Clin. Lab. 2012;58:645-657

ORIGINAL ARTICLE

Performance Evaluation of the Access® FT3 and FT4 Assays, Comparison With Immulite® and AxSYM®, and the Relationship to TSH Values

WOLFGANG KORTE [1,2], HANNA ENGLER [1], WALTER F. RIESEN [1], THOMAS BRINKMANN [3]

[1] Institut für Klinische Chemie und Hämatologie, Kantonsspital St. Gallen, St. Gallen, Switzerland
[2] Medizinische Fakultät, Universität Bern, Bern, Switzerland
[3] Unilabs, Geneva, Switzerland

SUMMARY

Background: **Different FT3 and FT4 assays report significantly different results. We compared the distribution of FT3 and FT4 in a cohort of Swiss patients measured with DxI 800, AxSYM, and Immulite 2000.**
Methods: **TSH, FT3, and FT4 values were measured in 1,938 serum samples. Patients were classified on the basis of their TSH values as low, normal, and high. For each class of TSH values, concordances of FT3 and FT4 results were determined among the three assays.**
Results: **For low TSH values in all three assays FT3 (FT4) concordance of DxI - AxSYM, DxI - Immulite, and AxSYM - Immulite was determined as 83.1%, 76.2% 68.5% (60.8%, 74.6%, 83.1%), for normal TSH as 89.2%, 79.0%, 75.3% (83.9%, 85.5%, 83.1%) and for elevated TSH as 78.0%, 86.0%, 78.0% (84.0%, 90.0%, 90.0%), respectively.**
Low FT4 concordance rates with DxI 800 were mainly caused by its FT4 upper reference limit of 14.1 pmol/L. Using a cut-off of 16.1 pmol/L concordances with AxSYM and Immulite were improved to 77.7% and 86.9% (low TSH), 92.5% and 96.2% (normal TSH), and 90.0% and 92.2% (high TSH). Low FT3 concordance rates with Immulite were caused by its low FT3 upper reference limit of 6.29 pmol/L as 11.6% of patient samples with normal TSH value showed unusually elevated FT3 results.
Conclusions: **We showed an overall good concordance of FT3 and FT4 results, when stratified according to corresponding TSH values and the appropriate reference range is used. However, our data also show that problems of interpretation of results based on numerical values have yet not been solved.**
(Clin. Lab. 2012;58:645-657. DOI: 10.7754/Clin.Lab.2011.110705)

KEY WORDS

Method evaluation; immunoassays, thyroid function tests, TSH, FT3, FT4

LIST OF ABBREVIATIONS

TSH - thyrotropin (thyroid stimulating hormone)
FT3 - free triiodothyronine
FT4 - free thyroxine
TPO - thyroid peroxidase
IQR - inter-quartile ratio

INTRODUCTION

Sensitive and precise methods for TSH are the recommended initial screening tool for suspected thyroid diseases. For identification and confirmation of thyroid disease as well as initiation and follow-up of therapy a combination of thyrotropin (TSH) and free triiodothyronine (FT3)/free thyroxine (FT4) is often used. However, lack of standardization of assays for free peripheral hormones remains a major draw-back for the interpretation of results [1-5].
The objective of the present study is to evaluate in a

Manuscript accepted November 27, 2011

large cohort of hospitalized and ambulatory patients, stratified upon the basis of TSH, if FT3 and FT4 results, and to provide concordant results expressed in relation to the normal values stated by different manufacturers.

MATERIALS AND METHODS

Instruments and Reagents

Patient samples were measured on the UniCel DxI 800 (Beckman Coulter Inc., Nyon, Switzerland) for routine thyroid analysis. The same day samples were measured on AxSYM (Abbott, Wiesbaden, Germany), and Immulite 2000 (Siemens, Munich, Germany).

Specimens

Non-identifiable left-over specimens with normal and pathological concentrations of TSH, FT3, and FT4 were collected from June to August 2006 at the Canton Hospital St. Gallen, Switzerland, from inpatients and outpatients (20.9% endocrinology, 15.2% general medicine, 13.7% outpatients, 12.2% surgery, 8.2% cardiology, 7.1% psychiatry, 4.6% oncology, 3.6% emergency department, 1.5% pediatrics, 1.2% geriatrics). All samples were assayed on the same day on all three analyzers. All samples were evaluated for anti-TPO antibodies on the DxI 800. Samples with positive anti-TPO were excluded.

Quality Control: Evaluation of analytical methods

Intra-assay and inter-assay imprecision were determined according to CLSI EP-5 guidelines [6] and were found as stated by the manufacturers' data sheet insert (data not shown).

Method comparison

For method comparison DxI 800 - AxSYM, DxI 800 - Immulite 2000, and AxSYM - Immulite 2000 linear regression was applied using the Passing-Bablok procedure [7].

Graphical classification of thyroid conditions

Graphical classification of thyroid conditions for each patient was performed on the basis of TSH/FT3 and TSH/FT4 scatter plots for DxI 800, AxSYM, and Immulite 2000 [2], whereby TSH-axis is log-scaled for better readability of the figures.

Concordance of different thyroid assay results

Concordance between different FT4 assays was calculated from contingency tables constructed pair-wise on combinations of TSH and FT4 obtained from three different instruments. Reflecting clinical decision, values were classified as normal, high or low with respect to the normal ranges of each directional insert. If both FT4 values were in the same expected functional class in relation to TSH results, the results of these two FT4 assays were considered as concordant (i.e. normal TSH and normal FT4, high TSH and low FT4, low TSH and high FT4). Results departing from this classification were tabulated as non-concordant (i.e. FT4 high in one assay but low in the other assay with normal TSH in both assays). Concordance scores were derived from these tables. The same methodology was followed for FT3 assay comparisons.

In addition, for all concordance results corresponding kappa values were calculated. Kappa relates concordant test results to concordant test results expected just by chance. It characterizes of inter-rater agreement and refers to reliability. Values can be interpreted as followed: poor <0.20, fair 0.21 - 0.40, moderate 0.41 - 0.6, good 0.61 - 0.80, very good >0.8 [8].

Ethics

Analytical method comparison was performed with anonymized, left-over samples. Swiss law does not require informed consent in this case.

Statistics

Comparisons between TSH, FT3, and FT4 from three different manufacturers were performed with Analyse-it® using non-parametric Passing-Bablok regression. Analysis of method concordances and kappa statistics were performed with JMP® (SAS Institute) software.

RESULTS

Method comparison of TSH

The large patient cohort with more than 1,500 serum samples measured with all three assays was highly representative of the wide spectrum of thyroid diseases. Only 2.79% of the patients had non-detectable TSH levels (<0.01 mU/L) and 5.43% exhibited overtly elevated concentrations (>10.0 mU/L). This wide disparity of thyroid function in our cohort was also seen in the wide ranges of FT3 and FT4 (Figure 1). Method dependant reference values for all three TSH assays are shown in Table 1.

