

Michela Bottani, Aasne K. Aarsand, Giuseppe Banfi, Massimo Locatelli, Abdurrahman Coşkun, Jorge Díaz-Garzón, Pilar Fernandez-Calle, Sverre Sandberg, Ferruccio Ceriotti and Anna Carobene*, on behalf of the European Federation of Clinical Chemistry and Laboratory Medicine Working Group on Biological Variation

European Biological Variation Study (EuBIVAS): within- and between-subject biological variation estimates for serum thyroid biomarkers based on weekly samplings from 91 healthy participants

<https://doi.org/10.1515/cclm-2020-1885>

Received December 29, 2020; accepted January 25, 2021;
published online February 9, 2021

Abstract

Objectives: Thyroid biomarkers are fundamental for the diagnosis of thyroid disorders and for the monitoring and treatment of patients with these diseases. The knowledge of biological variation (BV) is important to define analytical performance specifications (APS) and reference change values (RCV). The aim of this study was to deliver BV

estimates for thyroid stimulating hormone (TSH), free thyroxine (FT4), free triiodothyronine (FT3), thyroglobulin (TG), and calcitonin (CT).

Methods: Analyses were performed on serum samples obtained from the European Biological Variation Study population (91 healthy individuals from six European laboratories; 21–69 years) on the Roche Cobas e801 at the San Raffaele Hospital (Milan, Italy). All samples from each individual were evaluated in duplicate within a single run. The BV estimates with 95% CIs were obtained by CV-ANOVA, after analysis of variance homogeneity and outliers.

Results: The within-subject (CV_I) BV estimates were for TSH 17.7%, FT3 5.0%, FT4 4.8%, TG 10.3, and CT 13.0%, all significantly lower than those reported in the literature. No significant differences were observed for BV estimates between men and women.

Conclusions: The availability of updated, in the case of CT not previously published, BV estimates for thyroid markers based on the large scale EuBIVAS study allows for refined APS and associated RCV applicable in the diagnosis and management of thyroid and related diseases.

Keywords: analytical performance specification; biological variation; reference change value; thyroid.

*Corresponding author: Anna Carobene, Laboratory Medicine, IRCCS San Raffaele Scientific Institute, Via Olgettina 60, 20132 Milan, Italy, Phone: +39 02 26432850, E-mail: carobene.anna@hsr.it

Michela Bottani, IRCCS Istituto Ortopedico Galeazzi, Laboratory of Experimental Biochemistry & Molecular Biology, Milan, Italy

Aasne K. Aarsand, Department of Medical Biochemistry and Pharmacology, Haukeland University Hospital, Bergen, Norway; and Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), Haraldsplass Deaconess Hospital, Bergen, Norway

Giuseppe Banfi, IRCCS Istituto Ortopedico Galeazzi, Laboratory of Experimental Biochemistry & Molecular Biology, Milan, Italy; and Vita-Salute San Raffaele University, Milan, Italy

Massimo Locatelli, Laboratory Medicine, IRCCS San Raffaele Scientific Institute, Milan, Italy

Abdurrahman Coşkun, School of Medicine, Acibadem Mehmet Ali Aydınlar University, Istanbul, Turkey

Jorge Díaz-Garzón and Pilar Fernandez-Calle, Hospital Universitario La Paz, Madrid, Spain; and Quality Analytical Commission of Spanish Society of Clinical Chemistry (SEQC), Barcelona, Spain

Sverre Sandberg, Department of Medical Biochemistry and Pharmacology, Haukeland University Hospital, Bergen, Norway; Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), Haraldsplass Deaconess Hospital, Bergen, Norway; and Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway

Ferruccio Ceriotti, Clinical Laboratory, Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico, Milan, Italy. <https://orcid.org/0000-0002-0958-5354>

Introduction

Thyroid pathologies represent the most common endocrine disorders after diabetes. Laboratory tests, together with imaging methods, are fundamental for the diagnosis and management of these disorders in clinical practice [1]. Thyroid stimulating hormone (TSH) determination is recommended as a first screening test for the presence of overt and subclinical thyroid disorders, followed by free thyroxine (FT4), and free triiodothyronine (FT3) in patients who are suspected to have hyperthyroidism, as a second

step [1, 2]. Human calcitonin (CT) is mainly used as a marker for the presence and monitoring of medullary thyroid carcinoma (MTC), while thyroglobulin (TG) is mostly of value in the management of differentiated thyroid carcinoma (DTC) patients after treatment [1, 2].

Availability of reliable biological variation (BV) data [3] for these central thyroid markers is important to define analytical performance specifications (APS) [4, 5], and necessary to establish the reference change value (RCV), that may be used as a tool for monitoring patients in assessing what changes between two measurements can be explained by biological and analytical variation [6]. Global within-subject (CV_I) and between-subject (CV_G) BV estimates are available in the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) BV Database [7]. To deliver these estimates by meta-analysis, BV studies are appraised using the Biological Variation Data Critical Appraisal Checklist (BIVAC), which is based on 14 Quality Items (QIs), each of which is associated with a score *A*, *B*, *C*, and/or *D* [8]. BIVAC-compliant studies that satisfy the meta-analysis inclusion criteria (healthy adults, biweekly to monthly sampling, more than two samples per participant) are used to deliver the global CV_I and CV_G estimates reported in the database. For some thyroid biomarkers several studies of varying quality have been published, however, they are small-scale studies unable to assess e.g., differences between genders, and for CT, no BV data have been published previously.

The EFLM Working Group on BV (BV-WG) designed the European Biological Variation Study (EuBIVAS), which is a BIVAC compliant BV study that satisfies all the pre-analytical requirements of the critical appraisal checklist for BV studies [9, 10]. The EuBIVAS is a highly powered, large-scale study, involving 91 healthy subjects from five European countries [11, 12]. The aim of the present study was to deliver updated BV data, together with the derived APS and RCVs for TSH, FT4, FT3, TG, and CT.

