

Reference range for thyrotropin

Post hoc assessment

R. Larisch¹; A. Giacobino¹; W. Eckl¹; H.-G. Wahl²; J. E. M. Midgley³; R. Hoermann¹

¹Klinik für Nuklearmedizin, Klinikum Lüdenscheid, Germany; ²Institut für Laboratoriumsmedizin, Klinikum Lüdenscheid, Germany; ³North Lakes Clinical, Ilkley, UK

Keywords

TSH, reference range, thyroid hormones

Summary

Setting the reference range for thyrotropin (TSH) remains a matter of ongoing controversy. **Patients, methods:** We used an indirect method to determine the TSH reference range post hoc in a large sample. A total of 399 well characterised subjects showing no evidence of thyroid dysfunction were selected for definition of the TSH reference limits according to the method of Katayev et al.. To this end, the cumulative frequency was plotted against the individual logarithmic TSH values. Reference limits were calculated by extrapolating the middle linear part of the regression line to obtain the cut-offs for the 95% confidence interval. We also examined biological variation in a sample of 65 subjects with repeat measurements to establish reference change values (RCVs). **Results:**

Based on these, the reference interval obtained by the novel technique was in close agreement with the conventionally established limits, but differed significantly from earlier recommendations. **Discussion:** Following unverified recommendations could result in a portion of patients with subclinical thyroid dysfunctions being missed, an important consideration in a setting with a high prevalence of thyroid autonomy. **Conclusion:** Indirect post hoc verification of reference intervals from a large retrospective sample is a modern approach that gives plausible results. The method seems particularly useful to assess the adequacy and performance of reference limits reported or established by others in a particular setting. The present data should encourage re-evaluation of reference systems on a broader scale.

Schlüsselwörter

TSH, Referenzbereich, Schilddrüsenhormone

Zusammenfassung

Die vorliegende Arbeit befasst sich mit dem Referenzbereich des Thyreotropins (TSH). **Patienten, Methoden:** Mit der indirekten Methode von Katayev et al. wurde der Referenzbereich von TSH post hoc in einer Gruppe von insgesamt 399 gut untersuchten Personen ohne klinisch fassbare Zeichen einer Schilddrüsensyndrom erhoben. Dafür wurde die kumulierte Häufigkeit gegen die logarithmierten TSH-Werte aufgetragen. Der mittlere, lineare Bereich, der dem 95%-Konfidenzintervall der Regressionsgerade entsprach, war der Referenzbereich. Zudem wurde die biologische Variabilität der Referenzwerte in einer Gruppe von 65 Personen mit Mehrfachmessungen bestimmt. **Ergebnisse:** Der so bestimmte Referenzbereich stimmte gut mit aktuellen, konventionell erhobenen Referenzdaten überein. Allerdings ergaben sich signifikante Unterschiede zu früheren Empfehlungen. **Diskussion, Schlussfolgerung:** Die indirekte Post-hoc-Bestimmung des TSH-Referenzbereichs aus einer retrospektiven Stichprobe ist ein moderner Ansatz, der plausible Ergebnisse liefert. Die Methode erscheint gut geeignet, um vorgegebene Referenzwerte zu verifizieren. Die vorliegenden Daten sollten Kliniker motivieren, auf diese Weise ihre eigenen Referenzbereiche festzulegen.

Correspondence to:

Prof. Dr. Rolf Larisch
Department of Nuclear Medicine, Klinikum Luedenscheid, Paulmannshoeher Str 14
58515 Luedenscheid, Germany
E-mail: rolf.larisch@klinikum-luedenscheid.de

Thyrotropin-Referenzbereich Post-hoc-Festsetzung

Nuklearmedizin 2015;54:-
<http://dx.doi.org/10.3413/Nukmed-0671-14-06>
received: June 6, 2014
accepted in revised form: December 17, 2014
epub ahead of print: January 8, 2015

Modern thyroid laboratory evaluation of thyroid function relies heavily on the pituitary hormone thyrotropin (TSH), and the more subtle thyroid dysfunctions termed subclinical hypothyroidism or hyperthyroidism are exclusively based on TSH measurement (3). With the advent of third generation assays techniques have reached

a high standard of assay sensitivity and methodological reliability.