Patient TSH values from all three systems are reasonable comparable. The lower TSH reference value for DxI 800, AxSYM and Immulite 2000 corresponds to the 11.8th, 11.0th and 15.4th percentile of measured results, and the upper reference limit to the 93.7th, 91.0th and 92.7th percentile. The overall concordance (kappa value) of TSH results between DxI 800 - AxSYM, DxI 800 - AxSYM and AxSYM - Immulite 2000 were 96.4% (0.9273), 95.0% (0.8979), and 94.1% (0.8834). As the kappa values are >0.8, concordance of results for all three system combinations can be considered to be very good [8].

Method comparison of peripheral free hormones

FT3

The Immulite FT3 upper reference limit of 6.00 pmol/L represented by the upper red line in Figure 1, nearly cuts the 75th percentile of the box (50% of all results are within the box).

Table 1. Reference ranges for TSH, FT3, and FT4 on DxI 800 (Beckman Coulter), AxSYM (Abbott), and Immulite 2000 (Siemens/DPC) according to manufacturer's data insert. LRL, lower reference limit; URL, upper reference limit.

	TSH		FT3		FT4	
	LRL [mU/L]	URL [mU/L]	LRL [pmol/L]	URL [pmol/L]	LRL [pmol/L]	URL [pmol/L]
DxI 800	0.34	5.60	3.20	6.00	7.86	14.4
AxSYM	0.35	4.94	2.22	5.35	9.14	23.8
Immulite 2000	0.40	4.00	2.30	6.29	10.3	24.5

Table 2. Contingency table comparing FT3 and FT4 results obtained on UniCel DxI 800 (Beckman Coulter) vs. AxSYM (Abbott), DxI 800 vs. Immulite 2000 (Siemens/DPC), and Immulite 2000 vs. AxSYM in serum samples with normal TSH levels (n = 372) on all three instruments. Concordant results between two methods are highlighted in bold. Reference ranges for TSH and FT4 according to manufacturer's data sheet. LRL, lower reference limit; URL, upper reference limit.

A)

		Concordance / Kappa	AxSYM		
		FT3 83.1% 0.584 FT4 60.8% 0.325	FT3 FT4 <LRL	FT3 FT4 normal	FT3 FT4 >URL
DxI 800	FT3 FT4 <LRL		4 (3.08%) 3 (2.31%)	5 (3.85%) 1 (0.77%)	- -
	FT3 FT4 Normal		8 (6.15%) 1 (0.77%)	86 (66.2%) 50 (38.5%)	4 (3.08%) -
	FT3 FT4 >URL		- -	5 (3.85%) 49 (37.7%)	18 (13.9%) 26 (20.0%)

B)

		Concordance / Kappa	Immulite 2000		
		FT3 76.2% 0.480 FT4 74.6% 0.545	FT3 FT4 <LRL	FT3 FT4 normal	FT3 FT4 >URL
DxI 800	FT3 FT4 <LRL		3 (2.31%) 4 (3.08%)	6 (4.62%) -	- -
	FT3 FT4 Normal		2 (1.54%) 2 (1.54%)	75 (57.7%) 48 (36.9%)	21 (16.5%) 1 (0.77%)
	FT3 FT4 >URL		- -	2 (1.54%) 30 (23.1%)	21 (16.5%) 45 (34.6%)

C)

		Concordance / Kappa	AxSYM		
		FT3 68.5% 0.329 FT4 83.1% 0.637	FT3 FT4 <LRL	FT3 FT4 normal	FT3 FT4 >URL
Immulite 2000	FT3 FT4 <LRL		2 (1.54%) 4 (3.08%)	3 (2.31%) 2 (1.54%)	- -
	FT3 FT4 normal		10 (7.69%) -	69 (53.1%) 78 (60.0%)	4 (3.08%) -
	FT3 FT4 >URL		- -	24 (18.5%) 20 (15.4%)	18 (13.9%) 26 (20.0%)

Table 3. Contingency table comparing FT3 and FT4 results obtained on UniCel DxI 800 (Beckman Coulter) vs. AxSYM (Abbott), DxI 800 vs. Immulite 2000 (Siemens/DPC), and Immulite 2000 vs. AxSYM in serum samples with decreased TSH levels (n = 130) on all three instruments. Reference ranges for TSH and FT4 according to manufacturer's data sheet. LRL, lower reference limit; URL, upper reference limit.

A)

Concordance / Kappa		AxSYM		
FT3 83.1% 0.584 FT4 60.8% 0.325		FT3 FT4 <LRL	FT3 FT4 normal	FT3 FT4 >URL
DxI 800	FT3 FT4 <LRL	4 (3.08%) 3 (2.31%)	5 (3.85%) 1 (0.77%)	- -
	FT3 FT4 Normal	8 (6.15%) 1 (0.77%)	86 (66.2%) 50 (38.5%)	4 (3.08%) -
	FT3 FT4 >URL	- -	5 (3.85%) 49 (37.7%)	18 (13.9%) 26 (20.0%)

B)

Concordance / Kappa		Immulite 2000		
FT3 76.2% 0.480 FT4 74.6% 0.545		FT3 FT4 <LRL	FT3 FT4 normal	FT3 FT4 >URL
DxI 800	FT3 FT4 <LRL	3 (2.31%) 4 (3.08%)	6 (4.62%) -	- -
	FT3 FT4 Normal	2 (1.54%) 2 (1.54%)	75 (57.7%) 48 (36.9%)	21 (16.5%) 1 (0.77%)
	FT3 FT4 >URL	- -	2 (1.54%) 30 (23.1%)	21 (16.5%) 45 (34.6%)

C)

Concordance / Kappa		AxSYM		
FT3 68.5% 0.329 FT4 83.1% 0.637		FT3 FT4 <LRL	FT3 FT4 normal	FT3 FT4 >URL
Immulite 2000	FT3 FT4 <LRL	2 (1.54%) 4 (3.08%)	3 (2.31%) 2 (1.54%)	- -
	FT3 FT4 normal	10 (7.69%) -	69 (53.1%) 78 (60.0%)	4 (3.08%) -
	FT3 FT4 >URL	- -	24 (18.5%) 20 (15.4%)	18 (13.9%) 26 (20.0%)

FT3 method comparisons of DxI 800 - AxSYM (n = 703), DxI 800 - Immulite 2000 (n = 698), and AxSYM - Immulite 2000 (n = 664) are shown in Figure 2. The regression equations are y = 0.795x +1.289, y = 0.466x +2.257, y = 2.077x -1.941, respectively. Distribution of results in panel D and F is more scattered compared to panel B showing less correlation with the other two methods. Method dependant reference values for all three FT3 assays are shown in Table 1.

