Texila International Journal of Academic Research
Volume 3, Issue 2, Dec 2016

Harmonization of Thyroid Stimulating Hormone; an Alternative Approach

Article by Izza Abdulhafedh
Laboratory Scientist, the Nairobi Hospital, Nairobi, Kenya
Email: ihafedh2000@yahoo.co.uk

Abstract

Introduction: Comparison of TSH values by analysis of retrospective data obtained from daily internal quality control (IQC) runs performed concurrently on Architect i2000 and cobas e601; the two platforms used to run TSH test for patient samples at The Nairobi Hospital, Kenya.

Method: Data analysis was carried out using R project version 3.2.3. Data was analyzed both in combination (N=590) and segregated as Level 2 (N= 297) and level 3 (N=293).

Results: Shapiro-Wilk normality test returned normal distribution for level 2, abnormal distribution for level 3 and combined data on both platforms. Paired T-test for level 2 and Mann–Whitney U -Wilcoxon signed rank test for level 3 and combined data, indicated that the alternative hypothesis: true location shift is not equal to 0. Deming and Passing-Bablok regression analysis for level 2 and 3 showed a significant deviation in values between the two methods by a constant. Concordance correlation of the individual levels 2 and 3 is poor.

Discussion: The data demonstrates that there is a constant systematic error between the two methods, with cobas values reading higher each time, and that this difference is significant at both levels.

Conclusion: Regression analysis shows transferability of results from one method to another, however, the difference in means and the poor concordance correlation obliges a consideration of the TSH value results given by the lab especially so in the subclinical range (5 – 10) mIU/ml.

Keywords: *Thyroid Stimulating Hormone, TSH, Architect i2000SR, Cobas e601*

Introduction

A practical approach to harmonization of patients' TSH results, in a hospital where approximately 15000 TSH tests are run annually, on two different platforms. Both in and outpatient samples of varying demographics (age, gender, race etc) and purpose (diagnostic, monitoring or general medical checkup) are tested.

Credible method comparison studies are virtually non-existent in Kenya currently, in spite of the presence of a wide range of testing platforms in both public and private labs. At The Nairobi Hospital (TNH), a 350 bed hospital with a plan for further expansion; the lab receives inpatient and outpatient samples of all ages; in 2015 TNH laboratory ran an average of 1,315 TSH tests per month. A summary of TSH patient results from 13/1/2015 to 13/1/2016 as follows

A total of 15,784 patient samples tested for TSH of which:

- 750 results tested < 0.25 mU/L
- 14283 results were normal (0.25 – 5.0) mU/L
- 941 results > 5.0 mU/L

TSH is now concurrently run on either Roche's cobas e601 (electrochemiluminescence) or Abbott's Architect i2000SR (chemiluminescence). The decision to run a patient sample for TSH on either platform is random; both cobas and architect have a quality assurance program including daily IQC and EQA. A third party IQC material is used for daily IQC runs; Randox Premium Immunoassay Plus levels 2 and 3. The lab participates in the RIQAS program for both immunoassays and clinical chemistries.

The presence of two platforms in the lab, allows one to backup the other when confirming abnormal results. Prone to supplier inconsistencies in provision of reagents, the use of two different platforms has assisted in continuity of service. However, over time, two observations have emerged from running the tests on both platforms;

- IQC values differ in range for all levels of IQC and EQA
- Some abnormal high patient values when run on the other platform would give a different numerical value.

A dilemma is met when subclinical high or low TSH values are encountered where values from one platform reflect abnormality while the other is normal. Currently the course of action involves review of previous thyroid tests done, the result from the platform that performed the previous test is the one released for consistency in patient management. The lab has also not established its own reference ranges from the local population therefore uses those recommended by the manufacturers (0.25 to 5.0) mU/L. Each platform has a different range. The reference range used for patients is a combination of the two platforms; however the actual numerical value of patient tests has not been tackled. At any given time a patient test run on both platforms will have a different numerical value and this value is released but the question remains which one of the two values is a true reflection of the patient clinical status and if it is a follow up what is the impact of these values in managing thyroid disorders.

