Clinical Biochemistry 121-122 (2023) 110676

Contents lists available at ScienceDirect

Clinical Biochemistry

journal homepage: www.elsevier.com/locate/clinbiochem

Large method differences for free thyroid hormone assays in the hyperthyroid range can affect assessment of hyperthyroid status: Comparison of Abbott Alinity to Roche Cobas, Siemens Centaur and equilibrium dialysis LC-MS/MS

Alexander Bauer Westbye [a,b,*,1], Finn Erik Aas [a,b,1], Sandra R. Dahl [a,b], Svetlana N. Zykova [a,b], Oskar Kelp [a,b], Louise K. Dahll [a,b], Per M. Thorsby [a,b,c]

[a] Hormone Laboratory, Dept. Medical Biochemistry, Oslo University Hospital, Aker, PO Box 4959 Nydalen, 0424 Oslo, Norway
[b] Biochemical Endocrinology and Metabolism Research Group, Oslo University Hospital, Aker, PO Box 4959 Nydalen, 0424 Oslo, Norway
[c] Institute of Clinical Medicine, University of Oslo, PO Box 1171 - Blindern, 0318 Oslo, Norway

ARTICLE INFO

Keywords:
Thyroid hormone
FT4
FT3
Immunoassay
Equilibrium dialysis
LC-MS/MS
Antithyroid drugs

ABSTRACT

Background: Free T4 (FT4) determination is one of the most commonly performed biochemical tests in endocrinology. Treatment of thyroid dysfunctions is adjusted based on the severity of symptoms and biochemical test results. For Graves' hyperthyroidism, clinical guidelines recommend using FT4 as a (rough) guide to dose antithyroid drugs, together with other clinical information. It is well known that different platforms and methods give different FT4 results; however, large non-linear method differences at high FT4 concentrations are less well recognized. Current clinical guidelines do not make it clear that method differences in the hyperthyroid range can affect recommendations.

Method: Serum samples from patients with very low (biochemically hypothyroid) to very high (hyperthyroid) concentrations of FT4 and/or free T3 (FT3) were analyzed using Abbott Alinity and compared to concentrations measured using Roche Cobas, Siemens ADVIA Centaur (FT4 only) and an in-house equilibrium dialysis liquid chromatography tandem mass spectrometry (LC-MS/MS) method.

Results: Alinity measured markedly lower FT4 and FT3 concentrations compared to the other methods, particularly at high FT4 concentrations. Regression analysis indicated that Alinity FT4 had a non-linear (curved) relationship to FT4 measured by the other methods. The method differences affected guideline-recommended treatments for hyperthyroidism.

Conclusion: Measured free thyroid hormone concentrations are highly method-dependent, especially at high FT4 concentrations. Clinicians treating hyperthyroid patients should be aware that patients appear much less hyperthyroid from FT4-measurements performed using Alinity compared to Cobas or Centaur. Guideline-recommended antithyroid drug dosages based on FT4 (including multiples of the upper reference range) have to be adjusted to the FT4 method used. FT4 results from different methods should be clearly distinguished (e.g. separate lines) in medical records.

1. Introduction

Thyroid disorders are prevalent worldwide and affect up to 10 % of the middle-aged female population in developed countries [1,2]. Hyperthyroidism is caused by an excess of thyroid hormones (THs) in circulation, usually the result of overproduction in the thyroid gland. The prevalence of hyperthyroidism ranges from 0.2 to 1.3 %, is more common among women than men and the most common form of

Abbreviations: ED, Equilibrium dialysis; LC, Liquid chromatography; MS, Mass spectrometry; MS/MS, tandem MS; ED-LCMS, ED followed by LC-M S/MS; T3, triiodothyronine; T4, thyroxine; TH, thyroid hormone; FT3, Free T3; FT4, Free T4; ATD, antithyroid drugs; RI, Reference interval.

* Corresponding authors.
E-mail address: alwest@ous-hf.no (A.B. Westbye).
[1] Authors contributed equally.

https://doi.org/10.1016/j.clinbiochem.2023.110676
Received 11 August 2023; Received in revised form 13 October 2023; Accepted 14 October 2023
Available online 15 October 2023

hyperthyroidism is the autoimmune disease Graves' hyperthyroidism [3,4].

Diagnosis, treatment and follow-up of thyroid disorders relies heavily on thyroid function tests [1,2]. The primary clinical biochemical parameter is thyroid-stimulating hormone (TSH), and abnormal TSH values are typically followed up by measurement of free (non-protein bound) concentration of the TH thyroxine (Free T4, FT4) and sometimes triiodothyronine (FT3). Measurement of FT4 is therefore one of the most commonly performed biochemical tests in endocrinology.

The treatment goal for hyperthyroidism is to restore free TH concentrations to normal, and for Graves' disease patients this usually involve an antithyroid drug (ATD; e.g. methimazole, carbimazole or propylthiouracil), radioiodine or thyroid surgery. Management requires an evaluation of the patient's symptoms, thyroid gland size and biochemical test results. The primary biochemical parameter to infer the severity of thyrotoxicosis is FT4, often together with FT3 or total T3. Clinical guidelines for management of Graves' disease and UpToDate recommend (as a rough guide) to dose antithyroid medication based on measured FT4 concentration, together with clinical information (e.g. symptom severity). However, it is not clear from the guidelines that the FT4 method used can greatly affect the recommended dosage [5–8].