FT4

The DxI 800 FT4 upper reference limit of 14.1 pmol/L [9], represented by the upper red line in Figure 1, nearly cuts the 75th percentile of the box (50% of all results are within the box). By increasing to DxI 800 FT4 upper reference limit e.g. to 16.1 pmol/L (Figure 1, green line), as recently calculated in a clinical comparison study, [10] this analytical issue can be reduced.

Table 4. Contingency table comparing FT4 results obtained on UniCel DxI 800 (Beckman Coulter) vs. AxSYM (Abbott), DxI 800 vs. Immulite 2000 (Siemens/DPC), and Immulite 2000 vs. AxSYM in serum samples with increased TSH levels (n = 50) on all three instruments. Concordant results between two methods are highlighted in bold. Reference ranges for TSH and FT4 according to manufacturer's data sheet. LRL, lower reference limit; URL, upper reference limit.

A)

Concordance / Kappa		AxSYM		
FT3 78.0% 0.497 FT4 84.0% 0.673		FT3 FT4 <LRL	FT3 FT4 normal	FT3 FT4 >URL
DxI 800	FT3 FT4 <LRL	10 (20.0%) 14 (28.0%)	2 (4.00%) -	- -
	FT3 FT4 Normal	9 (18.0%) 4 (8.00%)	29 (58.0%) 28 (56.0%)	- -
	FT3 FT4 >URL	19 (38.0%) -	3 (62.0%) 49 (8.00%)	- -

B)

Concordance / Kappa		Immulite 2000		
FT3 86.0% 0.615 FT4 90.0% 0.788		FT3 FT4 <LRL	FT3 FT4 normal	FT3 FT4 >URL
DxI 800	FT3 FT4 <LRL	8 (16.0%) 13 (26.0%)	4 (8.00%) 1 (2.00%)	- -
	FT3 FT4 Normal	1 (2.00%) 1 (2.00%)	35 (70.0%) 31 (62.0%)	2 (4.00%) -
	FT3 FT4 >URL	- -	- 3 (6.00%)	- 1 (2.00%)

C)

Concordance / Kappa		AxSYM		
FT3 78.0% 0.508 FT4 90.0% 0.778		FT3 FT4 <LRL	FT3 FT4 normal	FT3 FT4 >URL
Immulite 2000	FT3 FT4 <LRL	9 (18.0%) 14 (28.0%)	- -	- -
	FT3 FT4 Normal	9 (18.0%) 4 (8.00%)	30 (60.0%) 31 (62.0%)	- -
	FT3 FT4 >URL	1 (2.00%)- -	1 (2.00%) 1 (2.00%)	- -

FT4 method comparisons of DxI 800 - AxSYM (n = 749), DxI 800 - Immulite 2000 (n = 843), and AxSYM - Immulite 2000 (n = 707) are shown in Figure 2. The regression equations are y = 1.016x -1.744, y = 0.724x - 0.731, y = 1.361x -1.133, respectively. Distribution of FT4 results in panel D and F is more scattered compared to panel B showing less correlation with the other two assays. Method dependant reference values for all three FT4 assays are shown in Table 1.

Method comparison of peripheral free hormones FT3 and FT4 on the basis of their relations with TSH

TSH versus FT3 and FT4

Thyroid conditions are shown in a graphical 2-dimensional distribution of results on the basis of a logTSH vs. FT3 (Figure 4, panel A, C, and E) and FT4 (Figure 4, panel B, D, and F). In each diagram the centre rectan-

Table 5. Overview of overall concordances and kappa values for DxI 800 vs. AxSYM and DxI 800 vs. Immulite 2000 and AxSYM vs. Immulite 2000 for FT3 (part A) and FT4 (part B) using different DxI 800 FT4 upper reference limits. 1) FT4 URL from 2006 datasheet; 2) FT4 URL as determined at UCL St. Luc, Louvain, Belgium [10]. TSH and FT3 reference ranges according to manufacturer's data sheet. LRL, lower reference limit; URL, upper reference limit.

A)

		Concordance (Kappa)		
FT4	TSH	DxI 800 vs. AxSYM	DxI 800 vs. Immulite 2000	AxSYM vs. Immulite 2000
URL 14.4 [1] pmol/L	<LRL	60.8% (0.325)	74.6% (0.545)	83.1% (0.637)
	normal	83.9% (0.213)	85.5% (0.224)	95.2% (0.338)
	>URL	84.0% (0.673)	90.0% (0.788)	90.0% (0.778)
URL 16.1 [2] pmol/L	<LRL	73.1% (0.472)	85.4% (0.723)	
	normal	90.3% (0.187)	94.6% (0.458)	
	>URL	90.0% (0.768)	92.0% (0.804)	

B)

		Concordance (Kappa)		
FT3	TSH	DxI 800 vs. AxSYM	DxI 800 vs. Immulite 2000	AxSYM vs. Immulite 2000
	<LRL	83.1% (0.584)	76.2% (0.480)	86.5% (0.328)
	normal	89.2% (0.208)	79.0% (0.064)	75.3% (0.014)
	>URL	78.0% (0.497)	86.0% (0.615)	78.0% (0.508)

gular section represents samples with normal TSH and normal FT3 and normal FT4 values. The FT4 diagram for DxI 800 clearly shows differences in the normal reference rectangle compared to AxSYM and Immulite 2000. Furthermore, for Immulite 2000, panel E shows many patient samples with normal TSH but unusually elevated FT3 values. This constellation of results was not detected for the other systems (panel A and C).

FT3 and FT4 concordances in relationship to normal, increased, and decreased TSH values

As shown in Table 2, we selected 372 patients with normal TSH in the DxI 800, AxSYM, and Immulite 2000 assay (panel A, B, and C). 81.2% had normal FT4 values measured with DxI 800 and AxSYM (panel A), and only 2.69% had low FT4 levels in both assays (concordant results). In contrast 12.6% showed elevated FT4 values with DxI 800 but normal with the AxSYM assay (discordant results). The overall concordance was 83.9%. Following the same method, the overall concordance between DxI 800 and Immulite 2000 (panel B) for the same patient samples was 85.5% and for Immulite 2000 vs. AxSYM was 83.1% (panel C).

Kappa values for all three comparisons were determined as only fair in a range from 0.21 to 0.34. An imbalanced distribution of discordant results was observed with 12.6% (panel A) and 11.6% (panel B) for elevated values measured by DxI 800 whereas AxSYM and Immulite 2000 reported normal results.

The same approach was used for comparison of FT3 results (Table 2). We selected the same 372 patients with normal TSH in all three systems (panel A, B, and C). 87.4% had normal FT3 values measured with DxI 800 and AxSYM (panel A), and only 1.88% had low FT3 levels in both assays (concordant results). The overall concordance was 89.2%. Following the same method, the overall concordance between DxI 800 and Immulite 2000 (panel B) for the same patient samples was only 79.0% and for Immulite 2000 vs. AxSYM only 75.3% (panel C). Kappa values for all three comparisons were determined as only poor in a range from 0.01 to 0.20. An imbalanced distribution of discordant results was observed with 15.1% (panel B) and 15.9% (panel C) elevated FT3 values measured by Immulite 2000 whereas DxI 800 and AxSYM reported normal results.

