IFCC Committee for Standardization of Thyroid Function Tests (C-STFT)
Meeting at AACC 2015, Atlanta, GA, USA, Monday July 27h (8:30 – 11:30 am)

PARTICIPANTS

The meeting attendance list is attached (Appendix A). The initials used in the minutes are contained in this list.
To avoid that all items dealt with in the meeting have to be repeated, the minutes are best read together with the accompanying slides (see Appendix B).

OPENING OF THE MEETING

The chair (LT) welcomed the meeting attendees. She started immediately with her slide presentation, more in particular the agenda (slide 2).
Editorial note:

- Dr. F. Mackenzie had sent before his apologies for the meeting.
- The Scientific Division of the IFCC has recently approved Dr. M. Patru (Ortho Clinical Diagnostics) for a first term as member of the IFCC "Committee on Standardization of Thyroid Function Tests (C-STFT)" (in replacement of Dr. F. Quinn, who resigned). Her time in office starts July 2015, ending December 31st, 2017.

(1) Familiarization Phase – Newcomers (by LT)

This phase applied for the 4 new companies who recently joined the project on invitation of LT. It was implemented as part of the step-up approach developed by C-STFT. LT summarized the applied design and results. She concluded that the Familiarization Phase worked adequately to (i) let familiarize the new companies with the C-STFT approach and (ii) allow judgement about the standardization status and quality of the new assays. In the case of the 4 newcomers, LT was happy that all got timely the permission to join the planned Phase IV study. Apart from that, most of them received recommendations to consider improvement of certain limitations of their assay.

No further questions were raised on this item by the audience.

(2) Phase IV method comparison studies for FT4 and TSH (by LT)

(a) Results

The results of the 2 method comparison studies, as reported to the manufacturers prior to the meeting, were summarized by LT. They formed the basis of a lively discussion.

FT4

- PS summarized the major observation: standardizing FT4 will be a major challenge for all manufacturers. The results in the euthyroid range are in closer agreement with the ED-ID tandem MS conventional reference measurement procedure (c-RMP) than in the other concentration ranges. However, the deviations from the c-RMP targets of all manufacturers (as a group) and amongst the manufacturers are per se critical. This should be an important topic in the education of all stakeholders.

- LT replied that this indeed will be an important part of the “risk-benefit” analysis implied by the FDA and started already by her (see below).

- JF mentioned there are 2 risks:
  - The reference interval (RI) will have to change.
  - There will remain aberrant results at the level of the individual sample (due to so-called sample-related effects), for which standardization will not account. LT confirmed that sample-related effects had been observed in Phase IV (see outliers in the difference plots). However, a cross check showed that different assays are distinctively susceptible (no particular sample giving issues for the majority of assays could be identified). PS replied that solving the sample-related effects could drastically impact/change an assay. LT argued

that solving these issues does not fall under the responsibility of the C-STFT but of each individual manufacturer. The strength of the conducted method comparison studies with clinical samples is that they can reveal the presence of these issues in certain assays, thus, most probably, some manufacturers worked already on these issues through the years. Both LT and JF stressed that it is to expect that, although manufacturers work on improving their assay, it will most probably never lead to the ideal assay for all kind of samples. Therefore, after standardization, it still will be important that laboratories know the limitations of their assays, and have a second (different) assay at hand for retesting samples with clinically aberrant results.

- GY stressed the need to know what the underlying cause is of some of the samples with aberrant results. LT replied that this might be difficult to find out because mostly the outliers represent a lack of specificity of an assay for (unknown) effects by the individual sample matrix.

- GY in addition stated that if these outlying results are not around the clinical decision limits, they may be of less importance. JF agreed with this from a clinical point of view, however, he recommended that these results should maybe not be taken into account upon recalibration of the assays, as they could negatively impact the quality and uncertainty of recalibration (confirmed by LT).

- MP asked for the possibility to receive a medication list of the donors of these samples. LT confirmed this list is available upon request; also the repository samples might be useful to get more insight into the origin of the sample-related effects.

- JF pinpointed the fact that primary healthcare pays more attention to TSH than FT4, so when addressing stakeholders, C-STFT might want to focus on TSH harmonization.

(b) Target setting approach (TSH)

LT discussed 2 possibilities to define the APTM: i) based on the harmonization panel only, as measured by 10 (11, see note) manufacturers or ii) based on both the harmonization and follow-up panel as measured by 4 manufacturers (= alternative approach). For both possibilities the advantages and disadvantages were discussed.
Note - Personal communication by LT: in a discussion in Atlanta with the company that did not yet submit measurements for Phase IV, LT was told the delay was due to instrumental malfunctioning encountered in the Familiarization Phase. They first wanted to fix this. Since apparently the problems are now solved, the company promised to do the Phase IV measurements ASAP. Dependent on them keeping this promise, their data will be included in the APTM for TSH based on 11 assays.

Discussion of this item:

- SM stated that all manufacturers will be biased in either approach of calculating the APTM, hence the first disadvantage of the alternative approach is not valid *(Somewhat "biased" target but in principle no problem (target is conventional, anyway and depends on the assays included in Phase IV)).* Moreover, the biggest advantage of the alternative approach is the direct link of the follow-up panel to the harmonization panel.
- In that regard MR agreed that the follow-up panel is very important and should be perfectly linked to the first panel (which is a fact using the alternative approach). Also the statistical answer to what the uncertainty on both panels is, will be an important factor in the decision making.
- LT confirmed the above and mentioned that for each follow-up panel, it will be the intention to measure it in overlap with the panel applying at that time. For this extra step, the uncertainty will be investigated by statisticians. On the request for when the 2nd follow-up panel is planned, LT answered this is difficult to anticipate (sufficient repository aliquots are left from the 1st follow-up panel, which only was measured by 4 companies), but she thinks maybe in about 2

years.

- GY asked why only 4 manufacturers measured the follow-up panel and how these manufacturers were selected. LT answered that i) only 4 manufacturers were selected in order not to deplete the follow-up panel too soon; enough material must be left for possible new-comers or retesting by manufacturers, if needed; ii) the 4 manufacturers were chosen based on their past results and performance (primarily in Phase III). Their results had a low scatter around the regression line and were nicely distributed around the APTM (above and below) over the complete measurement range.
- GB stressed there is not a huge difference between the 2 approaches to calculate the APTM (shown in a slide by LT), therefore, he would use the alternative approach.
- MR proposed to do the harmonization exercise by applying both approaches, and discuss, based on the actual results, which is the better approach, rather than on a pure theoretical discussion.
- GB added that in this way, the impact would indeed be better understandable.
- MR continued that this would also document to the FDA the rationale of the finally selected strategy.

