Report of the IFCC Working Group for Standardization of Thyroid Function Tests; Part 3: Total Thyroxine and Total Triiodothyronine

Linda M. Thienpont,1* Katleen Van Uytfanghe,1 Graham Beastall,2 James D. Faix,3 Tamio Ieiri,4 W. Greg Miller,5 Jerald C. Nelson,6 Catherine Ronin,7 H. Alec Ross,8 Jos H. Thijssen,9 and Brigitte Toussaint,10 for the IFCC Working Group on Standardization of Thyroid Function Tests

BACKGROUND: Because total thyroid hormone testing is performed on many automated clinical chemistry instruments, the IFCC Scientific Division commissioned the Working Group for Standardization of Thyroid Function Tests to include total thyroxine (TT4) and total triiodothyronine (TT3) in its standardization efforts.

METHODS: Existing SI-traceable reference measurement procedures (RMPs) were used to assign TT4 and TT3 values to 40 single-donor serum samples for subsequent use in a method comparison study with 11 TT4 and 12 TT3 immunoassays. Data from comparison of each immunoassay with the RMPs provided a basis for mathematical assay recalibration.

RESULTS: Seven TT4 assays had a mean bias within 10% of the RMP, but 2 deviated by an average of −12% and another 2 by +17%. All TT3 assays showed positive biases, 4 within and 8 outside 10%, up to 32%. Mathematical recalibration effectively eliminated assay-specific biases, but sample-related effects remained, particularly for TT3. Correlation coefficients with the RMPs ranged from 0.82 to 0.97 for TT4 and from 0.32 to 0.92 for TT3. The within-run and total imprecision ranges for TT4 were 1.4% to 9.1% and 3.0% to 9.4%, respectively, and for TT3 2.1% to 7.8% and 2.8% to 12.7%, respectively. Approximately one-half of the assays matched the internal QC targets within approximately 5%; however, we observed within-run drifts/shifts.

CONCLUSIONS: The study showed that of the assays we examined, only 4 TT4 but the majority of the TT3 assays needed establishment of calibration traceability to the existing RMPs. Most assays performed well, but some would benefit from improved precision, within-run stability, and between-run consistency.

Despite the well-known clinical rationale for measuring free rather than total thyroid hormone, limitations in current methods for the former necessitate retaining the measurements of total thyroxine (TT4)11 and total triiodothyronine (TT3) in the clinical laboratory (1, 2). For this reason, the mission statement of the IFCC Working Group for Standardization of Thyroid Function Test (WG-STFT) included the standardization of total thyroid hormone measurements. The achievement of this task was considered straightforward, in view of the availability of complete SI-traceable reference measurement systems (3–5). The systems comprise T4/T3 primary calibrators and trueness-based isotope dilution–liquid chromatography/tandem mass spectrometry (ID-LC/tandem MS) reference measurement procedures (RMPs). In addition, the feasibility of standardization (establishing SI traceability) of TT4/TT3 routine assays has already been demonstrated (6). The reference measurement systems (inclusive reference measurement services) are listed in the database of the Joint Committee for Traceability in Laboratory Medicine (7).

1 Laboratory for Analytical Chemistry, Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium; 2 Department of Clinical Biochemistry, Royal Infirmary, Glasgow, UK; 3 Stanford Medical School, Palo Alto, CA; 4 Department of Clinical Laboratory Medicine, Dokkyo University School of Medicine, Mibu, Tochigi, Japan; 5 Department of Pathology, Virginia Commonwealth University, Richmond, VA; 6 Department of Medicine, Loma Linda University School of Medicine, Loma Linda, CA; 7 Laboratoire de Neurobiologie, Centre National de la Recherche Scientifique, Marseille, France; 8 Department of Chemical Endocrinology, Radboud University Nijmegen Medical Center, Nijmegen, the Netherlands; 9 Laboratory of Endocrinology, University Medical Center Utrecht, Utrecht, the Netherlands; 10 Institute for Reference Materials and Measurements (IRMM), Joint Research Centre, European Commission, Geel, Belgium.

