

RESEARCH

How low can we (reliably) go? A method comparison of thyroid-stimulating hormone assays with a focus on low concentrations

Stan R Ursem^{1,2}, Anita Boelen^{1,2,3}, Jacquelin J Hillebrand^{1,2} and Wendy P J den Elzen^{4,5} and Annemieke C Heijboer^{1,2,3,6}

¹Department of Laboratory Medicine, Endocrine Laboratory, Amsterdam UMC Location University of Amsterdam, Meibergdreef, Amsterdam, The Netherlands

²Amsterdam Gastroenterology, Endocrinology & Metabolism, Amsterdam, The Netherlands

³Amsterdam Reproduction & Development, Amsterdam, The Netherlands

⁴Department of Laboratory Medicine, Laboratory Specialized Diagnostics & Research, Amsterdam UMC location University of Amsterdam, Meibergdreef, Amsterdam, The Netherlands

⁵Amsterdam Public Health, Amsterdam, The Netherlands

⁶Department of Laboratory Medicine, Endocrine Laboratory, Amsterdam UMC Location Vrije Universiteit Amsterdam, Boelelaan, Amsterdam, The Netherlands

Correspondence should be addressed to A C Heijboer: a.heijboer@amsterdamumc.nl

Abstract

Objective: International guidelines concerning subclinical hyperthyroidism and thyroid cancer advice absolute cut-off values for aiding clinical decisions in the low range of thyroid-stimulating hormone (TSH) concentrations. As TSH assays are known to be poorly standardized in the normal to high range, we performed a TSH assay method comparison focusing on the low range.

Methods: Sixty samples, selected to cover a wide range of TSH concentrations (<0.01 to 120 mIU/L) with oversampling in the lower range (<0.4 mIU/L), were used for the method comparison between three TSH immunoassays (Cobas, Alinity and Atellica). In addition, 20 samples were used to assess the coefficient of variation from duplicate measurements in these three methods.

Results: The TSH immunoassays showed standardization differences with a bias of 7–16% for the total range and 1–14% for the low range. This could lead to a different classification of 1.5% of all measured TSH concentrations <0.40 mIU/L measured in our laboratory over the last 6 months, regarding the clinically important cut-off value of TSH = 0.1 mIU/L. As the imprecision of the immunoassays varied from 1.6–5.5%, this could lead to a similar reclassification as the bias between immunoassays.

Conclusions: We established the standardization differences of frequently used TSH assays for the total and low concentration ranges. Based on the proportional bias and the imprecision, this effect seems to have limited clinical consequences for the low TSH concentration range. Nevertheless, as guidelines mention absolute TSH values to guide clinical decision-making, caution must be applied when interpreting values close to these cut-offs.

Key Words

- ▶ TSH
- ▶ thyroid-stimulating hormone
- ▶ method comparison
- ▶ subclinical hyperthyroidism

Introduction

Thyroid disorders are common. When a thyroid disorder is suspected, the standard procedure is measuring the serum or plasma thyroid-stimulating hormone (TSH). Accurate TSH measurements are important for timely diagnosis since symptoms can be subtle and also for monitoring therapy in hypo- or hyperthyroidism and assessing suppressive therapy in patients with thyroid cancer. Unfortunately, TSH assays are poorly standardized to date.

The IFCC Working Group on Standardization of Thyroid Function Tests established in 2010 standardization differences up to 39% between TSH assays and recommended further research into the concentration range close to the limit of quantification (1). The working group deemed the availability of TSH reference material in the mid- to short-term technically unlikely and now focusses on statistical harmonization (2). Currently, the grand majority of TSH assays are not yet harmonized, which means that absolute concentrations reported by laboratories using different assays are not comparable (3, 4). This could impact diagnosis, prognosis and management of disease, as has recently been shown for subclinical hypothyroidism (5).

The lower range of the TSH assay is important because in this pathophysiological range, often clinical decisions are made, and this is also where assay performance decreases (6). Several international guidelines mention absolute TSH concentrations for clinical decision-making such as in subclinical hyperthyroidism, thyroid cancer, pregnancy and assisted reproduction (7, 8, 9, 10, 11). However, to our best knowledge, research is lacking concerning the comparability of results in this low TSH range.

Therefore, we performed a method comparison for commonly used TSH immunoassays, with oversampling in the lower concentration range (<0.4 mIU/L) and established the imprecision in this low concentration range. Based on the standardization differences, we investigated the theoretical impact of the absolute cut-offs mentioned in the guidelines.