However, a still unresolved controversy has surrounded the determination of the reference range of TSH, particularly its upper limit (4, 5, 13, 19, 22, 27).

A lack of standardisation and harmonisation among the various commercially available methods has added to the controversy (24, 25). This makes it mandatory for each institution to establish its own reference ranges. The American Clinical and Laboratory Standards Institute has published guidelines how to proceed with the

task (28). Following a strict procedure, reference ranges are conventionally established in a well characterised and selected population that is considered disease free (28). The burden rests with the individual institution, particularly since large population studies have arrived at considerably varying reference ranges, even in disease free populations (2, 15, 26). For example, while the US NHANES study reported a reference range of 0.45–4.12 mIU/l, a study by Voelzke et al. found a reference range of 0.25–2.12 mIU/l in a German population (15, 26). Importantly, using either reference system would result in the misclassification of a substantial portion of subjects (26). This emphasises the magnitude of the problem and the requirement of adopting population-specific reference limits.

One of the problems in assessing TSH values is that data are not normally distributed in a given group of reference subjects. In order to overcome both this difficulty and problems with hidden pathologies the use of computerised indirect methods pioneered by Hoffmann and Katayev has been proposed (14, 17). Although several studies have validated the novel approach, the indirect methods have not been widely adopted for the routine purpose of validating reference ranges (7).

In this study, we aimed at verifying the TSH reference range in a large retrospective sample using an indirect method according to Katayev et al. (17).

- We have compared the indirect technique with the pre-established conventional reference range.
- Thereafter, we compared the resulting TSH reference range with recommendations of the manufacturer and with previously established reference intervals for the local population from other methods (18).

Patients and methods

Patients

Samples were collected as part of a prospective study conducted by the Department of Nuclear Medicine at Klinikum Luedenscheid (www.ClinicalTrials.gov, NCT 01969552). This part of the study was completed from spring to fall 2013, and in-

cluded 1258 adult subjects, 399 of which had no objective evidence for the presence of any thyroid dysfunction and were selected for further analysis in this study. We had excluded 428 patients after thyroid surgery or radioiodine treatment, 245 for taking thyroid medication, 110 for TPO antibody positivity, 49 for manifest thyroid dysfunction and 27 for sonographically and scintigraphically detected large hyperfunctioning nodules.

The characteristics of the study group are given in Results. All patients were seen on an ambulatory basis and had been referred to this specialised thyroid unit for the purpose of exclusion or treatment of various thyroid disorders. Non-laboratory assessment of all subjects included a detailed history and thyroid-related physical examination, routine performance of thyroid imaging by ultrasound, supplemented by scintigraphy in case a hyperfunctioning thyroid nodule was suspected or to be excluded.

Luedenscheid is located in a former area of mild iodine deficiency. Therefore, the prevalence of toxic adenomas, non-functioning thyroid nodules and multinodular goitre is high. The study was approved by the local Ethics Committee and patients gave written informed consent to participate.

Laboratory methods

All thyroid function tests were performed by the Institute of Laboratory Medicine of the hospital. Standard laboratory quality procedures were routinely employed, and regular participation in inter-laboratory tests were part of the quality management strategy. The laboratory is accredited by the German National Accreditation Body (DAkKS). Laboratory evaluation included measurement of FT4, FT3 and TSH. Thyroid autoantibodies (thyroid peroxidase antibodies (TPO-Abs) or TSH receptor antibodies (TSH-R Abs)) were measured in case of suspicion or for the purpose of exclusion of thyroid autoimmune disorders, and not routinely available on all subjects.