Most hospitals and clinical laboratories routinely measure FT4 and FT3 in serum using commercial, automated immunoassays. Generally, clinicians compare the results to established normal reference intervals (RIs) to infer if the patient is healthy. Although manufacturer-provided reference ranges for FT4 differ, most methods do not appear to differ dramatically at first sight – for example, the ranges for the lower limits are 9 to 11.9 pM and the upper limits 19 to 22.7 pM for Alinity, Cobas and Centaur FT4 (Table 1). Different clinical FT4 assays are known to not measure the same free TH concentration (i.e. report the same concentration value) for a sample, but have systematic biases. However, these differences are modest in the normal range. While there are several comparisons of modern immunoassays, they tend to be limited to normal, "healthy" populations (i.e. normal range) [9–12]. Little information is available on how methods behave above the normal range and whether there are differences with implications for clinical decisions. We report here how non-linear relationships between different clinical assays used for analyzing FT4 can affect assessment of hyperthyroid status, if method differences are not accounted for.

Table 1
Comparison of FT4 methods.

	Manufacturer RI (pM)[A]			Concentration corresponding to Alinity (pM)[B]	
RI limits	Alinity	Cobas (% increase)	Centaur (% increase)	Cobas (% increase)	Centaur (% increase)
Lower limit	9	11.9 (32 %)	11.5 (28 %)	10 (16 %)	11 (26 %)
Upper limit	19	21.6 (14 %)	22.7 (19 %)	31 (62 %)	27 (43 %)
1.5x upper limit	28.5	32.4 (14 %)	34.0 (19 %)	58 (103 %)	52 (81 %)
2x upper limit	38	43.2 (14 %)	45.4 (19 %)	92 (143 %)	85 (124 %)

Values are rounded.

[A] The manufacturer provided RI limits for FT4 provided with the tests, and the calculated values for 1.5- and 2-times the upper limits obtained my multiplying the upper limit with 1.5 or 2, respectively. In parenthesis, the calculated increase from Alinity to Cobas or Centaur concentration.

[B] Estimated corresponding concentration if reanalyzed using $Cobas_{FT4}$ or $Centaur_{FT4}$. The corresponding concentrations for samples measured with Alinity to have FT4 concentrations of 9, 19, 28.5 or 38 pM. In parenthesis, the percentage increase from Alinity to Cobas or Centaur. Concentrations converted from $Alinity_{FT4}$ using the regression equations from Fig. 1C and Fig. 1F.

2. Materials and methods

2.1. Samples

Samples for comparison of FT4 and FT3 were collected from patient samples received as part of routine diagnostic testing. Samples were collected haphazardly over several weeks from left-over material from clinical routine analysis, and supplemented with additional samples with high or low concentrations to ensure the collection covered a broad measurement range. Most samples were received from general practitioners in Oslo municipality (Norway) and the surrounding regions and from the Thyroid Outpatient Clinic (predominantly treating Graves' disease patients) at Oslo University Hospital. In general, the samples were first analyzed on Alinity, frozen and stored at −20 °C before analysis by the other methods. For the FT3 Alinity to Cobas comparison, samples were first analyzed on Cobas.

2.2. Immunological analyses

The immunological assays were performed on instruments in use for clinical routine analysis at Oslo University Hospital (OUS; Alinity and Cobas) or Fürst Medical Laboratory (Centaur) using accredited methods (OUS: NS-EN ISO/IEC 17025; Fürst: NS-EN ISO 15189). The assays were verified to perform satisfactorily by evaluation of precision, carry-over, and participation in an external quality control program (Labquality).

The immunological analyses were performed using the following instruments and reagent kits: Alinity i (Abbott Diagnostics; reagent kits: Free T4 and Free T3), Cobas (Roche Diagnostics; Elecsys FT4 III and Elecsys FT3 III) and ADVIA Centaur (Siemens Healthineers; ADVIA Centaur FT4).

The normal RI provided by Abbott for Alinity FT4 were 9.0 to 19 pM (pM = pmol/L; central 99 % interval) and Alinity FT3 2.4 to 6.0 pM (central 95 %). The central 95 % Alinity FT4 RI constructed by Abbott (10.3—17.8 pM) was presented on a scientific poster at EuroMedLab [13]. The manufacturer-provided RI for Cobas FT4 were 11.9 to 21.6 pM (central 95 %), Centaur FT4 11.5 to 22.7 pM (details of interval coverage were not provided by Siemens), and Cobas FT3 3.1 to 6.8 pM (central 95 %).

2.3. ED-LCMS analyses

Equilibrium dialysis liquid chromatography – tandem mass spectrometry (ED-LCMS) analysis was performed using a 96-well format method developed in-house, described in detail in the accompanying Supplementary Material. In short, samples were dialyzed using the 10 kDa molecular weight cutoff DispoEquilibrium Dialyzer against a HEPES buffer at 37 °C, essentially as previously reported [14]. Dialysate was purified using solid phase extraction, injected on a C18-column and separated by a H_2O:MeOH gradient using a Sciex ExionLC. Compounds were quantitated using scheduled multiple reaction monitoring using a Sciex QTRAP 6500+.

2.4. Software and data analysis

Data handling was performed in Microsoft Excel 2016 and R (v 4.2.2) [15] using RStudio (v 2022.07.02) [16]. Statistical analysis was performed in R/Rstudio: Bland-Altman plots [17] were prepared using "Tidyverse", including ggplot2 [18]. Passing-Bablok regression and visualization was performed using package *mcr* with setting method.reg = "PaBa" [19]. Polynomial regression was performed using *stats::lm* (weighted 1/x^2).