Material and methods

Serum samples

Leftover samples were collected from routine measurements in our tertiary center (Amsterdam UMC,

University of Amsterdam, Amsterdam, The Netherlands). Samples were selected to cover a wide range of TSH concentrations, with oversampling in the lower range, based on the measurements performed using the Cobas® assay (Roche Diagnostics) and were anonymized. Lithium heparin plasma samples were used for the analysis. Aliquots were stored at -20°C and measured for the current study within 2 months. Sixty samples were used for the method comparison and 20 samples were used for assessing the coefficient of variation.

Immunoassays

TSH was measured in one run using the automated electrochemiluminescence immunoassays of the following platforms: Cobas® (Roche Diagnostics) Elecsys TSH REF 08429324 190, Alinity® (Abbott Laboratories) TSH reagent kit G71292R02, and Atellica® (Siemens Healthcare Diagnostics) TSH3-UL REF 10995703. The Cobas and Alinity assays are standardized against the same WHO Second International Standard for human TSH (IRP 80/558) and the Atellica assay is standardized against the WHO Third International Standard for human TSH (IRP 81/565). The samples were included based on the initial measurement using the Cobas®. For the method comparison, all the samples were remeasured on the Cobas® platform in one run.

Statistical analysis

Statistical analysis was performed using Passing-Bablok regression to minimize the effect of outliers possibly caused by assay-specific interference. In addition, Spearman's rank correlation coefficient was used. The coefficient of variation from duplicate measurements was used to assess intra-assay variation using the within-subject standard deviation method. Statistical analysis was performed using MedCalc® version 18.5. *P*-values <0.05 were considered to reflect statistical significance. In addition, we estimated the effect of the greatest standardization difference on measured values around the absolute cut-off of TSH=0.1 mIU/L using the slope and intercept of the Passing-Bablok regressions for the low concentrations. We calculated the percentage that would cross the absolute cut-off when it would theoretically be measured by another assay. We used all reported TSH results in our academic hospital for 6 months (1 March 2022 until 1 September 2022).

Results

The distribution of the samples, as measured using the Cobas assay, is depicted in Supplementary Fig. 1 (see section on [supplementary materials](#) given at the end of this article). As the lower concentrations were oversampled, the median TSH concentration used for the method comparison was 0.24 mIU/L (interquartile range: 0.04–3.21 mIU/L; $n = 60$). For the duplicate measurements, the median was 0.25 mIU/L (interquartile range: 0.06–2.66 mIU/L; $n = 20$).

Passing–Bablok regression and correlation coefficients between the three methods are depicted in Fig. 1. The slope of the Passing–Bablok regressions indicates the standardization differences, as the y -intercept was negligible. The bias ranged from 7% to 16% in the complete measured range. For the samples with TSH <0.40 mIU/L, the bias differed between 1% and 14%, as is depicted in Fig. 1.

The imprecision of the assays increased as the concentration lowered and is shown in Table 1. Differences were seen between assays and concentrations, with the largest mean variation between duplicate measurements being 5.5% for the Atellica assay in the range [TSH] <0.40 mIU/L.

The Atellica and Cobas assay exhibited the largest bias in the lower concentration range (14%). Based on this bias, 1.5% of all TSH measurements with results <0.4 mIU/L (15 of the 998 measurements) in our

laboratory, using the Cobas assay, would cross the absolute cut-off and would be <0.10 mIU/L instead of ≥ 0.10 mIU/L, when measured with the Atellica assay.

In addition, we compared the initial TSH measurement on the Cobas, which we used to include the samples, to the remeasurements performed on the Cobas, but in one run. The Passing–Bablok regression was $\text{TSH}_{\text{Cobas, remeasurement}} = 0.935 \times \text{TSH}_{\text{Cobas, initial}} + 0.0003$; $\rho = 0.998$; $P < 0.0001$ for all concentrations, and for TSH <0.4 mIU/L: $\text{TSH}_{\text{Cobas, remeasurement}} = 0.924 \times \text{TSH}_{\text{Cobas, initial}} + 0.0013$; $\rho = 0.991$; $P < 0.0001$. This remeasurement would also lead to 1.5% of all TSH measurements <0.4 mIU/L to cross the absolute cut-off. So, the imprecision of one TSH assay causes as many patients crossing the cut-off as using a different TSH assay.