Materials and methods

The EuBIVAS study

For this project, six different laboratories situated in five European countries (Italy (Milan and Padua), Norway, Spain, Turkey, and the Netherlands) enrolled 91 healthy persons, consisting of two sex-separated subgroups consisting of 53 women and 38 men. At the first visit, the state of well-being of the enrolled subjects was evaluated by a questionnaire, allowing evaluation of their family and medical history, and lifestyle habits. Only subjects who fulfilled the inclusion criteria, described in [11], underwent phlebotomy for 10 consecutive weeks (April–June 2015); of these 91 participants, 77 completed all 10

collections, 10 completed nine, two completed eight, and two completed seven. Fasting blood were collected in serum tubes with clot activator (plastic, silicone coated, 10 mL (16 × 100 mm) [Becton Dickinson, USA, code 367820] and kept at room temperature (min. 30 min – max. 2 h). Samples were centrifuged at 3,000 g for 10 min, aliquoted, stored at –80 °C, and shipped frozen to the coordinating center (San Raffaele Hospital). Detailed information on the laboratories, exclusion/inclusion criteria for subjects' enrollment, and protocols for sample collection, handling, and storage, have been previously published in detail [11].

The EuBIVAS protocol was approved by the Institutional Ethical Review board of San Raffaele Hospital in agreement with the World Medical Association Declaration of Helsinki and by the Ethical board/regional Ethics Committee for each involved center. Informed consent was signed by all enrolled subjects.

Analytical method

All the analyses were performed at the San Raffaele Hospital (Milan, Italy) on the Roche Cobas e801 using the Immunoassay Electrochemiluminescence (ECLIA) reagents and calibrators as described in Supplemental Table 1. For each subject, samples were analyzed in duplicate in the same analytical run.

Data analysis

Data analyses were performed as previously described [13–15]. Briefly, to obtain CV_A (analytical variation) and CV_I BV estimates, data were analyzed using CV-ANOVA, an ANOVA method based on the CV-transformation of data [16]. The identification and removal of outliers were performed prior to analysis in order to achieve CV_A and CV_I homogeneity. The first was verified using the Bartlett test; the second with the Cochran test. The steady state of the participants was evaluated using the linear regression on the 10 pooled mean group sample concentrations for each measurand [17]. ANOVA was performed on the natural log-transformed data to deliver CV_G estimates. The presence of outliers between subjects was assessed by using the Dixon q-test, while normality was examined by the Shapiro–Wilk test.

The mean concentrations and the BV components, with the 95% CI, of each measurand were estimated for the whole study population and separately for men, women in fertile age below 50 years ($n=43$) and women in menopausal age above 50 years ($n=10$). For each measurand, differences in the mean concentrations, CV_I , and CV_G between subgroups were considered significant if the 95% CI did not overlap. When significant differences were observed between CV_I of men and women (overall group or <50 years), the lower of the two was chosen for APS calculation. If mean concentrations between men and women were significantly different, the lower of the two CV_G was applied in the APS [12].

APSs for the analytical imprecision (CV_{APS}), analytical bias (B_{APS}) were calculated according to:

$$CV_{APS} = \frac{1}{2} CV_I$$

$$B_{APS} = 0.25 \sqrt{CV_I^2 + CV_G^2}$$

RCV were estimated for an increase and a decrease using the formulas below, applying CV_A estimates based on duplicate measurement of study samples from all subjects:

$$\begin{aligned}
 SD_{A,\log}^2 &= \log_e(CV_A^2 + 1) \\
 SD_{I,\log}^2 &= \log_e(CV_I^2 + 1) \\
 SD_{\text{combined},\log} &= \sqrt{SD_{A,\log}^2 + SD_{I,\log}^2} \\
 RCV\% &= 100\% \times (e^{(\pm Z_{\alpha} \times \sqrt{2} \times SD_{\text{combined},\log})} - 1)
 \end{aligned}$$

where $Z_{\alpha}=1.65$ for the probability level of significant change set at 95%.

In addition, the number of samples required to calculate the homeostatic set point (NHSP) was calculated as follows:

$$\text{NHSP} = \left(Z \times \sqrt{CV_A^2 + CV_I^2} / D \right)^2$$

where $Z=1.96$ ($p<0.05$) and D is the desired percentage closeness to the homeostatic set point (set at 10 and 15%).

Data were analyzed using Excel 2016, XLSTAT (Statistical software for Excel), and IBM SPSS Statistic (version 20).

Results

Detailed information on participants' characteristics has been published [18]. Briefly, 20 subjects reported to be smokers, the physical activity and the alcohol intake was low-to-moderate, and the use of medications was limited. The median body mass index (BMI), median age, number, and gender of the included subjects are summarized in Supplemental Table 2. For all the five measurands, two women <50 years (one from Norway, one from Spain) and two women >50 years (both from Turkey) were eliminated from the study because of their TSH results being above the upper reference interval for our assay (>4.8 mIU/L), and they were later diagnosed with a thyroid disease, in line with these subjects having subclinical thyroid disease at the time of sampling for the EuBIVAS.

A woman from The Netherland with a mean TSH value of 5.0 mIU/L throughout the 10 EuBIVAS collections, was not excluded because retesting showed that she presently

has not signs of thyroid diseases, with normal thyroid hormones, negative anti-Thyroid peroxidase antibody and anti-TSH receptor.

No differences were found between mean values and BV estimates in the two female subgroups (females below and above 50 years), therefore only results from the overall female group are reported.

