TRABAJO ORIGINAL

A Comparison of Serum Testosterone (T) Measurements by Various Methods, Employing Current Validated Techniques or Otherwise, and Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS). A Multicentric Study

Comparación de los resultados de testosterona (T) por cromatografía líquida en tandem con espectrometría de masa (LC-MS/MS) y por kits comerciales convalidados o no por LC-MS/MS

Scaglia HE, Buccini G, Chichizola C, Colombani ME, Corazza N, Corthey C, Guevel L, Ibañez G, Lacoste, E, Nosetto S, Piaggio R, Riesco O, Sandoz, S, Scaglia J, Viola, AM, Wolfthal D, Zylbersztein CC

Hormone Determination Laboratory, Hospital Italiano, La Plata, ARGENTINA. The various participants in this study preformed their hormone determinations in their private clinical biochemistry practices.

ABSTRACT

Objective: To compare normal and hirsute women Testosterone (T) measurements performed at different laboratories by the same or different methods, and the gold standard method LC-MS/MS (Quest Diagnostics, USA).

Design: Prospective study.

Setting: Hormone Determination Laboratory, Hospital Italiano, La Plata, and each participating laboratory's private practice.

Patient(s): Blood samples were obtained from 23 individuals sorted into two groups, namely, normal women, n: 11(NW) and hirsute women, n: 12 (HW).

Interventions(s): None.

Main Outcome Measure(s): To evaluate whether serum T measurements obtained from each serum by the methods currently employed in our country, some of whose kits exhibit changes in previous presentations, some LC-MS/MS-validated and other non-validated ones are significantly different from those obtained by LC-MS/MS.

Result(s): None of the 11 NW showed high T values by LC-MS/MS. Two out of the 12 hirsute patients showed normal T values (LC-MS/MS). Methods and number of participating labs -shown between brackets were: in NW, 1st generation Architect (1), 2nd generation Architect (1); Immulite (1) Cobas (4); Access (1); Centaur (2); Immunotech-RIA (1); and, in HW, 2nd generation Architect (3); Immulite (3); Cobas (4); Access (1); Centaur (2); Immunotech-RIA (1). No false positives resulted from the assays performed.

No lab yielded false positive results in the NW group. No false positives were reported from the 10 hirsute women with increased T values by LC-MS/MS. False positives, though, resulted from two female hirsute patients with normal T values studied by four of the methods.

Statistically, the serum T measurements obtained were significantly different by Centaur in NW and, in HW, by Immulite and Centaur as compared to LC-MS/MS. In the Bland-Altman plot, Centaur and Cobas showed over 5 % of measurements outside the limits of agreement in the HW group. Assessment by p-Spearman resulted in divergences with LC-MS/MS for all methods in NW, whereas in the HW group there were none. When estimating sampling bias for each laboratory taking LC-MS/MS as the reference method and adopting a ± 6.4 % mean bias acceptability criterion for each method compared to LC-MS/MS, two of the techniques reviewed, 2nd generation Architect and Cobas, met the validation requirement satisfactorily. However, one

lab out of three using 2nd generation Architect failed to meet the validation requirement, while two out of four labs using Cobas also failed to meet the requirement. This demonstrates the great variability among methods, even when labs are employing the same technique.

Conclusion: From the clinical point of view, the methods currently used in our local environment yielded no false positives or false negatives and therefore did not misdiagnose hyperandrogenism. Still, Immulite, Centaur, RIA and Access did present false positives in two of the T-normal hirsute women. The relation of serum T measurements obtained by each method to measurements obtained by LC-MS/MS reveals that the dispersion of the results was larger with values under 0.3 ng/ml, quite close to the detection limit of the various techniques. **Rev Argent Endocrinol Metab 52:137-152, 2015**

No financial conflicts of interest exist.

Key words: testosterone, testosterone assay, LC-MS/MS

RESUMEN

Objetivos: Comparar los resultados de testosterona **(T)** obtenidos en diferentes laboratorios en mujeres normales (MN) e hirsutas (MH), empleando el mismo o diferentes métodos respecto a la técnica de cromatografía líquida en tándem con espectrometría de masa LC-MS/MS (gold standard) realizada en el laboratorio Quest (USA)

Protocolo: Estudio prospectivo

Estudio realizado en: Laboratorio de Determinaciones Hormonales; Hospital Italiano de La Plata y en los laboratorios privados de cada participante

Pacientes: Se obtuvieron muestras de sangre periférica en 23 mujeres agrupadas en 2 grupos: controles normales (n:11) y en mujeres hirsutas (n: 12)

Intervención: Ninguna.

Principales resultados a evaluar: Evaluar si con los métodos habituales empleados en nuestro medio con los diferentes kits comerciales, de los cuales algunos han sido convalidados por técnica "gold standard" y otros no, presentan diferencias significativas con los obtenidos por LC-MS/MS

Resultados: Los resultados obtenidos por LC-MS/MS mostraron que ninguna de las 11 MN tuvieron niveles aumentado de T y 2 MH tuvieron valores normales de T. Los métodos empleados y el número de laboratorios (entre paréntesis) que emplearon cada método fueron en MN Architect 1st generation (1) Architect 2nd generation (1); Immulite (1); Cobas (4); Access (1); Centauro (2); Immunotech-RIA (1). En las MH Architect 2nd generation (3); Immulite (3); Cobas (4); Access (1); Centauro (2); Immunotech-RIA (1). En el grupo de MN en ningún laboratorio (lab) se obtuvieron resultados falsos positivos. En el grupo de MH no se obtuvieron falsos negativos en las 10 hirsutas con valores aumentados de T por LC-MS/MS. En las 2 pacientes hirsutas con T normal en 4 métodos se obtuvieron falsos positivos

Estadísticamente los resultados fueron significativamente diferentes, en las MN por Centauro y en las MH por Immulite y Centauro. En el análisis de Bland-Altman Centauro y Cobas en las MH presentaron más del 5 % de los resultados fuera del límite de acuerdo. Resultados por p-Speerman todos los métodos fueron diferentes a LC-MSMS en las MN y no se obtuvieron diferencias en el grupo de MH. Evaluando el bias de cada muestra en cada laboratorio respecto a LC-MS/MS y adoptando el criterio de aceptabilidad de ± 6,4 % mean bias de cada método respecto al de LC-MS/MS, 2 de las metodologías estudiadas, Architect 2da generación y Cobas pasaron satisfactoriamente el requisito de validación, sin embargo de los 3 laboratorios que emplearon 2da generación, 1 no pasó el criterio de validación y de los 4 que usaron Cobas, 2 tampoco lo pasaron. Esto demuestra la gran variabilidad de los métodos aun entre lab que emplean la misma técnica.

Conclusiones: Desde el punto de vista clínico los métodos habitualmente empleados en nuestro medio, no sobrediagnosticaron o subdiagnosticaron hiperandrogenismo, por no presentar falsos positivos o negativos respectivamente. Sin embargo Immulite, Centauro, RIA y Access presentaron falsos positivos en las 2 hirsutas con T normal. En la relación de los resultados de cada muestra en cada método sobre el valor de LC-MS/MS referido a la concentración de T en ese suero por LC-MS/MS, la mayor dispersión de los resultados se observaron con valores menores de 0,3 ng/ml, muy cercano al límite de detección de las diferentes técnicas. **Rev Argent Endocrinol Metab 52:137-152, 2015**

Los autores declaran no poseer conflictos de interés

Palabras clave: testosterona, determinación de testosterona, LC-MS/MS

INTRODUCTION

Markers in the diagnosis of hyperandrogenemia with or without hirsutism in women include increased levels of serum T. It has been widely demonstrated that commercial immunoassays show great disparity at low T levels such as are usually found in women. Based on these findings, in a previous study we quantified T levels in aliquots of the same serum sample in 24 normal women (NW) and 15 hirsute women (HW). Aliquots of each sample were assayed in 18 laboratories by means of the current methods in use locally: 5 automated methods (Abbott: Axsym and Architect; Vidas; Siemens Immulite and Roche Elecsys); 3 manual coated tube RIA: Siemens, DSL, DiaSource and one in-house-developed RIA method by one of the laboratories following the consensus guidelines from the Endocrine Society[1]. Serum T measurements obtained for each serum by the methods currently employed in each lab were compared to those obtained by LC-MS/MS (Quest Diagnostics, USA)[2].

In conclusion, and in keeping with results obtained in previous studies[1,3-5], our findings clearly demonstrate that none of the immunoassays employed locally satisfactorily meets the degree of precision required for evaluating T concentrations under 3.0 ng/ml, as corresponds to the measurements obtained for women and children.

