European Journal of Endocrinology (2012) **166** 983–991 ISSN 0804-4643

CLINICAL STUDY

Comparison of serum testosterone and estradiol measurements in 3174 European men using platform immunoassay and mass spectrometry; relevance for the diagnostics in aging men

Ilpo T Huhtaniemi[1], Abdelouahid Tajar[2], David M Lee[2], Terence W O'Neill[2], Joseph D Finn[3], György Bartfai[4], Steven Boonen[5], Felipe F Casanueva[6,7], Aleksander Giwercman[8], Thang S Han[9], Krzysztof Kula[10], Fernand Labrie[11], Michael E J Lean[12], Neil Pendleton[13], Margus Punab[14], Alan J Silman[15], Dirk Vanderschueren[16], Gianni Forti[17], Frederick C W Wu[3] and the EMAS Group

[1]Department of Surgery and Cancer, Imperial College London, Hammersmith Campus, London, UK, [2]Arthritis Research UK Epidemiology Unit, Manchester Academic Health Science Centre, The University of Manchester, Manchester, UK, [3]Developmental and Regenerative Biomedicine Research Group, Andrology Research Unit, Manchester Academic Health Science Centre, Manchester Royal Infirmary, The University of Manchester, Manchester, UK, [4]Department of Obstetrics, Gynaecology and Andrology, Albert Szent-György Medical University, Szeged, Hungary, [5]Department of Geriatric Medicine, Katholieke Universiteit Leuven, Leuven, Belgium, [6]Department of Medicine, Santiago de Compostela University, Complejo Hospitalario Universitario de Santiago (CHUS), Santiago de Compostela, Spain, [7]CIBER de Fisiopatología Obesidad y Nutricion (CB06/03), Instituto Salud Carlos III, Santiago de Compostela, Spain, [8]Reproductive Medicine Centre, Skåne University Hospital, University of Lund, Malmö, Sweden, [9]Department of Endocrinology, University College London, London, UK, [10]Department of Andrology and Reproductive Endocrinology, Medical University of Lodz, Lodz, Poland, [11]Laboratory of Molecular Endocrinology and Oncology, Laval University, Quebec City, Quebec, Canada, [12]Department of Human Nutrition, University of Glasgow, Scotland, UK, [13]School of Community Based Medicine, The University of Manchester, Salford Royal NHS Trust, Salford, UK, [14]Andrology Unit, United Laboratories of Tartu University Clinics, Tartu, Estonia, [15]Arthritis Research UK, Chesterfield, UK, [16]Department of Andrology and Endocrinology, Katholieke Universiteit Leuven, Leuven, Belgium and [17]Endocrinology Unit, Department of Clinical Physiopathology, University of Florence, Florence, Italy

(Correspondence should be addressed to I T Huhtaniemi; Email: ilpo.huhtaniemi@imperial.ac.uk; F C W Wu; Email: frederick.wu@manchester.ac.uk)

(The details of the EMAS Group are presented in the acknowledgement section)

Abstract

Background: The limitations of serum testosterone and estradiol (E_2) measurements using non-extraction platform immunoassays (IAs) are widely recognized. Switching to more specific mass spectrometry (MS)-based methods has been advocated, but directly comparative data on the two methods are scarce.

Methods: We compared serum testosterone and E_2 measurements in a large sample of middle-aged/elderly men using a common platform IA and a gas chromatography (GC)–MS method, in order to assess their limitations and advantages, and to diagnose male hypogonadism. Of subjects from the European Male Aging Study ($n=3174$; age 40–79 years), peripheral serum testosterone and E_2 were analyzed using established commercial platform IAs (Roche Diagnostics E170) and in-house GC–MS methods.

Results: Over a broad concentration range, serum testosterone concentration measured by IA and MS showed high correlation ($R=0.93$, $P<0.001$), which was less robust in the hypogonadal range (<11 nmol/l; $R=0.72$, $P<0.001$). The IA/MS correlation was weaker in E_2 measurements ($R=0.32$, $P<0.001$, at E_2 <40.8 pmol/l, and $R=0.74$, $P<0.001$, at E_2 >40.8 pmol/l). Using MS as the comparator method, IA ascertained low testosterone compatible with hypogonadism (<11 nmol/l), with 75% sensitivity and 96.3% specificity. The same parameters with IA for the detection of low E_2 (<40.7 pmol/l) were 13.3 and 99.3%, and for high E_2 (>120 pmol/l) 88.4 and 88.6%.

Conclusion: A validated platform IA is sufficient to detect subnormal testosterone concentrations in the diagnosis of male hypogonadism. The IA used for E_2 measurements showed poor correlation with MS and may only be suitable for the detection of high E_2 in men.