TSH was measured with an automated direct chemoluminescence method (TSH3-Ultra, ADVIA Centaur XP, Siemens Healthcare Diagnostics, Erlangen, Ger-

many). The TSH3-Ultra assay is based on the 3. International Standard (WHO) for human TSH (IRP 81/565) with a range of linearity from 0.006 to 160.03 mIU/l and functional sensitivity of 0.008 mIU/l. Pooled serum samples were used for determining intra-assay ($n = 20$) and inter-assay imprecision (measured in duplicate over 10 consecutive days). Serum samples in the range from 0.52 mIU/l to 132.8 mIU/l showed coefficients of variation (CVs) from 1.4% to 2.4% (intra-assay imprecision) and 0.9% to 2.9% (inter-assay imprecision). At a TSH value of 0.52 mIU/l, the intra-assay CV was 1.4%, and the inter-assay CV 2.2%, whereas at a TSH concentration of 0.008 mIU/l (limit of quantitation, functional sensitivity) the inter-assay CV increased to 14.1%. FT3 and FT4 were measured with an automated competitive chemoluminescence method (FT4, FT3, ADVIA Centaur XP, Siemens Healthcare Diagnostics, Erlangen, Germany) with a range of linearity from 0.3 to 30.8 pmol/l (FT3) and 1.3 to 155 pmol/l (FT4). The reference interval used for FT3 was 3.1 to 6.8 pmol/l and 10 to 23 pmol/l for FT4. Serum samples with FT3 concentrations in the range from 2.9 to 14.2 pmol/l showed CVs from 2.4% to 3.1% (intra-assay imprecision) and CVs from 2.3% to 3.9% (inter-assay imprecision). Serum samples with FT4 concentrations in the range from 9.3 to 38.8 pmol/l showed CVs from 2.2% to 3.3% (intra-assay imprecision) and CVs from 2.5% to 4.0% (inter-assay imprecision).

The predefined reference range (0.47–3.73 mIU/l) that should be verified in this study was previously established following standard procedure (28). It stemmed from a local sample of 143 healthy subjects (97 women, 46 men, aged 51 ± 16 years) with no evidence for any thyroid dysfunction, negative anti-body screening (TPO-Ab) and unremarkable findings by ultrasound. Mean \pm SD values were for FT3 4.98 ± 0.60 , FT4 16.55 ± 2.21 and TSH 1.46 ± 0.66 .

Thyroid peroxidase antibodies (TPO-Abs) were measured with an automated competitive chemoluminescence method (Anti-TPO, ADVIA Centaur XP, Siemens Healthcare Diagnostics, Erlangen, Germany). Concentrations of TPO-Ab < 60

U/ml were considered TPO-Ab negative. Range of linearity of the anti-TPO assay is from 15 to 1300 U/ml. Serum samples with TPO-Ab concentrations in the range from 70.8 to 458.5 U/ml showed CVs from 1.3% to 6.8% (intra-assay imprecision) and CVs from 2.8% to 3.4% (inter-assay imprecision). TSH receptor antibodies (TSH-R Abs) were measured with a competitive ELISA (Anti-TSH-Receptor, EUROIMMUN AG, Luebeck, Germany). Concentrations of TSH-R Abs < 2.0 IU/l were considered TSH-R Ab negative. Range of linearity of the TSH-R Ab assay was from 0.2 to 20 IU/l. Serum samples with TSH-R Abs concentrations in the range from 2.8 to 15.8 IU/l show CVs from 2.4 % to 5.5% (intra-assay imprecision) and CVs from 6.5% to 13.0% (inter-assay imprecision).

Statistical methods

Descriptive data are shown as mean \pm standard deviation (SD). TSH values were non-normally distributed and conventionally converted to logarithmic units for all calculations. Following the method described by Katayev et al. (17) the cumulative frequencies of \ln TSH were related to the respective TSH value. The distribution was then used to calculate the regression line in the linear middle part of the curve. The linear part of the relationship was identified and selected by both visual inspection of the graphical representation and statistical means resulting in a correlation coefficient of $r > 0.99$. The optimised linear regression fitted by least square analysis over the linear part of the relationship was then used to derive the respective reference limits by extrapolation to the boundaries of the 95% confidence interval.

Mathematically, the regression line was described by the following equation,

$$y = a \cdot x + c + e$$

where y = the \ln TSH value, x = the cumulative frequency, a = the slope of the regression line, c = the intercept and e = an added small component of the acceptable linearity error (14). The upper and lower reference limits were calculated at a cumulative frequency of 2.5% or 97.5%, respectively.