It was agreed that manufacturers will receive both type of targets, so they can consider for both the effect of recalibration on their assay. However, in the end, the group has to come to an unanimous decision on which approach to use. Note by LT: the target based on 11 assays may be postponed until the end of August, because 1 company still had to report the Phase IV results.

- In addition, GY asked whether the WHO reference material would still be useful, i.e. for claims with regard to analytical sensitivity. LT explained that the APTM is related to the WHO due to the current calibration of the assays, which define the APTM. She also reminded of the fact that only using the WHO material does not lead to assays which are in agreement. The C-STFT approach adds one step to the traceability chain. This approach has been included in the draft of the new ISO 17511 (SM confirms).

(c) Next step after Phase IV (see slides) and discussion

- Establishment of a basic RI (n = 120) after recalibration of the FT4 and TSH assays – this should be the onset for further in-depth establishment of RIs by manufacturers using a higher n: MP asked the FDA representative what their recommendations are for the RI, is a common RI not sufficient? Are more than 120 subjects needed?
- TB answered that she thought that each manufacturer needs his own RI for verification purposes and that 120 subjects is the absolute minimum. Also RIs for subpopulations, e.g., in paediatrics, are needed.
- MR pointed to the big advantage of our study, in which all manufacturers will measure the same set of samples to establish their basic RI. This has never been done before, and is a real benefit of the work of the C-STFT.
- MP questioned again why the C-STFT RI couldn't be used and if verification is really needed.
- TB repeated it was recommended to validate an assay against this RI.
- SM stated that we all know that the RI will have to change and that we should profit from the advantages of our common RI: it will lower the risk as the RI will be the same over the different assays.
- PS added to the discussion that we shouldn't solely look at the standardization/harmonization status, but that there will be individual assay-specific differences. There are manufacturing limits for improvement in this regard. He also questioned whether it is worth measuring the samples with the "theoretically standardized/harmonized assay" (based on a mathematical function) and whether we shouldn't wait until the standardized/harmonized assays are really on the market, hence making use of the real recalibrated assay. This was later on in the meeting repeated by several other manufacturers.
- LT and MR stressed that the main purpose of the C-STFT exercise is to demonstrate that

common RIs are feasible and that recalibration is successful ("proof-of-concept"). In view of the aforementioned questions by the manufacturers, she had the impression that maybe the planned measurements for the RI came too early. However, she continued that, if these are done with the purpose of providing a proof-of-concept, they shouldn't be delayed too long. Eiditorial note: is it an option to send just one 0.5 mL aliquot to all manufacturers for the RI at this moment, and save the second aliquot for the proof of concept once the assays are effectively marketed?

- JR pointed to the challenges of the implementation phase, which will be huge.

(3) Homogeneity and stability study (by LT)

The results (homogeneity) and planning (stability) were presented; no discussion followed.

(4) Risk assessment (by LT)

The initiatives taken by UGent were presented; no discussion followed.

(5) Stability of performance of different platforms/assays (by KG)

Two online applications for mid- to long-term quality monitoring and developed by STT-Consulting and UGent were presented and discussed ("The Percentiler" and "The Flagger"). LT recalled why she had decided to use them in the C-STFT, i.e., to document the sustainability of FT4 standardization/TSH harmonization status established by the C-STFT activities (as requested by the FDA). She explained that to gain experience, the applications were already initiated in the pre-standardization/harmonization phase. All labs that participated for FT4 and TSH were either already participant in "The Percentiler" and "The Flagger" initially developed for clinical chemistry, or had accepted to join for FT4 and TSH only upon invitation by LT.

Discussion on this item:

- With regard to "The Flagger" (effect of analytical instability on the flagging rate), SN asked whether there is knowledge of the RIs that participants use. KG and LT replied that the UGent doesn't have access to this information, but relies on the RIs from the package inserts. Participants only report the percentage of results below (hypo) and above (hyper) the reference range limits.
- SN also asked how the data transition via the LIS (Laboratory Information System) works. KG and LT replied that this depends on the LIS used in the lab, and/or on the availability of an IT engineer in the lab. For example, in Belgium STT-Consulting and UGent have excellent contact with the provider of GLIMS (company name: MIPS), who helped with free-of-charge installation of data transition to the Percentiler and Flagger database. Other companies followed, however, some LIS providers ask for a fee when installing the transition. Others aren't capable at all to send the medians, and are not willing to work towards this. It would be advantageous if the information could be directly extracted from the middleware of the manufacturer's instruments. LT launched a call to the manufacturers to recruit more laboratories using their platforms/assays for participation.
- TB asked about the difference between inpatients and outpatients. LT explained that this difference mostly can be made by the LIS, depending on the coding of the samples. However, for several reasons it is not always possible to do so. In this case, the data are still useful, but only to investigate the stability of performance in that particular lab. If the distinction can be made, comparison to peers becomes possible, as the mid- to long term median of outpatients results reflects the typical value for an analyte in a healthy subject (usually it is very stable).

(6) Finances (by LT)

LT presented the financing plan for 2015.
Invoices will be sent latest by September.
No discussions followed.

(7) C-STFT website (by LT)

A website will be launched that catalogues all activities and results of the C-STFT group. In the meeting, there were no objections (yet) for the non-anonymous upload of the reports on the website as long as access is restricted to registered users (= C-STFT members + manufacturers), however, LT recalled that, apart from this, manufacturers had committed already to disclosure of results for publication of the Phase IV method comparison studies. Otherwise, there will be no change to get the data published.

(8) Future (by LT)

LT asked whether there is interest to expand the project with TT3/FT3? (IK) Interest for these analytes will be region-dependent, because the focus of primary healthcare is mostly on FT4 and TSH. In general everybody agreed that there are no steps needed, as they did not see a direct need from a clinical point of view.