2.5. Ethical statement

This study used left-over material received as part of routine diagnostic testing, without access to or recording of personal information

and did therefore not require approval by the Norwegian Regional Committees for Medical and Health Research Ethics (ref no. 586996).

3. Results

3.1. Abbott Alinity measured markedly lower FT4 concentrations than Roche Cobas and Siemens Centaur in the hyperthyroid range

Patient samples (n = 147), chosen to cover a large concentration range, were analyzed for FT4 concentrations using the two clinical analyzers Abbott Alinity ($Alinity_{FT4}$) and Roche Cobas ($Cobas_{FT4}$). Of the analyzed samples, 110 (75 %) were above the manufacturer provided RI (central 95 %) for Cobas, and 92 (63 %) or 99 (67 %) were above the manufacturer provided central 99 % RI or reported central 95 % RI [13], respectively, for Alinity. Eighteen samples were above the quantitative range of Cobas (upper limit of quantitation 100 pM) and were excluded from subsequent analysis. The measured concentrations of the remaining 129 samples were compared by Bland-Altman plots (Fig. 1A and Fig. S1A) and Passing-Bablok regression (Fig. 1B). The Bland-Altman plots clearly indicated that the assays did not produce similar concentration values: for most of the samples, Alinity measured substantially lower concentrations than Cobas and both absolute and percentage difference increased with increasing FT4 concentration (Fig. 1A and B). For samples at the upper end of the concentration range investigated (>25 pM $Alinity_{FT4}$), 29 of 30 samples (97 %) had a percentage difference in measured FT4 concentration greater than 50 %. For eight of the 30 samples (27 %) the difference was greater than 80 %. In addition, all the 18 excluded samples were measured to markedly higher concentrations on Cobas ($Cobas_{FT4}$ > 100 pM; quantitation limit) compared to their Alinity concentrations (range 34–62 pM; mean 44 pM).

The $Alinity_{FT4}$ and $Cobas_{FT4}$ correlated well (Fig. 1B, r = 0.946, Pearson), however closer examination of the Passing-Bablok regression plot (Fig. 1B) and the residuals (Fig. S1B) indicated a non-linear, curved trend. A second-order (quadratic) polynomial regression curve better described the relationship between $Alinity_{FT4}$ and $Cobas_{FT4}$ values (Fig. 1C) and produced more evenly spread residuals (Fig. S1C). Using