- Lower limit = $a \cdot 2.5 + e$
- upper limit is $97.5 \cdot a + e$

Tab. 1 Characteristics of the patients in this study (n = 399)

parameter (unit)	mean \pm SD
gender (%)	<ul style="list-style-type: none"> • 73 female • 27 male
age (years)	47 \pm 16
body mass index (kg/m ²)	27 \pm 5.6
smoking habits (%)	<ul style="list-style-type: none"> • 23 smokers • 77 non-smokers
thyroid volume (ml)	15 \pm 9
FT3 (pmol/l)	5.1 \pm 0.6
FT4 (pmol/l)	14.3 \pm 1.8
TSH (mIU/l)	1.6 \pm 1.0

where a and e were taken from the regression line established before.

The reference values obtained by the various methods were compared based on the reference change value (RCV). Statistical significance was established following standard procedure, as described by Fraser et al. (8).

$$RCV = 2^{1/2} \cdot Z \cdot (CV_a^2 + CV_i^2)^{1/2}$$

where Z is the probability selected for significance (The chosen Z value of 1.96 corresponds to a significance level of 0.05.), CV_a the analytic variation (between-run, inter-assay variation), and CV_i the within-subject biologic variation. Both parameters have been estimated based on our own data, and biological variation is reported in Results.

Statistical analyses were performed using R statistical package (version 3.0.3 for Mac) (21).

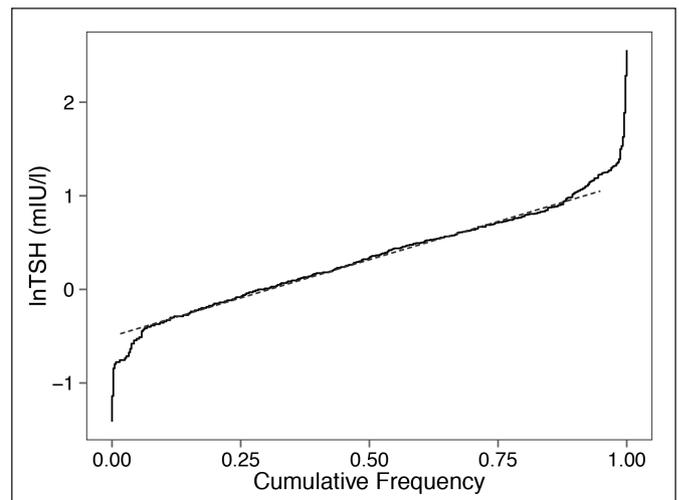
Results

Baseline characteristics of the 399 subjects included in the study are presented (► Tab. 1). Patients were predominately women (73%), aged 47 \pm 17 years (range 17–79 years). FT3 values were 5.1 \pm 0.6 pmol/l and FT4 14.3 \pm 1.9 pmol/l. TSH values ranged from 0.32 mIU/l to 9.81 mIU/l, with a mean of 1.61 \pm 0.97 mIU/l and a median of 1.40 mIU/l (IQR 0.92, 2.04).

The cumulative frequencies were related to the individual logarithmic TSH levels (► Fig. 1). As can be seen, the relationship follows a straight line over a wide segment in the middle part, changing sharply towards the ends. An optimised least square regression line was established over the linear part, as described in Methods. By extrapolating the regression line to the 95% confidence interval of the frequency distribution, respective lower and upper reference limits were derived. The reference interval established by this technique was 0.57–3.32 mIU/l.

Biological variation was estimated in a sample of 65 untreated euthyroid subjects (51 women, mean age 60 \pm 14 years) that were followed for 8.4 \pm 2.8 months. The CV_i for TSH was 19.2%. Fig. 2a shows a comparison of the intra-individual vs inter-individual spread (CV 55%) of the distribu-

Fig. 1 Cumulative frequency distribution of TSH versus the individual logarithmic TSH values: The linear middle part was analysed by linear least square regression (blue dashed line, $y = -0.48x + 0.016$, $r = 0.998$, $p < 0.0001$). Reference limits of 0.57 and 3.32 mIU/l TSH were calculated by extrapolation to 2.5% and 97.5%.



