

Variability of Thyroid Measurements from Ultrasound and Laboratory in a Repeated Measurements Study

Till Ittermann^a Adrian Richter^{a,d} Martin Junge^a Matthias Nauck^b
Astrid Petersmann^{b,e} Clemens Jürgens^a Harald Below^c
Carsten Oliver Schmidt^a Henry Völzke^a

^aInstitute for Community Medicine, University Medicine Greifswald, Greifswald, Germany; ^bInstitute for Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany; ^cInstitute of Hygiene and Environmental Medicine, University Medicine Greifswald, Greifswald, Germany; ^dGerman Rheumatism Research Center, Berlin, Germany; ^eInstitute for Clinical Chemistry, Interdisciplinary University Laboratory, University Medicine Göttingen, Göttingen, Germany

Keywords

Thyroid · Thyroid-stimulating hormone · Iodine · Thyroglobulin · Thyroid imaging · Variability · Measurement error

Abstract

Background: Variability of measurements in medical research can be due to different sources. Quantification of measurement errors facilitates probabilistic sensitivity analyses in future research to minimize potential bias in epidemiological studies. We aimed to investigate the variation of thyroid-related outcomes derived from ultrasound (US) and laboratory analyses in a repeated measurements study. **Subjects and Methods:** Twenty-five volunteers (13 females, 12 males) aged 22–70 years were examined once a month over 1 year. US measurements included thyroid volume, goiter, and thyroid nodules. Laboratory measurements included urinary iodine concentrations and serum levels of thyroid-stimulating hormone (TSH), free triiodothyronine (fT3), free thyroxine (fT4), and thyroglobulin. Variations in continuous

thyroid markers were assessed as coefficient of variation (CV) defined as mean of the individual CVs with bootstrapped confidence intervals and as intraclass correlation coefficients (ICCs). Variations in dichotomous thyroid markers were assessed by Cohen's kappa. **Results:** CV was highest for urinary iodine concentrations (56.9%), followed by TSH (27.2%), thyroglobulin (18.2%), thyroid volume (10.5%), fT3 (8.1%), and fT4 (6.3%). The ICC was lowest for urinary iodine concentrations (0.42), followed by fT3 (0.55), TSH (0.64), fT4 (0.72), thyroid volume (0.87), and thyroglobulin (0.90). Cohen's kappa values for the presence of goiter or thyroid nodules were 0.64 and 0.70, respectively. **Conclusion:** Our study provides measures of variation for thyroid outcomes, which can be used for probabilistic sensitivity analyses of epidemiological data. The low intraindividual variation of serum thyroglobulin in comparison to urinary iodine concentrations emphasizes the potential of thyroglobulin as marker for the iodine status of populations.

© 2020 The Author(s)
Published by S. Karger AG, Basel

Introduction

Variability of measurements in medical research and practice can be due to different sources, such as individuals, observers, readers, devices, or instruments and thus may be related to biological variation as well as to pre-analytical and analytical components of variation [1]. Such variability is not desired since it may increase the uncertainty of study findings. Estimating the magnitude of variability usually involves experiments of repeated measurements, e.g., in randomized blocked designs. However, such designs may be unsuitable for estimating the measurement error of human laboratory markers since intraindividual variability arises. Variations in individuals can occur due to diurnal, day-to-day, or seasonal variations. For example, individual urinary iodine excretion levels vary strongly from day to day and during the day because they mainly depend on previous food intake [2]. Diurnal and seasonal variations have also been reported for thyroid hormone levels [3–5]. Intra- and inter-individual biological variations for laboratory blood markers including thyroid-stimulating hormone (TSH), free thyroxine (fT4), free triiodothyronine (fT3), and thyroglobulin are reported in Ricós's table [6]. Observer and reader variations are particularly essential to evaluate ultrasound (US) techniques such as assessments of thyroid volume and thyroid nodules [7]. Likewise, US devices as well as laboratory measurement techniques may be prone to variations [8].

It is important to determine the measurement error of thyroid-related variables to estimate the potential bias in epidemiological and clinical studies and to provide measures of variation, which can be used for power calculations and probabilistic sensitivity analyses in future research. Against this background, we aimed to investigate the variation of thyroid-related variables derived from US and laboratory in a repeated measurements study, in which 25 individuals were examined twelve times within 1 year.

Subjects and Methods

The Study of Health in Pomerania (SHIP) repeated measurements study was conducted between June 2015 and July 2016 in the examination center of the SHIP located in Greifswald, Germany [9]. Twenty-five volunteers (13 females, 12 males) aged 22–70 years (mean age 44 years) were recruited from staff of the University Medicine Greifswald. Except for pregnancy there were no exclusion criteria applied in this study. Each of the participants was examined twelve times in intervals of approximately 1 month during the course of the study, which is comparable to previous stud-

ies on iodine variability [10]. Five individuals were exposed to thyroid medication intake during the course of the study; their individual progressions are shown in online supplementary Figure 1 (see www.karger.com/doi/10.1159/000507018 for all online supplementary material) and their medical histories are given in online supplementary Table 1. In one person levothyroxine was prescribed for 10 months at 250 mg and for the last 2 months of the study at 225 mg. In this individual we observed a strong decrease in serum TSH levels and increases in serum levels of fT3 and fT4 during the course of the study.

Age, sex, and sociodemographic data were assessed by questionnaire at the last visit of the participant. At each of the twelve examinations, thyroid US, blood sampling, and urine collection were performed. The examinations followed the same protocols as introduced in the population-based SHIP study [11, 12]. Thyroid US was performed with a portable device using a 13-MHz linear array transducer (Vivid-I; General Electrics, Frankfurt a.M., Germany). Thyroid volume was calculated as length \times width \times depth \times 0.479 (mL) for each lobe [13]. Goiter was defined as a thyroid volume >18 mL in women and >25 mL in men [14]. For thyroid volume all interobserver and interdevice variabilities showed mean differences (± 2 SD) of $<5\%$ ($<25\%$) [15]. Occurrence of thyroid nodules was defined as at least one nodular change >10 mm in US.

