TSH Comparison Between Chemiluminescence (Architect) and Electrochemiluminescence (Cobas) Immunoassays: An Indian Population Perspective

Rajarshi Sarkar

Received: 30 December 2012 / Accepted: 5 May 2013 / Published online: 26 May 2013
© Association of Clinical Biochemists of India 2013

Abstract Although 3rd generation TSH assays are the most widely used immunoassays, credible comparison studies, specially involving Indian sub-populations are practically non-existent. To compare the TSH measurements between chemiluminescence (Architect) and electrochemiluminescence (Cobas) immunoassays in an urban ambulatory Indian population, 1,615 subjects were selected randomly from the usual laboratory workflow, their TSH measured in Architect and Cobas and the paired data thus generated were statistically analysed. TSH values of Cobas were observed to be higher than the Architect values by 28.7 %, with a significant proportional difference between the two, but majority of the Cobas values (above 90 %) were within the limits of agreement with Architect values. In situations where both the instruments are in use simultaneously, a standardization of the methods is imperative, in larger interest of the patient populace.

Keywords TSH · Method comparison · Chemiluminescence · Electrochemiluminescence · Architect · Cobas

Introduction

Thyroid stimulating hormone (TSH), or thyrotropin has become quite ubiquitous in the test requests of all physicians'

prescriptions, and rightly so, as TSH is a very sensitive and specific parameter for assessing thyroid function and is particularly suitable for early detection or exclusion of disorders in the central regulating circuit between hypothalamus, pituitary and thyroid [1–5]. TSH is a glycoprotein of ~28 kDa, synthesized by the basophilic cells of anterior pituitary. It comprises of two non-covalently linked subunits— α , which is common to LH, FSH and hCG, and β , which is hormone specific, conferring its biological and immunological specificity [6]. Hence diagnostic antibodies to TSH are directed against its β -subunit and although immunoassay is the standard procedure for measurement of TSH, a 'generational' classification based on the lowest TSH value detectable with a 20 % or less interassay co-efficient of variation (CV) is often used to describe them. 1st generation (typically radioimmunoassay) has a detection limit of 1–2 $\mu\text{IU/mL}$ [7], 2nd generation (typically enzyme-linked immunosorbent assay) has a detection limit of 0.1–0.2 $\mu\text{IU/mL}$ while 3rd generation assays have a detection limit of 0.01–0.02 $\mu\text{IU/mL}$. 3rd generation assays can distinguish mildly subnormal TSH values, as seen in non-thyroid illness (NTI) from the very low values of frank hyperthyroidism, are helpful in classifying hyperthyroid subjects according to their degree of TSH suppression and in monitoring of thyroid cancer patients who require complete suppression of TSH [8–11]. In view of such advantages, 3rd generation assays are most widely used nowadays, but unfortunately, there is a dearth of credible comparison studies of these methods. The curse of this void is ultimately borne by the patients, as their clinicians are at a loss when they have to compare reports of TSH measured by two different methods. The objective of this study is to compare the values of TSH measured by two such 3rd generation assays—one based on chemiluminescence microparticle immunoassay (CMIA) by Architect (Abbott Diagnostics) and the other based on electrochemiluminescence immunoassay

R. Sarkar
Santosh Apartments, Flat No. 106, 245/1 Dum Dum Road,
Kolkata 700074, India

R. Sarkar (✉)
Department of Biochemistry, Drs. Tribedi & Roy Diagnostic
Laboratory Pvt. Ltd, 93 Park Street, Kolkata 700016, India
e-mail: sarkar825@yahoo.co.in

(eCLIA) by Cobas (Roche Diagnostics) on an urban ambulatory Indian population.

TSH CMIA is a two-step sandwich immunoassay [12]. In the 1st step, sample anti- β TSH antibody coated paramagnetic microparticles and TSH assay diluent in TRIS buffer are added. After washing under a magnetic field, anti- α TSH acridinium labelled conjugate is added in the second step. Pre-trigger, containing 1.32 % (w/v) H_2O_2 , and trigger, containing 0.35N-NaOH are then added to the reaction mixture. The resulting chemiluminescent reaction is measured as relative light units (RLUs). A direct relationship exists between the amount of TSH present in the sample and the RLUs detected. Evaluation and calculation of concentration of TSH are carried out by means of a calibration curve that was established using calibrators of known TSH concentration.