Preparation for FDA submission/implementation – timelines and requirements:

- LT asked what would be seen as realistic timeline for implementing the recalibrated assays (she referred to a similar question forwarded in the time of creatinine recalibration, where 18 months had been mentioned by manufacturers). IK estimated that 18 months is too little, 2/3 years is more realistic. PS pointed to the fact that regulatory approval in Japan and China takes up to 2 years.
- PS added that the timing will also depend on the priority the manufacturers will give to this project.
    - MR stated that a clear demand from clinicians is needed to get funding and priority. Moreover he thinks that implementation is way too early, e.g. at the ATA-meeting, there was no awareness on our activities. LT mentions that together with GB and JF she is trying to get into contact with the international TAs, which will attend the 15th International Thyroid Congress in Orlando (October 18-23, 2015). See http://www.thyroid.org/itc2015/
- According to TB, for FT4 new submissions will be needed, for TSH an internal assessment will tell whether a submission is needed.
- GY added that according to Chinese regulations, when the RI changes, clinical studies are needed.
- IK asked the FDA representative about the need of a predicate assay. In conclusion of the discussion that follows between the manufacturers, LT and TB, it was stated that the assay-before-recalibration can serve as the paper predicate to document the pre and post status, but that the C-STFT approach will replace the predicate when it comes down to documenting the standardization of the assays. TB also said that the FDA would have to discuss whether extra samples were needed to document standardization. However, all participants were surprised about this statement. They thought this to be an extra effort which would not bring additional information, in view of the history of efforts and work done by the C-STFT through the years (Phases I to IV).
- LT reminded that upon meeting with the FDA in 2014, it was emphasized that every manufacturer should submit its dossier for FDA clearance on an individual basis, however, with clear reference to the C-STFT group. This should allow FDA referees to deal with the submissions by the individual manufacturers in a similar way.

Summary of actions-to-take and pending questions to answer:

1. **UGent to send the TSH APTM values by the 2 approaches; manufacturers to do the harmonization exercise (=recalibration) by applying both targets, and to decide, after validation of the results, on the better approach; LT to take care of a common decision.**

2. **Manufacturers to do the FT4 standardization exercise by applying the targets set by the ED ID-LC/tandem MS c-RMP of UGent.**

3. **UGent to work with statisticians on the uncertainty of the TSH APTM targets, and the uncertainty of the follow-up panel in its link to the previous panel.**

4. **How and when will the RI-panels be measured by the manufacturers after recalibration of their assays? Preferred is within this and a few months, but will the second aliquot be reserved for verification once the recalibrated assays are effectively marketed?**

5. **Provide to LT-KG coordinates and contacts of routine laboratories that possibly can participate in the Percentiler/Flagger for FT4 ad TSH. Post-standardization/harmonization monitoring is requested by the FDA. Some companies are already well represented, others underrepresented, some not at all. LT does not directly see how to recruit herself more labs than the 62 (with 90 platforms) she has already brought on board.**

6. **After discussion with their management, the IVD company representatives will confirm that the Phase IV reports can be posted on the C-STFT website (www.ifcc-cstft.org) with disclosure of the results, but with access restricted to C-STFT members and IVD partners.**

7. **IFCC secretary (Mrs. P. Bramati) to send the invoices for 2015 (2nd stage of Phase IV samples; scientific secretariat 2015-'16).**

Minutes made by:
Katleen Van Uytfanghe, PhD with help of Kenneth Goossens, PhD-student

Minutes approved by Prof. Dr. Linda Thienpont, chair of the IFCC WG-STFT
Laboratory for Analytical Chemistry, Faculty of Pharmaceutical Sciences, UGent
Ottergemsesteenweg 460, B-9000 GENT, Belgium
Tel. +32 9 264 81 04
e-mail: linda.thienpont@ugent.be; katleen.vanuytfanghe@ugent.be;

Appendix A

Name	Affiliation	e-mail address
Adelmann, Annette (AA)	Beckman Coulter	amandelmann@beckman.com
Aoyagi, Katsumi (KA)	Fujirebio	km-aoyagi@fujirebio.co.jp
Beastall, Graham (GB)	IFCC EB and representative of the BTA	gbeastall@googlemail.com
Bosworth, Tracey (TB)	FDA/CDHR/OIR/DCID	tracey.bosworth@fda.hhs.gov
Botelho, Juli (JB)	Centers of disease control and prevention	jbotelho@cdc.gov
Chen, Vincent (VC)	Snibe	vincent.chen@snibe.com
Das, Barnali (BD)	Member of C-STFT (ACBI; Kokilaben Dhirubhai Ambani Hospital and Medical Research Institute at Mumbai, India)	drbarnalidas@rediffmail.com
Faix, Jim (JF)	Member of C-STFT (AACC; Montefiore Med Ctr)	jfaix@montefiore.org
Gillery, Philippe (PG)	Member of the IFCC SD	pgillery@chu-reims.fr
Goossens, Kenneth (KG)	University of Ghent, Belgium	kenneth.goossens@ugent.be
Hall, Christina (CH)	Fujirebio	christina.hall@fdab.com
Hishinuma, Akira (AH)	Dokkyo Medical University	a-hishi@dokkyomed.ac.jp
Kutschera, Iris (IK)	Diasorin Inc.	iris.kutschera@diasorin.it
Lin, Ji (JL)	Abbott Labs	Ji.lin@abbott.com
Marivoet, Stefaan (SM)	Tosoh Europe	Stefaan.Marivoet@tosoh.com
Narayanan, Shanti (SN)	Tosoh Bioscience Inc.	Shanti.Narayanan@tosoh.com
Ogden, Judy (JO)	Tosoh Bioscience Inc.	judy.ogden@tosoh.com
Patru, Maria-Magdalena (MP)	Ortho Clinical Diagnostics	mpatru11@its.jnj.com
Reid, John (JR)	Abbott	John.reid@abbott.com
Ribera, Ashley (AR)	Centers of disease control and prevention	aribera@cdc.gov
Rottmann, Michael (MR)	Member of C-STFT (Roche Diagnostics)	michael.rottmann@roche.com
Sibley, E.C. Paul (PS)	Corresponding member (Siemens Medical Diagnostics)	paul.sibley@siemens.com
Straseski, Joely	ARUP Labs / U. of Utah	joely.a.straseski@aruplab.com
Strelkow, Ilja (IS)	In.vent Diagnostica	i.strelkow@inventdiagnostica.de
Suga, Koichi (KS)	Tosoh Bioscience Inc.	Koichi.suga@tosoh.com
Suzuki, Masako (MS)	Tosoh	masako-suzuki-hb@tosoh.co.jp
Terry, Carla (CT)	Tosoh Bioscience Inc.	carla.terry@tosoh.com
Thienpont, Linda (LT)	Chair IFCC C-STFT	linda.thienpont@ugent.be
Van Uytfanghe, Katleen (KVU)	Scientific secretary IFCC C-STFT (University of Ghent, Belgium)	Katleen.VanUytfanghe@UGent.be
Vesper, Hubert (HV)	Centers of disease control and prevention	hvesper@cdc.gov
Yu, George (GY)	Mindray	yuzhiguang@mindray.com