Blood samples were taken between 7 a.m. and 11 a.m. in Vacutainer[®] serum gel tubes (BD, Franklin Lakes, NJ, USA). Serum TSH, fT3, and fT4 concentrations were analyzed within 6 h after sampling in the central laboratory of the University Medicine Greifswald by a homogeneous, sequential, chemiluminescent immunoassay based on LOCI[®] technology (Dimension Vista[®] System Flex[®] reagent cartridge; Siemens Healthcare Diagnostics Inc., Newark, DE, USA). The analytical measuring range was 0.005–100 mIU/L for TSH, 1.3–103.0 pmol/L for fT4, and 0.77–46.1 pmol/L for fT3. During the course of the study, internal quality control samples, which were conducted three times a day, revealed coefficients of variation (CVs) of 2.6% for TSH, 2.1% for fT4, and 2.1% for fT3. Thyroglobulin concentration measurements were performed in one batch in August 2016 with a solid-phase, chemiluminescent immunometric assay with a measurement interval from 0.9 to 300 ng/mL on the IMMULITE 2000 (Siemens Healthcare Diagnostics Inc.).

Urinary iodine concentrations were measured from spot urine samples by a photometric procedure (Photometer ECOM 6122; Eppendorf, Hamburg, Germany) with Sandell and Kolthoff reaction modified by Wawschinek [8, 16]. During the course of the study, the interassay CV for iodine was 4.2%.

Statistical Analyses

Continuous thyroid markers were expressed as median and dichotomous thyroid markers as percentage for each visit. The overall median was calculated as the mean of the twelve medians at the visits. A standard deviation (SD) was calculated for the twelve medians in comparison to the overall median. Furthermore, intraclass correlation coefficients (ICCs, continuous data) were assessed by one-way analysis of variance and Cohen's kappa (dichotomous data). The ICCs explain the amount of the total variance, which is related to the difference between the participants. If the ICC is 1, the variance of the respective marker would be fully related to the difference between the participants and values within one participant would not differ across the visits. ICC agreements

Table 1. Median values and percentages of thyroid measurements stratified by visit

	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Visit 10	Visit 11	Visit 12	Overall median	SD	ICC (95% CI)/kappa	MD ± SD
Iodine, µg/L	106	114	106	111	99	121	187	122	77	75	105	82	109	28.16	0.42 (0.26–0.58)	26.0±17.5
Iodine-creatinine ratio, µg/g	111	116	113	177	79	117	117	112	110	102	90	69	109	25.51	0.27 (0.13–0.42)	26.1±25.1
TSH, mIU/L	1.44	1.58	1.58	1.23	1.41	1.16	1.62	1.54	1.51	1.34	1.30	1.50	1.43	0.14	0.64 (0.49–0.78)	0.09±0.06
ft4, pmol/L	12.1	12.9	12.6	12.7	12.5	12.1	12.5	12.1	12.0	12.4	12.5	12.9	12.4	0.30	0.72 (0.60–0.84)	0.21±0.14
ft3, pmol/L	4.41	4.40	4.45	4.64	4.31	4.33	4.26	4.42	4.20	4.39	4.28	4.25	4.36	0.11	0.55 (0.40–0.71)	0.10±0.08
Thyroglobulin, ng/mL	14.1	13.9	14.0	10.9	18.2	17.1	12.8	12.8	14.1	13.0	12.2	11.6	13.7	2.02	0.90 (0.84–0.95)	0.85±0.62
Thyroid volume, mL	15.8	16.7	16.8	16.5	16.0	16.4	16.3	16.2	15.7	15.1	15.4	15.4	16.0	0.53	0.87 (0.81–0.94)	0.40±0.29
Goiter prevalence	16%	20%	20%	24%	24%	24%	24%	20%	20%	20%	20%	16%	20.7	2.75	0.64 (0.55–0.77)	–
Thyroid nodule prevalence	44%	48%	56%	52%	48%	48%	44%	48%	48%	50%	48%	46%	48.3	3.15	0.70 (0.59–0.77)	–

CI, confidence interval; ft3, free triiodothyronine; ft4, free thyroxine; ICC, intraclass correlation coefficient for continuous data; kappa, Cohen's kappa for dichotomous outcomes; MD, mean of the 66 mean differences between the visits and the corresponding SDs; SD, standard deviation of the medians of the twelve visits; TSH, thyroid-stimulating hormone.

were classified into the four groups “poor,” “good,” “very good,” and “excellent” according to the recommendations from Cicchetti [17]. Cohen's kappa explains the agreement of the twelve ratings at the twelve visits. For continuous variables, means of the 66 mean differences between the twelve visits and the belonging SDs were calculated. Homogeneity of variance between iodine and iodine-creatinine ratio was examined by a Levene test.

Furthermore, we calculated the individual CV defined as the individual SD of the twelve visits divided by the individual mean of the twelve visits for each of the 25 individuals and each biomarker. Individual CVs were plotted against the individual means. An overall CV was calculated for each biomarker as the mean of the CVs of the 25 individuals. Confidence intervals (CIs) were bootstrapped with 1,000 repetitions.