Discussion

In the present study, we performed a method comparison between three frequently used TSH immunoassays focusing on the low concentration range, as this range is important for clinical decision-making in case of (subclinical) hyperthyroidism and thyroid cancer.

The studied TSH immunoassays showed standardization differences with a bias between 7 and 16%. When analysing the standardization differences for the low range, the bias differed between 1 and 14%. Interestingly, the assays which were standardized

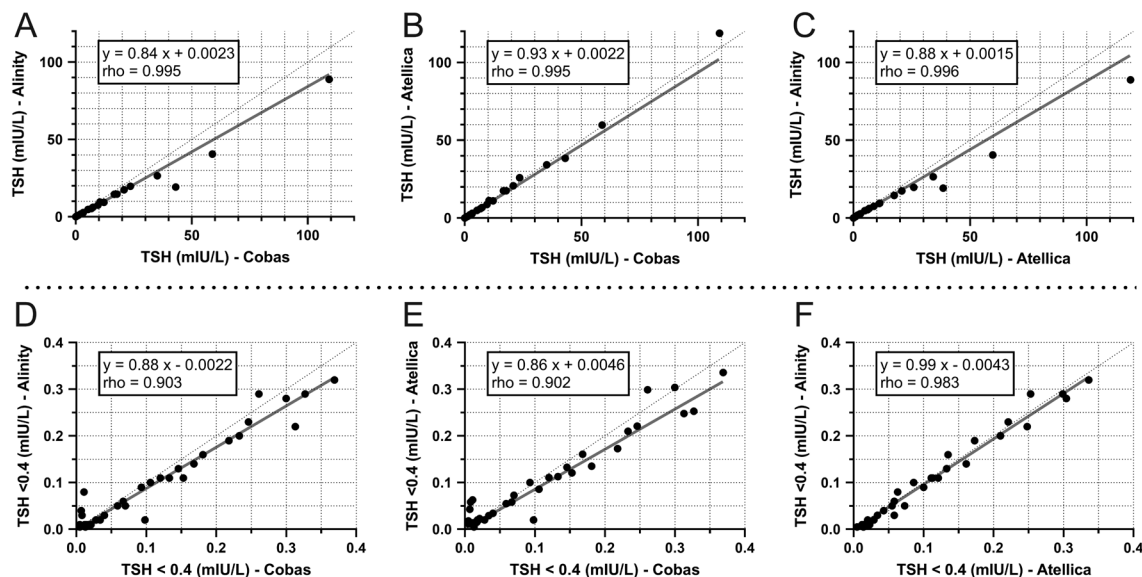


Figure 1

Passing–Bablok regression analysis for TSH for the total sample set ($n = 60$) (A, B, C) and for TSH <0.4 mIU/L ($n = 36$) (D, E, F). For all $P < 0.0001$.

Table 1 Coefficient of variation (CV) for duplicate TSH measurements.

Assay	Range (mIU/L)	CV (%)	n
Cobas	<0.4	3.0	12
	≥0.4	1.6	8
Alinity	<0.4	4.5	13
	≥0.4	3.5	7
Atellica	<0.4	5.5	13
	≥0.4	3.7	7

against the same WHO standard varied more from each other than compared to the assay with a different WHO standard. The imprecision differed between 1.6 and 5.5% and increased in the lower range of the assay.

Several guidelines use absolute cut-off TSH values in the low range for clinical decision-making (7, 8, 9, 10, 11). In the American and European guidelines for thyroid cancer and in the European guideline for subclinical hyperthyroidism an absolute TSH value of 0.10 mIU/L is mentioned as an important value above or below which recommendations are given concerning starting or adjusting medication and performing diagnostic testing (7, 8, 9). In order to establish whether the observed standardization differences translate into clinically relevant differences, we looked at all TSH results reported by our laboratory in the last 6 months. We calculated that 1.5% of all TSH measurements <0.4 mIU/L would theoretically lead to a different clinical outcome, concerning this absolute cut-off of TSH=0.1 mIU/L. Naturally, not all the measurements around 0.1 mIU/L concerned patients with subclinical hyperthyroidism or thyroid cancer.