Excused

Name	Affiliation	e-mail address
Mackenzie, Finlay	Member of C-STFT (ACB; UKNEQAS)	Finlay.Mackenzie@uhb.nhs.uk

Appendix B
Slides from the annual meeting in conjunction with the AACC 2015 Conference

IFCC Committee for Standardization of Thyroid Function Tests (C-STFT)

Annual meeting in conjunction with the AACC 2015 Conference

AACC 2015 ANNUAL MEETING & CLINICAL LAB EXPO
JULY 26-30, 2015 > ATLANTA, GEORGIA

Chair
Linda Thienpont
Linda.thienpont@ugent.be

Scientific Secretary
Katleen Van Uytfanghe
Katleen.vanuytfanghe@ugent.be

Agenda

Discussion items

Familiarization phase for newcomers

Phase IV method comparison study for FT4 and TSH

Homogeneity and stability study

Risk assessment

Stability of performance of different platforms/assays

Finances

Future – Other items?

Agenda

Familiarization phase for newcomers

Phase IV method comparison study for FT4 and TSH

Homogeneity and stability study

Risk assessment

Stability of performance of different platforms/assays

Finances

Future – Other items?

Familiarization phase – Newcomers

4 new IVD-manufacturers accepted the invitation to join

Fujirebio (Tokyo, Japan)

Mindray (Shenzhen, China)

Snibe (Shenzhen, China)

Maccura (Chengdu, China)

FUJIREBIO mindray maccura Snibe Diagnostic

Familiarization phase – Approach

"Step-Up"#

Step 1: High-volume samples

Panel (n = 20 samples) from apparently healthy donors
Protocol of Phase I (2 replicates in 3 runs with ≠ lots/calibrators)
FT4 targets by the ED ID-LC/tandem MS cRMP
TSH targets: APTM from results by 4 assays (Phase I-III)

➔ **Step-up?**

Step 2: Low-volume clinical samples with incl. of master calibrators for recalibration by manufacturer

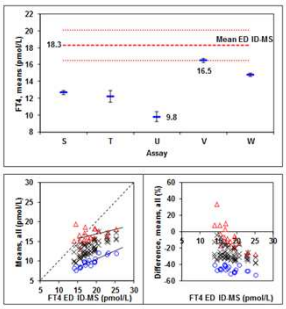
Leftover samples of Phase III
Similar measurement protocol (duplicates in 1 run) and targets

➔ **Step-up to Phase IV?**

#Van Uytfanghe K et al. Clin Chim Acta 2014;432:62-7.

Familiarization phase – Step 1

Status of standardization – FT4



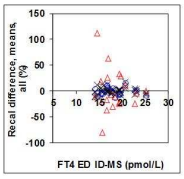
All assays but one measure FT4 much lower than the ED ID-MS cRMP (up to 46%)

Correlation is very good to reasonable for all assays (range r^2 : 0.93 – 0.80) but one (r^2 = 0.11)

Range of means is huge: 9.8 – 16.5 pmol/L

Familiarization phase – Step 1

Mathematical recalibration – FT4



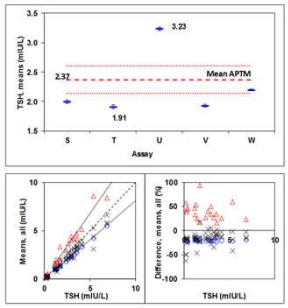
% Difference
Assay-specific calibration biases removed

Residual dispersion nearly entirely due to within-assay effects

For one assay, residual dispersion generally higher (to extreme high for some samples)

Familiarization phase – Step 1

Status of harmonization – TSH



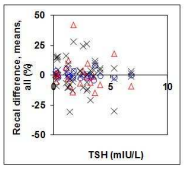
All but one assay deviate from the APTM by >10% (range from -22% to +40%)

Correlation is excellent to good (range r^2 : 0.998 – 0.92)

Range of means is considerable: 1.91 – 3.23 mIU/L

Familiarization phase – Step 1

Mathematical recalibration – TSH



% Difference
Assay-specific calibration biases removed

Residual dispersion nearly entirely due to within-assay effects

For 2 assays, residual dispersion higher (for 1 and 2 samples, resp.)

Familiarization phase – Step 1

Quality of the assays – FT4 and TSH

For most assays:

Precision – within- and total CV (%) – excellent

Accuracy – from IQC – good to reasonable

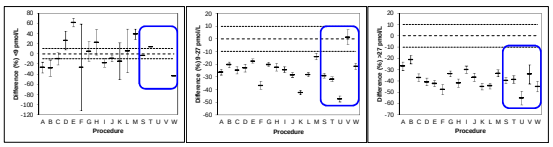
Stability – shifts and batch-to-batch variations – satisfactory, apart for some FT4 assays with drifting performance

Total random variation – % diff. of results of 1 R (but corrected for calibration bias) *vs* TE limits – good to reasonable for FT4 and TSH, but bad to borderline for 1 assay; in view of the low CV, "borderline for TSH" indicates high sample-related effects

➔ "Step-up" to the clinical samples of Phase III

Familiarization phase – Step 2

FT4



Observations mostly concordant with those in Step 1

Huge differences between assays

Considerable recalibration needed

Excellent to good performance in terms of precision, between-run differences, shifts and drifts

Except: Sample-related effects more pronounced on clinical samples (for some assays to a major extent)

Familiarization phase – Step 2

Sample-related effects – FT4

Sample-related effects		
	1.96 $SD_{\%\text{-res}}$ (Rep 1)	Outliers omitted
S	11.9	
T	12.4	1
W	15.5	1
U	17.1	1
V	36.7	#
TE-limit expanded to 11% for imprecision of ED-ID/MS; #no outliers omitted because too many		

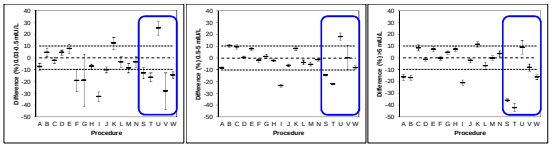
1.96 $SD_{\%\text{-res}}$ for 1 R (after correction for calibration bias) Reflects the combination of assay imprecision and sample-related effects

If CV_{wr} is low, a high 1.96$SD_{\%\text{-res}}$ reflects sample-related effects

Range: 11.9 to 36.7% (≥TE-limit)