Seasonal variations of the thyroid biomarkers were assessed by linear regression models using fractional polynomials accounting for potential nonlinear variations over time. For this, we used fractional polynomials up to degree 2 with all possible combinations of powers selected from the set (−2, −1, −0.5, 0, 0.5, 1, 2, 3). Afterwards, these models were compared using the log likelihood to determine the best-fitting model. If none of the fractional polynomial models fitted the data significantly better than the linear model, a linear association was assumed. Time of blood sampling was associated with thyroid biomarkers by mixed models with random intercept. For graphical illustration we plotted serum TSH levels against time of blood sampling and showed the individual variations as well as the mean variations of serum TSH levels over time of blood sampling. For the mean variation we used linear regression models in combination with fractional polynomials. Correlations of urinary iodine excretion levels with the other thyroid biomarkers were calculated as R^2 derived from linear regression models for each visit and for the median over the twelve visits. All analyses were conducted with Stata 16.0 (Stata Corp., College Station, TX, USA).

Results

Urinary iodine was the thyroid biomarker which varied most widely over the twelve visits (Table 1; Fig. 1). Standardization to creatinine significantly reduced the variation of urinary iodine levels ($p < 0.001$). The ICC demonstrated a fair agreement of the iodine measurements between the twelve visits, resulting in a mean difference of 26 µg/L. Likewise, serum ft3 levels showed fair agreement across the measurements, with a slightly higher ICC than for iodine, while the mean difference was 0.10 pmol/L with an SD of 0.08 pmol/L. Serum measurements of TSH and ft4 showed good agreement over the twelve visits, with mean differences of 0.09 mIU/L and 0.21 pmol/L and SDs of 0.06 mIU/L and 0.14 pmol/L, respectively. Very good agreements across the visits were observed for serum thyroglobulin level and thyroid volume. For the dichotomous outcomes goiter and thyroid nodule presence, the agreement was good, with SDs for the respective prevalence of 2.75 and 3.15.

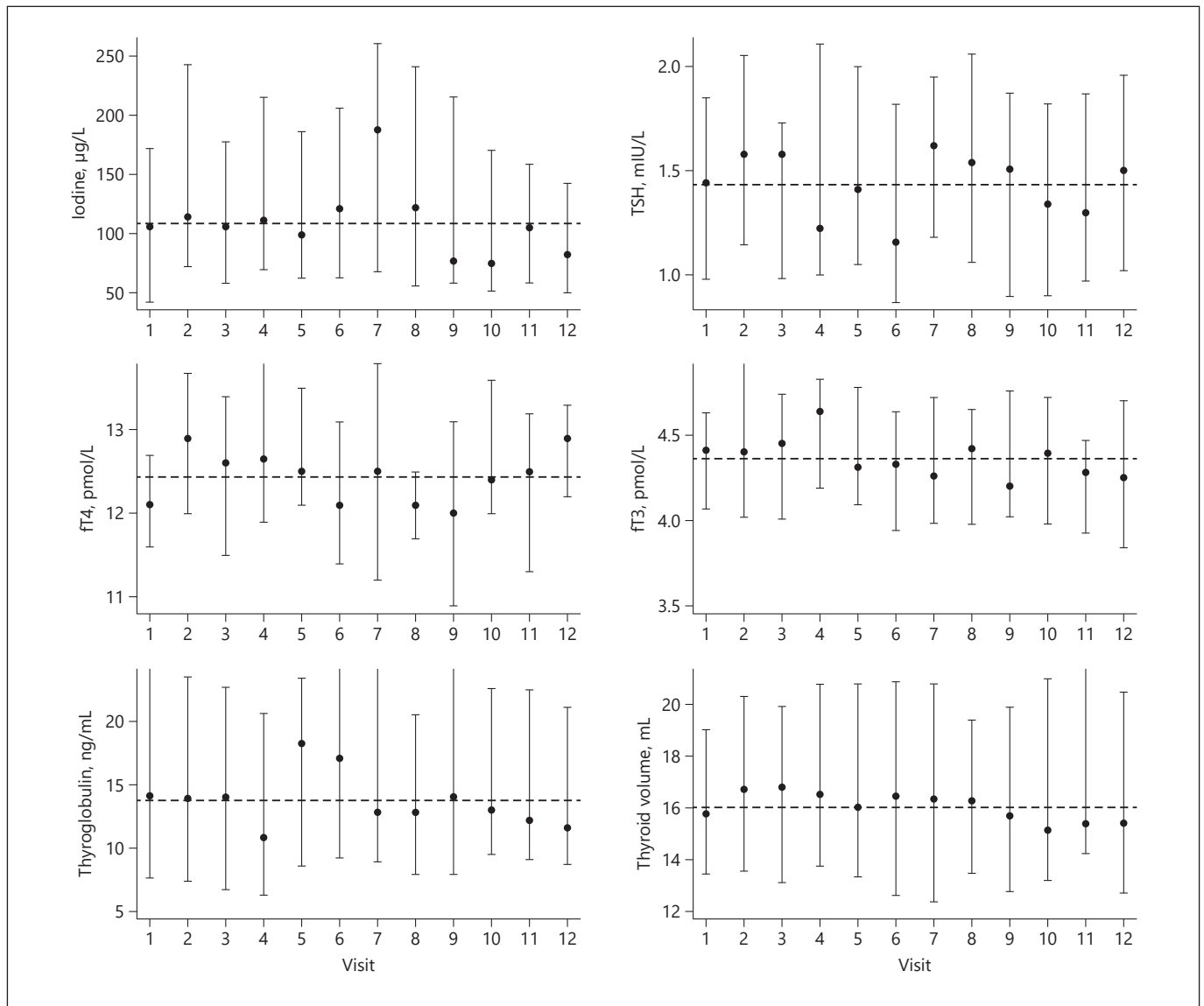


Fig. 1. Visit-specific medians (25th and 75th percentile) for continuous thyroid biomarkers. fT3, free triiodothyronine; fT4, free thyroxine; TSH, thyroid-stimulating hormone.