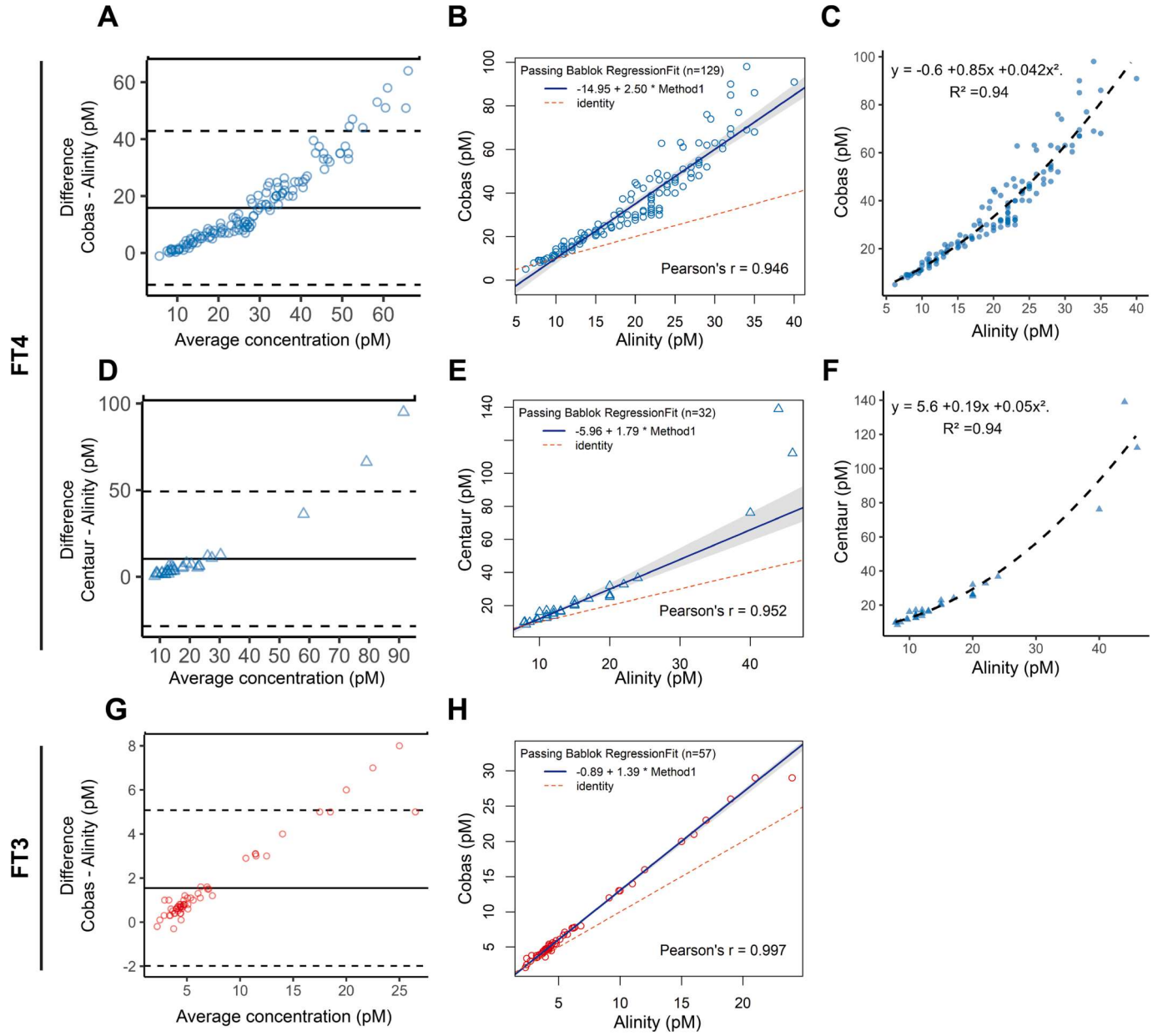

Fig. 1. Abbott Alinity FT4 and FT3 compared to Roche Cobas or Siemens Centaur. Comparison of measured FT4 concentrations (A to F) and FT3 (G to H) using Bland-Altman plots (A, D and G), Passing-Bablok regression (B, E and H) and second-order (quadratic) polynomial regression (C and F).

this equation, it can be seen that for samples at the lower reference limit (Alinity central 99 % RI), measured FT4 concentrations are comparable between the Cobas and Alinity (Table 1; 10 vs 9 pM, 16 % increased). However, for samples at the upper RI limit, Cobas measured a markedly higher concentration (31 vs 19 pM, 62 % increased) and for hyperthyroid samples at 1.5x- or 2-times the upper Alinity limit the difference is even greater (103 % and 143 % increased concentrations, respectively).

To investigate how Alinity compared to another commonly used clinical analyzer, an additional, smaller sample set (n = 32) was analyzed on Alinity and Siemens Centaur ($Centaur_{FT4}$). $Alinity_{FT4}$ and $Centaur_{FT4}$ exhibited a similar trend as observed in the comparison to $Cobas_{FT4}$: The two methods produced increasingly divergent results with increasing FT4 concentrations while at the low concentrations the methods produced much more similar results (Fig. 1D and S1D). Three of the samples had $Alinity_{FT4}$ concentrations above 25 pM, and all three (100 %) had a percentage difference in FT4 concentration above 50 %. Of those, two samples had a difference above 80 %.

The $Alinity_{FT4}$ and $Centaur_{FT4}$ results correlated well (Fig. 1E, r = 0.952), however again the inspection of the Passing-Bablok regression (Fig. 1E) and residuals (Fig. S1E) indicated a non-linear, curved relationship that appeared to be better described by a quadratic curve (Fig. 1F and Fig. S1F). Similar to Cobas, Centaur produced FT4 concentration values that were markedly higher than Alinity in the hyperthyroid range (Table 1).

As we describe in the Discussion (section 4.2), these method differences can have clinical implications.

3.2. Alinity measured lower FT3 concentrations than Cobas

FT3 concentrations are typically increased during thyrotoxicosis. We compared how FT3 concentrations measured by Alinity ($Alinity_{FT3}$) compared to Cobas ($Cobas_{FT3}$) by analyzing 60 samples covering a wide concentration range. Three samples were above the quantitative range of Alinity (upper limit of quantitation 30.72 pM; their $Cobas_{FT3}$ were 37, 38 and 40 pM) and were excluded from subsequent analysis. The difference between the two assays increased with increasing concentration (Fig. 1G, 1H and S1G), and for samples at the upper end of the concentration range ($Alinity_{FT3}$ > 15 pM), four of the five samples (80 %) had a percentage difference above 25 %. Passing-Bablok regression indicated the two methods were linearly related over the investigated range (Fig. 1H and Fig. S1H), and the results correlated well (r = 0.997, Pearson).

3.3. ED-LCMS-measured FT4 is non-linearly related to Alinity

ED-LCMS is by many considered the "gold standard" for free TH measurement [20], but unfortunately, the method is not widely available. We recently developed a 96-well format ED-LCMS method using offline sample preparation inspired by Yu et al. (2008) [14]. The method details and the validation for research purposes are described in detail in the Supplementary Information. In short, the method had a linear response from 2 to 202 pM for T3 and T4 in dialysis buffer and intra- and interassay imprecisions below 10 % and 15 %, respectively, for both FT3 and FT4 using pooled serum (Fig. S2C, D and Table S5). Furthermore, the reference ranges reported by [14] appeared to be transferable, indicating that the two methods measure comparable values, at least in the normal range (Fig. S2G).

The ED-LCMS method measured substantially higher FT4 (Fig. 2A and S3A) and FT3 (Fig. 2D and S3D) concentrations than the Alinity platform for most of the investigated range, consistent with other studies comparing automated immunoassays to ED-LCMS [12,21], including Alinity [10]. The absolute and percentage differences increased with increasing free hormone concentration. For samples at the upper end of the concentration range investigated (>15 pM $Alinity_{FT3}$ and > 25 pM $Alinity_{FT4}$), 12 of 14 (86 %) FT3 measurements had a percentage difference of more than 50 %, and 11 of 11 (100 %) FT4 measurements had a percentage difference of more than 100 %.