Familiarization phase – Step 2

TSH



Observations concordant with Step 1

Deviation (pos. & neg.) from the APTM >10%

Recalibration needed

Except: Assay quality, i.e., precision (at low end) and sample-related effects worse on clinical samples

Familiarization phase – Step 2

Sample-related effects – TSH

Sample-related effects		
	1.96 $SD_{\%-res}$ (Rep 1)	Outliers omitted
S	11.6	
T	11.6	1
W	15.7	3
U	27.3	5
V	72.0	#
TE-limit 22.8%; #outliers not omitted because too many		

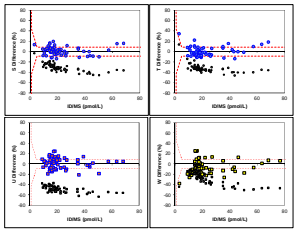
1.96 $SD_{\%-res}$ for 1 R (after correction for calibration bias)
Reflects the combination of assay imprecision and sample-related effects

A low CV_{wr} with a high 1.96$SD_{\%-res}$ is a good estimate for sample-related effects

Range: 11.6 to 72.0% (for 2 assays ≥TE-limit)

Familiarization phase – Step 2

FT4 Recalibration by manufacturers against ED-ID/MS targets

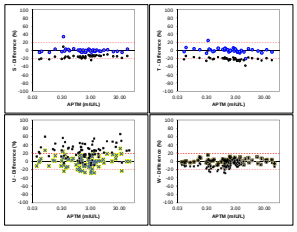


Successful: Assay calibration biases eliminated;
for some assays, optimization of recalibration in the low and/or high range would be beneficial

(NOTE: for one assay, no recalibration data were submitted)

Familiarization phase – Step 2

TSH Recalibration by manufacturers against Phase III APTM



Successful: Assay calibration biases eliminated;
remaining dispersion of the data nearly entirely due to within-assay random error, but the extent of dispersion varies

(NOTE: one assay did not submit recalibration data)

Familiarization phase – Conclusion

Apart from the aforementioned recommendations to improve certain limitations of their assay

All manufacturers were authorized to participate in Phase IV

Familiarization phase – Conclusion

Step-up approach# showed adequate

#Van Uytfanghe K et al. Clin Chim Acta 2014;432:62-7.

Agenda

Familiarization Phase for newcomers

Phase IV method comparison study for FT4 and TSH

Homogeneity and stability study

Risk assessment

Stability of performance of different platforms/assays

Finances

Future – Other items?

Phase IV method comparison – Design

FT4

Clinically relevant panel (n = 91)

Concentration range: 4.5 – 164 pmol/L (by ED-ID/MS)

Measurement protocol:

Sequence of samples: at random (selected by UGent)

Immunoassays: each manufacturer measures with master assay, only; 2 replicate measurements but in different runs; inclusion of master calibrators

ED-ID/MS: Measurements at least in triplicate in 3 independent runs

Individual results from immunoassays reported

Number of participating assays: 11

Sourcing of clinical samples

FT4/TSH

Two sources

Commercial companies
- **in.vent Diagnostica GmbH (Germany)**
- **Solomon Park Research Laboratories (SPRL) (USA)**

Clinicians
- **University of Sydney (Dr. P. Williams) (AU)**
- **University of Ghent (Dr. B. Lapauw), Louvain (Dr. B. Decallonne) & Brussels (Dr. B. Velkeniers) (BE)**
- **General hospitals Maria Middelares Ghent (Dr. P. Taelman), Sint-Jan Bruges (Dr. A. Van den Bruel) and O.L. Vrouw Aalst (Dr. P. Van Crombrugge) (BE)**
- **Private medical office in Yokosuka (Dr. A. Hishinuma) (JP)**

Ethical rules respected (de-ID of patients; informed consent)

Aliquoting of samples supplied by clinicians: by SPRL and in.vent

Sourcing of clinical samples

FT4/TSH (cont.)

In total 5 panels of samples collected

1. **FT4 standardization panel**
2. **TSH harmonization**
3. **TSH 1st follow-up**
4. **FT4 reference interval (RI)**
5. **TSH RI**

Information accompanying the samples: age, gender, diagnosis and/or treatment of the donating patients (stored on file at UGent and available on request)

Screening of the volunteers from SPRL before inclusion in the RI: TSH and anti-TPO titer (TOSOH AIA-2000) (courtesy: Narayanan S)

Sample storage at the National Institute for Biological Standards and Control (NIBSC) (UK) (courtesy: C. Burns)

Phase IV method comparison – FT4

Difference R1 - R2 (%)		
	Median (%)	SDdiff (%)
A	3.1	5.9
B	-0.4	4.1
C	-2.0	4.5
D	-1.1	5.6
E	0.2	3.6
F	4.1	5.6
G	0.6	3.9
H	1.4	2.5
I	3.1	2.7
J	1.0	5.0
L	0.9	4.4

Estimate for between-run imprecision and robustness of calibration

Median diff. >3% for 3 assays

Indicates between run calibration variability

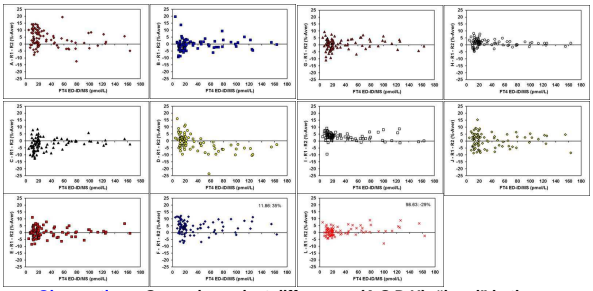
For other assays, differences concentration-dependent (see plots in next slide)

SDdiff

Range: 2.5% to 5.9%

Plots of differences between replicates

Overview – FT4



Observations: Conc.-dependent differences (A,C,D,H); "bend" in the eu- and hyperthyroid data (A,B,C)