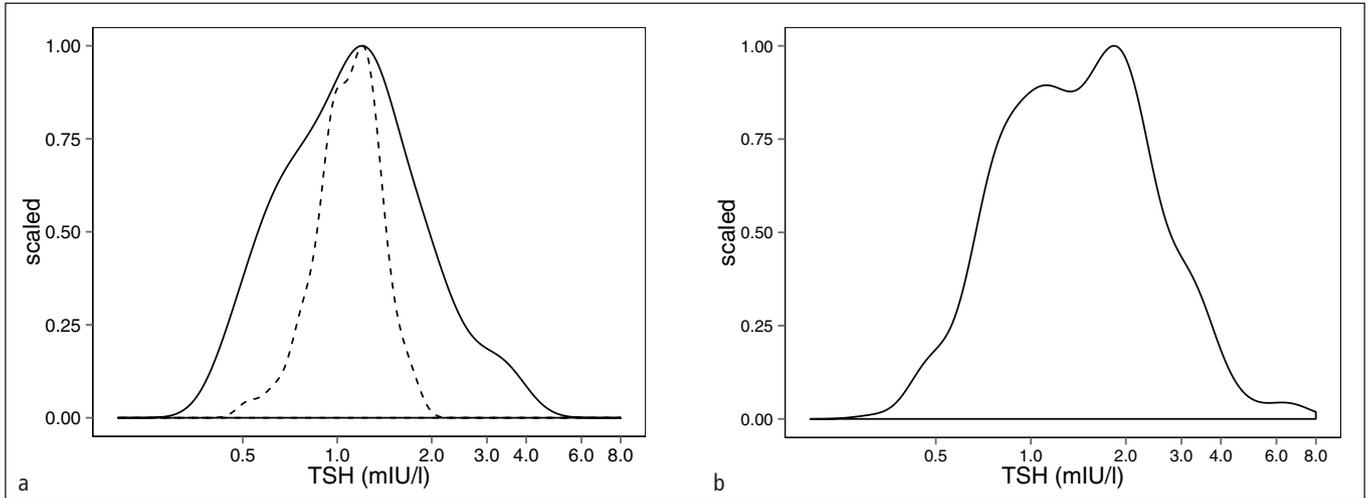


Fig. 2 Inter-individual (–) and intra-individual variation (---) of TSH
a) 65 euthyroid subjects; **b)** total collective (n = 399)

tion of TSH values in the 65 patients, and, for comparison, the distribution of TSH in the whole collective (► Fig. 2b).

The reference range obtained by this technique was compared with other predefined reference ranges that had either been previously established by conventional technique or were recommended by the manufacturer (► Tab. 2). Based on RCVs, which incorporate biological CV_i and analytical CV_a , the reference interval derived by the new technique was in close agreement with the conventionally established reference limits for the same assay (► Tab. 2). It was also in agreement with the previously established reference interval for the local population using another method, but there was significant deviation from a reference range proposed by the manufacturer of that method (► Tab. 2).

Discussion

By examining 399 subjects without evidence for thyroid dysfunction we have verified the reference range in a local population according to a method proposed by Katayev et al. (17). The established reference interval of 0.57–3.32 mIU/l was comparable to the predefined reference limits (laboratory established reference range) and the manufacturer's recommendation for the assay (► Tab. 2). Interestingly, the reference intervals were consistent with those obtained in the same population (not same sample) by a different method, but discrepant to a recommendation issued by the manufacturer of that method (► Tab. 2) (18). As for the between assay comparisons, it should be noted that both TSH assays tested are traceable to the WHO standard for TSH and the two methods have been reported to be in good agreement

(20). However, harmonisation among commercially available methods is not generally guaranteed (24).

Establishing valid reference ranges is of considerable importance and appropriate standardisation procedures have been recommended (28). Specifically for TSH, a controversy has recently emerged surrounding its reference interval, particularly the upper limit of the reference range. The cut-offs in this region proposed by various investigators are widely discrepant, ranging from 2 to 4 mIU/l. This dispute has been subject to a number of original articles and reviews, but remains unresolved (4, 5, 13, 19, 22, 27). While it is unlikely that the discrepancies are solely due to methodological aspects, it is important to consider the reliability of the reference interval used.

