Clinical Chemistry 51:1
161–168 (2005)

Endocrinology and Metabolism

Metrologic Traceability of Total Thyroxine Measurements in Human Serum: Efforts to Establish a Network of Reference Measurement Laboratories

Linda M. Thienpont,[1*] Katleen Van Uytfanghe,[1] John Marriot,[2] Peter Stokes,[2] Lothar Siekmann,[3] Anja Kessler,[3] David Bunk,[4] and Susan Tai[4]

Background: **Assuring/demonstrating metrologic traceability of in vitro diagnostics necessitates the availability of measurand-specific reference measurement systems (RMSs) and the possibility for industry to work with competent reference measurement laboratories (RMLs). Here we report the results of a European project to investigate the feasibility of developing a RMS for serum total thyroxine.**

Methods: **Four candidate RMLs (cRMLs) developed/implemented variants of a candidate reference measurement procedure (cRMP) based on isotope dilution–liquid chromatography–mass spectrometry. The sole constraint implemented was calibration with a common thyroxine primary calibrator. The RMPs were externally validated and assessed for comparability in round-robin trials using common samples, i.e., 5 lyophilized and 33 frozen native sera. At the same time, the performance of the cRMLs organized in a network was assessed. For uniform external quality assessment, common performance specifications were agreed on.**

Results: **All cRMLs performed the cRMPs with fulfillment of the predefined specifications: total and between-laboratory CVs ≤2.0% and 2.5%, respectively, and a systematic deviation ≤0.9%, estimated with a target assigned from the mean of means obtained by the cRMLs. The mean expanded uncertainty for value assignment to the native sera was 2.1%.**

Conclusions: **A network of cRMLs, with externally conformed competence to properly perform RMPs, has been established. Performance specifications were defined and will form the basis for admittance of new network members. A serum panel, successfully targeted during the validation process, is available for split-sample measurements with commercial routine measurement procedures. The model can now be used for other measurands for which traceability to the Système International d'Unités is needed.**

The concept of trueness-based standardization of measurements in laboratory medicine goes back to pioneering work done in the 1970s *(1–4)*. These activities were pursued at the scientific level in ambitious standardization projects by authoritative organizations such as the IFCC and the Centers for Disease Control and Prevention *(5–9)*. However, in practice, the standardization process did not receive attention until the European Directive on in vitro diagnostic medical devices (98/79/EC) became effective *(10)*. Indeed, since December 2003, manufacturers of diagnostic systems must demonstrate the metrologic traceability of the values assigned to the calibrators. Providing support to the European legislation, EN/ISO Standard 17511 describes the way to establish traceability of a routine measurement procedure *(11)*, i.e., by applying a reference measurement system (RMS).[5] For chemically

[1] Laboratory for Analytical Chemistry, Faculty of Pharmaceutical Sciences, Ghent University, Gent, Belgium.
[2] Laboratory of the Government Chemist, Teddington, United Kingdom.
[3] Institute for Clinical Biochemistry, University of Bonn, Bonn, Germany.
[4] National Institute of Standards and Technology, Gaithersburg, MD.
*Address correspondence to this author at: Laboratory for Analytical Chemistry, Faculty of Pharmaceutical Sciences, Ghent University, Harelbekestraat 72, 9000 Gent, Belgium. Fax 32-9-264-8198; e-mail linda.thienpont@ugent.be.
Received August 18, 2004; accepted October 27, 2004.
Previously published online at DOI: 10.1373/clinchem.2004.041772

[5] Nonstandard abbreviations: RMS, reference measurement system; (c)RMP, (candidate) reference measurement procedure; JCTLM, Joint Committee for Traceability in Laboratory Medicine; T_4, thyroxine; (c)RML, (candidate) reference measurement laboratory; ID-MS, isotope dilution–mass spectrometry; and IRMM, Institute for Reference Materials and Methods.

well-defined measurands, the system starts with the Système International d'Unités (SI units) and further consists of a multilevel hierarchy of higher order reference materials and reference measurement procedures (RMPs).