Figure 2 shows the individual CVs plotted against the individual means over the twelve visits. The mean CVs of these 25 individuals were 56.9% (95% CI 52.1–61.7%) for iodine, 27.2% (95% CI 17.4–37.1%) for TSH, 6.3% (95% CI 5.4–7.2%) for fT4, 8.1% (95% CI 7.0–9.1%) for fT3, 18.2% (95% CI 12.9–23.4%) for thyroglobulin, and 10.5% (95% CI 9.4–11.7%) for thyroid volume in the total population. Except for iodine-creatinine ratio, CVs were lower in individuals not taking thyroid medication and higher in those taking thyroid medication (Table 2). The intraindividual CVs of the serum markers in the 20 individuals

not on thyroid medication were within the 95% CIs reported in Ricós’s table [6].

Intraindividual progressions of thyroid biomarkers over the twelve visits are shown in Figure 3. In contrast to urinary iodine levels, serum thyroglobulin levels were relatively stable in most of the 25 individuals. Likewise, thyroid volume was also relatively constant over the 12 months, even though we observed some variations, particularly in individuals with larger thyroid glands. There was a trend towards higher urinary iodine levels in the winter months, but this effect was not statistically signifi-

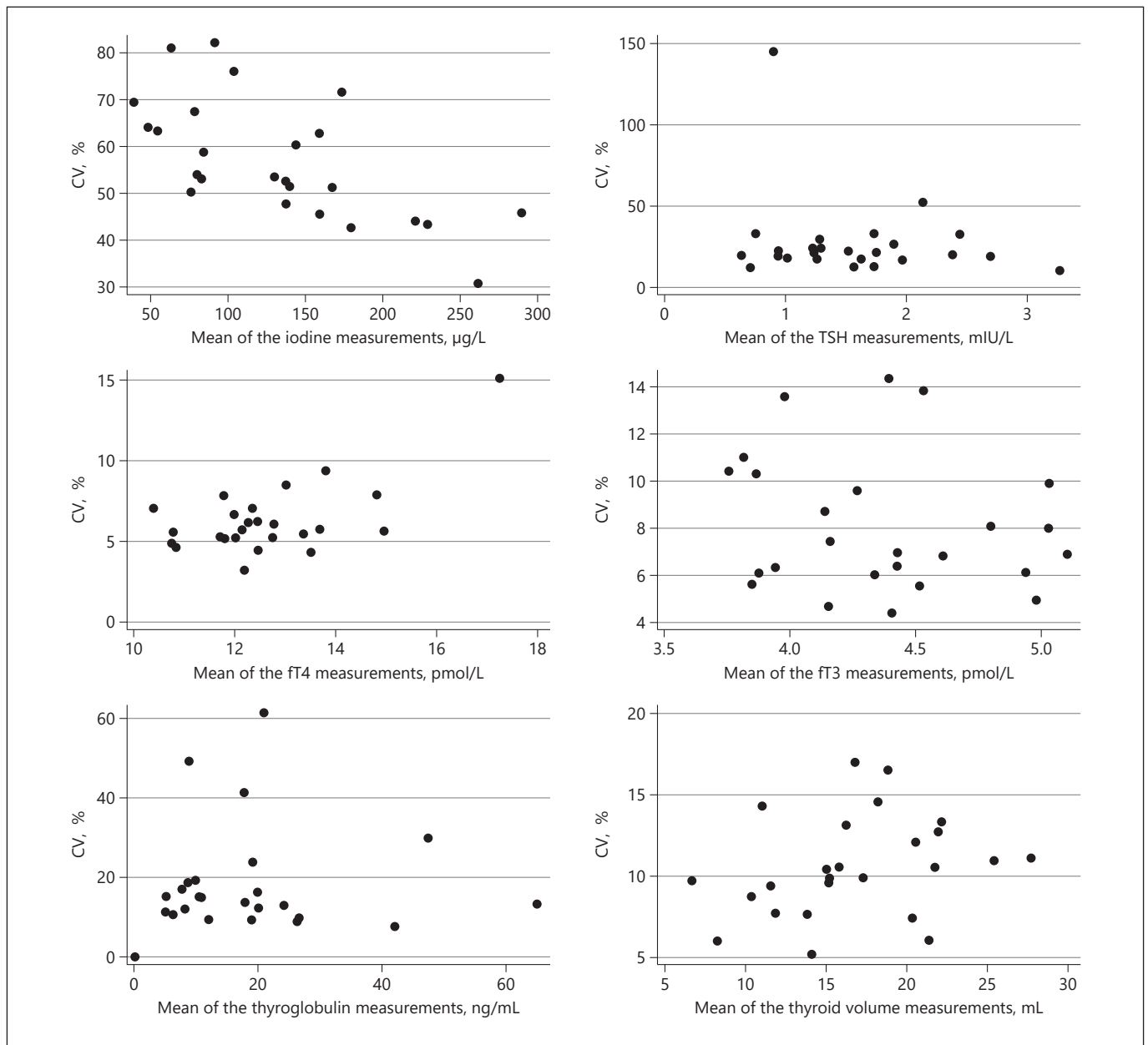


Fig. 2. Individual CVs plotted against the individual means. CV, coefficient of variation; fT3, free triiodothyronine; fT4, free thyroxine; TSH, thyroid-stimulating hormone.

cant ($p = 0.202$). Likewise, we observed no significant variations over the four seasons for the other thyroid biomarkers (data not shown).