European Journal of Endocrinology **166** 983–991

Introduction

Testosterone and estradiol (E_2) are the two most important sex steroids in men and women respectively, and their accurate determination in serum is of crucial importance in assessing gonadal function both in clinical management and research. Immunoassay (IA) methods have been the mainstay of sex steroid measurements since their advent in the late 1960s. Most of the time they provide rapid and economical information about circulating hormone concentrations. However, the accuracy and precision of testosterone IAs, especially at the low concentrations found in children, women, and hypogonadal men, remain

DOI: 10.1530/EJE-11-1051
Online version via www.eje-online.org

a concern (1, 2, 3, 4). While the majority of IAs estimate high (adult male) concentrations sufficiently well, they usually overestimate low (female) concentrations (5), thus reducing the specificity and sensitivity of diagnosis of female hyperadrogenism and male hypogonadism. There is no consensus on whether IAs for testosterone are able to reliably discriminate between eugonadal and hypogonadal men (1, 5, 6). Professional societies and individual investigators have therefore emphasized the need for improved standardized methods, as well as traceability of the standards, to overcome these problems in sex steroid measurements (3, 4, 5, 6, 7, 8, 9).

Even during the years of IA dominance, mass spectrometry (MS) was regarded as the 'gold standard' of steroid analysis, but due to its technical complexity, cost, and suboptimal sensitivity, it has only recently reached the methodological refinement required for a routine clinical chemistry laboratory. The recent technical improvements in instrumentation and the wider availability due to falling costs of equipment have made MS a competitive method with IA, having reached sufficient sensitivity yet maintaining its superior specificity in steroid hormone measurements. Therefore, opinions are being expressed to promote MS as the standard method for steroid hormone measurements (6, 9, 10). However, MS remains more expensive and labour intensive (requiring solvent extraction), shows similar lack of between-laboratory standardisation to IA (11, 12), and is currently still not accessible to all practitioners. It is therefore important to investigate in which clinical situations MS is necessary for the measurement of testosterone and when it is still sufficient to rely on IAs. For instance, while it is clear that measurements of testosterone by IA in children and women are unreliable, it is uncertain whether MS or IA should be the method of choice for the quantitation of testosterone to diagnose adult male hypogonadism.

Measurement of E_2 is more challenging than that of testosterone due to its much lower circulating concentration (50–100-fold less in men). Although E_2 measurements are less often required for men, high concentrations are of diagnostic importance in gynecomastia and the rare cases of feminizing tumors and aromatase excess (13). Low E_2 concentrations are important in the assessment of osteoporosis (14, 15) and cardiovascular diseases (16, 17), where the replacement of IAs with more specific and sensitive MS measurements is expected to be useful.

The European Male Aging Study research consortium (18) has measured testosterone and E_2 concentrations in the serum samples of a large cohort ($n=3174$) of 40–79-year-old men using both an established IA and MS method. This provided a unique opportunity to compare the results obtained with these two methods and to assess the applicability of each technique for clinical diagnostics and research.

Subjects and methods

Subjects and study design

A total of 3369 community-dwelling men aged 40–79 (mean $\pm$ S.D.: 60 ± 11) years were recruited from population registers in eight European centers (Florence, Italy; Leuven, Belgium; Lodz, Poland; Malmö, Sweden; Manchester, UK; Santiago de Compostela, Spain; Szeged, Hungary; and Tartu, Estonia). Details of the research protocol have been published elsewhere (18). Ethical approval for the study was obtained in accordance with local institutional requirements in each center, and written informed consent was obtained from the study subjects.

Hormone measurements

A single fasting morning (before 1000 h) venous blood sample was obtained and separated serum was stored at -80 °C. Measurements of testosterone and E_2 were carried out by the Modular E170 platform electrochemiluminescence IAs (Roche Diagnostics) and gas chromatography–MS (19, 20, 21, 22). Within- and between-assay coefficients of variation (CV) in IA measurements were 1.05 and 3.72% for testosterone (at 14.4 nmol/l human serum), and 5.2 and 9.1% for E_2 (at 0.071 nmol/l human serum) respectively. The male reference range on IA for testosterone was 10.4–34.6 nmol/l and for E_2 <200 pmol/l. In MS measurements, the intra- and interassay CV were 2.9 and 3.4% for testosterone (at 1.7 nmol/l human serum), and 3.5 and 3.7% for E_2 (at 0.07 nmol/l human serum) respectively. The average recovery for steroids following extraction on MS was 102 ± 3%, and the male reference ranges were 14.1–39.0 nmol/l for testosterone and 23–112 pmol/l for E_2.

Sex hormone-binding globulin (SHBG) was measured by the Modular E170 platform electrochemiluminescence IAs (Roche Diagnostics). Free testosterone concentrations were derived from total testosterone, SHBG, and albumin concentrations (23).

