Pilar Fernández-Calle*, Jorge Díaz-Garzón, William Bartlett, Sverre Sandberg, Federica Braga, Boned Beatriz, Anna Carobene, Abdurrahman Coskun, Elisabet Gonzalez-Lao, Fernando Marques, Carmen Perich, Margarida Simon and Aasne K. Aarsand, on behalf of the EFLM Working Group on Biological Variation and Task Group for the Biological Variation Database

Biological variation estimates of thyroid related measurands – meta-analysis of BIVAC compliant studies

https://doi.org/10.1515/cclm-2021-0904
Received August 13, 2021; accepted October 18, 2021;
published online November 15, 2021

Abstract

Objectives: Testing for thyroid disease constitutes a high proportion of the workloads of clinical laboratories worldwide. The setting of analytical performance specifications (APS) for testing methods and aiding clinical interpretation of test results requires biological variation (BV) data. A critical review of published BV studies of thyroid disease related measurands has therefore been undertaken and meta-analysis applied to deliver robust BV estimates.

Methods: A systematic literature search was conducted for BV studies of thyroid related analytes. BV data from studies compliant with the Biological Variation Data Critical Appraisal Checklist (BIVAC) were subjected to meta-analysis. Global estimates of within subject variation (CV_I) enabled determination of APS (imprecision and bias), indices of individuality, and indicative estimates of reference change values.

Results: The systematic review identified 17 relevant BV studies. Only one study (EuBIVAS) achieved a BIVAC grade of A. Methodological and statistical issues were the reason for B and C scores. The meta-analysis derived CV_I generally delivered lower APS for imprecision than the mean CV_A of the studies included in this systematic review.

Conclusions: Systematic review and meta-analysis of studies of BV of thyroid disease biomarkers have enabled delivery of well characterized estimates of BV for some, but not all measurands. The newly derived APS for imprecision for both free thyroxine and triiodothyronine may be

EFLM Working Group on Biological Variation: Pilar Fernández-Calle, Jorge Díaz-Garzón, William Bartlett, Sverre Sandberg, Anna Carobene, Abdurrahman Coskun, Aasne K. Aarsand. **EFLM Task Group for the Biological Variation Database**: Pilar Fernández-Calle, Jorge Díaz-Garzón, William Bartlett, Sverre Sandberg, Federica Braga, Boned Beatriz, Anna Carobene, Abdurrahman Coskun, Elisabet Gonzalez-Lao, Fernando Marques, Carmen Perich, Margarida Simon, Aasne K. Aarsand.

***Corresponding author: Pilar Fernández-Calle**, Analytical Quality Commission, Spanish Society of Laboratory Medicine (SEQCML), Barcelona, Spain; and Department of Laboratory Medicine, Hospital Universitario La Paz, Madrid, Spain,
E-mail: pfernandez.hulp@gmail.com
Jorge Díaz-Garzón, Analytical Quality Commission, Spanish Society of Laboratory Medicine (SEQCML), Barcelona, Spain; and Department of Laboratory Medicine, Hospital Universitario La Paz, Madrid, Spain
William Bartlett, Undergraduate Teaching, School of Medicine, University of Dundee, Dundee, Scotland
Sverre Sandberg, Department of Medical Biochemistry and Pharmacology, Norwegian Organization for Quality Improvement of Laboratory Examinations (NOKLUS), Haukeland University Hospital, Bergen, Norway
Federica Braga, Research Centre for Metrological Traceability in Laboratory Medicine (CIRME), University of Milan, Milan, Italy
Boned Beatriz, Analytical Quality Commission, Spanish Society of Laboratory Medicine (SEQCML), Barcelona, Spain; and Department of Laboratory Medicine, Hospital Royo Villanova, Zaragoza, Spain
Anna Carobene, Servizio Medicina di Laboratorio, Ospedale San Raffaele, Milan, Italy
Abdurrahman Coskun, Department of Medical Biochemistry Atasehir, School of Medicine, Acibadem Mehmet Ali Aydınlar University, Istanbul, Turkey
Elisabet Gonzalez-Lao and Carmen Perich, Analytical Quality Commission, Spanish Society of Laboratory Medicine (SEQCML), Barcelona, Spain
Fernando Marques, Analytical Quality Commission, Spanish Society of Laboratory Medicine (SEQCML), Barcelona, Spain; and Clinical Biochemistry Department, Metropolitan North Clinical Laboratory (LUMN), Germans Trias i Pujol University Hospital, Barcelona, Spain
Margarida Simon, Analytical Quality Commission, Spanish Society of Laboratory Medicine (SEQCML), Barcelona, Spain; and Department of Clinical Biochemistry, Hospital Universitario Badalona, Badalona, Spain
Aasne K. Aarsand, Department of Medical Biochemistry and Pharmacology, Haukeland University Hospital, Bergen, Norway

considered challenging. The high degree of individuality identified for thyroid related measurands reinforces the importance of RCVs. Generation of BV data applicable to multiple scenarios may require definition using "big data" instead of the demanding experimental approach.

Keywords: biological variation; biological variation data critical appraisal checklist (BIVAC); meta-analysis; thyroid biomarkers.

Introduction

Testing for thyroid disease constitutes a high proportion of the workloads of clinical laboratories worldwide [1, 2]. Different measurands are applied in the contexts of screening, diagnosis, and monitoring of thyroid disease states of varying aetiologies across all age groups. Interpretation of test results is complicated by the impacts of many physiological factors such as pregnancy, metabolic status, obesity, comorbidities that influence the metabolism of thyroid hormones, and their regulators [3–9]. This has led to the adoption of a diversity of decision limits for diagnosis and monitoring of thyroid disorders in different settings [10–12]. It follows that optimal analytical performance specifications (APS) for the measurement methods employed in the management thyroid disorders may also vary depending upon the setting; the selection APS is essential to enable choice of methods to guarantee good clinical care.

The first EFLM Strategic Conference, held in 2014 in Milan, defined three models to derive APS [13]: model 1, based on the effect of analytical performance on the clinical outcome; model 2, based on components of BV of the measurands and model 3, based on the state of the art of the measurement procedure, defined as the highest level of analytical performance technically achievable. Although the optimal approach for setting APS would be the adoption of clinical outcome-based studies [13] for many measurands, such studies are scarce. In this situation, the BV model has been proposed as an acceptable secondary approach to identify APS for many measurands. The adoption of this approach may be considered appropriate for thyroid hormone measurands [3–7].