In the case of CT, 22 subjects (two men and 20 women) were not included in the data analysis because their results were below the limit of quantitation (LOQ), while for TG, one man was excluded for the same reason. Considering the data for the whole study population after exclusion of the subjects described above, 2.7% of TSH results, 1.9% of FT3, 0.7% of FT4, 7.3% of TG, and 1.8% of CT were excluded to achieve CV_I and CV_A homogeneity. In addition, for TG three subjects were identified as outliers (two men and one woman) by the Dixon-q test, and thus not included in the CV_G estimate. The information about the detection and exclusion of outliers are summarized in the Supplemental Table 3. For all the measurands, the mean concentration of all samples did not change during the 10 weeks of the study (data not shown). Based on the Shapiro-Wilk test, the ln-transformed data for the whole population and for male and female subgroups were normally distributed. The results for each individual for TSH, FT4, FT3, CT, and TG, ordered by sex/age, are shown in Figures 1–5, respectively. Significant differences in concentrations were found among subjects from different countries for TSH (Turkey and the Netherland vs. the other countries) and TG (Turkey vs. the other countries) (Supplemental Figures 1 and 2).

The BV estimates of the whole study population and for the male and female subgroups are reported in Table 1. In Table 2, CV_{APS} and B_{APS} , RCVs, and NHSPs are presented for each measurand.

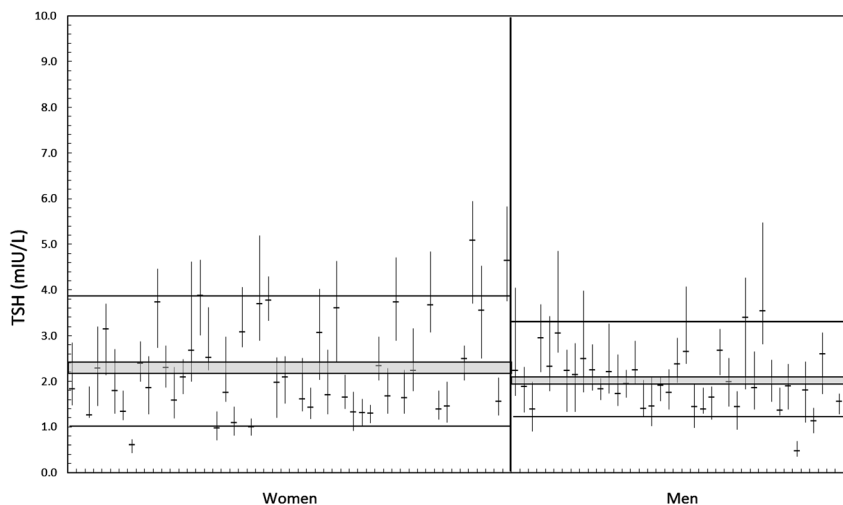


Figure 1: TSH according to age/sex. Median values (horizontal bars) and range (minimum-maximum) of thyroid stimulating hormone (TSH) for each individual ordered by sex and age after outliers' exclusion. Median age and range were 35 (22–59) years and 39 (21–69) years for males and females, respectively. The 95% CI of the mean, the 5th and the 95th percentiles are indicated by continuous lines for both men and women.

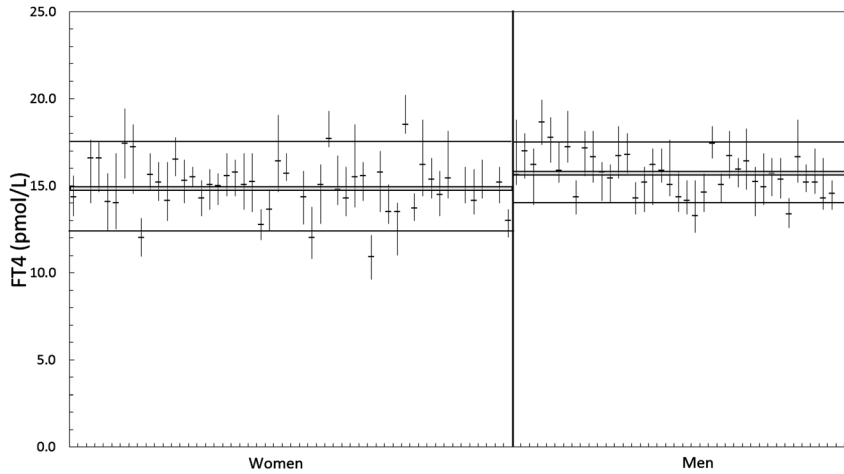


Figure 2: FT4 according to age/sex. Median values (horizontal bars) and range (minimum-maximum) of free thyroxine (FT4) for each individual ordered by sex and age after outliers' exclusion. Median age and range were 35 (22–59) years and 39 (21–69) years for males and females, respectively. The 95% CI of the mean, the 5th and the 95th percentiles are indicated by continuous lines for both men and women.

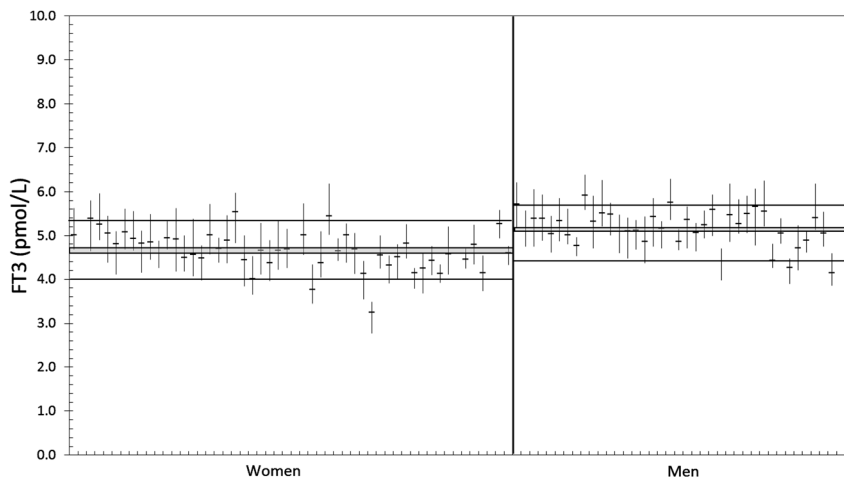


Figure 3: FT3 according to age/sex. Median values (horizontal bars) and range (minimum-maximum) of free triiodothyronine (FT3) for each individual ordered by sex and age after outliers' exclusion. Median age and range were 35 (22–59) years and 39 (21–69) years for males and females, respectively. The 95% CI of the mean, the 5th and the 95th percentiles are indicated by continuous lines for both men and women.

