

Mini Review

Ben Cowper*, Alicia N. Lyle, Hubert W. Vesper, Katleen Van Uytfanghe and Chris Burns

Standardisation and harmonisation of thyroid-stimulating hormone measurements: historical, current, and future perspectives

<https://doi.org/10.1515/cclm-2023-1332>

Received November 22, 2023; accepted January 18, 2024;

published online January 31, 2024

Keywords: thyroid-stimulating hormone; standardisation; reference materials; commutability; harmonisation

Abstract: Thyroid-stimulating hormone (TSH) is an important clinical marker in the diagnosis and management of thyroid disease. TSH measurements are reported in milli-International Units per Litre (mIU/L), traceable to a World Health Organisation (WHO) reference material. There is a wide variety of commercial immunoassays for TSH measurements available, which have historically been poorly harmonised due to a lack of commutability of the WHO reference materials with patient samples. This led to the recent development of a serum-based reference panel for TSH, traceable to the WHO reference material, available via the International Federation for Clinical Chemistry and Laboratory Medicine (IFCC), aimed at harmonisation of TSH immunoassays. This report describes recent developments in the TSH reference system, including establishment of the 4th WHO International Standard for TSH, and aims to clarify the relationship between the available reference materials and their intended uses. This 4th WHO IS is widely available and defines the unit of TSH activity, therefore its continued existence is of paramount importance, however it continues to show a lack of commutability with patient in many TSH immunoassays. This makes the C-STFT TSH panel, albeit available in restricted numbers, a critical resource to ensure better TSH assay harmonisation.

Introduction

Thyroid-stimulating hormone (TSH) is secreted from the pituitary gland and stimulates release of thyroid hormones; triiodothyronine (T3) and thyroxine (T4) from the thyroid gland. In return, thyroid hormones in the blood regulate the pituitary release of TSH. Secreted thyroid hormones are distributed throughout the human body, and elicit wide-ranging effects upon various metabolic, cardiovascular and developmental processes [1].

An interruption of this highly regulated feedback loop can cause a variety of diseases. Hypothyroidism (e.g. due to Hashimoto's thyroiditis, or iodine deficiency) is characterised by an underactive thyroid gland, causing low levels of thyroid hormones, leading to weight gain, fatigue, and hair loss, amongst other symptoms. Hyperthyroidism (e.g. due to Graves' disease or thyroid/pituitary adenoma) occurs where an overactive thyroid gives rise to high levels of thyroid hormones, causing symptoms including weight loss, palpitations and insomnia [2]. In each case interventions are available which aim to restore normal thyroid function [3, 4].

The measurement of TSH is a crucial component in the diagnosis of thyroid disorders, and the monitoring of ensuing therapy [5]. TSH is often the initial biomarker measured in clinical practice, and when blood TSH values indicate the presence of a disease, further investigations are initiated to establish a diagnosis. Hypo- and hyperthyroidism are usually associated with abnormally high or low TSH levels respectively.

A wide variety of commercial immunoassays for TSH measurement are available globally. It is widely accepted that TSH values of 0.3–0.5 and 4–5 mIU/L respectively represent the lower and upper limits of the “normal” TSH range; reflected by the listed reference intervals of these commercial immunoassays [6]. However, TSH immunoassays have historically been poorly harmonised, leading to

Hubert W. Vesper and Katleen Van Uytfanghe: on behalf of the IFCC Committee for Standardization of Thyroid Function Tests (C-STFT).

***Corresponding author: Ben Cowper**, Medicines and Healthcare Products Regulatory Agency (MHRA), Blanche Lane, South Mimms, Hertfordshire, EN6 3QG, UK, Phone: +441707641246, E-mail: ben.cowper@mhra.org

Alicia N. Lyle and Hubert W. Vesper, Centers for Disease Control and Prevention, Division of Laboratory Sciences, Atlanta, GA, USA

Katleen Van Uytfanghe, Laboratory of Toxicology, Ghent University, Ref4u, Gent, Belgium

Chris Burns, Medicines and Healthcare Products Regulatory Agency (MHRA), South Mimms, Hertfordshire, UK

the potential for inconsistent diagnosis and management of an abnormal TSH level between clinical laboratories, depending on the assay used. Clinicians have therefore needed to be mindful of assay-specific biases when interpreting “borderline” TSH results [5, 7].