The same approach was used for comparing FT3 and FT4 results with decreased TSH values (Table 3) and increased TSH values (Table 4).

Method comparison by classification of thyroid conditions on the basis of concordance tables using different DxI 800 FT4 upper reference limits (Table 5)

In a Belgian clinical evaluation, the DxI 800 FT4 upper reference value was estimated as 16.1 pmol/L [10]. We used this cut-off to recalculate the concordances of FT4 results between DxI 800 and AxSYM, and DxI 800 and Immulite 2000.

Concordance of FT4 results was improved for patient samples with normal TSH values from 83.9% to 92.5%

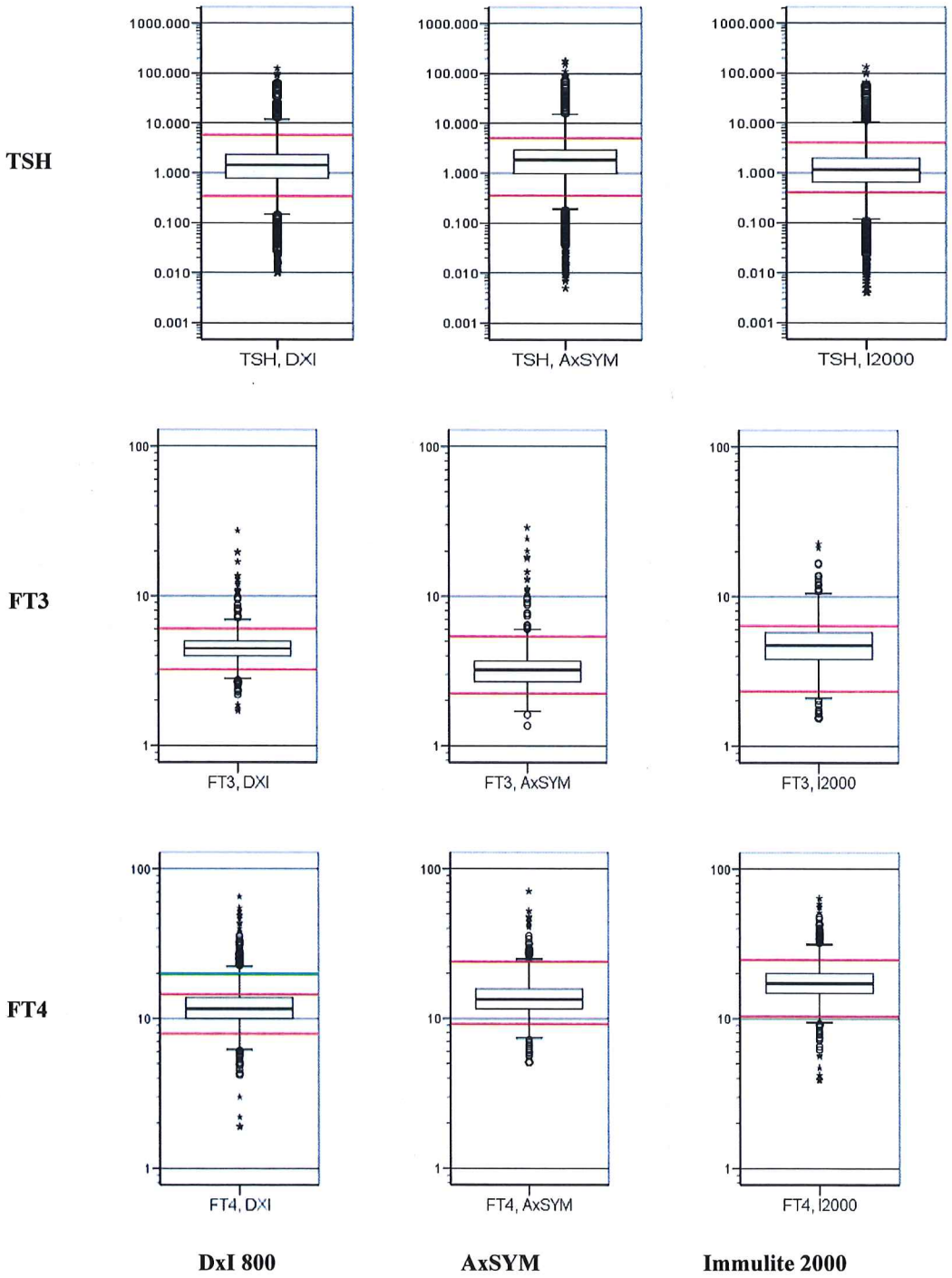

Figure 1. Box-and-whisker plots comparing TSH, FT3, and FT4 results determined on UniCel DxI 800 (Beckmann Coulter), AxSYM (Abbott), and Immulite 2000 (Siemens/DPC). Middle line represents median, box represents values from 25th to 75th percentile range. Symbols represent outliers with a distance more than 1.5 fold (empty circles) and 3 fold (stars) of IQR (interquartile range), respectively, outside of the box. The red lines indicate the upper and lower reference limits according to manufacturers' data insert. The green line for FT4 on DxI 800 represents the recommended upper reference limit of 16.1 pmol/L [10].