Passing-Bablok regression between Alinity and ED-LCMS (Fig. 2B

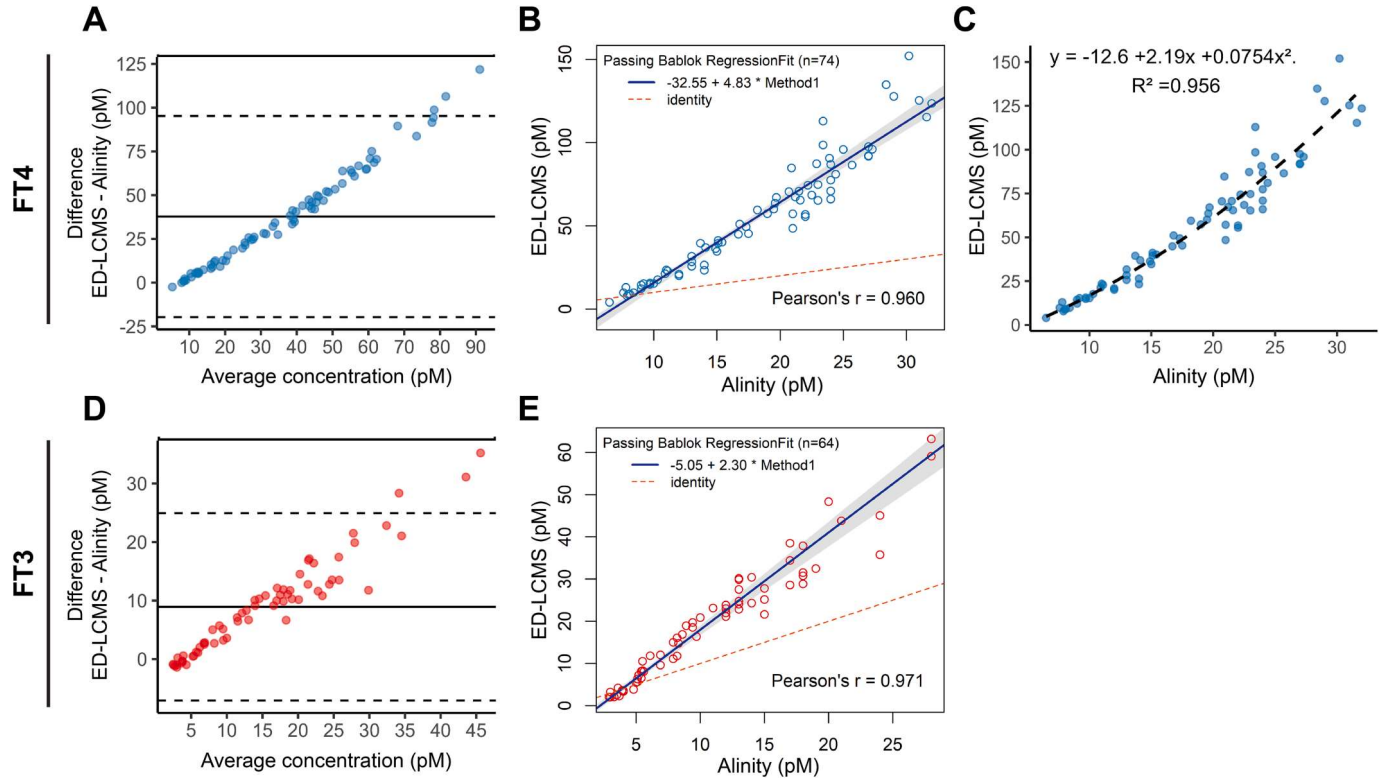

Fig. 2. Abbott Alinity FT4 and FT3 compared to equilibrium dialysis LC-MS/MS. Comparison of measured FT4 concentrations (A to C) and FT3 (D to E) using Bland-Altman plots (A and D), Passing-Bablok regression (B and E) and second-order (quadratic) polynomial regression (C).

and E) indicated that the values correlated (r = 0.960 and r = 0.971 for FT4 and FT3, resp. Pearson). Similar to the FT4 comparisons between Alinity and the other immunoassay methods, a non-linear trend was apparent between Alinity and ED-LCMS for FT4 (Fig. 2B and S3B). A curved, quadratic function appeared to better describe the relationship between the two methods (Fig. 2C and Fig S3C; see Fig S3F for Alinity as a function of ED-LCMS). In contrast, FT3 concentrations appeared to be linearly related over the investigated range (Fig. 2E and S3E).

4. Discussion

Several different vendors offer automated clinical platforms to measure FT4, however due to a lack of standardization or harmonization, the measured concentration value of the same sample can differ between different platforms. Method differences have been reported in the past, however these studies have mostly focused on healthy populations (normal range) [9–12]. In contrast, method differences above the normal range and implications for the treatment of hyperthyroid patients is little studied. This research project was a result of the needs of clinical endocrinologists at Oslo University Hospital (OUS) treating hyperthyroid patients, including Graves' disease patients, after a change of the FT4 method (from Delfia, discontinued by Perkin Elmer) to Alinity at the OUS Hormone Laboratory. The discrepancy between FT4 concentrations above the reference range turned out to be much greater than expected between Alinity and other immunoassay methods commonly in use in Norway, based on the information provided by the manufacturer and available studies.

4.1. Clinical FT4 assays differ much more in the hyperthyroid range than expected from reference interval comparisons

Our investigation initially focused on how Alinity concentration values compared to Cobas, the other platform in use at OUS, and was subsequently expanded to cover Centaur (together encompassing the most commonly used platforms in Norway). These two methods have manufacturer provided FT4 RI limits above Alinity, but all three methods have values that are within 3 pM (max. 32 % higher) at the lower limit and 4 pM (max. 19 % higher) at the upper limit Table 1– at first sight they can therefore appear to measure very similar concentrations. In contrast, we found large differences in measured FT4 concentrations in the hyperthyroid range between Alinity and Cobas/Centaur. The magnitude of the difference was not evident by comparing the RI of the tests, and we could not find publically available literature reporting these differences and their clinical implications for management of hyperthyroidism.

