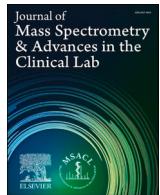




Contents lists available at ScienceDirect

# Journal of Mass Spectrometry and Advances in the Clinical Lab

journal homepage: [www.sciencedirect.com/journal/journal-of-mass-spectrometry-and-advances-in-the-clinical-lab](http://www.sciencedirect.com/journal/journal-of-mass-spectrometry-and-advances-in-the-clinical-lab)



## Practical considerations for accurate determination of free thyroxine by equilibrium dialysis

Ashley Ribera <sup>a,\*</sup>, Li Zhang <sup>a</sup>, Carla Ribeiro <sup>b</sup>, Norma Vazquez <sup>c</sup>, Janet Thonkulpitak <sup>c</sup>, Julianne C. Botelho <sup>a</sup>, Uliana Danilenko <sup>a</sup>, Katleen van Uytfanghe <sup>b</sup>, Hubert W. Vesper <sup>a</sup>

<sup>a</sup> Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, GA, United States

<sup>b</sup> Ref4U, Laboratory of Toxicology, Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium

<sup>c</sup> Battelle Memorial Institute, Atlanta, GA, United States



### ARTICLE INFO

#### Keywords:

Equilibrium dialysis  
Free thyroxine  
Thyroxine standardization  
Pre-analytical considerations

### ABSTRACT

**Background:** Free thyroxine (FT4) measurement is one of the most requested tests in patient care for diagnosing and treating thyroid-related illnesses. Equilibrium dialysis (ED) is considered the “gold standard” for FT4 measurement; however, several factors have a profound effect on the reliability of FT4 assays and require special consideration.

**Methods:** In the current study, we focused on evaluating critical factors that could contribute to reporting errors, such as adsorption of thyroxine (T4) to labware surfaces, stability of serum samples, stock solutions, and calibrator storage conditions, as well as the solvents used to prepare T4 solutions.

**Results:** The adsorption of T4 in ethanolic solutions and dialysates to labware surfaces can be reduced with the careful selection of pipette tips, test tubes, and 96-well plates. Adding pH modifiers to neat T4 solutions can improve its stability. FT4 in serum samples remains stable after exposure to four freeze–thaw cycles, 5 °C for 18–20 h, or –70 °C for a minimum of three years.

**Conclusion:** The presented study has demonstrated that the loss of analyte due to pre-analytical and analytical factors during operation of the FT4 reference measurement procedure (RMP) can be minimized by careful selection of all labware for sample preparation. It was found that the accuracy and imprecision of FT4 assays can be influenced by different types of dialysis devices, but acceptable alternatives to ED membranes were identified. This study demonstrates approaches to establish a FT4 method that is independent from specific suppliers and addresses critical pre-analytical and analytical factors important for FT4 measurements.

### 1. Introduction

Free thyroxine (FT4) measurement is an important part of thyroid function testing and is used to diagnose and treat thyroid disorders. Thyroxine (T4) is the precursor to the biologically active thyroid hormone triiodothyronine, which plays a crucial role in metabolism, temperature regulation, energy levels, heart rate, fertility, and fetal development [1,2]. FT4 can be measured using direct methods and immunoassays (IAs). Direct methods of measuring FT4 involve separating free T4 from protein-bound by either equilibrium dialysis (ED) or

ultrafiltration [3,4]. It is essential to preserve the endogenous equilibrium between free and bound T4 during isolation of the free hormone fraction in order to accurately quantify FT4 in patient samples.

The IAs usually do not apply physical separation of thyroxine and use antigen–antibody interactions to mimic actual FT4 separation. The preanalytical and analytical challenges of these IAs have been investigated in several studies [5,6]. Many factors can affect the reliability of FT4 IA measurements, such as abnormal binding proteins, protein binding displacers, heterophilic antibodies, autoantibodies, free fatty acids, assay antibodies, analogs, and/or serum dilution [7,8]. However,

**Abbreviations:** ED, equilibrium dialysis; FMED, Fast Micro-Equilibrium Dialyzer; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; IA, immunoassay; IFCC, International Federation of Clinical Chemistry and Laboratory Medicine; IS, Internal standard; LC-MS/MS, liquid chromatography tandem mass spectrometry; MIT, mono-iodotyrosine; MWCO, Molecular weight cutoff; qNMR, Quantitative nuclear magnetic resonance; RMP, Reference measurement procedure; SPE, solid phase extraction; WS, Working solution.

\* Corresponding author.

E-mail address: [aribera@cdc.gov](mailto:aribera@cdc.gov) (A. Ribera).

<https://doi.org/10.1016/j.jmsacl.2023.06.001>

Received 10 November 2022; Received in revised form 17 May 2023; Accepted 20 June 2023

Available online 23 June 2023

2667-145X/© 2023 The Authors. Published by Elsevier B.V. on behalf of The Association for Mass Spectrometry: Applications to the Clinical Lab (MSACL). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

there is very limited information available about pre-analytical and analytical challenges of ED-based methods. In this manuscript, we provide practical experience addressing these challenges when performing ED-based FT4 methods. FT4 is defined by the international conventional reference measurement procedure (RMP) based on ED combined with determination of the T4 concentration in the dialysate with a trueness-based isotope dilution-mass spectrometry method endorsed by the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) [9]. The principles of the consensus method endorsed by the IFCC are strictly maintained by laboratories conducting RMPs; however, these conventions only apply to the ED step of the procedure [3]. While the ED step is considered conventional and must be meticulously maintained, it is important to gain knowledge about what can be altered in this step without causing results to change. It is also important to understand which factors can lead to differences in results. This will ensure that the procedure can be maintained over time and does not rely on specific product manufacturers or supplies that may become discontinued in the future. In addition, principles and knowledge about pre-analytical and analytical factors can also be beneficial for clinical laboratories performing routine FT4 measurements using an ED-based approach.