Urinary iodine levels did not correlate strongly with any of the other thyroid biomarkers (Table 3). The strongest correlation was found with serum fT3 levels. The correlations of median individual urinary iodine levels over the twelve visits with individual median levels of the oth-

er thyroid biomarkers were not higher than the correlations at the single visits. Likewise, there were only weak correlations of mean iodine levels in the first half of the study, with mean levels of TSH ($R^2 = 0.028$), fT3 ($R^2 = 0.086$), fT4 ($R^2 = 0.007$), and thyroglobulin ($R^2 = 0.104$) in the second half of the study. Correlations of urinary iodine levels with thyroid biomarkers were even weaker when standardizing by urinary creatinine.

Table 2. Intraindividual CVs for thyroid biomarkers

	CV in % (95% CI)			Ricós's table [6]
	total population (<i>n</i> = 25)	population not on thyroid medication (<i>n</i> = 20)	population on thyroid medication (<i>n</i> = 5)	
Iodine, µg/L	56.9 (52.1–61.7)	56.1 (50.7–61.6)	60.0 (49.4–70.6)	
Iodine-creatinine ratio, µg/g	41.4 (37.5–45.3)	43.7 (39.9–47.5)	32.4 (23.2–41.6)	
TSH, mIU/L	27.2 (17.4–37.1)	20.3 (17.6–23.0)	55.0 (12.8–97.3)	15.9 (14.8–25.1)
fT4, pmol/L	6.3 (5.4–7.2)	5.7 (5.2–6.2)	8.8 (5.9–11.8)	7.7 (4.8–9.5)
fT3, pmol/L	8.1 (4.4–14.3)	7.6 (6.5–8.7)	10.1 (7.8–12.3)	6.0 (4.7–7.9)
Thyroglobulin, ng/mL	18.2 (12.9–23.4)	15.6 (12.0–19.2)	28.4 (8.1–48.7)	12.8 (10.0–17.1)
Thyroid volume, mL	10.5 (9.4–11.7)	10.4 (9.0–11.7)	11.4 (9.2–13.7)	

CI, confidence interval; CV, coefficient of variation; fT3, free triiodothyronine; fT4, free thyroxine; TSH, thyroid-stimulating hormone.

Table 3. Correlation of urinary iodine levels with thyroid biomarkers

	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Visit 10	Visit 11	Visit 12	Median over the 12 visits
TSH, mIU/L	0.04	0.07	0.03	0.03	0.02	0.03	0.08	0.08	0.01	0.01	0.06	0.01	0.05
fT4, pmol/L	0.01	0.06	0.01	0.02	0.03	0.00	0.10	0.03	0.01	0.01	0.18	0.01	0.04
fT3, pmol/L	0.00	0.08	0.07	0.10	0.00	0.06	0.44	0.00	0.12	0.03	0.11	0.07	0.11
Thyroglobulin, ng/mL	0.01	0.14	0.05	0.24	0.19	0.04	0.01	0.00	0.02	0.02	0.01	0.26	0.06
Thyroid volume, mL	0.12	0.02	0.02	0.01	0.00	0.02	0.08	0.01	0.00	0.01	0.00	0.03	0.05

Data are given as R^2 derived from linear regression with the respective thyroid biomarker as outcome and urinary iodine as exposure. fT3, free triiodothyronine; fT4, free thyroxine; TSH, thyroid-stimulating hormone.

Time of blood sampling was inversely associated with serum TSH levels. By each hour of time of the day, serum TSH levels were on average lower by 0.11 mIU/L (95% CI –0.18 to –0.03; $p = 0.006$). In Figure 4 this association is illustrated combined with the individual variations in serum TSH levels over time of blood sampling. For urinary iodine concentrations ($p = 0.436$) and serum levels of fT3 ($p = 0.352$), fT4 ($p = 0.252$), and thyroglobulin ($p = 0.555$) no such associations were observed.

Discussion

We investigated variations of thyroid measurements from 25 individuals, who were examined twelve times over 1 year. We observed that serum thyroglobulin levels and thyroid volume were the most stable thyroid markers within individuals, while the largest intraindividual variations were observed for urinary iodine levels. A seasonal trend was not observed for any of the considered thyroid

biomarkers. We found only weak correlations of urinary iodine levels with the other thyroid biomarkers even after combination of all twelve measurements. Time of blood sampling was only associated with serum TSH levels, but not with the other thyroid biomarkers.

According to the World Health Organization, median urinary iodine levels in population-based studies are the recommended marker to establish the iodine status in specific regions and countries [18]. In our study we found large variations in urinary iodine levels. In line with this, a previous study from Greenland concluded that 982 individuals are needed to establish the median urinary iodine level on the population level with a precision of 5% [19]. On the individual level, iodine excretion levels vary mainly due to differences in food intake, fluid uptake, and perspiration [2, 20]. Also, seasonal variations of iodine content in dietary products may explain individual variations in urinary iodine levels. A recent study demonstrated large variations of iodine content in milk peaking in spring [21]. In our study we also observed a trend towards