TSH eCLIA is a three-step sandwich immunoassay [13]. In the first step, sample is combined with a reagent containing biotinylated TSH antibody and a *N*-hydroxysuccinimide (NHS) ester of a modified ruthenium-tris(bipyridyl) $[\text{Ru}(\text{bpy})_3^{2+}]$ TSH-specific antibody in an assay cup. During a 9-min incubation step, antibodies capture the TSH present in the sample. In the second step, streptavidin-coated paramagnetic microbeads are added. During a second 9-min incubation, the biotinylated antibody attaches to the streptavidin-coated surface of the microbeads. After the second incubation, the reaction mixture containing the immune complexes is transported into the measuring cell; the immune complexes are magnetically entrapped on the working electrode, but unbound reagent and sample are washed away by a system buffer (Procell). In the ECL reaction, the conjugate is a ruthenium based derivative and tripropylamine (TPA); the chemiluminescent reaction is electrically stimulated to produce light at 620 nm. The amount of light produced is directly proportional to the amount of TSH in the sample. Evaluation and calculation of concentration of TSH is carried out by means of a calibration curve that was established using calibrators of known TSH concentration.

It may be noted the objective of this exercise is not to carry out a detailed method evaluation of either of the two automated processes. Such studies have already been carried out elsewhere elaborately [14–18]. The sole objective of this study is to determine the comparability of test results carried out on actual samples and the degree to which results of one method are transferable to the other.

Materials and Methods

This non-interventional, cross-sectional study was undertaken during the calendar year of 2012, in a standalone

diagnostic laboratory serving predominantly an urban ambulatory Indian population, on issuance of a statutory Ethical Committee approval.

Subjects

The subjects were selected randomly from the usual lab workflow. No exclusion criterion was applied with respect to demographics, drug intake or clinical condition, except for the fact that subjects under biotin supplementation were advised to undergo the tests 1 week after discontinuing the supplementation. Total number of subjects was 1,615. It may be noted that the study subset was not representative of the universal population served by the laboratory, as the objective was not to construct a biological reference interval; rather the conscious effort was to cover the whole of the measuring range of the TSH method, with special emphasis on the borderline low and borderline high regions. Hence, demographic data of the study population has not been included.

Samples

Venous blood samples were collected observing standard aseptic procedures and transferred to standard gel separator tubes [19, 20]. All samples were tested on the same day of collection. Grossly haemolysed, lipemic or icteric samples were excluded from the study.

Methods

TSH was primarily measured on the instrument Architect i2000SR (Abbott Diagnostics) and then repeat tested on the same date on the instrument Cobas 6000 (Roche Diagnostics). Both instruments are closed system, random access auto-analysers. Performance characteristics of both for TSH measurement are included in Table 1. The study was spaced out into a whole calendar year deliberately to even out the variations brought in by changes in lots of reagents, calibrators and in instrument calibrations. In particular, two lots of calibrators were used in each instrument, five lots of reagent for Architect and four lots of reagent for Cobas were used during this period. Method calibrations were done on each instrument roughly once every month for change of lots, as corrective measures for quality control outliers and after instrument calibrations, which were done once every 6 months. Two levels of 3rd party commercial control materials were run on every working day (Level 1 target value = $0.959 \mu\text{IU/mL}$, Level 2 target value = $8.13 \mu\text{IU/mL}$), on each instrument and standard Westgard warning and run rejection rules were applied.

Statistical Analyses

The paired data thus obtained were analysed by MedCalc Version 11.5.0 statistical software and segregated into the three clinically relevant groups guided by the biological reference interval (BRI) as provided by the Architect kit insert: Group A ($n = 117$), with values $<0.35 \mu\text{IU/mL}$; Group B ($n = 375$), with values $0.35\text{--}4.94 \mu\text{IU/mL}$ and Group C ($n = 1,123$), with values $>4.94 \mu\text{IU/mL}$, all based on Architect values. It is noteworthy that when BRI criteria of Cobas were applied to these three groups, 71 out of 117 (60.7 %) in Group A had values below $0.27 \mu\text{IU/mL}$ (conversely 39.3 % would be mis-diagnosed as hyperthyroid); 186 out of 375 (49.6 %) in Group B had values within the BRI of Cobas (consequently, 51.4 % of hypothyroid subjects would be mis-diagnosed as euthyroid). In Group C, all subjects had values above $4.20 \mu\text{IU/mL}$. None of the groups exhibited normal distribution. Hence each group was first subjected to the non-parametric Mann–Whitney U Test for independent samples [21] to measure the significance of difference, then examined for concordance correlation ρ_c [22], which evaluates the degree to which pairs of observations fall on the 45° line through the origin and contains a measurement of precision ρ and accuracy C_b ($\rho_c = \rho \cdot C_b$) and finally compared using two models of Regression: Deming [23], which takes measurement errors for both

