Revista Argentina de Endocrinología y Metabolismo
Copyright ® 2012 por la Sociedad Argentina de Endocrinología y Metabolismo
Vol 50 • Nº 1

TRABAJO ORIGINAL

Quantification of Testosterone (T) by 8 Immunoassays and by Liquid Chromatography -Tandem Mass Spectrometry (LC-MSMS) in Normal and Hirsute Women. A Multicenter Study

Cuantificación de testosterona (T) por 8 inmunoensayos y por cromatografía líquida en tandem con espectrometría de masa (LC MS/MS). En mujeres normales e hirsutas. Estudio multicéntrico

Scaglia HE, Aquilano DR, Buccini G, Chichizola C, Corazza N, Corthey C, Fuillerat S, Guevel L, Ibáñez G, Lacoste E, Mohr A, Nosetto S, Orsi M, Piaggio R, Riesco O, Sandoz S, Scaglia J, Viola AM, Wolfthal D, Zylbersztein C

Hormone Determination Laboratory, Hospital Italiano, La Plata, Argentina. The different participants in this study performed their determinations in their private laboratories

ABSTRACT

Objective: To compare T results in normal and hirsute women, obtained by different laboratories employing the same or different methods, including an in-home RIA, and the gold standard method LC-MS/MS. In addition, T results were referred to a curve obtained by 6 different pools that had been prepared on the basis of LC-MS/MS results.
Design: Prospective study
Setting: Hormone Determination Laboratory, Hospital Italiano, La Plata, and private practice of each participant laboratory.
Patient(s): Blood samples were obtained from 78 individuals sorted into 3 groups, namely, normal men (n:39), normal women (n:24) and hirsute women (n:15)
Interventions(s): None
Main Outcome Measure(s): To evaluate if the results obtained in each lab for each serum sample by the methods currently employed in our country are significantly different from those obtained by LC-MS/MS (Gold standard)
Result(s) One out of the 24 NW showed high T values by LC - MS/MS. In each lab, except in 1 (Architect) T results of this serum sample were normal. Two out of the 15 hirsute patients showed normal T values (LC - MS/MS). Method and number of labs -shown between brackets- and percentages of normal T results (false negatives) are described for each method as follows: Chemiluminescence: Axsym - Abbott (Axn) - (3) 85, Architect - Abbott - (Arch); (2) 70; Immulite - Siemmens - (IMM); (2) 42; Electrochemiluminescence - Elecsys - Roche- ((EQL); (4) 52; Fluorescent enzymatic - Vidas - Bio-Merieux - (Vidas) (1) 69; Manual coated tube radioimmunoassay (RIA): RIA - Siemmens Coat-a-Count (RIA S); (3) 64; RIA - DSL Inc (RIA DSL); (1) 31; RIA - DIASource - (DiaS); (1) 31; and in-Home RIA (in-H) (1) 12.
Statistically significant differences were obtained between different methods and against LC MS/MS. In-H method is the one that comes closest to 1 on the Weighted Deming regression and closest to zero on the SD intercept, (standard deviation of the constant in the straight line equation) indicating that the values match those obtained by LC - MS/MS. The values recorded by the various methods employed showed no significant modifications when plotted against a secondary standard curve.

Por error en la impresión de este trabajo en el Vol. 49 N° 4 de 2012, se vuelve a publicar el trabajo correcto.

Recibido: 18-04-2012 Aceptado: 04-09-2012

Correspondencia: Hugo E. Scaglia, Calle 59 N° 969 (1900) La Plata, Argentina
Telefax: 54 221 452 9577
info@iabe.com.ar

Conclusion(s) This indicates that the techniques in current use in our area underestimate hyperandrogenemia in these patients. Discrepancies are not due to the various calibration curves proposed in the corresponding commercial kits. The fact that the In-H technique affords finer results while employing a larger serum volume suggests that the disparities among the various commercial methods result from their limited sensitivity to the sample volumes they process. **Rev Argent Endocrinol Metab 50:3-18, 2013**

No financial conflicts of interest exist.

Key words: testosterone assay, different methods for T assay, T levels in normal women, T levels in hirsute women

RESUMEN

El diagnóstico de hiperandrogenemia requiere la demostración de niveles aumentados de Testosterona Total (TT) en suero. Los inmuno ensayos comerciales dan resultados divergentes a niveles bajos de TT como los obtenidos en mujeres. Valoramos los niveles de TT en 24 mujeres normales (MN) y 15 hirsutas (MH) en 18 laboratorios por métodos comúnmente empleados en nuestro medio. Los métodos y número de laboratorio que emplearon cada método se muestra entre paréntesis así como los porcentajes de T normal (falsos negativos) obtenidos en cada método fueron: Quimioluminiscencia: Axsym - Abbott (Axn)- (3) (85), Architect - Abbott - (Arch); (2) 70; Immulite - Siemmens - (IMM); (2) 42; Electroquimioluminiscencia - Elecsys - Roche- ((EQL); (4) 52; Enzimático acoplado a fluorescencia Vidas - Bio-Merieux - (Vidas) (1) 69, Radioinmunoiensayo en tubo recubierto (RIA): RIA - Siemmens (RIA S); (3) 64; RIA - DSL Inc (RIA DSL); (1) 31; RIA - DIASource - (DiaS); (1) 31; y un método desarrollado en uno de los laboratorios (in-H) (1) 12. Comparativamente a LC MS/MS los niveles fueron en todas las muestras significativamente más bajos por Axn y en 18 de las 24 MN por DiaS. En 7 casos; 3 por RIA S, 2 por IMM y 1 por EQL y Arch los valores de TT fueron superiores al límite superior de sus respectivos métodos. En todos los casos se obtuvo una gran variación entre los mismos y con diferentes métodos. Trece de las 15 MH tuvieron niveles altos de TT por LC MS/MS. De las MH con TT aumentada de acuerdo a la determinación por LC MS/MS entre el 12 y el 85 % de las mismas por los distintos métodos fueron normales, indicando que en la mayoría de los métodos habitualmente utilizados en nuestro medio subvaloran la hipernadrogenemia en estas pacientes. Estas diferencias se hacen más notorias a niveles más bajos de TT (Se obtuvieron valores normales en el 71 % de los casos con valores de TT entre 0,47 y 0,74 ng/ml y en el 38 % de los casos, con niveles de TT mayor a 0,98 ng/ml). En 9 muestras se determinó la TT empleando una curva en el rango de 0,21 a 6,44 ng/ml preparada con de una mezcla de 78 sueros cuyos valores fueron obtenidos por LC MS/MS. No se obtuvo una modificación significativa de los valores indicando que la diferencia entre los distintos métodos no es debida a las diferentes curvas de calibración de los kits comerciales. En conclusión ninguno de los métodos mayormente empleados en nuestro medio son aceptables para la evaluación de niveles menores a 1,5 ng/ml. **Rev Argent Endocrinol Metab 50:3-18, 2013**

Los autores declaran no poseer conflictos de interés.

Palabras clave: valoración de testosterona, diferentes métodos por valoración de testosterona, niveles de T en mujeres normales, niveles de T en mujeres hirsutas

INTRODUCTION

Determination of Total Testosterone (TT) or its free fraction (FT) or Bioavailable Testosterone (BAT) has proven a useful tool in the diagnosis of androgenic alterations. In males TT determination together with gonadotropins reveals information about testicular dysfunction[1,2]. In patients with prostatic cancer metastasis the treatment of choice involves gonadotropin-releasing hormone analogues or antiandrogen therapy[3]. Monitoring androgen levels is very important in these cases. In females some of the main clinical manifestations of hyperandrogenism include hirsutism, acne and/or loss of scalp hair. TT and FT determination proves essential in hyperandrogenism and, along with ovary ultrasound imaging and evaluation of menstrual cycle-related alterations, it permits patient streaming into idiopathic hirsutism[4], androgen-producing neoplasm, (late-onset) congenital adrenal hyperplasia[5] or polycystic ovary syndrome (PCOS)[6-8] groups.

The earliest TT immunoassays employed tritiated T3H as a tracer, extracted the steroid from the serum with solvents and purified the extract so as to eliminate proteins and molecules with related chemical structures which interfere with the reaction and thus improve result sensitivity and precision[9]. These assays were replaced by others using iodinated TT as tracer and, more recently, by non-radioactive tracers, like fluorescent, electrochemiluminescent, chemiluminescent and

colorimetric tracers. Such methods have become automated and their use widespread in clinical laboratories.

