Author Manuscript Author Manuscript Author Manuscript Author Manuscript

HHS Public Access
Author manuscript
Clin Biochem. Author manuscript; available in PMC 2024 April 05.

Published in final edited form as:
Clin Biochem. 2023 June ; 116: 42–51. doi:10.1016/j.clinbiochem.2023.03.010.

Development of an Equilibrium Dialysis ID-UPLC-MS/MS Candidate Reference Measurement Procedure for Free Thyroxine in Human Serum

Ashley Ribera[1], **Li Zhang**[1], **Amonae Dabbs-Brown**[1], **Otoe Sugahara**[1], **Krista Poynter**[1], **Katleen van Uytfanghe**[2], **Eri Shimizu**[3], **A E van Herwaarden**[4], **Julianne C. Botelho**[1], **Uliana Danilenko**[1], **Hubert W. Vesper**[1]

[1]Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control and Prevention, 4770 Buford Hwy NE, Atlanta, GA, 30341

[2]Laboratory for Analytical Chemistry, Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium

[3]Reference Material Institute for Clinical Chemistry Standards, Kawasaki, Kanagawa, Japan

[4]Department of Laboratory Medicine, Radboud University Medical Centre, Nijmegen, The Netherlands

Abstract

Background—Accurate and reliable measurement of human serum free thyroxine (FT4) is critical for the diagnosis and treatment of thyroid diseases. However, concerns have been raised regarding the performance of FT4 measurements in patient care. Centers for Disease Control and Prevention Clinical Standardization Programs (CDC-CSP) address these concerns by creating a FT4 standardization program to standardize FT4 measurements. The study aims to develop a highly accurate and precise candidate Reference Measurement Procedure (cRMP), as one key component of CDC-CSP, for standardization of FT4 measurements.

Methods—Serum FT4 was separated from protein-bound thyroxine with equilibrium dialysis (ED) following the recommended conditions in the Clinical and Laboratory Standards Institute C45-A guideline and the published RMP [23]. FT4 in dialysate was directly quantified with liquid chromatography-tandem mass spectrometry (LC-MS/MS) without derivatization. Gravimetric measurements of specimens and calibrator solutions, calibrator bracketing, isotope dilution, enhanced chromatographic resolution, and T4 specific mass transitions were used to ensure the accuracy, precision, and specificity of the cRMP.

[*]**Corresponding Authors:** Li Zhang, Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control and Prevention, 4770 Buford Highway MS F47, Atlanta, GA 30341. LZhang6@cdc.gov.

Disclaimer
The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of CDC. Use of trade names and commercial sources is for identification only and does not constitute endorsement by the U.S. Department of Health and Human Services, CDC.

Disclosure of Potential Conflicts of Interest
The authors declare no potential conflicts of interest.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Results—The described cRMP agreed well with the established RMP and two other cRMPs in an interlaboratory comparison study. The mean biases of each method to the overall laboratory mean were within ±2.5%. The intra-day, inter-day, and total imprecision for the cRMP were within 4.4%. The limit of detection was 0.90 pmol/L, which was sufficiently sensitive to determine FT4 for patients with hypothyroidism. The structural analogs of T4 and endogenous components in dialysate did not interfere with the measurements.

Conclusion—Our ED-LC-MS/MS cRMP provides high accuracy, precision, specificity, and sensitivity for FT4 measurement. The cRMP can serve as a higher-order standard for establishing measurement traceability and provide an accuracy base for the standardization of FT4 assays.

Keywords

Free thyroxine; FT4; equilibrium dialysis; LC-MS/MS; Reference Measurement Procedure; Interlaboratory comparison

1. INTRODUCTION

Thyroid hormones are important in the regulation of a series of physiological and biological processes, including energy metabolism, body temperature, heart rate, body weight, and brain and body development in infancy and childhood. [1, 2] Approximately 20 million people in the USA are afflicted with various types of thyroid disease, which collectively comprise the second most common set of endocrine disorders after diabetes. [3–6] Thyroxine (3, 3’, 5, 5’-tetraiodo-L-thyronine, T4) and triiodothyronine (3, 3’, 5-triiodo-L-thyronine, T3) are two tyrosine-based hormones produced and released by thyroid glands, with T4 being the predominant hormone in the circulation. [7, 8] Measurement of these hormones together with thyroid-stimulating hormone (TSH) is critical in diagnosis, classification, and treatment of thyroid diseases. The free, non-protein bound fraction of thyroid hormones in circulation are considered biologically and physiologically active, and serum free T4 (FT4) more closely correlates with disease states, such as hypo- and hyperthyroidism, than total T4. It is estimated that approximately 18 million FT4 tests are requested in the USA annually. [9, 10] In clinical laboratories, most FT4 assays are performed using immunoassays (IAs). [11, 12] Concerns regarding the accuracy and reliability of FT4 IAs have been raised. [13–16] One study showed that twelve out of thirteen IAs had negative bias to the equilibrium dialysis (ED)-based reference measurement procedure (RMP) ranging between −30.2 to −72.7% at high FT4 concentrations. [15] Another study, which included approximately 3,900 clinical laboratories, reported a difference ranging from −39.7% to 96.7% [14]. As a result, the need for standardization of FT4 tests has been stated by different stakeholders (www.hormoneassays.org, www.harmonization.net). The International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) Committee for Standardization of Thyroid Function Tests (C-STFT) is addressing this need. [17, 18] The Centers for Disease Control and Prevention (CDC) Clinical Standardization Program (CSP) has partnered with IFCC to create a standardization program for FT4 to improve accuracy, reliability, and comparability of current methods by using the reference system developed by the IFCC C-STFT. [19]

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

The IFCC C-STFT reference system includes a conventional RMP [20, 21] consisting of a well-defined procedure for ED based on the principle as described previously [20, 22, 23] and calibrators traceable to the International System of Units (SI). While the operational conditions for ED need to be followed as outlined in the conventional RMP, the actual measurement following ED can be performed by liquid chromatography tandem mass spectrometry (LC-MS/MS) or gas chromatography mass spectrometry (GC-MS), provided they operate at the level of a reference measurement procedure, fulfilling the stringent performance requirements established for the FT4 RMPs [24]. Several LC-MS, LC-MS/MS or GC-MS methods for thyroid hormones have been described [25–30]. GC-MS analysis was used to determine total T4 in serum and requires derivatization which limits its application. [26, 27, 29] LC-MS/MS allows for measurements without derivatization and has been applied in several routine FT4 methods [30–32]. The conventional FT4 RMP as endorsed by the IFCC and later recognized by the Joint Committee for Traceability in Laboratory Medicine (JCTLM) implemented two-dimensional LC separation [20]. The CDC RMP provides an alternative method for labs choosing one dimensional ultra-performance liquid chromatography due to availability of specific instrumentation or laboratory preference. The significant variability among FT4 assays can be minimized through standardization with assay recalibration against RMP as a first step. This creates the need for an operational FT4 RMP. We describe a FT4 RMP that employs ED as prescribed for the conventional RMP, uses certified primary reference materials as calibrators, and an optimized sample preparation procedure after ED followed by isotope dilution ultra-performance liquid chromatography-tandem mass spectrometry analysis (ID-UPLC-MS/MS).

2. Materials and Methods

2.1 Reagents

L-Thyroxine certified reference material IRMM-468 was obtained from the Joint Research Centre (Geel, Belgium). L-Thyroxine-$^{13}C_6$ (100 μg/mL) was procured from Sigma-Aldrich (St. Louis, MO). Custom HEPES [N-(2-hydroxyethyl) piperazine-N9-(2-ethylsulfonic acid)] dialysis buffer kits were from ABI Scientific (Sterling, VA). The composition of the HEPES dialysis buffer kits were elaborated in section 2.3.3 Analytical Solution. HEPES dialysis buffer (52.8 mmol/L, pH 7.4) was prepared by dissolving HEPES dialysis buffer kits into 1 L deionized water. Formic acid (98%, LC/MS grade), water with 0.1% formic acid (v/v) (26.5 mmol/L, LC/MS grade), methanol (HPLC Grade), acetonitrile (LC/MS grade), and ethanol (ACS grade) were obtained from Fisher Scientific (Suwanee, GA). Ammonium hydroxide (28.0–30.0% NH_3, w/w, 14.8–15.9 mol/L, Extra Pure) was purchased from ACROS Organics (Fair Lawn, NJ). All other reagents such as cyclohexane, ethyl acetate were at least of analytical grade and obtained from Fisher Scientific (Suwanee, GA). Fresh-frozen pooled and individual donor sera were acquired from Solomon Park Research Laboratories (Kirkland, WA), BioIVT (Westbury, NY) and in.vent Diagnostica GmbH (Hennigsdorf, Germany). Sera were collected and prepared according to protocol based on CLSI document C37. [33] Pooled and single donor serum samples were used as quality controls to assess performance of the RMP. These companies have IRB approvals to collect blood and obtained informed consent from donors. CDC's use of the blood is

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

consistent with the IRB approval and donor consent. No personal identifiers were provided to CDC.