Although three other EN/ISO standards, i.e., 15193, 15194, and 15195 *(12–14)*, cover the essential requirements of the elements of such a RMS, the diagnostics industry was initially stymied by how to identify them for each specified measurand; in other words, the legislation was in place but not the tools by which to comply. In this respect, two initiatives were taken. In the first initiative, the International Committee of Weights and Measures, the IFCC, and the International Laboratory Accreditation Cooperation agreed to cooperate toward the establishment of a Joint Committee for Traceability in Laboratory Medicine (JCTLM) *(15)*, with the mission statement to categorize higher order reference materials and RMPs and to establish criteria and processes assessing the competencies required for laboratories performing the RMPs. In the second initiative, the European Commission funded a project with the reference G6RD-CT-2001-00587, aiming at a RMS for thyroid hormone measurements as a model *(16)*. The project members consisted of expert laboratories in Belgium (Ghent University), the United Kingdom (Laboratory of the Government Chemist), Germany (University of Bonn), and the United States (NIST). The project's topic was chosen with great care via a worldwide survey conducted by the Scientific Division of IFCC to identify "measurands of priority" in laboratory medicine. As a result, it was considered that establishing metrologically traceable measurements of thyroid hormones would be of great support to the correct diagnosis and monitoring of therapeutic measures. With respect to the routine measurement of thyroxine (T_4), the free rather than the total hormone fraction was promoted as a prime diagnostic measurand *(17)*. However, the controversial analytical principle of some of the commercial immunoassays for free T_4 seems to have undermined the anticipated diagnostic potential *(18–20)*, with the consequence that even the most recently developed automated test systems still include total T_4. It was therefore decided that establishing SI traceability of the measurement of total T_4 would be the first part of the European project.

Here we describe our investigation of the feasibility of developing such a RMS. Although a primary T_4 calibrator was lacking initially, such a calibrator has since been established. A comprehensive description of the certification process for this calibrator according to the EN/ISO 15194 guidelines will be given elsewhere. We therefore report here only the development of RMPs and reference materials. We also deal with the project's efforts to organize the candidate reference measurement laboratories (cRMLs) in a network, which required establishment of and agreement on a process to assess proper performance of the members of the project group at the inception of the network and of candidate members subsequently applying for admittance.

Materials and Methods

RMPs

Isotope dilution–mass spectrometry (ID-MS) coupled to gas-liquid chromatography or HPLC was selected as the basis for the candidate RMPs (cRMPs) for serum total T_4. The project leader provided all participating laboratories with a common T_4 standard material and $^{13}C_6$- and $^{13}C_9$-labeled T_4 (Service de Chimie et Biochimie Appliquées, Faculté Polytechnique de Mons) for calibration and identification, respectively. The $^{13}C_6$- and $^{13}C_9$-T_4 materials contained <0.7% and 0.0% $^{13}C_0$-T_4, respectively. The unlabeled T_4 standard material was from a commercial source but was intended to become a primary reference material after certification of its purity under the coordination of the Institute for Reference Materials and Methods (IRMM) from the Joint Research Centre (Geel, Belgium). Because the certification process was not completed at the time of the development of the cRMPs, a purity of 100% was assumed. Apart from the instruction to calibrate only with the T_4 candidate primary calibrator, no further constraints were applied with respect to the design of the cRMPs. The laboratories at Ghent University and NIST used their previously developed and recently modified procedures based on isotope dilution–liquid chromatography–tandem MS with electrospray ionization *(21–23)*. The laboratory at the University of Bonn recently had transferred its isotope dilution–gas-liquid chromatography–MS cRMP *(24)* into an isotope dilution–liquid chromatography–electrospray ionization, single-stage MS variant (see the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol51/issue1/), whereas the laboratory at the Laboratory of the Government Chemist started from scratch *(25)*. Additional details, including information on liquid chromatography–MS instrumentation, measurement conditions, and calibration *(26, 27)*, are given in Tables 1 and 2 of the online Data Supplement.

MEASUREMENT PROTOCOL AND PERFORMANCE SPECIFICATIONS

Before starting the implementation and/or development of the cRMPs, the project laboratories agreed on a validation strategy. They adopted the specifications for the precision of a total T_4 RMP as published previously by a European Working Group, i.e., a maximum total CV of 2.0% for a measurement protocol consisting of duplicate analyses on three independent occasions *(28)*. In addition, it was specified that the difference between duplicates in a measurement series should not exceed 2.5% and that the within-run CV should be ≤1.5%. During the meetings held to evaluate progress of the work, the laboratories were required to disclose all encountered problems and to show representative reconstructed ion chromatograms. A laboratory could claim to have a cRMP available only after sufficient internal assessment to show that the predefined performance specifications were met. At that

stage, assessment of the comparability among the different cRMPs could be started.

reference materials

Two types of candidate reference materials were investigated to conduct the external validation of the cRMPs (see below): 5 lyophilized serum materials with different T_4 concentrations, purchased from the German Society of Clinical Chemistry and Laboratory Medicine (Bonn, Germany; ~500 vials containing 3 mL of serum for each concentration) and 33 frozen off-the-clot sera, each obtained from a single blood donation by an apparently healthy male or female donor. The frozen sera were purchased from Scantibodies Laboratory, Inc. All blood collections were carried out according to accepted protocols at blood banks regulated by the US Food and Drug Administration. One hour after collection and coagulation of the blood at room temperature, serum was isolated by centrifugation. Only those units found negative for the presence of antibodies to HIV I/II, syphilis, and hepatitis B surface antigen were shipped to Scantibodies (refrigerated between 2 and 8 °C) for further processing. T_4 standard material was added to 3 of the units to achieve hyperthyroid concentrations; the remaining 30 units, however, were native. No preservatives were added, but the sera were filtered through 0.22 μm filters to assure sterility and were fractionated into 1-mL portions in polypropylene vials. For each blood donation, a maximum of ~150 aliquots were available. After aliquoting, they were immediately stored at −70 °C and shipped to the participating laboratories on dry ice with continued storage at −70 °C.