The question arises as to how clinicians can practically assess the validity of the reference range they employ apart from rely-

method		reference		
		interval (mIU/l)	change values* (RCVs) (%)	deviation from conventional interval
established	in the present study	0.57 – 3.32	21.12	acceptable
	by conventional means	0.47 – 3.73	-	reference method
currently used		0.40 – 4.00	18.7	acceptable
recommended by Siemens Healthcare Diagnostics		0.55 – 4.78	17.28	acceptable
established previously by conventional means (18)		0.60 – 3.83	27.20	acceptable
recommended by Roche		0.27 – 4.20	74.13	lower limit unacceptable

* Any change exceeding the RCV of the reference method (53%) is significant.

Tab. 2

Comparison of reference intervals for TSH obtained by different methods

ing on this having been well established by the manufacturer or laboratory institution in the first place. Indeed, post hoc analyses and indirect methods have been proposed to serve this purpose (6, 10, 14, 17). In the present study, we used an indirect method to evaluate the reference range of TSH post hoc in a fairly large retrospective sample.

The close agreement between our pre-established reference range and the post hoc verification by the novel approach lends support to the validity of the indirect method. The feasibility and reliability of the technique have previously been confirmed by others (6, 7, 17), although Horowitz has been more critical (16). The technique appears to offer a practicable way to clinicians of obtaining confirmation of the validity of the reference system they are using for critical parameters such as TSH. The advantage of the novel approach comprises the relative ease of the procedure itself and the applicability to a large set of post hoc data, which is mostly readily available.

Limitations

Limitations relate to the presently limited experience with the method and to situations where the sample available is small, heterogeneity prevalent or the method poorly standardised. Our sample size was reasonably large, but not nearly as big as some pioneering studies of the method (7, 17). Because of this limitation and with a practice setting in mind, we made some changes to the protocol used by others in that we did not rely on the removal of pathological data by the statistical use of an outlier test, but incorporated a clinical screener eliminating readily known pathologies thereby strengthening the homogeneity of the collective. However, we did assess suitability of the data by ascertaining the robustness of middle part linearity and requiring a correlation coefficient exceeding 0.99 to pass the test. Following this approach, the sample may still comprise a small percentage of untested or hidden pathologies, which by their nature are likely to fall outside the middle segment and have little, if any, impact on the determination of extrapolated reference limits.

The new approach therefore still requires cautious interpretation and further confirmation. The method has been used to compare reference ranges in various subpopulations and to establish reference ranges for different ethnic groups (7). However, it should be noted that our intended use of the method was not as a substitute for the conventional technique, but for verification purpose of a pre-established reference interval. In case of clear discrepancy to the established reference system, scrutinising and revisiting the limits is advisable. A discrepancy observed in the present study of the reference range from the intervals originally proposed by the manufacturer was of concern. It re-emphasises the mandatory recommendation of establishing laboratory-defined reference values, although this poses a considerable burden on institutions performing the test. Not all may be in a position to define disease free controls and rules have officially been eased requiring only 20 healthy subjects for verification of manufacturer's recommendations (28). It may be interesting to see how the indirect method compares with those less stringent requirements, and whether the small numbers recommended in the direct approach provide sufficiently consistent statistical outcomes.

Improved control and harmonisation of thyroid function tests has become a topic of renewed interest following a recent report and comparative study involving 16 assays by the IFCC Working Group for Standardization of Thyroid Function Tests (24, 25). The report described the presence of assay-specific biases and an unexpectedly high variation among the tests. As a result, clinicians treating thyroid patients should be conscious of the limitations when interpreting thyroid parameters. Furthermore, indirect methods, such as the one described here, offer a tool of re-analysing data on a broad scale and targeting populations of interest. This would allow clinicians to network for standardisation purpose (17).

Another aspect relates to disease prevalence, which obviously is a function of the applied reference system. The use of inappropriate reference intervals that have not been specifically verified for the particular population may lead to a considerable overestimation or underestimation of the

true prevalence of thyroid dysfunction (26). Problems with the estimation of accurate reference intervals for TSH using existing techniques have been mentioned above. The regression method described seems to be better suited to deal with a skewed distribution of TSH values in disease free populations, because theoretically the cut-offs based on an optimised regression equation representing the area of normality are less subject to distribution bias. The non-normal distribution of TSH has been intensely debated. Some authors have attributed the phenomena to hidden thyroid disease, others disagreeing with that notion (11, 23).