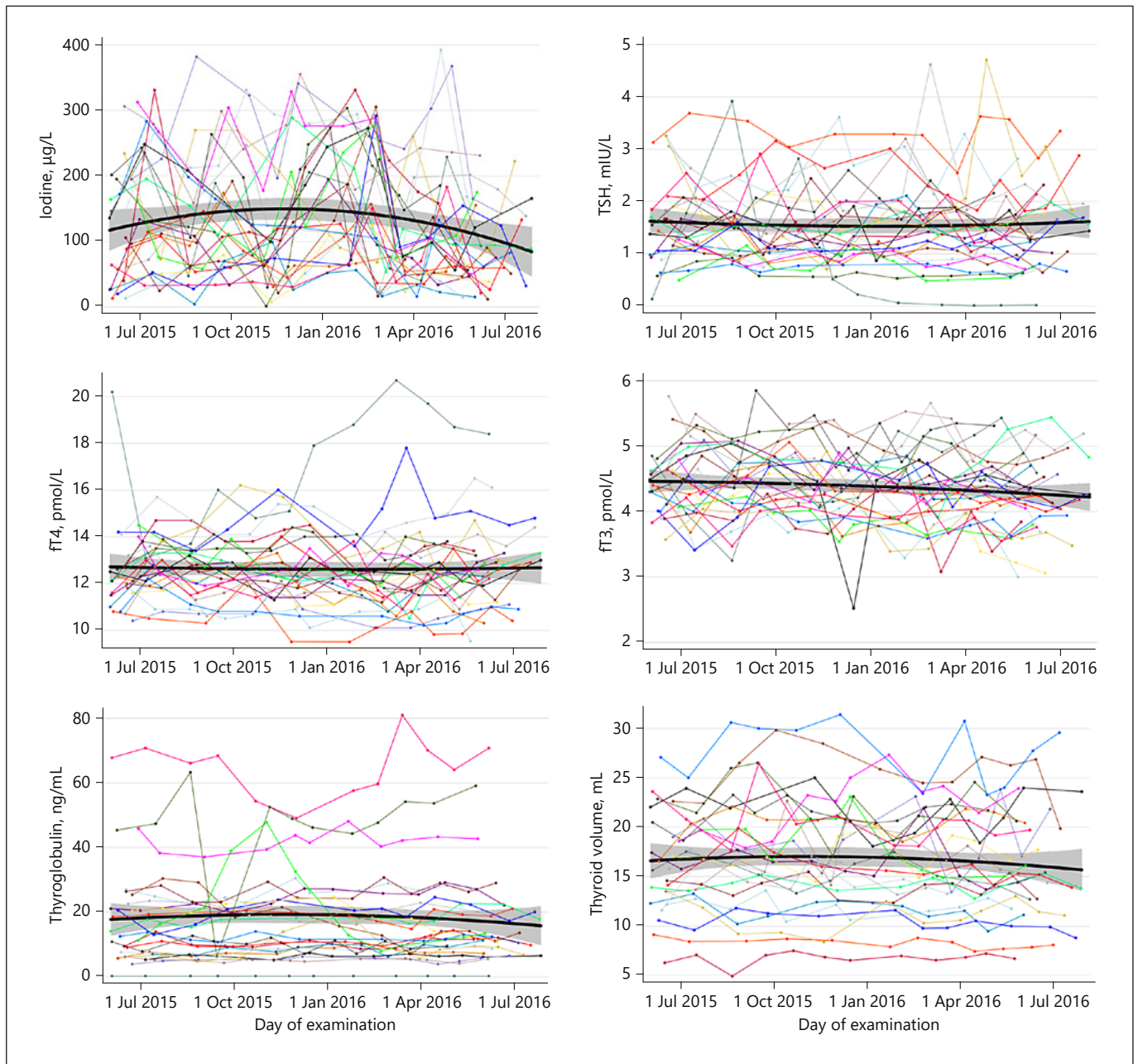


Fig. 3. Individual variation of thyroid biomarkers. fT3, free triiodothyronine; fT4, free thyroxine; TSH, thyroid-stimulating hormone.

a seasonal variation in urinary iodine levels, but this was not statistically significant.

Thyroglobulin, a protein important for the synthesis of thyroid hormones [20], has been discussed as a marker for iodine deficiency since thyroglobulin is more frequently released into the blood under iodine-deficient conditions [22, 23]. On the individual level, the variation of serum

thyroglobulin is much lower than that of urinary iodine. In individuals not on thyroid medication, the CV for thyroglobulin in our study was slightly higher compared to the CV reported in Ricós's table, but was within the 95% CI [6]. The Greenland study concluded that the number of samples for determination of an individuals' urinary iodine level is ten times higher than that for serum thyro-

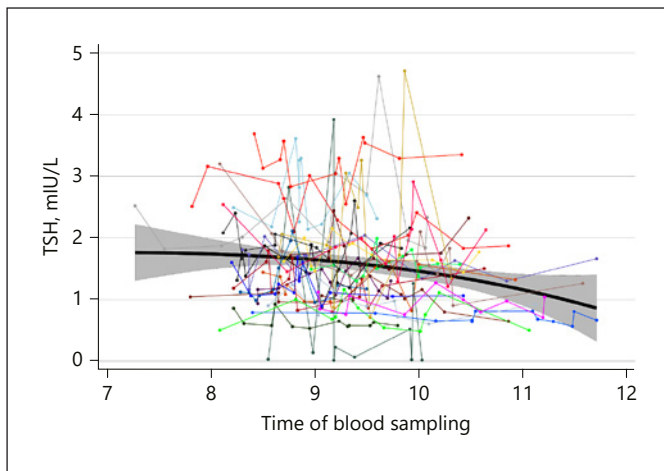


Fig. 4. Diurnal variation of serum TSH levels. TSH, thyroid-stimulating hormone.

globulin [19]. In line with this, the ICC in our analysis indicates that most of the variation of thyroglobulin is related to variation between individuals, demonstrating that serum thyroglobulin levels are quite stable within one individual over time. These findings stress the potential of thyroglobulin as a marker for iodine deficiency in the individual. However, in concordance with our study, previous data from 151 adults with 5 repeated measurements suggest only a weak correlation of serum thyroglobulin levels with urinary iodine concentrations [24]. This weak correlation may be explained by the high variation of urinary iodine within individuals. Even twelve measurements are not sufficient to determine an individual's iodine status [19]. The number of individuals included in our study meets the requirements for studies of variation [2, 25]. However, more individuals with different iodine supply may be needed to find stronger associations between urinary iodine and serum thyroglobulin concentrations.