In our previous study, and in line with earlier publications[6], we showed that it is an issue of paramount relevance to be able to compare testosterone measurements resulting from assays by various testing methods and by the same method implemented at different labs. As an attempt to standardize pre- and post analytical aspects of the assessment process and to establish a reliable, consistent reference range, which remains an enduring problem with laboratories, the Centers for Disease Control and Prevention (CDC)'s Hormone Standardization Project (CDC-HoST Program), National Center for Environment Health, Division of Laboratory Sciences, has put forward the following guidelines:
a) Implementing a standardization program to assay steroid hormones[7]
b) Establishing a common calibration standard, that is, a primary standard[8]
c) Specifying the matrix employed by commercial kit manufacturers to calibrate their immunoassays[9].

d) Referring the technique of choice to a reference method[10,11]. It is crucially important to accurately establish the reference ranges obtained from a population, which should, in turn, be well characterized with an adequate sample size of and using a well-established standardization procedure.

At present some laboratories supplying commercial kits for Testosterone testing claim in their inserts that results are validated with LC-MS/MS.

The aims of this study were as follows:
1. Ascertaining whether the serum T measurements obtained by testing with these new kits tie in with those obtained with a gold standard LC-MS/MS method. To that end, we carried out a fresh comparative study of T testing in normal and hirsute women using various commercial kits, some of which claim in their inserts that the method in question is LC-MS/MS-validated and others that do not specify such validation. The measurements thus obtained were statistically gauged between methods and by reference to Quest Diagnostics LC-MS/MS, in turn validated by the Centers for Disease Control and Prevention Hormone Standardization Program (CDC-Host Program), Clinical Laboratory Improvement Amendments (CLIA) and FDA.
2. Comparing T measurements in normal women by employing non-validated Abbott Architect, which we shall identify as 1st generation Architect, and 2nd generation Architect, this latter one validated by LC-MS/MS, so as to check whether the alterations made indeed significantly modify T values.

MATERIALS AND METHODS

Our multicentric study was carried out by 17 researchers in 14 clinical biochemistry laboratories in which blood samples corresponding to various clinical conditions were collected following previously established inclusion / exclusion criteria.

The subjects whose serum samples were analyzed in this study were not under any treatment whatsoever and did not present any endocrine disorders or non-endocrine disease that could interfere with T determination. Hirsute women showed an increased score according to Ferriman–Gallwey[12] criteria, with or without acne and / or loss of scalp hair, and with or without menstrual cycle alterations. Normal women did not present

hirsutism, had regular menstrual cycles and ovulatory progesterone levels measured in the luteal phase in some of them.

Blood samples were obtained from 25 individuals sorted into two groups, namely, normal women (n: 11) and hirsute women (n: 12). Two blood samples corresponding to 2 hirsute patients were discarded.

The blood samples were collected in tubes containing no anti-coagulants or preservatives and the serum was obtained by centrifugation. For all samples, 13 aliquots of at least 0.5 ml were obtained and kept at -20ºC until they were sent to each of the labs participating in the study for assay of serum T by the methods currently used by each participating lab. A 2ml aliquot was kept until all the samples had been collected, so as to have them all sent on together for assay by LC-MS/MS (Quest Diagnostics, USA).

The techniques employed are detailed as follows:

Normal women:
Abbott Architect - 1st generation: 1 lab;
Abbott Architect - 2nd generation: 1 lab;
Siemens Immulite: 1 lab;
Roche Cobas: 4 labs;
Beckman Access: 1 lab;
Centaur: 2 labs;
Bio Analytical Immunotech - manual coated tube RIA: 1 lab.

Hirsute women:
Abbott Architect - 2nd generation: 3 labs;
Siemens – Immulite: 3 labs;
Roche – Cobas: 4 labs;
Beckman – Access: 1 lab;
Centaur: 2 labs;
Bio Analytical Immunotech - manual coated tube RIA: 1 lab.

The calibration standards of the methods in this study correspond to their inserts. Of all the methods employed in this study, only second generation Architect – Abbott and Cobas – Roche report that they have been validated with LC-MS/MS.

Blood sampling for both groups was performed in two separate stages four months apart from each other. Within this intermediate period some of the labs changed their methods, which explains why in some cases the methodology differed from one group to another. The reason that there are three labs less in the NW group is that one of the labs originally used the Abbott Axsym technique and later changed to 2nd generation Abbott Architect.

Abbott Axsym was not used in the HW group, which, for the purpose of this study, resulted in the need to dispose of the measurements obtained in the NW group by such method. Another two labs were unable to process the samples due to technical reasons.

The measurements obtained from each serum by the methods currently employed in each lab were compared to those obtained by LC-MS/MS (Quest Diagnostics, USA)[12].

STATISTICAL STUDY

Descriptive statistics: mean, standard deviation (SD), median, minimum and maximum were calculated for each one of the methods in both experimental groups.

The serum T concentrations measured through each one of the 7 methods with normal women and of the 6 methods used with hirsute women were compared with those obtained by LC-MS/MS using paired-sample Wilcoxon non-parametric signed-rank test. Pearson's correlation coefficient was calculated among the values obtained by LC-MS/MS and by each one of the methods. The test was done in order to analyze whether the correlation was significantly different from 0 (zero).

In order to analyze the concordance of each one of the methods with one another and with LC-MSMS, Bland–Altman[13] plot was performed. Limits of agreement were calculated by estimating the mean difference ± 1.96 standard deviation (STD) of the differences.

Weighted Deming regression[14] was run on each of the different determination procedures from which LC-MS/MS was taken as independent variable (x) and the method as dependent variable (y). This equation permits finding the ratio between the two methods.

Overall mean bias was also calculated for each lab versus LC-MS/MS.

This research was carried out pursuant to the ethical standards of the Declaration of Helsinki.

RESULTS

T measurements were analyzed in serum samples from 23 individuals sorted out into 2 groups: NW (n: 11) and HW (n: 12). Quantifications of T were performed in 14 laboratories from aliquots of the same blood sample by applying both the testing

techniques customarily used in such labs and LC-MS/MS. This paper reports a comparative analysis of the measurements from NW and HW resulting from T testing by various methods in current use in our area against those obtained by LC-MS/MS.

Figure 1 shows results of T tested by LC-MS/MS on normal women and on patients with hirsutism, with or without acne and with or without cycle abnormalities (Quest Diagnostics, Nichols Institute). As can be seen, the method yielded increased T values for 10 out of the 12 hirsute women.

Figure 2 shows marked differences in the measurements of aliquots from 11 NW by the early version of Abbott Architect, 1st generation Architect, and by the modified LC-MS/MS-validated technique, 2nd generation Abbot Architect. As can be observed, there was a significant difference between the two methods.

Figures 3 to 7 reflect the relation between serum Testosterone levels obtained by different methods and by the different labs working with the same methodology (3 labs with 2nd generation Architect, 4 labs with Cobas, 3 labs with Immulite and 2 labs with Centaur). The figures also show comparative results between laboratories using the same method and LC-MS/MS.

With 2nd generation Architect (Figure 3) none of the serum samples yielded increased T levels (no false positives resulted from the assay) in the NW group. As for the HW group, whenever T levels were increased by LC-MS/MS, they were so by the former as well (there was no record of false positives, either). As regards the two hirsute patients with normal serum T values by LC-MS/MS, their measurements were also normal according to the standards referred to as normal values in the commercial kit insert (up to 0.55ng/ml). The HW group showed good matching among the different labs, between each other and by reference to LC-MS/MS, with measurements slightly above, though not significantly higher, in Lab 3.

No false positives or false negatives were found in the four labs testing by Cobas (Figure 4) in the NW group and the HW group respectively. Note, though, that there is greater dispersion in the NW group and very good matching among the participating labs and against LC-MS/MS (Figure 5).

Testing by Immulite (Figure 6) did not yield false positives in the NW group. No false negatives resulted from the HW group, either. In the latter group great dispersion was observed in the measurements from the various labs (bottom right-hand side of the figure). In Lab 3, sample 12, which had yielded normal values, revealed an increased T value (false positive) by LC-MS/MS.