The imprecision using the Cobas assay had the same repercussions on the different classification of samples around TSH=0.1 mIU/L as the greatest bias between different assays. For the Alinity and Atellica assays with a slightly higher imprecision for TSH <0.4 mIU/L, the impact on the clinical management difference would be even higher than the immunoassay bias. Although the standardization differences for the total concentration range are evident, we consider these standardization differences too small to be clinically relevant in the low concentration range. Nonetheless, it is important to be cautious when interpreting concentrations close to a clinical cut-off value, especially if the methods are poorly standardized and the thresholds are not assay specific.

The widespread use of arbitrary dichotomous cut-offs based on continuous data has limitations. Considering

imprecision and bias, combined in total laboratory error or measurement uncertainty, are crucial when establishing clinical cut-offs. Total laboratory error or measurement uncertainty encompasses variability throughout the entire testing process, including pre-analytical, analytical and post-analytical factors. Additionally, within-subject biological variation contributes to variability in results and should also be considered when interpreting laboratory results. Both the total laboratory (analytical) variation and within-subject biological variation (estimated at 17.7% for TSH, source: <https://biologicalvariation.eu/search?query=TSH>) contribute to the critical difference, also called the reference change value, the smallest difference between sequential laboratory results that is associated with a real change in the patient. By accounting for these factors, a better understanding of the limitations associated with arbitrary dichotomous cut-offs is achieved, enabling more informed and accurate interpretations of assay results.

Over time, a trend towards smaller standardization differences for TSH assays is seen in multiple studies (1, 12, 13, 14). Comparing our results to other method comparisons is generally complicated by the limited mention of the assay specifics. For some method comparisons, it is difficult to analyse results side by side because of different statistical methods, such as linear regressions instead of Passing-Bablok regressions or because of anonymization of the used assays (1, 15, 16). The imprecision is known to increase with lower concentration, which we also observed (2, 16). The clinical management difference we observed is smaller than in previous studies assessing the effect on clinical and subclinical hypothyroidism. This may be explained by the proportional difference in low vs high concentrations and improvements in bias between assays over time (5, 15, 17).

Taken together, we performed a method comparison of commonly used TSH assays and focused on the lower range of TSH concentrations. The bias in the lower concentration range showed a maximum of 14% between methods. Based on this proportional bias and the imprecision we estimated the theoretical effect on measured TSH concentrations in the last 6 months in our lab. This effect seems to have limited clinical consequences for the low TSH concentration range. Nevertheless, as guidelines use absolute TSH values to guide clinical decision-making, caution must be applied with TSH values close to these cut-offs.

Supplementary materials

This is linked to the online version of the paper at <https://doi.org/10.1530/ETJ-23-0123>.

Declaration of interest

Authors state no conflict of interest.

Funding

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

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Author contribution statement

All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Acknowledgements

We would like to thank the technicians of the laboratory for endocrinology, Amsterdam Medical Centers, Amsterdam, The Netherlands.