In the past twenty years, diagnostics as part of evidence based medicine has witnessed an exponential increase in lab users' expectations in terms of turnaround times and quality results. This has led to increase in testing capacity/menus and patient volumes. Supplier dynamics also have an important impact in the lab's decision on the type of equipment procured. All these factors combined have led to the existence of several platforms in the lab. The current trend in automation is geared towards high-throughput, modular, robotic systems that incorporate both immunoassay and clinical chemistry analyzers into one instrument. Most recently, liquid chromatography-tandem mass spectrometry (LC-MS/MS) methods have been developed to measure total and free thyroid hormones as well as Thyroglobulin. These techniques however are technically complex and cannot be automated because they involve specimen pretreatments. Tests for TT4, TT3, THBR, TSH, Tg, TPOAb and TgAb using non-isotopic (primarily chemiluminescent) signals are currently available on a variety of immunoassay analyzer platforms that employ bar-coding, multiple-analyte random-access, primary tube sampling, auto-dilution, STAT testing and computerized data output.

Laboratories primarily select an analyzer to perform thyroid testing on the basis of instrument menu and operating costs, and only secondarily recognize that there are differences in the functional performance of different methods. Although the move to automation is seen as cost-effective, the consolidation of a diversity of immunoassay tests onto one platform has led to a transfer of thyroid testing from small, specialized laboratories to the general chemistry laboratory setting. This centralization has resulted in a loss of laboratory expertise for the clinical interpretation of thyroid tests. This has negatively impacted the ability of laboratory staff to discuss reasons for discordant test results with physicians(Spencer, 2013)

General objectives

Generally this study sought to harmonize TSH test in TNH laboratory for better clinical management of thyroid disorders in patients that use the lab's services

Specific objectives

- To gather and analyze IQC TSH data sets, run concurrently on Architect i2000 and COBAS e601 platforms
- To compare the two methods currently in use for testing patients and
- To recommend a course of action for the better management of thyroid disorders

Literature review

Following is a review of comparisons made for TSH on different platforms by other studies. The review will also highlight the importance of consistency in results for patient management.

The thyroid is one of the glands of the endocrine system. The TSH test is the most accurate test for diagnosing both hyperthyroidism and hypothyroidism. Thyroid hormones affect metabolism, brain development, breathing, heart and nervous system functions, body temperature, muscle strength, skin dryness, menstrual cycles, weight, and cholesterol levels. Thyroid function tests (TFTs) are used to diagnose thyroid disorders. Following an abnormal TFT result, a combination of imaging tests, such as ultrasound of the thyroid, a thyroid scan, or a radioactive iodine uptake test, can also be used to find the cause of thyroid disorders. Hyperthyroidism is caused by the presence of excess thyroid hormone in the bloodstream; thyroid hormones include free (FT3) or total tri-iodothyronine (TT3) and free (FT4) or total thyroxine (TT4). Symptoms of hyperthyroidism include increased speed of bodily functions leading to weight loss, sweating, rapid heart rate, and high blood pressure, among other. Hypothyroidism is a disorder that occurs when the thyroid doesn't make enough thyroid hormone for the body's needs. Without adequate thyroid hormone, many of the body's functions slow down. People may have symptoms such as fatigue, weight gain, and cold intolerance((NIDDK), 2014)

Harmonization in the broad sense is the overall process of achieving comparability of results among clinical laboratory measurement procedures that measure the same parameter.

The first generation of TSH assays used between 1965 and 1985 were based on radioimmunoassay (RIA) methodology that had limited functional sensitivity (~ 1.0 mIU/L)(Odell et al., 1965, Utiger, 1965, Yalow and Berson, 1996) Because these RIA-era TSH methods were too insensitive to detect TSH in all euthyroid subjects, their clinical utility was limited to the diagnosis of primary hypothyroidism. The more sensitive immunometric assay (IMA) methodology (also called "sandwich" or "noncompetitive" methodology) became available in the mid-1980s. These IMA techniques are based on the excess antibody approach of Miles and Hales originally reported in the 1960s but did not become widely adopted until advances in monoclonal antibody technology allowed the large-scale production of specific antibodies in the 1980s. Mechanistically, these IMA methods employed an excess of TSH monoclonal antibody, bound to a solid support (bead, tube, magnetic microparticle or adsorption gel) that captured TSH from the serum specimen during a 20 to 120 minute incubation period. A different poly- or monoclonal TSH antibody targeted to a different TSH epitope(s) and labeled with an isotopic (I-125) or non-isotopic signal was then added followed by a further incubation and removal of unbound constituents by washing. The signal bound to the solid support was quantified as being directly proportional to the serum TSH concentration in the test sample. Later modifications to this basic concept included the use of chimeric monoclonal antibodies to reduce interference by heterophilic antibodies and the use of Avidin-Biotin and magnetic particle separation techniques. By 1990, IMA non-isotopic methods had replaced most TSH RIA methods and as a result of inherently greater assay sensitivity and specificity resulted in narrowing the TSH reference range by reducing glycoprotein hormone cross-reactivity and improving precision. Currently, most TSH testing is performed on automated immunoassay platforms employing advanced IMA technology.