In the two labs using Centaur (Figure 7 top left) the measurements obtained for all the samples from the NW group were higher than those resulting from LC-MS/MS testing. In both labs, sample
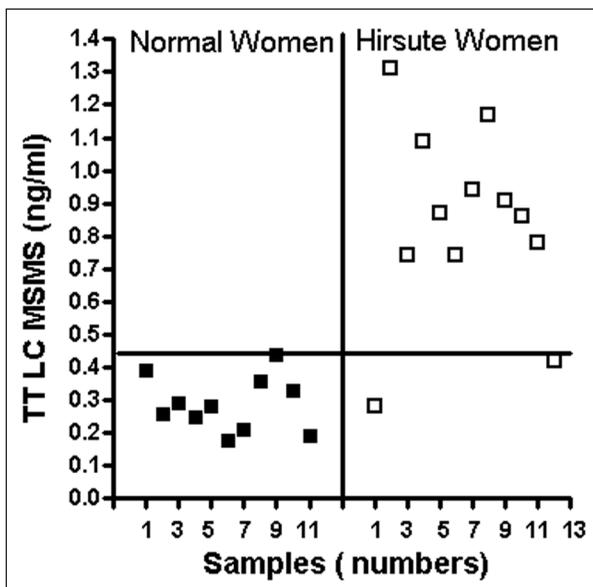

Figure 1. Measurements of TT tested by LC-MS/MS in normal women and in patients with hirsutism, with or without acne and with or without cycle abnormalities (Quest Diagnostics, Nichols Institute).
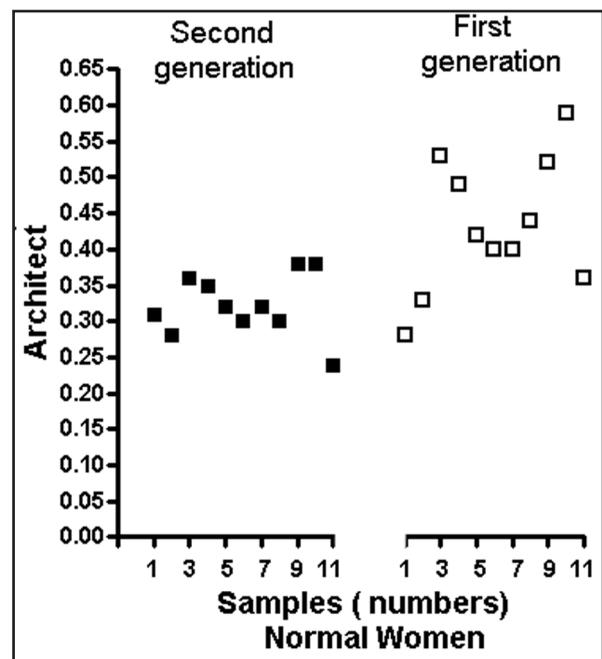

Figure 2. Testosterone measurements in same serum aliquots from 11 normal women as determined by 1st generation Architect (right side of figure) against 2nd generation Architect (left side of figure).
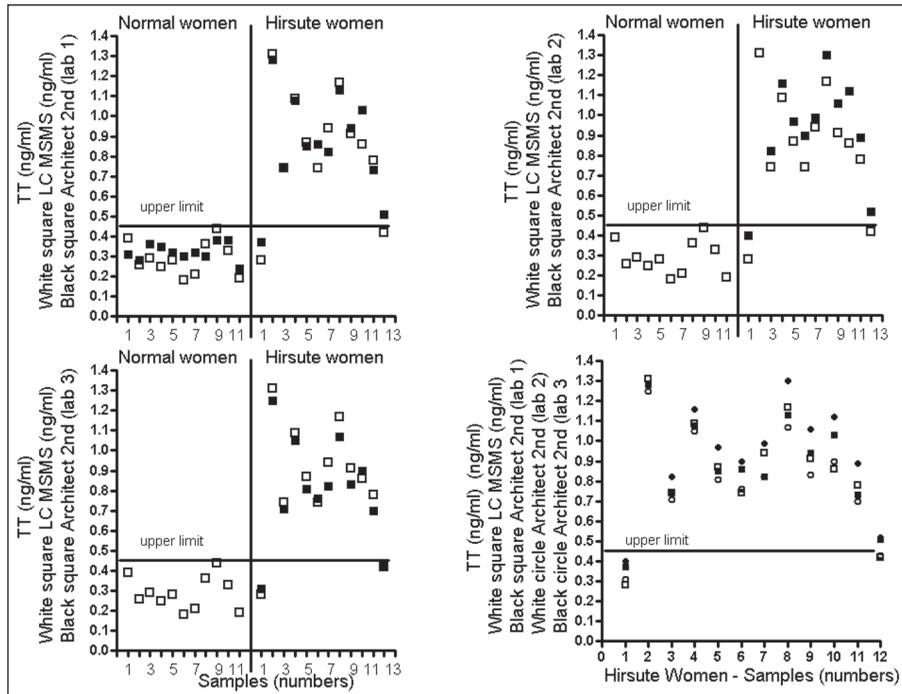
Figure 3. shows the relation between Testosterone levels obtained by the three laboratories (lab1, lab 2 and lab 3, top left, top right and bottom left, respectively) using Abbott and 2nd generation Architect commercial kits (■), and LC-MS/MS (□) both in normal and hirsute women. The straight line of each set indicates the upper limit for normal women obtained by LC-MS/MS. Bottom right is shown the relation between every LC-MS/MS-determined serum and each of the three labs involved for the hirsute women group. (LC-MS/MS □; Architect 2nd generation, laboratories 1■, 2 ○ and 3 ●, respectively)
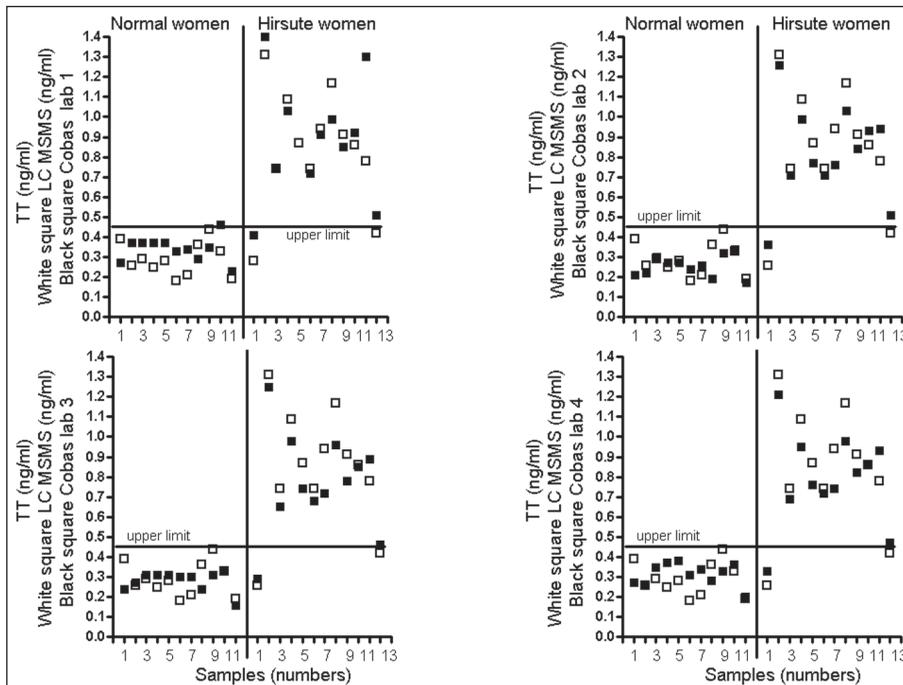