stability study of the reference materials

The candidate reference materials were subjected to a stability study. The protocol was from the IRMM and consisted of a short- and long-term isochronous study for the lyophilized sera, i.e., storage for 2 weeks at 40 and 4 °C, with −20 °C as the reference temperature, and for 12 and 18 months at 4 and −20 °C, with −70 °C as the reference temperature. The frozen sera were tested after long-term storage at −70 °C (also during 12 and 18 months) vs −160 °C as the reference point. In each study, two measurements per time point were to be performed by ID-MS. A one-way ANOVA was used to evaluate the data.

external validation of the performance of the cRMPs

External validation of the cRMPs was done in several round-robin trials with blind measurement of common samples. In the first and second intercomparisons, the laboratories were asked to analyze the above-described lyophilized sera in duplicate on three independent occasions. Results from all individual measurements were assessed for fulfillment of the predefined specifications. Subsequently, for each set of results the arithmetic mean, SD, and CV were calculated, and afterward, the mean of means and the within- and between-laboratory CVs were calculated. The latter was done via a one-way ANOVA. In deciding whether there was sufficient comparability of the RMPs, the limit for the between-laboratory CV was predefined as 2.5%. For the final validation of the cRMPs, the panel of 33 frozen sera was measured. The measurements were divided among the cRMLs in such a way that finally three sets of duplicate results would be available per serum. Note that for these measurements, duplicates meant measurement of each serum in two independent measurement series.

validation of the performance of the cRMLs in a network

The round-robin trials were also used to assess the performance of the cRMLs in a network. The assessment criteria were the above-mentioned precision specifications (total and between-laboratory CV) extended with a limit for systematic deviation of 0.9% *(28)*. The latter was estimated from repeated measurements of the two lyophilized serum materials targeted in the second round-robin trial by the four cRMLs. In addition, the capability to deliver measurement results before a deadline was a criterion to judge adequate performance of a network member.

calculation of the uncertainty of values assigned to frozen off-the-clot sera

The measurements of the frozen samples by the four cRMLs enabled the assignment of total T_4 target values and the estimation of individual measurement uncertainties. The latter was done in compliance with the *Guide to the Expression of Uncertainty in Measurement* (GUM) *(29)*. Because the uncertainty of the purity of the calibrator was not yet known, it was provisionally set to zero; hence, only the uncertainty attributable to measurement was considered. A single-factor ANOVA gave evidence for the fact that the latter consisted of both the within-laboratory imprecision and the between-laboratory variation. The former was estimated for each cRML as the within-group SD calculated from the individual serum measurement values (duplicates per serum) in a single-factor ANOVA. The within-laboratory SD, divided by the square root of 2, allowed calculation of the uncertainty of each cRML's reported mean serum value (uncertainty expressed in percentage compared with the mean of the mean values for each serum). Combination of these uncertainties for the measurements by the three (occasionally two) laboratories that targeted the same serum (via the square root of the summed squares) gave the combined uncertainty of the serum's assigned value (the mean of the three mean values) attributable to the within-laboratory variation. The uncertainty attributable to the between-laboratory variation was estimated for each serum from the deviation between the highest and lowest reported mean values (also expressed as a percentage). The mean of all 33

deviations was then used to estimate the between-laboratory uncertainty according to the rectangular distribution; hence, a constant value was assumed. The final combined uncertainty was estimated from combination of the within- and between-laboratory uncertainties (quadratic addition and square root). The expanded uncertainty was calculated with a coverage factor based on the Student *t*-distribution (95% confidence level).

Results

For the first round-robin trial, the four individual sets of results (for each set, n = 6) could all be used because the maximum total CV was 1.6% (clearly below the 2.0% specification) and none of the duplicates exceeded the limit of a 2.5% difference. The three lyophilized serum materials measured in that trial had concentrations of 94.7 ± 2.7, 133.5 ± 4.0, and 163.6 ± 4.6 nmol/L (mean of means ± 95% confidence interval), respectively. One-way ANOVA estimated a maximum within-laboratory CV of 1.0% and a maximum between-laboratory CV of 2.0%, which met the preset specifications. The results are represented graphically in Fig. 1.