2.2 Apparatus

ED temperature was monitored using an RTR-501 data logger with PT1000 probe from T&D Corporation (Matsumoto, Japan). Dialysis was performed in a 104L Heratherm IMH80 Advanced Protocol Microbiological Incubator from ThermoFisher Scientific (Waltham, MA). A Micro-Equilibrium Dialyzer System, consisting of 1 mL PTFE dialysis cells, cell racks and spacers, cell stoppers and emptying plugs, and 5 kDa MWCO regenerated cellulose membranes were obtained from Harvard Apparatus (Holliston, MA). Racks of dialysis cells were continuously rotated during the incubation period with bottle/ tube Rollers from ThermoFisher Scientific (Waltham, MA). Density measurements were performed on a DM500 density meter from Anton Paar (Graz, Austria). Sep Pak C18 1cc cartridges were acquired from Waters (Milford, MA). Isolute SLE+ 1 mL tabless supported liquid extraction (SLE) cartridges and a PRESSURE+ manifold were obtained from Biotage (Charlotte, NC). Samples were analyzed using a Shimadzu LC-30AD high-performance liquid chromatography (HPLC) module (Kyoto, Japan) coupled with an AB Sciex API 5500 Triple Quad Mass Spectrometer (Framingham, MA, USA).

2.3 Preparation of measuring system and analytical portion

2.3.1 Calibration—A primary calibrator stock solution (Stock Solution A) was prepared gravimetrically from certified T4 primary reference material IRMM-468. All dilutions were performed gravimetrically. The concentrations were expressed as mass fractions (e.g. mg/g or ng/g). Ethanol containing 1.7% ammonium hydroxide (v/v) (0.26 mol/L NH_3 in ethanol) was used as the solvent for preparation of stock, intermediate, and working solutions. T4 primary reference material (100 mg) was weighed and transferred to an amber glass bottle and dissolved with the help of sonication in 100 mL solvent (Stock Solution A: 1 mg/g). The intermediate solution was prepared by diluting 100 µL of Stock Solution A in 100 mL solvent (Stock Solution B: 1 µg/g). Calibrator working solutions (WS) were prepared by diluting up to 100 µL of Stock B solution in 100 mL of solvent to reach a concentration of 1 ng/g. Internal standard (IS) stock solutions were prepared using $^{13}C_6$-T4 material. Gravimetric procedures, as described above, were followed for dilution to obtain IS working solutions (ISWS) at a concentration of 0.1 ng/g. Calibration curves were prepared fresh prior to analysis using the bracketing technique previously described to meet the rigorous requirements for imprecision and bias for the cRMP [34, 35]. Five-point calibration curves were prepared by adding 20.0–60.0 pg of $^{13}C_6$-T4 to 39.5 pg T4 to obtain mass ratios from 0.66 to 2 of unlabeled to labeled T4, encompassing the approximately 1:1 ratio of T4 to its labeled IS in unknown samples. To accomplish this, we first acquired the orientational serum FT4 values using a routine FT4 assay prior to spiking with labeled T4. Calibration curves were prepared in triplicate. To evaluate the linear range of calibration, a 7-point expanded calibration curve with concentrations 1.29, 6.44, 12.9, 25.7, 64.4, 129, and 258 pmol/L (1, 5, 10, 20, 50, 100, and 200 pg/mL), was prepared and spiked with a fixed amount of ISWS solution (50 pg in each calibrator). These calibrators were evaporated under an N_2 stream and reconstituted with 800 µL of 10% acetonitrile (v/v) in water containing 0.1%

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(v/v) formic acid (1.92 mol/L acetonitrile and 26.5 mmol/L formic acid in water) before LC-MS/MS analysis.

2.3.2 Structure of analytical series—

Samples are processed together with 2 batch quality control (QC) samples in the same run. 3 sets of calibrators are prepared for each batch. Approximately 20 patient samples are processed in one run. To precaution against carryover, blanks of running buffer, referred to as run blanks, are spaced approximately every 8 samples. For additional details, refer to supplementary Table S1.

2.3.3 Analytical solution—

HEPES dialysis buffer was prepared by dissolving one HEPES dialysis buffer kit containing 52.8 mmol/L HEPES, 91.6 mmol/L sodium chloride, 1.65 mmol/L potassium phosphate, 2.68 mmol/L potassium chloride, 1.12 mmol/L magnesium sulfate heptahydrate, and 5.00 mmol/L urea in 1L of deionized water and supplementing with 8.00 mmol/L sodium azide and 1.90 mmol/L calcium chloride. A pH-adjusting buffer containing 776 mmol/L HEPES was also prepared by dissolving 9.25 g HEPES in 50 mL deionized water. The pH of HEPES dialysis buffer and pH-adjusting buffer were measured using a Hanna HI4221 laboratory research grade pH/mV pH meter (Smithfield, RI) and adjusted to pH 7.4 ± 0.03 at 37 °C with 10 mol/L NaOH or glacial acetic acid from Fisher Scientific (Suwanee, GA).

2.4 Operation of measuring system

2.4.1 Equilibrium Dialysis—

ED was performed using the conditions outlined in the conventional RMP [20, 22, 23]. Frozen sera (1 mL each) were thawed at room temperature for 1 hour with continuous rotation at 10–20 rpm, then buffered to pH 7.4 ± 0.03 at 37 °C with minimal volume (no more than one-tenth the volume of serum) of pH adjusting buffer. Dialysis membranes (5 kDa MWCO, regenerated cellulose) were pretreated by immersing 3 times in deionized water and twice in dialysis buffer for 15 minutes each. Dialysis cells were assembled with the pretreated dialysis membrane in between the two PTFE half cells and compressed in racks between spring-loaded spacers to prevent leakage (Figure 1). One compartment of the dialysis cell was filled with 1 mL of pH-adjusted serum, and the other compartment was filled with the same volume of HEPES dialysis buffer. Then, the dialysis cells were continuously rotated at 25 rpm in a convection oven at 37.0 ± 0.5 °C for a total of 4 hours after temperature recovery. The temperature within the oven was monitored remotely using a temperature data logger. After the 4-hour incubation, dialysate samples were collected into glass test tubes. All results presented in this manuscript were obtained using the cRMP that included the ED step unless specified otherwise.

2.4.2 Dialysate preparation—

A routine FT4 assay (Roche Diagnostics, Indianapolis, IN) provided the approximate concentrations of FT4 in the samples before ED. Dialysate samples were spiked with appropriate amounts of ISWS solution, according to approximate estimations by the routine assay, to achieve T4/T4-$^{13}C_6$ mass ratios close to 1:1. Then, T4 in the dialysate was isolated from the sample matrix by solid-phase extraction (SPE), according to the previously published method [20], followed by liquid-liquid extraction (LLE) with ethyl acetate at a 1:1 volume ratio of dialysate to ethyl acetate. Prior to SPE, dialysate samples with IS were equilibrated by shaking at 1600 rpm at room temperature

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

for 60 minutes, then acidified with 20 μL 5% formic acid in water (v/v) (1.33 mol/L). Flow rates during SPE were approximately 1–2 drops/second for all steps. SPE cartridges (50 mg C18 sorbent, 1 mL capacity, Waters, Milford, MA) were conditioned sequentially with 2 mL of acetonitrile and 2 mL of water with 0.1% formic acid (v/v) (26.5 mmol/L). Sample dialysates were loaded onto C18 SPE cartridges, eluates were collected, then eluates were passed through the cartridges a second time to maximize recovery. SPE cartridges were washed with 1 mL water with 0.1% (v/v) formic acid (26.5 mmol/L) followed by 1 mL 25% (v/v) acetonitrile in water with 0.1% (v/v) formic acid (4.79 mol/L acetonitrile and 26.5 mmol/L formic acid). T4 was eluted with 1 mL 50% (v/v) acetonitrile in water containing 0.1% (v/v) formic acid (9.59 mol/L acetonitrile and 26.5 mmol/L formic acid). Eluates were acidified by adding 30 μL of 30% HCl (9.56 mol/L) before LLE with ethyl acetate. The eluates were combined with 1 mL ethyl acetate and subjected to vortex mixing for 5 minutes at 1600 rpm. Aqueous and organic phases were separated under refrigerated (4°C) centrifugation at 1000 × g for 2 min. The FT4-containing organic phase was isolated with glass Pasteur pipettes and placed into a new glass tube. The extraction was repeated with an additional 1 mL ethyl acetate. The combined organic extracts were dried under nitrogen flow and reconstituted with 200 μL of 10% (v/v) acetonitrile in water containing 0.1% (v/v) formic acid (1.92 mol/L acetonitrile and 26.5 mmol/L formic acid).

Three alternate extraction procedures were tested and compared to the aforementioned procedure (Procedure ‘A’). Procedure ‘B’ utilized serum dialysates that were acidified with 65 μL 35% formic acid in water (v/v) (9.28 mol/L) before extracting hydrophobic components twice with 0.9 mL cyclohexane. Organic extracts were discarded, and aqueous extracts were further acidified with 35 μL 35% formic acid in water (v/v) (9.28 mol/L) before extracting T4 twice with 1 mL ethyl acetate. In procedure ‘C’, sample dialysates were acidified with 100 μL 35% formic acid in water (v/v) (9.28 mol/L) and extracted twice with 1 mL ethyl acetate. In procedure ‘D’ sample dialysates were acidified with 100 μL 35% formic acid in water (v/v) (9.28 mol/L) and transferred to 1 mL SLE cartridges (Biotage. Charlotte, NC). Positive pressure of nitrogen gas was applied briefly to initiate loading of samples onto SLE substrate. After a 5-minute equilibration period, T4 was extracted by adding 5 mL ethyl acetate to the cartridges in two 2.5 mL increments. The alternative extraction procedures (methods B-D) were evaluated for extraction efficiency and matrix effects, as described in the method validation section. Liquid transfer was automated using a Hamilton Microlab Starlet (Reno, NV).