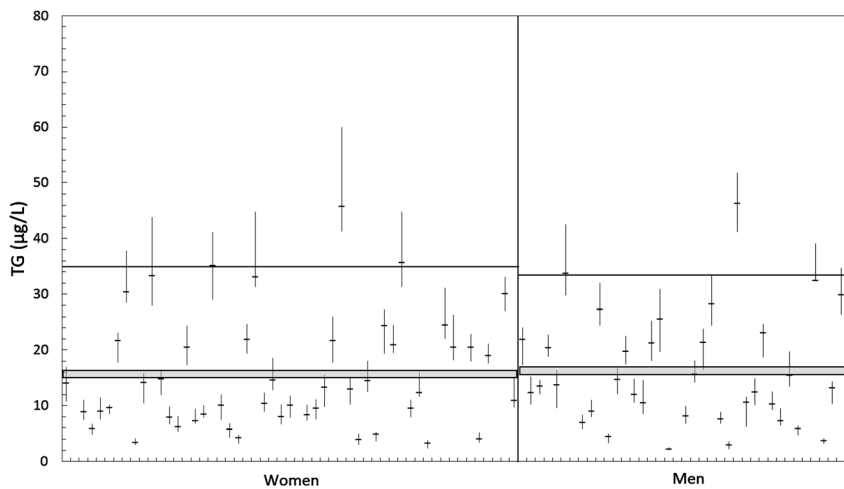


Figure 4: TG according to age/sex. Median values (horizontal bars) and range (minimum-maximum) of thyroglobulin (TG) for each individual ordered by sex and age after outliers' exclusion. Median age and range were 35 (22–59) years and 39 (21–69) years for males and females, respectively. The 95% CI of the mean, the 5th and the 95th percentiles are indicated by continuous lines for both men and women.

The mean serum concentration for TSH was significantly higher in women than in men; while FT3, FT4, and CT serum concentrations were significantly higher in men compared to women (see Table 1). Thus, for these four

measurands the lowest CV_G estimate was applied in the APS calculation, while for TG the overall CV_G estimate was used (as highlighted in Table 1). Results for CV_I estimates were similar between men and women, thus the overall CV_I

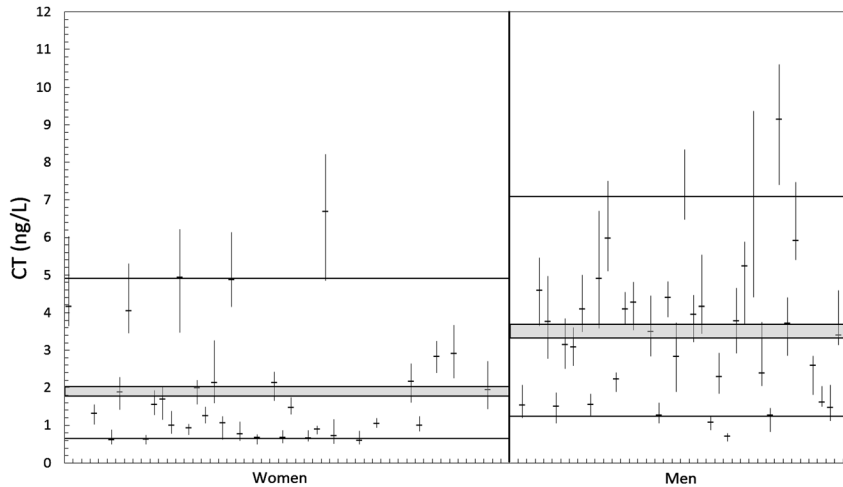


Figure 5: CT according to age/sex. Median values (horizontal bars) and range (minimum-maximum) of calcitonin (CT) for each individual ordered by sex and age after outliers' exclusion. Median age and range were 35 (22–59) years and 39 (21–69) years for males and females, respectively. The 95% CI of the mean, the 5th and the 95th percentiles are indicated by continuous lines for both men and women.

Table 1: EuBIVAS within-subject (CV_i) and between-subject (CV_G) biological variation estimates^a, with 95% CIs, for thyroid stimulating hormone (TSH), free triiodothyronine (FT3), free thyroxine (FT4), thyroglobulin (TG), and calcitonin (CT).

		Number of individuals	Total number of results	Mean number of samples/individuals	Mean number of replicates/samples	Mean value (95% CI)	CV_A , % (95% CI) ^b	CV_i , % (95% CI)	CV_G , % (95% CI)
TSH, mIU/L	All	85	1,568	9.33	1.96	2.20 (2.14–2.25)	1.46 (1.39–1.54)	17.7 (16.8–18.7)	
	M	38	702	9.37	1.94	2.05 (1.99–2.11)		18.9 (17.5–20.5)	35.9 (28.9–48.0)
	W	47	866	9.30	1.96	2.31 (2.23–2.40)		16.8 (15.7–18.1)	46.3 ^c (38.9–62.1)
FT3, pmol/L	All	86	1,579	9.29	1.95	4.85 (4.84–4.90)	1.77 (1.70–1.80)	5.0 (4.8–5.3)	
	M	38	691	9.24	1.94	5.15 (5.11–5.19)		4.6 (4.3–5.0)	8.0 (6.5–10.5)
	W	48	888	9.33	1.96	4.64 (4.61–4.68)		5.3 (5.0–5.7)	9.8 (8.0–12.3)
FT4, pmol/L	All	86	1,597	9.40	1.95	15.3 (15.2–15.4)	1.71 (1.60–1.80)	4.8 (4.5–5.0)	
	M	38	697	9.32	1.94	15.7 (15.6–15.8)		4.5 (4.2–4.9)	7.5 (6.1–9.9)
	W	48	900	9.46	1.97	14.9 (14.8–15.1)		5.0 (4.6–5.3)	10.2 (8.4–12.7)
TG, µg/L	All	86	1,557	9.10	1.98	15.9 (15.3–16.5)	3.74 (3.56–3.94)	10.3 (9.8–10.9)	79.0^c (66.3–98.3)
	M	37	681	9.27	1.97	16.5 (15.6–17.4)		9.3 (8.5–10.1)	80.5 (62.5–117.6)
	W	49	865	8.88	1.98	15.6 (14.8–16.4)		10.7 (10.0–11.5)	78.8 ^c (62.6–105.8)
CT, ng/L	All	68	1,297	9.56	1.99	2.77 (2.65–2.89)	3.84 (3.60–4.10)	13.0 (12.3–13.9)	
	M	36	684	9.53	1.99	3.52 (3.36–3.69)		12.7 (11.7–13.8)	65.8^c (51.6–92.0)
	W	32	613	9.59	1.99	1.93 (1.79–2.08)		13.4 (12.3–14.7)	79.0 ^c (60.1–115.9)