* Address correspondence to this author at: Faculteit Farmaceutische Wetenschappen, Harelbekestraat 72, B-9000 Gent, Belgium. E-mail Linda.thienpont@ugent.be.

Received November 6, 2009; accepted February 23, 2010. Previously published online at DOI: 10.1373/clinchem.2009.140228

11 Nonstandard abbreviations: TT4, total thyroxine; TT3, total triiodothyronine; WG-STFT, Working Group for Standardization of Thyroid Function Tests; ID-LC/tandem MS, isotope dilution–liquid chromatography/tandem mass spectrometry; RMP, reference measurement procedure; IQC, internal QC; CVw, within-run CV; CVt, total CV.
Materials and Methods

SERUM PANEL
We selected donors by screening a group of 200 apparently healthy individuals (Solomon Park Research Laboratories, Kirkland, WA, www.solomonpark.com) to obtain a panel of 40 TT4 and 37 TT3 single-donor serum samples with a maximum range of free T4 values. The decision to limit the number of samples to 40 was based on economic considerations (because RMPs were time-consuming and expensive). The panel was produced according to the C37-A protocol from the CLSI (except that the serum was not filtered or pooled) (10). Details, including an overview of donor characteristics, can be found in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol56/issue6 and is in common for all 3 parts of this report.

ASSIGNMENT OF TARGET VALUES
TT4 and TT3 concentrations of the serum samples with ID-LC tandem MS were measured at Ghent University as described elsewhere (11–13). In the current study, all measurements were done with an Acquity TQD MS coupled to an Acquity UPLC system (Waters). The RMPs were performed according to predefined quality specifications (14). The measurements were performed at least in duplicate (TT4) or triplicate (TT3) in independent runs. For these protocols, the expanded uncertainty values for the measurements were estimated to be 3.3% for TT4 and 2.8% for TT3. For details and modifications, see the online common Data Supplement (shared by all 3 parts of this WG-STFT report) and the Data Supplement to this article.

METHOD COMPARISON STUDY
The study protocol required duplicate measurements in 3 runs with preferably 3 different combinations of instruments/calibrator lots/reagent lots, controlled with manufacturers’ internal QC (IQC) materials (see Table 1 for manufacturer conformance to intended design).

STUDY PARTICIPANTS
In total, 8 diagnostic companies participated in the study, with 11 immunoassays for TT4 and 12 for TT3, all performed on automated multianalyte platforms. For details on the platforms and assays used, see the online Data Supplement for this article.

DATA ANALYSIS
Data analysis was preceded by inspection of the results plotted in a scatter diagram. This inspection led to exclusion from regression and correlation analysis of samples with the highest TT4 and the lowest TT4 and TT3 concentrations, because these samples would be too influential. The remaining concentration range is referred to as the reduced range.

Data analysis started with the assessment of the current TT4 and TT3 standardization status. This process was done in 3 different ways: First, each assay mean TT4 and TT3 concentration for the samples (covering the reduced range) was calculated and graphically compared to the mean of the RMP ± 10% (mean values in the y axis, codes of the assays in the x axis). This graph also included the 1-sided 95% CIs of the mean of the assays. These values were calculated from the SDs obtained by ANOVA from the means of the different sets of single results (2 sets per run, 3 runs in total) and the respective degrees of freedom from the Satterthwaite approximation. Second, Deming regression analysis was performed, comparing each assay mean from 6 replicates per sample to the mean of the RMP results. The outcome was displayed in a scatter plot (data for the RMP in the y axis, for the tested assays in the x axis), showing the regression equations/lines of the assays that provided the most discrepant results. Note that although the regression analysis was done only for the reduced data range, the lines were extended to cover the complete concentration range. Third, the percentage difference of each assay mean per sample from the mean by the RMP (complete range) was calculated and displayed.

Improvement of standardization was investigated by mathematical recalibration of each assay’s results (mean of 6 replicates; reduced range). Recalibration was performed by use of the reverse relationship between the method comparison data (x = means of individual assay results, y = mean of the RMP results). The equation used to fit the x and y data was obtained by use of Deming regression. Note, however, that the data sets, particularly the percentage differences of the mean of the assays from the mean of the RMP, were tested for outliers with the Grubbs test (95% probability level) before the reverse regression analysis was per-
formed. Although outliers were removed from regression analysis, they were not excluded from recalibration. Inclusion of the outliers was necessary to allow comparison of the quality of performance of the assays, e.g., in the recalibrated difference plots, in which the emphasis was on the number of differences outside the total error limits (see below).