Statistical analysis

From the total of 3369 participants, 150 were excluded because of prevalent pituitary or testicular diseases or current use of medications that could affect pituitary/testicular function (testosterone, dehydroepiandrosterone, antiandrogens, GnRH agonists, glucocorticoids, and psycholeptic agents) or interfere with sex steroid clearance or measurements (e.g. anticonvulsants). The reason for exclusion was their expected interference with the use of testosterone values in the diagnosis of late-onset hypogonadism (LOH; see Results). Of the remaining men, 3174 had complete data on testosterone and 3016 on E_2 with both IA and MS and were included in this analysis.

The analysis consisted of descriptive statistics to assess subject characteristics, where the data were presented as mean and S.D. for continuous variables and count (percentage for discrete variables). Distributions of total testosterone and total E_2 with both techniques were plotted via histogram. Spearman correlation measure was used to test the correlation for total testosterone and total E_2 between the two assays and within each assay. Agreement between the two assays was explored using the Bland–Altman plot (24) for limits of agreement, and bias estimation was used; this plots the % difference between MS and IA (i.e. $100\times$ (IA−MS)/MS) against the average of the two assays ((IA+MS)/2). Deviations from $\pm 20\%$ were used as the limits of bias.

Deming regression technique (25), which takes into account any measurement errors in the hormones, was used to additionally compare hormone concentrations between the two assays. Sensitivity and specificity of the IA measurement, using MS as the comparator method, were calculated to further explore the diagnostic accuracy of IA.

Results

Cohort characteristics

Characteristics of the analysis cohort of 3174 men are shown in Table 1. Mean (S.D.) age of the men was 59.7 (11.0) years. The recruitment was carried out from a random general population that was relatively healthy, as shown by a variety of characteristics. Of these, 21.4% were current smokers, and in 27.1% at least one co-morbid condition was reported, which included self-reported heart conditions, high blood pressure, bronchitis, asthma, peptic ulcer, epilepsy, diabetes, cancer, liver conditions, kidney conditions, prostate diseases, and thyroid disorders. The mean (S.D.) body mass index was 27.7 (4.1) kg/m^2 and the mean (S.D.) waist circumference was 98.4 (11.1) cm.

Mass spectrometry vs immunoassay

The mean (S.D.) testosterone concentrations were very similar between IA and MS: 16.5 (5.80) and 16.6 (5.95) nmol/l respectively. E_2 concentrations were, on average, higher with IA than MS: 92.9 (28.7) and 74.1 (25.1) pmol/l respectively. There was a good agreement in the distribution of results between the two assays for testosterone (Fig. 1a). In contrast, with E_2 (Fig. 1b), there were more samples with concentration below 70 pmol/l with MS than IA, and above this concentration there were more E_2 samples by IA than by MS.

Bland–Altman plot

For testosterone (Fig. 2a), there was little bias between the two methods at mean concentrations of the paired values (MS, IA) ranging from 0.175 to 46.21 nmol/l. The mean IA–MS difference (negative bias) was a low and nonsignificant −0.036 (95% confidence interval (95% CI), −0.113 to 0.040) nmol/l, with 95% limits of agreement of −4.36 to 4.29 nmol/l. Here, 9% of the testosterone concentrations by IA were more than 20% higher than those measured by MS, and 3% were over 20% lower. There was no significant trend in the relationship between the percentage bias and the average testosterone concentration of the two methods. Spearman correlation between the percentage bias and the average testosterone concentration of the two methods was 0.01 ($P=0.508$). This confirms that there was no concentration-dependent loss of agreement between the two methods of testosterone quantification.

For E_2 (Fig. 2b), there was a significant mean percentage difference (positive bias for IA) of 18.77 (95% CI, 18.11 to 19.43) pmol/l between the IA and MS measurements, with 95% limits of agreement of −18.5 to 56.1 pmol/l. The range of mean concentrations of the paired values (MS, IA) was from 17.01 to 254.65 pmol/l. The average discrepancy between the concentrations of the two methods (bias) was high; 58% of the E_2 concentrations by IA were over 20% higher

Table 1 Cohort characteristics ($n=3174$).

	Mean (S.D.)	**Count** (%)	**95% CI**	**(5–10–50–95–97.5)th Centiles**
Age (years)	59.67 (10.96)			
BMI (kg/m^2)	27.66 (4.10)			
Waist circumference (cm)	98.42 (11.07)			
Serum testosterone (nmol/l) MS	16.58 (5.95)		16.37, 16.78	(8.3–9.7–15.9–27.8–30.6)
Serum testosterone (nmol/l) IA	16.54 (5.80)		16.34, 16.74	(8.1–9.7–15.9–27.1–29.6)
Serum estradiol (pmol/l) MS	74.13 (25.09)		73.26, 75.01	(40.8–46.7–70.3–119.8–133.1)
Serum estradiol (pmol/l) IA	92.90 (28.73)		91.90, 93.90	(52.7–59.7–90.0–143.8–160.6)
Current smokers		673 (21.40)		
One morbidity		845 (27.07)		
Two or more morbidities		711 (22.75)		

MS, mass spectrometry; IA, immunoassay.