Figure 4. shows the relation between Testosterone levels obtained by the four labs (labs 1, 2, 3 and 4 top left, top right, bottom left and bottom right, respectively), using the Roche commercial kit -Cobas (■) and LC-MS/MS (□) in both normal and hirsute women. The straight line of each set indicates the upper limit for normal women obtained by LC-MS/MS.
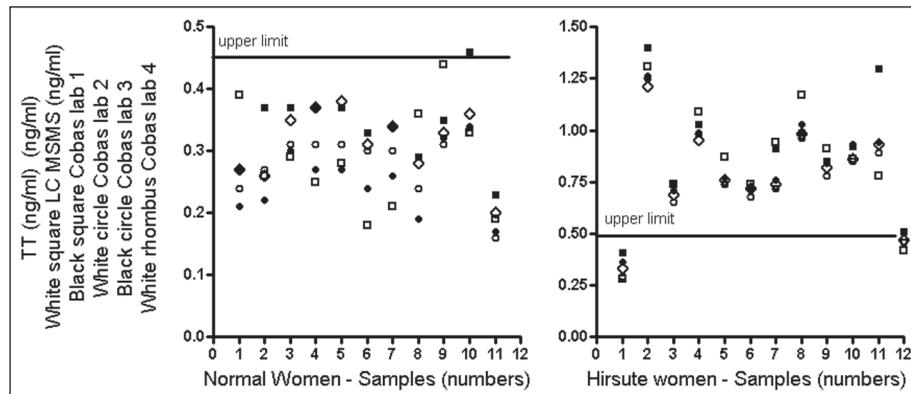
Figure 5. Relation of Testosterone levels in each LC-MS/MS-determined serum (□) and for each of the four labs, using the Roche commercial kit -Cobas (lab 1 ■), (lab 2 ○); (lab 3 ●) (lab 4 ◇) both in normal women (left) and in hirsute women (right).
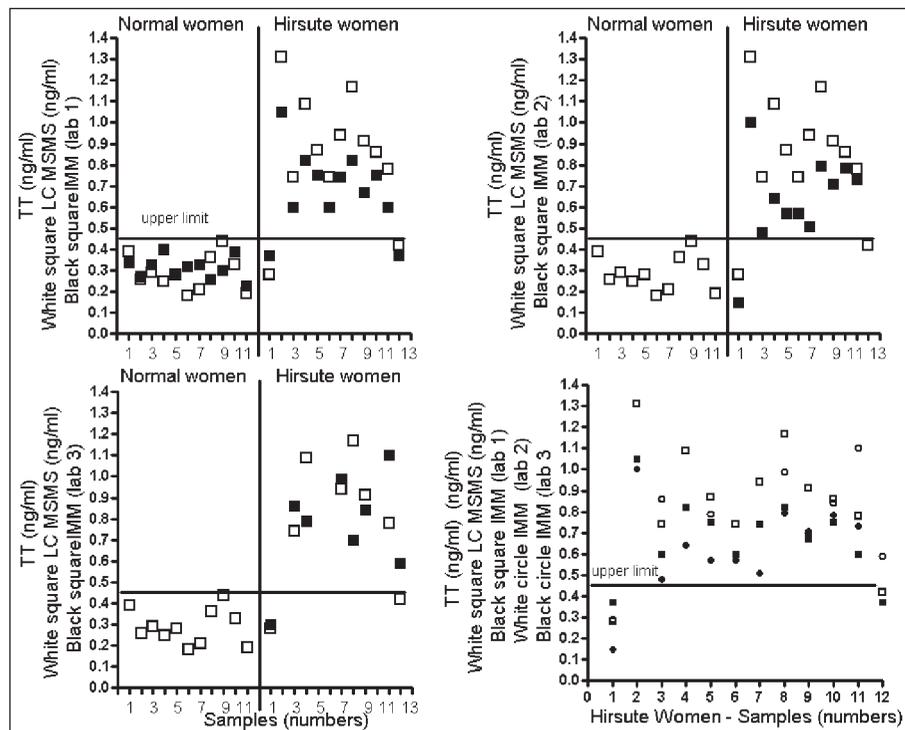

Figure 6. shows the relation between Testosterone levels obtained by the three labs (lab 1, 2 and 3 top left, top right and bottom left, respectively), utilizing the Siemens-Immulite commercial kit (■) and LC-MS/MS (□) both in normal and hirsute women. The straight line of each set indicates the upper limit for normal women obtained by LC-MS/MS. Bottom right is shown the relation of each LC-MS/MS-determined serum and each of the three labs involved for the hirsute women group (LC-MS/MS □; Immulite, laboratories 1■, 2 ● and 3 ○, respectively)

12 in the HW group showed increased T levels by Centaur compared to LC-MS/MS, and also by reference to the average normal values obtained in the course of the present study (left-hand side of Figure 7) and to the cut-off values of the kit insert (false positive).

Measurements obtained by Access technique in the NW group did not present false positives. In the HW group, samples 1 and 12, which had presented normal T values by LC-MS/MS, showed higher values by Access than the normal values obtained in this study as can be observed in the
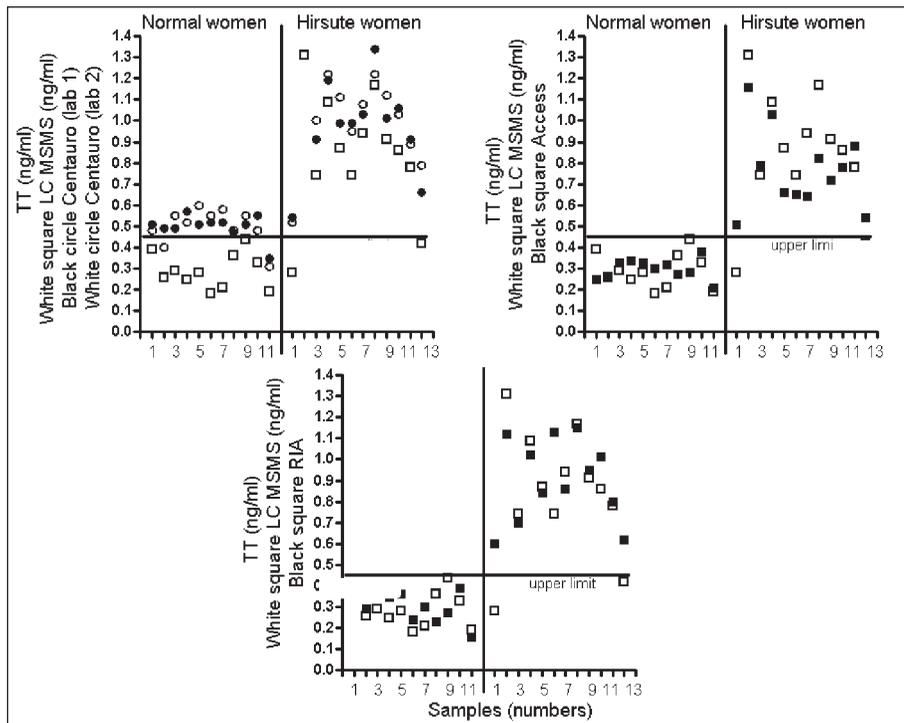
Figure 7. shows the relation between Testosterone levels obtained both in normal and hirsute women by two labs using Centauro method, lab 1● and lab 2 ○ (top left), another lab using Access (top right) and another one using manual RIA by means of an Immunotech kit (bottom left), and by LC-MS/MS (□) in normal and hirsute women. The straight line of each set indicates the upper limit for normal women obtained by LC-MS/MS.

Figure 8. Relation between Testosterone levels obtained by various methods and by LC–MS/MS in normal women (Y axis) compared to the level obtained by LC MSMS (X axis). The different symbols in each set of measurements indicate the various laboratories that employed the same methodology.