Various techniques were used for evaluation of the quality of performance of the assays. The $r^2$ values, obtained from Pearson correlation analysis for each set of singlicate measurements (note that each assay produced 6 sets of singlicate results in total), were averaged and used to rank the assays according to descending correlation as a measure of scatter in the results. The most representative set of singlicate data for each assay was chosen based on its $r^2$ value being closest to the mean $r^2$. As another measure of assay performance, we used the difference in concentration units ($y$) of the most representative set of singlicate results, recalibrated by using the same procedure described before, plotted against the mean of the RMP ($x$). The number of differences outside the total error limits was used to estimate the influence of variation attributable to assay imprecision and sample-related effects (the combined random error components). The total error limits were based on the allowable total error from the biologic variation concept (15) (total error goal = 7.0% for TT4 and 12.0% for TT3, from desirable imprecision = 2.5%.

### Table 1. Measurement design, imprecision and IQC data (sorted by increasing CVt).

<table>
<thead>
<tr>
<th>Code</th>
<th>Run information</th>
<th>CV, %</th>
<th>Deviation IQC target, %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Instrument no.</td>
<td>Calibration no.</td>
<td>Reagent no.</td>
</tr>
<tr>
<td><strong>TT4 assays</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>K</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>P</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>F</td>
<td>3</td>
<td>1c</td>
<td>3</td>
</tr>
<tr>
<td>M</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>L</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>H</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td><strong>TT3 assays</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>J</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>K</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>C*</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>I*</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>F</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>L</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>N</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>M</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

* Per convention in statistics, CVw was taken because CVt < CVw.

b Different targets.

c Two calibrators.

d Mean of range.

* Not applicable.

f Two runs.
for TT4 and 4.4% for TT3; desirable bias = 3.0% for TT4 and 4.8% for TT3); however, the limits were expanded to 8% for TT4 and 13% for TT3 for the imprecision of the respective RMPs.

The general quality of performance of the assays was reflected by attributes such as imprecision [within run CV (CVw; calculated from the duplicates and pooled for the 3 runs) and total CV (CVt; calculated from the first and second replicate of each run and pooled)], between-run calibration consistency (investigated from the percent difference of the IQC results from the target), stability of runs (documented by plotting of the percent differences of the replicates vs the sample number), and systematic differences between runs (investigated by comparing the percent differences of the means of the different runs).

For Deming regression analysis we used CBstat software (version 5.1, K. Linnet, www.cbstat.com); all other statistical analyses and plotting were done with Microsoft Excel® 2002 (office.microsoft.com).

### Results

#### FOREWORD

The presented results have been communicated in detail with the participating assay manufacturers. To maintain confidentiality, the discussions were always conducted in an anonymous way, with arbitrary alphabetical coding of the assays and with no relation between the codes used to designate the TT4 and TT3 assays. However, manufacturers always knew the code(s) assigned to their own assay(s).

#### STATUS OF STANDARDIZATION AND RECALIBRATION

The outcomes of the TT4 and TT3 method comparison study are graphically summarized in Fig. 1. Fig. 1A shows the individual assay TT4 mean values for the samples covering the reduced concentration range from 53 to 124 nmol/L, as determined by the RMP. Four of the 11 investigated assays exceeded the 10% limits around the mean of the RMP, 2 in the positive and 2 in the negative direction. The scatter plot (Fig. 1B) shows the most extreme Deming regression lines/equations. In the percent-difference plot (Fig. 1C), again the extreme differences are emphasized: the circles and triangles indicate the assays that deviated by an average of −12% and +17% from the RMP, respectively.