methods into account and Passing–Bablok [24], which requires no special assumptions regarding the distribution of the samples and the measurement errors. Based on these findings, Bland–Altman plots [25], where the differences (or alternatively the ratios) between the two techniques are plotted against the averages of the two techniques, were constructed for each group.

Results

Group A revealed a Mann–Whitney U statistic $Z (=2.790)$ with $P = 0.0053$, hence not significant, i.e. means of Architect and Cobas data are not different (vide Table 2). Concordance correlation generated a coefficient, $\rho_c = 0.6047$, signifying poor correlation (vide Table 3). Deming regression (Table 4) returned an intercept $A = -0.05584$ and a slope $B = 1.7975$ indicating absence of any constant difference but a presence of a significant proportional difference towards y -axis, i.e. Cobas values. Passing–Bablok regression calculated an intercept $A = -0.006184$ and a slope $B = 1.3963$, signifying negligible constant difference but significant proportional deviation to y -axis. Cusum test revealed no significant deviation from linearity ($P = 0.34$). Bland–Altman plot (Fig. 1) showed that Cobas values were on a higher side with a mean

Table 1 Performance specifications of Architect i2000 SR and Cobas 6000 for TSH method

Performance indicator	Architect	Cobas
Analytical sensitivity (value lying 2 SD above the zero calibrator in a repeatability study):	$0.0025 \mu\text{IU/mL}$	$0.005 \mu\text{IU/mL}$
Functional sensitivity (lowest analyte concentration that can be reproducibly measured with a across-run precision of 20 % CV):	$0.0036 \mu\text{IU/mL}$	$0.014 \mu\text{IU/mL}$
Analytical specificity	<10 % Cross reactivity in human serum samples containing TSH in the normal range was observed with FSH $\leq 500 \text{ mIU/mL}$, LH $\leq 500 \text{ mIU/mL}$ and hCG $\leq 200,000 \text{ mIU/mL}$	For the monoclonal antibodies used, the following cross-reactivities were found: LH 0.038 %, FSH 0.008 %, hGH and hCG no cross-reactivity
Analytical measurement range	$0.0025\text{--}100 \mu\text{IU/mL}$ (or up to $1,000 \mu\text{IU/mL}$ for 10 fold diluted samples)	$0.005\text{--}100 \mu\text{IU/mL}$ (or up to $1,000 \mu\text{IU/mL}$ for 10 fold diluted samples)
Precision	Within run -1.1 to 5.0 %CV, across-runs -1.9 to 5.3 %CV	Within run -1.1 to 3.0 %CV, Across-runs -3.2 to 7.2 %CV
Accuracy (as per evaluations through one year on a commercial 3rd party EQA programme)	$1.8\text{--}4.6$ %	$1.2\text{--}5.6$ %
Interferences	<10 % interferences were observed by haemolysis up to 500 mg/dL of Hb, lipaemia up to $3,000 \text{ mg/dL}$ of triglycerides, icterus up to 20 mg/dL of bilirubin and 12 g/dL of protein	No significant interferences were observed by haemolysis up to 1 g/dL of Hb, lipaemia up to $1,500 \text{ mg/dL}$ of intralipid, icterus up to 41 mg/dL of bilirubin, up to 25 ng/mL of biotin, 2 g/dL of IgG, 0.5 g/dL of IgM and $3,250 \text{ IU/mL}$ of rheumatoid factors
Biological reference intervals	$0.35\text{--}4.94 \mu\text{IU/mL}$ ($n = 549$)	$0.27\text{--}4.20 \mu\text{IU/mL}$ ($n = 516$)

difference of 0.07 $\mu\text{IU/mL}$. Although majority of values (96.58 %) fell within the limits of agreement, which are defined as the mean difference plus and minus 1.96 times the standard deviation of the differences, the regression line of differences detected a proportional difference with positivity towards Cobas.