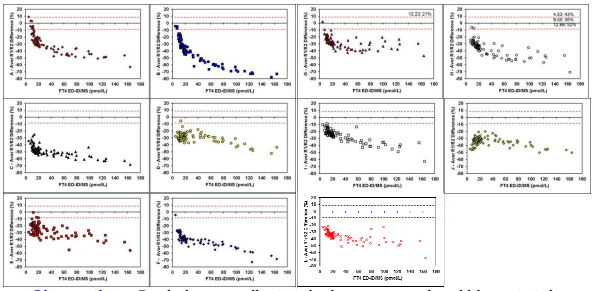
Phase IV method comparison – FT4

Assay Deviation from ED-ID/MS	Median (%) <10	Median (%) >10 <25	Median (%) >25 <100	Median (%) >100 pmol/L
A	-5.9	-24.6	-39.1	-49.0
B	-16.0	-34.2	-57.7	-72.7
C	-37.6	-46.6	-52.8	-62.9
D	-29.1	-27.4	-29.2	-45.5
E	-23.5	-22.3	-31.7	-42.4
F	-28.4	-37.9	-45.6	-58.7
G	-18.8	-27.2	-35.7	-30.2
H	-24.5	-32.0	-46.2	-48.9
I	-16.7	-22.5	-32.7	-41.6
J	-40.8	-32.4	-34.3	-46.1
L	-26.4	-33.7	-40.2	-45.2

Observation: Deviations strongly conc.-dependent

Plots of assay deviations from ED-MS

Overview – FT4



Observations: Deviations smallest at the low conc. end and biggest at the high end; range of dev. <10 pmol/L: ~-6 to -41%, >100 pmol/L ~-30 to -73%

Phase IV method comparison – FT4

Assay Residuals after fitting to ED-MS (%)	Median (%)	Robust SD (%)#	Fit
A	0.5	8.0	4 PL
B	0.5	5.7	4 PL
C	-1.1	7.3	4 PL
D	-0.7	6.5	4 PL
E	1.1	9.3	4 PL
F	0.6	4.8	4 PL
G	-1.1	6.6	4 PL
H	-2.3	8.2	4 PL
I	-0.7	5.8	4 PL
J	0.3	6.6	4 PL
L	0.3	7.2	4 PL

#Median absolute difference x 1.4826

Median deviation (fit)
Overall reasonable (apart from assay H); Fit: 4PL for all assays

NOTE: Specific fit used for recalibration will be the responsibility of each manufacturer

Robust SDdiff (%)
The lower it is, the more reliable the recalibration will be

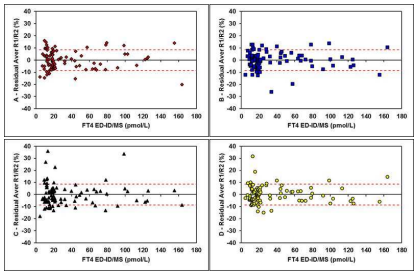
NOTE: Underestimates the data variability in case of outliers or non-random distribution (See plots)

Phase IV method comparison – FT4

Assay correlation to ED-MS (<25 pmol/L)	Fit	r^2
A	Linear	0.90
B	Linear	0.90
C	Linear	0.88
D	Linear	0.94
E	Linear	0.87
F	Linear	0.96
G	Linear	0.88
H	Linear	0.70
I	Linear	0.97
J	Linear	0.96
L	Linear	0.96

Plots of differences of the residuals

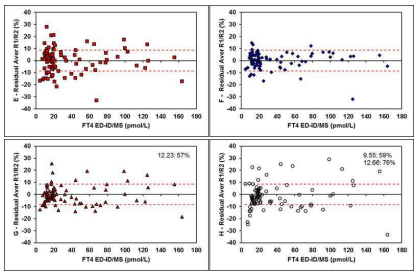
FT4 – Differences (mean of replicates; %) from ED-ID/MS *vs* TE limits (±8.6%, n = 2)#



#https://www.westgard.com/biodatabase1.htm

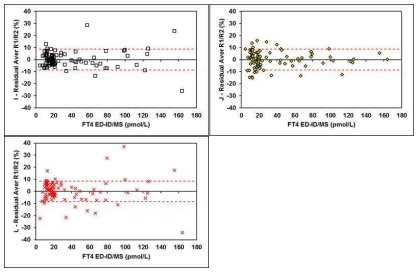
Plots of differences of the residuals

FT4 – Differences (mean of replicates; %) from ED-ID/MS *vs* TE limits (±8.6%, n = 2) (cont.)



Plots of differences of the residuals

FT4 – Differences (mean of replicates; %) from ED-ID/MS *vs* TE limits (±8.6%, n = 2) (cont.)



Phase IV method comparison – Design

TSH

Harmonization panel (n = 101 clinical samples)

First follow-up panel (n = 95)

Measurement range

Harmonization: ~0.002 to 75 mIU/L (APTM from 11 assays); 1 sample at 174 mIU/L
Follow-up: <0.002 to 118 mIU/L (APTM from 4 assays); 1 sample at 149 mIU/L and 1 at 164 mIU/L

Measurement protocol and reporting: Similar to FT4

NOTE: Follow-up and harmonization panel measured by 4 assays in the same run with random sequence of samples

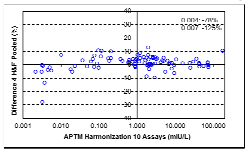
Phase IV method comparison – Design

TSH

Target setting: APTM calculated from 10 assays by a robust factor method# (zero's excluded, as well as 1 assay)

Alternative: APTM from 4 assays after pooling of their data for the harmonization and follow-up panel

NOTE: Difference between the two targets within ~±10%



#Stöckl et al. Clin Chem Lab Med 2014;52:965-72.

Phase IV method comparison – Design

TSH alternative for target setting

Disadvantage

Somewhat "biased" target but in principle no problem (target is conventional, anyway and depends on the assays included in Phase IV)

Fewer assays that contribute to the target

Advantage

Target estimate based on ~ 2x the sample-size with better distribution of concentrations:

- **25 samples <0.01 mIU/L (allows reliable insight into assay sensitivity)**
- **7 samples >75 mIU/L (allows a better estimate of assay calibration in that range)**

Phase IV method comparison – Design

TSH

Further advantage

Follow-up panel needs no value transfer from the harmonization (U follow-up = U harmonization panel)

NOTE

Important is the "fit-quality" of both targets to be used for recalibration (will be investigated from U calculations with our statistician)

Phase IV method comparison – Design

Target setting approach is open for discussion by the group

?