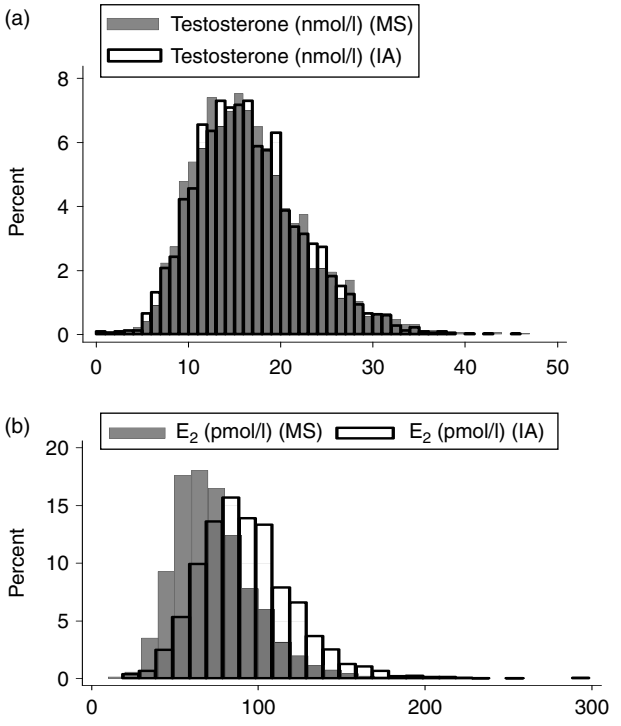

Figure 1 Distribution of testosterone (panel a), and E_2 (panel b) concentrations, as measured with MS and IA.

than by MS, and 3% E_2 concentrations by IA were more than 20% lower than by MS. Hence, IA grossly overestimated the E_2 levels. Figure 2b shows also a trend of the relationship between the bias (percentage difference) and the average concentration of the MS and IA methods. The positive bias increased as the average of E_2 decreased with a Spearman correlation of -0.07 ($P<0.001$), i.e. showing an inversely concentration-dependent positive bias for IA vs MS.

Deming regression

Figure 3 shows the scatter plots of testosterone (panel a) and E_2 (panel b) for the two methods, as well as the results from the Deming regression. The agreement between testosterone concentrations measured by IA and MS was close to the line of best fit ($y=x$, i.e. the line of equality), 0.97 (95% CI, 0.96 to 0.99), and the intercept was 0.41 nmol/l (95% CI, 0.18 to 0.65). For E_2, the agreement between the two techniques deviated considerably from the line of equality, with a slope of 1.19 (95% CI, 1.15 to 1.25) and an intercept of 4.28 pmol/l (95% CI, 0.75 to 7.81).

Table 2 presents the correlation coefficients between the MS and IA measurements of testosterone and E_2 at different concentrations of the hormones. With testosterone, the correlation coefficient was 0.93 in the entire cohort, 0.92 with testosterone concentrations >8 nmol/l, and 0.69 at testosterone concentrations <8 nmol/l. Using Deming regression, the agreement between testosterone concentrations measured by IA and MS with testosterone levels >8 nmol/l was close to the line of best fit, with a slope of 0.97 (95% CI, 0.95 to 0.99), and an intercept of 0.42 nmol/l (95% CI, 0.15 to 0.70). The agreement between testosterone concentrations with testosterone levels <8 nmol/l deviated more from the line of best fit, with a slope of 1.71 (95% CI, 1.04 to 2.37), and an intercept of -4.27 nmol/l (95% CI, -8.52 to -0.02). With E_2, the correlation coefficient between the MS and IA measurements was 0.76 in the entire cohort, and 0.74 at E_2 levels above 40.8 pmol/l, but only 0.32 at concentrations below 40.8 pmol/l. Using Deming regression, the agreement between E_2 concentrations measured by IA and MS with E_2 concentrations either above 40.8 pmol/l or below 40.8 pmol/l was divergent from the line of best fit, i.e. with E_2 >40.8 pmol/l, the slope was 1.24 (95% CI, 1.18 to 1.30) and the intercept was 0.04 pmol/l (95% CI, -4.13 to 4.21); with E_2 ≤ 40.8 pmol/l, the slope was 7.89 (95% CI, 3.17 to 12.6) and the intercept was -214 pmol/l (95% CI, -382 to -46.5). Significant, though less robust, correlations were also found between the testosterone and E_2 concentrations,