Statistical treatment of Group B gave a Mann–Whitney U statistic Z (=9.514) with $P < 0.0001$, hence significant, i.e. means of Architect and Cobas data are different (vide Table 2). Concordance correlation $\rho_c = 0.6669$, signalling poor correlation (vide Table 3). Deming regression (Table 4) stats were as follows: intercept $A = -0.2621$, slope $B = 1.4190$, indicating an ignorable constant deviation but a significant y-axis proportional deviation. Passing–Bablok regression numbers are listed thus: intercept $A = -0.1123$ and slope $B = 1.3904$ revealing no constant difference but significant y-leaning proportional deviation. Cusum test for linearity revealed no significant deviation from linearity. Bland–Altman plot (Fig. 2) showed that Cobas values were on a higher side with a mean difference of 1.00 $\mu\text{IU/mL}$. Although majority of values (94.4 %) fell within the limits of agreement, most of them lie *below* the *line* of equality signifying a systematic deviation towards Cobas and the regression line of differences detected a proportional difference with positivity towards Cobas.

Similar trends were evident with Group C too. Mann–Whitney U statistic Z (=9.366) with $P < 0.0001$, hence significant, i.e. means of Architect and Cobas data are different (vide Table 2). Concordance correlation $\rho_c = 0.8479$, signifying poor correlation (vide Table 3). Deming regression (Table 4) figures are presented thus: intercept $A = 1.4617$, slope $B = 1.2539$, hence indicating significant systematic (both constant and proportional) deviation towards y-axis (Cobas). However, Passing–Bablok analysis (intercept $A = 0.1734$, slope $B = 1.2991$) revealed no constant deviation but indicated a proportional deviation of similar magnitude and direction. Bland–Altman plot (Fig. 3) returned similar observations as in the other two groups: mean difference of 9.8 $\mu\text{IU/mL}$ with majority of values (94.48 %) lie

Table 2 Findings of the Mann–Whitney U Test for independent samples applied to the three groups

	Group A	Group B	Group C
Sample size, n	117	375	1,123
Median			
Architect	0.150	3.290	29.690
Cobas	0.194	4.290	38.850
Mann–Whitney U	5,400.0	42,087.5	486,643.0
Test statistic Z (corrected for ties)	2.790	9.514	9.366
Two-tailed probability	$P = 0.0053$	$P < 0.0001$	$P < 0.0001$

Median value of Cobas is on the higher side in all the groups

Table 3 Concordance correlation findings of the three groups

	Group A	Group B	Group C
Sample size, n	117	375	1,123
Concordance correlation coefficient	0.6047	0.6669	0.8479
Pearson ρ (precision)	0.7587	0.9116	0.9513
Bias correction factor C_b (accuracy)	0.7970	0.7316	0.8913

Concordance correlation coefficient, $\rho_c = \rho \cdot C_b$

within the limits of agreement but below the line of equality, indicating a systematic bias towards Cobas and a regression line of difference revealing a proportional deviation, again towards Cobas.

Discussion

A detailed review of available literature confirmed the findings of this study along predictable lines: of particular importance in this context are the studies of Hendriks et al. [26] and Rawlins and Roberts [27]. Method comparison experiment for TSH using 50 patient samples by Hendriks et al. between AxSym (an earlier congener of Architect, which uses FPIA, vis-à-vis CMIA of Architect) and Elecsys 2010 (an earlier precursor of Cobas e-series, using the same eCLIA methodology) generated a Passing–Bablok slope of 1.37 and intercept of $-0.02 \mu\text{IU/mL}$. Rawlins and Roberts laid emphasis on the lower limit of quantification by determining the respective functional sensitivities but also included a method comparison study using 104 patient samples which returned a slope of 1.33 and an intercept of 0.01 $\mu\text{IU/mL}$ by Passing–Bablok analysis, between Immulite 2000 (which they have shown to be comparable with Architect) and E170 (which is also an earlier version of Cobas e-series, using the same technology). It may be worth mentioning that the present exercise generated similar figures as these two but in contrast, has involved a far greater number of subjects ($n = 1,615$), has employed instruments of the latest versions available (Architect i2000 SR and Cobas 6000) and most importantly, has focussed on the impact of the comparability in the clinically relevant ‘borderline low’ and ‘borderline high’ regions. From the findings of the statistical analyses of the data, it may be inferred that a systematic difference existed between the values of TSH measured by Architect and Cobas. The mean value of Cobas was higher than that of Architect all across the span of data with Mann–Whitney P values in the significant range. According to McBride’s descriptive scale for values of the concordance correlation coefficient (<0.9 : Poor, 0.9 – 0.95 : Moderate, 0.951 – 0.99 : Substantial, >0.99 : almost perfect), all three groups of data exhibited poor strength of agreement. By and large, both models of