Estimates of within-subject (CV_I) and between-subject (CV_G) biological variation for many different measurands have been collated for inclusion in the EFLM Biological Variation Database [14]. The quality of studies, and veracity of the resultant estimates of BV, impact on the validity and safety of their clinical applications are considered. To address this, a standard for evaluating BV studies, the Biological Variation Data Critical Appraisal Checklist (BIVAC) was published in 2018 by the EFLM Working Group on Biological Variation and the Task Group on the Biological Variation Database (TG-BVD) [15]. Furthermore, a meta-analysis approach to deliver global estimates based on BIVAC compliant studies was developed [15].

The aim of the study reported here is, following a systematic literature search, to review published BV studies of thyroid disease related measurands, verify those which are BIVAC compliant and, enable the collation of data sets of sufficient quality to apply meta-analysis to produce robust BV estimates. Those estimates have then been used to determine; index of individuality (II), reference change value (RCV) [16] and BV based APS [17] applicable to thyroid stimulating hormone (TSH), free thyroxine (fT4), thyroxine (T4), free triiodothyronine (fT3), triiodothyronine (T3), thyroid stimulating immunoglobulin (TSI), thyroglobulin (TG), anti thyroid peroxidase antibody (anti TPO Ab), anti thyroglobulin antibody (anti TG Ab) and thyroxine binding globulin (TBG), with attention given to different population subgroups and states of health.

Materials and methods

A literature search for BV studies of thyroid related analytes published up until 13/09/2021 was conducted as previously described [18, 19]. Briefly, searches were carried out in PubMed, using as key terms the analyte in question with each of the following combinations: "within-subject*", "between-subject*", "within-person*", "between-person*", "interindividual*", "inter-individual*" and "intraindividual*", "intra-individual*", where the asterisk denotes "biological variation" "variation", "coefficient of variation", and "CV". The retrieved publications are in this study identified by the reference number they have been given in the EFLM Biological Variation Database (BVD), as shown in Supplementary Table 1, with a subscript indicating different subgroups from the same study.

Applying the BIVAC criteria, each study is appraised against 14 quality items (QI), which focus on the pre-analytical procedures, measurand measurement procedure, applied statistical methods, and presentation of data. Each QI may be assigned an A, B, C or D score, indicating increasing non-compliance; the lowest score of any QI defines the overall BIVAC grade. BV estimates derived from studies that receive a D score are considered unsuitable for use in clinical practice. A summary of Quality items of the Biological Variation Data Critical Appraisal Checklist and their achievable scores is provided in Supplementary Table 2. Two appraisers assessed the studies independently; a third appraiser checked the evaluation of the scores where there was not full agreement to enable a consensus score to be reached. A weighted median approach delivered the global CV_I and CV_G estimates [20]; the overall BIVAC grade together with the inverse width of the confidence interval (CI) were used as weighting factors. Meta-analysis was performed as defined by the paper published by Aarsand et al. [15], on data from BIVAC compliant studies performed in adults including all studies with BIVAC grades A, B and C fulfilling the

following inclusion criteria: age 18–75 years, >2 subjects and >2 samples per subject, BV estimates given as CV and a numerical estimate of CV_A included. For studies reporting more than one BV estimate from the same set of samples, derived by different analytical methods, only one estimate was included [19]. The CV_I and CV_G estimates derived from the meta-analyses were used to determine APS for CV_A and bias/systematic error [17], RCV [16] to enable assessment of significance of change between two measurements at a predetermined probability, and the index of individuality (II) to enable assessment of the utility of population based reference interval [20]. The following formulae were applied:

$$CV_A = 0.5\,(CV_I)$$

$$\text{Bias} = 0.5\left(CV_I^2 + CV_G^2\right)^{0.5}$$

$$\text{RCV} = 100\,\%\left(\exp\left(\pm Z\,2^{1/2}\left(CV_{LnA}^2 + CV_{LnI}^2\right)^{1/2}\right)\right);$$

This enables calculation of asymmetric RCVs, where CV_{Ln} refers to ln-transformed data = $(\ln(1 + CV^2))^{1/2}$ and "Z" refers to the Z-score equal to the number of standard deviations appropriate for the selected probability. To calculate the RCV, every laboratory must apply its own CV_A. For illustrative purposes, we applied the mean CV_A estimate derived from all the studies included in the meta-analysis and a Z value of 1.64 (probability level 95%) to calculate the RCVs in this study,

$$II = CV_I/CV_G$$

Probability of curves for the significance of percentage change between consecutive results were constructed using Microsoft Excel for upward and downward change using the RCV formula shown above.

Results

The systematic review identified 17 relevant BV studies (Supplementary Table 1). In Table 1, the number of subgroups contained within each study for each measurand, the associated BIVAC grades, the number of studies performed in healthy and non-healthy subjects and the healthy subgroups included in the meta-analysis are shown.

The European Biological Variation Study (EuBIVAS) was the only one allocated an A grade on application of BIVAC [21]. All the other studies were classified as B or C, mainly due to methodological and statistical issues. None of the studies was classified as D, the lowest BIVAC grade indicating that BV data are unreliable. Results of BV estimates derived from meta-analysis together with the II, RCV and APS are provided in Table 2.

Figures 1–4 display graphically the CV_I values with 95% confidence limits for TSH, T4, fT4, T3, fT3 and TG, highlighting inclusions and exclusions from the meta-analysis and associated BIVAC grades of the studies. The RCV values presented in Table 2 enable identification of percentage change at a p value of 0.05 for an increase or decrease in the measurand. Figure 5 shows probability curves for the significance of percentage change between consecutive fT4 results. In this plot, the meta-analysis derived CV_I for fT4 (4.8%, 95% CI 4.6–7.0) and the mean CV_A of the included studies (Table 2) were used as constants. By varying the percentage change to identify values for Z, the probability of significance of the percent change calculation was made possible. From the plot, a change in fT4 from 20.5 to 17.4 pmol/L represents a fall of 15.1% (exceeds RCV for a fall in measurand of −13.7%) which is significant at a probability between 0.95 and 0.975, whereas the same percentage rise from 20.5 to 23.6 (less than the significant RCV for a rise in the measurand of +15.9%) fails to reach significance. If the CV_A is set lower, i.e. at the proposed BV derived APS for imprecision of 2.4%, the decrease and increase of RCV become −11.7% and +13.2%, respectively, and, the same change of 15.1% from a baseline of 20.5 pmol/L is observed to be of greater significance, i.e. the downward (decrease) change is associated with a probability of significant change >0.975 and the upward (increase) change now significant at a probability of >0.95.