For all these reasons these immunoassays for TT determination are easy to use, relatively cheap and, being automated, they allow simultaneous processing of a large number of samples. For TT levels in males, methods for TT determination are, generally speaking, adequately sensitive; yet they are relatively inaccurate. In hypogonadic men, prepubertal boys or girls, or normal (NW) or hirsute women (HW), the methods fail to present enough sensitivity for the TT levels expected in these cases, hence their modest clinical usefulness.

On account of these difficulties the Council of the Endocrine Society appointed a task force to review the problem and be able to make recommendations about it on the basis of the conclusions drawn. Assessment work involved reviewing the literature, gathering data, evaluating the College of American Pathologists survey data and reaching consensus towards the final discussion and drafting of the outcome document[9].

The main conclusions arrived at point out that:

- direct assay by RIA, ELISA or chemiluminescence for the methods available for measuring TT are technically simple, quick and inexpensive and can be automated. However, TT concentration is often under or overestimated, susceptible to matrix effects, not standardized and only relatively accurate, with values at TT 3.0 ng/ml.
- RIA following extraction and chromatography permits the use of relatively large volumes for assay increased sensitivity. Yet, the procedure is susceptible to matrix effects, proves labor-intensive, demands considerable processing time and requires a high level of technical expertise.

A summary of the key findings and recommendations is as follows:

- In the absence of other information, direct assays (those performed on whole serum) perform poorly at low TT concentrations (i.e. in women, children and hypogonadic men) and should be avoided.
- Assays after extraction and chromatography, followed by either mass spectrometry (MS) or immunoassay, are likely to furnish more reliable results and are currently preferred.

The task force recommends that this methodology be adopted in as much as it improves the accuracy and precision of androgen evaluation; it also considers that the latter criteria should prevail over others of simplicity and economy in TT determination[9].

OBJECTIVES

1. Even though sound, thorough quality control check-ups are done in our country by comparing the different TT evaluation methods, these do not match the sample values obtained by liquid or gas chromatography -tandem mass spectrometry (LC-MS/MS), considered the gold standard. Thus, drawing on these recommendations, this study set out to compare the results of TT obtained by different laboratories employing the same or different methods, including an in-Home (in-H) with prior extraction and chromatographic purification, and the gold standard method LC-MS/MS.
2. In addition, the TT results were referred to a curve obtained by 6 different pools that had been prepared on the basis of LC-MS/MS results. This was performed with a view to correcting discrepancies between the labs.

MATERIALS AND METHODS

The multicenter study was carried out in 18 laboratories in which blood samples corresponding to various clinical situations 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 TT determination. HW showed an increased score according to Ferriman – Gallwey[10] criteria, with or without acne and/or loss of scalp hair, and with or without menstrual cycle alterations. NW did not present hirsutism, had regular menstrual cycles and progesterone levels after ovulation measured in the luteal phase in some of them.

Blood samples were obtained from 39 individuals sorted into 2 groups, namely, NW (n:24) and HW (n:15). The blood samples were collected in tubes containing no anti-coagulants or preservatives and the serum was obtained by centrifugation. For all samples, 20 aliquots of at least 0.5 ml were obtained and kept at -20 °C until they were sent to each of the labs participating in the study. Eighteen aliquots of each serum were sent

for analysis to each lab intervening this study. Out of each aliquot from each serum sample of the eighteen sent, one part was analyzed by the methods in current use in each participating lab, and the remaining amount was stored at -20 °C to be re-quantified. A 2 ml aliquot was kept until all the samples had been collected, so as to have them all sent on together for analysis by LC-MS/MS; another 2 ml aliquot was set aside at -20 °C for performing various mixtures using 78 sera obtained from 39 normal men, and the serum samples from NW and HW which were within the concentration range 0.21-6.44 ng/ml on the basis of the values yielded by LC-MS/MS. From these serum samples six serum pools were obtained and used as secondary calibrators for each method used for TT determination. The mixture was performed with values in ng/ml greater than 6; between 4 and 6; between 2 and 4; between 0.7 and 2; between 0.3 and 0.7; and less than 0.3. The mean ± SD for each mixture respectively (ng/ml) was 6.44 ± 0.68; 4.70 ± 0.63; 2.94 ± 0.50; 1.07 ± 0.37; 0.44 ± 0.13 y 0.21 ± 0.06.

Table 1 shows the individual values for each serum and those for the mixtures performed according to the concentration limits defined in each table, as well as the mean and SD of each mixture performed.

TT concentrations for each of the 78 sera were measured by the methods currently in use in each of the 18 laboratories participating in the study, according to the relevant package inserts, and later repeated by the same method but with reference to the secondary standard prepared according to the description above:

The techniques employed are detailed as follows:

5 automated immunoassays:

MEIA, Microparticle Enzyme Immuno Assay - Axsym - Abbott (Axn)- 3 labs

Chemiluminescence - Architect - Abbott - (Arch); 2 labs

Chemiluminescence - Immulite - Siemens - (IMM); 2 labs

TABLE 1. Calibration curve the serum mixtures prepared according to the values obtained by LC-MS/MS (secondary calibrators)

	POOL F	POOL E		POOL D	POOL C		POOL B	POOL A
	Higher	Between	ml used	Between	Between	ml used	Between	Lowest
	6.0 ng/ml	4.0-6.0 ng/ml	in pool	2.0-4.0 ng/ml	0.7-2.0 ng/ml	in pool	0.3-0.7 ng/ml	0.3 ng/ml
	6.8	4.1		3.7	0.72	4.0	0.40	0.23
	6.0	4.1	6.0	3.1	0.71	5.0	0.60	0.30
	6.6	4.4	7.0	3.7	0.99	3.5	0.56	0.26
	6.0	4.0		2.2	0.74	2.5	0.35	0.28
	5.4	5.7	7.0	2.7	0.93	1.5	0.39	0.16
	7.4	5.0	5.0	2.5	1.50	2.0	0.41	0.23
	6.9	5.2		2.7	1.40	2.0	0.50	0.19
		5.1		3.0	1.60	4.0	0.69	0.16
				2.2			0.35	0.13
				3.0			0.29	0.13
				3.7			0.32	0.26
				2.8			0.31	0.30
				2.7			0.34	0.25
				3.0			0.35	0.18
				2.6			0.41	0.16
				2.8			0.43	0.19
				2.7			0.43	
				3.9			0.53	
							0.54	
							0.68	
							0.47	
Media	6.44	4.70		2.94	1.07		0.44	0.21
SD	0.68	0.63		0.50	0.37		0.13	0.06

Note: Mixtures A, B, D and F were performed with equal volumes of sera. The C and E were used for mixing with the volumes specified The concentration was calculated as the average of the each mixture

Electrochemiluminescence - Elecsys - Roche- (EQL); 4 labs

Fluorescent enzymatic - Vidas - Bio-Merieux - (Vidas) 1 lab

3 manual coated tube radioimmunoassay (RIA)

RIA - Siemens (RIA S); 3 labs

RIA - DSL Inc (RIA DSL); 1 lab

RIA - DIASource – (DiaS); 1 lab,

and one in-H developed by one of the participating laboratories[11], the description of which is as follows:

8000 dpm (T3H) is added to the 0.5 ml serum and 1ml serum samples for assays on males and on women respectively in order to assess recovery and TT is extracted with 2 ml ether-hexane (4:1). The anti-T Polyclonal Rabbit Antibody was obtained by injecting T-6-oxime: BSA; T3H ~30000 dpm/0.1 ml was used as tracer; the standard curve was prepared by successive dilutions of a serum mixture whose TT concentration was measured by LC-MS/MS, with values from 0.15 to 9.6 ng/ml. Dextran-coated charcoal was used to separate the antibody-bound from the free hormone.

Purification of the serum extract was done by chromatography on 80-200 mesh (Fisher Sci Co) neutral alumina column previously activated by successive washes with Ethanol, Methanol, Cl2CH2: Methanol 1:1and Cl2CH2. Chromatographic separation was achieved by elution 1) with 5 fractions of 1 ml Ethanol 1.2 % in Hexane (androstenedione) and 2) with 5 fractions of 1 ml Ethanol 3 % in Hexane. Radioactivity is measured in one aliquot of each fraction and the solvent is evaporated from the elution in the tube with the highest radioactivity concentration. The extract was resuspended in 0.8 ml buffer and RIA was performed in 0.2 ml and 0.4 ml aliquots corresponding to 100 and 200 male and female serum microliters respectively. The method permits quantification of TT levels in a certain standard curve zone for TT values between 0.6 and 7.2 ng/ml with a coefficient of variation lower than 6 %, with the following conditions: within the said range the values in the two dilutions used ran parallel to the curve, and, in such range of the curve, it is possible to reliably assess TT levels as from 0.3 ng/ml in men and as from 0.15 ng/ml in women.