Table 4 Regression analysis findings of the three groups

	Group A		Group B		Group C	
	Deming	Passing–Bablok	Deming	Passing–Bablok	Deming	Passing–Bablok
Intercept A	−0.05584	−0.006184	−0.2621	−0.1123	1.4617	0.1734
95 % CI of intercept	−0.1274 to 0.01568	−0.0088 to −0.00083	−0.4529 to −0.07125	−0.2468 to 0.02780	0.7087 to 2.2146	−0.1968 to 0.4962
Slope B	1.7975	1.3963	1.4190	1.3904	1.2539	1.2991
95 % CI of slope	1.2137–2.3814	1.3417–1.4375	1.3512–1.4869	1.3390–1.4416	1.2226–1.2852	1.2787–1.3193
Cusum test for linearity	NA	$P = 0.34$	NA	$P = 0.05$	NA	$P = 0.15$

In the cusum test for linearity, P values obtained for all three groups are not significant; hence there is no significant deviation from linearity in any group

Fig. 1 Bland–Altman plot for Group A. Majority of points (96.58 %) lie within the limits of agreement (mean $\pm$ 1.96 SD), albeit *below* the line of equality. Regression line of difference indicates a proportional deviation. Mean ratio and mean % difference for this set of data are 0.81 and 30.7 % respectively

regression analysis predicted minimal constant deviation but substantial proportional deviation towards Cobas values, with slopes varying from 1.2539 to 1.7975. Bland–Altman plots reinforced the prediction of the regression analyses, but also revealed that majority of the points (>90 %) lay within the limits of agreement in all the groups. Hence, it may be opined that, according to the Bland–Altman plot interpretation, in majority of the instances (above 90 %), values of TSH measured by one method were transferable to the other.

But in view of the significant difference in mean values and the poor concordance correlation returned by all the three clinically significant groups, certain practical issues need attention. It is often discussed that proportional differences most commonly occur due to improper assignment of the amount of substance in the calibrator: if a calibrator has less analyte than is labelled, all the unknown measurements would be on the higher side; more analyte than is labelled would similarly result in negative error. It would seem obvious that this form of error can be

Fig. 2 Bland–Altman plot for Group B. Majority of points (94.4 %) lie within the limits of agreement (mean $\pm$ 1.96 SD), but *below* the line of equality. Regression line of difference indicates a proportional deviation. Mean ratio and mean % difference for this set of data are 0.76 and 28.5 % respectively

eliminated by using the same calibrator for both the methods. But as pointed out by Hendriks et al. (2000), calibration against the same standard does not automatically guarantee inter-method agreement. The differences occur at two levels: firstly, during establishment of a method, even though the same recombinant TSH molecule is used as the standard, diagnostic antibodies directed to it, monoclonal or polyclonal, are raised differently by individual manufacturers, which recognizes different epitopes of the recTSH with different avidity. Secondly, pituitary TSH present in human sera occur as different glycoforms in health and disease which are recognized differently by the diagnostic antibodies, thereby generating a new level of assay heterogeneity. Indeed in the current scenario, both instruments operate as closed systems and use dedicated barcoded calibrators, both traceable to the WHO Second International Reference Preparation 80/558. The other option is method standardization: as is evident from the

Fig. 3 Bland–Altman plot for Group C. Majority of points (94.48 %) lie within the limits of agreement (mean $\pm$ 1.96 SD), but *below* the line of equality. Regression line of difference indicates a proportional deviation. Mean ratio and mean % difference for this set of data are 0.78 and 26.8 % respectively