Deviations from the time, temperature, buffer composition, dilution, and pH conventions of ED may affect the affinity of binding proteins in samples and profoundly alter FT4 concentration. This can impact the accuracy and imprecision of results. The materials and devices used for ED can also influence FT4 present in the sample at equilibrium; T4 can be adsorbed on the surface of sample containers, membranes, and other labware that is in contact with patient samples [10]. Previous studies with radioactive T4 have suggested that labware choice plays an important role in maintaining T4 concentration in solutions [11]. In addition, dialysis devices have inherent properties that can affect dialysis time and dilution [12]. The limited availability of suitable dialysis devices and membranes necessitates careful selection of alternatives to ensure continuity of measurement using the reference system and avoid dependence on the dialysis device itself as a convention for the RMP. Storage and preparation conditions of samples and calibration solutions may also influence accuracy of patient results [13,14]. This study aims to evaluate pre-analytical factors that may influence FT4 concentration in the context of an ED-based FT4 RMP. Observations regarding solution stability and FT4 recovery are applicable to routine FT4 methods as well. The methods described here are meant to serve as a guide to evaluating potential sources of inaccuracy and imprecision in FT4 methods. The study was conducted at two locations operating independent FT4 RMPs: CDC Hormone Reference Laboratory and at the Reference Laboratory at Ghent University (Ref4U). Previously completed and ongoing studies have demonstrated excellent comparability between these two RMPs [15,16].

## 2. Materials and methods

### 2.1. Materials

Where applicable, different, interchangeable materials and methods used for analysis conducted at either CDC or Ref4U are listed. L-Thyroxine certified reference material IRMM-468 was obtained from the Joint Research Centre (Geel, Belgium) and Sigma-Aldrich® (St. Louis, MO, USA) [17]. L-Thyroxine-<sup>13</sup>C<sub>6</sub> (100 µg/mL) and 99.96 isotopic purity dimethyl sulfoxide (DMSO)-d<sub>6</sub> were procured from Sigma-Aldrich® (St. Louis, MO, USA) for use by CDC. L-Thyroxine-<sup>13</sup>C<sub>9</sub> was from the Service de Chimie et Biochimie Appliquées, Faculté Polytechnique de Mons (Belgium) or Sigma-Aldrich (St. Louis, MO, USA) for use by Ref4U. All serum materials were purchased from Solomon Park Research Laboratories (Burien, WA). Solomon Park Research Laboratories has IRB approvals to collect blood and obtained informed consent from donors. Use of blood by CDC is consistent with the IRB approval and donor consent. No personal identifiers were provided to CDC. A 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) dialysis buffer was prepared

according to the published procedures [18]. Custom HEPES dialysis buffer kits were from ABI Scientific (Sterling, VA, USA). HEPES pH adjusting buffer was prepared to contain 776 mM HEPES. All samples and buffers were adjusted to pH 7.4 ± 0.03 at 37 °C before use with analytical-grade hydrochloric acid or 10 N sodium hydroxide. All reagents used were of analytical grade or better. Borosilicate glass 16 × 100 mm culture tubes were purchased from DWK Life Sciences (Milville, NJ, USA). Silanized clear borosilicate glass 2 mL vials were purchased from Thermo Fisher Scientific (Waltham, MA, USA). Polypropylene 2 mL 96-well plates were purchased from Arctic White LLC (Bethlehem, PA, USA). Sep-Pak 1 mL C18 solid-phase extraction (SPE) cartridges were purchased from Waters™ (Milford, MA, USA). Positive displacement 1 mL pipette tips were purchased from Gilson™ (Middleton, WI, USA). Four commercially available polytetrafluoroethylene (PTFE) dialysis cells: the Dianorm® Macro 1, the Dianorm® Micro 1S, the Dianorm® Micro, and the Fast Micro-Equilibrium Dialyzer®; and a Multi-Equilibrium Dialyzer were purchased from Harvard Apparatus (Holliston, MA, USA). Dianorm® regenerated cellulose membranes with 5 kDa and 10 kDa cutoffs were purchased from Harvard Apparatus. Spectra/Por regenerated cellulose 3.5 kDa membranes were purchased from Repligen (Boston, MA, USA). Quantitative <sup>1</sup>H-Nuclear Magnetic Resonance spectroscopy (qNMR) was performed by the Complex Carbohydrate Research Laboratory at the University of Georgia on a Bruker 600.06 MHz spectrometer with a 5 mm cryoprobe (Billerica, MA). Analysis by liquid chromatography tandem mass spectrometry (LC-MS/MS) was performed on either a Waters™ Acquity TQD mass spectrometer system with electrospray ionization source coupled with a Waters™ Acquity liquid chromatograph at Ref4U (Milford, MA, USA), or a Shimadzu™ LC-30AD HPLC module (Kyoto, Japan) coupled with an AB SCIEX Triple Quad™ API 5500 Mass Spectrometer at CDC (Framingham, MA, USA). Analytical parameters for both systems are listed in supplementary tables S3 and S4.

### 2.2. Preparation of stock solutions and calibrators

Certified L-thyroxine primary reference material IRMM-468 (Joint Research Centre, Geel, Belgium) was used to prepare calibrator solutions for FT4 quantification. All solutions were prepared gravimetrically. The stock solution, intermediate solution, and working solution (WS) of T4 used by Ref4U were prepared as described in supplementary table S1. Monoiodotyrosine (MIT) was added to intermediate (0.1 µg/g T4) and working (1 ng/g T4) calibrator solutions prepared at Ref4U at a concentration 5000 times higher than the T4 concentration as a protective carrier [19]. All solutions prepared at CDC contained a 1.7% solution of ammonium hydroxide in ethanol as solvent unless otherwise noted. Ammonium hydroxide (28.0–30.0% NH<sub>3</sub>, Extra Pure) was purchased from ACROS Organics (Fair Lawn, NJ). The stock solution, intermediate solution and working solution of T4 used by CDC were prepared as described in supplementary table S2. Similarly, solutions of isotopically labelled internal standard (ISWS) were prepared gravimetrically to a final concentration of 0.1 ng <sup>13</sup>C<sub>6</sub>-T4/g solvent.