Table 1: Number of studies for each measurand with BIVAC grades and number of subgroups.

	n studies	n subgroups	Subgroups grades (A/B/C)	Number of healthy/Non-healthy subgroups	Number of healthy subgroups included in meta-analysis (BIVAC grades)
TSH	12	13	1A/1B/11C	10/3	7 (1A-1B-5C)
fT4	9	9	1A/1B/7C	5/4	5 (1A-1B-3C)
T4	8	11	11C	8/3	5 (5C)
fT3	6	6	1A/1B/4C	5/1	4 (1A-1B-2C)
T3	8	10	10C	7/3	6 (6C)
TG	5	8	1A/1B/6C	8/0	4 (1A-1B-2C)
Anti TPO Ab	2	2	2C	1/1	1 (1C)
Anti TG Ab	2	2	2C	1/1	1 (1C)
TSI	10	11	11C	8/3	5 (5C)
TBG	2	1	1C	1/0	0

Detailed references of the studies are listed on the EFLM website and enable users to understand the intricacies and characteristics of the studies.

Table 2: Biological variation estimates derived from meta-analysis of BIVAC compliant studies performed in healthy adults, with associated II, RCV and APS.

		Meta-analysis derived BV estimates				APS	
	Mean CV_A, %	CV_I (95% CI)	CV_G (95% CI)	II	RCV, %	CV	Bias
TSH	6.4	17.9 (16.2–29.3)	36.2 (33.4–47.0)	0.49	−35.5 to +55.1	9.0	10.1
fT4	4.2	4.8 (4.6–7.0)	8.2 (7.5–10.8)	**0.59**	−13.8 to +16.0	2.4	2.4
T4	5.7	6.4 (4.9–6.9)	12.0 (11.0–12.2)	**0.53**	−18.1 to +22.0	3.2	3.4
fT3	3.1	5.1 (4.7–6.2)	8.2 (8.0–10.5)	0.62	−13.0 to +14.9	2.6	2.4
T3	6.0	9.6 (6.9–10.4)	10.7 (5.0–15.3)	0.9	−23.1 to +30.1	4.8	3.6
TG	5.5	10.7 (10.3–15.4)	81.0 (29.2–84.9)	**0.13**	−24.4 to +32.2	5.4	20.4
Anti TPO Ab[a]	10.6	11.3 *(10.0–12.8)	147 *(113.1–210.1)	**0.08**	−30.2 to +43.2	5.7	36.9
Anti TG Ab[a]	9	8.5 *(7.0–10.1)	82 *(63.1–117.2)	**0.1**	−25.0 to +33.3	4.3	20.6
TSI	8.8	21.2 (14.8–29.3)	35.0 (24.0–48.4)	0.61	−41.1 to +69.7	10.6	10.2
TBG[a]	5.2	4.4 *(3.2–5.6)	12.6 *(9.2–19.7)	**0.35**	−14.7 to +17.2	2.2	3.3

[a]Only one study fulfilled the inclusion criteria of the meta-analysis. Estimates are those derived from this only study; bold and underlined II below 0.6 indicates high individuality of the measurand. CV_I, within-subject coefficient of variation; CV_G, between-subject coefficient of variation; II, index of individuality; RCV, reference change value; APS, analytical performance specifications; CI, confidence interval. The RCV values presented enable identification of change at a probability of 0.05 (95% certainty) between decrease or increase of the concentration of the measurand.

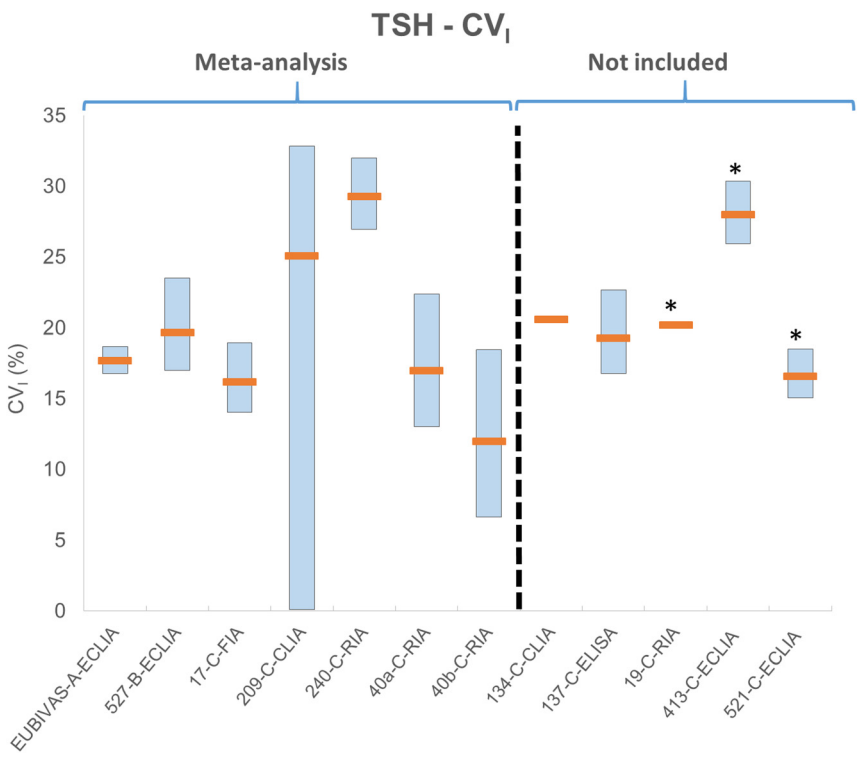

Figure 1: Estimates of CV_I for thyroid stimulating hormone (TSH) with 95% CI.
Studies included and not included in meta-analysis. Black dotted vertical line divides the studies in those included in meta-analysis and those not included. Studies based on non-healthy subjects are marked by*. X-axis: the reference number reflecting the study identifier recorded in EFLM-BVD and Supplementary Table 1, a subscript indicating different subgroups from the same study, followed by the BIVAC grade and an analytical method code for the method used in each study. Box plot orange horizontal bars represent the median of the CVI estimates and blue boxes represent their corresponding 95% CI. ECLIA, electrochemoluminiscence; CLIA, chemoluminiscence; FIA, fluoroimmunoassay; RIA, radioimmunoassay; ELISA, enzymoimmunoassay.