Goiter and thyroid nodule prevalence varied both by around 3 percentage points over the twelve visits. The mean variation of thyroid volume over the twelve visits was 0.7 mL, with larger thyroids varying more than smaller ones. In our study we used two identical thyroid US devices using the ellipsoid method, which were regularly quality controlled. The main issue of thyroid US volumetry is that the whole thyroid does not fit on one US image, making it difficult for the observer to precisely measure the thyroid volume. This applies particularly for larger thyroids. Furthermore, differences in positioning the transducer may lead to different resolutions of the US images. Thus, the observer may be the main source of bias

for thyroid volume measurements. Compared to our results, much larger interobserver variations for thyroid volume have been described by Brauer et al. [7] based on data from 42 participants and 3 observers. In our study, we also had 3 observers, but the mean differences were aggregated over the twelve visits. Furthermore, our study was designed to clone the workflow and logistics of SHIP [9]. Thus, the reported variations in our study are a mixture of intra- and interobserver variations. This may explain the much smaller variation in our study compared to the one by Brauer et al. [7].

Similar to our findings, a circadian rhythm of TSH was reported in a study conducted in 13 individuals with blood samples drawn half-hourly over a period of 24 h [26] as well as in another study using 20,019 blood samples collected in an outpatient clinic over 3 years [3]. In both studies a peak of serum TSH levels until 9 a.m. in the morning was reported. This peak may be explained by an earlier pulsatile secretion of TSH from the hypothalamus, which is controlled by dopaminergic and somatostatinergic mechanisms [27]. For the other thyroid biomarkers we did not detect any diurnal variation over the morning, even though a study from the 1970s reported a diurnal variation of fT3 similar to the one reported for TSH [28]. That study, however, investigated diurnal variations only in 5 individuals, and the quality of fT3 assays has significantly improved since then [29].

The variations reported in our paper may be used for probabilistic sensitivity analyses in clinical studies. One strategy would be to randomly draw values from a normal distribution with a mean of 0 and an SD derived from the mean deviations reported in our analyses. For example, random values, derived from a normal distribution with a mean of 0 and an SD of 0.09, are added to the measured TSH value for each participant of a study. Analyses will then be repeated with the updated TSH variable. This procedure should be repeated at least 100 times (depending on the magnitude of the variation) to get an impression of the precision of the results. The results of these sensitivity analyses can be reported as the mean estimate of the 100 estimates derived from the 100 runs of the sensitivity analyses. CIs for the mean estimates can be calculated similar to how they are calculated in multiply imputed data sets [30].

Our study was performed in an examination center which has more than 20 years of experience in conducting thyroid US and epidemiological studies. Our laboratory is accredited and regularly participates in external quality programs. A limitation of our study is that our participants, recruited from staff of the University Medicine

Greifswald, were relatively healthy, even though 5 participants were under thyroid treatment. Potentially, variations of thyroid biomarkers may be higher in other population groups. However, in our study the variations in individuals not exposed to thyroid medication were comparable to those reported in Ricós's table [6] and in previous studies [10, 19]. Variations were higher in individuals exposed to thyroid medication, which may be related to variations in response to thyroid therapy. Particularly, in one individual exposed to 250 mg levothyroxine, we observed a decrease in TSH and increases in fT3 and fT4 levels during the course of the study.

Our study provides measures of variation for thyroid outcomes which can be used for probabilistic sensitivity analyses of epidemiological data. The low intraindividual variation of serum thyroglobulin in comparison to urinary iodine concentrations emphasizes the potential of thyroglobulin as marker for the iodine status of populations.

Acknowledgement

The authors thank the participants of the SHIP measurement error study as well as the staff of the examination center and of the Central Laboratory of the University Medicine Greifswald.