Results obtained for each serum by the methods currently employed in each lab were compared to those obtained by LC-MS/MS (Quest labs, USA). TT determinations were repeated by the very same methods but this time using as calibration curve the serum mixtures prepared according to the values obtained by LC-MS/MS (secondary calibrators) as it is described in the paragraph above. These 6 serum mixtures were included as samples in each one of the runs of the various intervening labs and each participant reported the dose values obtained. For calculation purposes, we have identified each one of these values as “counts”.

On the basis of each dose value obtained by LC-MS/MS as dose values for a calibration curve, and taking the dose values obtained by the various laboratories as “counts”, a calibration curve was constructed for each laboratory in which the dosage values / TT counts obtained for each patient were extrapolated. Results calculated with the curve thus constructed unify the dose values reported and therefore relieve us from following the recommendations attached to the specific insert in every method.

STATISTICAL STUDY

As regards the statistical analysis of the Axn, EQL, IMM, Arch and RIA S method, the measurements corresponding to the respective laboratories using these methods were averaged for every sample. Such average was taken as the TT concentration of the method in question.

Descriptive statistics: mean, standard deviation (SD), median, minimum and maximum were calculated for each one of the 9 methods.

The TT concentrations measured for each one of the 9 methods were compared with those obtained by LC-MS/MS using paired-sample Wilcoxon nonparametric signed-rank test. Fisher’s exact two-tailed test was run, too.

Spearman’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 is significantly different from 0 (zero).

In order to analyze the concordance of each one of the 9 methods with one another and with LC-MS/MS, Bland-Altman[12] analysis was performed. Limits of agreement were calculated by estimating the mean difference ± 1.96 standard deviation (STD) of the differences.

Weighted Deming regression[13] was run on each of the 9 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 values obtained by both methods.

This research was done according to the ethical standards of the Declaration of Helsinki.

RESULTS

Figures 1 and 2 show comparative results for NW of the various immunoassays contrasted against those obtained by LC-MS/MS. Only one NW presented TT values by LC-MS/MS above the normal range (> 0.45 ng/ml). TT levels in this sample were normal with all immunoassays except for one participant with the Arch technique.

Table 2 shows the values of TT, median, mean, SD and concentration range (ng/ml) obtained both by LC MS/MS and by immunoassay for the samples from NW and HW. The results of the statistical analysis of the data are also entered. For normal and HW "n" was 24 and 15 respectively except for DiaS at 21 and 13 respectively for the two groups.

Figure 3 shows TT values determined by LC-MS/MS in the 15 HW. As it can be seen, the method yielded increased TT values for 13 out of the 15 HW.

Figures 4 and 5 present a comparison of the results by the various immunoassays versus those resulting from the tests by LC-MS/MS in NW. From the HW with high TT scores by LC-MS/MS, increased values resulted in: 2 cases by Axn 5 cases by EQL, 7 cases by IMM, 5 by Arch, 8 by RIA S, 3 by DiaS, 2 by Vidas and 12 cases by In-H. A further characteristic of the results obtained by the different methods in contrast with LC-MS/MS is that the 2 HW with normal TT by the latter method, TT values in one case by IMM, by RIA S and DiaS values were above the upper cut-off values suggested for those methods. Table 3 shows in percentages the HW with normal TT for whom increased TT values were obtained by LC-MS/MS. For the results obtained by the In-House method

TABLE 2. Testosterone values obtained by 9 testosterone immunoassay and LC MS/MS for samples from normal women (up) and hirsutes women (down).

Methods	LC MSMS	Axn	EQL	IMM	Arch	Ria S	Ria DSL	DiaS	In H	Vidas
Mean	0.27	0.22	0.34	0.26	0.33	0.36	0.30	0.56	0.29	0.23
SD	0.10	0.10	0.21	0.18	0.14	0.18	0.18	0.24	0.10	0.15
Median	0.28	0.21	0.37	0.16	0.35	0.33	0.24	0.59	0.30	0.19
Minimum	0.13	0.07	0.08	0.02	0.12	0.15	0.12	0.15	0.14	0.10
Maximum	0.53	0.44	0.85	0.62	0.62	0.95	0.87	1.00	0.50	0.79
p(Wilcoxon)		0.013	0.107	0.710	0.017	0.017	0.560	0.000	0.534	0.043
Bland-Altman analysis:										
mean difference		0.05	-0.07	0.01	-0.06	-0.09	-0.03	-0.29	-0.02	0.03
agreement limits		-0.10	-0.40	-0.26	-0.28	-0.41	-0.30	-0.70	-0.19	-0.17
Sig bilateral		0.00	0.001	0.00	0.002	0.029	0.001	0.022	0.001	0.00

Methods	LC MSMS	Axn	EQL	IMM	Arch	Ria S	Ria DSL	DiaS	In H	Vidas
Mean	0.74	0.50	0.83	0.87	0.80	0.88	0.76	0.56	0.29	0.23
SD	0.33	0.21	0.52	0.44	0.64	0.43	0.60	0.24	0.10	0.15
Median	0.69	0.47	0.69	0.75	0.55	0.77	0.55	0.59	0.30	0.19
Minimum	0.35	0.09	0.34	0.23	0.41	0.44	0.31	0.15	0.14	0.10
Maximum	1.50	0.92	2.05	1.95	2.87	1.92	2.69	1.00	0.50	0.79
p(Wilcoxon)		0.011	0.359	0.066	0.679	0.015	0.187	0.000	0.534	0.043
Bland-Altman analysis:										
mean difference		0.25	-0.09	-0.13	-0.06	-0.14	-0.02	-0.29	-0.02	0.03
agreement limits		-0.35	-0.62	-0.61	-0.90	-0.50	-0.74	-0.70	-0.19	-0.17
Sig bilateral		0.098	0.000	0.000	0.001	0.000	0.000	0.289	0.000	0.000

Axn: Abbott - Axsym; EQL: Roche - Elecsys; Imm: Siemens Immulite; Arch: Abbott - Architect; RIA S: Siemens RIA; RIA DSL: RIA DSL Inc; DiaS; DiaSource RIA; In H: In home RIA; Vidas: Bio-Merieux

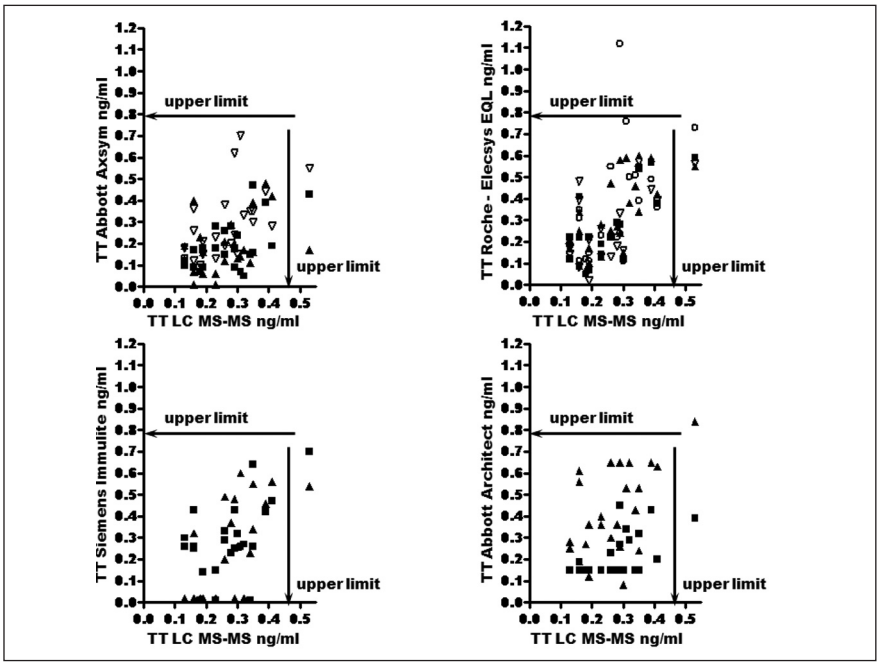

Figure 1. Relation between Testosterone levels obtained by various methods and by LC-MS/MS in normal women. The different symbols in each set of results indicate the various laboratories that employed the same methodology. The arrows in the upper section of each graph indicate the upper cut-off limits for each method according to its insert. The arrows on the right of each set indicate the upper limit for normal women obtained by LC-MS/MS.
Note: Axn: Abbott-Axsym; EQL: Roche-Elecsys; Imm: Siemens Immulite; Arch: Abbott-Architect.