2.4.3 UPLC–MS/MS analysis—Chromatographic separation was performed on a reversed phase UPLC column (Hypersil GOLD C18 50×2.1 mm, 1.9 μm; Fisher Scientific, Suwanee, GA), that was protected with a guard column, (Hypersil GOLD C18 Drop-In Guard Cartridges, 2.1 mm, Fisher Scientific, Suwanee, GA) and heated to 55 °C. Analytes on the column were eluted by a linear gradient of 90–25% mobile phase A (0.1% formic acid in H_2O, v/v) and 10–75% mobile phase B (100% methanol) over 8 minutes at a flow rate of 0.25 mL/min. The column was washed for 2.5 minutes at 98% mobile phase B, then equilibrated at initial conditions for 1.5 minutes before subsequent injections. The mass spectrometry analysis was performed in positive ion mode with the following parameters: ion spray voltage at 5500 V, source heater temperature at 650 °C, curtain gas at 20 psi,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

ion source gas 1 at 80 psi, ion source gas 2 at 60 psi, and collision gas at 10 psi. Selected reaction monitoring (SRM) was used to detect the following ion transitions: *m/z* 777.7 →731.6 (quantitation ion, QI) and *m/z* 777.7 → 604.7 (confirmation ion, CI) for T4, and *m/z* 783.7 → 737.6 (QI) and *m/z* 783.7 → 610.7 (CI) for the IS (T4-$^{13}C_6$). All LC-MS/MS data were recorded at unit mass resolution.

2.4.4 Blanking—
A run blank is analyzed by LC-MS/MS at least every 8 samples and evaluated for potential carryover.

2.4.5 Data processing—
Analyst® software 1.6 (Applied Biosystems) was used for operation and data analysis of the LC–MS/MS. Area count ratios (area count analyte/ area count IS) of samples were used for calculating analyte concentrations based on the calibration curve at the same run. The final concentration of FT4 in serum dialysate was expressed in pmol/L by converting mass fraction concentration (pg T4/g serum dialysate) to molar concentration (pmol/L) using the measured sample dialysate density.

2.5 Method validation

2.5.1 Accuracy and precision—
Because FT4 serum- or plasma-based certified reference materials are not available, accuracy of the cRMP was evaluated through interlaboratory comparison among the IFCC endorsed FT4 RMP at Ref4U, Laboratory of Toxicology, Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium, (JCTLM DB identification number: C8RMP1); CDC cRMP; cRMPs performed at the Reference Material Institute for Clinical Chemistry Standards, Kawasaki, Kanagawa, Japan (ReCCS); cRMPs performed at the Department of Laboratory Medicine, Radboud University Medical Centre, Nijmegen, the Netherlands (RUMCN). A set of 20 euthyroid single donor sera were selected for a comparison covering a concentration range of 14.1 to 24.2 pmol/L. The samples were analyzed in 3–4 independent measurements with independent ED for each replicate. The results produced by CDC cRMP were compared with those measured by the recognized RMP (for accuracy evaluation) and two cRMPs (for verification purposes). The CDC cRMP biases to each of the participating laboratories were evaluated using pairwise comparison. The mean percent bias to the FT4 RMP at Ghent University was determined following the protocol from the Clinical and Laboratory Standards Institute (CLSI) document EP 09 A3. [36] In addition, results of each laboratory were plotted against the mean of the four comparative laboratories, and a Deming regression analysis was performed for the interlaboratory comparison to estimate the difference between the means of each laboratory with the mean of four laboratories. To evaluate the method accuracy at hypo-, hyperthyroidism ranges, CDC cRMP was further compared to an established RMP (Ghent University) for individual donor sera with lower and higher FT4. A total of 4 samples with FT4 concentrations 3.98, 12.0, 52.2, and 86.0 pmol/L were analyzed in singlicate over 2–3 independent analytical runs and the mean bias to the mean FT4 concentration among values measured by CDC cRMP and the RMP at the Ghent University was determined.

Precision was evaluated by analyzing serum-based control materials with low (5.26 pmol/L), medium (18.9 pmol/L), and high (36.0 pmol/L) FT4 concentrations in 2–5 replicates

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

inclusive independent ED for each replicate, respectively over at least 5 days. The intra-assay, inter-assay, and total percent coefficients of variation (CVs) were determined, according to the principles described in CLSI EP 5-A3.[37]

2.5.2 Measurement uncertainty—Potential sources of uncertainty were evaluated and used to calculate the standard uncertainty and expanded uncertainty, according to The International Organization for Standardization (ISO) Guide to the Expression of Uncertainty in Measurement 2008. [38] The estimated variance of Type A uncertainty was obtained from imprecision of the repeated measurements inclusive independent ED for each replicate over 5–22 days, and Type B uncertainty was estimated from uncertainties in the purity of the primary reference material, inaccuracy in the weighing of each component, and the measurement of the serum dialysate density. Type A and B uncertainties were combined quadratically to determine the standard uncertainty. Expanded uncertainty, at the 95% confidence level, was determined by multiplying standard uncertainty by a coverage factor, k=2. Low, medium, and high levels of serum-based material were used for the assessment. Samples were measured in 2–5 replicates inclusive independent ED for each replicate on each day for 5–22 days.

2.5.3 Specificity and sensitivity—The QI/CI of T4 was used to evaluate any potential interferences. The QI/CI of T4 in 20 single-donor serum samples were compared with those of the neat (matrix free) calibrators prepared in 1.7% ammonium hydroxide (v/v) in ethanol (0.26 mol/L NH_3 in ethanol). As suggested by CLSI C62-A, the criterion to confirm the absence of interferences was a difference of less than 20%. [39] The mean QI/CI ratio of 12 replicates for each calibrator was used for calculations. Thyroxine was identified when the chromatographic peak had the same retention time as the IS (±2.5%). A group of structural analogs (Supplementary Table 2) of T4, with relative molecular masses close to T4 and T4-$^{13}C_6$, were tested for potential quantitation interferences. The absence of peaks for QI and CI transitions resulting from structural analogs at the same retention time of T4 was used to identify lack of interferences.

The limit of detection (LOD) of T4 measurement in dialysis buffer by LC/MS/MS without ED procedure was estimated according to the Taylor Method. [40] Briefly, spiked dialysis buffer at 5 concentration levels from 1.29–64.5 pmol/L(1–50 pg/mL) were prepared and analyzed in duplicate on 3 different days (n=6). The standard deviation of each spiked dialysis buffer was plotted against the concentration. The estimated standard deviation when the concentration is extrapolated to zero concentration is represented as S_0. LOD were 3 times S_0. Limit of quantitation (LOQ) of T4 measurement in dialysis buffer was defined as the minimum concentration meeting maximum imprecision of 5% in repeated measurements of the spiked dialysis buffer study.

2.5.4 Linearity assessment—Linearity range of T4 measurement in dialysate samples by LC-MS/MS was evaluated following the principles described in CLSI document EP6-A [41] Seven levels of calibrator working solutions (1.29, 6.44, 12.9, 25.7, 64.4, 128, and 257 pmol/L) were measured over the course of 4 days and the ratios of T4 QI to $^{13}C_6$-T4 QI area counts were plotted against thyroxine concentrations. Linearity of the measurement range (the range between lowest and highest calibrator concentrations) was assessed using

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

residuals and linear and polynomial fitting models. The mean sum of squared residuals (ASSR) and the mean relative sum of squared residuals (RASSR) from 4 sets of calibration curves analyzed over 4 days were used to choose the best fitting regression model from among linear and polynomial models. The ASSR was calculated from the sum of the squared differences between the model predicted and observed Y values where X represents the calibrator concentrations and Y represents the instrument responses (ratios of area counts to IS). The RASSR was calculated by dividing the ASSR by the average of the Y values. Mean slope and mean intercept were calculated using results obtained from 4 independent calibration curves. All calculations were performed using SAS (SAS Version 9.4, SAS Institute Inc., NC, USA).

2.5.5 Matrix effect—The IS compensated matrix effect (ME) on quantitation was evaluated, according to procedures described previously. [35, 42] The serum dialysate samples were compared with matrix-free neat samples in ethanol. Calibration curves with concentrations 4.02, 8.04, 16.1, 32.2, 48.3, and 64.4 pmol/L (3.13, 6.25, 12.5. 25.0, and 50.0 pg/mL) were prepared in pooled serum dialysate and ethanol. IS, at a concentration of 19.3 pmol/L (15.0 pg/mL), was added to each calibrator solution. The calibrators in serum dialysate were prepared according to extraction procedures A, B, C, and D. The calibrators in ethanol were dried under nitrogen flow and reconstituted with 200 μL of 10% acetonitrile in water containing 0.1% (v/v) formic acid. The area count ratios (T4/IS) of calibrators in the matrix were compared to those of the neat calibrators prepared in ethanol (matrix free). The sample ME was determined with the following equation: ME% = B/A × 100, where “B” is the area count ratios obtained from samples in matrix, and “A” is the area count ratios in matrix free samples. The slopes of the calibration curves in matrix were compared with the slopes of matrix free calibration curves to assess the influence of ME on the slope of curves.