References

- 1 Montgomery DC. *Design and analysis of experiments*. 10th edition. New York: John Wiley & Sons; 2017.
- 2 Andersen S, Karmisholt J, Pedersen KM, Laurberg P. Reliability of studies of iodine intake and recommendations for number of samples in groups and in individuals. *Br J Nutr*. 2008 Apr;99(4):813–8.
- 3 Andersen IB, Brasen CL, Christensen H, Noehr-Jensen L, Nielsen DE, Brandslund I, et al. Standardised resting time prior to blood sampling and diurnal variation associated with risk of patient misclassification: results from selected biochemical components. *PLoS One*. 2015 Oct;10(10):e0140475.
- 4 Leppäluoto J, Sikkilä K, Hassi J. Seasonal variation of serum TSH and thyroid hormones in males living in subarctic environmental conditions. *Int J Circumpolar Health*. 1998; 57(Suppl 1):383–5.
- 5 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 (Oxf)*. 1997 May; 46(5):587–98.
- 6 Ricós C, Alvarez V, Cava F, García-Lario JV, Hernández A, Jiménez CV, et al. Current databases on biological variation: pros, cons and progress. *Scand J Clin Lab Invest*. 1999 Nov; 59(7):491–500.
- 7 Brauer VF, Eder P, Miehle K, Wiesner TD, Hasenclever H, Paschke R. Interobserver variation for ultrasound determination of thyroid nodule volumes. *Thyroid*. 2005 Oct; 15(10):1169–75.
- 8 Ittermann T, Johner S, Below H, Leiterer M, Thamm M, Remer T, et al. Interlaboratory variability of urinary iodine measurements. *Clin Chem Lab Med*. 2018 Feb;56(3):441–7.
- 9 Völzke H, Alte D, Schmidt CO, Radke D, Lorbach R, Friedrich N, et al. Cohort profile: the Study of Health in Pomerania. *Int J Epidemiol*. 2011 Apr;40(2):294–307.
- 10 Andersen S, Pedersen KM, Pedersen IB, Laurberg P. Variations in urinary iodine excretion and thyroid function. A 1-year study in healthy men. *Eur J Endocrinol*. 2001 May; 144(5):461–5.
- 11 Khattak RM, Ittermann T, Nauck M, Below H, Völzke H. Monitoring the prevalence of thyroid disorders in the adult population of Northeast Germany. *Popul Health Metr*. 2016 Nov;14(1):39.
- 12 Völzke H, Lüdemann J, Robinson DM, Spiekler KW, Schwahn C, Kramer A, et al. The prevalence of undiagnosed thyroid disorders in a previously iodine-deficient area. *Thyroid*. 2003 Aug;13(8):803–10.
- 13 Brunn J, Block U, Ruf G, Bos I, Kunze WP, Scriba PC. Volumetric analysis of thyroid lobes by real-time ultrasound. *Dtsch Med Wochenschr*. 1981 Oct;106(41):1338–40. German.
- 14 Gutekunst R, Becker W, Hehrmann R, Olbricht T, Pfannenstiel P. Ultrasonic diagnosis of the thyroid gland. *Dtsch Med Wochenschr*. 1988 Jul;113(27):1109–12. German.
- 15 Völzke H, Werner A, Wallaschofski H, Friedrich N, Robinson DM, Kindler S, et al. Occupational exposure to ionizing radiation is associated with autoimmune thyroid disease. *J Clin Endocrinol Metab*. 2005 Aug;90(8): 4587–92.
- 16 Lorenz-Wawschinek O, Tiran B, Eber O, Langsteger W. Photometric determination of iodine in urine. *Exp Clin Endocrinol*. 1994; 102:57–8.
- 17 Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess*. 1994;6(4):284–90.

Statement of Ethics

All participants provided informed written consent. The study was approved by the ethics committee of the University of Greifswald and followed the Declaration of Helsinki.

Disclosure Statement

The authors declare that they have no conflict of interest.

Funding Sources

The study was supported by Euthyroid, a project funded by the European Union's Horizon 2020 research and innovation program under grant agreement number 634453.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection, and data analysis were performed by T. Ittermann, C. Jürgens, M. Nauck, A. Petersmann, H. Below, and H. Völzke. The first draft of the manuscript was written by T. Ittermann and A. Richter and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

- 18 World Health Organization. Assessment of iodine deficiency disorders and monitoring their elimination. A guide for programme managers. Third edition. 2007. https://apps.who.int/iris/bitstream/handle/10665/43781/9789241595827_eng.pdf.
- 19 Andersen S, Noahsen P, Westergaard L, Laurberg P. Reliability of thyroglobulin in serum compared with urinary iodine when assessing individual and population iodine nutrition status. *Br J Nutr*. 2017 Feb;117(3):441–9.
- 20 Pearce EN, Caldwell KL. Urinary iodine, thyroid function, and thyroglobulin as biomarkers of iodine status. *Am J Clin Nutr*. 2016 Sep;104(Suppl 3):898S–901S.
- 21 O’Kane SM, Pourshahidi LK, Mulhern MS, Weir RR, Hill S, O’Reilly J, et al. The effect of processing and seasonality on the iodine and selenium concentration of cow’s milk produced in Northern Ireland (NI): implications for population dietary intake. *Nutrients*. 2018 Mar;10(3):10.
- 22 Knudsen N, Bülow I, Jørgensen T, Perrild H, Ovesen L, Laurberg P. Serum Tg – a sensitive marker of thyroid abnormalities and iodine deficiency in epidemiological studies. *J Clin Endocrinol Metab*. 2001 Aug;86(8):3599–603.
- 23 Zimmermann MB, Aeberli I, Andersson M, Assey V, Yorg JA, Jooste P, et al. Thyroglobulin is a sensitive measure of both deficient and excess iodine intakes in children and indicates no adverse effects on thyroid function in the UIC range of 100–299 µg/L: a UNICEF/ICCIDD study group report. *J Clin Endocrinol Metab*. 2013 Mar;98(3):1271–80.
- 24 Ma ZF, Venn BJ, Manning PJ, Cameron CM, Skeaff SA. The sensitivity and specificity of thyroglobulin concentration using repeated measures of urinary iodine excretion. *Eur J Nutr*. 2018 Jun;57(4):1313–20.
- 25 Fraser CG, Harris EK. Generation and application of data on biological variation in clinical chemistry. *Crit Rev Clin Lab Sci*. 1989;27(5):409–37.
- 26 Weeke J. Circadian variation of the serum thyrotropin level in normal subjects. *Scand J Clin Lab Invest*. 1973 May;31(3):337–42.
- 27 Behrends J, Prank K, Dogu E, Brabant G. Central nervous system control of thyrotropin secretion during sleep and wakefulness. *Horm Res*. 1998;49(3–4):173–7.
- 28 Weeke J, Gundersen HJ. Circadian and 30 minutes variations in serum TSH and thyroid hormones in normal subjects. *Acta Endocrinol (Copenh)*. 1978 Dec;89(4):659–72.
- 29 Thienpont LM, Van Uytvanghe K, Beastall G, Faix JD, Ieiri T, Miller WG, et al.; IFCC Working Group on Standardization of Thyroid Function Tests. Report of the IFCC Working Group for Standardization of Thyroid Function Tests; part 2: free thyroxine and free triiodothyronine. *Clin Chem*. 2010 Jun;56(6):912–20.
- 30 Carpenter JR, Kenward MG. *Multiple imputation and its application*. First edition. Chichester: Wiley; 2013.