Figure 9. Relation between Testosterone levels obtained by various methods and by LC–MS/MS in hirsute women (Y axis) compared to the level obtained by LC MSMS (X axis). The different symbols in each set of measurements indicate the various laboratories that employed the same methodology.
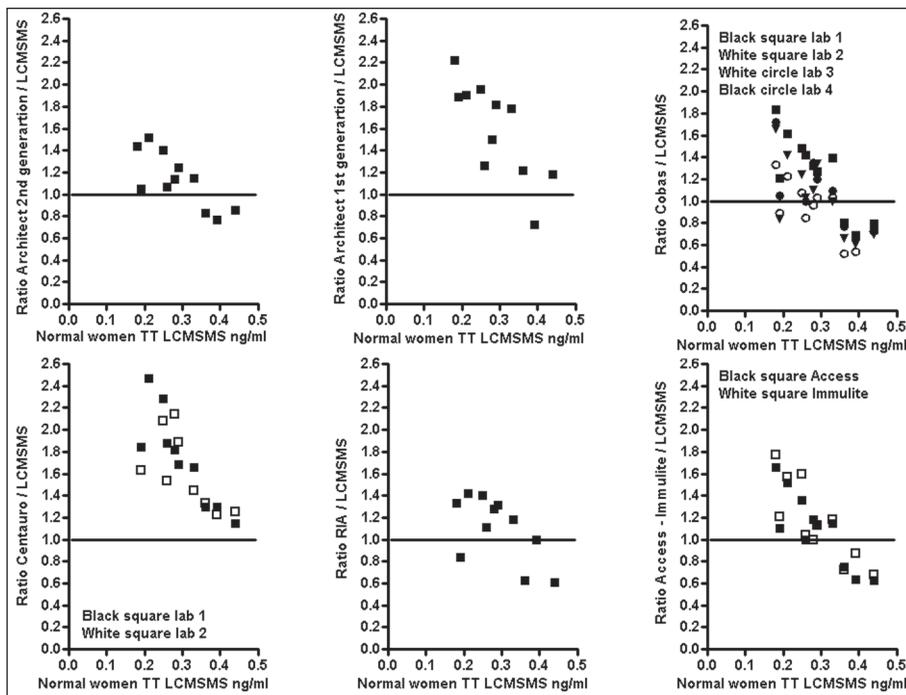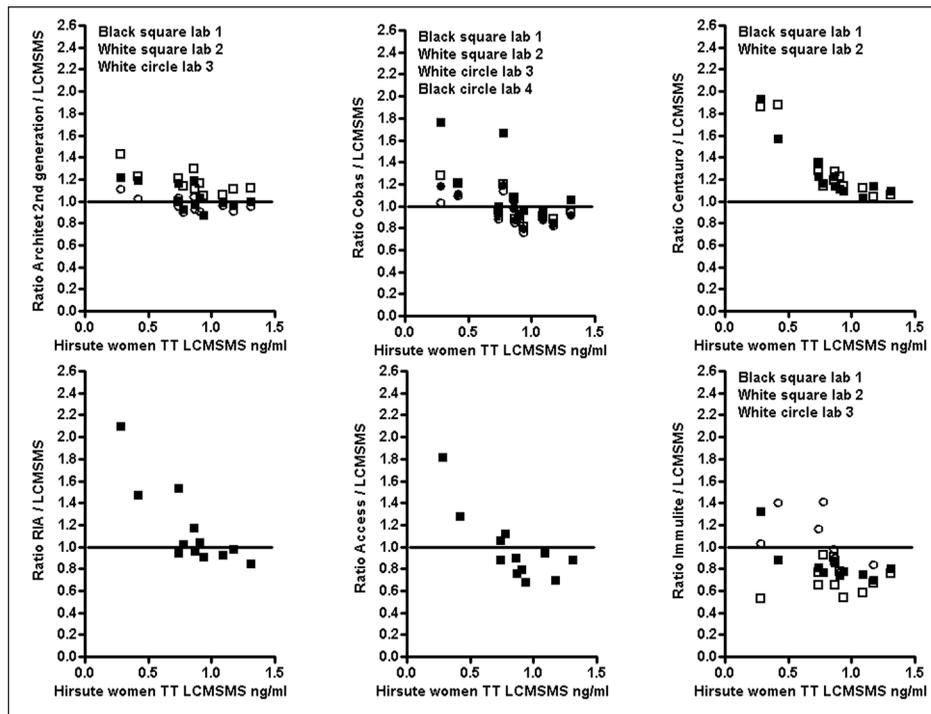
left-hand side of the figure (false positive, Figure 7 top right).

Measurements by RIA (Figure 7 bottom left) did not exhibit any false positives in the NW group and, in the two hirsute patients with normal T values by LC-MS/MS (cases 1 and 11), increased T values were obtained against those of LC-MS/MS and even against the average normal values resulting from the present study (false positives).

On running Wilcoxon paired-sample test there were significant differences in measurements by 1st generation Architect and Centaur by reference to LC-MS/MS (p<0.05) in the NW group, whereas in the HW group there were differences for Architect (Lab 2), Cobas (Lab 3), Centaur and Immulite (both Labs 2 and 3) (See Tables 1 and 2)

By Bland-Altman plot T measurements from the NW group obtained by 1st generation Architect and Immulite (Lab 1), and from the HW group, by 2nd generation Architect (Lab 2), Centaur (Lab 2), Immulite (Lab 1) and Cobas (Lab 1) revealed one T value outside the limit of acceptability calculated as the difference of the mean SD ±1.96 of the difference between the measurement by LC-MS/MS and those for each method (Tables 3 and 4).

As regards Spearman's correlation coefficient, results were not statistically significant (p> 0.05 in all cases) in the NW group. However, with hirsute women all the methods showed a significant correlation (p< 0.05) (Tables 3 y 4). It is worth noting that in the NW group all the measurements fell close to the method quantification limit, which causes greater scattering of results.

The results of Weighted Deming regression are presented in Table 5. It can be seen that the values match those obtained by LC-MS/MS, since the closer to 1 the value of the slopes (and the closer to zero the values of the intercept) the closer the values of x and y, i.e. the method in question and the gold standard.

Figure 8 shows how measurements by each method relate to those by LC-MS/MS in the NW group and how they compare to the concentration obtained by LC-MS/MS. Greater scattering of results is apparent in all cases by reference to the ideal value 1, which proves excessive for 1st generation Architect and Centaur when T levels fall below 0.3 ng/ml.

Figure 9 shows how HW measurements by each method relate to those obtained by LC-MS/MS and how they compare to the LC-MS/MS concentration

TABLE 1. Testosterone immune assay values obtained by 11 different laboratories and LC MSMS from normal women

	LC MSMS	Arch 2 G	Arch 1 G	RIA	Acc	Cobas lab 1	Cobas lab 2	Cobas lab 3	Cobas lab 4	Cent lab 1	Cent lab 2	Imm
Mean	0.29	0.32	0.43	0.31	0.30	0.34	0.25	0.28	0.31	0.50	0.50	0.31
SD	0.08	0.04	0.09	0.08	0.05	0.06	0.05	0.05	0.06	0.06	0.08	0.05
n	11	11	11	11	11	11	11	11	11	11	11	11
Median	0.27	0.32	0.42	0.30	0.30	0.35	0.26	0.30	0.33	0.51	0.52	0.32
Min	0.18	0.24	0.28	0.16	0.21	0.23	0.17	0.16	0.20	0.35	0.31	0.23
Max	0.44	0.38	0.59	0.39	0.38	0.46	0.34	0.33	0.38	0.57	0.60	0.4
"p"		0.213	**0.011**	0.507	0.721	0.100	0.423	0.760	**0.300**	**0.003**	0.003	0.359

Note: LCMSMS Liquid Chromatography tandem Mass Spectrometry. Min: Minimum; Max: Maximum
"p" Wilcoxon study
Arch 2G Architect 2nd generation (Abbott); Arch 1 G: Architect 1st generation (Abbott); Cobas (Roche)
RIA: Radioimmunoassay (Immunotech); Acc: Access; Cent: Centauro; IMM:Immulite.
p< 0.05 statistically different

TABLE 2. Testosterone immune assay values obtained by 13 different laboratories and LC MSMS from hirsute women

	LC MSMS	Arch 2 G lab 1	Arch 2 G lab 2	Arch 2 G lab 3	RIA	Acc	Cobas lab 1	Cobas lab 2	Cobas lab 3	Cobas lab 4	Cent lab 1	Cent lab 2	Imm lab1	Imm lab2
Mean	0.84	0.86	0.97	0.80	0.90	0.77	0.89	0.82	0.77	0.79	1.01	1.04	0.68	0.63
SD	0.29	0.26	0.30	0.26	0.19	0.19	0.30	0.23	0.25	0.23	0.25	0.25	0.19	0.22
n	12	12	12	12	12	12	11	12	12	12	12	12	12	11
Median	0.87	0.86	0.98	0.82	0.91	0.75	0.91	0.81	0.76	0.79	1.00	1.06	0.71	0.64
Min	0.28	0.37	0.40	0.31	0.60	0.51	0.41	0.36	0.29	0.33	0.54	0.52	0.37	0.15
Max	1.31	1.28	1.47	1.25	1.15	1.16	2.48	1.25	1.26	1.21	1.43	0.54	1.05	1.00
"p"		0.625	**.002**	0.31	0.56	0.182	0.385	0.410	**0.028**	0.091	**.002**	**.002**	**.004**	**.003**

Note: LCMSMS Liquid Chromatography *tandem* Mass Spectrometry. Min: Minimum; Max: Maximum
"p" Wilcoxon study
Arch 2G Architect 2nd generation (Abbott); RIA: Radioimmunoassay (Immunotech); Acc: Access; Cent: Centauro;
IMM:Immulite;Cobas (Roche). p< 0.05 statistically different

obtained. It is clear that for Architect and Roche there was a significant LC-MS/MS correlation. For both labs using Centaur, although excellent matching was achieved between the two, in all cases the relation was greater than 1. Immulite displayed great dispersion among the values obtained and Access and RIA showed a greater dispersion at concentrations below 0.5 ng/ml.

DISCUSSION

The comparative study of serum T obtained from the same normal and hirsute women aliquots, tested at different labs and in relation to each other, versus the Quest Diagnostics method using turbulent flow liquid chromatography-tandem mass spectrometry[14] revealed significant differences for all the methods in the NW group, though not for all those in the HW group.

In hirsute women, T testing proves an essential diagnostic parameter in defining the physiological condition to be used jointly with clinical evaluation and ultrasound imaging. Indeed, hirsutism is associated with various conditions, most of which are defined by biochemical hyperandrogenism, that is, by levels of circulating androgens. As a first evaluation parameter of androgenic action Dehydroepiandrosterone-sulfate is assessed as a precursor of mainly adrenal origin and Total Testosterone and/or free Testosterone.