2.5.6 Extraction recovery—Extraction recovery was evaluated as described previously. [42] Serum dialysate samples were extracted by following the sample preparation procedure described above. Extraction recovery for T4 was determined by comparing the peak area of labeled IS in samples spiked post-extraction (“B”) with area of labeled IS in samples spiked pre-extraction (“C”). Percent recovery was calculated as C/B × 100.

3. RESULTS AND DISCUSSION

FT4 RMP are needed to assign target values to serum materials used to assist with assay calibration and verification of successful implementation of metrological traceability as required by ISO 17511:2020 [43]. For FT4, because the analyte is defined by the RMP, patient results are traceable to the ED-based RMP and the FT4 RMP is calibrated with the primary thyroxine reference material that is traceable to SI. The availability of FT4 RMPs is critical as no certified reference materials for FT4 are available. To address this need, CDC developed a new highly accurate and precise FT4 cRMP that complies with the conventional RMP for ED conditions followed by analysis of FT4 in dialysate by LC-MS/MS.

The ED conditions outlined in the conventional RMP were critical to preserve the endogenous equilibrium between free and protein-bound T4 in sera during sample preparation, and these conditions constrain membrane and ED device selection. [20] The

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

original publication describing the ED step not only mentions certain material manufactures but also the characteristics of the materials (i.e., regenerated cellulose). Any material that meets those characteristics can be used. Therefore, to maintain independence of the method from a specific supplier and continuity of the method, any future changes in device or membrane selection should be verified by comparison to an established FT4 RMP. Additional studies of dialysis devices and membrane type are being conducted and will be presented in a separate manuscript.

The time required for serum samples to reach equilibrium at 37 °C was determined by monitoring the change in T4 concentration over a period of 21 hours (Supplementary Figure 1). The time to reach equilibrium was assessed as the time to reach a plateau in FT4 concentration, which was observed after 4 hours. While the conventional ED step for phase separation of FT4 from protein-bound T4 must be followed as described in the conventional RMP, quantitation of FT4 in the dialysate can be achieved by various suitable analytical methods, such as GC-MS or LC-MS, provided that the performance characteristics meet requirements for FT4 RMPs [20, 24].

Ideally, as a FT4 RMP is defined by specific operational parameters, accuracy and consistency need to be ensured through a network of reference laboratories performing the conventional ED steps in the same manner. The described cRMP is highly accurate and in excellent agreement with recognized RMP performed at Ghent University as well as with two other candidate RMPs comprising the FT4 reference laboratory network. The CDC cRMP mean bias to the established RMP operated at Ghent University was −1.9% determined with 20 individual donor samples (Table 1), which is within the 2.5% bias limit for cRMPs defined previously [20, 24]. CDC cRMP was also in excellent agreement with two cRMPs participating in the interlaboratory study. The CDC cRMP biases from pairwise comparison to RUMCN and ReCCS cRMPs were 2.5% and −0.6%, respectively, suggesting excellent agreement of cRMPs (Table 1). The means of each laboratory were then compared with the overall mean of four laboratories. The bias of each laboratory to the overall mean of four laboratories is presented in Figure 2a, and the mean bias of each lab to the overall 4-laboratory mean is within 2.5% (Table 1). Consistently, Deming regression analysis showed no significant differences ($p>0.05$) between the mean of CDC cRMP to the overall mean of four laboratories (Figure 2b, Table 1). The accuracy of CDC cRMP was further evaluated using pairwise comparison to Ghent University RMP for measurements of individual sera at hypo- and hyperthyroidism ranges (Table 2). The mean percent bias to the Ghent University was −1.8 ± 3.6%, which is consistent with performance of the measurements of euthyroid sera. The high level of accuracy is achieved through closely following conventional ED conditions in the established FT4 RMP [20], use of certified primary reference material IRMM-468 as calibrators, which provides measurement results traceable to SI in accordance with 17511:2020 [43], and utilization of gravimetric instead of volumetric measurements. Furthermore, calibrator bracketing and use of an adjusted ratio of T4/IS at 1:1 minimizes potential inaccuracy due to ion suppression [44–46].

The presented cRMP is precise with intra-assay and inter-assay imprecisions (%CV) of less than 3.6% for QC materials tested at low, medium, and high FT4 (Table 3). The combined, or total, imprecision (%CV) was less than 4.4%. Inter-assay, intra-assay, and combined

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

imprecision of our cRMP were comparable to the imprecision of the JCTLM-listed FT4 RMP at the Ghent University (2.4%, 2.8%, and 3.7%, respectively) [20] and within the 5% requirement for FT4 RMPs [24]. When the cRMP participated in the interlaboratory comparison study, the mean CV% (± 95% CI) for all sample measurements was 3.3 ± 1.1%, which was consistent with the performance characteristics of the cRMP. Measurement uncertainty was assessed at low, medium, and high serum levels associated with hypo-, eu-, hyperthyroidism. Type B uncertainty was minimized by using primary IRMM-468 standards with a purity of 98.7% and gravimetric measurements in place of volumetric measurements during the preparation of standards and samples. The relative expanded uncertainty was 8.8% at each level (Table 4).

The described cRMP is highly specific. Chromatographic peaks for QI and CI transitions of T4 and T4-$^{13}C_6$ were not present in the blank dialysis buffer sample indicating no interferences or carryover from materials and reagents (Supplementary Figure S2). Assessment of potential interferences from specific compounds was determined by comparing the retention times of T4 with those of a series of structural analogs. None of the compounds tested interfered with T4 (Supplementary Table S2). T3 and reverse T3 (rT3) showed baseline separation from T4 (Figure 3). When an IS-only sample was monitored, no additional peaks were observed in the analyte signal, confirming the absence of isotopic interferences from the use of the selected IS (Supplementary Figure S2). To assess potential interferences from unknown compounds, the mean QI/CI ratio determined from neat T4 calibrator solutions was compared to the QI/CI ratio determined in 20 individual donor samples (FT4 concentrations: 11.9 – 29.5 pmol/L [9.27–22.9 pg/mL]). The mean QI/CI ratio of the calibrators measured in triplicate over 4 days was 4.47 (95% CI of 4.44–4.50). The mean QI/CI ratio of individual donor samples was 4.52 (95% CI of 4.48–4.56), well within the ±20% criterion suggested by CLSI C62-A. [39].

The LOD and LOQ are sufficiently low to reliably measure samples from hypo-, eu-, and hyperthyroid individuals. The LOD and LOQ of the cRMP were 0.9 pmol/L (0.70 pg/mL) and 1.60 pmol/L (1.24 pg/mL), respectively. The extrapolated SD used in LOD assessment at zero concentration was determined as shown in Supplementary Figure S3. The LOD and LOQ are similar to the LOD (0.5 pmol/L) and LOQ (1.3 pmol/L) reported for the JCTLM-listed RMP at Ghent University [23].

The cRMP was determined to be linear over the analytical measurement range which covers hypo-, eu-, and hyperthyroid patient samples (1.29–258 pmol/L (1–200 pg/mL)) with R^2>0.995 with no significant nonlinear relationship observed (Supplementary Figure S4). To evaluate sample ME on quantitation, the slope, R^2, and ME% of the calibration curves in matrix were compared to those of the neat calibration curve (Table 5). Calibrators in matrix were prepared by method A. The slopes of dialysate matrix and matrix free calibration curves were all close to 1. Calibration curves of T4 in serum dialysate were linear over the range of 4.07–64.6 pmol/L (3.16–50.2 pg/mL) (R^2>0.995). The difference of the slope of calibration curves prepared with method A to the neat calibration curve was less than 1% and the IS compensated ME% was 98.8% for method A, indicating that calibration is not affected by the presence of serum dialysate matrix and that there is no matrix influence on the peak area ratios. Calibrators may be prepared in ethanol solutions

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

without any significant matrix effect. In the presented method, calibrator solutions were prepared yearly. The stability of calibrator solutions (Stock A, Stock B, and WS) prepared in 1.7% ammonium hydroxide (v/v) in ethanol (0.26 mol/L NH_3 in ethanol) was confirmed by HPLC for Stock A and Stock B, and LC-MS/MS for WS, respectively over an 18-month period of storage at −20 °C.

For alternative extraction methods (B-D), matrix effects were evaluated in a similar manner (Table 5). No additional interfering peaks at the same retention time of T4 were observed in chromatograms for samples prepared by methods B–D. The slopes of dialysate-based calibration curves were all close to 1 for methods B–D. In addition, the mean difference of the slopes of the calibration curves in matrix to the neat calibration curve slope were −0.1–2.0%. The IS compensated ME% of each method was close to 100%. These observations indicated that methods B–D did not result in an ME that would negatively influence quantitation.

The extraction recoveries measured in triplicate for methods A–D were 85.4 ± 5.1%, 93.7 ± 4.5%, 95.0 ± 5.5%, and 92.2 ± 4.0%, respectively. The extraction efficiencies of the four methods were equally suitable for sample preparation. Preliminary evaluation of methods B–D demonstrate the flexibility of choosing different validated extraction methods during sample preparation. This may be advantageous during method transfer between laboratories in the future, however thorough validation of each new extraction method will be needed, including performance comparison with the established RMP.