References

- 1 Thienpont LM, Van Uytvanghe K, Beastall G, Faix JD, Ieiri T, Miller WG, Nelson JC, Ronin C, Ross HA, Thijsen JH, *et al.* Report of the IFCC working group for standardization of thyroid function tests; part 1: thyroid-stimulating hormone. *Clinical Chemistry* 2010 **56** 902–911. (<https://doi.org/10.1373/clinchem.2009.140178>)
- 2 Thienpont LM, Van Uytvanghe K, Van Houcke S, Das B, Faix JD, MacKenzie F, Quinn FA, Rottmann M, Van den Bruel A & IFCC Committee for Standardization of Thyroid Function Tests (C-STFT). A progress report of the IFCC committee for standardization of thyroid function tests. *European Thyroid Journal* 2014 **3** 109–116. (<https://doi.org/10.1159/000358270>)
- 3 Thienpont LM, Van Uytvanghe K, De Grande LAC, Reynders D, Das B, Faix JD, MacKenzie F, Decallonne B, Hishinuma A, Lapauw B, *et al.* Harmonization of serum thyroid-stimulating hormone measurements paves the way for the adoption of a more uniform reference interval. *Clinical Chemistry* 2017 **63** 1248–1260. (<https://doi.org/10.1373/clinchem.2016.269456>)
- 4 Vesper HW, Van Uytvanghe K, Hishinuma A, Raverot V, Patru MM, Danilenko U, van Herwaarden AE & Shimizu E. Implementing reference systems for thyroid function tests – a collaborative effort. *Clinica Chimica Acta; International Journal of Clinical Chemistry* 2021 **519** 183–186. (<https://doi.org/10.1016/j.cca.2021.04.019>)
- 5 Kalaria T, Fenn J, Sanders A, Ford C & Gama R. Clinical concordance assessment should be an integral component of laboratory method comparison studies: a regression transference of routine clinical data approach. *Clinical Biochemistry* 2022 **103** 25–28. (<https://doi.org/10.1016/j.clinbiochem.2022.02.008>)
- 6 Beckett G & MacKenzie F. Thyroid guidelines - are thyroid-stimulating hormone assays fit for purpose? *Annals of Clinical Biochemistry* 2007 **44** 203–208. (<https://doi.org/10.1258/000456307780480945>)
- 7 Biondi B, Bartalena L, Cooper DS, Hegedüs L, Laurberg P & Kahaly GJ. The 2015 European Thyroid Association guidelines on diagnosis and treatment of endogenous subclinical hyperthyroidism. *European Thyroid Journal* 2015 **4** 149–163. (<https://doi.org/10.1159/000438750>)
- 8 Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, Pacini F, Randolph GW, Sawka AM, Schlumberger M, *et al.* 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid* 2016 **26** 1–133. (<https://doi.org/10.1089/thy.2015.0020>)
- 9 Filetti S, Durante C, Hartl D, Leboulleux S, Locati LD, Newbold K, Papotti MG, Berruti A & ESMO Guidelines Committee. Thyroid cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* 2019 **30** 1856–1883. (<https://doi.org/10.1093/annonc/mdz400>)
- 10 Alexander EK, Pearce EN, Brent GA, Brown RS, Chen H, Dosiou C, Grobman WA, Laurberg P, Lazarus JH, Mandel SJ, *et al.* 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid* 2017 **27** 315–389. (<https://doi.org/10.1089/thy.2016.0457>)
- 11 Poppe K, Bisschop P, Fugazzola L, Minziori G, Unuane D & Weghofer A. 2021 European Thyroid Association guideline on thyroid disorders prior to and during assisted reproduction. *European Thyroid Journal* 2021 **9** 281–295. (<https://doi.org/10.1159/000512790>)
- 12 Hendriks HA, Kortlandt W & Verweij WM. Analytical performance comparison of five new generation immunoassay analyzers. *Ned Tijdschr voor Klin Chemie* 2000 **25** 170–177.
- 13 Rawlins ML & Roberts WL. Performance characteristics of six third-generation assays for thyroid-stimulating hormone. *Clinical Chemistry* 2004 **50** 2338–2344. (<https://doi.org/10.1373/clinchem.2004.039156>)
- 14 da Silva VA, de Almeida RJ, Cavalcante MP, Pereira Junior LA, Reis FM, Pereira MF, Kasamatsu TS & Camacho CP. Two thyroid stimulating hormone assays correlated in clinical practice show disagreement in subclinical hypothyroidism patients. *Clinical Biochemistry* 2018 **53** 13–18. (<https://doi.org/10.1016/j.clinbiochem.2017.12.005>)
- 15 Kalaria TR, Sanders A, Ford C, Buch H, Fenn JS, Ashby HL, Mohammed P & Gama RM. Biochemical assessment of adequate levothyroxine replacement in primary hypothyroidism differs with different TSH assays: potential clinical implications. *Journal of Clinical Pathology* 2022 **75** 379–382. (<https://doi.org/10.1136/jclinpath-2020-207316>)
- 16 Clerico A, Ripoli A, Fortunato A, Alfano A, Carrozza C, Corrales M, Dittadi R, Gessoni G, Migliardi M, Rizzardi S, *et al.* Harmonization protocols for TSH immunoassays: a multicenter study in Italy. *Clinical Chemistry and Laboratory Medicine* 2017 **55** 1722–1733. (<https://doi.org/10.1515/cclm-2016-0899>)
- 17 Kalaria T, Sanders A, Fenn J, Ashby HL, Mohammed P, Buch HN, Ford C & Gama R. The diagnosis and management of subclinical hypothyroidism is assay-dependent – implications for clinical practice. *Clinical Endocrinology* 2021 **94** 1012–1016. (<https://doi.org/10.1111/cen.14423>)

Received 21 June 2023

Accepted 8 August 2023

Available online 8 August 2023

Version of Record published 14 September 2023