The first IMA methods that used a radioisotopic signal (I-125) were designated "immunoradiometric assays", or IRMAs. Subsequent IMA methods adopted a variety of non-isotopic signals that gave rise to a lexicon of terminology to distinguish between assays using different signals. For example, immunoenzymometric assays (IEMA) used enzyme signals; immunofluorometric assays (IFMA) used fluorophors as signals, immunochemiluminometric assays (ICMA) used chemiluminescent molecules as signals and immunobioluminometric assays (IBMA) used bioluminescent signal molecules. This explosion of methodology led to a range of IMAs with competing claims for sensitivity. Initially, the IMA methods were designated as "sensitive", "highly sensitive", "ultrasensitive" or "supersensitive" assays – terms used to distinguish the new IMA methodology from the older insensitive RIA methods then still in use. This descriptive nomenclature

was confusing and led to a debate concerning the meaning of “sensitivity”. After it became evident that it was the between-run precision of the method that was the best determinant of assay sensitivity, a new parameter “functional sensitivity” became adopted. Functional sensitivity has been defined as the TSH value associated with a 20 percent coefficient of variation (CV) established from assays run over a 6 to 8 week period (a typical clinical interval used to assess TSH changes in an out-patient setting) (Baloch et al., 2003) Both manufacturers and clinical laboratories have now adopted this functional sensitivity definition as the lowest reporting limit for TSH assays. The new nomenclature also defines each generation as having a ten-fold difference in functional sensitivity. For example, RIA methods with functional sensitivities between 1 and 2 mIU/L are designated as “first generation”. IMA methods with functional sensitivities between 0.1 and 0.2 mIU/L are designated as “second generation”. Third generation TSH methods with functional sensitivities between 0.01 and 0.02 mIU/L are typically automated and non-isotopic and have become recognized as necessary to meet the current standard of care(Spencer, 2013)

In the last two decades, TSH assay sensitivity has been further enhanced by the adoption of non-isotopic (chemiluminescent and fluorescent) signals that are inherently more sensitive than I-125 and offer the additional advantage of being easier to automate. By 1990, IMA non-isotopic methods had replaced most TSH RIA methods and as a result of inherently greater assay sensitivity and specificity resulted in narrowing the TSH reference range by reducing glycoprotein hormone cross-reactivity and improving precision. Currently, most TSH testing is performed on automated immunoassay platforms employing advanced IMA technology(Spencer, 2013)

The Abbott architect i2000SR uses chemiluminescence 2-step assay method while the Roche cobas e601 uses electrochemiluminescence one step assay method. "Chemiluminescent" processes entail the creation of luminescent species by chemical transfer of energy while "Electrochemiluminescence" entails creation of luminescent species electrochemically (Williams, 2015)

The Architect *i*2000SR chemiluminescence method incorporates an acridinium derivative tracer; using microparticles as solid phase. After exposure to pre-trigger and trigger reagent, the acridinium undergoes a decomposition reaction and the emitted light is amplified and processed. The Cobas e601 incorporates an electrochemiluminescence detection cell. The streptavidin-coated paramagnetic beads are coupled to the ruthenium-labeled antigen-antibody complex. After the addition of tripropylamine, a voltage is applied, and the resulting luminescence is measured(Sarkar, 2014)

Difference in values on platforms using third generation assays is now a point of focus in harmonization of TSH and standardization of thyroid hormones (Sarkar, 2014, Williams, 2015, Serdar et al., 2015)

In a study on an Indian urban subpopulation, specifically comparing 1,615 patient TSH values on these two analyzers, it was found that a systematic difference existed between the two. The study also suggested that 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. The main drawback of this method is that it assumes that one of the two instruments is generating correct values and the other is not (Sarkar, 2014)

In a different study; data sets obtained from four different immunoassay analyzers, poor and statistically different correlation was observed between analyzers at TSH values ranging from (1.0 to 10.0) uIU/mL compared with FT4. The study concluded that these variations between analyzers may affect the clinical decisions especially in the evaluation of subclinical hypothyroidism, clinicians and laboratory specialists should be aware of these situations (Hendriks et al., 2000)