By comparing with previously established FT4 RMP, we provide an alternative set of extraction and chromatographic conditions in the presented cRMP. Using our chromatographic conditions T4 could be baseline resolved from T3, rT3 and other interference within 12 min (Figure 3), which presents a simplified alternative to published 2-dimensional chromatography [20]. The described procedure does not require the use of monoiodotyrosine (MIT) or diiodotyrosine (DIT) as carriers during calibrator preparation which can simplify preparation of standards and patient samples and potentially prevent introducing T4 from contaminated MIT or DIT. We achieved this with addition of ammonium hydroxide as a pH modifier to the calibrator solutions, which reduced adsorption of FT4 to containers. An automation system (Hamilton Microlab Starlet) was introduced for liquid handling during sample preparation to reduce labor intensity and human errors.

4. CONCLUSION

We developed a new fit-for-purpose FT4 cRMP based on an ED-ID-LC-MS/MS procedure. The accuracy and reliability of the cRMP are supported by comparison to the established RMP recognized by JCTLM, with the described FT4 cRMP displaying comparable analytical performance characteristics. The cRMP is also in excellent agreement with the entire reference laboratory network that consists of one established and three candidate RMPs. Automated sample preparation was implemented to improve the throughput and efficiency of the method. The method sensitivity and quantitation dynamic range allow for measuring samples from hypo-, eu-, and hyperthyroid individuals. The described cRMP can

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

meet the requirements of CDC CSP to provide accurate FT4 value assignment for human sera to be used in the standardization program.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Norma Vazquez, Janet Thonkulpitak, Barrett Brister, and Chui Tse for their contributions to the CDC Clinical Reference Laboratories, support from the International Federation for Clinical Chemistry (IFCC) Committee for the Standardization of Thyroid Function Tests (C-STFT), and Clark Coffman, Tatiana Buchannan, and Lynn Collins of the CDC Clinical Standardization Programs.

ABBREVIATIONS

T4	Thyroxine or 3, 3’, 5, 5’-tetraiodo-L-thyronine
T3	3,3′,5-triiodo-l-thyronine
rT3	Reverse T3
FT4	Free T4
TSH	Thyroid-stimulating hormone
ED	Equilibrium dialysis
LC-MS/MS	Liquid Chromatography/Tandem Mass Spectrometry
GC-MS	Gas chromatography mass spectrometry
UPLC	Ultra-performance liquid chromatography
HPLC	High-performance liquid chromatography
JCTLM	Joint Committee for Traceability in Laboratory Medicine
C-STFT	Committee for Standardization of Thyroid Function Tests
IFCC	International Federation of Clinical Chemistry and Laboratory Medicine
CLSI	Clinical Laboratory and Standards Institute
CDC-CSP	Centers for Disease Control and Prevention Clinical Standardization Programs
cRMP	Candidate Reference Measurement Procedure
IS	Internal standard
QC	Quality control
LLE	Liquid-liquid extraction

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

SPE	Solid-phase extraction
ME	Matrices effect
LOD	Limit of detection
LOQ	Limit of quantitation
SRM	Selected reaction monitoring
QI	Quantitation ion
CI	Confirmation ion
CV	Coefficient of variation
MIT	Monoiodotyrosine
DIT	Diiodotyrosine

REFERENCES

[1]. Brent GA, Mechanisms of thyroid hormone action, J Clin Invest 122(9) (2012) 3035–43. [PubMed: 22945636]

[2]. Mullur R, Liu YY, Brent GA, Thyroid hormone regulation of metabolism, Physiol Rev 94(2) (2014) 355–82. [PubMed: 24692351]

[3]. American Thyroid Association, General Information/Press Room, 2020. https://www.thyroid.org/media-main/press-room/.

[4]. Chaker L, Ligthart S, Korevaar TI, Hofman A, Franco OH, Peeters RP, Dehghan A, Thyroid function and risk of type 2 diabetes: a population-based prospective cohort study, BMC Med 14(1) (2016) 150. [PubMed: 27686165]

[5]. Hage M, Zantout MS, Azar ST, Thyroid disorders and diabetes mellitus, J Thyroid Res 2011 (2011) 439463. [PubMed: 21785689]

[6]. Hollowell JG, Staehling NW, Flanders WD, Hannon WH, Gunter EW, Spencer CA, Braverman LE, Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III), J Clin Endocrinol Metab 87(2) (2002) 489–99. [PubMed: 11836274]

[7]. Pilo A, Iervasi G, Vitek F, Ferdeghini M, Cazzuola F, Bianchi R, Thyroidal and peripheral production of 3,5,3’-triiodothyronine in humans by multicompartmental analysis, Am J Physiol 258(4 Pt 1) (1990) E715–26. [PubMed: 2333963]

[8]. Larsen PR, Thyroid-pituitary interaction: feedback regulation of thyrotropin secretion by thyroid hormones, N Engl J Med 306(1) (1982) 23–32. [PubMed: 7031472]

[9]. Thienpont LM, Van Uytfanghe K, Poppe K, Velkeniers B, Determination of free thyroid hormones, Best Pract Res Clin Endocrinol Metab 27(5) (2013) 689–700. [PubMed: 24094639]

[10]. HHS, Medicare Payments for Clinical Laboratory Tests in 2014: Baseline Data, 2015. https://oig.hhs.gov/oei/reports/oei-09-15-00210.pdf. 2020).

[11]. Midgley JE, Direct and indirect free thyroxine assay methods: theory and practice, Clin Chem 47(8) (2001) 1353–63. [PubMed: 11468222]

[12]. Favresse J, Burlacu MC, Maiter D, Gruson D, Interferences With Thyroid Function Immunoassays: Clinical Implications and Detection Algorithm, Endocr Rev 39(5) (2018) 830–850. [PubMed: 29982406]

[13]. Sapin R, d’Herbomez M, Free thyroxine measured by equilibrium dialysis and nine immunoassays in sera with various serum thyroxine-binding capacities, Clin Chem 49(9) (2003) 1531–5. [PubMed: 12928239]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[14]. Steele BW, Wang E, Klee GG, Thienpont LM, Soldin SJ, Sokoll LJ, Winter WE, Fuhrman SA, Elin RJ, Analytic bias of thyroid function tests: analysis of a College of American Pathologists fresh frozen serum pool by 3900 clinical laboratories, Arch Pathol Lab Med 129(3) (2005) 310–7. [PubMed: 15737023]

[15]. De Grande LAC, Van Uytfanghe K, Reynders D, Das B, Faix JD, MacKenzie F, Decallonne B, Hishinuma A, Lapauw B, Taelman P, Van Crombrugge P, Van den Bruel A, Velkeniers B, Williams P, Thienpont LM, I.C.f.S.o.T.F. Tests, Standardization of Free Thyroxine Measurements Allows the Adoption of a More Uniform Reference Interval, Clin Chem 63(10) (2017) 1642–1652. [PubMed: 28720678]

[16]. Lee RH, Spencer CA, Mestman JH, Miller EA, Petrovic I, Braverman LE, Goodwin TM, Free T4 immunoassays are flawed during pregnancy, Am J Obstet Gynecol 200(3) (2009) 260 e1–6.

[17]. Thienpont LM, Beastall G, Christofides ND, Faix JD, Ieiri T, Miller WG, Miller R, Nelson JC, Ross HA, Ronin C, Rottmann M, Thijssen JH, Toussaint B, Measurement of free thyroxine in laboratory medicine--proposal of measurand definition, Clin Chem Lab Med 45(4) (2007) 563–4. [PubMed: 17439341]

[18]. Thienpont LM, Faix JD, Beastall G, Standardization of Free Thyroxine and Harmonization of Thyrotropin Measurements: A Request for Input from Endocrinologists and Other Physicians, Thyroid 25(12) (2015) 1379–80. [PubMed: 26413772]

[19]. Vesper HW, Van Uytfanghe K, Hishinuma A, Raverot V, Patru MM, Danilenko U, van Herwaarden AE, Shimizu E, Implementing reference systems for thyroid function tests - A collaborative effort, Clin Chim Acta 519 (2021) 183–186. [PubMed: 33933427]

[20]. Van Houcke SK, Van Uytfanghe K, Shimizu E, Tani W, Umemoto M, Thienpont LM, IFCC international conventional reference procedure for the measurement of free thyroxine in serum: International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) Working Group for Standardization of Thyroid Function Tests (WG-STFT)(1), Clin Chem Lab Med 49(8) (2011) 1275–81. [PubMed: 21675941]

[21]. Thienpont LM, Beastall G, Christofides ND, Faix JD, Ieiri T, Jarrige V, Miller WG, Miller R, Nelson JC, Ronin C, Ross HA, Rottmann M, Thijssen JH, Toussaint B, Proposal of a candidate international conventional reference measurement procedure for free thyroxine in serum, Clin Chem Lab Med 45(7) (2007) 934–6. [PubMed: 17617044]

[22]. CLSI, Measurement of Free Thyroid Hormones; Approved Guideline. Clinical Laboratory and Standards Institute document C45-A [ISBN 1-56238-548-8]. CLSI, Wayne, PA, 2004.

[23]. Van Uytfanghe K, Stockl D, Ross HA, Thienpont LM, Use of frozen sera for FT4 standardization: investigation by equilibrium dialysis combined with isotope dilution-mass spectrometry and immunoassay, Clin Chem 52(9) (2006) 1817–21. [PubMed: 16840583]

[24]. Van Uytfanghe K, Standardization and Harmonization Laboratory Updates IFCC C-STFT 2020 Meeting, 2020.