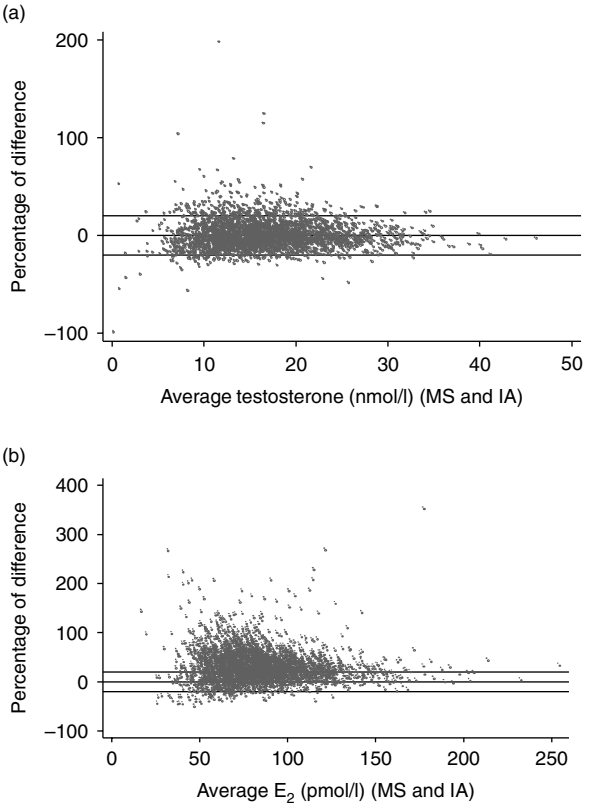

Figure 2 Bland–Altman plots of testosterone (panel a) and E_2 (panel b). Y-axis depicts the % difference between values of the two measurements ($100\times$(IA$-$MS)/MS). The horizontal lines are 0 and -20 and $+20$%. The mean difference between the two assays in the testosterone values was 0.77% and of the E_2 values, 30.1%.

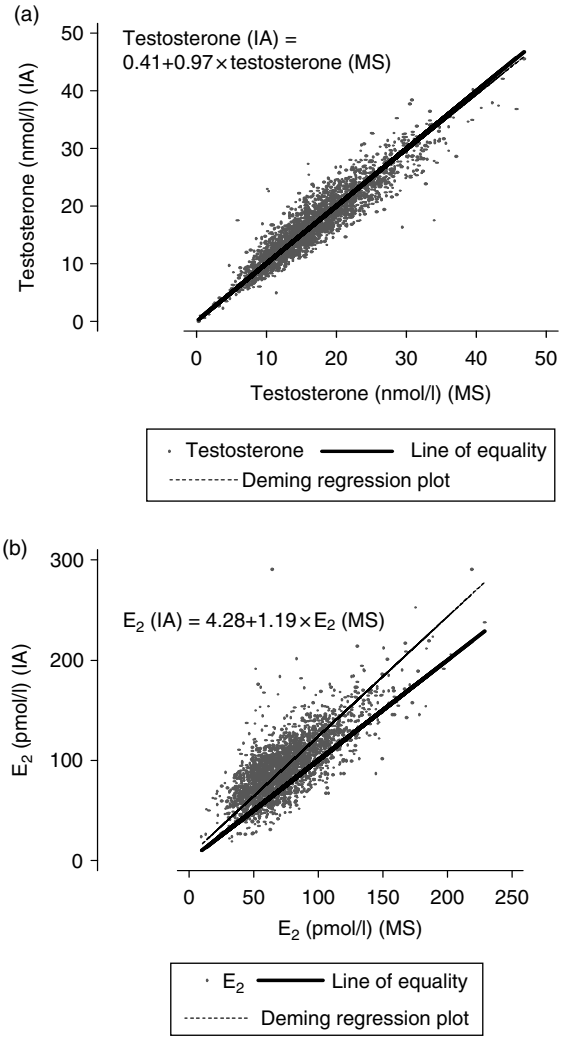

Figure 3 Scatter plots for testosterone (panel a) and E_2 (panel b) with IA and MS, and the Deming regression results.

which were weaker or nonsignificant in the IA/IA and IA/MS comparisons.

Using MS as the comparator method, we then assessed the sensitivity (% of true positives) and specificity (% of true negatives) of IA to detect low testosterone concentrations by MS at defined thresholds and to identify patients fulfilling the diagnostic criteria of LOH, i.e. low testosterone in combination with three sexual symptoms (reduced morning erections and sexual thoughts, and erectile dysfunction; Table 3) (26). The sensitivity and specificity of IA to detect total testosterone <11 nmol/l were 75.0 and 96.3% respectively. To detect total testosterone <8 nmol/l, the sensitivity and specificity of IA increased slightly to 76.9 and 98.3%. If the presence of symptoms of androgen deficiency (three sexual symptoms) in addition to a total testosterone <11 nmol/l and calculated free testosterone <220 pmol/l were used as the criteria, the sensitivity and specificity of identifying LOH with the IA testosterone measurements increased to 85.5 and 99.4%. If the threshold levels of testosterone were decreased to <8 nmol/l (together with the three sexual symptoms), the respective parameters increased even further to 92.3 and 99.8%.