Bland–Altman plots, the mean ratio (Architect/Cobas) of Groups A, B and C are 0.81, 0.76 and 0.78 (30.7, 28.5 and 26.8 when expressed as percentages) respectively. Arithmetic mean of these three values is 0.783 (or 28.7 %). If the instrument factor of Cobas (which normally is 1.0) is put as 0.783, the TSH values would be correspondingly lower and would match those of Architect. Conversely, if the instrument factor of Architect is changed to 1.277 ($=1/0.783$), its TSH values would match with those of Cobas. Changes in values of quality control specimen and Levey–Jennings charts must be interpreted accordingly. However, while reporting for externally derived challenge materials, such as for interlaboratory comparisons, the same correction factors must be applied to arrive at the original values. The main drawback of this method is that it assumes as one of the two instruments is generating correct values and the other is not. Another feasible approach of harmonization of the methods is described by Theinpont et al. [28]. In this process, results of a panel of human serum samples obtained by each method are recalibrated using the best-fitting reverse regression equation of the method comparison data. This recalibration data would then be used to establish traceability of calibration to the mean values for the panel of serum samples. This process would be repeated with a new panel overlapping with the current panel; so that the new panel's mean values can be matched to the harmonization fixed point of the first panel, and so on. Even this process is not devoid of major concerns: each recalibration of an assay would entail changes in the analytical measurement range, values for calibrators, controls and reference intervals for various populations. When such standardization is not possible and a clinical laboratory generates reports of TSH measured by the two systems, namely Architect and Cobas,

the report format should contain the particular name of the instrument and method which generated the TSH value.

While comparing the BRI of TSH by both the methods, both of which were determined on Western populations, it is noteworthy that BRI of Cobas (0.27–4.20 μ IU/mL) is actually on the lower side than the BRI of Architect (0.35–4.94 μ IU/mL). Interestingly, all the aforementioned method comparison studies were undertaken on Western populations and all of them revealed a bias towards instruments from the Cobas platform. Indeed, a mathematical estimation by Rawlins and Roberts (2004) fixed BRIs of Architect and E170 as 0.39–4.0 and 0.54–5.3 μ IU/mL respectively, which is in agreement with the data presented here. But when the TSH values of an Indian population are compared, not only in the range of BRI, but also below and above that interval, the Cobas values were consistently obtained at a higher side. Agreed that BRIs are non-transferable from population to population, but why and how did this discrepancy occur? As has been mentioned earlier, TSH assays are affected by the health status of the subjects. In acute illness, TSH reference interval widens to 0.1–20 μ IU/mL; in central hypothyroidism abnormal TSH glycoforms predominate leading to falsely high results; heterophile antibodies render results to be falsely increased with gross inter-assay differences; similar interferences are known to occur with TSH auto-antibodies. Of course, there might be several other factors of immunological interferences at play in this current scenario, starting from matrix effects of the reaction solution [29, 30] and the surface where the reaction is occurring to cross-reacting substances like the digitalis-like immunoreactive factor (DLIF) or other competing immunospecific antibodies to the analyte in question i.e. TSH [31]. What exactly is the cause of this variation remains to be uncovered and possibly requires employment of molecular level investigation, which, unfortunately is beyond the scope of this exercise.

Acknowledgements The author is greatly indebted to Dr. Subhendu Roy, Director, Drs. Tribedi & Roy Diagnostic Laboratory, and Dr. Debasis Banerjee, Department of Haematology, Drs. Tribedi & Roy Diagnostic Laboratory, without whose moral and intellectual support, this study would not have been possible.

References

1. Wheeler MH, Lazarus JH. Diseases of the Thyroid. London: Chapman and Hall Medical; 1994. p. 109–15.
2. Surks MI, Chopra IJ, Mariash CN, Nicoloff JT, Solomon DH. American Thyroid Association guidelines for the use of laboratory tests in thyroid disorders. JAMA. 1990;263:1529–32.
3. Keffer JH. Preanalytical considerations in testing thyroid function. Clin Chem. 1996;42(1):125–35.
4. Ladenson PW. Optimal laboratory testing for diagnosis and monitoring of thyroid nodules, goitre and thyroid cancer. Clin Chem. 1996;42(1):183–7.
5. Nicoloff JT, Spencer CA. The use and misuse of the sensitive thyrotropin assays. J Clin Endocr Metab. 1990;71:553–8.