In the second round-robin trial, the four sets of results could again be accepted for the calculation of the mean of means (± 95% confidence interval), i.e., 96.74 ± 1.06 and 130.0 ± 1.2 nmol/L. In this trial, the maximum within-laboratory CV was estimated to be 0.5%, and the maximum between-laboratory CV was estimated to be 0.6% (Fig. 2).

In the final round-robin trial, the cRMPs were validated for measurement of the frozen, off-the-clot sera. For internal accuracy control of these measurements, the two lyophilized sera of the second round-robin trial and their target values (mean of means) were used. On average, 7–10 measurement series were needed by each cRML to complete the duplicate measurements of the frozen sera. The systematic deviation estimated from these repeated measurements ranged from −0.2% to −0.8%. Note that on the basis of the described internal accuracy control measures, one laboratory decided to withdraw its results for four sera. Through this final validation trial, each serum was assigned a target value from the mean of three sets of duplicates (except for the above-mentioned sera, for which only two sets of duplicates were available). Because it is the intention to use these sera in split-sample measurements with routine immunoassays, the individual concentrations, covering a total T_4 concentration range between ~50 and 300 nmol/L, cannot be published here. The within-laboratory SDs for measurement of the sera by the cRMLs, as estimated from a single-factor ANOVA, were 1.18, 0.96, 0.61, and 0.64 nmol/L, respectively. The mean between-laboratory uncertainty was estimated to be 0.55%. Use of these values in the estimation of the expanded relative uncertainty gave a mean of 2.1% (range, 1.5–2.9%).

One-way ANOVA of the data for the short- and long-term stability assessments of the lyophilized and

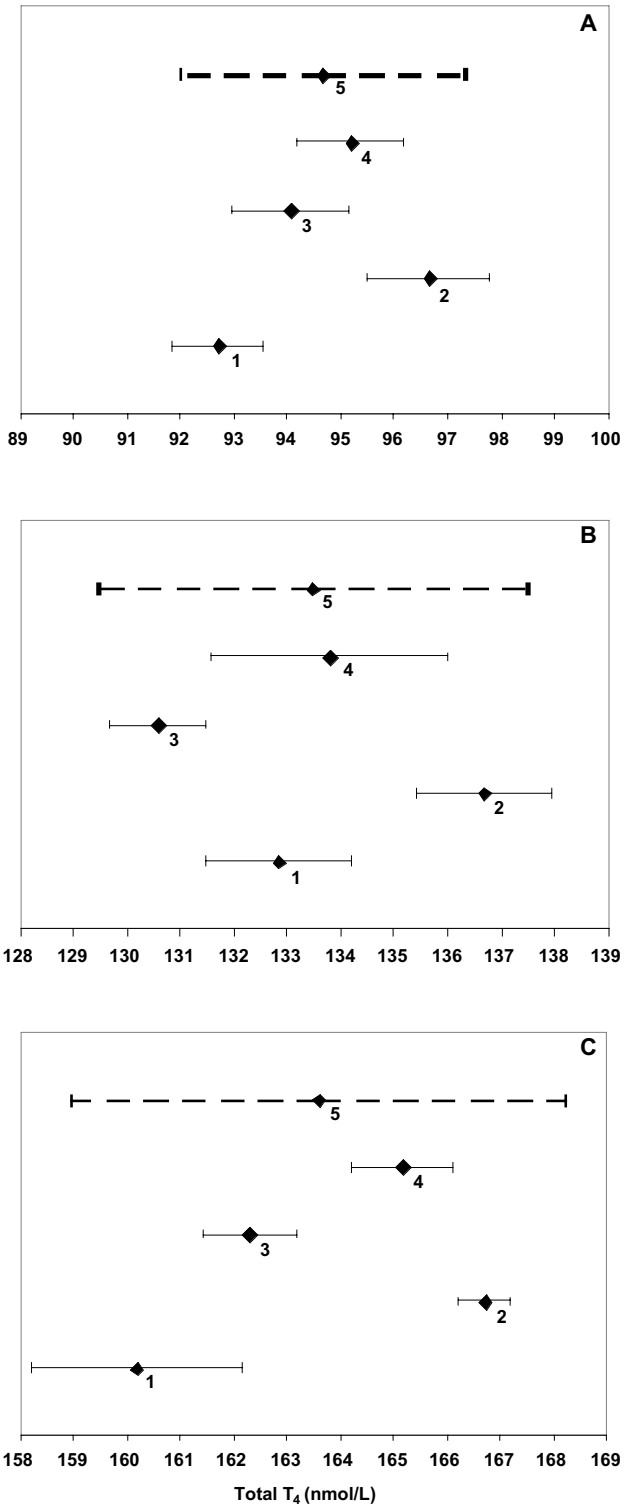

Fig. 1. Bar graphs representing the arithmetic mean of each laboratory's results (*points 1–4*) and the mean of means (*point 5*) from the four accepted sets of results, each with its 95% confidence interval (*error bars*).