When comparing reference intervals and establishing statistical significance in differences observed both analytical variation and biological variation have to be considered. The reference change value (RCV) incorporates both aspects (8). To this end, biological variation was derived from our own panel of patients. The intra-individual spread we observed was only approximately one third as wide as the inter-individual variation in the population confirming earlier data by Andersen et al. (1). This means that a population-based reference range for TSH, even if well established, might fail to detect thyroid malfunction at the individual level (13). For instance, an association between high normal thyroid function and atrial fibrillation has been described (12). In the Rotterdam study, subjects with lower TSH within the euthyroid range had an almost twofold increased risk of atrial fibrillation, compared with those in the highest quartile (12). This might raise the suspicion that a portion of the subjects classified euthyroid actually suffered from subclinical hyperthyroidism.

Indeed, our lower reference limit had to be raised in the evaluation process. A strict reliance on recommendations could result in missing a portion of patients with subclinical hyperthyroidism. Defining the lower limit of the reference range is an important consideration, particularly in a nuclear medicine setting with a high prevalence of thyroid autonomy and frequent decisions about radioiodine application. The same is true for the upper reference limit and subclinical hypothyroidism, which has been the subject of a continuing

debate (19). We note that the controversy may not be readily resolved, because differences between methods or populations that are discernible based on RCVs are relatively wide and the intra-individual variation is much more narrow than the inter-individual variation (►Fig. 2a). Indirect methods tend to produce a more narrow reference interval based on the nature how they work, compared to conventional techniques. Hence, we recognise the existence of a grey zone of uncertainty that surrounds the reference limits and favour a more narrow reference range as a diagnostic tool followed by additional testing, e.g. antibody measurement and thyroid imaging (13). It is, however, clear that the treatment decision is based on more than reference limits and takes into account the patient's history and clinical symptoms. Determination of individual set-points on the basis of repeated TSH measurements may be more appropriate for a more personalised approach, but are currently not well established (9).

Conclusion

In this study we describe the application of an indirect post hoc method according to Katayev and coworkers (17) for the assessment of the laboratory reference intervals of TSH. Reference intervals derived by this approach were plausible, comparing well with the predefined reference range in the local population, but in part differed significantly from manufacturer's recommendations. The method complements the conventional direct sampling approach and offers additional advantages over the conventional technique related to the retrospective design, ease of use (apart from refined statistical modelling), resource savings, targeted subgroup analyses and better suitability for slightly skewed distributions. This should warrant further exploration and propagate the broader application of the technique in the field.

The clinical relevance seems to extend particularly to the diagnosis and treatment of thyroid autonomy.

It does, however, not overcome limitations with the interpretation of TSH values that extend beyond the reference interval. These as always will require a more sophisticated follow-up process to determine the final diagnostic decision.