TABLE 3. Bland Altman ans Spearman correlation coefficient of the values of Testosterone by different methods compared to LCMSMS from normal women

	Arch 2 G lab 1	Arch 1 G lab 2	RIA	Acc	Cobas lab 1	Cobas lab 2	Cobas lab 3	Cobas lab 4	Cent lab 1	Cent lab 2	Imm lab 1
Bland-Altman analysis											
Mean difference	0.144	0.144	0.016	0.008	0.052	-0.035	-0.009	0.025	0.210	0.211	0.025
SD	0.071	0.109	0.092	0.097	0.098	0.084	0.090	0.094	0.090	0.109	0.094
Values outside the agreement limits mean ±1.96SD	0	1 (9.1%)	0	0	0	0	0	0	0	0	1 (10%)
Spearman coefficient	0.494	0.328	0.383	-0.023	0.121	0.301	0.201	0.073	-0.129	-0.037	0.729
sig bilateral – p (p Spearman)	**0.122**	**0.325**	**0.245**	**0.947**	**0.723**	**0.369**	**0.553**	**0.818**	**0.705**	**0.914**	**0.729**

Note: LCMSMS Liquid Chromatography tandem Mass Spectrometry. Arch 2G Architect 2nd generation (Abbott);
Arch 1 G: Architect 1st generation (Abbott); RIA: Radioimmunoassay (Immunotech); Acc: Access; Cent: Centauro;
Imm:Immulite; Cobas (Roche)

TABLE 4. Bland Altman ans Spearman correlation coefficient of the values ot Testosterone by different mthods compared to LCMSMS from hirsute women

	Arch 2 G lab 1	Arch 2 G lab 2	Arch 1 G lab 3	RIA	Acc	Cobas lab 1	Cobas lab 2	Cobas lab 3	Cobas lab 4	Cent lab 1	Cent lab 2	Imm lab 1	Im lab 2
Bland-Altman analysis													
Mean difference	0.019	0.124	-0.039	0.058	-0.078	0.179	-0.025	-0.072	-0.054	0.163	0.197	-0.164	-0.250
SD	0.083	0.055	0.054	0.173	0.176	0.683	0.104	0.098	0.105	0.062	0.830	0.115	0.137
Values outside the agreement limits mean ±1.96SD	0	1 (8.3%)	0	0	0	1 (9.1%)	0	0	0	0	1 (8.3%)	1 (8.3%)	0
Spearman coefficient	0.851	0.939	0.928	0.715	0.666	0.825	0.867	0.858	0.865	0.931	0.954	0.933	0.669
sig bilateral – p (p Spearman)	0.000	0.000	0.000	0.009	0.018	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.024

Note: LCMSMS Liquid Chromatography tandem Mass Spectrometry. Arch 2G: Architect 2nd generation (Abbott);
RIA: Radioimmunoassay (Immunotech); Acc: Access; Cent: Centauro; Imm: Immulite; Cobas (Roche)

The gold standard method for measuring free T is equilibrium dialysis, a procedure that is not routinely feasible in all clinical biochemistry labs. Alternatively, and with similar results to those of dialysis, FT can be calculated by using an equation based on the law of mass action[1] for which T value is one of the analytes. A wide range of papers describe how to calculate the distribution of circulating serum T[16-18]; there are also numerous links on the Internet to testosterone calculators, all of which employ T concentration as a constitutive element in the various formulas. Therefore, a proper calculation requires that serum T be measured by a validated, acceptably accurate technique using duly approved methods.

TABLE 5. Results of the regression equation weighted Deming. Lc MSMS method taking as the independent variable (x) and each method (y) as the dependent variable

| Normal women | | Hirsute women | |
Method	SD Intercept	Method	SD Intercept
Architec 2nd generation	0.069	Architec 2nd generation lab 1	0.014
Architec 1st generation	6.823	Architec 2nd generation lab 2	0.031
RIA	0.204	Architec 2nd generation lab3	0.011
Access	0.129	RIA	0.083
Cobas lab 1	0.203	Immulite lab 1	0.096
Cobas lab 2	0.101	Immulite lab 2	0.027
Cobas lab 3	0.208	Cobas lab 1	-1.310
Cobas lab 4	0.214	Cobas lab 2	0.052
Centauro lab 1	0.271	Cobas lab 3	0.255
Centauro lab 2	0.656	Cobas lab 4	0.033
Immulite	0.138	Centauro lab 1	0.036
		Centauro lab 2	0.004
		Access	0.172

TABLE 6. Bias of testosterone immuno assay by different methods respect to the LCMSMS from hirsute women. Criterion of satisfactory performance was ±6.4 % mean bias to the LC MSMS

Méthods	Arch 2nd Gen	ArchArch 2nd Gen	2nd Gen	RIA	IMM	IMM	Cobas	Cobas	Cobas	Cobas	Cent	Cent	Access
	lab 1	lab 2	lab 3		lab 1	lab 2	lab 1	ab 2	lab 3	lab 4	lab 1	lab 2	
%	2.3	14.7	4.6	6.8	-19.5	-31.5	18.9	-3.0	-8.5	-6.4	19.3	23.3	-9.2

Note: LCMSMS Liquid Chromatography tandem Mass Spectrometry. Arch 2nd Gen: Architect 2nd generation (Abbott); RIA: Radioimmunoassay (Immunotech); Acc: Access; Cent: Centauro; Imm: Immulite; Cobas (Roche)

As regards the commercial kits Abbott 1st and 2nd generation Architect, there are significant disparities. The former, a method not validated by LC-MS/MS, reveals major differences compared to the latter, 2nd generation Architect, and is substantially different from LC-MS/MS. The adjustments made in developing the new version, whose insert states that its results are validated by LC-MS/MS performed with Quest Diagnostics, afford better results. The samples in the present study were referred to Quest Diagnostics for validation. Due to the incongruent measurements obtained by the 1st generation kit, and to the fact that Abbott discontinued use of this version, the assays thus performed have been ruled out of the discussion of the measurements obtained.

The statistical analysis showed higher concordance with LC-MS/MS in relation to our earlier work. In the NW group, the Wilcoxon signed-rank test showed that only the results yielded by Centaur were statistically different from those obtained by LC-MS/MS, while in the HW group there were divergences for Immulite and Centaur techniques. According to Bland-Altman analysis, Centaur, Immulite and Cobas presented more than 5 % of measurements outside the agreement limits in the HW group. Regression analysis of the combined data for both groups showed significant correlation coefficients for all the methods compared to LC-MS/MS, with 95 % confidence interval of the slopes including 1, except for Immulite, which implies that the values match those obtained by

LC-MS/MS. Still, the Spearman correlation coefficient revealed that none of the methods used matched LC-MS/MS in the NW group, while all of them matched LC-MS/MS within the HW group.

The centrality of dependable serum T determination relates to the fact that T level is a good marker in studying a range of medical conditions. Numerous papers have reported hirsutism to be the clinical manifestation of various disorders. Azziz et al[19] studied 1000 hirsute female patients and described the following pathological conditions and their associated frequencies: androgen-secreting neoplasm: 0.23 %; classical adrenal hyperplasia: 0.69 %; nonclassical adrenal hyperplasia: 2.06 %; hyperandrogenism-insulin-resistance and acantosis nigricans (HAIR-AN): 3.78 %; Polycystic Ovarian Syndrome (PCOS): 82.02 %; hyperandrogenism-hirsutism/ ovulatory dysfunction: 6.75 %; idiopathic hirsutism: 4.47 %. PCOS is clearly the highest incidence condition among female patients with hirsutism and one for which, as for idiopathic hirsutism, defining biochemical and clinical hyperandrogenism, high serum DHEA-S and/or T, Tbio, or free T and increased Ferriman-Gallwey score respectively, take on crucial diagnostic value. Even though androgen assessment proves important for the former conditions too, clinical features other than hirsutism, along with further lab tests and imaging, are relevant tools in accurately diagnosing them.

Idiopathic hirsutism has been reported for women with clinical hirsutism and normal androgen levels, with no menstrual dysfunction and with progesterone levels in the luteal phase compatible with ovulatory cycles[20]. Obtaining accurate T/ free T and DHEA-S measurements is then critical for diagnosis. Non-validated techniques resulting in wrongly increased values (false positives) would turn diagnosis of idiopathic hirsutism into hyperandrogenism-hirsutism/ ovulatory dysfunction, which would ultimately involve a different prognosis and treatment.