It is therefore important to ensure harmonisation of TSH measurements between assays, laboratories, and geographical regions, to ensure consistency in the diagnosis and management of thyroid disorders. Having harmonised assays allows for the development of generally accepted common reference intervals (RIs) and clinical decision limits, ideally derived across different assays, thereby accounting for the remaining inter-assay variation.

The TSH reference system

TSH reference materials to establish International Units

TSH is a heterogeneous glycoprotein, of which the different isoforms occurring in human blood and tissues are not yet well characterised [8]. Furthermore, a robust method for quantitation of total TSH or specific isoforms in serum in mass units has yet to be achieved. Consequently, serum TSH measurements are reported in milli-International Units per Litre (mIU/L). The mIU of TSH activity is defined by the World Health Organisation (WHO) reference material, which is available from the National Institute for Biological Standards and Control (NIBSC) catalogue of the UK Medicines and Healthcare Products Regulatory Agency (MHRA). The 1st International Reference Preparation (IRP), coded 68/38, was established in 1975, and was characterised by immunoassays and bioassays. This was replaced in 1983 with the 2nd IRP, 80/558, designated specifically for use in immunoassays [9], followed by the 3rd International Standard (IS), 81/565, established in 2003 [10]. The latest iteration of the

standard, the 4th IS, 81/615, was established by WHO in 2023 [11]. The three most recent preparations, 80/558, 81/565, and 81/615, were sequentially established as the WHO IS across a 40-year period, however they were prepared at approximately the same time (in the early 1980s), using the same batch of bulk TSH, extracted from human pituitary gland [9]. Each replacement standard has been value-assigned in mIU per ampoule, calibrated against the existing standard, thus maintaining traceability of unitage through the product lifecycle (Figure 1) [9–11].

Although a common global calibrator for TSH measurements has been available from WHO since the 1970s and this calibrator has been widely used to calibrate commercial TSH immunoassays, for many years TSH measurements were observed to be poorly harmonised across different assays [12]. Research studies have found that certain reference material preparations can show very different instrument responses than patient samples even though they are believed to have the same analyte concentration. These findings have led to the definition of ‘commutability’; the equivalence of the relationship between the results of different measurement procedures for a reference material and for patient samples [13]. It is therefore assumed that the current and previous WHO standards have had poor commutability [14, 15]. This is likely due to use of pituitary TSH to prepare the WHO standards. Previous studies have shown that pituitary TSH contains predominantly lower-complexity glycoforms, which differ from the highly-sialylated forms which predominate in serum TSH [16–18]. Therefore, the TSH contained within the WHO standards is unlikely to be fully structurally representative of TSH contained in patient serum samples and thus may not exhibit equivalent immunological responses across different commercial immunoassays which make use of different anti-TSH antibodies aimed at detecting serum TSH. Calibration using standards containing pituitary TSH is therefore likely to have introduced “bias” in TSH measurements across the various assays, resulting in an observed lack of harmonisation.

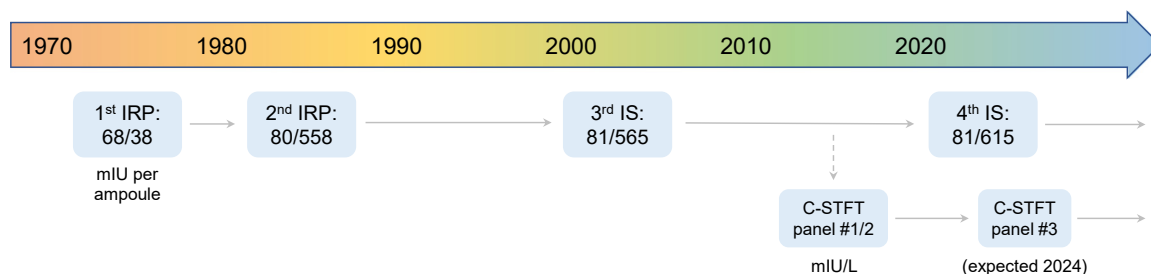


Figure 1: A timeline of TSH measurement standards, including traceability of measurement units through WHO preparations and C-STFT panels. IRP, International Reference Preparation; IS, International Standard; C-STFT, Committee for Standardisation of Thyroid Function Tests.