^aResults were assessed for men and women. Results in bold were used to estimate APS and RCV for the whole population in Table 2. ^bAnalytical variation (CV_A) estimates were based on CV-ANOVA of duplicate analysis of all study samples. ^cNot normal distribution for original data, normal if ln transformed.

Table 2: EuBIVAS analytical performance specification (APS) for imprecision (CV_{APS}) and bias (B_{APS}), reference change value (RCV), and the number of samples required to estimate the homeostatic set point (NHSP) for thyroid stimulating hormone (TSH), free triiodothyronine (FT3), free thyroxine (FT4), thyroglobulin (TG), and calcitonin (CT) based on the EuBIVAS BV estimates as reported in Table 1.

	$CV_{APS}\%$ ^a	$B_{APS}\%$ ^b	RCV% ^c (Decrease; Increase)	NHSP ^d 10%	NHSP ^d 15%
TSH	8.9	10.0	-33.6; 50.7	13	6
FT3	2.5	2.4	-11.6; 13.1	2	1
FT4	2.4	2.2	-11.2; 12.6	1	1
TG	5.2	20.0	-22.5; 29.0	5	3
CT	6.5	16.8	-27.0; 36.4	8	4

^a $CV_{APS} = \frac{1}{2} CV_I$, ^b $B_{APS} = 0.25(CV_I^2 + CV_G^2)^{0.5}$. ^cRCV were calculated as described in the text delivering asymmetric values for rise and fall at the probability level of 95% for significant unidirectional change, applying CV_A estimates based on duplicate measurement of all study samples. ^dNHSP ($Z^*(CV_A^2 + CV_I^2)^{1/2}/D$)² where D is the allowed percentage deviation from the true homeostatic set point, and Z is 1.96 (for a p-value <0.05). NHSPs associated with 10%, and 15% deviations from the true homeostatic set points are calculated.

estimates were used for RCV and NHSP calculation (Table 2).

Discussion

The current clinical guidelines for thyroid diseases diagnosis and management [19–21] outline serum TSH as the most sensitive marker of overt or subclinical thyroid diseases. With abnormal TSH results, FT4 should be examined, especially for hypothyroidism, and/or FT3, in the case of hyperthyroidism and thyrotoxicosis. TSH concentrations are modified even by minimal changes in the thyroid hormones (THs) levels and altered TSH concentrations occur during hyperthyroidism or hypothyroidism onset even before free THs changes [2]. However, serum TSH levels may be influenced by numerous factors, such as ethnicity, gender, age [1, 2, 22] and pregnancy; for the latter specific reference interval related to age of gestation are required [23, 24]. Diurnal and circannual variations in TSH levels have also been reported [22], and smoking, obesity and use of medications or dietary supplements (as a source of excessive iodine intake for some individuals) may affect TSH levels [22–24]. Therefore, it is not surprising that TSH mean concentrations showed significant differences between EuBIVAS participants from the different countries (see Supplementary Figure 1). We have previously reported lifestyle differences between the country-defined EuBIVAS study cohorts [18]. This includes a high number of smokers and the highest mean BMI in the Turkish cohort [18], which

may explain the higher TSH results that were found in this group. Thus, the TSH CV_G estimates and APS for bias based on the EuBIVAS data may not be applicable to all populations and should be used with caution.

The current TSH reference interval, especially the upper limit, is still under debate [22, 25, 26], due to the possible high prevalence of subclinical thyroid diseases that makes it difficult to identify truly healthy subjects. In our study, we excluded four subjects who had a TSH levels above our laboratory's upper reference limit (>4.8 mIU/L) and who later were diagnosed with thyroid disease, in line with subclinical disease at the time of sampling. 22 subjects had results between 2.5 and 4.8 mIU/L, but all other thyroid biomarkers were normal, and these subjects were not excluded. The lack of harmonization between different analytical assays for TSH, FT4, and FT3 also influences the reference intervals in use in laboratories and limits the clinical applicability of the above cited current guidelines for thyroid diseases diagnosis and management. In this context, the possibility to obtain the personal reference interval, as recently published by Coskun et al. [27], might be a step forward in advancing personalized medicine. The International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) Working Group for Standardization of Thyroid Function Tests is presently working on the standardization and harmonization of TSH, FT3, and FT4 measurements [28, 29].