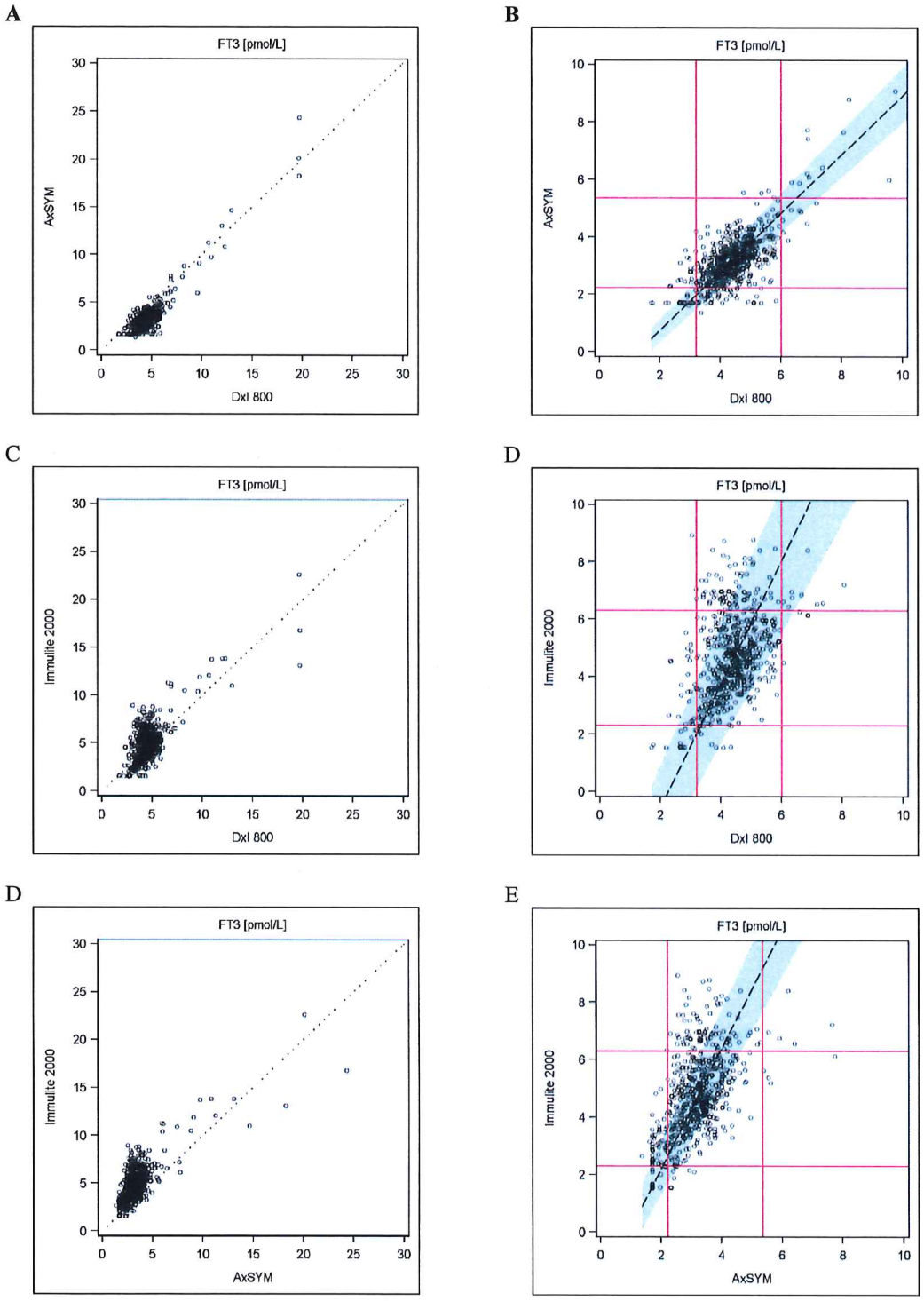

Figure 2. FT3 Method Comparison.
UniCel DxI 800 (Beckman Coulter) vs. AxSYM (Abbott), n = 703 (A, B); UniCel DxI 800 vs. Immulite 2000 (Siemens/DPC), n = 698 (C, D); and AxSYM (Abbott) vs. Immulite 2000 (Siemens/DPC), n = 664 (E, F). A, C, and E illustrate all data points with dotted identity line; B, D, and F use smaller concentration ranges with Passing-Bablok regression (dotted line) and 95% confidence interval (blue area) calculated with FT3 values between 2.0 and 10.0 pmol/L. Slope and offset for DxI 800 vs. AxSYM, 0.9750 and 1.288; DxI 800 vs. Immulite 2000, 0.4663 and 2.2567; Immulite 2000 vs. AxSYM, 2.0769 and -1.9408.