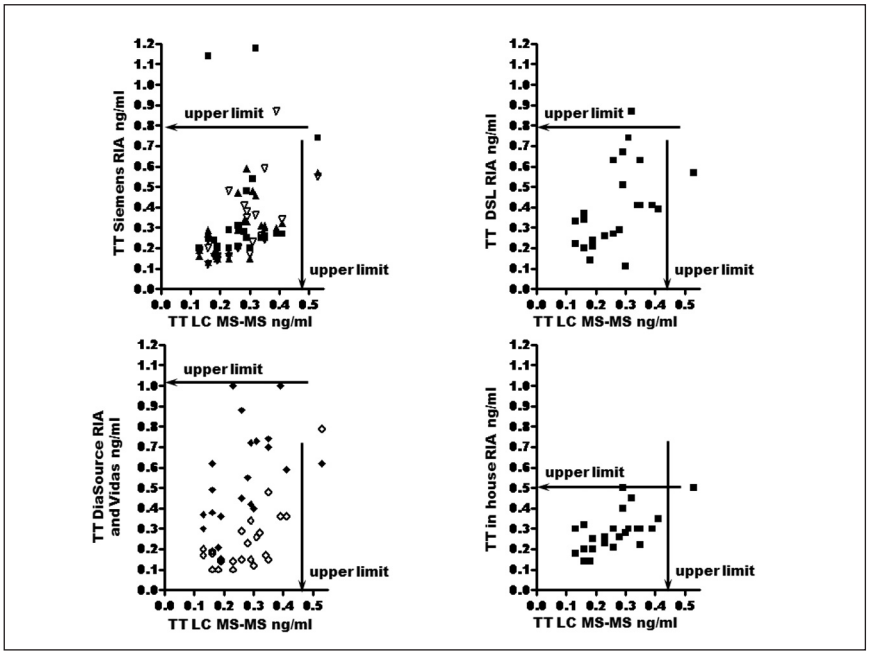

Figure 2. Relation between Testosterone levels obtained by various methods and by LC-MSMS in normal women. The different symbols used for the set of results obtained by the Siemens RIA method indicate the various laboratories that employed the same methodology. The results expressed in black diamonds and white diamonds stand respectively for results obtained by DiaSouce RIA and by Vidas methods. The arrows in the upper section of each graph indicate the upper cut-off limits for each method according to its insert. The arrows on the right of each set indicate the upper limit for normal women obtained by LC-MS/MS.
Note: Vidas: Bio-Merieux; RIA S: Siemens RIA; RIA DSL: RIA DSL Inc; In H: In home RIA; Dia S: DiaSource RIA.

discrepancy by reference to LC-MS/MS was 12 % while for the remaining methods it ranged between 31 and 85 %. The greatest discrepancies occurred at the lowest TT concentrations. For increased TT levels by LC-MS/MS within the range 0.47 to 0.74 ng/ml (cut-off value for this method is 0.45 ng/ml) the average discrepancy was 71 % while for values by LC-MS/MS above 0.9 ng/ml discrepancy was 38.5 %.

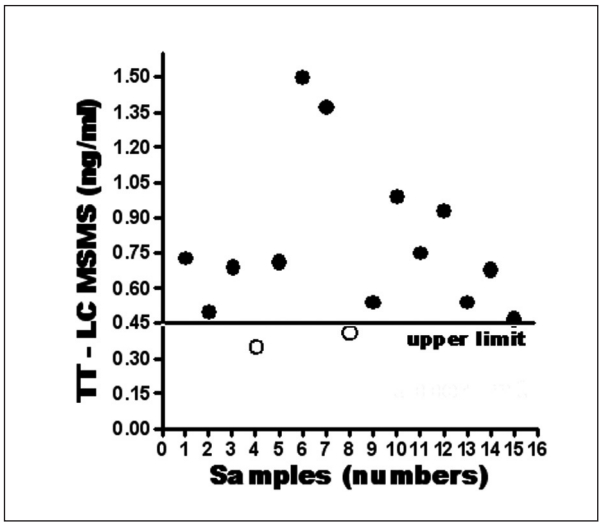

Figure 3. Results of TT tested by LC-MS/MS in 15 patients with hirsutism, acne with or without cycle abnormalities (Quest Diagnostics, Nichols Institute).

Significance levels for paired-sample Wilcoxon non-parametric signed-rank test are shown in Table 2. For methods Axn, EQL, Imm, Arch and RIA S methods employed by two or more labs the average determination value was used. P-values below 0.05 ($p < 0.05$) were considered statistically significant. In NW the results yielded by methods Axn, Arch and RIA S, DiaS and Vidas methods differ significantly from those by LC-MS/MS. For both groups p was below 0.05 ($p < 0.05$). In the HW group TT values by Axn and DiaS showed no significant correlation vs. those by LC-MS/MS (Spearman's correlation 0.324 and 0.292 respectively and bilateral Sig 0.24 and 0.332 respectively). Other correlations proved statistically significant ($p < 0.05$).

Weighted Deming regression was run on each one of the 9 quantification methods taking LC-MS/MS as independent variable (x) and each method (y) as dependent variable. The regression

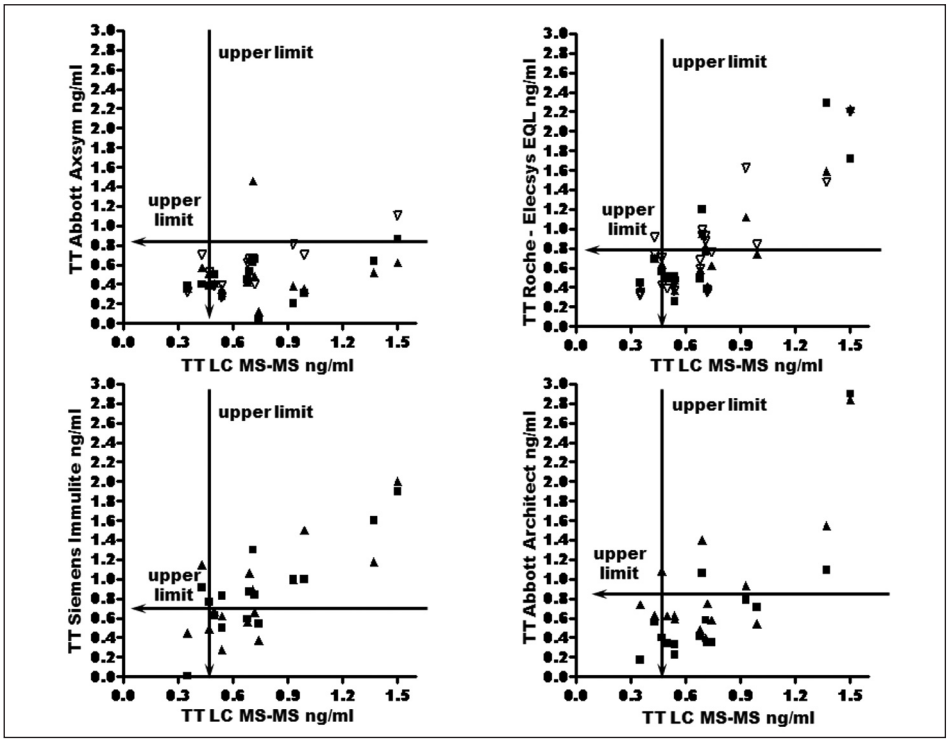

Figure 4. Relation between Testosterone levels obtained by various methods and by LC-MS /MS in hirsute women. The different symbols in each set of results indicate the various laboratories that employed the same methodology. The arrows in the upper section of each graph indicate the upper cut-off limits for each method according to normal women values obtained in this study. The arrows on the right of each set indicate the upper limit for normal women obtained by LC-MS/MS.
Note: Axn: Abbott - Axsym; EQL: Roche - Elecsys; Imm: Siemens Immulite; Arch: Abbott - Architect.