In the EuBIVAS population, significant differences in mean concentrations of TSH, FT4, and FT3 between men and women were observed (see Table 1), as has previously been described only for TSH [30]. The EuBIVAS-based APS for imprecision (CV_{APS}) (Table 2) are lower than those as of now reported in the EFLM BV database (10.6, 3.9, and 3.0% for TSH, FT4, and FT3, respectively) [7]. EuBIVAS has delivered high-quality data on BV, using a large cohort of samples collected under optimal conditions, fully BIVAC compliant [8]. As a consequence, the global estimates reported in the EFLM DB [7], will be inevitably influenced by the new meta-analysis calculation after the inclusion of the EuBIVAS based estimates. It should be recalled that the APSs are used for both internal quality control and for external quality assurance (EQA) [4]. Thus, if EQA providers recalculate their APSs based on new and tighter data, it will be more difficult for participating laboratories to satisfy the new quality specification. Generally, EQA providers should calculate their APSs in the same way and use the EFLM BV database as their source for BV data [31, 32]. The CV_A estimates for imprecision obtained in our study are clearly lower than the CV_{APS} . This indicates that the analytical system applied in our study likely is within these requirements. However, considering the EuBIVAS

strict pre-analytical protocol and that the CV_A estimates are based on duplicate analysis of all study samples, a routine setting with long-term analysis may give rise to CV_A estimates exceeding the EuBIVAS APS.

Until now 10, seven, and five papers have been published and included in the EFLM database for TSH, FT4, and FT3, respectively [7]. In addition, another study with BV estimates for TSH, FT4, FT3, and TG has recently been published by Mairesse et al. [33]. The EuBIVAS provides lower BV estimates than those previously published, which delivers not only stricter APS, but also smaller RCV, which would identify smaller changes as being within the expected variation caused by analytical and biological variation. If using the TSH, FT4, and FT3 CV_A and CV_I estimates derived from the EuBIVAS, the RCVs (95% probability for a significant unidirectional change) were 50.7, 12.6, and 13.1%, respectively (see Table 2). In clinical practice, this means that for a subject with 2.5 mIU/L of TSH, an increase to 3.8 mIU/L in consecutive TSH result could be explained simply by biological and analytical variations. This can also explain a subject with 16.7 pmol/L an increase to 18.8 pmol/L, and with 5.4 pmol/L an increase to 6.1 pmol/L, for FT4 and FT3, respectively, in an adult patient. However, it is important to bear in mind that results within the RCV does not rule out that such a change is of clinical importance. It is important to take into account that the EuBIVAS RCVs were obtained using the specific EuBIVAS CV_A based on analysis of duplicate samples, and, for this reason, cannot be considered as universal values thus underlining that each laboratory has to calculate its own RCVs using relevant CV_A estimates.

Another information for the management of thyroid diseases is the calculation of the number of samples required to obtain reliable estimates of the TSH, FT4, and FT3 homeostatic set points. Using the EuBIVAS BV estimates for TSH, 13 samples are required to obtain an approximation of 10%, and six for a 15% approximation, in line with the higher variability observed for TSH, which underlines the challenges in interpreting TSH results. The limit of 10 and 15%, arbitrarily chosen, hint clinician that, considering the cost benefit, depending on the patient needs might decide the number of requested tests. For FT4 and FT3, on the other hand, only one sample is sufficient for the 15% approximation (see Table 2). Robust BV data for TSH may play an important role to improve the TSH measurement specificity and reliability for screening patients with subclinical/overt thyroid diseases, for monitoring patients under therapy and, in some cases, for the identification of patients with an increased risk of bone fractures, thyroid cancer, and cardiovascular diseases. Reliable BV estimates can also be a fundamental tool for clinicians in

the correct identification of subjects who require substitutive or suppressive TSH therapy, thus underlining that an adequate training about the usefulness of BV data in the clinical decisions is of importance for endocrinologists, and for giving recommendations on how many repeated samples are required be taken to obtain the patient's set point.

TG is an important DTC tumor marker that is used to evaluate the effectiveness of treatment and for monitoring patients with recurrent disease after partial or total thyroid ablation. Moreover, increased serum TG concentrations can also be found in subjects with other thyroid diseases, such as thyroiditis, increased thyroid activity, goiter, or Graves' disease [4, 34]. The EuBIVAS CV_I estimate for TG (all subjects: 10.3% (95% CI; 9.8–10.9)) is lower than the estimates recently reported by Mairesse et al. (15.4%, 95% CI; 13.3–18.3) [33] and by Feldt-Rasmussen, even if for the latter, 12.8%, is not significantly different based on evaluation of the 95% CI; 10.0–17.1 [35]. EuBIVAS, being highly powered and fully BIVAC-compliant, delivers lower RCVs and stricter APS than those based on non-BIVAC-compliant studies with clinical implications both for monitoring patients and ensuring that analytical methods are fit for purpose. However, it must be kept in mind that a high number of results had to be excluded to obtain homogeneity of the data. The EuBIVAS based RCV is 29.0% (95% probability for significant unidirectional change) (see Table 2), this means that for a subject with a TG baseline value of 35.2 $\mu\text{g/L}$, an increase to 45.4 $\mu\text{g/L}$ could be explained by biological and analytical variations alone. TG measurement is a useful marker for patient monitoring after DTC tumor removal and treatment [34], and the increase of TG in serial consecutive measurements during follow up seems to be more informative than only one TG evaluation after treatment [36, 37]. The NHSP calculated for TG based on the EuBIVAS data is five samples with a 10% approximation and three with 15% approximation. Thus, BV estimates can be influential to improve the TG measurement specificity for monitoring patients with thyroid diseases, potentially also for DTC patients during the follow up. However, it is important to keep in mind that our data have been derived from a healthy population and are based on weekly samplings, whereas for monitoring of thyroid diseases, less frequent samplings intervals are indicated. Moreover, TG measurements post thyroidectomy, may not have a “homeostatic set point” due to expected growth of tumor cells and may have different physiology and so caution should be used in these RCVs.