Phase IV method comparison – TSH

Difference R1 - R2 (%)	Median (%)	SDdiff (%)
A	-0.5	2.8
B	3.7	5.8
C	1.1	9.7
D	0.8	23.4
E	0.0	24.6
F	0.6	8.0
G	-1.0	14.5
H	0.3	2.6
I	0.3	2.9
J	-1.2	4.9
L	-2.0	7.0
Range: >0.02 to <75 mIU/L		

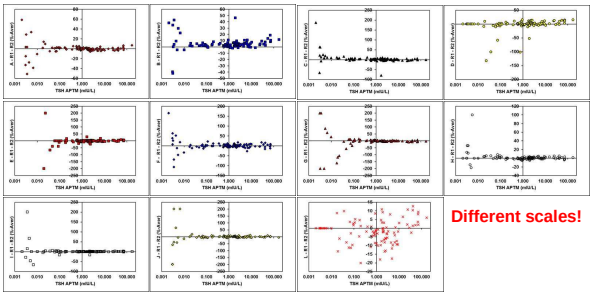
Estimate for between-run imprecision and robustness of calibration

Median diff. >3% for assay B; shows also conc.-dependent diff. at the low/high end; is similar for assay L at high conc. (see plots in next slide)

SDdiff
Range: 2.6% to 24.6%
Note: High values (~15 to 25% for assay D, E, G) typically due to increased SDdiff >0.02 to <0.15 mIU/L. Zero differences (see plots in next slide) caused by truncation of results <LoQ

Plots of differences between replicates

Overview – TSH

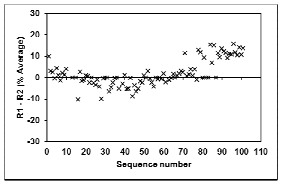


Different scales!

Observations: Conc.-dependent sign. diff. >3% (B); similar sign. diff. at high end (L); High SDdiff (~15-25%) >0.02 to <0.15 mIU/L (D,E,G)

Plots of differences between replicates

TSH



Certain assays show deviations which depend on the measurement sequence

Phase IV method comparison – TSH

Assay Deviation from APTM	Median (%)	SDdiff (%)	Robust SD (%)
A	0.5	5.1	2.8
B	-6.7	7.5	6.1
C	19.4	20.8	21.1
D	2.5	21.1	6.1
E	4.1	21.5	17.4
F	-6.3	7.6#	6.5
G	-1.2	10.4#	4.1
H	-5.1	13.4	9.3
I	7.8	10.2	11.1
J	7.4	7.1	6.0
L	-13.9	6.9	5.1
Range: >0.02 to <75 mIU/L; #1 Outlier			

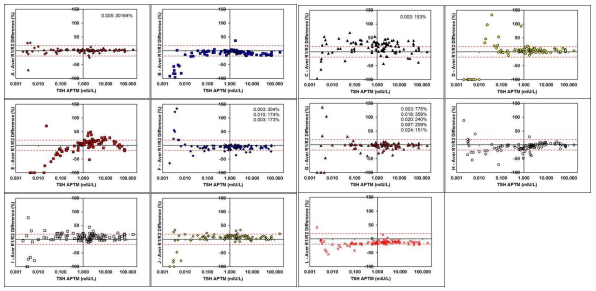
Median deviation (%)
For most assays – apart from 2 – within 10%

SDdiff (%)
Range from 5 to ~20%, the latter due to higher conc.-dependent dev. (D) or over whole range (C)/combined effects (E)

Robust SDdiff (%)
The lower, the more reliable the recalibration
NOTE: Underestimates the data variability ("outliers" or non-random distribution)

Plots of assay deviations from APTM

Overview – TSH



Observations: Median deviation >10% (C,L); High SDdiff (~20%) (C,D,E)

Phase IV method comparison – TSH

Assay residuals after fitting to APTM (%)	Median (%)	Robust SD (%)	Fit
A	-0.3	2.8	W-linear
B	5.0	7.0	W-linear
C	2.3	16.5	4 PL
D	-0.9	5.7	W-linear
E	0.3	7.6	4 PL
F	0.4	8.2	W-linear
G	0.2	5.2	W-linear
H	-0.8	8.9	4 PL
I	-1.7	9.7	W-linear
J	0.7	6.0	W-linear
L	1.0	5.9	W-linear
Range: >0.02 to <75 mIU/L			

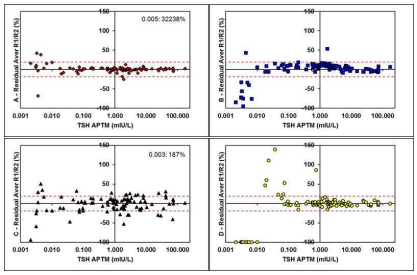
Fit (median deviation)
Overall reasonable (apart from assay B)
NOTE: Manufacturers responsible for fit

Robust SDdiff (%)
The lower, the more reliable recalibration

NOTE: Distribution of res. indicative for LoQ; issues if variation of res. sign. increases below a certain conc.; res. leveling off to pos. or neg. values, indicate suboptimal fit in that range

Plots of differences of the residuals

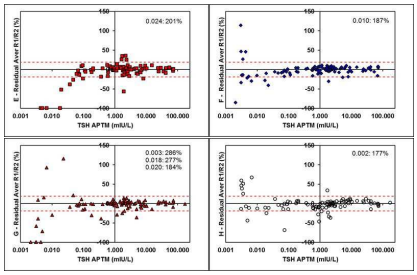
TSH – Differences (mean of replicates; %) from APTM *vs* TE limits (±19.6%, n = 2)#



#https://www.westgard.com/biodatabase1.htm

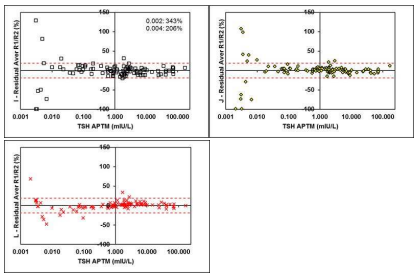
Plots of differences of the residuals

TSH – Differences (mean of replicates; %) from APTM *vs* TE limits (±19.6%, n = 2) (cont.)



Plots of differences of the residuals

TSH – Differences (mean of replicates; %) from APTM *vs* TE limits (±19.6%, n = 2) (cont.)