Fig. 1D shows that for the reduced TT3 range (1.1–1.9 nmol/L; concentrations determined by the RMP) all mean values of 12 assays were positively biased compared to the mean of the RMP, 4 within and 8 outside the 10% goal. The scatter plot (Fig. 1E) documents the extreme Deming regression equations. The percent-difference plot (Fig. 1F) shows that the differences varied from good agreement with the mean of the RMP to a considerable deviation (compare the circles with an average bias of only +1.0% to the triangles with an average deviation as much as +32%).

The mathematical recalibration of the results for the reduced range was effective in removing the biases of the assays (Fig. 2). The representation of this result can best be appreciated by comparison of the location of the circles and triangles in Fig. 2 with those in Fig. 1, C and F. However, note that in the plot for TT3 (Fig. 2B), 10 samples, indicated by squares, document excessive combined random error components in 3 assays (A, F, and J).

With regard to the number of outliers excluded from regression analysis, there were 5 for TT4, i.e., 1 in each of 3 assays (C, G, and M) and 2 in assay K. For TT3, 6 outliers were eliminated, 1 for assay J and 5 for A.

#### ASSAY QUALITY

**Performance compared to the RMP (data in the reduced range).** The assays were ranked according to decreasing $r^2$ values as a measure of scatter in results compared to the RMP (for details, see online Supplemental Tables 1 and 2). For TT4, the highest and lowest correlating assays had $r^2$ values of >0.95 and <0.83, respectively. For TT3 the highest $r^2$ value was 0.92, the lowest values ranged between 0.51 and 0.32.

With regard to the quality of performance as derived from the difference plots, Fig. 3 (top, TT4) shows the assays with the highest ($r^2 = 0.97$), moderate ($r^2 = 0.94$), and poorest correlation ($r^2 = 0.82$), having 1, 4, and 12 differences outside the total error limits of ±8%, respectively. Note that outlier testing before recalibration resulted in elimination of 3 values in 2 assays (1 for H; 2 for K).

Fig. 3 (bottom, TT3) shows analogous plots for the assays with the highest (0.92), moderate (0.79), and lowest (0.32) mean $r^2$ values having 0, 1, and 7 differences outside the total error limits of ±13%, respectively. Note that the poorest correlating assay was the one for which excessive combined random error effects were observed. For TT3, 9 outliers were identified in 4 assays (1 for E and J; 3 for F; 4 for A).

**General performance characteristics.** The imprecision data derived from the study-defined protocol for measurement of the patients’ samples are summarized in Table 1. The data are sorted by increasing CVt values. For some TT4 assays, the CVw is used in the CVt column, because CVt < CVw (a commonly used convention in statistics because CVt < CVw by chance). Note that Table 1 also lists the number of instruments, calibrators, and reagent lots used. The majority of manufacturers used 2 or 3 different reagent lots in the 3 runs.
Fig. 1. (A), (TT4); (D) (TT3): assay means (±1-sided 95% CIs) vs the mean by the RMPs. The x axis gives the codes of the different assays, the dotted lines represent the mean of the RMP ±10%. For the assays differing >10% from the mean of the RMP, the numerical value of the mean is listed. (B), (TT4); (E), (TT3): scatter plot (x = mean of the RMP, y = mean of singlicate results per assay) with indication of the line of equality (dotted) and the most extreme Deming regression lines/equations. The results for the most deviating assays are indicated by circles and triangles; all other assays are indicated with the same symbol, X. (C), (TT4); (F), (TT3): percent-difference plot with indication of the strongest negatively (circles) and positively (triangles) biased assays. Note that (B), (C), (E), and (F) are extended to show the complete range (10–221 nmol/L for TT4, 0.6–1.9 nmol/L for TT3).
Note, however, that because of these experimental differences between manufacturers, caution is necessary when comparing the values for total imprecision. Most companies used 3 different IQC samples; however, some used only 1 material. With the exception of a few assays, they used the same target values for different reagent and calibrator lots. The IQC targets were matched within a limit of approximately 5% by several assays, whereas other assays showed considerable biases over all controls or at certain concentrations.