Even though there is no consensus about the use of CT as a screening method for patients with thyroid nodules for MTC diagnosis or with hereditary MTC forms (about 25% of MTC), some data indicates its potential as a sensitive prognostic and diagnostic marker in both pre- and

post-operative phases of thyroid C cells diseases, especially for MTC [36]. The lack of agreement is especially due to the fact that, in case of diagnostic uncertainty, the CT screening results requires the association with CT stimulation assay with pentagastrin or calcium gluconate as confirmatory tests to increase the CT specificity. For adults, values >1,000 ng/L of CT revealed a positive predictive value of 100% for MTC and are thus highly suspicious for MTC presence, while CT values <10 ng/L indicate no C cell diseases, and values >20 ng/L for basal CT are positive predictive values in the diagnosis of MTC [38]. In the EuBIVAS population, circulating levels of CT were significantly lower in women than in men and about 24% of the subjects had CT values lower than the LOQ of our assay, which are both characteristics already described in literature [39]. To the best of our knowledge, no studies have reported BV estimates for CT. The CT RCV value, calculated based on EuBIVAS CV_A and CV_I data, was 36.4% (see Table 2). This indicates that an increase in CT results from 15.0 to 20.5 ng/L could be explained simply by biological and analytical variations, considering an RCV with 95% probability for significant unidirectional change. The CT NHSPs were eight and four with approximation of 10 and 15% respectively, indicating that for CT more than one measurement could be needed for clinical decision if the concentration is near the action limit. Due to the lack of BV studies for CT, the EuBIVAS BV estimates together with RCVs and NHSPs are of fundamental importance for correct interpretation of CT results, and useful to create more specific range of basal CT concentrations for the healthy population. In addition, these results will be useful to create cut-offs when monitoring MTC patients after surgery or in response to therapy.

Study limitations

The analyses were performed using only one manufacturer's reagents, but it is unlikely that this will affect the BV estimates for the same measurand. The samples were continuously stored at $-80\text{ }^\circ\text{C}$ and thawed only prior to analysis but were kept at room temperature for from 30 min to a maximum of 2 h before analyses were performed. Data are derived from a healthy adult population, and the use of this data, especially for RCVs applied to diseased populations must be done with caution. Furthermore, BV estimates from the EuBIVAS were based on weekly samplings, which must be taken into consideration when used in clinical situations where samplings would be far less frequent, especially for measurands which may be influenced by season, such as for TSH. Sample collection

for the EuBIVAS was standardized, with all samples being drawn between 08.00 and 10.00 am. Considering the diurnal variation of especially TSH [22], this must be taken into account when using our BV estimates for deriving RCVs.

Conclusions

Our study delivers updated, and for CT not previously available, BV based estimates for thyroid markers based on the large scale EuBIVAS study. No differences in CV_I estimates between men and women or between countries were observed, but differences in mean TSH concentrations between different country-specific subgroups. The EuBIVAS based RCVs, NHSP, and APS may be instrumental in assuring quality of the laboratory analysis of thyroid markers and in the diagnosis and monitoring of overt or subclinical thyroid pathologies, follow-up of thyroid tumor patients.

Acknowledgments: We thank Roche for kindly providing us with the kits necessary for carrying out the study. We would also like to thank study participants and EuBIVAS partners for their contributions: Gerhard Barla, William Bartlett, Elena Guerra, Thomas Røraas, Niels Jonker, Mario Plebani, Una Ørvim Sølvik, Marit Sverresdotter Sylte, Mustafa Serteser, Francesca Tosato, and Ibrahim Unsal.

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 EuBIVAS protocol was approved by the Institutional Ethical Review board of San Raffaele Hospital in agreement with the World Medical Association Declaration of Helsinki and by the Ethical board/regional Ethics Committee for each involved center. Informed consent was signed by all enrolled subjects.

References

1. Soh SB, Aw TC. Laboratory testing in thyroid conditions – pitfalls and clinical utility. *Ann Lab Med* 2019;39:3–14.
2. Esfandiari NH, Papaleontiou M. Biochemical testing in thyroid disorders. *Endocrinol Metab Clin N Am* 2017;46:631–48.
3. Carobene A. Reliability of biological variation data available in an online database: need for improvement. *Clin Chem Lab Med* 2015; 53:871–7.