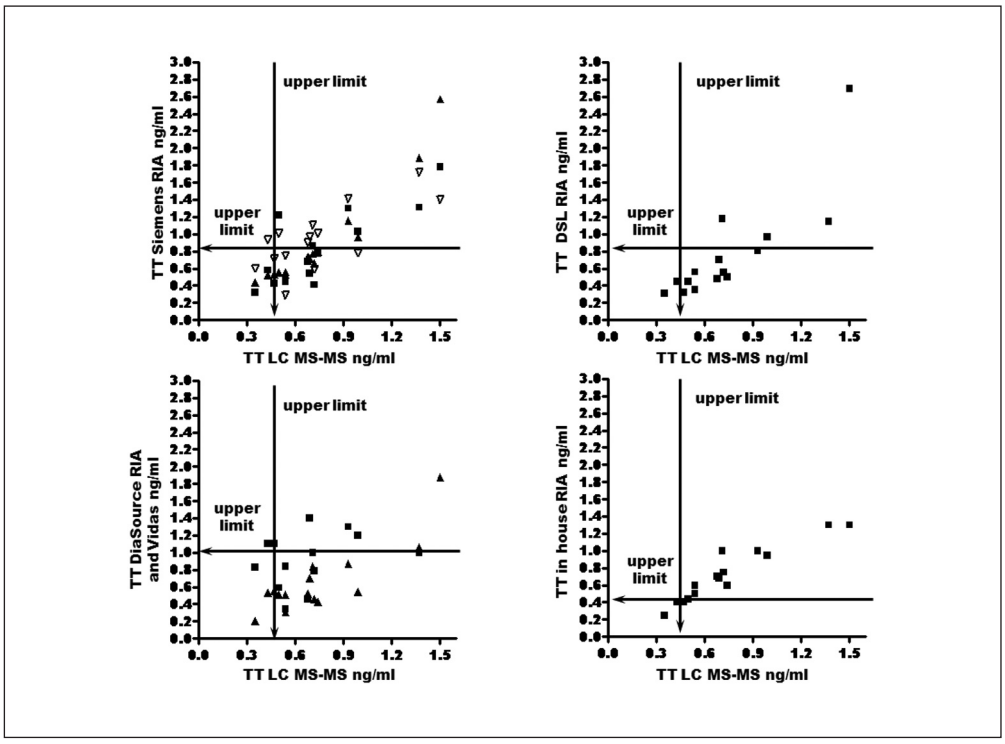

Figure 5: Relation between Testosterone levels obtained by various methods and by LC-MS/MS in hirsute women. The different symbols used for the set of results obtained by the Siemens RIA method indicate the various laboratories that employed the same methodology. Black squares and black triangles stand respectively for results obtained by DiaSouce RIA and by Vidas methods. The arrows in the upper section of each graph indicate the upper cut-off limits for each method according to normal women values obtained in this study. The arrows on the right of each set indicate the upper limit for normal women obtained by LC-MS/MS.
Note: Vidas: Bio-Merieux; RIA S: Siemens RIA; RIA DSL: RIA DSL Inc; In H: In home RIA; Dia S: DiaSource RIA.

TABLE 3. Percentages the hirsute women with normal testosterone levels for whom increased testosterone values were obtained by LC-MS/MS

Method (# of labs):	1[3]	2[4]	3[2]	4[2]	5[3]	6[1]	7[1]	8 (1)	9 (1)
Samples with normal T (%):	85	52	42	70	64	31	31	12	69

Note: [1]Axn: Abbott - Axsym; [2]EQL: Roche - Elecsys; [3]Imm: Siemens Immulite; [4]Arch: Abbott - Architect;
5: RIA S: Siemens RIA; 6: RIA DSL: RIA DSL Inc 7: Dia S: RIA DiaSource; 8: In H: In home RIA; 9: Vidas: Bio-Merieux
Number between brackets represent the number of lab which used the same technique

was not calculated for Axn since it did not prove concordant in any group. This equation allows the ratio between the two methods to be established. Table 4 for each shows the results obtained for each method both in the NW group as in the HW group. The In-H method is the one that comes closest to 1 in the equation and to SD intercept also has close to zero indicating that the values match with the LC- MS/MS method.

The results from NW samples analyzed by the ratio method are shown in Figure 6. As can be seen, Vidas, Axn, and RIA S underestimated TT for the range of female concentrations tested. Wilcoxon test yielded significant differences in values by LC-MS/MS ($p < 0.02$ for Arch and RIA S and $p < 0.042$ for Vidas). DiaS overestimated results in the range of female concentrations tested. Values for women's TT

TABLE 4. Results of the regression equation weighted Deming LC MSMS method taking as the independent variable (x) and each method (y) as the dependent variable. For each method, the (n) for normal women (NW) is 24 and hirsutism women (HW) is15, except for DiaSource, NW (n: 21) and HW (n:13).

Method	Group	Equation	SD Intercep
Elesys (EQL)	NW	y: -0.226 + 2.104 x	0.2000
	HW	y: -0.334 + 1.572 x	0.1281
Siemens Immulite IMM)	NW	y: -0.234 + 1.843 x	0.1132
	HW	y: -0.115 + 1.328 x	0.1739
Architect (Arch)	NW	y: -0.048 + 1.405 x	0.1002
	HW	y: -0.629 + 1.924 x	0.7314
Siemens Coat-a-tube	NW	y: -0.148 + 1.877 x	0.3144
(Ria S)	HW	y: -0.082 + 1.296 x	0.1008
DSL Inc (RIA DSL)	NW	y: -0.178 + 1.791 x	0.1931
	HW	y: -0.590 + 1.821 x	0.5541
DiaSource (DiaS)	NW	y: -0.012 + 2.149 x	0.2775
	HW	y: -0.048 + 1.410 x	1.1361
In home methods (In H)	NW	y: 0.0186+1.002 x	0.0708
	HW	y: -0.002 + 0.976 x	0.0854
Bio-Merieux (Vidas)	NW	y: -0.185 + 1.560 x	0.1444
	HW	y: -0.242 + 1.214	0.3076

concentration were widely dispersed for the other immunoassays.

Following Bland and Altman's plots, limits of agreement specified the mean difference ±1.96 STD for each method, considering at outside the limits of agreement more than 5 % of results lying outside the limits in the quotient graph. The mean ± SD of the TT concentrations ratio as well as the limits for each method are shown in Figure 6. Two results were obtained (over 5 %) that fell outside the limits for EQL, IMM, Arch, whereas it was concordant for DSL, and In-H.

The results for the HW samples analyzed by the ratio method are shown in Figure 7. Wilcoxon test yielded significant differences by LC-MS/MS ($p < 0.02$) for RIA S. Figure 7 also shows the mean ± SD of the TT concentrations ratio as well as the limits of agreement for each method according to Bland and Altman's plots. Results for EQL and Arch (more than 5 %) fell outside the limits. As for IMM values, even though they do not fall outside concordance bands, one of them is on the line and several others in the neighboring area. Values by RIA DSL proved concordant. Still, most of the values are plotted above the baseline, which suggests that values by this method are higher than those obtained by LC-MS/MS. Like results were obtained by DiaS, RIA S, In-H and Vidas.

Figure 8 shows the results of the secondary curve obtained by the different methods. As this curve was used as a secondary standard and the various sera which had previously yielded discordant values by LC-MS/MS were reanalyzed, the results were not significantly modified (Table 5a y 5b). Given that the In-H method employs as standard curve a serum pool assessed by LC-MS/MS, the results of the samples reanalyzed with the secondary curve were identical to those previously obtained.

DISCUSSION

The comparative study of the TT concentrations obtained from aliquots of the same NW and HW blood samples showed that there is great qualitative and quantitative variability among the different methods in current use in our area. A comparison of the results obtained revealed significant differences between these methods and Quest Lab using turbulent flow liquid chromatography / tandem mass spectrometry[14]. Similar results have been previously published[9, 15-17].

In the NW group, 1 of the 15 NW showed TT levels above the upper cut-off limit for LC-MS/MS. Given that the inclusion criterion criteria in this group had been for women to present normal cycles, have no hirsutism and not being under any treatment whatsoever that could affect TT levels or testosterone transport, an increase in SHBG by contraceptives or other drugs inducing such effect must be ruled out. As this was a double-blind study, every participant lab had the sera sent to the project coordination centre; the aliquots were separated and given different labels, and then referred back to the intervening labs and for LC-MS/MS. Once all the methods were run for all the samples, one of the samples within the NW group was found to present an increased TT score by LC-MS/MS. The trial on TT transport on this patient could not continue due to an insufficient serum volume or the impossibility of drawing a new sample, with the hypothesis for his occurrence being an endogenous SHBG increase that accounts for the absence of hirsutism in the presence of free testosterone levels. This increase went undetected in all but one of the labs (Arch) while by the In-H method the score lay on the limit of the upper normal range. As for the samples resulting in normal