In the E_2 assay, the ability of IA to detect concentrations below 40.8 pmol/l (the lowest 5th centile for E_2 by MS) had a sensitivity of only 13.3%, with a specificity of 99.3% (Table 3). The IA performance to detect a concentration below 61.2 pmol/l (the lowest tertile for E_2 by MS) had a slightly better sensitivity of 25.6% and a specificity of 96.1%. In contrast, the sensitivity (88.6%) and specificity (88.4%) of E_2 IA were clearly better to detect high E_2 concentrations (>119.8 pmol/l; the highest 5th centile for E_2 by MS). Hence, IA performed especially poorly at low E_2 concentrations and grossly overestimated them, as also seen in Figs 2b and 3b.

Discussion

Our study provides thus far the largest comparative data on testosterone and E_2 measurements by IA and MS in serum samples of over 3000 men. Using MS as the comparator method for testosterone and E_2 measurements, we can conclude that testosterone measurements by IA offer good accuracy at all concentrations found in eugonadal as well as hypogonadal men. In contrast, IA provides acceptable estimates of E_2 only at the higher concentrations detected. Importantly, our data do not confirm that our platform IA for testosterone lack sensitivity and specificity in the hypogonadal range. The correlation of testosterone values between IA and MS measurements was high in the entire assay cohort, 0.93, and when testosterone level by MS was >8 nmol/l, 0.92. However, when testosterone concentrations were <8 nmol/l, the correlation was clearly lower, 0.69, indicating poorer accuracy of one or both of the methods. As compared with MS, the sensitivity of testosterone IA to detect at low testosterone concentrations (either <11 or <8 nmol/l) was 75–77%, and the specificity to detect a normal testosterone concentration was 96–98%.

The above figures are probably underestimates because the comparator MS method is not free of variability either (11, 12). Hence, the performance of both assays may contribute to the degradation of correlation. IAs have the known problems with antibody specificity, matrix effects, and lack of linearity and functional sensitivity. The various MS methods are not identical, use diverse procedures, calibrations and technologies, and are not totally free of influence of interfering substances. Because we omitted 150 samples from men with pituitary or testicular diseases and their treatments from the analysis, our measurements do not

Table 2 Correlation coefficients between testosterone and E_2 measurements by MS and IA.

	Testosterone MS	Testosterone IA	E_2 MS	E_2 IA
(A) Testosterone and E_2; the entire analysis sample				
Testosterone MS	1	0.93***	0.49***	0.32***
Testosterone IA		1	0.47***	0.37***
E_2 MS			1	0.76***
E_2 IA				1
(B) Testosterone (<8 nmol/l)				
Testosterone MS	1	0.69***	0.39***	0.19*
Testosterone IA		1	0.21*	0.16
E_2 MS			1	0.68***
E_2 IA				1
(C) Testosterone (>8 nmol/l)				
Testosterone MS	1	0.92***	0.46***	0.31***
Testosterone IA		1	0.44***	0.37***
E_2 MS			1	0.76***
E_2 IA				1
(D) E_2 (<40.8 pmol/l; lowest 5th centile)				
Testosterone MS	1	0.91***	0.41***	0.01 ($P=0.87$)
Testosterone IA		1	0.30***	0.08 ($P=0.32$)
E_2 MS			1	0.32***
E_2 IA				1
(E) E_2 (>40.8 pmol/l)				
Testosterone MS	1	0.93***	0.45***	0.28***
Testosterone IA		1	0.43***	0.34***
E_2 MS			1	0.74***
E_2 IA				1

*$P<0.05$, ***$P<0.001$.

take into account all potential interferences in the clinical samples.

Interestingly, in the context of clinical management, when the sexual symptoms were combined with low testosterone concentrations to diagnose symptomatic LOH, the sensitivity of detection by IA increased from 75 to 85.5% with testosterone <11 nmol/l (+free testosterone <220 pmol/l) and further to 92.3% with testosterone <8 nmol/l, probably by eliminating the impact of some functionally irrelevant borderline or erroneous testosterone concentrations (between 8 and 11 nmol/l). This emphasizes the importance of combining testosterone concentration and symptoms in the diagnosis of LOH. We can thus conclude that the IA used in our study is sufficiently sensitive and specific to discriminate between normal and low testosterone concentrations in men suspected to have LOH. However, it has to be emphasized that the testosterone IA we used was of good quality, having passed with acceptable accuracy a rigorous standardization procedure (http://www.cdc.gov/labstandards/hs.html). All IAs used in clinical testosterone measurements are unlikely to have the same high quality.