A reference method should be chosen for the comparative method. Reference method infers a high quality method whose results are known to be correct through use of a reference method procedure (RMP). Any differences between a test method and a reference method are assigned to the test

method, i.e., the errors are attributed to the test method because the correctness of the reference method is well documented. In the absence of RMP for TSH, a comparative method will be used

The term “comparative method” is a more general term and does not imply that the correctness of the method has been documented. Most routine laboratory methods fall into this latter category. Any differences between a test method and a routine method must be carefully interpreted. If the differences are small, then the two methods have the same relative accuracy. If the differences are large and medically unacceptable, then it is necessary to identify which method is inaccurate. Recovery and interference experiments can be employed to provide this additional information (Westgard et al., 1999)

Methodology of study

Analysis of retrospective data obtained from daily IQC runs performed concurrently on Architect i2000 and Cobas e601. IQC material used is Randox premium immunoassay plus levels 2 and 3.

Statistical analysis of data involved data sets obtained for both IQC levels separately and combined. The data was tested for normality using the Shapiro test, followed by the OLS method for IQC level 2 which showed normal distribution but was of minimal use when comparing all the data. Levels 2, 3 and the combined data were analyzed using both Deming and the Passing-Bablok regression methods. Deming method takes into consideration measurement errors of both methods while Passing-Bablok assumes no special considerations for the data when calculating. Confidence intervals and ‘P’ values were generated. Lastly concordance correlation was also calculated.

Study area

This study was carried out at The Nairobi Hospital’s main laboratory where TSH testing is done on Architect i2000 by the chemiluminescence method (CLIA) and Cobas e601 by the electrochemiluminescence method (eCLIA) (Table 1)

Study population

Only IQC data was handled for the study; however mention was be made of some patient data on which observations led to this study. The target population of the study is the patients that present for testing at TNH laboratory.

Study duration

Data included IQC run on a daily basis for about one and a half years (January 2015- June 2016)

Sampling method and size

Using the EQA sample RIQAS immunoassay cycle 43 sample 9, specifically chosen for the mean value’s closeness to the subclinical upper limit of TSH in patients, the following information was extracted from the results returned and used to calculate p value for level of significance when comparing mean values from the two instruments (Table 2)

Comparison of the two means for level of significance produced a p value of < 0.001, this justifies that a sample size N=590 used for this study, is sufficient to demonstrate whether a difference exists or not.

Collection of data

590 data sets were obtained by retrieving TSH values of daily IQC runs on both analyzers from concurrently run IQC material over a period of about one and a half years. The IQC material is run once daily therefore data sets were distributed over several events during the period, including different IQC lot numbers, reagent lots, calibrations carried out during lot changeovers, after instrument preventive maintenance and as corrective action for QC outliers.

Over the period that data was collected, three different IQC lots had been used for level 2 and two different lots for level 3. Standard Westgard warning and rejection rules were applied for IQC runs on

each analyzer. EQA data shows TSH passed and consistent results obtained for both analyzers during the period

Data analysis

Data analysis was carried out using R project version 3.2.3(R core team, 2015). 590 paired data sets were obtained and analyzed both in combination (N=590) and segregated as Level 2 (N= 297) and level 3 (N=293). Shapiro test was used to determine distribution patterns. Significance of difference was measured using paired T-test for level 2 and Mann-Whitney U-Wilcoxon test for level 3 and combined data. Regression analysis for the combined data and individual levels 2 and 3 was carried out using Deming method which takes into consideration measurement errors of both methods and Passing-Bablok method which assumes no special considerations for the data; in addition OLS method was used for regression analysis of level 2 since it exhibited normal distribution. Based on this analysis Bland-Altman plots were constructed for each data group; level 2, level 3and combined. Concordance correlation coefficient (CCC) was calculated to test level of agreement between the paired data.

Observations and findings

Distribution

Shapiro-Wilk normality test, using R, returned normal distribution for level 2 and abnormal distribution for level 3 and combined data on both platforms (**Table 6**)

Significance of difference

Paired T-test to measure the significance of difference for level 2 indicated that the alternative hypothesis: true difference in means is not equal to 0 and the mean of the differences = 0.46 (**Table 6**).

Mann–Whitney *U* -Wilcoxon signed rank test for level 3 and combined data, indicated that the alternative hypothesis: true location shift is not equal to 0 in both cases (**Table 6**).