[25]. De Brabandere VI, Hou P, Stockl 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 12(16) (1998) 1099–103. [PubMed: 9737016]

[26]. Hantson AL, De Meyer M, Guerit N, Simultaneous determination of endogenous and 13C-labelled thyroid hormones in plasma by stable isotope dilution mass spectrometry, J Chromatogr B Analyt Technol Biomed Life Sci 807(2) (2004) 185–92.

[27]. Siekmann L, Measurement of thyroxine in human serum by isotope dilution mass spectrometry. Definitive methods in clinical chemistry, V, Biomed Environ Mass Spectrom 14(11) (1987) 683–8. [PubMed: 2962677]

[28]. 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 48(4) (2002) 637–42. [PubMed: 11901062]

[29]. Thienpont LM, De Brabandere VI, Stockl D, De Leenheer AP, Development of a new method for the determination of thyroxine in serum based on isotope dilution gas chromatography mass spectrometry, Biol Mass Spectrom 23(8) (1994) 475–82. [PubMed: 7918690]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[30]. Yue B, Rockwood AL, Sandrock T, La'ulu SL, Kushnir MM, Meikle AW, Free thyroid hormones in serum by direct equilibrium dialysis and online solid-phase extraction--liquid chromatography/tandem mass spectrometry, Clin Chem 54(4) (2008) 642–51. [PubMed: 18258669]

[31]. Tanoue R, Kume I, Yamamoto Y, Takaguchi K, Nomiyama K, Tanabe S, Kunisue T, Determination of free thyroid hormones in animal serum/plasma using ultrafiltration in combination with ultra-fast liquid chromatography-tandem mass spectrometry, J Chromatogr A 1539 (2018) 30–40. [PubMed: 29395158]

[32]. Soldin SJ, Soukhova N, Janicic N, Jonklaas J, Soldin OP, The measurement of free thyroxine by isotope dilution tandem mass spectrometry, Clin Chim Acta 358(1–2) (2005) 113–8. [PubMed: 16018881]

[33]. Danilenko U, Vesper HW, Myers GL, Clapshaw PA, Camara JE, Miller WG, An updated protocol based on CLSI document C37 for preparation of off-the-clot serum from individual units for use alone or to prepare commutable pooled serum reference materials, Clin Chem Lab Med 58(3) (2020) 368–374. [PubMed: 31665109]

[34]. Botelho JC, Ribera A, Cooper HC, Vesper HW, Evaluation of an Isotope Dilution HPLC Tandem Mass Spectrometry Candidate Reference Measurement Procedure for Total 17-beta Estradiol in Human Serum, Anal Chem 88(22) (2016) 11123–11129. [PubMed: 27744701]

[35]. Botelho JC, Shacklady C, Cooper HC, Tai SS, Van Uytfanghe K, Thienpont LM, Vesper HW, Isotope-dilution liquid chromatography-tandem mass spectrometry candidate reference method for total testosterone in human serum, Clin Chem 59(2) (2013) 372–80. [PubMed: 23213081]

[36]. CLSI, Measurement Procedure Comparison and Bias Estimation Using Patient Samples; Approved Guideline - Third Edition EP09-A3 [ISBN 1-56238-888-6]. CLSI, Wayne, PA, 2013.

[37]. CLSI, Evaluation of Precision of Quantitative Measurement Procedures: approved guideline EP05A3 [ISBN: 1-56238-968-8]. CLSI, Wayne, PA, 2014.

[38]. ISO, Uncertainty of measurement–Part 3: Guide to the expression of uncertainty in measurement. International Organization of Standardization; Geneva, Switzerland:, 2008.

[39]. CLSI, CLSI: Liquid chromatography-mass spectrometry methods; approved guideline. CLSI document C62-A., 2014.

[40]. Taylor JK, Quality Assurance of Chemical Measurements, Lewis Publishers: Boca Raton, FL, Chapter 9 (1987) 78–82.

[41]. CLSI, Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach; Approved Guideline-first edition. Wayne, PA: Clinical and Laboratory Standards Institute, 2003.

[42]. Matuszewski BK, Constanzer ML, Chavez-Eng CM, Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS, Anal Chem 75(13) (2003) 3019–30. [PubMed: 12964746]

[43]. ISO(17511), In vitro diagnostic medical devices — Requirements for establishing metrological traceability of values assigned to calibrators, trueness control materials and human samples, 2020.

[44]. Sargent M, Harte R, Harrington C, Guidelines for Achieving High Accuracy in Isotope Dilution Mass Spectrometry (IDMS), in: Sargent M, Harte R, Harrington C (Eds.), Guidelines for Achieving High Accuracy in Isotope Dilution Mass Spectrometry (IDMS), The Royal Society of Chemistry 2002, pp. 1–34.

[45]. Tai SS, Xu B, Welch MJ, Phinney KW, Development and evaluation of a candidate reference measurement procedure for the determination of testosterone in human serum using isotope dilution liquid chromatography/tandem mass spectrometry, Anal Bioanal Chem 388(5–6) (2007) 1087–94. [PubMed: 17530229]

[46]. Heuillet M, Lalere B, Peignaux M, De Graeve J, Vaslin-Reimann S, Pais De Barros JP, Gambert P, Duvillard L, Delatour V, Validation of a reference method for total cholesterol measurement in human serum and assignation of reference values to proficiency testing samples, Clin Biochem 46(4–5) (2013) 359–64. [PubMed: 23219742]

[47]. CLSI, Preliminary evaluation of quantitative clinical laboratory measurement procedures: approved guideline EP10-A3 [ISBN 1-56238-622-0]. CLSI, Wayne, PA, 2006.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

