar behavior.

Comparison of the imprecision—bias plots computed from the results for different periods is another approach to making evident the possible changes in average analytical performances of the participant laboratories. Figures 1 and 2 show, for \( T_3 \) and \( T_4 \), the imprecision—bias plots of the second and the fourth six-month periods for the same laboratories that participated in all the latter three periods. It can be seen that, on the average, these data are closer to the origin on the plot for the fourth six-month period. This corresponds to an improved analytical reliability. In fact, the percentage of laboratories assaying \( T_3 \) with a reliability better than 12% is increased from 48 to 61%, and for \( T_3 \) assay the percentage of laboratories displaying a poor reliability (>24%) is decreased from 18 to 5%.

Figure 3 shows data on bias and imprecision for \( T_3 \) and \( T_4 \) during the fourth six-month period for the nine kits most used by participants (from six to 27 laboratories), to-
This shows the comparison between the last and the second period; the participating laboratories during the first period were too few (n = 42). Closed circles represent the mean performance of each participant; open circles have been used for coincident performances of two or more participants. The inset tables report the percentage of laboratories for which the analytical reliability (distance from the origin) fell within the indicated ranges. One outlier (bias = 20%, imprecision = 50%) deleted from the right plot.

This analysis points out that, on average, the reliability of the T₄ kits is better than that of the T₃ kits. In fact, four T₄ kits exhibit a reliability of about 10% or better, while no T₃ kit does this well. The relatively poorer performances of the T₃ kits in comparison with T₄ kits seem ascribable more to the presence of larger biases than to poorer precision. The four kits in which the bound/free separation is carried out by the use of antibody-coated tubes (BDI, CLA, DPC1, SPA, see legend to Figure 3) exhibit performances as discrepant as those found for kits based on different separation techniques. This suggests that analytical reliability does not directly reflect different methodological approaches but rather it depends on the technology of kit production and its control.

Discussion

These data confirm the usefulness of setting up EQCS and of continuing them for long periods. Interlaboratory surveys seem to be the only means by which one can reliably and quantitatively evaluate the relative performance of a radioimmunoassay. The average performance derived from EQCS results is useful to the respective participant laboratories as a reference point against which to compare their own performance. Participation in an EQCS must not be considered as a substitute of the internal quality-control procedures carried out within each laboratory, because such procedures allow the daily evaluation of reliability.

The results of EQCS make it possible to inter-compare performances of different assays (in this paper we compared assay for T₃ with assay for T₄) and to monitor possible trends in their average performances. The data reported here demonstrate an improvement in assays for T₃ and T₄ during the last EQCS period. Although this change is a relatively small one, we conclude that the average performances of these assays is good in comparison with that attained by other radioimmunoassays. Possibly this improvement can be partly ascribed to participation in the EQCS.

Another remarkable feature of the EQCS is to provide some basis for an evaluation of the available methods/kits. In fact, almost all laboratories now use kits for radioimmunoassays, and these are to be selected on the basis of analytical reliability rather than their practicability only. On the other hand, the single user can not easily perform, within his own laboratory, an experimental evaluation of the ever-increasing number of such commercially available
kits. Therefore EQCS reports represent a valuable aid, because evaluation of kit performance is continuously updated and obtained from a large number of results produced in different laboratories. Such evaluations also represent a useful source of information and a stimulus for kit manufacturers to improve the reliability of their products.

References