The difference between the two analyzers is significant at both levels 2 and 3

Regression analysis

Regression analysis was conducted using Methcomp package in R (Bendix Carstensen, 2015)

An analysis of level 2 using the OLS method gave an intercept = 1.83, slope = 0.33, Residual standard error: 0.1038 on 291 degrees of freedom, Multiple R-squared = 0.1096, adjusted R-squared= 0.1066, p-value: 6.33e-09. As architect values increase by 1 cobas values increase by 0.33 units and when the TSH value on architect is zero the value on cobas is 1.83. This means the readings of the two methods differ by a constant.

Since OLS method is not applicable to values that are not normally distributed, Deming and Passing-Bablok methods were applied to all the three data sets to better interpret the comparison.

Deming regression for level 2 gave an intercept of 0.55 and slope of 0.96, level 3 an intercept of 2.34 and slope of 0.96, and combined data returned an intercept of 0.28 and a slope of 1.10; this signifies a significant deviation in values between the two methods by a constant; this constant is higher at level 3 than level 2. Of note though is that for all three data sets the slopes are close to 1, meaning there is no great shift from the 45º line.

Passing-Bablok method, for level 2 returned an intercept of 0.46 and slope of 1, level 3 an intercept of 3.12 and a slope of 0.9, and the combined data an intercept of 0.29 and a slope = 1.08.

These findings are summarized in Table 4

Concordance

The data was also examined for concordance correlation coefficient (CCC) using R (Lawrence and Lin, 1989, Lin, 2011). See Table 5

The **correlation coefficient** of two variables in a data sample is a normalized measurement of how the two are linearly related. If the correlation coefficient is close to 1, it would indicate that the variables are positively linearly related and the scatter plot falls almost along a straight line with positive slope. For -1, it indicates that the variables are negatively linearly related and the scatter plot almost falls along a straight line with negative slope. And for zero, it would indicate a weak linear relationship between the variables. (Yau, 2009-2016).

Bland–Altman plots, where the differences between the two techniques are plotted against the averages of the two techniques, were obtained (Figure 1).

Discussion

Constant systematic errors are systematic deviations estimated as the average differences between the 2 methods. The presence of a constant systematic error indicates that one method measures consistently higher or lower in comparison with the other method. Proportional systematic error means that the differences between the 2 methods are proportionally related to the level of measurements (Jensen and Kjelgaard-Hansen, 2006).

The data clearly demonstrates that there is a constant systematic error between the two methods, with cobas values reading higher each time, and that this difference is significant at both levels. The constant error is demonstrated by the intercepts obtained by the regression analysis by different methods returning values that differ slightly, due to the difference in the calculation formula of each and considerations taken into account for each method, but showed positivity towards Cobas values (Table 3)

OLS method was only applied for level 2 since it is normally distributed, the result depicts a positive linear relationship between the two methods with positivity towards cobas values; this means cobas values are higher than architect values by a constant.

Concordance correlation of the individual levels 2 and 3 is poor. Even though the combined data gives a strong correlation; this is not of clinical significance since the range under consideration is the sub-clinical level of TSH where symptoms are not yet apparent in the patient. It is therefore more prudent to consider the differences between the levels individually.

Results returned by Deming show a slope of almost one for both the combined data and the two levels separately; this means both methods are in agreement however the intercepts show a y-leaning proportional deviation i.e. cobas values are higher than architect values.

Similarly, regression results by Passing-Bablok method returned slope values of almost 1 for all the data sets showing no constant difference but the intercepts indicate a significant y-leaning proportional deviation of all the values i.e. cobas values are higher than architect values.

The Bland Altman plots show the values distributed away from the equality line “0” and are all below zero signifying that for all paired data, again this is a demonstration that cobas values were higher (Figure 1)

Interestingly the BRI of architect is reported slightly higher than that of cobas (Table 1), it is beyond the scope of this study to further investigate how these values were obtained but they are now questionable in view of the results of this and other similar studies.

Conclusion

Both regression models show minimal constant deviation but a significant proportional deviation towards cobas values and these are the same conclusions Sarkar (2014) made, further inferring that much as the regression analysis shows transferability of results from one method to another, the difference in means and the poor concordance correlation obliges a consideration of the practicality in patient management of thyroid disorders especially so in the subclinical range.

Considering the current normal range (0.25 - 5.0) UIu/ml, it is clear that patient management will differ according to the value reported; this is especially significant in TSH values up to 10UIu/ml, considered the subclinical range where symptoms of hypothyroidism are not yet apparent and can only be detected by the lab test.