Funding

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

Conflict of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

References

- Andersen S, Pedersen KM, Bruun NH, Laurberg P. Narrow individual variations in serum T(4) and T(3) in normal subjects: a clue to the understanding of subclinical thyroid disease. *J Clin Endocrinol Metab* 2002; 87: 1068–1072.
- Aoki Y, Belin RM, Clickner R, Jeffries R et al. Serum TSH and total T4 in the United States population and their association with participant characteristics: National Health and Nutrition Examination Survey. *Thyroid* 2007; 17: 1211–1223.
- Baloch Z, Carayon P, Conte-Devolx B et al. Laboratory medicine practice guidelines. Laboratory support for the diagnosis and monitoring of thyroid disease. *Thyroid* 2003; 13: 3–126.
- Brabant G, Beck-Peccoz P, Jarzab B et al. Is there a need to redefine the upper normal limit of TSH? *Eur J Endocrinol* 2006; 154: 633–637.
- Dickey RA, Wartofsky L, Feld S. Optimal thyrotropin level: normal ranges and reference intervals are not equivalent. *Thyroid* 2005; 15: 1035–1039.
- Dorizzi RM, Giannone G, Cambiaso P et al. Indirect Methods for TSH Reference Interval: At Last Fit for Purpose? *Am J Clin Pathol* 2011; 135: 167–174.
- Feng Y, Bian W, Mu C et al. Establish and verify TSH reference intervals using optimized statistical method by analyzing laboratory-stored data. *J Endocrinol Invest* 2014; 37: 277–284.
- Fraser CG, Hyltoft Peterson P, Larsen ML. Setting analytical goals for random analytical error in specific clinical monitoring situations. *Clin Chem* 1990; 36: 1625–1628.
- Goede SL, Leow MK-S, Smit JWA, Dietrich JW. A novel minimal mathematical model of the hypothalamus-pituitary-thyroid axis validated for individualized clinical applications. *Math Biosci* 2014; 249: 1–7.
- Grossi E, Colombo R, Cavuto S et al. TheRELAB project: a new method for the formulation of reference intervals based on current data. *Clin Chem* 2005; 51: 1232–1240.
- Hamilton TE, Davis S, Onstad L, Kopecky KJ. Thyrotropin levels in a population with no clinical, autoantibody, or ultrasonographic evidence of thyroid disease: implications for the diagnosis of subclinical hypothyroidism. *J Clin Endocrinol Metab* 2008; 93: 1224–1230.
- Heeringa J, Hoogendoorn EH, van der Deure WM et al. High-normal thyroid function and risk of atrial fibrillation: the Rotterdam study. *Arch Intern Med* 2008; 168: 2219–2224.
- Hoermann R, Midgley JEM. TSH measurement and its implications for personalised clinical decision-making. *J Thyroid Res* 2012; 2012: 1–9.
- Hoffmann RG. Statistics in the practice of medicine. *JAMA* 1963; 185: 864–873.
- Hollowell JG, Staehling NW, Flanders WD et al. Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey. *J Clin Endocrinol Metab* 2002; 87: 489–499.
- Horowitz GL. Estimating reference intervals. *Am J Clin Pathol* 2010; 133: 175–177.
- Katayev A, Balciza C, Seccombe DW. Establishing reference intervals for clinical laboratory test results: is there a better way? *Am J Clin Pathol* 2010; 133: 180–186.
- Larisch R, Tawo PL, Hoermann R. TSH, FT3 und FT4 bei Personen ohne Schilddrüsenfunktionsstörung. Jahrestagung der Deutschen Gesellschaft für Nuklearmedizin 2010; P31.
- Laurberg P, Andersen S, Carlé A et al. The TSH upper reference limit: where are we at? *Nat Rev Endocrinol* 2011; 7: 232–239.
- Rawlins ML, Roberts WL. Performance characteristics of six third-generation assays for thyroid-stimulating hormone. *Clin Chem* 2004; 50: 2338–2344.
- R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing 2013; www.R-project.org/
- Surks ML, Goswami G, Daniels GH. The thyrotropin reference range should remain unchanged. *J Clin Endocrinol Metab* 2005; 90: 5489–5496.
- Surks MI, Hollowell JG. Age-specific distribution of serum thyrotropin and antithyroid antibodies in the US population: implications for the prevalence of subclinical hypothyroidism. *J Clin Endocrinol Metab* 2007; 92: 4575–4582.
- Thienpont LM, Van Uytvanghe K, Beataill G et al. Report of the IFCC Working Group for Standardization of Thyroid Function Tests; part 1: thyroid-stimulating hormone. *Clin Chem* 2010; 56: 902–911.
- Thienpont LM, Van Uytvanghe K, Beataill G et al. Report of the IFCC Working Group for Standardization of Thyroid Function Tests; part 2: free thyroxine and free triiodothyronine. *Clin Chem* 2010; 56: 912–920.
- Völzke H, Alte D, Kohlmann T et al. Reference intervals of serum thyroid function tests in a previously iodine-deficient area. *Thyroid* 2005; 15: 279–285.
- Wartofsky L, Dickey RA. The evidence for a narrower thyrotropin reference range is compelling. *J Clin Endocrinol Metab* 2005; 90: 5483–5488.
- Wayne PA: Clinical and Laboratory Standards Institute. Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline. CLSI document C28-A3. Clinical and Laboratory Standards Institute 2008.