The whole is represented for analysis of three lyophilized samples (*A–C*) in round-robin trial 1.

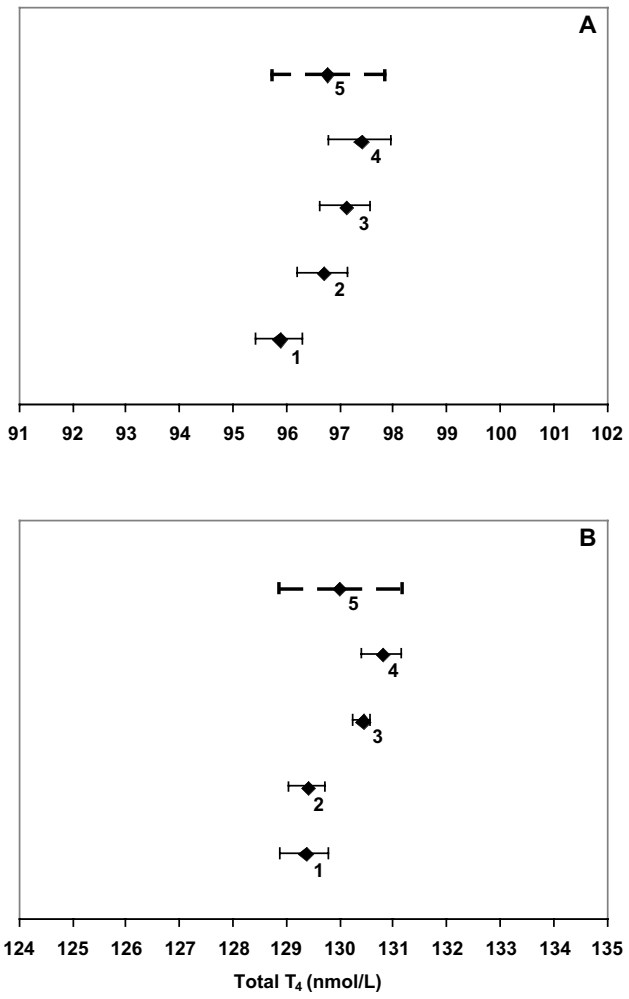

Fig. 2. Bar graphs representing the arithmetic mean of each laboratory's results (*points 1–4*) and the mean of means (*point 5*) from the four accepted set of results, each with its 95% confidence interval (*error bars*).

The whole is represented for analysis of two lyophilized samples (*A* and *B*) in round-robin trial 2.

frozen sera did not reveal a significant difference between the storage conditions ($P > 0.05$).

Discussion

DEVELOPMENT AND/OR IMPLEMENTATION OF cRMPs FOR TOTAL T_4 IN SERUM

It was indisputable that only ID-MS would qualify as the applicable method for the development of a cRMP for total T_4 in serum *(11, 30)*. In addition, the requirements described in EN/ISO Standard 15193 had to be considered *(12)*. In line with the philosophy of the former Bureau Communautaire de Référence of the European Commission, the participating laboratories had the freedom to develop their own variant of an ID-MS cRMP. Although this was inconsistent with the concept of transferability testing of a cRMP, as for example, adopted in the hemoglobin A_{1c} project *(31)*, the decision in the European project is justifiable on the basis that all participating laboratories had sufficient previous experience with cRMPs *(15)* and from the belief that the agreement among results obtained with different cRMPs is the strongest argument for their analytical validity. For this reason, from the start of the project it was clear that round-robin trials would be a key part of the validation process.

REFERENCE MATERIALS

It is obvious from this study that serum-based matrix reference materials are an important tool in the process of validation of newly developed/implemented measurement procedures and/or of internal accuracy assessment for cRMPs. They also play a prime role in establishing the relationship of routine measurement procedures with the SI unit *(11)*. Whereas their commutability is not an issue in combination with matrix-independent RMPs, it is of utmost importance with regard to routine measurement procedures [see, e.g., Ref. *(32)*]. The European project members, representing the diagnostics industry, therefore urged that reference materials with matrices closely matching patient sera be included in this feasibility study. Use of such materials in the traceability chain at the level of the manufacturer's calibrators is not only permissible by the EN/ISO 17511, but is even recommended for validation of metrologically traceable calibration. From this perspective, the project selected authentic, single-donation, native off-the-clot sera with concentrations as evenly distributed as practicable over the whole of the measuring interval for serum total T_4. Because it was considered unethical to collect blood from individuals suffering from hyperthyroidism, T_4 was added to three donations to obtain materials with increased T_4 concentrations. The only manipulations that all serum donations underwent were sterile filtration to avoid bacterial contamination and freezing for logistic reasons.