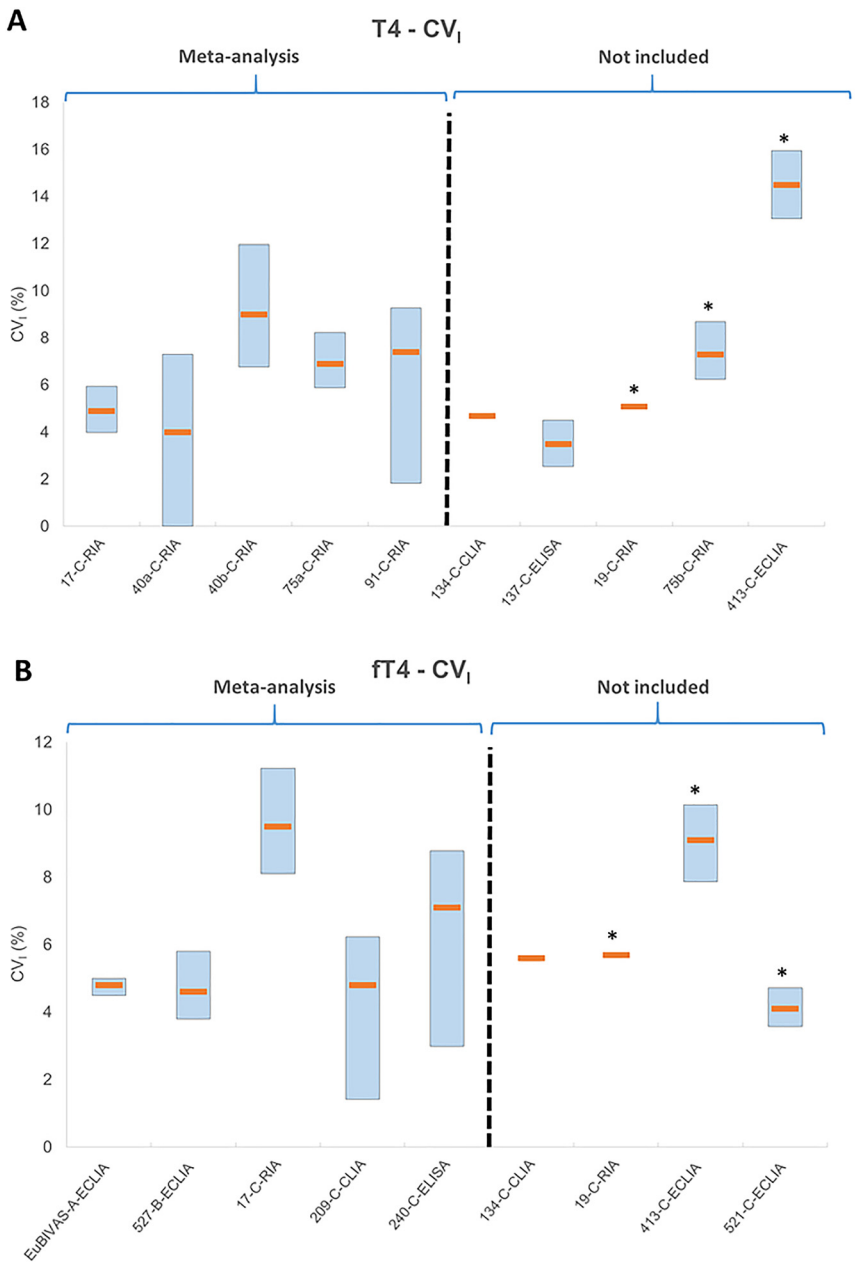

Figure 2: Estimates of CV_I for A) total T4 (T4) and B) free-T4 (fT4) with 95% CI.
Studies included and not included in meta-analysis. Black dotted vertical line divides the studies in those included in meta-analysis and those not included. Studies based on non-healthy subjects are marked by*. X-axis: the reference number reflecting the study identifier recorded in EFLM-BVD and Supplementary Table 1, a subscript indicating different subgroups from the same study, followed by the BIVAC grade and an analytical method code for the method used in each study. Box plot orange horizontal bars represent the median of the CVI estimates and blue boxes represent their corresponding 95% CI. ECLIA, electrochemoluminiscence; CLIA, chemoluminiscence; FIA, fluoroimmunoassay; RIA, radioimmunoassay; ELISA, enzymoimmunoassay.

Discussion

The investigation and management of actual and potential thyroid disorders incorporates interpretation of a range of measurand estimates against reference limits or consensus decision limits [2, 3]. Valid clinical interpretation of test results requires an understanding of the magnitude of BV and how it may differ in various settings. Those decisions are affected by the performance of the assays used to generate the results, which again is informed by knowledge of the magnitude of BV exhibited in a measurand in any clinical scenario.