Subclinical thyroid disease is, by its very nature, a laboratory diagnosis. Patients with subclinical disease have few or no definitive clinical signs or symptoms of thyroid dysfunction. Thus, it is critically important that the normal reference range for TSH be standardized and that laboratories engage in appropriate quality control procedures to ensure that the results they report are accurate and reproducible (Surks et al., 2004)

Currently the laboratory has the appropriate quality assurance program in place; this study shows that there is a significant difference between the two methods evidenced by the poor correlation and the significant proportional deviation towards cobas values. Furthermore even patient values show the same pattern of difference that eventually led to this study.

Figures and tables

Table 1. Performance specification of Architect i2000SR and Cobas 6000 as adopted from (Sarkar, 2014)

Performance indicator	Architect	Cobas
Analytical sensitivity (value lying 2 SD above the zero calibrator in a repeatability study):	0.0025 μIU/mL	0.005 μIU/mL
Functional sensitivity (lowest analyte concentration that can be reproducibly measured with a across-run precision of 20 % CV):	0.0036 μIU/mL	0.014 μIU/mL
Analytical specificity	<10 % Cross reactivity in human serum samples containing TSH in the normal range was observed with FSH ≤ 500 mIU/mL, LH ≤ 500 mIU/mL and hCG ≤ 200,000 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	0.0025–100 μIU/mL (or up to 1,000 μIU/mL for 10 fold diluted samples)	0.005–100 μIU/mL (or up to 1,000 μIU/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)	1.8–4.6 %	1.2–5.6 %
Interferences	<10 % interferences were observed by haemolysis up to 500 mg/dL of Hb, lipaemia up to 3,000 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 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 IU/mL of rheumatoid factors
Biological reference intervals	0.35–4.94 μIU/mL (n = 549)	0.27–4.20 μIU/mL (n = 516)

Table 2. EQA RIQAS immunoassay cycle 43 sample 9 data used for estimation of sample size

RIQAS Cycle43 sample 9	**ARCHITECT i2000 SR**	**COBAS e601**
Mean	6.451	7.371
SDI	-1.10	0.06
N (sample size from which the mean was calculated)	125	148

Table 3. A summary of the data as obtained from Microsoft Excel 2007

	ARCHITECT	**COBAS**
Mean	8.55	9.66
Standard Error	0.27	0.30
Median	12.61	14.54
Mode	14	2.5

Standard Deviation	6.57	7.20
Sample Variance	43.18	51.83
Kurtosis	-1.85	-1.88
Skewness	0.08	0.06
Range	18.91	20.79
Minimum	1.67	2.21
Maximum	20.58	23
Count	590	590
Confidence Level(95.0%)	0.53	0.58

Table 5 Regression analysis findings

	Level2		**Level 3**		**Combined**	
Sample size (n)	293		297		590	
	Deming	Passing-Bablok	Deming	Passing-Bablok	Deming	Passing-Bablok
Intercept	0.55	0.46	2.34	3.12	0.28	0.29
95% CI of intercept	-0.10 to 1.19	0.06 to 0.79	1.25 to 3.42	1.94 to 4.22	0.20 to 0.37	0.27to 0.32
Slope	0.96	1	0.96	0.90	1.10	1.08
95% CI of slope	0.64 - 1.27	0.84 to 1.2	0.89 - 1.03	0.83 - 0.98	1.09 - 1.10	1.08-1.09
Pearson's r		0.33		0.84		0.996

Table 5. Concordance correlation (Lawrence and Lin, 1989); A bias correction factor measures how far the best-fit line deviates from a line at 45 degrees. No deviation from the 45 degree line occurs when C.b = 1

	Level 2	**Level 3**	**Combined data**
Sample size	293	297	590
Concordance Correlation Coefficient (CCC)	0.03	0.50	0.98
CCC Range	0.02- 0.05	0.45 – 0.55	0.976 – 0.981
Pearson’s r (precision)	0.33	0.84	0.996
Bias Correction factor (accuracy)	0.10	0.60	0.98

Table6. A summary of the initial statistical analysis of the data. Shapiro test for distribution pattern, paired T-test for significance of difference in normally distributed data (level 2) and Mann-Whitney U-Wilcoxon analysis for data not normally distributed (level 3 and the combined data)