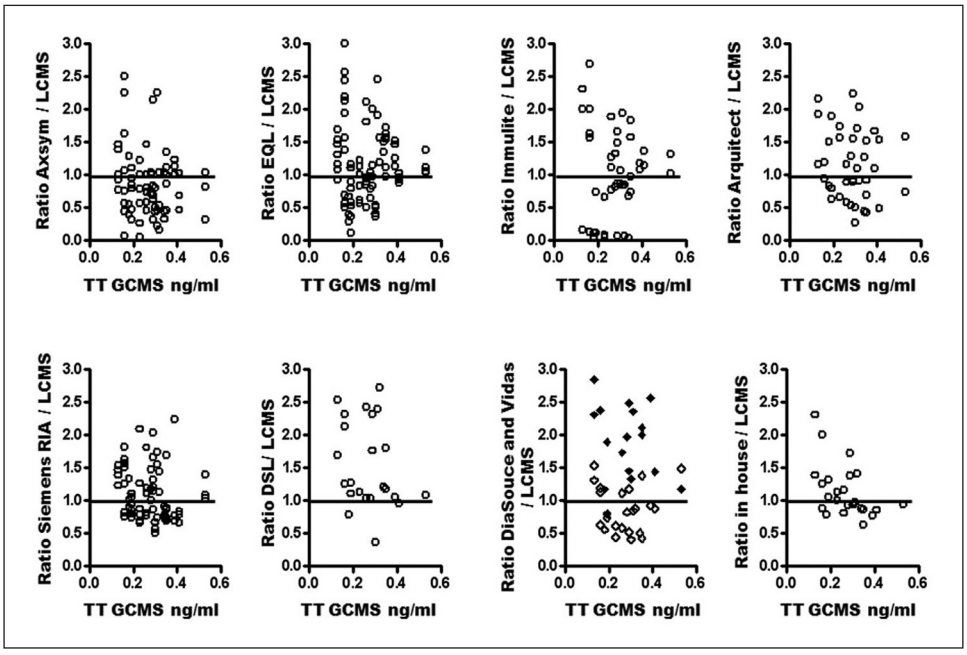

Figure 6. Testosterone levels in normal women obtained by the various immunoassays and LC–MS/MS methods as evaluated by the ratio method. .The y axis represents the ratio testosterone concentration by immunoassay / testosterone concentration by LC-MS/MS and the x axis represents the testosterone concentration measured by LC-MS/MS. Each point represents an individual value. Results obtained by Axn, EQL, Imm, Arch and RIA S correspond to those obtained in 3, 4, 2, 2 and 3 laboratories respectively employing such techniques. RIA DSL , DiaS, Vidas and an In-H correspond to a single one lab employing this methodology. The results expressed in black diamonds and white diamonds stand respectively for results obtained by DiaSouce RIA and by Vidas methods.
Note: Axn: Abbott - Axsym; EQL: Roche - Elecsys; Imm: Siemens Immulite; Arch: Abbott - Architect; Vidas: Bio-Merieux; RIA S: Siemens RIA; RIA DSL: RIA DSL Inc; In H: In home RIA; Dia S: DiaSource RIA.

TT by LC-MS/MS, on the other hand, increased scores were obtained in one sample for one of the labs employing EQL, in three samples for two of the labs using RIA S and in one sample by DSL. It must be noted that the methods used by more than one lab yielded widely dispersed values. No single factor can clearly explain such disparity; it could be due to differences among the various lots of commercial kit or in the calibration of the various types of equipment employed.

TT levels in hirsute patients by LC-MS/MS were high except in two cases showing normal scores. The aim of this study was to compare TT results obtained by different methods and by LC-MS/MS; for this reason the hirsute patients' previous diagnoses were not documented. According to the Androgen Excess Society (AES)' Guideline for defining PCOS, in which the syndrome is described as a predominantly hyperandrogenic disorder, these two women could constitute cases of hyperandrogenism without ovary dysfunction (Phenotype K), idiopathic hirsutism (Phenotype O) or PCOS without hyperandrogenemia (Phenotype E, F or H)[18].

TT values (range, median, mean and SD) showed high dispersion both in the HW and in the NW groups. This was seen not only among labs using different methods but also among those sharing the same technique. By comparison with the values obtained by LC-MS/MS, the differences among all the methods were considerable.

In order to determine if a given sample from the HW group had yielded normal or increased scores, the upper cut-off values obtained in the NW group for each method were taken as upper cut-off limits. All the methods employed in this study were found to underestimate the number of patients with hyperandrogenism in HW when comparing their scores with those obtained by LC-MS/MS. This gap widens when TT levels are lower. This mismatch is due to the fact that methods using liquid or gas chromatography, or better

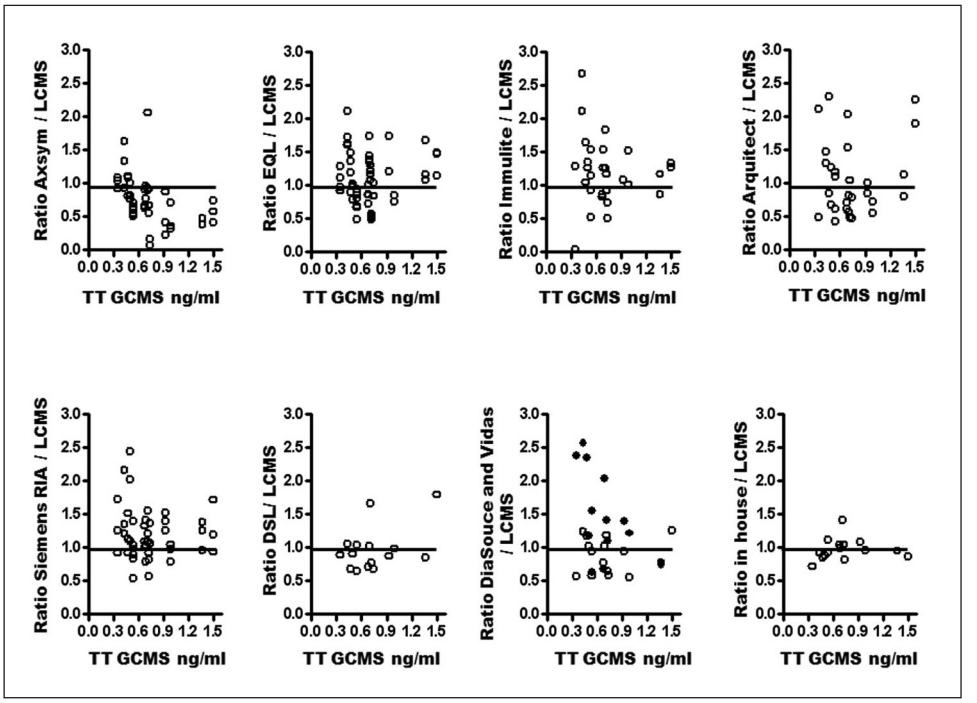

Figure 7. Testosterone levels in hirsute women obtained by various immunoassays and LC-MS/MS as evaluated by the ratio method. The y axis represents the ratio testosterone concentration by immunoassay / testosterone concentration by LC-MS/MS and the x axis represents the testosterone concentration measured by LC-MS/MS. Each point represents an individual value. Results obtained by Axn, EQL, Imm, Archt and RIA S correspond to those obtained in 3, 4, 2, 2 and 3 laboratories respectively employing such techniques. RIA DSL, DiaS, Vidas and In-H correspond to a single one lab employing this methodology. The results expressed in black circle and white circle stand respectively for results obtained by DiaSouce RIA and by Vidas methods. Note: Axn: Abbott - Axsym; EQL: Roche - Elecsys; Imm: Siemens Immulite; Arch: Abbott - Architect; Vidas: Bio-Merieux; RIA S: Siemens RIA; RIA DSL: RIA DSL Inc; In H: In home RIA; Dia S: DiaSource RIA.

still, those using an isotope-dilution - tandem MS/MS (ID GC-MS) measurement procedure suffer neither interference from cross reactions nor matrix effects; it is on account of these traits that they have been considered reference procedures for the evaluation of steroid immunoassays[19,20]. In a similar study, Ognibene et al. demonstrated that highly dispersed results were obtained at very low TT concentrations or, with some systems, notably with Immulite 2000, there was a marked underestimation of values. The results in the latter study revealed that only Architect i2000 proved concordant with ID GC-MS[21] at low TT concentration levels. Our results, on the other hand, differ from those presented by Ognibene, since for the intervening labs using Arch the scores obtained by this method did not correlate with those by LC-MS/MS. This discrepancy could possibly be due to the different types of HW sample analyzed.

In short, the various immunoassays employed for determination of TT in NW and HW yielded differing scores from those by LC-MS/MS. Particularly among HW, in qualitative terms, there was a lower number of women with hyperandrogenemia. This, from a clinical standpoint, turns out to be a problem when it comes to phenotyping individuals.