### 2.3. Stability of T4 in stock and calibration solutions

Aliquots of 50 pg/mL (64.7 pmol/L) T4 solutions were prepared in solvent A (100% ethanol), solvent B (1.7% ammonium hydroxide in ethanol), solvent C (4% acetic acid in ethanol), and solvent D (4% formic acid in ethanol). The short-term stability of T4 in the above solvents was evaluated over 7 days at room temperature to determine the adsorption of FT4 to the glass tubes used for sample dialysate collection and calibration curve preparations. Aliquots of 3 mL of each solution were placed into glass sample tubes. On days 0, 1, 2, 5, and 7, 200 µL aliquots were taken from the 3 mL solution in glass tubes and 100 µL of ISWS was added to each aliquot. The T4 in the samples was quantified using LC-MS/MS (Sciex API 5500 system).

Long-term stability of T4 stock and calibration solutions was

determined using HPLC coupled with a UV spectrophotometric detector (HPLC-UV). A 1 mg/g T4 solution in DMSO-*d*<sub>6</sub> was prepared and its concentration was confirmed by qNMR. This solution (“RS-A”) was used to verify the concentrations of Stock A and Stock B T4 solutions. A 5-point reference calibration curve of 5.08–27.4 µg/g T4 in 1.7% ammonium hydroxide (NH<sub>4</sub>OH) in ethanol containing 2% v/v DMSO from RS-A was freshly prepared, and the absorbance of these solutions at 250 nm was used to verify the concentration of T4 in the calibration solutions Stock A and Stock B by linear regression. Dilutions of Stock A containing 17.0 µg/g T4 (17.3 µmol/L, “UV-A”) and Stock B containing 0.424 µg/g T4 (0.432 µmol/L, “UV-B”) in 1.7% NH<sub>4</sub>OH in methanol containing 2% (v/v) DMSO were prepared in triplicate and diluted 1:1 with 1% (v/v) formic acid prior to analysis by HPLC-UV. The concentrations of T4 WS were below the limit of detection of the UV detector used, so they could not be directly confirmed; instead, a new 1 ng/g T4 WS was freshly prepared from the verified Stock B solution, and its gravimetric concentration was compared to the concentration calculated by linear regression from calibrators prepared from the WS being verified.

#### 2.4. Stability of FT4 in serum, serum dialysates, and extracted samples

Freeze-thaw (-70 °C to room temperature) stability was evaluated by measuring FT4 of serum samples subjected to four freeze-thaw cycles before sample preparation. The long-term stability of serum at -70 °C was evaluated periodically for 3 years. Stability at 5 °C was evaluated by comparing serum samples thawed and prepared immediately from -70 °C to those thawed at 5 °C overnight (18–20 hrs) before sample preparation. To determine the extent of FT4 adsorption in dialysate, 1 mL aliquots from a serum dialysate pool were transferred into glass sample tubes and spiked with 100 µL ISWS at 0, 24, and 48 h. T4 was extracted from serum and dialysate samples prior to analysis by LC-MS/MS as described previously [19]. The on-board stability of extracted samples at 5 °C was evaluated by repeated analysis of extracted samples via LC-MS/MS after they had been placed in an auto-sampler at 5 °C for up to 4.8 days. Stability was assessed based on principles discussed in Section 2.7. Statistical analysis.

#### 2.5. Evaluation of adsorption of T4 to the labware surface

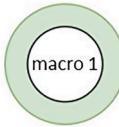
The percent of thyroxine recovered after exposure to commonly used labware was determined according to a previously published procedure [20]. A schematic representation of the adsorption testing method is shown in [supplementary figure S1](#). Two sets of samples, one spiked with isotopically labelled internal standard before exposure to labware and one spiked after, were prepared using 25 pg/mL (32.4 pmol/L) T4 solutions in three different solvent compositions to evaluate recovery. To prepare set 1, which served as a control set, an aliquot of 25 pg/mL (32.4 pmol/L) T4 solution prepared in either solvent A, solvent B, or 10% acetonitrile in water with 0.1% formic acid (solvent E) was added to each of the labware listed in [Table 2](#). A single pipette was used to transfer half of the initial T4 solution volume to a new test tube before adding an equal volume of 25 pg/mL (32.4 pmol/L) <sup>13</sup>C<sub>6</sub>-T4. Five replicates of each control sample were prepared per solvent. Set 2 was prepared similarly to set 1, except the T4 solutions were transferred to test tubes and spiked with internal standard after the solutions A, B or E were exposed to the tested labware an additional one or five times. Identical tubes and pipettes were used in the same quantities as the control set for all samples in set 2 to minimize the impact of T4 adsorption on these common surfaces on the T4 recovery. All solutions and sample extracts were dried under nitrogen flow at room temperature and reconstituted in 200 µL 10% acetonitrile in water with 0.1% formic acid. All samples were analyzed by LC-MS/MS according to the reference method procedure at CDC [16]. The percent of T4 recovered after exposure to labware was determined according to the formula 100\*B/A, where A is the mean T4/<sup>13</sup>C<sub>6</sub>-T4 peak area ratios of set 1 and B is the peak area ratios of set 2.

#### 2.6. Robustness of the conventional equilibrium dialysis methods

The variation resulting from the use of different dialysis cells and membranes, as well as the time required to reach equilibrium during dialysis, were evaluated for robustness of the conventional equilibrium dialysis method. FT4 was measured according to the reference measurement procedure for FT4 endorsed by the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) committee for Standardization of Thyroid Function Tests (C-STFT) [19]. Briefly, the

**Table 1**

Equilibrium dialysis cells and membranes used for ruggedness testing.