1st Follow-up panel – TSH

Target setting
By the selected 4 assays after recalibration either

- **Against the harmonization APTM (10 assays)**

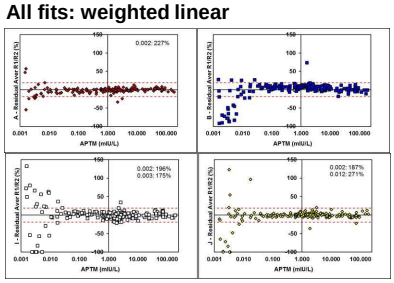
or

- **Alternatively, against the APTM from the 4 assays after pooling of their data for the harmonization and follow-up panel**

1st Follow-up panel – TSH

Observations for the 4 assays
(pooling of the data for the harmonization and follow-up panel)

All fits: weighted linear



Slope range:
0.856 to 1.063
Intercept range:
-0.001 to 0.007

All have good fit >0.01 mIU/L

Residuals
widen/level off <0.01 mIU/L, only
➔ Data suggest justified choice

Next step after Phase IV – FT4/TSH

Recalibration of assays
By manufacturers based on the included master calibrators

Reference interval (RI)
Panel of 120 samples from apparently healthy Americans to be measured by recalibrated assays; FT4 target values are currently set by ED-ID/MS at UGent

➔Proof-of-concept for success of standardization/ harmonization and feasibility to use a common RI

➔Basis for further establishment by manufacturers of new RIs after standardization/harmonization

Agenda

Familiarization phase for newcomers

Phase IV method comparison study for FT4 and TSH

Homogeneity and stability study

Risk assessment

Stability of performance of different platforms/assays

Finances

Future – Other items?

Homogeneity study

FT4/TSH

Subset of samples not included in any of the panels (similar specs for collection/processing in parallel with the Phase IV samples)

Six samples collected/processed by SPRL, 6 by in.vent (selected to cover the eu-, hypo- and hyperthyroid concentration range)

Measurement design similar to that used in the studies for certified reference materials (CRMs) from the EC# (8 randomly chosen aliquots of a sample measured in singleton *vs* 8 measurements out of a pool of 4 extra randomly chosen aliquots of the same sample; single aliquots and pool measured in an alternating sequence within run)

Study conducted by Roche (Cobas) (courtesy: M. Rottmann)

Outcome: statistical testing confirmed acceptance of H_0 (= homogeneity) (*p* of F-test (95%) > 0.05), except for 1 sample

#The certification of progesterone in two lyophilized serum materials, CRM 347 and CRM 348. EUR 12282 EN

Stability study

FT4/TSH

Duration: 2 years

Nine FT4 and 9 TSH samples (samples selected to cover the complete concentration range)

Study design#: compare the values for samples stored for different time periods at the effective storage temperature (-70°C) *vs* at the reference temperature (liquid N_2); 4 storage time points – time 0, 8, 16 and 24 months at -70°C; to avoid any complication due to measurement errors/variation, all samples will be measured at the same time at the end. Therefore the design is such that at time points 8 and 16, one box of samples need to be moved from the reference to the storage temperature

Measurements will be done by Roche (Cobas) (courtesy: M. Rottmann)

Stability monitoring started

#Discussed with I. Zegers and H. Schimmel from the Institute for Reference Materials and Measurements (IRMM) (Geel, Belgium)

Agenda

Familiarization phase for newcomers

Phase IV method comparison study for FT4 and TSH

Homogeneity and stability study

Risk assessment

Stability of performance of different platforms/assays

Finances

Future – Other items?

Risk assessment

Rationale

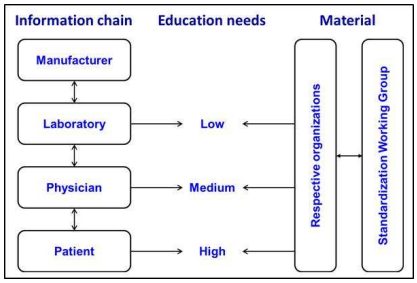
Do not "a-priori" assume that standardization or harmonization will be beneficial for the patient

Do assessment of risks/actions to mitigate risks

UGent started to solicit input for risk analysis/actions at the different stakeholder interfaces

Risk assessment

Information chain, education needs and elaboration of educational material



Risk assessment

Undertaken actions

Manufacturers

1st Action: Templates at https://drive.google.com/folderview?id=0B47uofxuvldnWENKUFhSZ21pUGs&usp=sharing

2nd Action: Questionnaire asking about communication lines in case of RI changes and/or implementation of global standardization (Tip: refer to past cases such as HbA1c and creatinine)

Laboratories and laboratory medicine societies (IFCC Member Societies)

Ask about potential risks and actions to mitigate:
- **Directors of big laboratories (ARUP, Mayo, LaCorp, UGent)**
- **Directors of laboratories nominated by IFCC Member Societies**

Risk assessment

Undertaken actions

Physicians and clinical societies
- **Case studies sent to Belgian GPs**
- **Attendance of scientific clinical meetings (LT/KVU, questionnaire)**
- **Launch of a one-pager in several "clinical" journals (to establish contact with clinical societies and their members, and ask their input for risk analysis) (authors: LT, J. Faix and G. Beastall)**

Patient Organizations
- **Approach Thyroid Foundations (TFs): TF Intern. (Canada), UK TF (courtesy: G Beastall), the Belgian and Swedish TF (courtesy: P. Lakwijk)**
- **Publish a one-pager in their respective magazines (authors: LT, J. Faix and G. Beastall)**
- **Present on our work at the 15th International Thyroid Congress (ITC), Oct. 18-23, 2015, Orlando, FL (courtesy: G. Beastall; J. Faix)**

Risk assessment

Undertaken actions

Regulation

First contacts made with the Chinese FDA (Beijing Institute of Medical Device Testing (BIMT) and Beijing Center for Medical Device Quality Supervision and Testing) (courtesy: P. Sibley from Siemens)

NOTE: BIMT has been assigned the administrative secretariat of the SAC/TC136 (=Technical Committee of Clinical Laboratory Testing and In Vitro Diagnostic Test Systems of Standardization Administration of China), which mirrors the ISO/TC212

Agenda

Familiarization phase for newcomers

Phase IV method comparison study for FT4 and TSH

Homogeneity and stability study

Risk assessment

Stability of performance of different platforms/assays

Finances

Future – Other items?

Stability of performance of platforms/assays

The Percentiler

The Flagger

EMPOWER IVD GLOBE

By Kenneth Goossens, PhD

Agenda

Familiarization phase for newcomers

Phase IV method comparison study for FT4 and TSH

Homogeneity and stability study

Risk assessment

Stability of performance of different platforms/assays

Finances

Future – Other items?

C-STFT website

www.ifcc-cstft.org

Agenda

- Familiarization phase for newcomers
- Phase IV method comparison study for FT4 and TSH
- Homogeneity and stability study
- Risk assessment
- Stability of performance of different platforms/assays
- Finances

Future – Other items?

Future

- TT3 – FT3?
- Phase IV: publication(s)
- RI studies
- Prepare FDA submission
- Prepare for implementation
- Other items?

Acknowledgements/Thanks

Dietmar Stöckl
Prof. S. Van Aelst
OpenAnalytics
Clinicians

Laboratories willing to participate in the Percentiler/Flagger