Shifts and drifts in patient sample results within a run were seen in 9 TT4 and 7 TT3 assays, specifically, after the measurement of the mid-run IQC samples (measurement sequence: start IQC, samples 1 to 40 for TT4 or 1 to 37 for TT3; mid-run IQC, samples 40 or 37 to 1; end IQC, see the online Data Supplement). For TT4, 9 assays showed shifts between replicates: 2 assays in all 3 runs (shifts in the order of 10%–12% and 2.5%–5%, respectively), 4 assays in 2 runs (2.5%–10%, with 1 assay also showing a drift in 1 run), and 3 assays in only 1 run.

For TT3, similar shifts between replicates (varying between 2.5% and 12%) were observed for 7 assays (for details, see online Supplemental Fig. 3 for TT4 and Supplemental Fig. 6 for TT3). Systematic differences between the 2 most deviating runs were <10% for all TT4 assays, whereas for TT3 they were >10% for 3 assays (E, 11.8%; L, 12.3%; and M, 21%). The systematic differences had influence on the CVt, in particular for the TT3 assays L and M (see Table 1).

Discussion

In total, 11 TT4 and 12 TT3 immunoassays were validated in a method comparison study against trueness-based ID-MS RMPs. The agreement of the immunoassay results with the RMPs was assessed against limits of ±10% (see Fig. 1, A and D). For TT4, the majority of the assays met the 10% goal; however, 4 required recalibration. In contrast, 8 of the 12 examined TT3 assays had their means outside the ±10% limit, and thus needed recalibration. This contrast between the current standardization status of TT4 and TT3 assays was also shown in the scatter and percent-difference plots. For TT4, the symmetrical location of the extreme Deming regression lines along the line of equality (Fig. 1B) and of the percent differences along the 0 difference line (Fig. 1C) demonstrate the generally good agreement of the immunoassays with the RMP. In contrast, the equivalent TT3 plots (Fig. 1, E and F) show that all assays were positively biased up to 32%. In view of the tight systematic error limits derived from the biological variation of TT4 (3%) and TT3 (4.8%) (15), the empirical goal of 10% may be lowered.

To demonstrate the clinical impact of the extent of standardization differences among the most discrepant assays, we used the generated Deming regression equations to calculate (Fig. 1, B and E) that TT4 concentrations of 53–124 nmol/L measured by use of the RMPs translated to TT4 concentrations of 49–106 nmol/L measured by assay M but to 57–151 nmol/L by assay H. TT3 concentrations of 1.1–1.9 nmol/L by the RMPs translated to concentrations of 1.0–2.0 nmol/L by assay C and 1.4–2.5 nmol/L by assay K. Considerable overlap in ranges was noted among the assays.

Compared with previously performed large-scale studies, the overall picture for TT4 revealed in this study was similar (6, 16). For TT3, however, results of only 1 study showed a similar positive bias for the majority of assays (6), whereas results of another study demonstrated a negative bias (16). It should be men-
tioned, however, that the ID-MS comparison method used in the latter study was not claimed to be a RMP (nor was it listed in the database of the Joint Committee for Traceability in Laboratory Medicine) (7).

By mathematical recalibration we were able to correct the systematic deviations from the RMPs for both TT4 and TT3 (Fig. 2). In the case of TT3, such recalibration would decrease the all-assay mean by approximately 13%.

In addition, after recalibration, a dispersion in measured values remained owing to the combined within-run imprecision and sample-related effects. The effects were excessive for 3 TT3 assays (squares in Fig. 2B) that had grossly increased results (ranging from +20% to +74%) for 10 samples. The reason for this finding is unknown (see similar problems described for free T3 in part 2). Remeasurement by the RMP of the sample aliquots used for the assays of concern demonstrated no increased TT3 values, nor was there an association with other biochemical results (see Table 1 in the online common Data Supplement). It remains to be investigated whether the observed effect may be due to the clotting process used in the C37-A protocol (for a short outline, see the online Data Supplement; for details see the C37-A guideline) (10). In some cases, depending on the outcome, this protocol, which was originally developed for cholesterol analysis, may require verification for T3 (total and free). Alternatively, the 3 concerned manufacturers may wish to investigate the robustness of their assay design, because problems occurred only with their TT3 assays (1 assay was also problematic for free T3, see part 2).