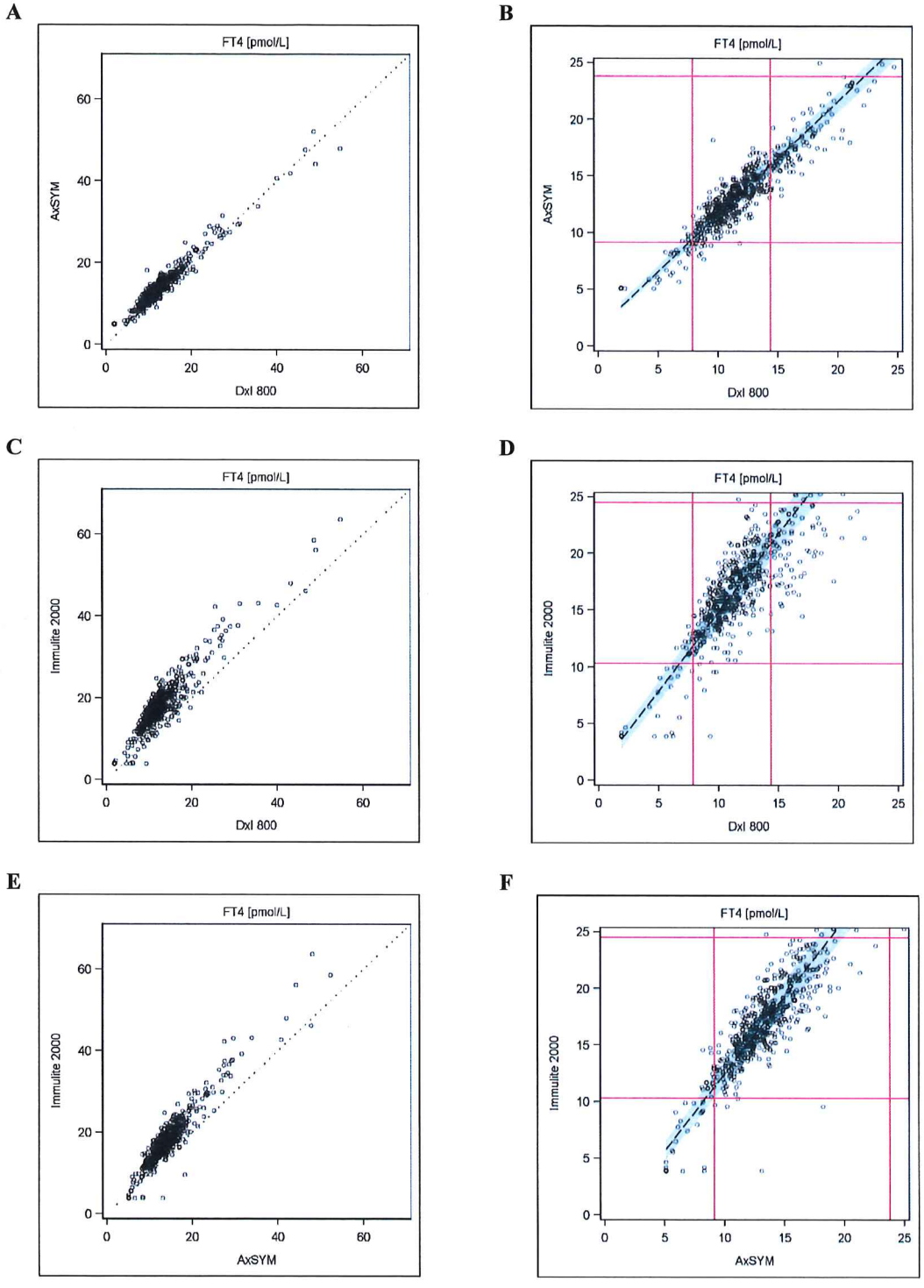

Figure 3. FT4 Method Comparison.
UniCel DxI 800 (Beckman Coulter) vs. AxSYM (Abbott), n = 749 (A, B); UniCel DxI 800 vs. Immulite 2000 (Siemens/DPC), n = 843 (C, D); and AxSYM (Abbott) vs. Immulite 2000 (Siemens/DPC), n = 707 (E, F); A, C, and E illustrate all data points; B, D, and F use smaller concentration ranges with Passing-Bablok regression (dotted line) and 95% confidence interval (blue area) calculated with FT4 values between 2.0 and 25.0 pmol/L. The green vertical line (B) represents the DxI 800 FT4 upper reference limit of 16.1 pmol/L [10]. Slope and offset for DxI 800 vs. AxSYM, 1.0164 and -1.7443; DxI 800 vs. Immulite 2000, 0.7241 and -0.7310; Immulite 2000 vs. AxSYM, 1.3611 and -1.1333.

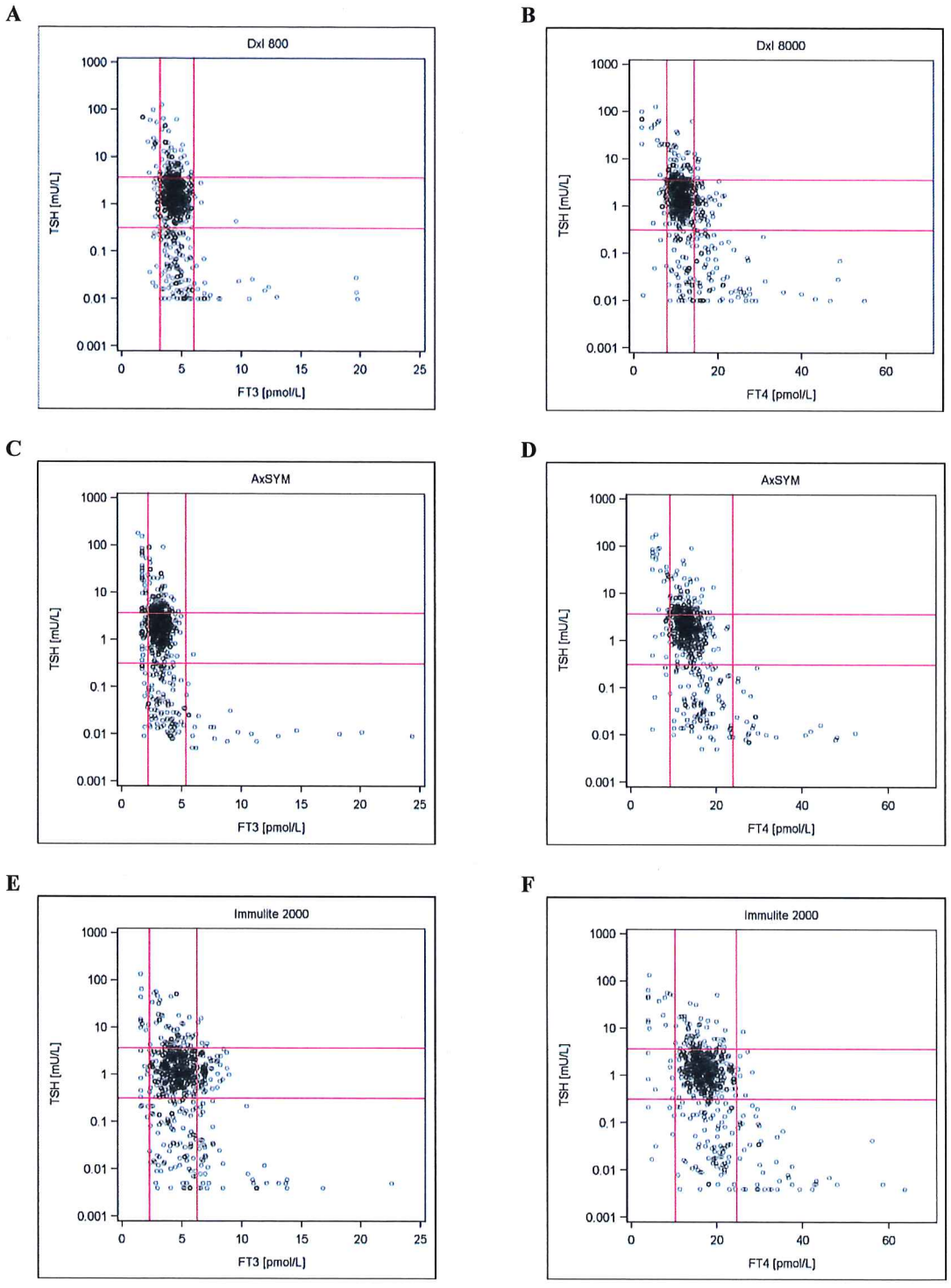

Figure 4. Correlation of logTSH values with linear FT3 and FT4 concentrations using 595 serum samples. LogTSH vs. FT3 (A) and FT4 (B) for DxI 800 (Beckmann Coulter); logTSH vs. FT3 (C) and FT4 (D) for AxSYM (Abbott); logTSH vs. FT3 (E) and FT4 (F) for Immulite 2000 (Siemens/DPC). The red lines indicate the lower and upper reference limit for TSH (horizontal) and FT3 and FT4 (vertical) according to manufacturer's data insert. The green line indicates the proposed DxI 800 FT4 upper reference of 16.1 pmol/L [10].

(DxI 800 vs. AxSYM), and from 85.5% to 96.2% (DxI 800 vs. Immulite 2000).
For samples with decreased TSH values concordances improved from 60.8 to 77.7% (DxI 800 vs. Immulite 2000) and from 74.6 % to 86.9% (DxI 800 vs. Immulite 2000). In addition FT4 concordance improved for patient samples with elevated TSH levels from 84.0% to 90.0% (DxI 800 vs. AxSYM) and from 90.0% to 92.0% (DxI 800 vs. Immulite 2000).