TSH reference materials to harmonize measurements (harmonisation panel)

To overcome the limitations of the WHO International Standard, the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) established the Working Group for Standardisation of Thyroid Function Tests (WG-STFT) in the late 2000s (later reclassified as an IFCC Committee; C-STFT) with a remit to develop resources and processes to improve the standardisation of measurements of thyroid hormones and TSH [19]. By 2015, the Committee had developed a putative reference panel for TSH aimed at improving harmonisation of TSH measurements. The panel included approximately 100 serum samples, spanning a range of TSH concentrations (including those indicative of hypo- and hyperthyroidism). The panel samples were value assigned in mIU/L (thereby traceable to the WHO reference material) via an “all procedure trimmed mean” (APTM) approach, in which samples were analysed by as many different immunoassay methods as possible, followed by statistical analysis to assign a TSH concentration (APTM) to each sample [20]. This panel of samples was developed with the intention that it ultimately serves as the primary reference material for calibration of TSH immunoassays. The use of serum samples maximises the likelihood of commutability of the panel with patient samples, and it was therefore anticipated that the use of the panel to recalibrate commercial immunoassays would lead to an improvement in harmonisation. Indeed, the concept was extensively demonstrated in a series of reports and publications from 2010 to 2017 [6, 21–23]. The approach is now described in ISO 21151:2020 *in vitro* diagnostic medical devices – requirements for International harmonisation protocols establishing metrological traceability of values assigned to calibrators and human samples [24].

The 1st and 2nd C-STFT panels were prepared and analysed in parallel. The samples comprising the 1st panel were value-assigned via the APTM approach using immunoassays traceable to the WHO standard, thus maintaining traceability of the mIU/L values. Samples in the “follow-up” 2nd panel were assigned TSH values alongside the 1st panel. Following successful proof of principle, the 2nd panel was made available for manufacturers to recalibrate their assays in 2016. Future replacements of the panel will adopt the same strategy, with analysis of the candidate panel alongside the existing panel, against which the TSH concentrations of the candidate panel samples will be assigned, thus continually maintaining traceability of the panel to the mIU defined by the WHO standards (Figure 1).

Present day TSH immunoassay standardisation

Both the C-STFT harmonization panel and the WHO International Standard are required to maintain the reference system for TSH and are available to laboratories and IVD manufacturers. To ensure continuity, the WHO, via the UK MHRA/NIBSC, established a replacement, the 4th WHO IS (81/615) for TSH in 2023 [11]. The evaluation of this material provided an opportunity to assess the current harmonisation of commercial TSH immunoassays, including the impact of the introduction of the C-STFT panel in 2016 and the potential impact of using the WHO IS as an assay calibrator.

The collaborative study to evaluate the candidate 4th IS for TSH was performed in collaboration with IFCC C-STFT and the CDC Clinical Standardization Programs, in conjunction with the establishment of a 3rd harmonization panel [11]. The 2nd C-STFT panel samples were analysed by 15 manufacturers of TSH immunoassays, alongside the existing 3rd IS, 81/565, and candidate 4th IS, 81/615. Assays were performed using ARCHITECT i2000, AccuraSeed®, AutoLumo A2000 Plus, Access 2, LIAISON® XL, LUMIPULSE® G1200, VIDAS®, i3000, CL-6000i, cobas® e801, VITROS® XT Integrated System, ADVIA Centaur® XP, MAGLUMI X3, HISCL-5000 and AIA-2000 analysers. The resulting data provided insight into the current status of TSH measurement harmonisation across the included assays, as per current assay calibrations. In addition, by expressing reported TSH estimates for C-STFT panel samples relative to the WHO preparations, it was also possible to determine what would be the impact of assay calibration against a WHO IS.