Given that hair growth and characteristics are androgen-dependent, normal-androgen-level hirsutism has not yet been conclusively established. Based on experimental studies, various theories have been put forward, including:

1. The hair follicle produces several factors, among others: epidermal growth factor (EGF); platelet derived growth factor (PDGF); nerve growth factor (NGF); bone morphologic protein (BMP); vascular endothelial growth factor (VEGF); fibroblast growth factor (FGF); tissue growth factor-α (TGF-α) and insulin-like growth factor type 1 (IGF-1). The latter increases the expression of 5-alpha reductase with normal androgen levels[21]. Higher in-situ levels of IGF-1 could possibly increase androgenic activity in the hair follicle.

2. It has been posited that with normal androgen levels, interleukin 6 (IL 6) and tissue growth factor beta (TGF-beta) mediated by their membrane receptors, could induce androgen-receptor phosphorylation and thus trigger an androgen response[22].

3. It has been experimentally demonstrated that while peptide-receptor antagonists associated to the parathormone (R-PTHrP) cause hair loss, antagonists produce a counter effect[23].

4. It has been demonstrated that contraction in the number of androgen receptor exon 1 CAG repeat polymorphisms induces androgen action with normal circulating androgen levels due to a mechanism yet to be fully elucidated[24].

As for PCOS, the Rotterdam expert conference held in 2003 arrived at a consensus on the diagnostic criteria to be followed. The conclusions drawn were published in 2004[25].

The patient should meet at least two of the following three criteria:
1. Oligo- or anovulation
2. Clinical and/or biochemical hyperandrogenism
3. Polycystic ovaries

Criteria 1 and 2 or 2 and 3 combined could happen without hirsutism and normal DHEA-S levels in some patients. In such cases, T testing proves crucial. The presence of false negatives in such a determination would lead to PCOS underdiagnosis, and therefore to the possible ensuing metabolic consequences for the patients involved.

The Androgen Excess Society (AES) position statement regards PCOS as an ovarian dysfunction-associated hyperandrogenism syndrome[26], the primary condition being hyperandrogenism, both clinical (hirsutism, acne, alopecia) and/or biochemical (hyperandrogenemia), including one or both of the following features:
1. Ovarian dysfunction: a) oligo-anovulation: cycles lasting under 27 days or over 34 days; b) oligoovulation: cycles lasting between 27 and 34 days with P4 levels < 4 ng/ml.
2. Polycystic ovary imaging (12 or more follicles 2 to 9 mm in diameter in one or both ovaries and/ or increased ovary volume (> 10ml).

AES guidelines again demonstrate the relevance of accurately determining androgen levels in order to characterize possible PCOS phenotypes, particularly among patients whose condition does not involve hirsutism.

Both for idiopathic hirsutism and for PCOS, associated diseases that could present excess androgens should be ruled out. These include: thyroid dysfunction, hyperprolactinemia, nonclassical adrenal hyperplasia (CYP21), Cushing Syndrome, androgen-secreting neoplasms, hyperandrogenism, insulin-resistance and acantosis nigricans (HAIR-AN).

From the qualitative point of view, the current study yielded significant differences compared to our earlier work2. Only four of the methods resulted in false positives, that is to say, increased T values in the two hirsute patients with normal serum T levels by LC-MS/MS. No false positives were found in the NW group except for Centaur methodology.

However, in quantitative terms, and particularly in the NW group, the methods employed –all of them currently used locally- showed substantial differences in resulting values compared against those of LC-MS/MS.

In order to perform measurements that could be easily performed with the commercially available kits either by manual or automated procedures, Tiel Groenetege and colleagues[27] assayed measurements from the serum extracted in ether and resuspended in the matrix for various techniques compared to LC-MS/MS and found a fair amount of matching to the gold standard method.

In this study we performed a duplicated ether extract sample for each aliquot in the NW group. One of the aliquots was resuspended in a phosphate buffer solution. Each participating lab was sent the buffer-suspended sample and the duplicate dry extract to be later resuspended in the assay matrices. None of the reported measurements altered the differences recorded by LC-MS/MSwith either of the two protocols, except for 2nd generation Architect, for which the ether extract was resuspended in the solvent supplied by Abbott for T testing (these measurements are not shown in this paper)

At present, researchers insist on the need to employ LC-MS/MS-validated methods so as to accurately assess circulating T levels within groups with serum T levels close to the quantification limits (lower threshold), such as women, children hypogonadal men and aged men.

The issue has been addressed in a number of other studies reported in the literature, which are outlined as follows:
1) T results from two labs using two LC-MS/MS-validated methods were compared to a RIA[28];
2) Thirteen steroids were tested simultaneously by LC-MS/MS and RIA in women with PCOS[29];
3) Serum T and estradiol measurements were compared in aged men by using an in-house method of gas chromatography tandem mass spectrometry and chemoluminescence[30];
4) Pediatric and adult reference intervals were defined by LC-MS/MS and by comparison with immunological methods[31] as well as in hipogonadal men with hypogonadotrophic hypogonadism and panhypopituitarism[32].

In line with the conclusions presented in this paper, the findings reported in the studies above generally seem to demonstrate that some of the methods currently used would yield clinically satisfactory results.

From a statistical point of view, though, establishing the most efficient statistical methodology to determine acceptability of a given technology turns out to be highly complex matter. Various methods employed in this study (non-parametric paired-sample Wilcoxon, Pearson's correlation coefficient, Bland-Altman and Weighted Deming regression) failed to yield similar differences by reference to LC-MS/MS for certain methods. This lack of consistency on defining acceptability for some methodologies prevents us from drawing any definitive conclusions as regards which statistical methods should be employed so as to be able to validate serum T measurement in hirsute women.

The Centers for Disease Control and Prevention's Hormone Standardization Project (CDC-HoSt Program) set out to standardize serum T testing with a view to improving on diagnosis and preventing T- associated disorders. (For further information contact them at standardization@cdc.gov or visit their official website http://www.cdc.gov/labstandards).

It would be most desirable for every lab to comply with the recommendations and requirements laid down by CDC allowing routinely used methodology to be approved, and thus be able to certify the procedure under review if it meets the ± 6.4 mean bias performance criterion compared to the method of reference. Now, since this condi-

tion is not easy to fulfill, we have had to adopt the CDC criterion and tried to validate the measurements obtained by the various methods in each lab respectively on the basis of the measurements obtained from twin serum samples from hirsute women tested by LC-MS/MS at Quest Diagnostics and validated at CDC in May 2012 and revalidated in 2013 and 2014.

Table 6 shows the mean bias percentages for each lab by reference to LxC-MS/MS. As can be seen, two of the techniques reviewed, 2nd generation Architect and Cobas, satisfactorily passed the validation test. Still, from the three labs using 2nd generation Architect, one failed the meet the validation requirement, as did another two out of the four employing Cobas.

The foregoing demonstrates the variability of inter-lab results, which could possibly be due to divergences between the different commercial kit batches used by the various labs, to method calibration, or to instrument-inherent factors. This is particularly relevant, given that upon certification of the method CDC claims that it is both the supplier's and the participant's responsibility to ensure that the results remain consistent during the one-year period when validity holds.

In conclusion, the results obtained demonstrate that serum T measurements performed with the latest kits in normal women or hirsute women differed substantially from the results obtained locally in an earlier study with methods manufactured by the same suppliers.

The most significant improvements, regardless of statistical analyses, were as follows:
1) All serum T measurements by LC-MS/MS in the NW group fell within the normal range in all cases. Within this group, all the values reported by every lab were also normal, that is to say, there were no false positives, which cancelled the possibility of overdiagnosing clinical hyperandrogenism.
2) As regards the measurements by LC-MS/MS in hirsute women, whenever increased T values were obtained, none of the cases reported by the participating labs included normal values, that is to say, no false negative were reported, which cancelled underdiagnosis of hyperandrogenism.
3) For the two hirsute women with normal serum T levels (by LC-MS/MS), only four of the methods yielded values with higher T levels, in other words, false positives, which induced overdiagnosis /misdiagnosis of hyperandrogenism.

Our next study will be aimed at a larger number of cases in both groups. In the HW group, it will set out to assess T incidence in the various conditions and disorders for which hirsutism is a clinical manifestation as well as T significance in characterizing some of the disease phenotypes, particularly PCOS. Over and above, we expect to continue to be engaged in a number of private and public educational pursuits and support activities, working alongside national and international agencies, with a view to insisting that pharmaceutical companies keep on improving their commercial test kits, as some commercial labs have already done. In this way, clinical biochemistry labs will be able to enhance the reliability of their reports on serum T values.