For the purpose of validation and internal accuracy control of the cRMPs, the project used lyophilized sera. The advantage of this type of material lies in the fact that it is available in large batches and is prepared to last for several years. The rather restricted number of vials purchased for this project (500 per concentration) should not be seen to be in conflict with the latter statement because the current project was only a pilot study investigating the feasibility of establishing a RMS for total T_4. The preparation of a large batch of commonly available lyophilized reference materials may be a future task for institutes such as the IRMM. On the other hand, with respect to the number of 1-mL aliquots of the single-donation sera, the diagnostics industry should realize that these materials are available only in limited stock sizes (the certification process with the cRMLs and the stability study consumed ~90 of the 150 mL that was the maximum obtained per donation).

The isochronous stability study performed for both types of serum-based matrix reference materials allowed us to infer that the sera remain sufficiently stable to last

for several years after certification, if stored under appropriate conditions.

validation of the cRMPs

The results of the external validations (Figs. 1 and 2) demonstrate not only that all laboratories were successful in fulfilling the predefined total CV requirements, but that the results obtained with the four cRMPs agreed well (see the between-laboratory CVs for the first and second round-robin trials). Nevertheless, the progress in intercomparability from the first to the second trial was remarkable (the deviation between the lowest and highest sets of results was reduced from 4.7% to 1.6%). This was most probably attributable to an increase in skill in performing the cRMPs. The outcome of the internal accuracy control further shows that the cRMLs were able to maintain the accuracy of their performance over a longer period with fulfillment of the 0.9% limit for systematic deviation. With respect to the assignment of target values by different cRMLs, the expanded relative uncertainties calculated for the frozen sera demonstrate that this is possible with sufficient reliability. Nevertheless, a significant ANOVA suggested that both the within- and between-laboratory variation contributed to the measurement uncertainty. Generally speaking, they were in the same order of magnitude, but there were sera for which the contribution by the within-laboratory imprecision was negligible compared with the between-laboratory imprecision, whereas the converse was never the case.

The assumption that the within-laboratory SD is constant for the measurement of all sera is justified from the fact that for each serum the same absolute amount of analyte is taken through analysis. Which contribution was the highest depended on the serum: the higher the concentration, the more dominant the between-laboratory effect became. This is because the latter was assumed constant. Although the between-laboratory variation might again be explained by small differences in the skill levels in the individual cRMLs, the most probable cause lies in small systematic differences in the measurement procedures (e.g., degree of sample purification, chromatographic, and MS resolution). The latter effects come to light only when many samples, each with its individual matrix, are measured.

organization of the cRMLs in a network

A premise to the success of the concept of ensuring/demonstrating SI traceability via RMSs is not only the availability of validated RMP(s) but also of competent RMLs. As explained earlier, it is not the manufacturer who should be left with the decision about the competence of a RML. Therefore, the JCTLM established objective criteria and processes to do this in a sufficiently transparent way through the publication of a database *(15)*. This is one mechanism applicable to existing RMLs. However, because it is foreseeable that the needed RML capacity will be high in the near future, a second mechanism is required to ensure that different cRMLs perform at the same level of analytical quality. Only if this can be guaranteed will a manufacturer be confident in the analytical services provided by whichever cRML available at the time of SI-traceable value assignment is needed. In this respect, it has long been recommended that the best mechanism consists of organizing cRMLs in a network *(28, 33)* where they are regularly subjected to external performance assessments. Because at present only a few networks exist, e.g., the IFCC hemoglobin A_{1c} and the Cholesterol Network *(31, 34)*, it became an additional aim of the European project to organize the four cRMLs into a network for measurement of serum total T_4. The underlying idea was that the network would not only be a tool to assess proper performance of the cRMLs in the time span of the project, but should be maintained after the project. Hence, it would create an opportunity for increased RML capacity by setting a system in place to assess laboratories that apply for admittance in the near future.

As can be inferred from the results presented in this study, a network with regular organization of external quality assessment is an excellent mechanism to show that existing network members or potential new members perform the cRMPs at the same level of analytical quality. It also cannot be emphasized enough that common internal accuracy control materials with reliable target values are the key to proper performance of a cRMP, even in an experienced cRML, as well as to maintaining the analytical level in a RML network. The fact that in this study one of the laboratories decided to withdraw its results on the basis of the internal accuracy measures can be used as an argument for this statement. Hence, for the future of the total T_4 network, it will be necessary to provide the cRMLs with new materials in time to perform overlapping runs with the old material.