In and above the normal range, Alinity measured a lower FT4 concentration compared to the two other immunoassay (and the ED-LCMS) methods investigated. The difference increased substantially, in a non-linear fashion, with increasing FT4 concentration (Table 1, Fig. 1C and F). While the method differences in and below the RI are substantial from a laboratory standpoint, they are less substantial from a clinical viewpoint. In contrast, clinicians should be aware of the behavior of the different methods at high FT4 concentrations, as the large concentration differences can have implications for clinical decisions, as described below (section 4.2).

It is not clear what makes the different assays perform as they do. However, there are multiple parameters that are important for the performance of immunoassays, including the specific antibodies and their concentrations, buffer composition, additives, solid phase, incubation time and temperature. Designing immunoassays for free hormones is in general difficult as the equilibrium between the free and bound form are easily disturbed during analysis (e.g. by small changes in temperature or pH). FT4 assays are probably particularly challenging due to the high ratio of bound to free thyroxine. However, the particular reasons why different methods differ in their measured FT4 concentration are beyond the scope of this investigation.

4.2. Clinical implications of method differences for the management of overt hyperthyroidism

Clinical guidelines for management of hyperthyroidism recommend that the initial dosage of ATD used to lower TH production should be targeted to the severity of hyperthyroidism. This is to ensure sufficient reduction in TH synthesis while minimizing side effects [5,8]. The subsequent titration of drug dosage also uses FT4 concentrations to monitor the normalization of free TH concentrations (TSH typically responds too slowly to be of clinical use). The current American guideline for hyperthyroidism management recommends, as a rough guide, dosage by comparison of measured FT4 to the upper limit of the method's RI:

> "initial MMI [methimazole] daily dosing: 5–10 mg if free T4 is 1–1.5 times the upper limit of normal; 10–20 mg for free T4 1.5–2 times the upper limit of normal; and 30–40 mg for free T4 2–3 times the upper limit of normal. These rough guidelines should be tailored to the individual patient, incorporating additional information on symptoms, gland size, and total T3 levels where relevant. Serum T3 levels are important to monitor initially because some patients normalize their free T4 levels with MMI but have persistently elevated serum T3, indicating continuing thyrotoxicosis." - [8]

Dosage based on comparison to the upper reference limit is also recommended by UpToDate [6,7], while the current Norwegian guideline recommends ATD dosage based on (absolute) FT4 concentration [22].

What is not clear from these recommendations is that the guideline-recommended drug dosage for a patient can greatly depend on the FT4 method used, due to the different response at high FT4 concentrations (Fig. 1C, Fig. 1F and Fig. 2C). Importantly, comparison of a hyperthyroid patient's FT4 concentration to the upper reference limit of the method does not "harmonize" the different methods: the concentration two times the upper reference limit of the Alinity, Cobas and Centaur RI are quite similar, and differ by less than 20 % relative to Alinity (Table 1). However, if a sample that is measured to be 2 times the upper reference limit using Alinity ($Alinity_{FT4}$ 38 pM) is reanalyzed by Cobas and Centaur, those concentrations ($Cobas_{FT4}$ 92 pM; $Centaur_{FT4}$ 85 pM; from our regression equations) would not be 2 times, but 4.3 (Cobas) and 3.7 times (Centaur) the upper reference limits for the respective methods (Table 2). Similar results are also obtained if the upper limit (17.76 pM) of the narrower 95 % RI reported by Abbot [13] is used instead (Table 2).

Therefore, the apparent severity of thyrotoxicosis inferred from FT4 measurements is highly dependent on the FT4 method used. Clinicians therefore need to adapt the rough clinical guides for dosage of ATD to the specific FT4 method used, and to interpret this together with symptom severity and T3 concentrations when making clinical decisions for Graves' disease patients.

4.3. Medical records should clearly distinguish FT4 methods

Long-term follow-up of patients with thyroid disease can involve a change of analytical method used to monitor the patient, for example when a patient is transferred from the hospital (inpatient) to ambulatory care or a general practitioner. Electronic medical records (patient journals) should clearly differentiate values for FT4 that were obtained using different methods and labs, e.g. by displaying the results in different rows/lines clearly labeled with the method and/or lab. Otherwise, FT4 results for hyperthyroid patients obtained by different methods are prone to result in clinical confusion and, worst case, in treatment errors. In our experience, a lack of such differentiation after a change of method resulted in clinical endocrinologists that were puzzled by the unexplained (apparent) changes in the FT4 concentration of their hyperthyroid patients. Upon awareness of the issue, immediate changes to the reporting format were implemented. Our laboratory also

Table 2
Influence of FT4-method on apparent thyrotoxicosis severity: example using a sample measured at twice the upper RI limit of Alinity.