From a statistical point of view, the equation obtained by weighted Deming regression analysis in NM and in HW was greater than 1 for all methods except In-H. The intercept also showed high dispersion, indicating that the methods do not adequately correlate with LC-MS/MS. The better concordance demonstrated by In-H could be due to the fact that the methodology allows for a greater sample volume to be processed, with a consequent improved analytical sensitivity. The lower sensitivity of the various immunoassays at low TT concentrations could possibly be a non-concordance factor with LC-MS/MS.

An analysis of Wilcoxon non-parametric test, performed to compare the TT concentrations

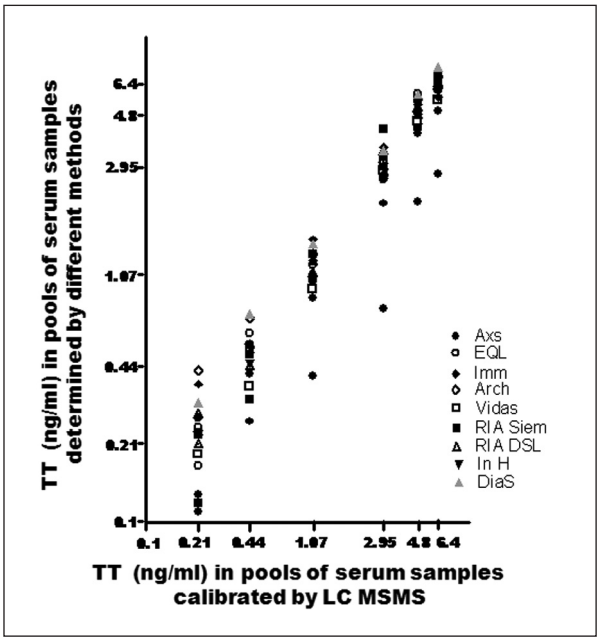

Figure 8. Six serum pools obtained from 76 sera: 37 men, 24 normal women and 15 hirsute women used as secondary calibrators for every method used for Testosterone determination. Results were obtained by LC-MS/MS. The mixture was achieved with values (ng/ml) greater than 6; between 4 and 6; between 2 and 4; between 0.7 and 2; between 0.3 and 0.7, and smaller than 0.3. The mean value ± DS for every mixture respectively (ng/ml) was 6.41±0.73; 4.80±0.70; 2.95±0.52; 1.07±0.37; 0.44±0.13 and 0.21±0.06.
Note: Axn: Abbott - Axsym; EQL: Roche - Elecsys; Imm: Siemens Immulite; Arch: Abbott - Architect; Vidas: Bio-Merieux; RIA S: Siemens RIA; RIA DSL: RIA DSL Inc; In H: In home RIA; Dia S: DiaSource RIA.

measured for each one of the methods by reference to LC-MS/MS, revealed significant differences ($p < 0.05$) in the NW group by Axn, Arch, RIA S, DiaS and Vidas, and so it did among HW by Axn, RIA S and DiaS.

Bland-Altman's analysis, too, shows a lack of concordance between the results obtained by LC-MS/MS and by the different immunoassays both for NW and for HW, which goes to prove there is a fair amount of imprecision in the latter methods.

From a clinical point of view, most –if not all- of the immunoassays evaluated in this study failed to draw adequate distinctions in TT levels within the NW range so as to actually tell normal from abnormal androgen concentrations. One of the possible reasons for this defective correlation between the immunoassays and LC-MS/MS could be found in the methods' detection limits and their low functional sensitivity. That is why, while not proving thoroughly validated, In-H, which allows extraction of greater serum volumes, was the method to yield the best correlation coefficient and fewest false negatives for the HW group (normal TT values in samples with increased values by LC-MS/MS).

There is no universally acknowledged TT calibration standard (9). On this account, we set up a secondary reference standard on the basis of serum mixes whose concentrations had been determined by LC-MS/MS. The reassessment of the NW and HW samples calculated according to this preparation did not significantly modify the results. This would suggest that it is not actually the standard that is responsible for dissimilarities between chromatographic and immunoassay methods and among immunoassays themselves.

As it has been previously described, what could account for poor concordance between the immunoassays and LC-MS/MS is a cross reaction of the antibodies with related steroids[22] or of drugs interacting with SHBG[23] in direct immunoassays. Tracers, particularly non-tritiated ones depending on SHBG affinity and certain types of antibody employed[24, 25] could also be responsible for low concordance rates. Extractive and chromatographic purification procedure on serum samples of NW following the recommendations attached to the specific insert in every method did not improve the high results variability observed among the different techniques available[26].

TT determination has long been used as a useful tool in the diagnosis of hypo- and hyperandrogenemia-associated conditions. A major downside, though, turns out to be the poor comparability among the results obtained by the various methods. With that in mind, every effort should be directed towards standardization of the quantifying process. Special emphasis should be placed on pre- and post analytical aspects such as the reference range. The Centers for Disease Control and Prevention, National Center for Environment Health, Division of Laboratory Sciences have recently released a project for standardization of steroid hormone assessment[27]. This study proposes a common calibration standard, that is, a primary standard[28], the matrix employed by kit manufacturers to calibrate their immunoassay, sample preparation[29], validation of results to a reference technique[30-31], as well as clear-cut, well-adjusted values for an adequate number of subjects in a well-characterized population, utilizing a well-defined procedure. In the future a study of

TABLE 5A. Testosterone level results in 5 serum samples. The levels of which were determined by LC MS/MS and recorded for each of the samples were determined by various laboratories according to the package inserts for each method (I) and repeated by calculating them with the calibration standards prepared on the basis of the values obtained by LC MS/MS (LC).

	Sample1 0.30 ng/ml		Sample 2 0.26 ng/ml		Sample 3 3.10 ng/ml		Sample 4 3.10 ng/ml		Sample 5 0.54 ng/ml	
Method	I	LC	I	LC	I	LC	I	LC	I	LC
Axn	0.36	0.34	0.48	0.50	2.90	3.40	2.60	3.06	0.65	0.71
Axn	0.20	0.43	0.10	0.14	2.30	5.58	1.50	3.88	0.31	0.75
Axn	0.19	0.29	0.30	0.41	2.83	3.05	2.78	3.00	0.40	0.52
EQL	0.24	0.18	0.29	0.24	3.10	3.23	3.00	3.13	0.61	0.58
EQL	0.11	0.16	0.24	0.27	3.67	3.32	3.57	3.22	0.52	0.52
EQL	0.16	0.13	0.29	0.25	3.47	3.21	3.52	3.26	0.65	0.58
EQL	0.17	0.13	0.28	0.25	3.29	3.46	3.22	3.39	0.57	0.56
IMM	0.23	0.21	0.40	0.32	3.30	3.47	3.00	3.40	0.70	0.64
IMM	0.22	0.12	0.46	0.33	3.54	2.94	3.45	2.85	0.87	0.67
Arch	0.40	0.19	0.45	0.24	3.46	3.14	3.43	3.11	0.66	0.44
Vidas	0.13	0.13	0.20	0.22	3.66	3.82	2.99	3.13	0.52	0.64
RIA S	0.10	0.20	0.36	0.46	3.20	2.43	3.23	2.45	0.38	0.48
RIA S	0.15	0.14	0.29	0.26	3.03	2.89	3.05	2.92	0.50	0.44
RIA DSL	0.12	0.12	0.26	0.26	4.00	4.00	4.00	4.00	0.60	0.60
RIA S	0.28	0.21	0.41	0.33	3.10	2.82	3.45	3.16	0.66	0.56
In H	0.29	0.28	0.28	0.28	3.12	3.10	3.11	3.10	0.50	0.51
DiaS	0.23	0.10	0.42	0.27	3.80	3.17	4.10	3.43	0.65	0.46

Note: Axn: Abbott - Axsym; EQL: Roche - Elecsys; Imm: Siemens Immulite; Arch: Abbott - Architect; Vidas: Bio-Merieux; RIA S: Siemens RIA; RIA DSL: RIA DSL Inc; In H: In home RIA; Dia S: DiaSource RIA Samples 3 and 4 are the same. Participants did know it when the samples were processed.

TABLE 5B. Testosterone level results in 4 serum samples The levels of which were determined by LC MS/MS and recorded for each of the samples were determined by various laboratories according to the package inserts for each method (I) and repeated by calculating them with the calibration standards prepared on the basis of the values obtained by LC MS/MS (LC).