The quality of performance as derived by correlation analysis and the number of differences outside the biological total error limits after mathematical recalibration were fairly good for the majority of the examined TT4 and TT3 assays. However, 4 TT4 assays (L, H, B, and M in online Supplemental Fig. 2) showed a higher number of differences exceeding the 8% goal. The same applies for 3 TT3 assays (see J, F, and A in online Supplemental Fig. 5), showing grossly positive differences outside the 13% goal (compare the scale of the y axis in the difference plots with that of the other assays). The aforementioned assays may require improvement because assays are exposed to more sources of error under the conditions of routine clinical laboratory testing, and the nature of analyzed samples is diverse (samples from sick patients compared to the apparently healthy individuals in this study).

With regard to precision, the ranges for CVw and CVt observed for the TT4 assays were comparable to those reported in previous studies (17, 18). None of the TT4 assays and only a few of the TT3 assays met the biological CVt goals for patient monitoring, i.e., 2.5% and 4.4%, respectively (15), suggesting that improvements in assay precision are needed. Although this conclusion is concordant with that of the TT4 study by Steele et al. (18), it is in contrast with that of Roberts et
al., who found that all investigated TT4 assays complied with the biological imprecision goal (17). In the latter study, however, a different percent goal than that used by the WG was adapted and may have contributed to this discrepancy (1).

With regard to the within-run stability of the assays, the number of TT4 and TT3 assays that showed shifts and drifts was rather high. These results indicate that there is a need for improvement of the assays in this respect. In addition, the IQC performance in terms of between-run consistency of calibration could be improved for those assays with considerable biases, because the biological bias limits of both TT4 and TT3 are small (15). For this reason, as we indicated earlier, a 5% standardization goal for these 2 analytes is desirable.

In conclusion, comparison of results of 11 TT4 and 12 TT3 immunoassays to RMP results for a panel of human samples showed that standardization of TT4 measurements would be beneficial for 4 assays, whereas for TT3 measurements the majority of assays need standardization. The study results showed that a majority of the examined assays had acceptable quality of performance for measurement of a set of samples from nondiseased individuals. However, some assays would benefit from improvements in precision, within-run stability, and between-run consistency. Although the potential effectiveness of standardization was demonstrated only by mathematical recalibration of the data, the IFCC recommends that manufacturers take this study as an incentive to comply with the scientific and legal requirement of establishing calibration traceability of their assays to the existing TT4 and TT3 reference measurement systems (7). As described in the ISO 17511 standard, establishment of calibration traceability can be accomplished by performing an analogous method comparison study, but with inclusion of master calibrators and value reassignment (19).

Author Contributions: All authors contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures of Potential Conflicts of Interest: Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:

Employment or Leadership: W.G. Miller, AACC and CLSI.

Consultant or Advisory Role: None declared.

Stock Ownership: None declared.

Honoraria: None declared.

Research Funding: None declared.

Expert Testimony: None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

Acknowledgments: The Chair of the IFCC WG-STFT (L.M. Thienpont) is grateful to (names in alphabetical order): M. Banci and G. Markowitz (DiaSorin S.p.A, Italy); G. Baudino (bioMérieux, France); L. Dodge (Mayo Foundation, USA); R. Janzen, L. Mansbach, and Y. Xue (Siemens Healthcare Diagnostics, USA); M. Kasai (TOSOH Corp., Japan); M. Minihan (Olympus Life Science Europa GmbH, Ireland); D. Montague (Ortho-Clinical Diagnostics, UK); F. Quinn (Abbott Diagnostics, USA); A. Rockwood and B. Yue (ARUP Laboratories, USA); M. Rottmann (Roche Diagnostics GmbH, Germany); C. Walton and T. Brinkmann (Beckman Coulter Inc., USA and Europe), acting as representative on behalf of their organizations. Their efforts to review and provide comments on the manuscript are highly appreciated. The 9 organizations sponsored (all contributed equally) the study in terms of sample procurement and funding for the assignment with the reference method values. L.M. Thienpont is also indebted to D. Stockl for the graphical and statistical handling of the data and the expert advice in the editing of the report.

References