	Alinity 99 % RI[A]		Alinity 95 % RI[A]	
Method	Concentration, pM[B]	Thyrotoxicosis, magnitude[C]	Concentration, pM[B]	Thyrotoxicosis, magnitude[C]
$Alinity_{FT4}$	38	2.0	36	2.0
$Cobas_{FT4}$	92	4.3	83	3.8
$Centaur_{FT4}$	85	3.7	75	3.3

Values are rounded.
[A] Type of Alinity FT4 reference interval used for comparison: kit provided (central 99%) or the central 95% RI reported by Abbot [13].
[B] Method-specific FT4-concentrations. The corresponding (calculated) concentrations for a sample measured to 38 pM $Alinity_{FT4}$ (2-times the upper limit) if reanalyzed by $Cobas_{FT4}$ or $Centaur_{FT4}$.
[C] The magnitude of thyrotoxicosis (times-changed upper RI-limit) using the three different methods. Calculated from the method-specific concentration (Table 2) divided by the upper RI limit of the respective method (Table 1).

contacted requesting physicians to inform them about method differences at high FT4 concentrations.

In summary, the severity of thyrotoxicosis inferred by comparing a patient's FT4 concentration, either the absolute value or compared to the upper reference limit, is highly method dependent. Clinicians need to be aware of this when treating hyperthyroid patients. However, we emphasize that we do not claim that any immunoassay method is superior or inferior to monitor thyroid disease.

4.4. Most FT4 immunoassays measure lower FT4 concentrations compared to ED-LCMS

Most (if not all) FT4 immunoassays measure a lower FT4 concentration ("under-recover") in healthy individuals compared to ED-LCMS, considered the "gold standard" for FT4 measurement [10,12]. For several immunoassays, the difference to ED-LCMS is exacerbated at higher FT4 concentrations, while some immunoassays were reported to measure a higher FT4 concentration at very low concentrations [21]. $Alinity_{FT4}$ was shown to measure approximately 66 % (and $Cobas_{FT4}$ 88 %) of the $ED\text{-}LCMS_{FT4}$ concentration in one study of a healthy group [10]. We observed the slightly lower value of 47 % (at $Alinity_{FT4}$ 12.5 pM). The difference between Alinity and ED-LCMS increased at higher FT4 concentrations, similar to the overall trend reported for multiple FT4 methods by De Grande et al. [21]. For our samples near the highest concentration included in the comparison, the $Alinity_{FT4}$ was only approximately 20 % of the $ED\text{-}LCMS_{FT4}$ (30 pM $Alinity_{FT4}$ equated to 120 pM $ED\text{-}LCMS_{FT4}$; Fig. 2C). We note that we unfortunately were not able to confirm the accuracy (bias) of our ED-LCMS method due to a lack of suitable reference material. However, our method appeared to produce values in general agreement with the method reported by Yue et al. (2008) [14] (Supplemental Results and Discussion).

4.5. Limitations to our study

One limitation to our study is that most samples were first analyzed on Alinity on arrival, frozen and then subsequently analyzed by the other assay (except for the FT3 Alinity to Centaur comparison). This could systematically influence the measurements; however, we do not believe this to be a major issue for the following reasons: free thyroid hormone concentrations were reported to be stable for at least three freeze–thaw cycles [23–25]. The documentation provided for Alinity (07P70, revised Feb 2018), Cobas (09043276500, 2023–07, V 2.0) and Centaur FT4 kits (10629962_EN Rev. 11, 2023–03) specify that samples can be subjected to one freeze–thaw cycle. Our own freeze–thaw experiments using $Alinity_{FT4}$ also indicate that FT4 is relatively robust to freeze–thaw cycles (Fig S4A). Furthermore, sub-group analysis of the samples that were first analyzed on Cobas indicate a similar, curved method difference between Alinity and Cobas (Fig S4B).

4.6. Conclusion

FT4 immunoassays were not linearly related and measured very different concentrations for severely hyperthyroid patients. Clinicians treating hyperthyroid patients (e.g. Graves' disease) should be aware that patients appear much less hyperthyroid from FT4 measurements performed using Alinity compared to Cobas or Centaur. Some guidelines recommended to use FT4 as a (rough) guide to prescribe the dosage of ATD, however the FT4 method used can affect the recommended dose (even if multiples of upper reference range are used). They therefore have to be adjusted to the specific FT4 method. FT4 results from different methods should be clearly distinguished (e.g. separate lines) in medical records to avoid confusion or errors.

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.

Acknowledgements

The authors thank Guro S. Clementz and Tove L. Gulbrandsen for assistance with sample handling and the laboratory information management system, and Liv H Bakke, Tone Smetop and Yilu Ye for assistance on and maintenance of the LC-MS/MS system. We thank Fürst Medical Laboratory for performing the Centaur FT4 measurements.

Funding

This study was funded by the Hormone Laboratory, and 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.clinbiochem.2023.110676.

References

[1] D. Salvatore, T. Davies, M. Schlumberger, I. Hay, P. Larsen, Thyroid Physiology and Diagnostic Evaluation of Patients With Thyroid Disorders, in: S. Melmed, K. Polonsky, P. Larsen, H. Kronenberg (Eds.), Williams Textbook of Endocrinology, Elsevier, Philadelphia, 2016.
[2] P.M. Sluss, F.J. Hayes, Laboratory Techniques for Recognition of Endocrine Disorders, in: S. Melmed, K.S. Polonsky, P.R. Larsen, H.M. Kronenberg (Eds.), Williams Textbook of Endocrinology, Elsevier, Philadelphia, PA, USA, 2016, pp. 77–107.
[3] P.N. Taylor, D. Albrecht, A. Scholz, G. Gutierrez-Buey, J.H. Lazarus, C.M. Dayan, O. E. Okosieme, Global epidemiology of hyperthyroidism and hypothyroidism, Nat. Rev. Endocrinol. 14 (5) (2018) 301–316.