	Sample 6 0.35 ng/ml		Sample 7 3.73 ng/ml		Sample 8 3.73 ng/ml		Sample 9 0.28 ng/ml	
Method	I	LC	I	LC	I	LC	I	LC
Axn	0.58	0.62	3.48	4.03	4.60	5.21	0.27	0.23
Axn	0.43	1.11	1.60	4.10	1.99	4.93	0.32	0.78
Axn	0.61	0.76	3.02	3.28	3.15	3.44	0.44	0.57
EQL	0.64	0.61	3.94	4.14	3.80	3.98	0.92	0.91
EQL	0.66	0.64	4.38	3.95			0.51	0.51
EQL	0.69	0.62	4.21	3.90	4.16	3.85	0.65	0.58
EQL	0.67	0.67	3.81	4.01	3.83	4.04	0.56	0.55
IMM	0.45	0.36	4.00	4.10	3.40	3.60	0.90	0.92
IMM	0.76	0.58	4.37	3.78	4.11	3.50	0.56	0.42
Arch	0.77	0.55	3.62	3.29	3.61	3.28	0.71	0.49
Vidas	0.41	0.50	3.81	3.98	4.08	4.28	0.45	0.55
RIA S	0.28	0.38	3.36	2.53	5.16	4.13		
RIA S	0.39	0.34	3.38	3.33	3.22	3.13	0.52	0.46
RIA DSL	0.38	0.38	3.00	3.00	3.00	3.00	0.10	0.10
RIA S	0.52	0.43	3.51	3.22	4.07	3.76	0.63	0.53
In H	0.38	0.36	3.70	3.70	3.68	3.70	0.21	0.26
DiaS	0.69	0.49	3.10	2.57	3.10	2.57	0.68	0.49

Note: Axn: Abbott - Axsym; EQL: Roche - Elecsys; Imm: Siemens Immulite; Arch: Abbott - Architect; Vidas: Bio-Merieux; RIA S: Siemens RIA; RIA DSL: RIA DSL Inc; In H: In home RIA; Dia S: DiaSource RIA Participants did know it when the samples were processed Samples 7 and 8 are the same. Participants did know it when the samples were processed.

this type will permit validation of the various immunoassays used for different endocrine diseases in routine clinical practice.

In conclusion, similarly to results previously obtained elsewhere and, as it has been widely discussed in the course of this communication, our results clearly demonstrate that none of the immunoassays in current use locally present sufficient precision for quantifying concentrations lower than 1.5 ng/ml, such as those in normal, hirsute women and children levels.

Acknowledgments: The authors thank Delia Garrido and Maria del L. Calcagno for support with the statistical analyses of our data and Ana Moldero for manuscript preparation.

REFERENCES

1. **Wang C, Swerloff RS.** Evaluation of testicular function. Ballieres Clin Endocrinol Metab 6:405-434, 1992
2. **Knoblovits P, Suárez SM, Scaglia HE, Zylbersztein CC, Litwak LE, Costanzo PR.** Hypothalamic pituitary gonadal axis function in men with type 2 diabetes mellitus. Annual Meeting of Endocrine Society. P3-218, 2011
3. **Kuhn JM, Billebaud T, Navratil H, Moulonguet A, Fiet J, Grise P, Louis J-F, Costa P, Husson L-M, Dahan R, Bertagna C, Edelstein R.** Prevention of the transient adverse effects of a gonadotropin-releasing hormone analogue (buserelin) in metastatic prostatic carcinoma by administration of an antioandrogen (nilutamide). N Engl J Med 321:413-418, 1989
4. **Azzis, R. Carmina E, Sawaya M:** Idiopathic Hirsutism. Endocrine Reviews 21 (4):347-362; 2000
5. **Azzis R, Sanchez LA, Knochenhauer ES, Moran C, Lazenby J, Stephens KC, Taylor K, Boots LR.** Androgen excess in women: experience with over 1000 consecutive patients J Clin Endocrinol Metab 89: 453-462, 2004
6. **Zawadski JK, Dunaif A.** Diagnostic criteria for polycystic ovary syndrome: Toward a rational approach. In Dunaif A, Givens JR, Haseltine FP, Merriam GR, eds. Polycystic Ovary Syndrome, Boston: Blackwell Scientific, 377-384, 1992
7. The Rötterdam ESHRE/ASRM- sposored PCOS consensus workshop group. Revised 2003 consensus on diagnostic criteria and long term health risks to PCOS. Hum Reprod 19:41-47, 2004
8. **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 POCS: the complete task force report. Fert Steril 91:456-488, 2009
9. **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
10. **Ferriman D, Gallwey JD.** Clinical assessment of body hair growth in women J Clin Endocrinol Metab 21:1440-1447, 1961
11. **Scaglia HE, Zylbersztein CC, Aquilano DR.** Increased analytical sensitivity of a testosterone ria using chromatography - purifeid ether extracs. its application for evaluating hipogonadal patients. 13th International Congress of Endocrinology. Medimond International Procedings. Brazil, November 8-12, pag 205-209, 2008
12. **Bland JM, Altman DG.** Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307-310, 1986
13. **Linnet K.** Neccesary sample sixe for method comparison studies based on regression analysis. Clin Chem. 45:882-894, 1999
14. **Grant RP, Cameron C, Mackenzie-McMurter S.** Generic serial and parallel on-line direct - injection using turbulent flow liquid chromatography/tandem mass spectrometry. Rapid Commun Mass Spectrom. 16:1785-1792, 2002
15. **Boots LR, Potter S, Potter D, Azziz R.** Measurement of total serum testosterone levels using commercially avaible kits: high degree of between - kit variability.Fert Steril 69:286-292, 1998
16. **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
17. **Fitzgerald RL, Herold DA.** Serum total testosterone, immunoassay compared with negative chemical ionization gas chromatography. Clin Chem 42:749-755, 1996
18. **Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Futterweit W, Janssen OE, Legro RS, Norman RJ, Taylor AE, Witchel SF.** 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
19. **Siekmann L.** Determination of steroid hormones by the use of istope dilution-mass spectrometry: a definitive method in clinical chemistry. J Steroid Biochem 11:117-123, 1979
20. **Lawson AM, Gaskell SJ, Hjelm M.** International Federation of Clinical Chemistry (IFCC). Office for Reference Methodos and Materials (ORMM) Methodological aspects on quantitative mass spectrometry used for accuracy control in clinical chemistry. J Clin Chem Clin Biochem 23:433-441, 1985
21. **Ognibene A, Drake CJ, Jeng KY, Pascucci TE, Hsu S, Luceri F.** A new modular chemiluminescence immunoassay analyser evaluated. Clin Chem Lab Med 38:254-260, 2000
22. **Bodlaender, P.** No SHBG interference with coat -

A tube total testosteronedirect RIA kit. Clin Chem 36:173, 1990

23. **Chattoraj S.** Endocrine fuction. Tietz NW eds. WB Saunders Philadelphia Fundamentals of clinical chemistry 699-817, 1976.
24. **Pugeat MM, Dunn JF, Nisula BC.** Transport of steroid hormones: Interaction of 70 drugs with testosterone - binding globulin and corticosteroid - binding globulin in human plasma. J Clin Endocrinol Metab 53:69-75, 1981
25. **Micallef JV, Hayes MM Ltif A, Ahsan R, Sufi SB.** Serum binding of steroid tracers and its possible effects on direct steroid immunoassay. Ann Clin Biochem 32:566-574, 1995
26. **Scaglia HE, Role N, Neira F, Fuillerat S, Corthey C, Riesco O, Scaglia J, Zylbersztein CC, Chichizola C and Aquilano DR.** Extractive and chromatographic procedures on serum samples of normal women for testosterone measurement did not improve the high results variability observed among available methods. The Endocrine Society's 88 th Annual Meeting. Boston, Massachusetts, USA, june 24-27, 2006. Abstract P2-623
27. **Vesper HW, Botelho JC.** Standardization of testosterone measurements in humans. J Steroid Bioch Mol Biology 121:513-519, 2010
28. **Vesper HW, Thienpont LM,** Traceability in laboratory medicine. Clin Chem 55:1067-1075, 2009
29. **Botelho JC, Shacklady C, Razdan R Wesper HW**. Impact of sample preparation procedures of testosterone measurement by isoopic-dilution liquid chromatography-tandem mass spectrometry. Clin Chem 54 Abstract 122, 2008
30. **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 isoopic-dilution liquid chromatography-tandem mass spectrometry. Clin Chem 54:1290-1297, 2008
31. **Vesper HW, Bhasin S, Wang C, Tai SS, Dodge LA, Sinh RJ, Nelson J, Ohorodnik S, Clarke NJ, Parker Jr CR, Razdan R, Monsell EA, Myers GL.** Interlaboratory comparison study of serum total testosterone measurements performed by mass spectrometry methods. Steroids 74:498-503, 2009