Cell types				
Macro 1	Macro 1S	Micro	Fast Micro-Equilibrium Dialyzer	
				
Max. volume (ml)	1.36	1.15	0.25	1.00
Working volume (ml)	1.00	1.00	0.2	1.00
Min. volume (ml)	0.70	0.60	0.125	-
Membrane area (cm <sup>2</sup> )	4.50	11.30	2.0	2.30
Membrane area/working volume ratio	4.50	11.30	10	2.30

Dialysis cells and membranes used for ruggedness testing. Three equilibrium dialysis cell types (Macro 1S, Micro, and FMED) were evaluated by the reference measurement procedure at Ref4U by comparing results obtained using the validated dialysis cells routinely used by the method (Macro 1) to results of the same set of samples measured with Macro 1S, Micro, and FMED.

**Table 2**

Percent recovery of T4 in neat solutions after exposure to common labware surfaces.

A		Recovery of T4 after exposure to common lab items, mean percent recovered $\pm$ SD					
Type of labware	Exposures	96-well 2 mL plates	n	14 mL clear glass tubes	n	Silanized clear 2 mL glass vials	n
Solvent A,	1	101 $\pm$ 2.5	5	98.7 $\pm$ 3.0	5	100 $\pm$ 7.8	5
100% Ethanol	5	98.3 $\pm$ 1.6	5	98.6 $\pm$ 2.5	4	94.5 $\pm$ 4.5	5
Solvent B,	1	100 $\pm$ 2.2	5	99.2 $\pm$ 3.3	5	99.9 $\pm$ 3.2	5
1.7% ammonium hydroxide in ethanol	5	101 $\pm$ 0.1	4	101 $\pm$ 2.4	5	97.0 $\pm$ 3.2	5
Solvent E,	1	98.2 $\pm$ 2.8	5	100 $\pm$ 1.0	5	98.3 $\pm$ 1.1	5
10% acetonitrile with 0.1% formic acid	5	93.7 $\pm$ 2.8	5	96.8 $\pm$ 2.9	5	89.9 $\pm$ 3.9	5

B Mean percent difference between 1 and 5 exposures						
	96-well 2 mL plates	p-value ( $\alpha=0.05$ )	14 mL clear glass tubes	p-value ( $\alpha=0.05$ )	Silanized clear 2 mL glass vials	p-value ( $\alpha=0.05$ )
Solvent A	−2.8 $\pm$ 1.6	0.11	−0.1 $\pm$ 2.5	0.95	−5.3 $\pm$ 4.5	0.15
Solvent B	1.5 $\pm$ 0.1	0.21	2.1 $\pm$ 2.4	0.32	−2.9 $\pm$ 3.2	0.15
Solvent E	−4.6 $\pm$ 2.9	0.14	−3.2 $\pm$ 2.9	0.07	−8.6 $\pm$ 4.0	<0.05

Neat T4 solutions prepared in either solvent A (100% ethanol), solvent B (1.7% ammonium hydroxide in ethanol), or solvent E (10% v/v acetonitrile in water with 0.1% formic acid) were exposed to common labware (2 mL polypropylene 96-well plates, borosilicate glass 16  $\times$  100 mm test tubes, or silanized clear 2 mL borosilicate autosampler vials) 1 and 5 times.

A). The percent of T4 recovered after exposure to labware was determined according to the formula  $100^*B/A$ , where B is the mean area ratio of samples spiked with internal standard post-exposure to labware, and A is the mean area ratio of the solutions spiked with internal standard pre-exposure.

B). The mean percent difference of recovery after 1 exposure to recovery after 5 exposures was evaluated for each labware and solvent using the paired *t*-test ( $\alpha = 0.05$ ).

dialysis cells listed in Table 1 were composed with a 5 kDa regenerated cellulose membrane between two PTFE half-cells. Additional Macro 1 cells were assembled with 10 kDa, 5 kDa or 3.5 kDa membranes for comparison. Variations of membrane molecular weight cut-off (MWCO) and suppliers tested were limited to regenerated cellulose due to documented binding of T4 to polysulfone and polyethersulfone membranes [3]. Serum was adjusted to  $pH 7.4 \pm 0.03$  at  $37^{\circ}C$  by adding a maximum of 10% of the serum volume as HEPES pH adjusting buffer to each serum sample. pH-adjusted serum was loaded into the serum compartment of each cell type listed in Table 1. The equal volume of HEPES dialysis buffer was added to the opposite half-cell (buffer compartment). Due to differences in membrane surface area, cells were incubated at  $37^{\circ}C$  for 3–6 h; the time required to reach equilibrium was determined for each cell type as the period where T4 levels in dialysates remained constant. Dialysates were collected in tared glass vials containing an appropriate amount of  $^{13}C$ -labeled T4 as internal standard equivalent to the thyroxine concentration of the dialysate. T4 was extracted from dialysate matrix components and concentrated prior to analysis by LC-MS/MS according to previously published procedures [19].

## 2.7. Statistical analysis

The Student's paired *t*-test was used to determine the significance of difference in FT4 concentration among samples that had undergone 1 or 4 freeze–thaw cycles, to determine the difference in recovery after 1 and 5 exposures to common labware among different solvents, and to determine the significance of difference between the T4 Stock B solution gravimetric concentration and the concentration determined by HPLC-UV after 1.5 years of storage. All other stability tests were evaluated by linear regression according to the principles listed in the Clinical and Laboratory Standards Institute (CLSI) document EP25-A [21]. Statistical differences between sample means obtained from the dialysis cell and dialysis membrane types listed in Tables 5a and 5b were evaluated using one-way repeated measures ANOVA. Pairwise differences between sample means among dialysis device and membrane type were evaluated as needed using *p*-values adjusted for multiple comparisons. Statistical significance was evaluated at a 95% significance level for all tests. Data analysis was done using Microsoft Excel® with the Analyse-it® add-in or the R statistical environment in R Studio®.

## 3. Results and discussion

Strict adherence to the ED step of the FT4 RMP is necessary to ensure equilibrium dialysis assays preserve endogenous equilibrium and produce accurate, reproducible FT4 measurements [3,19]. The composition of dialysis buffer, ED temperature, and time were not part of this investigation. Previously published literature discusses the importance of preserving conventional buffer composition that is close to the composition of the ultrafiltrate of normal human serum; departure from established buffers may result in inconsistent results [3,18]. In addition to conventions for dialysis time, temperature, and buffer composition, labware and solvents chosen during analysis can present further challenges towards reliable measurements. This study provides an example and investigates the influence of these pre-analytical and analytical factors on FT4 measurement results.