DISCUSSION

Different thyroid assays with different reference values report significantly different results and therefore patient's results are not interchangeable [11-14]. This has been known for almost 40 years without a suitable answer being found [15-17]. Therefore, implementation of new assays leads to confusion as clinicians must adapt to different numerical values and normal ranges especially in case of thyroid assays where a combination of three different assays (TSH, FT3, and FT4) is often used for diagnosis, treatment, and follow up of thyroid disorders. As standardization of these assays is still under discussion [18-28] laboratories have to work with these differences of different thyroid assays. Therefore, it is mandantory to validate FT3 and FT4 assays in their relationship to TSH.
With the Access thyroid assays another methodology with different numerical results and normal values was introduced in 2003. Using a very large cohort of Swiss patient samples we have evaluated the Access FT3 and FT4 assays in relationship to TSH and compared the assays with assays of two other manufacturers. We showed that all three TSH assays are comparable. Patient TSH values were reasonably comparable with a concordance of >90% between all systems and very good kappa values of >0.8.
The Access FT4 reference limits were re-evaluated in 2006 and found to be 7.86 - 14.4 pmol/L [9]. However, clinicians and endocrinologists considered this FT4 upper reference limit too low. In the same year a multicentric reference study with serum samples of 763 apparently healthy individuals (aged 18 - 65 years) from Germany, France, and Italy confirmed the Access FT4 upper reference limit of 14.4 pmol/L [29]. However, it was shown that the FT4 median values and central 95% ranges were significantly different in these European countries. Despite the well known differences in nutritional iodine-deficiencies and iodine uptake due to regional variations of total iodine content of food (two coast areas in France and Italy, one mountain region in Germany) the authors suggested a common Access FT4 reference interval of 7.8 - 14.4 pmol/L. This is an inappropriate estimation and contrary to good clinical practice to calculate local reference ranges.
Passing-Bablok regression comparing all three instruments showed a good correlation for DxI 800 vs. AxSYM, whereas DxI 800 vs. Immulite 2000, and AxSYM vs. Immulite 2000 exhibit a wider scatter of results (Figure 3) due to the higher imprecision of Immulite 2000 results.
Classification of thyroid conditions on the basis of a logTSH vs. free hormone diagram is a well accepted method [30]. Even if this graphical 2-dimensional probability distribution of TSH and FT3 or FT4 reference is not ideal as the widths of the TSH and FT3 or FT4 reference intervals derived from this bivariate distribution are mutually interdependent [31], the graphical plot allows a rapid understanding of differences in the division of the area into 9 rectangular sections. The logTSH/FT4 diagram for UniCel DxI 800 clearly shows differences in the normal reference rectangle compared to logTSH/FT4 of AxSYM and Immulite 2000 (Figure 4). The main reason for this difference is the DxI 800 FT4 upper reference limit of 14.1 pmol/L [9]. This effect is also visible in the Box-Whisker plot comparing FT4 results (Figure 1). The upper reference limit represents only the 78.2th percentile of all measured FT4 values. By increasing the DxI 800 FT4 upper reference limit to 16.1 pmol/L as recently determined in a clinical comparison study [10] the upper FT4 reference limit would represent the 86.2th percentile for the Swiss cohort.
Concordances between different FT4 assays were calculated from contingency tables. We plotted pairwise combinations of TSH and FT4 results obtained from three different instruments for decreased TSH values on all instruments (Table 2), normal TSH values (Table 3), and increased TSH values (Table 4). For normal TSH values, concordances between FT4 results ranged between 83.9% and 85.5% of the three system combinations. Although the concordances are relatively high, the kappa values determined exhibit only fair concordances between 0.213 and 0.338. One of the reasons is an unequal distribution of discordant results in contingency tables. Most of the discordant results are located in one box of the table e.g. for DxI 800 - AxSYM (Table 3, panel A) 12.6% of discordant results have a normal FT4 result on AxSYM and an elevated FT4 result on DxI 800. This effect is even more prominent for FT4 discordant results for decreased TSH values. Here FT4 concordance is only 60.8% with a fair kappa value of only 0.3258 because 37.7% of discordant results are within one box (AxSYM normal, DxI 800 increased, Table 2, panel A). In contrast, the concordance for AxSYM - Immulite 2000 is 83.1% with a moderate kappa value of 0.637. Here, 15.4% of discordant results derive from one FT4 combination (AxSYM normal, Immulite 2000 increased).
By increasing the DxI 800 FT4 upper reference limit e.g. to 16.1 pmol/L (representing the 86.2th percentile of our Swiss cohort) as recently calculated in a Belgian clinical comparison study [10], we were able to increase concordances and kappa values significantly. The overall percentage of concordance (kappa) between DxI 800 and AxSYM, and DxI 800 and Immulite 2000 was improved to 86.6% (0.494) and 91.9% (0.711), respectively.

For FT3 we observed an even worse effect of discordant results measured with Immulite 2000 for patient samples with normal TSH values. Concordance with AxSYM and DxI 800 values was only 75.3% and 79.0%. Imbalanced distribution of discordant values resulted in poor kappa values between 0.01 and 0.06. This was mainly caused by a too low FT3 upper reference limit for Immulite 2000.

There is an ongoing debate to lower the TSH upper limit of reference range regardless of the method used. Currently, some institutions use an upper limit for the TSH reference range of 5.0 mU/L while others use 4.0 or 3.0 mg/L. Some experts recommend an upper limit of 2.5 mU/mL [32-34]. We have recalculated the concordance tables by lowering the TSH upper reference limit for all three assays to 3.0 mU/L. However, this does not influence concordance results (kappa) for the different assay combinations (data not shown).

Our results show that the calculated overall concordances between all three systems are acceptable, even with some cases of discordance. However, the kappa values calculated indicate only a weak overall agreement for the three systems. Three reasons might explain these differences: 1. a strong disagreement of the reference limits, 2. the imprecision of the assays, and 3. an unequal distribution of discordant results. An indicator for the imprecision is the scatter in the logTSH vs. FT3/FT4 plot (Figure 4), this could also be shown in Bland-Altman plots or in the width of ellipse in Z-score analysis. Furthermore, our analytical classification is based on the assumption that some of the patients might have non-thyroidal pathologies. In these cases there is discordance between TSH and FT3/FT4 values and TSH alone cannot be used for pathophysiological classification.

In conclusion, the overall agreement of the methods investigated is acceptable with regard to the calculated concordances. However, there is a method-specific disagreement in reference limits; there are also relevant differences in the imprecision of some of the assays. Although the measuring techniques (e.g. GC or LC-IDMS) are capable of measuring free T3 and free T4 accurately, they do not solve the preanalytical problems e.g. specific separation of T3 and T4 from blood matrix. Therefore, harmonization of results has not been achieved. According to good laboratory practice, changing methods of thyroid function tests still requires evaluation of method-specific and population-specific reference limits.

Acknowledgment:

We are deeply grateful to Dr. Thomas Keller, Acomed statistik, Leipzig, Germany (www.acomed-statistics.com) for statistical analysis and advice. We thank Beckman Coulter Inc., Nyon, Switzerland, for providing reagents free of charge and we thank Grainne Delany, Bielefeld, Germany, for linguistic advice.

Declaration of Interest:

The authors declare that they have no competing interests.