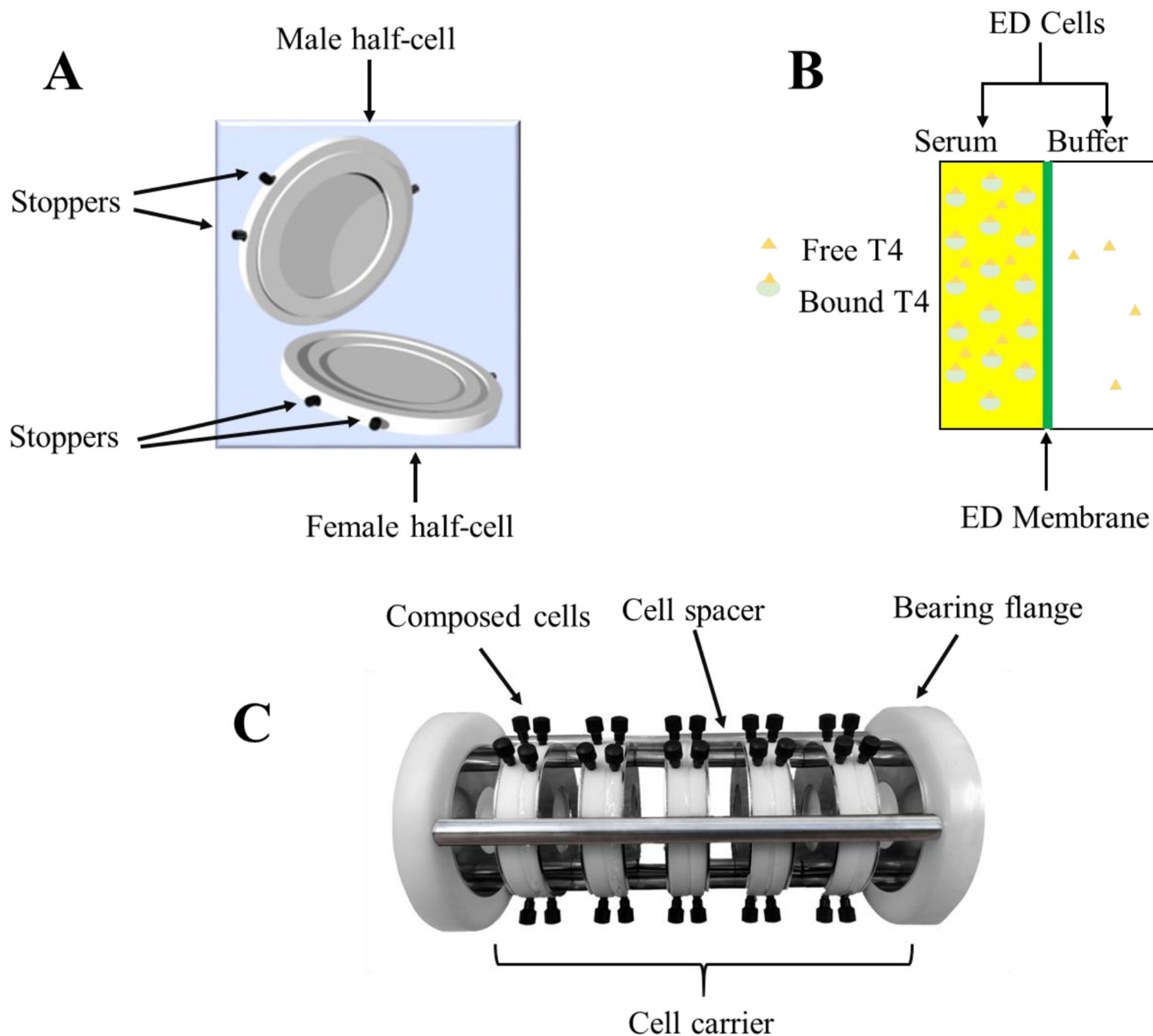


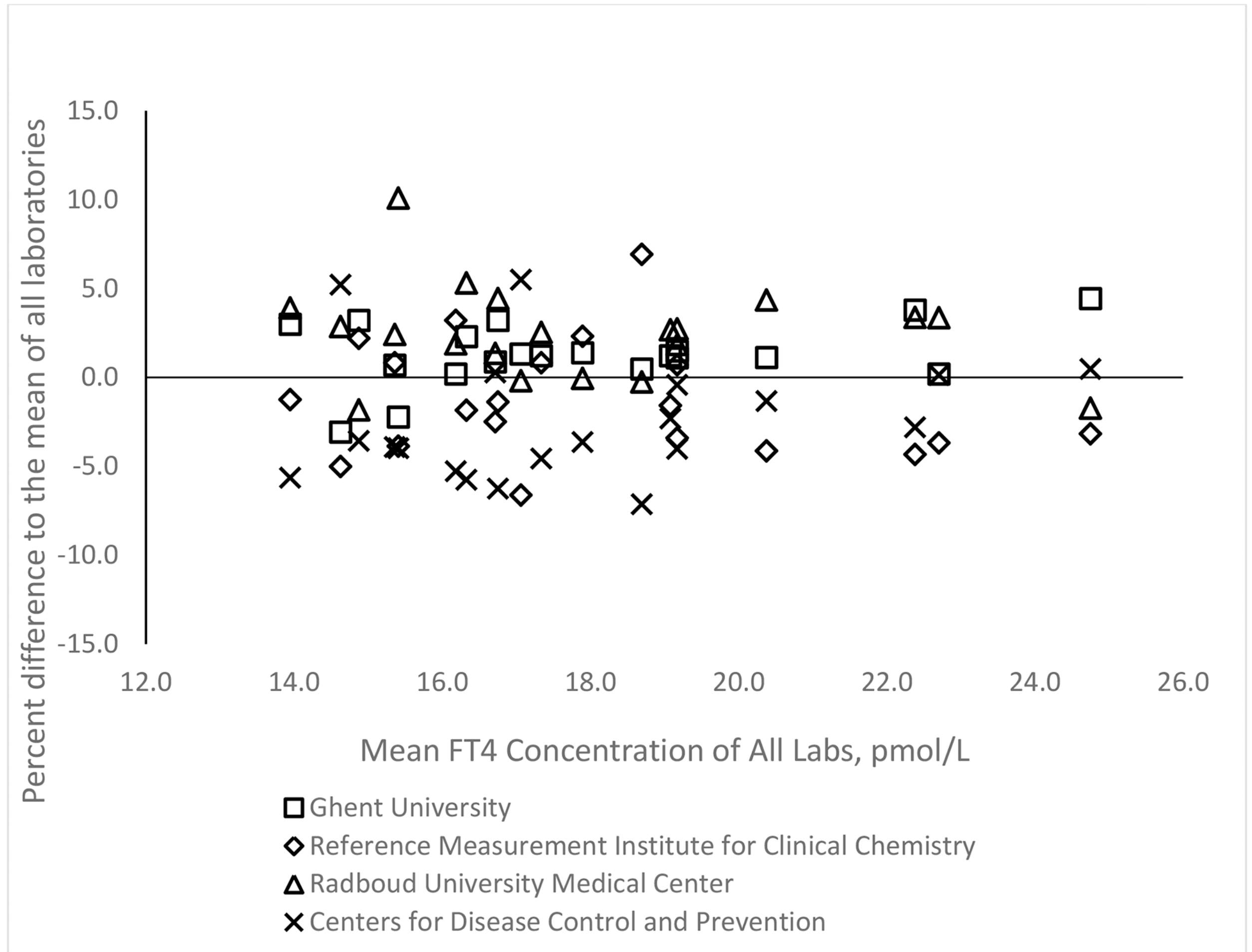
Figure 1.

Construction of equilibrium dialysis device for serum FT4 measurements

A): Dialysis cells used in cRMP; B): Dialysis cells were constructed by assembling the dialysis membrane between two PTFE half cells. C): The constructed dialysis cells were mounted in the cell carriers. One compartment of the dialysis cell was filled with serum, and the other compartment was filled with HEPES dialysis buffer.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

A.



Author Manuscript Author Manuscript Author Manuscript Author Manuscript

B.

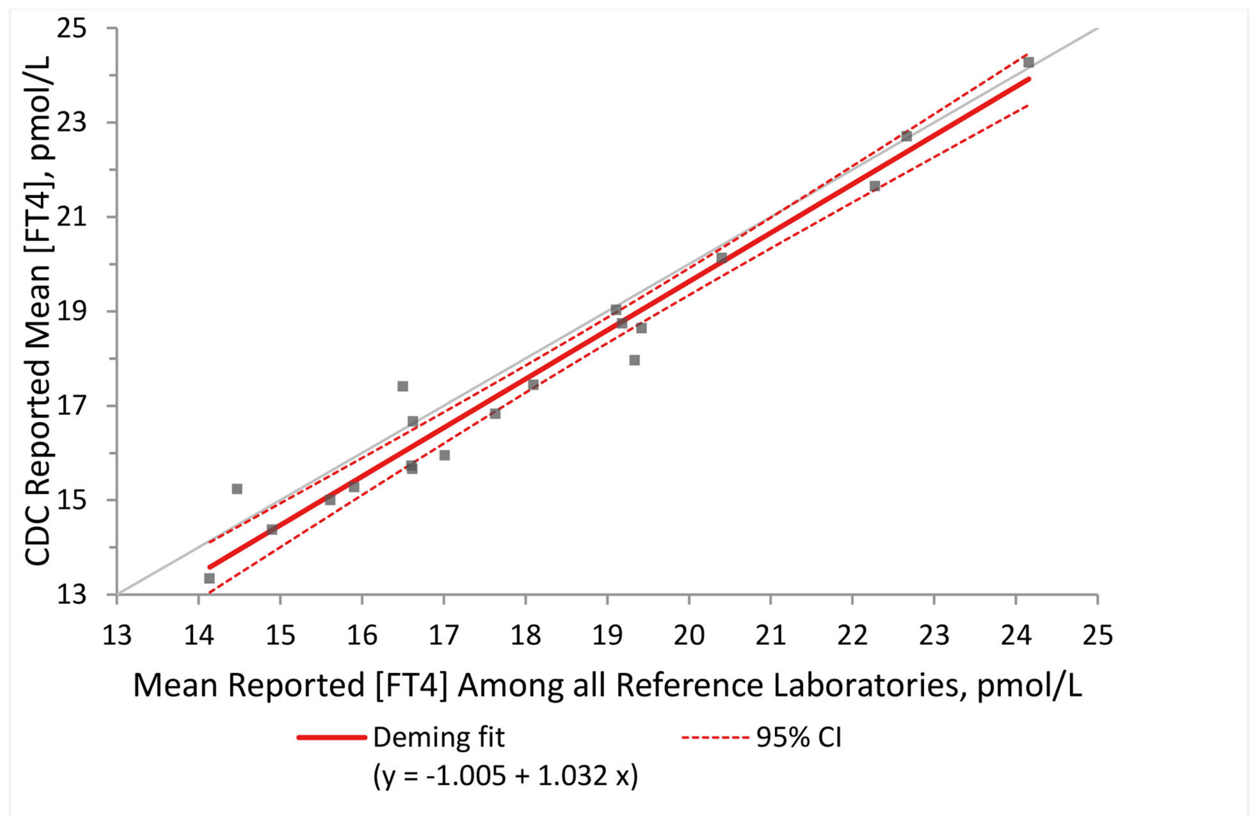


Figure 2.

Interlaboratory comparison of serum FT4 measurements

A set of 20 euthyroid single donor sera were measured independently by reference laboratories at the University of Ghent (Ugent, Belgium), the Centers for Disease Control and Prevention (CDC, USA), the Radboud University Medical Center Nijmegen (RUMCN, Netherlands), and the Reference Material Institute for Clinical Chemistry Standards (ReCCs, Japan).

(A) The percent differences of each laboratory were plotted against the mean of the four comparative laboratories.

(B) Deming regression analysis of the CDC candidate RMP and the mean FT4 concentration measured by all four reference laboratories.