### 3.1. Adsorption of T4 to labware surfaces

The recovery of neat T4 in various solvents among commonly used labware is summarized in Table 2. The change in concentration of neat T4 solutions prepared in either 100% ethanol (solvent A), 1.7% NH<sub>4</sub>OH in ethanol (solvent B), or 10% acetonitrile in water with 0.1% formic acid (solvent E) was determined after one and five consecutive exposures of 0.5 mL of each solution to polypropylene 2 mL 96-well plates, untreated clear borosilicate glass test tubes, or silanized clear borosilicate glass vials. The recovery of T4 in solvent A or B from 96-well plates or test tubes was higher than the recovery after exposure to silanized vials, and the recoveries among all labware tested using solvent E were lower than for solvents A or B. The mean percent difference between one exposure and five exposures to each labware for the three solvents is summarized in Table 2B. No significant difference in recovery was found for solvents A, B, or E after one and five exposures to 96-well plates or test tubes. There was no significant difference in the recovery of T4 in solvent A and B in silanized vials after one and five exposures; however, there was a significant difference (*p*-value  $< 0.05$ ) when comparing the recovery of T4 in solvent E using silanized vials, which suggests it is necessary to pre-screen labware that will be in direct contact with neat solutions of T4 to avoid analyte loss.

### 3.2. Stability of T4 in stock and calibration solutions

Changes in T4 concentration and its association with storage temperature, duration, and solvent are summarized in Table 3. The percent difference of T4 concentrations determined by HPLC-coupled UV spectrophotometry compared to gravimetric concentrations was  $-0.2\%$  for Stock A (after 2.9 years) stored at  $-70^{\circ}C$  and  $3.3 \pm 4.6\%$  for Stock B

**Table 3**

Short-term and long-term stability of T4 in neat solutions. Short-term stability of neat T4 solutions prepared in either solvent A (Ethanol), solvent B (1.7% ammonium hydroxide in ethanol), solvent C (Ethanol with 4% acetic acid), or solvent D (Ethanol with 4% formic acid) was evaluated daily from 1 to 7 days at room temperature (RT). Percent change in T4 concentration was determined as the percent difference in measured concentration of each solution compared to the initial solution concentration. Long-term stability of neat T4 solutions prepared in solvent B (1.7% ammonium hydroxide in ethanol) was evaluated over 2.6 years at 20 °C. Percent change in T4 concentration ( $\pm$ SD) was determined as the percent difference in measured concentration of each solution compared to the initial solution concentration at time 0 for serum or the initial gravimetric concentration for neat T4 solutions.

Solution	Solvent	Test Condition	Duration	Initial [T4] (95% CI)	n	Final [T4] (95% CI)	n	Percent Difference $\pm$ SD	p-value ( $\alpha = 0.05$ )
T4 Stock A	B	Storage at -70 °C	2.9 years	450 $\mu$ g/g	–	449(436–462)	3	-0.2 $\pm$ 0.6	0.98
T4 Stock B	B	Storage at -70 °C	1.5 years	0.424 $\mu$ g/g	–	0.438(0.411–0.466)	4	3.3 $\pm$ 4.6	0.19
T4 WS	B	Storage at -20 °C	2.6 years	1.05 ng/g	–	1.06(1.04–1.08)	5	1.0 $\pm$ 3.3	0.70
50 pg/mL T4	A	Storage at RT	7 days	46.2 pg/mL (40.6–51.8)	3	46.4(42.6–50.3)	3	0.5 $\pm$ 3.4	0.46
50 pg/mL T4	B	Storage at RT	7 days	47.3 pg/mL (42.8–51.8)	3	48.9(46.7–51.1)	3	3.4 $\pm$ 1.9	0.06
50 pg/mL T4	C	Storage at RT	7 days	50.0 pg/mL (42.7–57.3)	3	48.3(43.5–53.1)	3	-3.4 $\pm$ 3.9	0.59
50 pg/mL T4	D	Storage at RT	7 days	48.7 pg/mL (42.3–55.0)	3	41.8(33.7–49.9)	3	-14.1 $\pm$ 6.7	0.001

(after 1.5 years) stored at the same temperature. Concentration of the T4 WS was within  $1.0 \pm 3.3\%$  of the gravimetric concentration when determined by linear regression using a WS prepared 2.6 years previously, suggesting that it is stable during this time period when stored at -20 °C. Solutions of T4 (50 pg/mL) in solvents A, B, and C were within -3.4–3.4% of the expected gravimetric concentration after 7 days at room temperature, confirming their stability with no significant observed T4 loss during storage. Solutions of T4 prepared in solvent D were significantly reduced in 7 days (p-value: 0.001); the decrease in T4 concentration of 14.1% from day 1 to day 7 may be due to esterification of formic acid in the presence of alcohol, which would change the solution pH over time and thus make it unsuitable as a solvent when preparing T4 solutions under these conditions. Based on these data, future T4 stock and WS should be prepared in either 100% ethanol or 1.7% NH<sub>4</sub>OH in ethanol and stored at -20 °C to prevent stability-related changes to the T4 concentration.

### 3.3. Stability of FT4 in serum and serum extracts

Changes in T4 concentration with changing temperature and storage duration conditions are summarized in Table 4. The mean percent difference in T4 concentration among serum samples with FT4 concentrations of 16.1–20.2 pmol/L between one and four freeze–thaw cycles was  $-0.5 \pm 3.6\%$ , indicating that the serum was not significantly changed after undergoing four freeze–thaw cycles before sample preparation. FT4 levels in serum were stable at 5 °C for 18–20 h. Repeated analysis of dialysate extracts from serum with 15.7–25.7 pmol/L stored in an auto-sampler at 5 °C for 4.8 days was reproducible, with a mean percent difference of  $0.6 \pm 2.5\%$ . FT4 in serum stored at -70 °C

remained relatively unchanged from the initial concentration after storage for 3.3 years, indicating stability when stored at this temperature and duration. FT4 concentration in serum dialysate samples after 48 h at room temperature was within  $-1.0 \pm 8.9\%$  of the initial concentration, indicating there was no significant loss of T4 in the collection tubes and that dialysate solutions are stable for at least 48 h at room temperature.