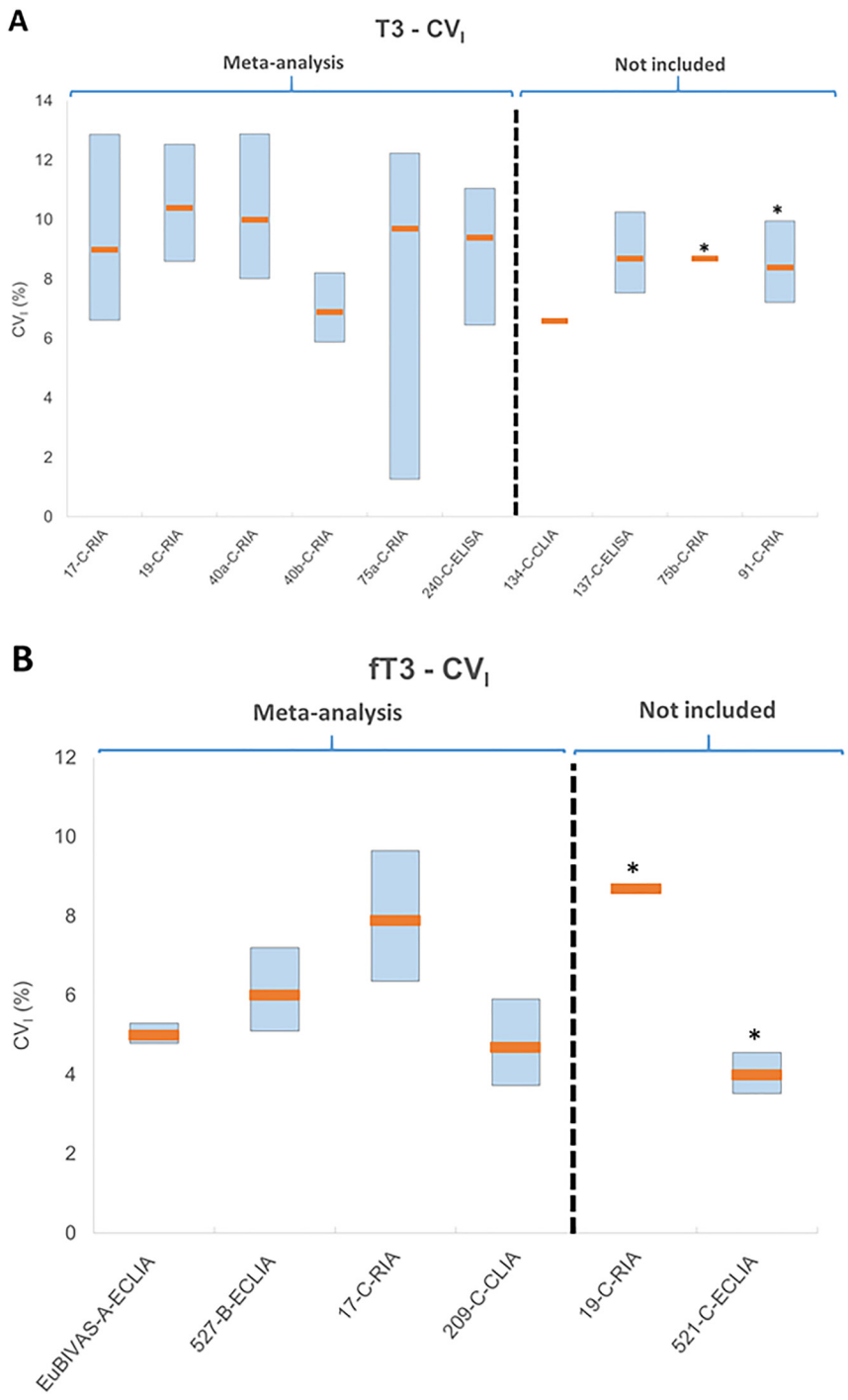

Figure 3: Estimates of CV_I for A) total T3 (T3) and B) free-T3 (fT3) with 95% CI.
Studies included and not included in meta-analysis. Black dotted vertical line divides the studies in those included in meta-analysis and those not included. Studies based on non-healthy subjects are marked by*. X-axis: the reference number reflecting the study identifier recorded in EFLM-BVD and Supplementary Table 1, a subscript indicating different subgroups from the same study, followed by the BIVAC grade and an analytical method code for the method used in each study. Box plot orange horizontal bars represent the median of the CVI estimates and blue boxes represent their corresponding 95% CI. ECLIA, electrochemoluminiscence; CLIA, chemoluminiscence; FIA, fluoroimmunoassay; RIA, radioimmunoassay; ELISA, enzymoimmunoassay. *Non-healthy subjects (A) 75-b: Healthy pregnant women.

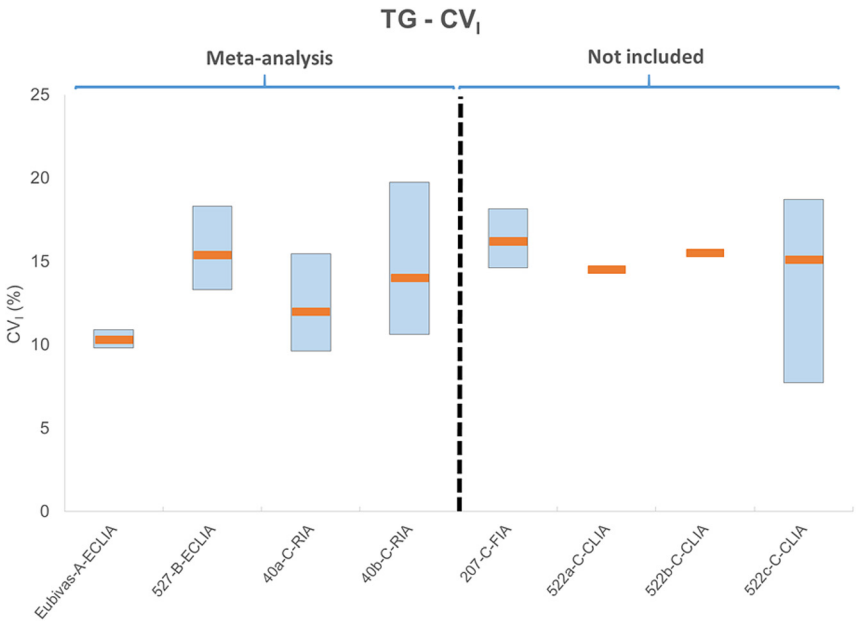

Figure 4: Estimates of CV_I for thyroglobulin (TG) with 95% CI.
/Studies included and not included in meta-analysis. Black dotted vertical line divides the studies in those included in meta-analysis and those not included. Studies based on non-healthy subjects are marked by*. X-axis: the reference number reflecting the study identifier recorded in EFLM-BVD and Supplementary Table 1, a subscript indicating different subgroups from the same study, followed by the BIVAC grade and an analytical method code for the method used in each study. Box plot orange horizontal bars represent the median of the CVI estimates and blue boxes represent their corresponding 95% CI. ECLIA, electrochemoluminiscence; CLIA, chemoluminiscence; FIA, fluoroimmunoassay; RIA, radioimmunoassay; ELISA, enzymoimmunoassay. 522a: Inuit; 522-b: Caucasian; 522-c: All (95 healthy).

Meta-analysis of the data from quality assessed studies of BV has enabled derivation of robust BV data required clinically. The quality of the contributing studies is assured by application of the BIVAC, enabling identification of reliable data sets for inclusion in the meta-analysis to deliver the CV_I and CV_G estimates; these data enable calculation of the II and RCV for thyroid related measurands and will aid interpretation of results from systems meeting the APS defined by the same BV estimates.

Quality and characteristics of thyroid BV studies

The systematic literature review identified only one study, the EuBIVAS, achieving the highest BIVAC grade A [21]; one paper from Mairesse et al. [22] was graded B, while the remainder were classified as C (Supplementary Table 1). The highest number of studies were identified for TSH (n=12) and TSI (n=10). Eight BV subgroups from five studies for TG were identified: four of those were included in meta-analysis, one A, one B and two C studies. There was, however, little observable differences in the data described in these studies and in the four studies not included in the meta-analysis. Meta-analysis was not possible for in anti TPO Ab, anti TG Ab and TBG as only one qualifying paper (paper 40, BIVAC Category C) [23] was found. Caution should therefore be exercised when making use of these BV estimates clinically for these measurands because they do not come from meta-analysis but from a single study.

Many factors may impact on the observed estimates of BV of a measurand including population studied, disease status, technology applied, and timing. The characteristics of the study populations such as age, gender and health status all serve to impact upon observed variation. In addition, analytical method standardization and performance, in terms of imprecision and specificity, might be expected to impact upon the value of a measurand [20]. Clearly technological developments have delivered generational changes in analytical systems deployed, meaning that such performance characteristics of the assays have greatly improved with time. Finally, the rhythmical changes described for thyroid hormones mean that differences in timing of sampling and duration of the study, may result in between study variance of BV estimates. All of these factors have to be considered as part of the review process.