In conclusion, the study conducted in the framework of European project G6RD-CT-2001-00587 focused on the development of an RMS for total T_4 in serum as a model measurand. Different variants of ID-MS cRMPs are now available and are performed with predefined analytical specifications in four cRMLs organized in a network. The next phase in the project will be the development of a technical implementation plan. This will comprise nomination of the cRMPs and cRMLs for review by the JCTLM and potential inclusion in the database. In this way, the diagnostics industry will become acquainted with their availability. The network activities also will need to be continued according to the mechanism developed here. However, in view of the expected need for increased RML capacity, it is hoped that other laboratories will show interest in network membership. The decisive step to demonstrate that the model RMS works toward establishing SI traceability of routine measurement procedures will be the organization of split-sample measurements with routine commercial total T_4 procedures. This step

recently has been initiated. The invitation to participate was not restricted to European in vitro diagnostic companies but was launched worldwide because SI traceability is an issue of interest to all continents. At least 15 companies accepted the invitation and received aliquots of the native serum panel. The data from these parallel measurements will be interpreted in a method-comparison study by graphical and statistical techniques. Last but not least, continuation of this project toward the development of a RMS for free T_4 would be desirable.

Certain commercial equipment, instruments, or materials are identified in this report to adequately specify the experimental procedure. Such identification does not imply recommendation or endorsement by NIST, nor does it imply that the materials or equipment identified are the best available for the purpose. The European authors gratefully acknowledge financial support by the European Commission of Project G6RD-CT-2001-00587. All authors are also grateful for the constructive discussions with the IFCC representative, J. Thijssen, and the other project members. Last but not least, we are indebted to the attention by the project's coordinator, R. Lequin, for ensuring smooth performance of the administrative and financial affairs.

References

1. Cali JP. An idea whose time has come. Clin Chem 1973;19: 291–3.
2. Uriano GA, Gravatt CC. The role of reference materials and reference methods in chemical analysis. CRC Crit Rev Anal Chem 1977;6:361–411.
3. Tietz NW. A model for a comprehensive measurement system in clinical chemistry. Clin Chem 1979;25:833–9.
4. Stamm D. Reference materials and reference methods in clinical chemistry. J Clin Chem Clin Biochem 1979;17:283–97.
5. Maas AH, Weisberg HF, Burnett RW, Muller-Plathe O, Wimberley PD, Zijlstra WG, et al. Approved IFCC methods. Reference method (1986) for pH measurement in blood. International Federation of Clinical Chemistry (IFCC), Expert panel on pH, blood gases and electrolytes. J Clin Chem Clin Biochem 1987;25:281–9.
6. Marcovina SM, Albers JJ, Dati F, Ledue TB, Ritchie RF. International Federation of Clinical Chemistry standardization project for measurements of apolipoproteins A-I and B. Clin Chem 1991;37: 1676–82.
7. Thienpont LMR, De Leenheer AP, Siekmann L, Kristiansen N, Linsinger T, Schimmel H, et al. The characterisation of cortisol concentrations in a serum reference panel (IRMM/IFCC No. 451). EUR 19007 EN. IRMM/IFCC Information, Reference Materials. Brussels, Belgium: European Commission, Directorate General, Joint Research Centre, 1999.
8. Siekmann L, Bonora R, Burtis CA, Ceriotti F, Clerc-Renaud P, Ferard G, et al. IFCC primary reference procedures for the measurement of catalytic activity concentrations of enzymes at 37 degrees C. Part 1. The concept of reference procedures for the measurement of catalytic activity concentrations of enzymes. Clin Chem Lab Med 2002;40:631–4.
9. McNamara JR, Leary ET, Ceriotti F, Boersma-Cobbaert CM, Cole TG, Hassemer DJ, et al. Point: status of lipid and lipoprotein standardization. Clin Chem 1997;43:1306–10.
10. Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices. Off J European Communities 1998;Dec 7:L331/1–L331/37.
11. In vitro diagnostic medical devices—measurement of quantities in samples of biological origin—metrological traceability of values assigned to calibrators and control materials. Standard reference EN/ISO 17511. Brussels, Belgium: Comité Européen de Normalisation/Technical Body 140, 2003.
12. In vitro diagnostic medical devices—measurement of quantities in samples of biological origin—presentation of reference measurement procedures. Standard reference EN/ISO 15193. Brussels, Belgium: Comité Européen de Normalisation/Technical Body 140, 2002.
13. In vitro diagnostic medical device—measurement of quantities in samples of biological origin—description of reference materials. EN/ISO 15194. Brussels, Belgium: Comité Européen de Normalisation/Technical Body 140, 2002.
14. In vitro diagnostic medical device—measurement of quantities in samples of biological origin—requirements for reference measurement laboratories. EN/ISO 15195. Brussels, Belgium: Comité Européen de Normalisation/Technical Body 140, 2003.
15. Bureau International des Poids et Mesures. Joint Committee for Traceability in Laboratory Medicine. Declaration of cooperation between the CIPM, IFCC and ILAC, for the establishment of a Joint Committee for Traceability in Laboratory Medicine. http://www.bipm.org/en/committees/jc/jctlm/declaration.html. JCTLM Database of Reference Materials and Reference Measurement Procedures. http://www1.bipm.org/en/committees/jc/jctlm/jctlm-db (accessed October 2004).
16. European Commission—5th Framework Programme. Competitive and Sustainable Growth Project G6RD-CT-2001-00587 (2001–2005). Feasibility studies for the development of reference measurement systems for thyrotropin (TSH) and for free thyroxine (FT4), and validation of reference measurement systems (procedure and material) for thyroxine (T4) and triiodothyronine (T3) in human serum (2001–2005). Brussels, Belgium: European Commission, 2000.
17. Mendel CM. The free hormone hypothesis: a physiologically based mathematical model [Review]. Endocr Rev 1989;10:232–74.
18. Ekins R. Validity of analog free thyroxine immunoassays. Clin Chem 1987;33:2137–52.
19. Ekins R. Measurement of free hormones in blood [Review]. Endocr Rev 1990;11:5–46.
20. Ekins R. Analytical measurements of free thyroxine [Review]. Clin Lab Med 1993;13:599–630.
21. De Brabandere VI, Hou P, Stöckl D, Thienpont LM, De Leenheer AP. Isotope dilution-liquid chromatography/electrospray ionization-tandem mass spectrometry for the determination of serum thyroxine as a potential reference method. Rapid Commun Mass Spectrom 1998;12:1099–103.
22. Tai SS, Sniegoski LT, Welch MJ. Candidate reference method for total thyroxine in human serum: use of isotope-dilution liquid chromatography-mass spectrometry with electrospray ionization. Clin Chem 2002;48:637–42.
23. Van Uytfanghe K, Stöckl D, Thienpont LM. Development of a simplified sample pretreatment procedure as part of an isotope dilution-liquid chromatography-tandem mass spectrometry candidate reference measurement procedure for serum total thyroxine [Letter]. Rapid Commun Mass Spectrom 2004;18:1539–40.
24. Siekmann L. Measurement of thyroxine in human serum by isotope dilution mass spectrometry. Definitive methods in clinical chemistry, V. Biomed Environ Mass Spectrom 1987;14:683–8.
25. Hopley CJ, Stokes P, Webb KS, Baynham M. The analysis of thyroxine in human serum by an 'exact matching' isotope dilution