Estimated regression statistics ± 95% CI and associated p-value were 1.03 ± 0.09 ($p = 0.48$) and −1.01 ± 1.80 ($p = 0.25$) for the slope and intercept, respectively.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

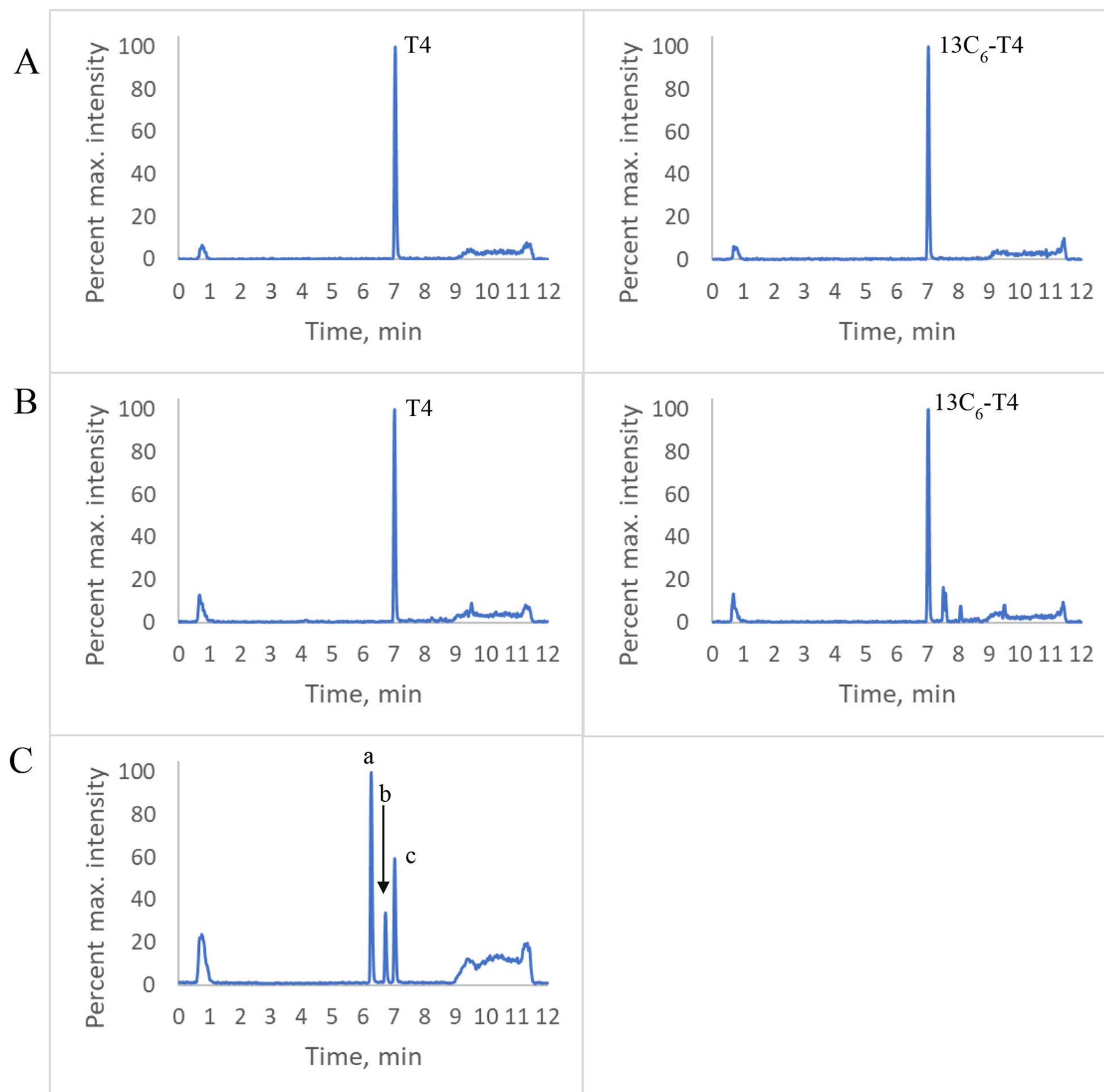


Figure 3.

Representative chromatograms of (A) neat sample (T4 concentration of 64.5 pmol/L, 2.6 pg on-column); (B) native serum dialysate sample (T4 concentration of 20.6 pmol/L, 2.4 pg on-column); and (C) TIC showing the separation of a) (T3), b) (rT3), and c) (T4) in neat sample (T4 concentration of 64.4 pmol/L, 2.5 pg on-column).

(T3, rT3, and T4 in neat samples were prepared in 10% acetonitrile in water containing 0.1% (v/v) formic acid, and serum dialysate samples were extracted with C18 SPE followed by LLE. Neat samples and serum dialysate samples were analyzed by UPLC-MS/MS, as described in the method section.)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1.

Interlaboratory comparison and Deming regression analysis results.

Laboratory	Slope		Intercept		Correlation (R^2)	Bias to Overall Mean (%) ± 95% CI	CDC Bias from Pairwise Comparison (%) ± 95% CI
	Estimate (95%CI)	p-value	Estimate (95%CI)	p-value			
UGent	1.07 (0.98–1.15)	0.11	−0.96 (-2.48–0.55)	0.20	1.00	1.3 ± 0.80	−1.9 ± 1.0
CDC	1.03 (0.94–1.13)	0.48	−1.01 (-2.80–0.79)	0.25	0.98	−2.5 ± 1.5	n/a
RUMCN	1.00 (0.88–1.12)	0.99	0.45 (-1.62–2.51)	0.66	0.99	2.5 ± 1.2	2.5 ± 1.1
ReCCS	0.97 (0.86–1.07)	0.51	0.36 (-1.45–2.18)	0.68	0.98	−1.3 ± 1.4	−0.6 ± 1.4

(A set of 20 euthyroid single donor sera were measured independently by reference laboratories at the University of Ghent (UGent, Belgium), the Centers for Disease Control and Prevention (CDC, USA), the Radboud University Medical Center Nijmegen (RUMCN, Netherlands), and the Reference Material Institute for Clinical Chemistry Standards (ReCCs, Japan) in 2018. The results of each lab were plotted against the mean of the four laboratories. Deming regression analysis was performed for the interlaboratory comparison. Mean bias was calculated as the percent difference of each laboratory to the mean of the 4 laboratories. The CDC bias was calculated as the percent difference of the CDC value to the pairwise mean.)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2.

Accuracy of the CDC cRMP when measuring hypo- and hyperthyroid samples

	CDC cRMP reported result	Ghent RMP reported result	Paired mean	CDC cRMP Bias to Paired Mean	
	Mean [FT4] ± SD, pmol/L	**Mean [FT4] ± SD, pmol/L**	**pmol/L**	**Mean Difference (%)**	**n**
Hypothyroid samples	4.13 ± 0.1	3.88 ± 0.3	4.01	3.1	2
	11.8 ± 0.9	12.1 ± 0.1	12.0	−1.0	3
Hyperthyroid samples	49.3 ± 2.2	54.3 ± 1.0	51.8	−4.8	3
	83.7 ± 3.9	87.8 ± 2.7	85.7	−2.4	3

Compared with the measurement of established FT4 RMP from University of Ghent, the mean difference (± 95% CI) between the established method and the CDC cRMP for two hypothyroid and two hyperthyroid samples was −1.8 ± 3.6%, which is consistent with the observation for euthyroid samples.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3:

Precision of cRMP measurements of serum FT4.

Materials	N	Concentration pmol/L	Imprecision Intra-Assay CV %	Inter-Assay CV %	Total CV %
Euthyroid serum (Level 1)	22	18.9	2.6	2.3	3.5
Hypothyroid serum (Level 2)	5	5.26	2.9	3.3	4.4
Hyperthyroid serum (Level 3)	7	36.0	2.1	3.6	4.1

(Imprecision was evaluated by analyzing three levels of serum-based QC materials in 2–5 replicates (inclusive independent ED) Intra-assay, inter-assay, and total imprecision of euthyroid range over >20 days were calculated according to CLSI EP5-A3[37]. Imprecision of two serum samples outside of the euthyroid range were analyzed over 5–7 days according to CLSI EP10-A3[47]).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 4:

Estimation of expanded uncertainties of cRMP measurements of serum FT4.

	Level 1	**Level 2**	**Level 3**
Concentration (pmol/L)	18.9	5.26	36.0
N	22	5	7
Relative Standard Uncertainty, Type A (%CV)	3.5	4.4	4.1
Relative Standard Uncertainty, Type B (%CV)	0.36	0.36	0.36
Combined Relative Standard Uncertainty, %	3.5	4.4	4.1
Coverage Factor	2	2	2
Relative Expanded Uncertainty, %	7.0	8.8	8.2

(Type A uncertainty was obtained from imprecision of the repeated measurements (n=5–22); Type B uncertainty was estimated from uncertainties in the purity of the primary reference material, inaccuracy in the weighing, and the measurement of the serum dialysate density.)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5.

Matrix effect evaluation for calibration curve (4.02–64.4 pmol/L).

	Neat	Method A	Method B	Method C	Method D
Slope (Mean ± SD)	1.00 ± 0.02	1.01 ± 0.04	1.02 ± 0.04	1.00 ± 0.005	1.02 ± 0.03
R^2	1.00	0.997	0.999	0.996	0.999
Difference to Neat Slope	-	−0.6%	2.8%	−0.2%	1.7%
Mean Matrix Effect	-	98.8 ± 12%	103 ± 10%	101 ± 26%	103 ± 16%

Calibration curves with 6 levels (4.02, 8.04, 16.1, 32.2, 48.3, and 64.4 pmol/L) were prepared in ethanol or pooled serum dialysate. The calibrators in the serum dialysate were prepared according to the extraction procedures A (SPE followed by LLE using ethyl acetate), B (sequential LLE using cyclohexane followed by ethyl acetate), C (LLE using ethyl acetate), and D (SLE), respectively. The matrix-based calibration curves were compared with neat calibration curves to determine the bias of their slopes as compared to neat slopes, as well as mean matrix effect.)