6. Pierce JG. The subunits of pituitary thyrotropin. Their relationship with other glycoprotein hormones. *Endocrinology*. 1971;89: 1331–44.
7. Spencer CA. Clinical uses and limitations of rapid TSH assays. *Med Lab Prod* 1988; 17–19.
8. Bayer MF. Performance criteria for appropriate characterization of “(highly) sensitive” thyrotropin assays. *Clin Chem*. 1987;33: 630–1.
9. Hay ID, Bayer MF, Kaplan MM, Klee GG, Larsen PR, Spencer CA. American Thyroid Association assessment of current free thyroid hormone and thyrotropin measurements and guidelines for future clinical assays. *Clin Chem*. 1991;37:2002–8.
10. The National Academy of Clinical Biochemistry: Standards of Laboratory Practice. Laboratory Support for the Diagnosis & Monitoring of Thyroid Disease. NACB, 1996.
11. Hay ID, Klee GG. Linking medical needs and performance goals: clinical and laboratory perspectives on thyroid disease. *Clin Chem*. 1993;39:1519–24.
12. TSH Reagent Kit Literature for Architect System, Abbott Laboratories. Longford, Abbott Park, Wiesbaden:2010.
13. TSH Reagent Kit Literature for Cobas e601, Roche Diagnostics. Mannheim:2011.
14. Kahapola-Arachchige KM, Hadlow N, Wardrop R, Lim EM, Walsh JP. Age-specific TSH reference ranges have minimal impact on the diagnosis of thyroid dysfunction. *Clin Endocr*. 2012;77:773–9.
15. Aldrimer M, Ridefelt P, Rödö P, Niklasson F, Gustafsson J, Hellberg D. Reference intervals on the Abbot Architect for serum thyroid hormones, lipids and prolactin in healthy children in a population-based study. *Scand J Clin Lab Invest*. 2012;72:326–32.
16. Kazerouni F, Amirasouli H. Performance characteristics of three automated immunoassays for thyroid hormones. *Casp J Int Med*. 2012;3(2):400–4.
17. Çımrın D, Şişman AR, Kenesarı Y, Küme T, Akan P, Cöcer C. Comparison of Siemens Immulite and Abbott Architect TSH measurement methods. *Türk Klinik Biyokimya Dergisi*. 2011;9(3):97–103.
18. Sánchez-Carbayo M, Mauri M, Alfayate R, Miralles C, Soria F. Analytical and clinical evaluation of TSH and thyroid hormones by electrochemiluminescent immunoassays. *Clin Biochem*. 1999; 32(6):395–403.
19. World Health Organization. Laboratory Biosafety Manual. 3rd ed. Geneva: World Health Organization; 2004.
20. Clinical and Laboratory Standards Institute. Protection of laboratory workers from occupationally acquired infections: approved guidelines-third edition. CLSI document M29-A3. Wayne: Clinical and Laboratory Standards Institute; 2005.
21. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat*. 1947;18(1):50–60.
22. Lin LK. A concordance correlation coefficient to evaluate reproducibility. *Biometrics (International Biometric Society)*. 1989;45(1):255–68.
23. Deming WE. Statistical adjustment of data. New York: Wiley; 1943.
24. Passing H, Bablok W, Bender R, Schneider B. A general regression procedure for method transformation. *J Clin Chem Clin Biochem*. 1988;26(11):783–90.
25. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327(8476):307–10.
26. Hendriks HA, Kortlandt W, Verweij WM. Analytical performance comparison of five new generation immunoassay analyzers. *Ned Tijdschr Klin Chem*. 2000;25:170–7.
27. Rawlins ML, Roberts WL. Performance characteristics of six third-generation assays for thyroid-stimulating hormone. *Clin Chem*. 2004;50:2338–44.
28. Theinpont LM, Uytfanghe KV, Beattall 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. Bjerner J, Nustad K, Norum LF, Olsen KH, Børmer OP. Immunometric assay interferences: incidence and prevention. *Clin Chem*. 2002;48:613.
30. Valdes R Jr, Miller TI. Increasing the specificity of immunoassays. *J Clin Immunoassay*. 1992;15:87.
31. Kricka LJ. Interference in immunoassay—still a threat. *Clin Chem*. 2000;46:1037.