Most of the studies found in the literature search were performed before the year 2000 and consequently the

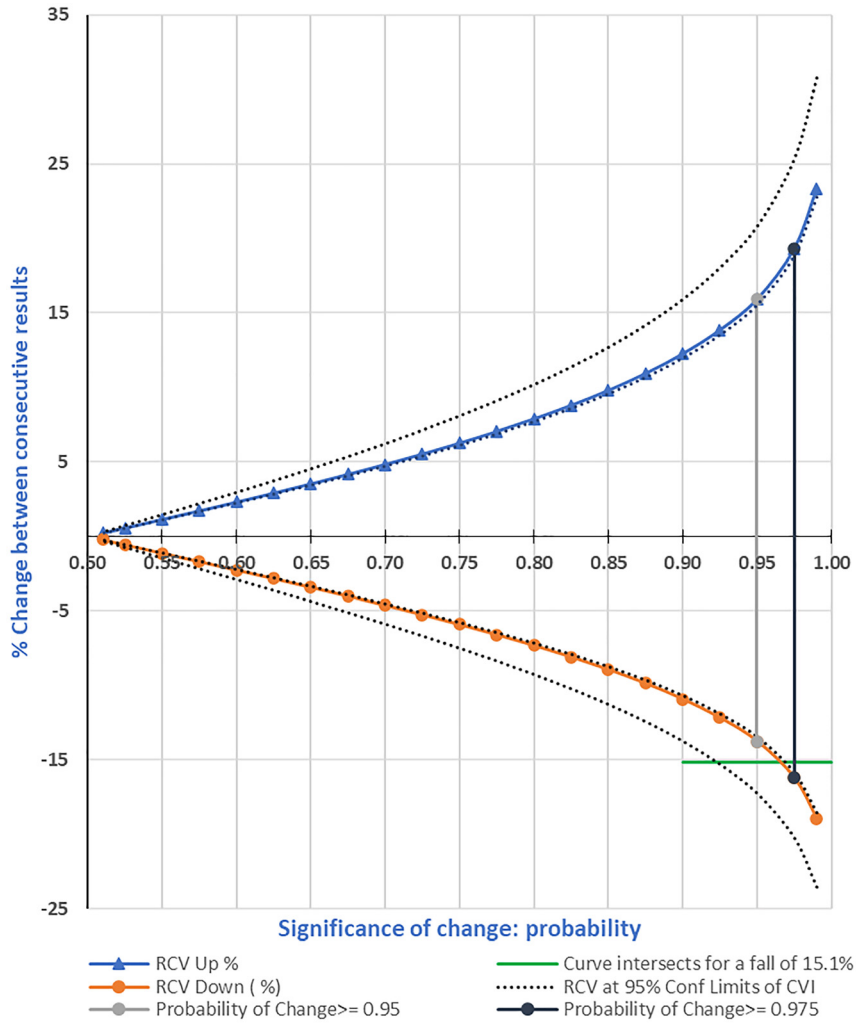

Figure 5: Probability curves for the significance of change between consecutive FT4 results.
Curves were constructed using a rearrangement of the RCV formula to calculate values for Z for varying percentage reference change values (RCV), using the CV_I (meta-analysis result of 4.8%) and the average CV_A estimate based on the included studies (4.2%) as constants. Curves are also drawn for the upper and lower confidence limits of the CV_I estimate (4.6%, 7.0%). The green horizontal line illustrates a value corresponding to a decrease in FT4 concentration of 15.1%, which can be seen to intersect the down curve that corresponds to a probability of >0.95 and <0.975.

analytical methods (mostly RIA) might not be expected to be fully comparable in terms of analytical specificity with those currently in use (ECLIA, CLIA). Taking TSH as an example, we identified two CLIA and four ECLIA studies in comparison with four RIA. In general, the review shows that similar BV estimates were obtained, regardless of the analytical method used. The mean CV_A obtained when pooling CV_A estimates from all studies was 6.4% (RIA CV_A mean: 9.9%, ECLIA: CV_A mean: 2.1% and CLIA CV_A mean: 13.5%, the latter was derived from only one paper [paper 209; analysis performed in a Siemens Centaur]) [24]. Mean TSH concentrations ranged from 1.34 to 2.5 μUI/mL in healthy people. In general, minimal impact on BV estimates for other measurands was observed attributable to the analytical method type applied. This enables use of historical BV data.

Diurnal and seasonal variations in TSH and thyroid hormone measurands have been documented previously [25–28]. A meta-analysis by Kuzmenko et al. [27] indicated that seasonal changes are seen in TSH, fT3 and fT4, but not for T4. This could lead to differences in BV estimates depending upon study duration.

Three of 12 the studies of TSH identified for this review had a study duration extending to one year. Two of them (paper 240; RIA was performed in healthy individuals and paper 413; ECLIA in children [29, 30]), reported higher CV_I (29% and 28%, respectively) than studies of shorter duration. This may be consistent with super-imposed seasonal

variation being included in the BV estimates. The third study (paper 521, employing ECLIA in a study spanning 13 months) [31] identified a CV_I of 16.6%. However, this study population included diseased individuals (subclinical hypothyroidism). None of these studies assessed for trends in their data (Supplementary Table 1), which is a BIVAC criterion. Similar findings applied to T3, fT3, T4, and fT4 in the papers where these hormones are included. Clarification of these relationships requires more studies covering a complete period of one year, with focus on potential trends over time. There is potential here to analyse big data sets from a laboratory information systems to ascertain impact at a population level and extract estimates of BV [31, 32].

Turning to diurnal variations, Liyanage et al. [26] recorded a within-day variability in TSH and fT4 in Sinhalese adults that requires further characterization. This delivers chronobiological complexity as a confounding factor to be understood and accounted for when deriving and applying BV data. Diurnal variation in TSH variation was also identified in a recently published NHANES III US survey, with the peak occurring during the night and the nadir, which approximates to 50% of the peak value, occurring between 10:00 and 16:00 [22]. The authors of this study [25] proposed that this type of BV is not of great significance in terms of clinical interpretation of results as most samplings take place between 08:00 and 18:00 avoiding the night time peak. It will, however, contribute additional higher estimates of BV if the time of day study subjects are sampled is imprecise at an individual or group level. Sviridonova et al. [28] found that the morning median TSH value in the patients with subclinical hypothyroidism was higher than in the afternoon (5.83 mU/L vs. 3.79 mU/L), and they concluded that, according to the TSH reference interval, hypothyroidism could not be diagnosed in about 50% of the cases in the afternoon. Here again there may be a future role for analysis of big data sets to understand the significance of these complex variation states [32, 33].

Our systematic review indicated that state of health has little impact on the CV_I of the measurands reviewed (Figures 1–4). A cautionary note here is that this observation is based on a smaller number of studies in specific disease states such as subclinical hypothyroidism or well controlled hypothyroidism, and it cannot be assumed that this is true of all diseases [34]. Further studies, applying big data approaches could be useful to obtain BV in populations not easily studied in the traditional way, such as children.