A similar assessment of the E_2 measurements did not reveal as good correlations as with testosterone. In the entire cohort, the IA/MS correlation was 0.76, and it was 0.32 with E_2 concentrations <40.8 pmol/l and 0.74 at E_2 levels >40.8 pmol/l. It is expected that the assay performance for E_2 is worse at molar levels that are, on average, 0.4% of those of testosterone. In particular, the sensitivity of IA to detect low E_2 concentrations was poor, at 13–25%. Accordingly, IA grossly overestimated the low E_2 values. This seriously hampers the usefulness of the IA data on E_2 at low concentrations. However, the sensitivity and specificity of IA to detect E_2 concentrations in the highest 5th centile (>120 pmol/l) were acceptable (88.6 and 88.4% respectively).

Testosterone is still considered the standard assessment tool in the diagnostic approach of men with low bone density. However, with serum E_2 concentrations being more closely associated with BMD than those of testosterone (14, 15) in men, and with MS-based assays allowing more accurate and sensitive measurements at low concentrations of E_2, their measurement is becoming increasingly useful. When comparing the clinical applicability of E_2 data in studies of BMD, Khosla *et al.* (27) concluded that although the MS data provide more accurate measurements in men, the applicability of the E_2 IA data for bone data is generally valid. Hence, the necessity of switching E_2 measurements from IA to MS is somewhat relative, admitting that the latter technique yields more accurate, but not necessarily clinically more useful results. Our data, however, show that serum testosterone and E_2 concentrations are not highly correlated. A case can be made to develop clinical algorithms incorporating accurate measurement of E_2 as part of the evaluation of osteoporosis in men. Moreover, recent epidemiologic studies in men and women have demonstrated associations between low sex hormone concentrations (including E_2) and the risk of cardiovascular disease in both sexes (16, 17, 28, 29),

Table 3 Sensitivity and specificity of IA in detection of low testosterone levels (panels A and B), diagnosing LOH (panels C and D), and detection of low and high E_2 (panels E, F and G), using MS as the reference method.

	MS								IA	
	Low testosterone			Oh by MS			E_2 (pmol/l)			
	Yes	No	Total	Yes	No	Total	≤	>	Sensitivity (%)	Specificity (%)
(A) Detection of low testosterone[a]										
Low testosterone by IA									75.0	96.3
Yes	405	96	501							
No	135	2538	2673							
Total	540	2634	3174							
(B) Detection of low testosterone[b]										
Low testosterone by IA									75.4	98.5
Yes	98	47	145							
No	32	2997	3029							
Total	130	3044	3174							
(C) Diagnosis of LOH[c,h]										
LOH by IA									85.5	99.4
Yes				53	17	70				
No				9	2884	2893				
Total				62	2901	2963				
(D) Diagnosis of LOH[d,h]										
LOH by IA									88.5	99.8
Yes				23	6	29				
No				3	2931	2934				
Total				26	2937	2963				
(E) Detetion of low E_2 <40.8 pmol/l[e]										
E_2 by IA									13.2	99.3
≤40.8							21	20		
>40.8							138	2995		
Total							159	3015		
(F) Detection of low E_2 <61.2 pmol/l[f]										
E_2 by IA									25.6	96.1
≤61.2							271	82		
>61.2							787	2034		
Total							1058	2116		
(G) Detection of high E_2 >119.83 pmol/l[g]										
E_2 by IA									88.6	88.4
<119.83							2667	18		
>119.83							349	140		
Total							3016	158		

[a]Total testosterone <11 nmol/l.
[b]Total testosterone <8 nmol/l.
[c]Total testosterone <11 nmol+free testosterone <220 pmol/+3 sexual symptoms.
[d]Total testosterone <8 nmol/+3 sexual symptoms.
[e]The lowest 5th centile for E_2 by MS.
[f]The lowest tertile for E_2 by MS.
[g]The highest 5th centile for E_2 by MS.
[h]Total number of cases smaller because data on sexual symptoms were not available from 211 men.

suggesting another indication where more reliable methods for E_2 measurement should be used. It should be acknowledged that in everyday clinical practice, the higher concentrations of E_2 in men that may occur in gynecomastia and the rare cases of feminising tumors and aromatase excess can be discriminated with sufficient accuracy by IA.