As described in the study report [11], the reported TSH concentrations of the C-STFT 2nd panel samples relative to assay calibrators, across the 15 immunoassays, were reasonably well harmonised. Reported mean TSH concentrations for each sample analysed by each method were expressed relative to the study median value (across all methods) for that sample, to assess the degree of bias evident for each sample by each method. A relative TSH concentration of 100 % would indicate exact agreement between the individual method result and the study median for a given sample, i.e. zero bias, whereas a value above or below would indicate a degree of positive or negative bias. For results reported relative to kit standards/calibrators, one method yielded negative median and mean % bias over all samples values of –11.6 and –11.2 % respectively, across fifty C-STFT samples, whilst another method yielded positive median and mean % bias

values of +12.1 and +9.5 % respectively. Otherwise, bias values for all other methods were within ± 6 %.

Overall, these results are suggestive that the current harmonisation of TSH measurements across different assay platforms is relatively good (Figure 2A), which suggests that several manufacturers have harmonised their assays following availability of the C-STFT sample panel. For example, concerted efforts have been made to harmonise TSH immunoassays from manufacturers in Japan using the IFCC C-STFT panel, led by various national medical and regulatory associations [25]. It is also expected that similar efforts have or will be made to improve TSH harmonisation in other geographical areas. This should, in turn, enable clinicians to interpret “borderline” TSH levels with increased confidence and assurance in the future.

The recent improvements in TSH harmonisation are exemplified by recalculating C-STFT panel sample TSH measurements relative to the 4th IS, 81/615 (i.e. effectively reverting to the assay calibration landscape prior to C-STFT establishment). When results, recalibrated to 81/615 for each sample analysed by each method, are then expressed relative to the study median, the maximum negative and positive biases observed across the 50 samples increase to $-21.6/-22.4$ % (median/mean bias) and $+24.4/+18.2$ %, whilst only five methods yield bias values within ± 6 %. Therefore, overall, immunoassay harmonisation would likely be significantly diminished by calibration using the WHO standard (Figure 2B).