4. Sandberg S, Fraser CG, Horvath AR, Jansen R, Jones G, Oosterhuis W, et al. Defining analytical performance specifications: consensus statement from the 1st strategic conference of the European Federation of Clinical Chemistry and Laboratory Medicine. *Clin Chem Lab Med* 2015;53:833–5.
5. Aarsand AK, Røraas T, Bartlett WA, Coşkun A, Carobene A, Fernandez-Calle P, et al. Harmonization initiatives in the generation, reporting and application of biological variation data. *Clin Chem Lab Med* 2018;56:1629–36.
6. Fraser CG. Reference change values. *Clin Chem Lab Med* 2011;50: 807–12.
7. Aarsand AK, Fernandez-Calle P, Webster C, Coskun A, Gonzales-Lao E, Díaz-Garzón J, et al. EFLM biological variation database. Available from: <https://biologicalvariation.eu/> [Accessed Dec 2019].
8. 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. *Clin Chem* 2018;64:501–14.
9. Carobene A, Guerra E, Locatelli M, Ceriotti F, Sandberg S, Fernandez-Calle P, et al. Providing correct estimates of biological variation-not an easy task. The example of S100- β protein and neuron-specific enolase. *Clin Chem* 2018;64:1537–9.
10. Carobene A, Guerra E, Marqués-García F, Boned B, Locatelli M, Coşkun A, et al. Biological variation of morning serum cortisol: updated estimates from the European biological variation study (EuBIVAS) and meta-analysis. *Clin Chim Acta* 2020;509: 268–72.
11. Carobene A, Strollo M, Jonker N, Barla G, Bartlett WA, Sandberg S, et al. Sample collections from healthy volunteers for biological variation estimates' update: a new project undertaken by the working group on biological variation established by the European Federation of Clinical Chemistry and Laboratory Medicine. *Clin Chem Lab Med* 2016;54:1599–608.
12. Carobene A, Aarsand AK, Guerra E, Bartlett WA, Coşkun A, Díaz-Garzón J, et al. European Biological Variation Study (EuBIVAS): within- and between-subject biological variation data for 15 frequently measured proteins. *Clin Chem* 2019;65:1031–41.
13. Carobene A, Marino I, Coşkun A, Serteser M, Unsal I, Guerra E, et al. The EuBIVAS project: within- and between-subject biological variation data for serum creatinine using enzymatic and alkaline picrate methods and implications for monitoring. *Clin Chem* 2017;63:1527–36.
14. Aarsand AK, Díaz-Garzón J, Fernandez-Calle P, Guerra E, Locatelli M, Bartlett WA, et al. The EuBIVAS: within- and between-subject biological variation data for electrolytes, lipids, urea, uric acid, total protein, total bilirubin, direct bilirubin, and glucose. *Clin Chem* 2018;64:1380–93.
15. Carobene A, Guerra E, Locatelli M, Cucchiara V, Briganti A, Aarsand AK, et al. Biological variation estimates for prostate specific antigen from the European Biological Variation Study; consequences for diagnosis and monitoring of prostate cancer. *Clin Chim Acta* 2018;486:185–91.
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: 725–36.
17. Ceriotti F, Díaz-Garzón Marco J, Fernández-Calle P, Maregnani A, Aarsand AK, Coskun A, et al. The European Biological Variation Study (EuBIVAS): weekly biological variation of cardiac troponin I estimated by the use of two different high-sensitivity cardiac troponin I assays. *Clin Chem Lab Med* 2020;58:1741–7.
18. Bottani M, Banfi G, Guerra E, Locatelli M, Aarsand AK, Coşkun A, et al. European Biological Variation Study (EuBIVAS): within- and between-subject biological variation estimates for serum biointact parathyroid hormone based on weekly samplings from 91 healthy participants. *Ann Transl Med* 2020;8:855.
19. Garber JR, Cobin RH, Gharib H, Hennessey JV, Klein I, Mechanick JL, et al. Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and The American Thyroid Association. *Endocr Pract* 2012;18:988–1028.
20. 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. *Eur Thyroid J* 2015;4:149–63.
21. Ross DS, Burch HB, Cooper DS, Greenlee MC, Laurberg P, Maia AL, et al. 2016 American thyroid association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid* 2016;26:1343–421.
22. Razvi S, Bhana S, Mrabeti S. Challenges in interpreting thyroid stimulating hormone results in the diagnosis of thyroid dysfunction. *J Thyroid Res* 2019;2019:4106816.
23. Lazarus J, Brown RS, Daumerie C, Hubalewska-Dydejczyk A, Negro R, Vaidya B. 2014 European thyroid association guidelines for the management of subclinical hypothyroidism in pregnancy and in children. *Eur Thyroid J* 2014;3:76–94.
24. Alexander EK, Pearce EN, Brent GA, Brown RS, Chen H, Dosiou C, et al. 2017 Guidelines of the American thyroid association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid* 2017;27:315–89.
25. Brabant G, Beck-Peccoz P, Jarzab B, Laurberg P, Orgiazzi J, Szabolcs I, et al. Is there a need to redefine the upper normal limit of TSH? *Eur J Endocrinol* 2006;154:633–7.
26. Wartofsky L, Dickey RA. The evidence for a narrower thyrotropin reference range is compelling. *J Clin Endocrinol Metab* 2005;90: 5483–8.
27. 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* 2020. <https://doi.org/10.1093/clinchem/hvaa233> [Epub ahead of print].
28. Thienpont LM, Van Uytendange K, Beastall G, Faix JD, Ieiri T, Miller WG, et al. Report of the IFCC working group for standardization of thyroid function tests; part 1: thyroid-stimulating hormone. *Clin Chem* 2010;56:902–11.
29. Thienpont LM, Van Uytendange K, Beastall G, Faix JD, Ieiri T, Miller WG, et al. Report of the IFCC working group for standardization of thyroid function tests; part 2: free thyroxine and free triiodothyronine. *Clin Chem* 2010;56:912–20.
30. Andersen S, Bruun NH, Pedersen KM, Laurberg P. Biologic variation is important for interpretation of thyroid function tests. *Thyroid* 2003;13:1069–78.

31. Carobene A, Franzini C, Ceriotti F. Comparison of the results from two different external quality assessment schemes supports the utility of robust quality specifications. *Clin Chem Lab Med* 2011; 49:1143–9.
 32. Haeckel R, Wosniok W, Kratochvila J, Carobene A. A pragmatic proposal for permissible limits in external quality assessment schemes with a compromise between biological variation and the state of the art. *Clin Chem Lab Med* 2012;50:833–9.
 33. 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* 2020. <https://doi.org/10.1111/cen.14356> [Epub ahead of print].
 34. Algeciras-Schimnich A. Thyroglobulin measurement in the management of patients with differentiated thyroid cancer. *Crit Rev Clin Lab Sci* 2018;55:205–18.
 35. Feldt-Rasmussen U, Hyltoft Petersen P, Blaabjerg O, Hørder M. Long-term variability in serum thyroglobulin and thyroid related hormones in healthy subjects. *Acta Endocrinol* 1980;95: 328–34.
 36. Baudin E, Do Cao C, Cailleux AF, Leboulleux S, Travagli JP, Schlumberger M. Positive predictive value of serum thyroglobulin levels, measured during the first year of follow-up after thyroid hormone withdrawal, in thyroid cancer patients. *J Clin Endocrinol Metab* 2003;88:1107–11.
 37. Huang SH, Wang PW, Huang YE, Chou FF, Liu RT, Tung SC. Sequential follow-up of serum thyroglobulin and whole body scan in thyroid cancer patients without initial metastasis. *Thyroid* 2006;16:1273–8.
 38. Costante G, Meringolo D. Calcitonin as a biomarker of C cell disease: recent achievements and current challenges. *Endocrine* 2020;67:273–80.
 39. d’Herbomez M, Caron P, Bauters C, Do Cao C, Schlienger JL, Sapin R, et al. Reference range of serum calcitonin levels in humans: influence of calcitonin assays, sex, age, and cigarette smoking. *Eur J Endocrinol* 2007;157:749–55.
-
- Supplementary Material:** The online version of this article offers supplementary material (<https://doi.org/10.1515/cclm-2020-1885>).