### 3.4. Comparison among different dialysis equipment types

Macro 1S, Micro, and Fast Micro-Equilibrium Dialyzer (FMED) cells were tested for their ability to reach equilibrium within five hours and compared to results for the Macro 1 cells validated for use by the FT4 reference method under identical conditions, as summarized in Table 5A. Reaching equilibrium was determined by observing no significant change in concentration of serum samples between the fourth and fifth hour, which was determined using the Student's paired *t*-test. No significant difference was found between FT4 when dialyzing a sample (18 pg/g) for three, four, five or six hours (within run, n = 5 per time point, p = 0.58,  $\alpha = 0.05$ ) for the Macro 1S and between four and five hours for the FMED cells (within run, n = 8 per time point, p = 0.35,  $\alpha = 0.05$ ), suggesting that all were able to reach equilibrium within five hours. Sample FT4 concentrations (8.1–17 pg/g) from these cells were compared to the results of the dialysis cells used for the FT4 RMP and evaluated based on the limits for imprecision (<5%) and bias ( $\pm 2.5\%$ ) established for the RMP [15]. The Macro 1S cells met the bias and imprecision criteria for the RMP with a mean bias of 1.2% to the results of the Macro 1 cells used by the RMP and imprecision of 2.1% calculated over two days (n = 16). The Micro dialysis cells and FMED cells did not

**Table 4**

Stability of FT4 in serum and serum extracts. Serum samples were evaluated to discern the freeze–thaw (-70 °C to room temperature) stability over four cycles before sample preparation in comparison to the same serum samples that were thawed and prepared immediately from frozen (1 freeze–thaw cycle). The mean [FT4] in pmol/L from both sets of samples were compared using the student's paired *t*-test ( $\alpha = 0.05$ ). Long term serum stability during storage at -70 °C, stability of serum dialysate extracts during analysis by LC-MS/MS, and short-term stability of serum stored at 5 °C was evaluated according to CLSI EP25-A [21].

Sample ID	Test Condition	Duration	[FT4], pmol/L (95% CI)		Mean Percent Difference $\pm$ SD				
			Initial Condition	n	Final Condition	n	Individual Sample	All samples	p-value ( $\alpha = 0.05$ )
Sample 1	Freeze-thaw	4 cycles	20.2(19.6–20.7)	6	19.9(18.7–21.1)	3	-1.3 $\pm$ 2.3	-0.5 $\pm$ 3.6	0.19
Sample 2			19.2(18.3–20.2)	6	19.8(16.9–22.7)	3	3.0 $\pm$ 6.0	0.38	
Sample 3			18.8(17.6–19.9)	6	19.7(16.4–22.9)	3	4.8 $\pm$ 6.9	0.11	
Sample 4			16.3(15.6–17.0)	7	15.5(14.3–16.7)	8	-4.7 $\pm$ 8.8	0.26	
Sample 5			16.1(15.4–16.7)	7	15.7(14.7–16.6)	8	-2.6 $\pm$ 7.0	0.39	
Sample 6			18.9(17.9–19.9)	7	18.5(17.9–19.0)	9	-2.2 $\pm$ 3.8	0.32	
Sample 1	Storage at -70 °C	3.3 years	19.6(19.2–19.9)	5	19.8(19.2–20.3)	6	1.0 $\pm$ 2.7	1.0 $\pm$ 2.7	0.46
Sample 2	Storage at 5 °C	18–20 h	19.1(18.7–19.5)	17	19.3(18.7–19.9)	12	1.4 $\pm$ 5.0	-0.4 $\pm$ 1.8	0.40
Sample 3			18.8(18.4–19.1)	17	18.6(17.9–19.3)	12	-1.2 $\pm$ 5.9	0.50	
Sample 4			15.6(15.1–16.1)	16	15.7(15.1–16.3)	9	0.8 $\pm$ 4.9	0.73	
Sample 5			16.0(15.5–16.4)	17	15.6(14.9–16.2)	9	-2.5 $\pm$ 5.5	0.28	
Sample 3	Stability during analysis at 5 °C	4.8 days	17.9(17.0–18.9)	3	17.5(15.1–19.9)	3	-2.3 $\pm$ 4.3	0.6 $\pm$ 2.5	0.47
Sample 7			15.7(14.5–16.9)	3	16.0(15.3–16.8)	3	2.0 $\pm$ 1.5	0.40	
Sample 8			33.2(25.4–41.1)	3	33.9(28.0–39.8)	3	2.2 $\pm$ 2.7	0.92	
Sample 9	Storage at RT	48 h	12.2(10.6–13.8)	3	12.1(9.39–14.8)	3	-1.0 $\pm$ 8.9	-1.0 $\pm$ 8.9	0.90

**Table 5**

Results of testing equilibrium dialysis cells (A) and membrane MWCO (B) used for ruggedness testing.