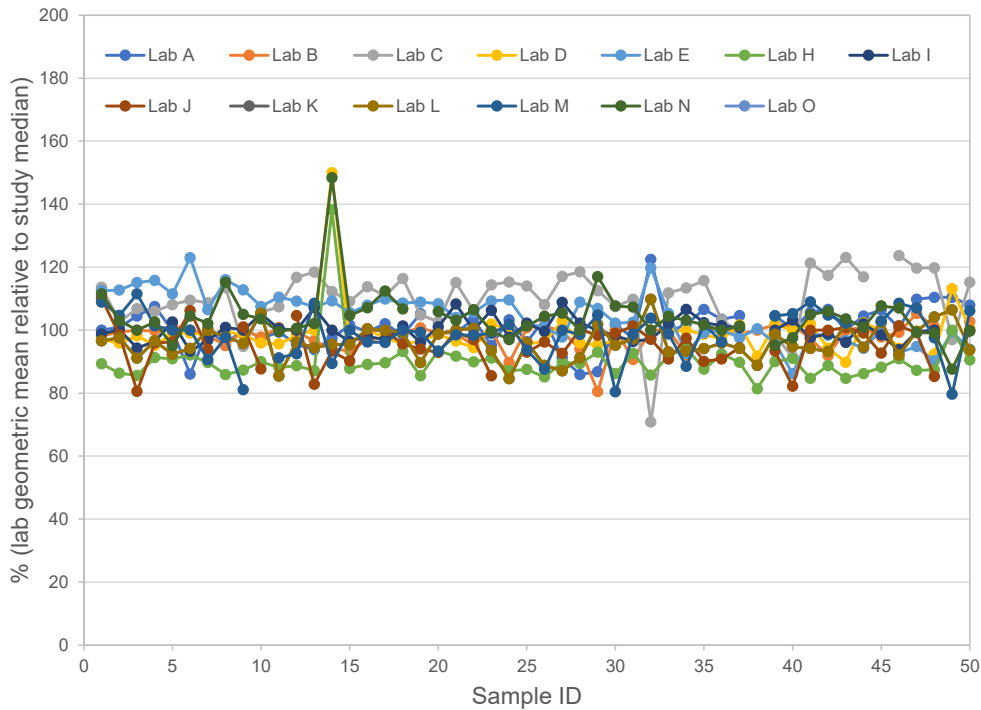
The future of TSH standardisation

Despite its known limitations, the use of pituitary derived TSH has remained the most feasible strategy for development of WHO standards for TSH over the previous 50 years. WHO standards are typically formulated with stabilisers and lyophilised to promote long-term stability and facilitate low-cost global shipping at ambient temperature. They are ideally prepared in sufficiently large batches to last for 10+ years and contain a sufficiently high concentration of active ingredient to ensure suitability for use as a calibrator at the upper end of assay working ranges. For example, the 4th IS for TSH (81/615) contains 11.7 mIU of TSH activity per ampoule, which can be reconstituted in as little as 1 mL diluent to achieve a concentration of 11.7 mIU/mL, which is approximately 2,500-fold greater the upper end of the “normal” TSH range in human serum (4–5 mIU/L). Sourcing sufficiently large volumes of serum, containing sufficiently high levels of TSH, to develop a WHO standard would be very challenging, hence the continued use, to date, of more readily-available pituitary TSH preparations.

The 4th IS, 81/615, was established in October 2023 in sufficient numbers to last 15+ years. However, based on the findings of the study to establish this material, discussed herein, it is not advised that this material be used as a standalone calibrator for TSH immunoassays, as would typically be recommended for a WHO IS, without due consideration of its commutability. The IFCC C-STFT panel is an invaluable resource which has led to a significant improvement in TSH immunoassay harmonisation in recent years and theoretically it alone could be used as a standalone calibrator of TSH immunoassays in the future. However, due to the aforementioned difficulty sourcing significant quantities of serum from multiple donors containing varied TSH concentrations from across the measurement range, it is not feasible to produce the panel in sufficiently large numbers for use as a single global calibrator. Discontinuation of the WHO IS for TSH and signposting to the IFCC C-STFT panel would very likely lead to an impractically and unsustainably high demand for the latter. Therefore, the co-existence of the WHO IS and IFCC C-STFT for TSH is likely to remain the case for the foreseeable future. The WHO IS may yet find use as an assay calibrator, where acceptable commutability is concluded, and will also prove useful as a continuous reference for assay monitoring during its lifecycle (e.g. before and after minor modifications which are not expected to affect antibody recognition). Meanwhile, the C-STFT panel can be used more sparingly to set assay harmonisation, if required, and subsequently ensure continued harmonisation in the event that significant changes are made which may affect antibody binding.

Whilst there are upwards of 15 major TSH immunoassay manufacturers established within the C-STFT network, who are therefore aware of the availability and utility of the C-STFT panel, any new or emerging manufacturers of TSH assays may naturally seek to use the WHO IS to calibrate their assay, perhaps on advice from a national or regional regulatory authority. Careful and cautionary messaging therefore needs to be provided alongside the WHO IS regarding its potential lack of commutability in a developing TSH immunoassay. The study to evaluate the 4th IS, 81/615, indicated that it does in fact have good commutability with patient samples in some immunoassays, and it is therefore possible that this would also be the case for new and future assays. However, manufacturers of new TSH assays would be advised to conduct a rigorous commutability study to establish that the WHO IS would be an appropriate calibrator for their assay and, regardless of the outcome, it is imperative that they should also be made aware of the existence of the C-STFT panel and be provided with access to this panel to maintain assay harmonisation.