Method		**Architect**			**Cobas**		
		Level 2	Level 3	Combined	Level 2	Level 3	Combined
N		293	297	590	292	297	590
Shapiro-Wilk normality test	W	1.0	0.78	0.76	1.0	0.75	0.73
	p value	0.5023	<2.2e-16	<2.2e-16	0.2469	< 2.2e-16	<2.2e-16
Distribution		Normal	Not Normal	Not Normal	Normal	Not Normal	Not Normal
Paired T-test for Level 2	T	62.39		NA		NA	
	Df	292					
	p value	< 2.2e-16					
	CI	95%					
	mean of differences	0.46					
Mann-Whitney U-Wilcoxon	NA		V	44228		V	174210
			p value	< 2.2 e-16		p value	< 2.2 e-16
			true location shift	not zero		true location shift	not zero

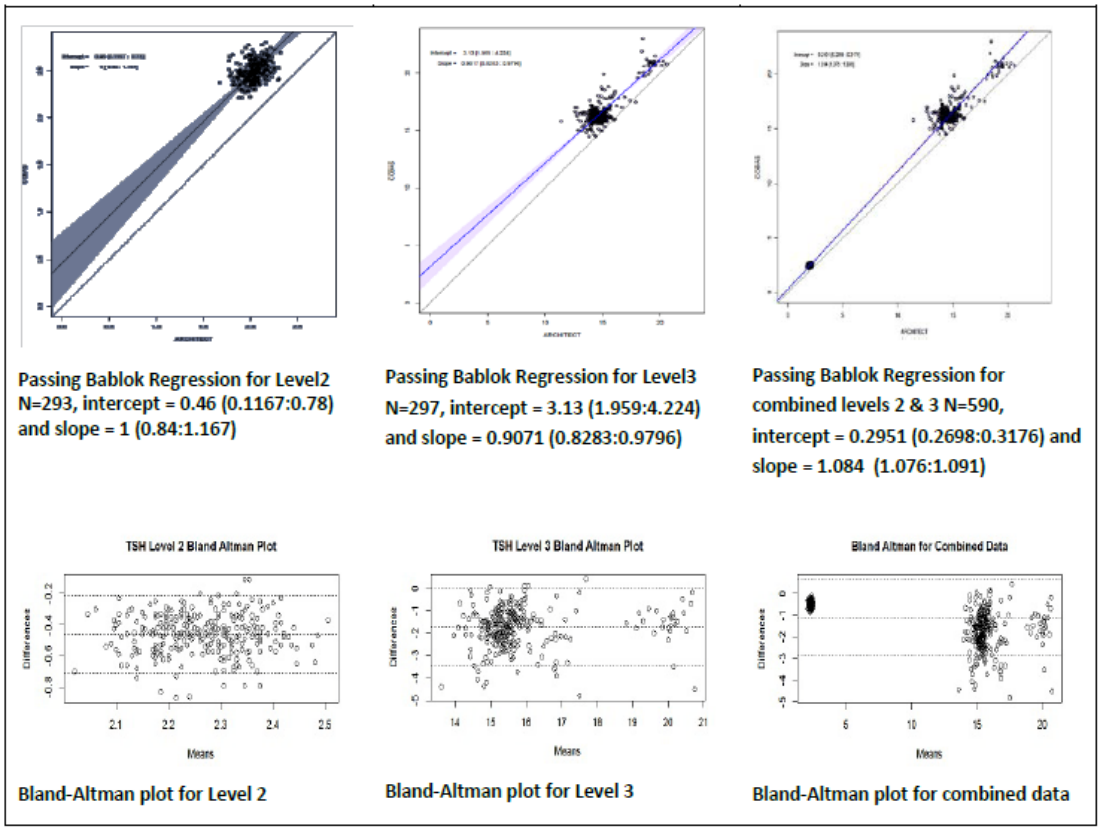

Figure 1. Passing-Bablok regression analysis and bland altman plots for level 2, 3 and combined data

Acknowledgement

No study, no matter how brief, is really achieved single handed. I might have worked alone in gathering information, data and putting together the final report, but was in constant consultation seeking advice from Drs. Christopher Gontier in writing the report and all it entails, and Dr. Paul Makau for introducing and teaching me the statistical application R and the referencing application Endnote. It is with immense gratitude that I acknowledge their support and help.

I thank The Nairobi Hospital (TNH) for my job and opportunity to carry out the study, the TNH Bio-Ethics and Research Committee for permission to use hospital data, my supervisor in special chemistry Mr. John Mugo for supporting the data with his own observation and experience, my biochemistry supervisor Mrs Felister Wazome for organizing time off for me to attend exams and live sessions during the coursework, my colleagues for their moral support and my family for believing in me. A special thanks to all the student coordinators at Texila American University for their online support. Full online learning is not easy without such support.