A	Cell Type	Macro 1	Macro 1	Macro 1S	Micro 0.2	FMED
Membrane MWCO	5 kDa	10 kDa	5 kDa	5 kDa	5 kDa	
N	16	8	16	14	8	
%CV	–	2.6%	2.1%	5.4%	20.9%	
Mean % bias	–	0.3%	1.2%	5.5%	–18.1%	
p-value ( $\alpha = 0.05$ )	–	1.00	0.97	0.01	<0.0001	

B	3.5 kDa			5 kDa			10 kDa		
	Mean, pg/mL (pmol/L)	% CV	N	Mean, pg/mL (pmol/L)	% CV	N	Mean, pg/mL (pmol/L)	% CV	N
12.3 (15.9)	2.8	6	12.6 (16.3)	3.9	7	12.4 (16.1)	5.1	14	
11.6 (15.0)	2.6	6	12.3 (15.9)	2.9	8	12.1 (15.7)	5.4	17	
14.7 (19.0)	2.1	6	14.5 (18.8)	2.7	9	14.5 (18.8)	3.3	21	
14.6 (18.9)	3.2	8	14.7 (19.0)	4.6	9	14.2 (18.4)	4.2	21	

<sup>a</sup>Multiple dialysis device types (Macro 1S, Micro, and Fast Micro-Equilibrium Dialyzer or FMED cells) were tested to analyze their capability to reach equilibrium within 5 h using sample FT4 concentrations. The samples from these cells were compared to results of the dialysis cells used for the FT4 RMP at Ref4U under identical conditions. Comparison of sample means among dialysis device and membrane type using 1-way repeated measures ANOVA indicated significant differences among the dialysis devices tested ( $p < 0.05$ ). Pairwise testing with adjustment for multiple comparisons was done to determine which sample means of the 4 devices tested were significantly different from the means of the reference device ("Macro 1"). Adjusted p-values indicate significant differences in sample means for the FMED and Micro 0.2 devices.

<sup>b</sup>Regenerated cellulose membranes with MWCO 3.5, 5, and 10 kDa were tested during equilibrium dialysis of 4 serum samples (ranging in T4 concentration from 15.0 to 19.0 pmol/L) using the Macro 1S dialysis cells. Comparison of sample means among membrane MWCO was determined using 1-way repeated measures ANOVA. No significant differences in sample means among all samples tested ( $p = 0.42$ ) and all membranes tested ( $p = 0.12$ ) were observed at the 5% significance level.

produce results consistent with the existing RMP and were found to be unsuitable for use in the RMP under stated conditions. The bias to the RMP for both the Micro and FMED cells were outside of the  $\pm 2.5\%$  limit (5.5% and –18.1%, respectively). The imprecision of the Micro and FMED cells were also  $> 5\%$  in both cases (5.4% and 20.9%, respectively). The increased imprecision and bias to the RMP of the FMED cells may be due to their almost two-fold decrease in membrane surface area and increased chamber depth of these devices which may slow diffusion across the membrane [12]. The higher variability observed when using FMED cells routinely contradicts previous observations of stable FT4 concentration during fourth and fifth hours of dialysis during testing which suggests that equilibrium may not be reached within five hours; additional testing at longer dialysis times is suggested before determining this device's suitability for FT4 measurement.

The imprecision of measurement using 10 kDa membranes with the Macro 1 dialysis cells was 2.6%. The mean bias for Macro 1 cells with 10 kDa membranes compared to those with 5 kDa membranes was 0.3%. Interchangeability of 10 kDa and 5 kDa membranes was confirmed in a separate experiment, where the influence of membrane MWCO on sample FT4 concentration was assessed for serum with mean FT4 concentrations of 12.5 pg/mL (16.1 pmol/L, "Sample A"), 12.2 pg/mL (15.7 pmol/L, "Sample B"), 14.6 pg/mL (18.8 pmol/L, "Sample C"), and 14.5 pg/mL (18.6 pmol/L, "Sample D") using the Macro1S dialysis cells. Results from one-way repeated measures ANOVA are shown in Table 5b. Comparison of Sample A–D data collected using 3.5, 5, or 10 kDa MWCO

membranes indicates no significant difference in sample mean concentration among the samples tested between the 3 different membrane types at the  $\alpha = 0.05$  significance level ( $p = 0.12$ ). This study demonstrates that several membranes can be used interchangeably if conventional steps such as ED temperature and buffer are maintained, making the FT4 RMP independent of a specific manufacturer or membrane type as long as a comparison study for the new type of membrane demonstrates good agreement. It is important to note that the mean percent dialysate volumes recovered for 3.5 kDa, 5 kDa, and 10 kDa membranes were  $85 \pm 6\%$ ,  $77 \pm 8\%$ , and  $59 \pm 12\%$  of the original serum volume respectively; this could be an important consideration for measurement sensitivity or for volume-critical measurements such as density measurement.

The results demonstrated in this study can serve as a useful guide for laboratories developing not only FT4 RMPs, but also routine ED-based methods. They can be of use to invitro diagnostic device (IVD) manufacturers, as some of the results on T4 absorption and stability can be applied to calibrator preparation and storage.

#### 4. Conclusion

The reference measurement system for FT4 follows strict conventions for ED time, pH, and temperature to ensure reproducibility of results across all reference measurement procedures. Deviations from these conventions can alter the endogenous free-bound T4 equilibrium; furthermore, adsorption of T4 to labware used in preparing calibrators and patient samples can negatively influence FT4 measurements and lead to inaccurate results. Based on our testing, untreated borosilicate glass and either ethanol or ethanol with ammonium hydroxide as a pH modifier were selected for use at CDC when preparing calibrators due to higher recovery observed for these neat T4 solutions. However, due to potential differences in availability and lot-to-lot variations of materials, it is recommended that each laboratory individually consider the impact of their specific equipment and materials on their FT4 measurements.

Studying what can be changed during the ED step brings the method closer to the ultimate goal of independence from specific manufacturer devices and membranes, thereby helping to ensure continuity of the reference measurement system should these devices become unavailable. Careful selection of dialysis device, calibrator, sample preparation labware, and solvent is necessary in order to prevent any loss of FT4 during the critical steps of FT4 measurement before this loss can be accounted for by the addition of an internal standard.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgments

The authors gratefully acknowledge support and contributions to the development of the CDC RMP for FT4 from the International Federation for Clinical Chemistry (IFCC) Committee for the Standardization of Thyroid Function Tests (C-STFT); and Dr. Barrett Brister, Chui Tse, Amonae Dabbs-Brown, Krista Poynter, Otoe Sugahara, Clark Coffman, Tatiana Buchannan, Eric Edwards, and Lynn Collins at CDC.