method with liquid chromatography/tandem mass spectrometry. Rapid Commun Mass Spectrom 2004;18:1033–8.
26. Thienpont LM, Van Nieuwenhove B, Stöckl D, De Leenheer AP. Calibration for isotope dilution mass spectrometry—description of an alternative to the bracketing procedure. J Mass Spectrom 1996;31:1119–25.
27. Siekmann L, Breuer H. Determination of cortisol in human plasma by isotope dilution-mass spectrometry. Definitive methods in clinical chemistry. J Clin Chem Clin Biochem 1982;20:883–92.
28. Thienpont LM, Franzini C, Kratochvila J, Middle J, Ricós C, Siekmann L, et al. Analytical quality specifications for reference methods and operating specifications for networks of reference laboratories. Recommendations of the European EQA-Organizers Working Group B. Eur J Clin Chem Clin Biochem 1995;33:949–57.
29. International Organization for Standardization. ISO guide to the expression of uncertainty in measurement (GUM). Geneva: ISO, 1995.
30. Stöckl D, Franzini C, Kratochvila J, Middle J, Ricos C, Siekmann L, et al. Analytical specifications of reference methods. Compilation and critical discussion (From the members of the European EQA-organizers Working Group B) [Review]. Eur J Clin Chem Clin Biochem 1996;34:319–37.
31. Jeppsson JO, Kobold U, Barr J, Finke A, Hoelzel W, Hoshino T, et al. International Federation of Clinical Chemistry and Laboratory Medicine (IFCC). Approved IFCC reference method for the measurement of HbA1c in human blood. Clin Chem Lab Med 2002; 40:78–89.
32. Miller WG, Kaufman HW, McLendon WW, editors. College of American Pathologists Conference XXIII. Matrix effects and accuracy assessment in clinical chemistry. Arch Pathol Lab Med 1993;117:343–436.
33. Siekmann L, Doumas BT, Thienpont L, Schumann G. Reference materials and reference measurement systems in laboratory medicine. Networks of reference laboratories. Eur J Clin Chem Clin Biochem 1995;33:1013–7.
34. Myers GL, Kimberly MM, Waymack PP, Smith SJ, Cooper GR, Sampson EJ. A reference method laboratory network for cholesterol: a model for standardization and improvement of clinical laboratory measurements. Clin Chem 2000;46:1762–72.