Above all I thank God for allowing me the inspiration, capacity and energy to do it all.

References

[1]. BALOCH, Z., CARAYON, P., CONTE-DEVOLX, B., DEMERS, L. M., FELDT-RASMUSSEN, U., HENRY, J. F., LIVOSLI, V. A., NICCOLI-SIRE, P., JOHN, R., RUF, J., SMYTH, P. P., SPENCER, C. A. & STOCKIGT, J. R. 2003. Laboratory medicine practice guidelines. Laboratory support for the diagnosis and monitoring of thyroid disease. Thyroid, 13, 3-126.

[2]. BENDIX CARSTENSEN, L. G., CLAUS EKSTROM AND MICHAL FIGURSKI 2015. MethComp: Functions for Analysis of Agreement in Method Comparison Studies. R package version 1.22.2.

[3]. FAIX, J. D., AND THIENPONT, LINDA M 2013. Thyroid-Stimulating Hormone; Why Efforts to Harmonize Testing Are Critical to Patient Care. Clinical Laboratory News.

[4]. HENDRIKS, H. A., KORTLANDT, W. & VERWEIJ, W. M. 2000. Standardized comparison of processing capacity and efficiency of five new-generation immunoassay analyzers. Clin Chem, 46, 105-11.
[5]. JENSEN, A. L. & KJELGAARD-HANSEN, M. 2006. Method comparison in the clinical laboratory. Veterinary Clinical Pathology, 35, 276-286.
[6]. LAWRENCE, I. & LIN, K. 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics, 255-268.
[7]. LIN, Y. Y. A. L. 2011. Statistical Tools for Measuring Agreement. 0.8-1 ed.
[8]. (NIDDK), N. I. O. D. A. D. A. K. D. 2014. Thyroid Tests [Online]. NIH Publication No. 14–6284. Available: http://www.endocrine.niddk.nih.gov/ [Accessed January 1st 2016].
[9]. ODELL, W. D., WILBER, J. F. & PAUL, W. E. 1965. Radioimmunoassay of thyrotropin in human serum. The Journal of Clinical Endocrinology & Metabolism, 25, 1179-1188.
[10]. R CORE TEAM 2015. R: A language and environment for statistical computing 3.2.3 ed. Vienna, Austria: The R Foundation for Statistical Computing.
[11]. SARKAR, R. 2014. TSH Comparison Between Chemiluminescence (Architect) and Electrochemiluminescence (Cobas) Immunoassays: An Indian Population Perspective. Indian Journal of Clinical Biochemistry, 29, 189-195.
[12]. SERDAR, M. A., İSPIR, E., OZGURTAS, T., GULBAHAR, O., CIRACI, Z., PASAOGLU, H. & KURT, I. 2015. Comparison of four immunoassay analyzers for relationship between thyroid stimulating hormone (TSH) and free thyroxine (FT4)/[Dört immünolojik ölçüm cihazının tiroid stimüle edici hormon (TSH) ve serbest tiroksin (FT4) arasındaki ilişki yönünden karşılaştırılması]. Turkish Journal of Biochemistry, 40, 88-91.
[13]. SPENCER, C. A. 2013. Assay of thyroid hormones and related substances. In: DE GROOT LJ, C. G., DUNGAN K, ET AL., EDITORS: ENDOTEXT [INTERNET]. (ed.). South Dartmouth (MA): MDText.com, Inc.; 2000.
[14]. SURKS, M. I., ORTIZ, E., DANIELS, G. H. & ET AL. 2004. Subclinical thyroid disease: Scientific review and guidelines for diagnosis and management. JAMA, 291, 228-238.
[15]. UTIGER, R. D. 1965. Radioimmunoassay of human plasma thyrotropin. Journal of Clinical Investigation, 44, 1277.
[16]. WESTGARD, J. O., BARRY, P. L., QUAM, E. F. & EHRMEYER, S. S. 1999. Basic method validation: training in analytical quality management for healthcare laboratories, Westgard Quality Corporation.
[17]. WILLIAMS, D. P. 2015. 1400 Update of Status of standardisation of TFTs Royal Prince Alfred Hospital and University of Sydney.
[18]. YALOW, R. S. & BERSON, S. A. 1996. Immunoassay of endogenous plasma insulin in man. Obesity research, 4, 583-600.
[19]. YAU, C. 2009-2016. R Tutorial An Introduction to R Statistics. r-Tutor.com.