Application of BV data

If a measurand exhibits a low II (<0.6), the utility of population based reference intervals is limited for detection of clinically significant results. In this situation, it is better to use the subject under investigation as their own point of reference [35] and to employ RCV to identify significant change between two consecutive test results. A change greater than the RCV indicates significant change at a given probability warranting further action. When the II is low, a patient may have consecutive test results that fall within the population reference range that exhibit a difference that exceeds the RCV. RCV is generally reported as a percentage. Ideally, the II should be greater than 1.4 for population based reference intervals to be applicable, at II values between 0.6 and 1.4 they are of more limited utility. The lower the value of II, then the higher individuality of the measurand and the lower the utility of population based reference intervals.

As seen in Table 2, the II values indicate high degrees of individuality for all the measurands, with the greater majority having values of less than or about 0.6. This finding indicates the use of RCV or personalized reference intervals is preferable in the interpretation of consecutive test results in an individual [17, 20, 35].

Correct interpretation of significance of change in results is clearly important and RCV is considered a valid concept to enable this. We introduce probability curves as a possible tool under development to aid in the process of interpretation. They enable an objective assessment of the statistical significance of an observed change in a measurand. It is of fundamental importance to understand that valid application of RCV, and the derived probability curves for any measurand, requires consideration of the magnitude and constancy of CVA; RCV being a function of two variables, CV_I and CV_A. If applying RCVs clinically, the user must first ascertain whether the CV_I value used in the formula reflects that of the population to which the RCV is to be applied. Secondly, the CV_A utilised must reflect that of the method used to deliver the result locally, ideally meeting the APS goal for imprecision. The curves, an example of which is shown in Figure 5 are then only generally applicable across a range of measurand values if the CV_A is constant across that range. If the imprecision profile of the analytical system shows significant variance along the length of the calibration curve, then the RCV becomes a measurand/magnitude dependent variable. Modern laboratory information management systems

could be programmed to appropriately to account for imprecision changes to deliver RCV in real time and used to flag change along with statistical significance. Transportability of RCVs across systems requires well defined and robust estimates of CV_I and equivalence of APS.

The CV_I estimates of free thyroid hormones (fT3 and fT4) appear lower than the estimates for total hormones (T3, T4) concentrations. Thyroid hormones circulate largely protein bound. Free hormone concentrations reflect an equilibrium between protein bound fractions, tissue metabolism and thyroid production. It is logical to assume that, as the free hormones are the biologically active fraction of the total and exert feedback control over thyroid hormone production, that they may be more tightly controlled via homeostatic mechanisms.

Limited availability of data for the thyroid antibody-related analyses leaves a requirement for more studies or again investigation of big data approaches to deliver robust BV data.

Updated APS based upon meta-analysis derived BV data are shown in Table 2. The meta-analysis derived CV_I deliver in general APS for imprecision (CV) that is lower than the mean CV_A estimate of the studies included in this systematic review, excepting TSH and TSI. The lower estimates of CV_I for both fT4 and fT3 deliver APS which may be considered challenging even for the most recent state of the art analytical methods.

Use of analytical systems of equivalent analytical performance and meeting the APS identified in Table 2, will facilitate transportability of RCV values across geography, time and analytical methods used.

In conclusion, this study has delivered a systematic review of published studies of BV of thyroid disease related measurands. An objective assessment of the quality of the studies published by application of the BiVAC has enabled the collation of data sets of sufficient quality to apply meta-analysis and of delivery well characterized estimates of BV for some, but not all measurands studied and reviewed. For some important analytes there were few studies of measurands required for patient care in the many diverse and complex clinical scenarios requiring laboratory testing for healthcare. The review, and subsequent meta-analysis, of experimentally derived data sets, has enabled derivation of BV data estimates that are suitable for the many applications of the data; those including the setting of APS and determination of significance of change in serial results. It is important that manufacturers consider the significance of the new APS (e.g. fT4 and fT3) as it may challenge the performance requirements of their current devices [36].

However, of the many papers graded only the EuBIVAS study fully complied with BIVAC to be classified as Grade A quality. A potentially high degree of complexity has been identified in the BV of the thyroid biomarkers reviewed. Chronobiological factors, the unavailability of BIVAC grade A studies, and the requirement for BV applicable to the many clinical scenarios across varying time scales delivers challenges. Big data approaches for definition may provide an alternative to the stringent experimental approach to definition of BV data required. The prevalence of thyroid testing in populations delivers large volumes of well-characterized data in clinical laboratory databases to support this approach.

Research funding: None declared.
Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.
Competing interests: Authors state no conflict of interest.
Informed consent: Informed consent was obtained from all individuals included in this study.
Ethical approval: The local Institutional Review Board deemed the study exempt from review.

References

1. Vanderpump MPJ. Epidemiology of thyroid disorders. In: Luster M, Duntas LH, Wartofsky L, editors. The thyroid and its diseases: a comprehensive guide for the clinician. Cham: Springer International Publishing; 2019:75–85 pp.
2. Ladenson PW, Singer PA, Ain KB, et al. American Thyroid Association guidelines for detection of thyroid dysfunction. Arch Intern Med 2000;160:1573–5.
3. Bunch DR, Firmender K, Harb R, El-Khoury JM. First- and second-trimester reference intervals for thyroid function testing in a US population. Am J Clin Pathol 2020. https://doi.org/10.1093/ajcp/aqaa165 [Epub ahead of print].
4. Zhang Y, Wu W, Liu Y, Guan Y, Wang X, Jia L. The impact of TSH levels on clinical outcomes 14 days after frozen-thawed embryo transfer. Pregnancy Childbirth 2020;20:677.
5. Abbas W, Adam I, Rayis DA, Hassan NG, Lutfi MF. Thyroid hormones profile among obese pregnant Sudanese women. J Clin Transl Res 2020;8:14–9. [eCollection 16 Jul 2020].
6. Murillo-Llorente M, Fajardo-Montañana C, Pérez-Bermejo M, Vila-Candel R, Gómez-Vela J, Velasco I. Intra-individual variability in TSH levels of healthy women during the first half of pregnancy. Endocrinol Diabetes Nutr 2017;64:288–94.
7. Punda A, Škrabić V, Torlak V, Gunjača I, Boraska Perica V, Kolčić I, et al. Thyroid hormone levels are associated with metabolic components: a cross-sectional study. Croat Med J 2020;61:230–8.
8. Brenta GJ. The association between obesity and the thyroid: is the "Chicken or the Egg" conundrum finally solved? Clin Endocrinol