References:

1. Stockigt J. Assessment of thyroid function: towards an integrated laboratory - clinical approach. Clin Biochem Rev 2003;24:109-22.
2. Demers LM, Spencer CA, eds. Laboratory Medicine Practice Guidelines: Laboratory Support for the Diagnosis and Monitoring of Thyroid Disease. National Academy of Clinical Biochemistry (NACB) 2002, p. 38. www.nacb.org.
3. AACE/AME Task Force on Thyroid Nodules. American Association of Clinical Endocrinologists and Associazione Medici Endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules. Endocr Pract 2006;12:63-102.
4. Gharib H, Tuttle RM, Baskin HJ, Fish LH, Singer PA, McDermott MT. Subclinical thyroid dysfunction: a joint statement on management from the American Association of Clinical Endocrinologists, the American Thyroid Association, and the Endocrine Society. J Clin Endocrinol MeTable 2005;90:581-5.
5. American College of Obstetrics and Gynecology. ACOG practice bulletin. Thyroid disease in pregnancy. Number 37, August 2002. American College of Obstetrics and Gynecology. Int J Gynaecol Obstet 2002;79:171-80.
6. Approved guideline - evaluation of precision performance of clinical chemistry devices. EP5-A 1999;19:2. Villanova PA: National Committee for Clinical Laboratory Standardisation.
7. Passing H, Bablok W. A new biometrical procedure for testing the equality of measurement from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry. Part. I. J Clin Chem Clin Biochem 1983;21:709-20.
8. Altman DG. Practical Statistics for Medical Research, Chapman & Hall/CRC, Boca 1991, p. 404.
9. Plouvier E, Alliot L, Bigorie B, Kowalski C, Medeau V, Thuillier F. Necessity to define thyroid reference values for a better clinical interpretation. Ann Biol Chem 2011;69:77-83.
10. Fillee C, Cumps J, Ketelslegers JM. Comparison of three free T4 (FT4) and free T3 (FT3) immunoassays in healthy subjects and patients with thyroid diseases and severe non-thyroidal illness. Clin Lab 2012;58:725-736.
11. Rawlins ML, Roberts WL. Performance characteristics of six third generation assays for thyroid stimulating hormone. Clin Chem 2004;50:2338-44.
12. Holm SS, Hansen SH, Faber J, Staun-Olsen P. Reference methods for the measurement of free thyroid hormones in blood: evaluation of potential reference methods for free thyroxine. Clin Biochem 2004;37:85-93.
13. Thienpont LM, van Uytfanghe K, Beastall G, et al. Report of the IFCC Working Group for Standardization of Thyroid Function Tests: part 2: free thyroxine and free triiodothyronine. Clin Chem 2010;56:912-20.

14. Thienpont LM, van Uytfanghe K, Beastall G, et al. Report of the IFCC Working Group for Standardization of Thyroid Function Tests: part 3: total thyroxine and total triiodothyronine. Clin Chem 2010;56:921-9.

15. Ellis SM, Ekins RP. Proceedings: The radioimmunoassay of free (diffusible) T3 and T4 concentration in serum. J Endocrinol 1973; 59:43.

16. Ekins RP, Jackson T, Sinha A, Edwards P. Principles of free hormone measurement. J Endocrinol Invest 1986;9 Suppl 4:3-15.

17. Ekins RP, Edwards P, Jackson T, Geiseler D. Interpretation of labeled-analog free hormone assay. Clin Chem 1984;30:491-3.

18. Ekins RP. Measurement of free hormones in blood. Endocrine Rev 1990;11:5-46.

19. Nelson JC, Wilcox RB. Analytical performance of free and total thyroxine assays. Clin Chem 1996;42:146-54.

20. Ekins RP. The science of free hormone measurement. Proc UK NEQAS Meeting.1998;3:35-59.

21. Christofides ND, Wilkinson E, Stoddart M, Ray DC, Beckett GJ. Assessment of serum thyroxine binding capacity-dependent biases in free thyroxine assays. Clin Chem 1999;45:520-5.

22. Sapin R. Serum thyroxine binding capacity-dependent bias in five free thyroxine immunoassays: assessment with serum dilution experiments and impact on diagnostic performance. Clin Biochem 2001;34:367-71.

23. Midgley JE. Direct and indirect free thyroxine assay methods: theory and practice. Clin Chem 2001;47:1353-63 [Review].

24. Fritz KS, Wilcox RB, Nelson JC. A direct free thyroxine (T4) immunoassay with the characteristics of a total T4 immunoassay. Clin Chem 2007;53:911-5.

25. Holm SS, Hansen SH, Faber J, Staun-Olsen P. Reference methods for the measurement of free thyroid hormones in blood: evaluation of potential reference methods for free thyroxine. Clin Biochem 2004;37:85-93 [Review].

26. Clinical and Laboratory Standards Institute (CLSI, formerly NCCLS). Measurement of Free Thyroid Hormones; Approved Guideline. CLSI document C45-A (ISBN 1-56238-548-8). CLSI, Wayne, Pennsylvania, 2004.

27. Thienpont LM, van Uytfanghe K, Beastall G, et al. Report of the IFCC Working Group for Standardization of Thyroid Function Tests: part 1: thyroid stimulating hormone. Clin Chem 2010;56: 902-11.

28. Van Uytfanghe K, Stöckl 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 2006;52:1817-21.

29. d'Herbomez M, Jarrige V, Darte C. Reference intervals for serum thyrotropin (TSH) and free thyroxine (FT4) in adults using the Access Immunoassay System. Clin Chem Lab Med 2005;43:102-5.

30. Demers LM, Spencer CA, eds. Laboratory medicine practice guidelines: laboratory support for the diagnosis and monitoring of thyroid disease. Washington (DC): National Academy of Clinical Biochemistry; 2002, p. 38.

31. Ross HA, den Heijder M, Hermus ARMM, Sweep CGJ. Composite interval for thyroid-stimulating hormone and free thyroxine, comparison with common cutoff values, and reconsideration of subclinical thyroid disease. Clin Chem 2009;55:2019-25.

32. Spencer CA, Takeuchi M, Kazarosyan M, MacKenzie F, Beckett GJ, Wilkinson E. Interlaboratory/intermethod differences in functional sensitivity of immunometric assays of thyrotropin (TSH) and impact on reliability of measurement of subnormal concentrations of TSH. Clin Chem 1995;41:367-74.

33. Grebe SK. Effects of reducing the upper limit of normal TSH values. JAMA 2003;290:3195-6.

34. Laurberg P, Anderson S, Carlé A, Karmisholt J, Knudsen N, Pederson IB: The TSH upper reference limit: where are we at? Nat Rev Endocrinol 2011;7:232-9.

Correspondence:
Prof. Dr. Wolfgang Korte
Institut für Klinische Chemie und
Laboratoriumsmedizin
Kantonsspital, Haus 04
9007 St. Gallen, Switzerland
Tel.: +41 71494 39 73
Fax: +41 71494 39 00
Email: wolfgang.korte@ikch.ch