A



B

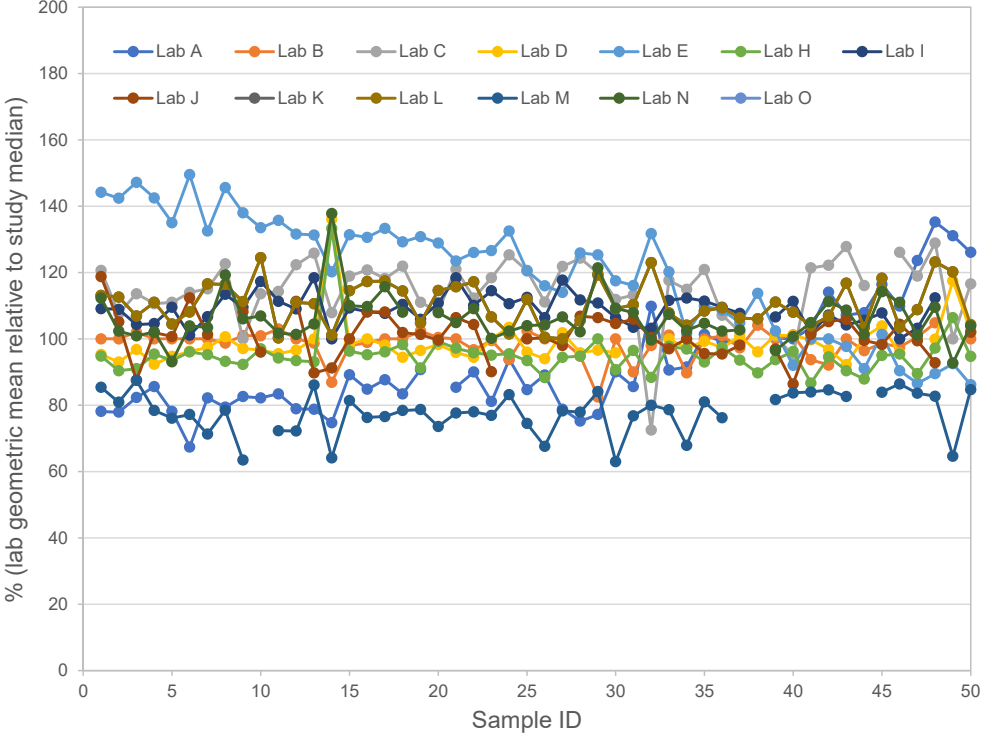


Figure 2: Lab GM values for each sample in each assay expressed relative (as a %) to the study median for the same sample. % values are derived from results reported relative to assay/kit calibrators (A) and after re-calibration against the 4th IS, 81/615 (B).

Acknowledgments: The authors acknowledge the contributions to the establishment of the described TSH reference materials of the IFCC C-STFT, its members Dr. V. Raverot, Dr. S. L. Andersen and Prof. A. Hishinuma, and the collaborative study participants.

Research ethics: Not applicable.

Informed consent: Not applicable.

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention/the Agency for Toxic Substances and Disease Registry. Use of trade names is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention, the Public Health Service, and the US Department of Health and Human Services.

Author contributions: The authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: The authors state no conflict of interest.

Research funding: None declared.

Data availability: The raw data can be obtained on request from the corresponding author.