Metabol 2021. https://doi.org/10.1210/clinem/dgab291 [Epub ahead of print].
9. Mahdavi M, Amouzegar A, Mehran L, Madreseh E, Tohidi M, Azizi F. Investigating the prevalence of primary thyroid dysfunction in obese and overweight individuals: tehran thyroid study. BMC Endocr Disord 2021;21:89.
10. Feldt-Rasmussen U, Klose M. In: clinical strategies in the testing of thyroid function. In: Feingold KR, Anawalt B, Boyce A, Chrousos G, de Herder WW, Dungan K, et al., editors. Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc.; 2020.
11. Caiulo S, Corbetta C, Di Frenna M, Medda E, De Angelis S, Rotondi D, et al. Newborn screening for congenital hypothyroidism: the benefit of using differential TSH cutoffs in a two-screen program. J Clin Endocrinol Metab 2020. https://doi.org/10.1210/clinem/dgaa789 [Epub ahead of print].
12. Geno KA, Reed MS, Cervinski MA, Nerenz RD. Evaluation of thyroid function in pregnant women using automated immunoassays. Clin Chem 2021;67:772–80.
13. Ceriotti F, Fernandez-Calle P, Klee GG, Nordin G, Sandberg S, Streichert T, et al. Criteria for assigning laboratory measurands to models for analytical performance specifications defined in the 1st EFLM Strategic Conference. Clin Chem Lab Med 2017;55:189–94.
14. Aarsand AK, Fernandez-Calle P, Webster C, Coskun A, Gonzales-Lao E, Diaz-Garzon J, et al. The EFLM biological variation database. Available from: https://biologicalvariation.eu/ [Accessed 31 May 2021].
15. Aarsand AK, Røraas T, Fernandez-Calle P, Ricos C, Díaz-Garzón J, Jonker N, et al. The biological variation data critical appraisal checklist: a standard for evaluating studies on biological variation. European federation of clinical chemistry and laboratory medicine working group on biological variation and task and finish group for the biological variation database. Clin Chem 2018;64:501–14.
16. Røraas T, Støve B, Petersen PH, Sandberg S. Biological variation: the effect of different distributions on estimated within-person variation and reference change values. Clin Chem 2016;62:725–6.
17. Fraser CG, Harris EK. Generation and application of data on biological 11 variation in clinical chemistry. Crit Rev Clin Lab Sci 1989;27:409–37.
18. Díaz-Garzón J, Fernández-Calle P, Minchinela J, Aarsand AK, Bartlett WA, Aslan B, et al. Biological variation data for lipid cardiovascular risk assessment biomarkers. A systematic review applying the biological variation data critical appraisal checklist (BIVAC). Clin Chim Acta 2019;495:467–75.
19. González-Lao E, Corte Z, Simón M, Ricós C, Coskun A, Braga F, et al. Systematic review of the biological variation data for diabetes related analytes. European federation of clinical chemistry and laboratory medicine working group on biological variation and task group for the biological variation database. Clin Chim Acta 2019;488:61–7.
20. Fraser CG, Sandberg S. Biological variation. In: Rifai N, Horvath AR, Wittwer CT, editors. Tietz textbook of clinical chemistry and molecular biology, 6th ed.; 2017:157–70 pp.
21. Bottani M, Aarsand AK, Banfi G, Locatelli M, Coşkun A, Díaz-Garzón J, et al. European Biological Variation Study (EuBIVAS): within- and between-subject biological variation estimates for serum thyroid biomarkers based on weekly samplings from 91 healthy participants EUBIVAS Thyroid. Clin Chem Lab Med 2021 [Epub ahead of print].
22. Mairesse A, Wauthier L, Courcelles L, Luyten U, Burlacu MC, Maisin D, et al. Biological variation and analytical goals of four thyroid function biomarkers in healthy European volunteers. Clin Endocrinol 2021;94:845–50.
23. Feldt-Rasmussen U, Petersen PH, Blaabjerg O, Horder M. Long-term variability in serum thyroglobulin and thyroid related hormones in healthy subjects. Acta Endocrinol 1980;95:328–34.
24. Ankrah-Tetteh T, Wijeratne S, Swaminathan R. Intraindividual variation in serum thyroid hormones, parathyroid hormone and insulin-like growth factor-1. Ann Clin Biochem 2008;45:167–9.
25. Thyrotropin/thyroid stimulating hormone (TSH) measurement. Medscape; 2003. Available from: https://www.medscape.com/viewarticle/452667_4 [Accessed 15 May 2021].
26. Liyanage YSH, Siriwardhana ID, Dissanayake M, Dayanath BKPT. Study on diurnal variation in TSH and freeT4 levels of healthy adults. Sri Lanka J Diabetes Endocrinol Metabol 2018;8:8–16. [Accessed 31 Jul 2021].
27. Kuzmenko NV, Tsyrlin VA, Pliss MG, Galagudza MM. Seasonal variations in levels of human thyroid-stimulating hormone and thyroid hormones: a meta-analysis. Chronobiol Int 2021;38:301–17.
28. Sviridonova MA, Fadeyev VV, Sych YP, Melnichenko GA. Clinical significance of TSH circadian variability in patients with hypothyroidism. Endocr Res 2013;38:24–31.
29. Maes M, Mommen K, Hendrickx D, Peeters D, D'Hondt P, Ranjan R, et al. Components of biological variation, including seasonality, in blood concentrations of TSH, TT3, FT4, PRL, cortisol and testosterone in healthy volunteers. Clin Endocrinol 1997;46:587–98.
30. Oladipo O, Nenninger DA, Parvin CA, Dietzen DJ. Intraindividual variability of thyroid function tests in a pediatric population. Clin Chim Acta 2010;411:1143–5.
31. Karmisholt J, Andersen S, Laurberg P. Analytical goals for thyroid function tests when monitoring patients with untreated subclinical hypothyroidism. Scand J Clin Lab Invest 2010;70:264–8.
32. Loh TP, Ranieri E, Metz MP. Derivation of pediatric within-individual biological variation by indirect sampling method. Am J Clin Pathol 2014;142:657–63.
33. Jones GRD. Estimates of within-subject biological variation derived from pathology databases: an approach to allow assessment of the effects of age, sex, time between sample collections, and analyte concentration of reference change values. Clin Chem 2019;65:579–88.
34. LaFranchi S. Sick-euthyroid syndrome. Decision Support in Medicine LLC; 2013. Available from: https://www.cancertherapyadvisor.com/home/decision-support-in-medicine/pediatrics/sick-euthyroid-syndrome/. [Accessed 21 Jul 2021].
35. Coşkun A, Sandberg S, Unsal I, Cavusoglu C, Serteser M, Kilercik M, et al. Personalized reference intervals in laboratory medicine: a new model based on within-subject biological variation. Clin Chem 2021;67:374–84.
36. Giovannini S, Zucchelli GC, Iervasi G, Iervasi A, Chiesa MR, Mercuri A, et al. Multicentre comparison of free thyroid hormones immunoassays: the Immunocheck study. Clin Chem Lab Med 2011;49:1669–76.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/cclm-2021-0904).