[4] T.F. Davies, S. Andersen, R. Latif, Y. Nagayama, G. Barbesino, M. Brito, A. K. Eckstein, A. Stagnaro-Green, G.J. Kahaly, Graves' disease, Nat. Rev. Dis. Primers 6 (1) (2020) 52.
[5] G.J. Kahaly, L. Bartalena, L. Hegedus, L. Leenhardt, K. Poppe, S.H. Pearce, 2018 european thyroid association guideline for the management of graves' hyperthyroidism, Eur Thyroid J 7 (4) (2018) 167–186.
[6] D.S. Ross, Accessed 2023, Graves' Hyperthyroidism in Nonpregnant Adults: Overview of Treatment 12 (2022) 09, https://www.uptodate.com/contents/graves-hyperthyroidism-in-nonpregnant-adults-overview-of-treatment.
[7] D.S. Ross, Accessed 2023, Thionamides in the Treatment of Graves' Disease 15 (2022) 09, https://www.uptodate.com/contents/thionamides-in-the-treatment-of-graves-disease.
[8] D.S. Ross, H.B. Burch, D.S. Cooper, M.C. Greenlee, P. Laurberg, A.L. Maia, S. A. Rivkees, M. Samuels, J.A. Sosa, M.N. Stan, M.A. Walter, 2016 american thyroid association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis, Thyroid 26 (10) (2016) 1343–1421.
[9] J.H. Barth, A. Luvai, N. Jassam, W. Mbagaya, E.S. Kilpatrick, D. Narayanan, S. Spoors, Comparison of method-related reference intervals for thyroid hormones: studies from a prospective reference population and a literature review, Ann. Clin. Biochem. 55 (1) (2018) 107–112.
[10] H.I. Jansen, A.E. van Herwaarden, H.J. Huijgen, R.C. Painter, J.J. Hillebrand, A. Boelen, A.C. Heijboer, Pregnancy disrupts the accuracy of automated fT4 immunoassays, Eur Thyroid J 11 (6) (2022).
[11] J. Pelanti, T. Lamberg, T. Salopuro, C. Pussinen, J. Suvisaari, L. Joutsi-Korhonen, C. Schalin-Jantti, O. Itkonen, M. Anttonen, Changing immunochemistry platforms: thyroid function test comparison and reference intervals based on clinical needs, J. Appl. Lab. Med. 7 (6) (2022) 1438–1444.
[12] L.M. Thienpont, K. Van Uytfanghe, G. Beastall, J.D. Faix, T. Ieiri, W.G. Miller, J. C. Nelson, C. Ronin, H.A. Ross, J.H. Thijssen, B. Toussaint, I.W.G.o.S.o.T.F.Tests, report of the IFCC Working group for standardization of thyroid function tests; part 2: free thyroxine and free triiodothyronine, Clin. Chem. 56 (6) (2010) 912–920.
[13] A. Meyer, R. Müller, T. Brühl, M. Hoffmann, J. Bartz, G. Hoffmann, I. Stavchanska, G. Danzl, S. Zenk, Estimation of Reference Intervals for TSH, fT3, and fT4 on Alinity for Two Regionally Separated Laboratories in Germany: Comparison of Three Different Indirect Methods (poster), EuroMedLab, Munich, Germany, 2022.
[14] 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.
[15] R Core Team R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing 2023 Vienna, Austria.
[16] RStudio Team RStudio: Integrated Development for R, RStudio 2020 Boston, MA, USA.
[17] J.M. Bland, D.G. Altman, Statistical methods for assessing agreement between two methods of clinical measurement, Lancet 1 (8476) (1986) 307–310.
[18] H. Wickham, M. Averick, J. Bryan, W. Chang, L. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. Pedersen, E. Miller, S. Bache, K. Müller, J. Ooms, D. Robinson, D. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, H. Yutani, Welcome to the Tidyverse, Journal of Open Source Software 4 (43) (2019) 1686.
[19] S. Potapov, F. Model, A. Schuetzenmeister, E. Manuilova, F. Dufey, J. Raymaekers, mcr: Method Comparison Regression, 2023.
[20] 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.
[21] L.A.C. De Grande K. Van Uytfanghe D. Reynders B. Das J.D. Faix F. MacKenzie B. Decallonne A. Hishinuma B. Lapauw P. Taelman P. Van Crombrugge A. Van den Bruel B. Velkeniers P. Williams L.M. Thienpont Standardization of Free Thyroxine Measurements Allows the Adoption of a More Uniform Reference Interval 63 10 2017 1642 1652.
[22] L.K. Dahll, B.G. Nedrebø, I. Norheim, S. Hammerstad, R. Jorde, Graves' hypertyreose [Graves' hyperthyrodisms], 2022. https://metodebok.no/endokrinologi. (Accessed 15.09.2023.
[23] J.J. Hillebrand A.C. Heijboer E. Endert Effects of repeated freeze-thaw cycles on endocrine parameters in plasma and serum Ann Clin Biochem 54 2 2017 289 292.
[24] T. Mannisto, E. Suvanto, H.M. Surcel, A. Ruokonen, Thyroid hormones are stable even during prolonged frozen storage, Clin. Chem. Lab. Med. 48 (11) (2010) 1669–1670, author reply 1671–2.
[25] A. Ribera, L. Zhang, C. Ribeiro, N. Vazquez, J. Thonkulpitak, J.C. Botelho, U. Danilenko, K. van Uytfanghe, H.W. Vesper, Practical considerations for accurate determination of free thyroxine by equilibrium dialysis, J Mass Spectrom Adv Clin Lab 29 (2023) 9–15.