References

- Muller R, Liu Y-Y, Brent GA. Thyroid hormone regulation of metabolism. *Physiol Rev* 2014;94:355–82.
- Leso V, Vetrani I, De Cicco L, Cardelia A, Fontana L, Buonocore G, et al. The impact of thyroid disease on the working life of patients: a systematic review. *Int J Environ Res Publ Health* 2020;17:4295.
- Jonklaas J, Bianco AC, Bauer AJ, Burman KD, Cappola AR, Celi FS, et al. Guidelines for the treatment of hypothyroidism: prepared by the American Thyroid Association task force on thyroid hormone replacement. *Thyroid* 2014;24:1670–751.
- Sundares V, Brito JP, Want Z, Prokop LJ, Stan MN, Murad MH, et al. Comparative effectiveness of therapies of Graves' hyperthyroidism: a systematic review and network meta-analysis. *J Clin Endocrinol Metab* 2013;98:3671–7.
- Van Uytanghe K, Ehrenkranz J, Halsall DM, Hoff K, Loh TP, Spencer CA, Kohrle J, et al. Thyroid-stimulating hormone and thyroid hormones (triiodothyronine and thyroxine): an American Thyroid-Association-Commissioned review of current clinical and laboratory status. *Thyroid* 2023;33:1013–28.
- Thienpont LM, Van Uytanghe K, De Grande LAC, Reynders D, Das B, Faix JD, et al. Harmonization of serum thyroid-stimulating hormone measurements paves the way for the adoption of a more uniform reference interval. *Clin Chem* 2017;63:1248–60.
- Mirjanic-Azaric B, Jerin A, Radic Z. Thyroid stimulating hormone values of clinical decisions of hypothyroidism measurement by three different automated immunoassays. *Scand J Clin Lab Invest* 2020;80:151–5.
- Szkudlinksi MW, Fremont V, Ronin C, Weintraub BD. Thyroid-stimulating hormone and thyroid-stimulating hormone receptor structure-function relationships. *Physiol Rev* 2002;82:473–502.
- Gaines Das RE, Bristow AF. The second International Reference Preparation of thyroid-stimulating hormone, human, for immunoassay: calibration by bioassay and immunoassay in an international collaborative study. *J Endocrinol* 1985;104:367–79.
- Rafferty B, Gaines Das RE. Report of an international collaborative study of the proposed 3rd International Standard for thyroid-stimulating hormone, human, for immunoassay. WHO/BS/03.1975 2003.
- Cowper B, Partridge K, Rigsby P, Lyle A, Vesper H, Van Uytanghe K, et al. 4th WHO International Standard for thyroid-stimulating hormone (TSH), human, pituitary. WHO/BS/2023.2454 2023.
- Thienpont LM, Uytanghe K, Beastall G, Faix JD, Ieiri T, Miller GW, et al. Report of the IFCC working Group for standardisation of thyroid function tests; part 1: thyroid-stimulating hormone. *Clin Chem* 2010;56:902–11.
- Vesper H, Miller GW, Myers GL. Reference materials and commutability. *Clin Biochem Rev* 2007;28:139–47.
- Vesper HW, Myers GL, Miller GW. Current practices and challenges in the standardisation and harmonization of clinical laboratory tests. *Am J Clin Nutr* 2016;104:907–12.
- Faix JD, Miller GW. Progress in standardizing and harmonizing thyroid function tests. *Am J Clin Nutr* 2016;104:913–7.
- Constant RB, Weintraub BD. Differences in the metabolic clearance of pituitary and serum thyrotropin (TSH) derived from euthyroid and hypothyroid rats: effects of chemical deglycosylation of pituitary TSH. *Endocrinology* 1986;119:2720–7.
- Trojan J, Theodoropoulou M, Usadel KH, Stalla GK, Schaaf L. Modulation of human thyrotropin oligosaccharide structures – enhanced proportion of sialylated and terminally galactosylated serum thyrotropin isoforms in subclinical and overt primary hypothyroidism. *J Endocrinol* 1998;158:359–65.
- Donadio S, Pascual A, Thijssen JHH, Ronin C. Feasibility study of new calibrators for thyroid-stimulating hormone (TSH) immunoprocures based on remodeling of recombinant TSH to mimic glycoforms circulating in patients with thyroid disorders. *Clin Chem* 2006;52:286–97.
- International Federation for Clinical Chemistry and Laboratory Medicine (IFCC) Committee for Standardisation of Thyroid Function Tests (C-STFT). Available from: <https://ifcc-cstft.org/>, <https://ifcc.org/ifcc-scientific-division/sd-committees/c-stft/>.
- Stockl D, Van Uytanghe K, Van Aelst S, Thienpont LM. A statistical basis for harmonization of thyroid-stimulating hormone immunoassays using a robust factor analysis model. *Clin Chem Lab Med* 2014;52:965–72.
- Miller GW, Myers GL, Lou Gantzer W, Kahn SE, Schonbrunner ER, Thienpont LM, et al. Roadmap for harmonization of clinical laboratory procedures. *Clin Chem* 2011;57:1108–17.
- Van Houcke SK, Thienpont LM. “Good samples make good assays” – the problem of sourcing clinical samples for a standardization project. *Clin Chem Med Lab* 2013;51:967–72.
- Van Houcke SK, Van Aelst S, Van Uytanghe K, Thienpont LM. Harmonisation of immunoassays to the all-procedure trimmed mean – proof of concept by use of data from the insulin standardization project. *Clin Chem Lab Med* 2013;51:e103–5.
- ISO 21151 International Organisation for Standardisation. In vitro diagnostic medical devices – Requirements for International harmonisation protocols establishing metrological traceability of values assigned to calibrators and human samples. Geneva: ISO; 2020. ISO 21151:2020.
- IFCC 2016 General Conference notes. 20181109_-_minutes_c-stft_meeting_-_budapest.pdf (ifcc-cstft.org).