#### 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.

### Funding disclosure

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jmsacl.2023.06.001>.

### References

- [1] R. Mullur, Y. Liu, G.A. Brent, Thyroid hormone regulation of metabolism, *Physiol. Rev.* 94 (2) (2014) 255–282, <https://doi.org/10.1152/physrev.00030.2013>.
- [2] “Thyroid Patient Information,” 2022. [Online]. Available: <https://www.thyroid.org/thyroid-information/>. [Accessed 2022].
- [3] CLSI, *Measurement of Free Thyroid Hormones; Approved Guideline*, Clinical and Laboratory Standards Institute, Wayne, Pennsylvania, 2004.
- [4] J.E. Midgley, Direct and indirect free thyroxine assay methods: theory and practice, *Clin. Chem.* 47 (8) (2001) 1353–1363, <https://doi.org/10.1093/clinchem/47.8.1353>.
- [5] A. Ardabiligazir, S. Afshariyamchou, D. Mir, I. Sachmechi, Effect of high-dose biotin on thyroid function tests: case report and literature review, *Cureus* 10 (6) (2018) e2845.
- [6] J. Choi, S. Gyu Yun, “Comparison of biotin interference in second- and third-generation roche free thyroxine immunoassays, *Ann. Lab. Med.* 40 (3) (2020) 274–276, <https://doi.org/10.3343/alm.2020.40.3.274>.
- [7] J. Favresse, M.-C. Burlacu, D. Maiter, D. Gruson, Interferences with thyroid function immunoassays: clinical implications and detection algorithm, *Endocr. Rev.* 39 (5) (2018) 830–850, <https://doi.org/10.1210/er.2018-00119>.
- [8] R. Sapin, M. d'Herbomez, Free thyroxine measured by equilibrium dialysis and nine immunoassays in sera with various serum thyroxine-binding capacities, *Clin. Chem.* 49 (9) (2003) 1531–1535, <https://doi.org/10.1373/49.9.1531>.
- [9] L.M. Thienpont, G. Beastall, N.D. Christofides, J.D. Faix, T. Ieiri, V. Jarrige, W. G. Miller, R. Miller, J. Nelson, C. Ronin, H.A. Ross, M. Rottmann, J.H. Thijssen, B. Toussaint, Proposal of a candidate international conventional reference measurement procedure for free thyroxine in serum, *Clin. Chem. Lab. Med.* 45 (7) (2007) 934–936, <https://doi.org/10.1515/cclm.2007.155>.
- [10] B. Yue, A.L. Rockwood, T. Sandrock, S.L. La'ulu, M.M. Kushnir, A.W. Meikle, Free thyroid hormones in serum by direct equilibrium dialysis and online solid-phase extraction–liquid chromatography/tandem mass spectrometry, *Clin. Chem.* 54 (4) (2008) 642–651, <https://doi.org/10.1373/clinchem.2007.098293>.
- [11] J.A. Kennedy, G.S. Besses, Comparison of adsorption of iodine-131-thyroxine to glass and plastic containers, *J. Nucl. Med.* 8 (3) (1967) 226–228.
- [12] C.M.B. Ribeiro, Robustness against changes in the design for equilibrium dialysis of the free thyroxine reference measurement procedure, Howest - Hogeschool West-Vlaanderen, 2018.
- [13] E.N. Behrend, R.J. Kemppainen, D.W. Young, Effect of storage conditions on cortisol, total thyroxine, and free thyroxine concentrations in serum and plasma of dogs, *J. Am. Vet. Med. Assoc.* 212 (10) (1998) 1564–1568.
- [14] L. Nye, T.H. Yeo, V. Chan, D. Goldie, J. Landon, Stability of thyroxine and triiodothyronine in biological fluids, *J. Clin. Pathol.* 28 (11) (1975) 915–919, <https://doi.org/10.1136/jcp.28.11.915>.
- [15] H.W. Vesper, K. Van Uytfanghe, A. Hishinuma, V. Raverot, M.M. Patru, U. Danilenko, A.E. van Herwaarden, E. Shimizu, Implementing reference systems for thyroid function tests - A collaborative effort, *Clin. Chim. Acta* 519 (2021) 183–186, <https://doi.org/10.1016/j.cca.2021.04.019>.
- [16] A. Ribera, L.i. Zhang, A. Dabbs-Brown, O. Sugahara, K. Poynter, K. van Uytfanghe, E. Shimizu, A.E. van Herwaarden, J.C. Botelho, U. Danilenko, H.W. Vesper, Development of an equilibrium dialysis ID-UPLC-MS/MS candidate reference measurement procedure for free thyroxine in human serum, *Clin. Biochem.* 116 (2023) 42–51.
- [17] B. Toussaint, H. Schimmel, C.L. Klein, M. Wiergowski, H. Emons, Towards the certification of the purity of calibrant reference materials for thyroid hormones: a chicken and egg dilemma, *J. Chromatogr. A* 1156 (1–2) (2007) 236–248, <https://doi.org/10.1016/j.chroma.2006.11.095>.
- [18] J.C. Nelson, R.T. Tomei, Direct determination of free thyroxin in undiluted serum by equilibrium dialysis/radioimmunoassay, *Clin. Chem.* 34 (9) (1988) 1737–1744, <https://doi.org/10.1093/clinchem/34.9.1733>.
- [19] S.K. Van Houcke, K. Van Uytfanghe, E. Shimizu, W. Tani, M. Umemoto, L. M. Thienpont, 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), *Clin. Chem. Lab. Med.* 49 (8) (2011) 1275–1281, <https://doi.org/10.1515/cclm.2011.639>.
- [20] B.K. Matuszewski, M.L. Costanzer, C.M. Chavez-Eng, Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS, *Anal. Chem.* 75 (13) (2003) 3019–3030, <https://doi.org/10.1021/ac020361s>.
- [21] CLSI, *Evaluation of Stability of In Vitro Diagnostic Reagents; Approved Guideline*, Clinical and Laboratory Standards Institute, Wayne, Pennsylvania, 2009.