REFERENCES

1. **Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H**. Position Statement: Utility, limitations and pitfalls in measuring testosterone; An Endocrine Society position statement. J Clin Endocrinol Metab 92:405-413, 2007

2. **Scaglia HE, Aquilano DR, Buccini G, Chichizola C, Corazza N, Corthein C, Fuillerat S, Guevel L, Ibañez G, Lacoste E, Mohr A, Nosetto S, Orsi M, Piaggio R, Riesco O, Sandoz S, Scaglia J, Viola AM, Wolfthal D, Zylbersztein CC.** Quantification of testosterone (T) by 8 immunoassay and by liquid chromatography-tandem mas spectrometry (LC- MS/MS) in normal and hirsute women. A multicentric study. Rev Argent Endocrinol Metab, 50:3-18; 2013

3. **Boots L R, Potter S, Potter D, Azziz R** Measurement of total serum testosterone levels using commercially available kits: high degree of between- kit variability. Fert Steril 69:286-292, 1998

4. **Taieb J, Mathian B, Millot F, Patricot M-C, Mathieb E, Queyrel N, Lacroix I, Somma-Delpero C, Boudou P.** Testosterone measurement by 10 immunoassay and by isotope-dilution gas chromatography-mass spectrometry in sera from 116 men, women and children. Clin Chem 49:1381-1395, 2003.

5. **Fitzgerald RL, Herold DA.** Serum total testosterone, immunoassay comnpared with negative chemical ionization gas chromatography. Clin Chem 42:749-755, 1996

6. **Grant RP, Cameron C, Mackensie-McMurter S.** Generic serial and parallel on-line direct-injection using turbulent flow liquid chromatography/tandem mass spectrometry. Rapid Commun Mass Spetrom 16:1785-1792, 2002

7. **Vesper HW, Botelho JC.** Standardization of testosterone measurements in humans. J Steroid Bioch Mol Biology 121:513-519, 2010

8. **Vesper HW, Thienpont LM.** Traceability in laboratory medicine. Clin Chem 55:1067-1075, 2009

9. **Botelho JC, Shacklady C, Razdan R Wesper HW.** Impact of sample preparation procedures of testosterone measurement by isotopic-dilution liquid chromatography-tandem mass spectrometry. Clin Chem 54 Abstract 122, 2008

10. **Thienpont LM, Van Uytfanghe AL. Blincko S, Ramsay CS,Xie H, Doss RC, Keevil BG, Owen LJ, Rockwood AI, Kushnir MM, Chun KY, Chandler HP, Field PM, Sluss PM.** State-of-the-art of serum testosterone measurement by isotopic-dilution liquid chromatography-tandem mass spectrometry. Clin Chem 54:1290-1297, 2008

11. **Vesper HW, Bhasin S, Wang C, Tai SS, Dodge LA, Sinh RJ, Nelson J, Ohorodnik S, Clarke NJ, Parker Jr CR, Razdan R, Monsell E A, Myers GL.** Interlaboratory comparison study of serum total testosterone measurements performed by mass spectrometry methods. Steroids 74:498-503, 2009

12. **Ferriman D, Gallwey J.D**. Clinical assessment of body hair growth in women. J Clin Endocrinol Metab 21:1440-1447, 1961

13. **Bland J. M, Altman DG.** Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307-310, 1986

14. **Linnet K.** Necessary sample size for method comparison studies based on regression analysis. Clin Chem. 45:882 - 894, 1999

15. **Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Futterweit W.** Position statement: criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: An Androgen Excess Society guideline. J Clin Endocrinol Metab 91:4237-4245, 2006

16. **Dunn JK, Nisula BC y Roadbard D.** Transport of steroids hormones. Binding of 21 endogenous steroids to both testosterone-binding globulin and corticosteroid- binding globulin in human plasma. J Clin Endocrinol Metab 53:58-68, 1981

17. **Sodergárd R, Backstroom T, Schnbhan V, Cartensen H.** Calculation of free and bound fractions of testosterone and estradiol 17 beta to human plasma protein at body temperature. J Steroid Biochem 16:801-810, 1982

18. **Riesco, OF; Aquilano, DR y Scaglia, HE**. Descripción de un método simple para la determinación de Testosterona libre y biodisponible. Evaluación crítica y aplicaciones en la clínica endocrinológica. Acta Bioquímica Clínica Latinoamericana vol. XXX-VI, No 1, 5-26,2002

19. **Azziz R, Sanchez LA, Knochenhauer ES, Moran C,Lazenby J, Stephens K, Taylor K, Boots LR**. Androgen Excess in Women: Experience with over 1000 consecutive patients. J Clin Endocrinol Metab 89:453-462, 2004

20. **Azzis, R, Carmina E, Sawaya E.** Idiopathic Hirsutism. Endocrine Reviews 21(4):347-362, 2000.

21. **Thiboutot DM,** Normal physiology of the pilosebaceus unit. in Azzizz R, Nestler JE,and Dewailly D. Androgen excess disorders in women (eds). Lippincott – Raven. New York. 1997, pag 103-115

22. **Heinlein CA, Chang C**. Androgen receptor (AR) Coregulator: An overview. Endocrine Reviews 23:175-201, 2002 (Publicado on line en 2013)

23. **Holick MF.** PTHrp in alopecia. Endocrine Society - 83rd annual meeting Abst: S29.

24. **Scaglia, J.** Receptor de los andrógenos, gen que lo codifica. Mecanismo de acción y patologías asociadas. Revista Argentina de Andrología 17:34-41, 2008

25. **Chang J; R Azziz R, Legro R, Dewailly D, Franks S, Fausen B, Balen B.** (Scientific Committee) Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. The Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Scientific Committee: Fert Steril 81:19-25, 2004

26. **Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Futterweit W, Janssen OE, Legro RS, Norman RJ, Taylor AE, Witchel SF.** The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report. Fertil Steril: 91:456-458, 2009.

27. **Tiel Groenestege, WM, Bui HN, ten Kate J , Menheere PPCA, Oosterhuis WP, VaderHL, Heijboer AC,Marcel J.W. Janssen MJW**. Accuracy of First and Second Generation Testosterone Assays and Improvement through Sample Extraction. *Clinical Chemistry* 2012; v. 58,p.1154-1156 (to the editor) Published April 26, 2012

28. **Legro RS, Schlaff WD, Diamond MP, Coutifaris C, Casson PR, Brzyski RG, Christman GM, Trussell JC, Krawetz SA, Snyder PJ, Ohl D, Carson SA, Steinkampf MP, Carr BR, McGovem PG, Cataldo NA, Gosman GG, Nestler JE, Myrers ER, Santoro, Eisenberg E, Zhang M, Zhang H.** for Reproductive Medicina Network. Total testosterone assay in women with Polycystic Ovary Syndrome. Precision and correlation with hirsutism. J Clin Endocrinol Metab 95:5305-5313, 2012. Published on line 2/jul 2013

29. **Keeffe CC. Goldman MM, Zhang K, Clarke N, Reitz RE, Welt CK.** Simultaneous measurement of thirteen steroid hormones in women with Polycystic Ovary Syndrome and Control women using liquid Chromatography-tandem mass spectrometry. PloSOne 8:9 (4) e 93805, 2014. Published on line doi:10.1371/journal.pone 0093805 e collection, 2014

30. **Huhtaniemi IT, Tajar A, Lee DM, O`Neill TW, Finn JD, Bartfal G, Boonen S, Casanueva FF, Giwercman A, Han TS, Kula K, Labrie F, Lean MEJ, Punab M, Silman AJ, Vanderschueren D, Forti G, Wu FCW,** EMAS group. Comparison of serum testosterone and estradiol measurements in 3174 European men using platform immunoassay and mass spectrometry; relevance for the diagnostics in aging men. European Journal of Endocrinology 166:983-991, 2012

31. **Kushnir MM, Blamires T, Rockwood AL, Ronerts WL, Yue B, Erdogan E, Bunker AM, Wayne Meikle A.** Chromatography-tandem mass spectrometry assay for androstenedione, dehydroepiandrosterone and testosterone with pediatric and adult reference intervals. Clin Chem 56:1138-1147, 2010

32. **Giton F, Trabado S, Maione L, Sarfati J, Le Bouv Y, Brailly-Tabard S, Fiet J, Young J.** Sex steroids, precursors and metabolites deficiencies in men with isolated hypogonadotipic hypogonadism and panhypopituitarism: a GCMS-based comparative study. J Clin Endocrinol Metab. 2015 Feb;100(2):E292-6. doi: 10.1210/jc.2014-2658. Epub 2014 Nov 13.