Although MS, in general, is more specific and has lower intra- and interassay variability than IA, it faces similar inter-laboratory variability issues as IA (11, 12). All MS methods are not equal; like IA, they represent a heterogeneous group of measurements with significant differences in performance. One study comparing several established MS methods for the determination of testosterone in serum found overall CV of up to 33% at low concentrations, up to 15% at >1.5 nmol/l, and 1.4–11.4% at concentrations >3.5 nmol/l (11). Nevertheless, the variability in testosterone results with MS methods in most comparisons is substantially smaller than those reported for platform IAs (1, 5, 30, 31). Whether this difference translates into improved clinical relevance requires additional data and experience. A very recent study comparing total testosterone assays in women concluded that the results obtained by IA and MS were comparable, and there is significant variability and poor precision also between various MS methods at low

levels (12). Hence, switching from IA to MS is not a guaranteed solution to improve the quality of sex steroid measurements at low concentrations. Improvements in performance and standardization in platform IAs are feasible alternatives that are already being implemented by some manufacturers. It is a major investment to abandon IA technology in favor of MS, and the reasons for this must be tangible and supported by evidence rather than conjecture. Our results suggest that, at least in men, IA can be as good as MS in the clinically important discrimination between eugonadal and hypogonadal men, especially when combined with clinical signs of androgen deficiency. The variability and imprecision of E_2 measurements by MS is smaller than by IA, and it is clear that MS is superior to IA in the measurement of this hormone, especially at low concentrations. It seems prudent to conclude that the selection of an assay should be driven by the measurement performance in light of the clinical need and not by assay technology.

In conclusion, the comparison of measurements of serum testosterone and E_2 in the largest cohort so far of adult male samples indicates that clinically relevant results on serum testosterone for the diagnosis of hypogonadism can be obtained both with well-validated IA and MS assays. Our findings do not support a mandatory requirement, on either analytical or clinical grounds, to switch from good-quality IAs to MS in the measurements of testosterone in male subjects. In contrast, clinicians should be aware of the unreliability of apparently low E_2 results in men obtained by IA. Finally, assay performance is more important than assay technology.

Declaration of interest

I T Huhtaniemi consulted for Ferring Pharmaceuticals, Denmark. F C W Wu consulted for Bayer-Schering Healthcare, Germany; Akzo-Nobel (Organon), The Netherlands; Ferring Pharmaceuticals, Denmark; Pierre-Fabre Medicaments, France; Ardana Biosciences, UK; Procter & Gamble, United States; and Lilly-ICOS, United States, and has also received research grant funding support from Bayer-Schering Healthcare, Germany; Bayer Schering; Lilly-ICOS; and other companies. All other authors have nothing to declare.

Funding

This work was supported by the Commission of the European Communities Fifth Framework Programme, 'Quality of Life and Management of Living Resources' grant QLK6-CT-2001-00258. Dr S Boonen is a senior clinical investigator of the Fund for Scientific Research (Flanders, Belgium; F.W.O. – Vlaanderen) and holder of the Novartis Leuven University Chair in Gerontology and Geriatrics. Dr Vanderschueren is a senior clinical investigator supported by the Clinical Research Fund of the University Hospitals Leuven, Belgium.

Acknowledgements

The authors wish to thank the men who participated in the eight countries, the research/nursing staff in the eight centers: C Pott (Manchester), E Wouters (Leuven), M Nilsson (Malmö), M del Mar Fernandez (Santiago de Compostela), M Jedrzejowska (Lodz), H-M Tabo (Tartu) and A Heredi (Szeged) for their meticulous data collection, and C Moseley (Manchester) for data entry and project co-ordination. Non-financial support from Arthritis Research UK and the NIHR Manchester Biomedical Research Centre is acknowledged. The EMAS Group: Florence (G Forti, Luisa Petrone, Giovanni Corona); Leuven (D Vanderschueren, S Boonen, Herman Borghs); Lodz (K Kula, Jolanta Slowikowska-Hilczer, Renata Walczak-Jedrzejowska); London (I T Huhtaniemi); Malmö (A Giwercman); Manchester (F C W Wu, A J Silman, T W O'Neill, J D Finn, Philip Steer, A Tajar, D M Lee, Stephen Pye); Santiago (F F Casanueva, Mary Lage, Ana I Castro); Szeged (G Bartfai, Imre Földesi, Imre Fejes); Tartu (M Punab, Paul Korrovitz); Turku (Min Jiang).

References

1 Wang C, Catlin DH, Demers LM, Starcevic B & Swerdloff RS. Measurement of total serum testosterone in adult men: comparison of current laboratory methods versus liquid chromatography–tandem mass spectrometry. *Journal of Clinical Endocrinology and Metabolism* 2004 **89** 534–543. (doi:10.1210/jc.2003-031287)

2 Moal V, Mathieu E, Reynier P, Malthiery Y & Gallois Y. Low serum testosterone assayed by liquid chromatography–tandem mass spectrometry. Comparison with five immunoassay techniques. *Clinica Chimica Acta* 2007 **386** 12–19. (doi:10.1016/j.cca.2007.07.013)

3 Rosner W & Vesper H. Toward excellence in testosterone testing: a consensus statement. *Journal of Clinical Endocrinology and Metabolism* 2010 **95** 4542–4548. (doi:10.1210/jc.2010-1314)

4 Vesper HW & Botelho JC. Standardization of testosterone measurements in humans. *Journal of Steroid Biochemistry and Molecular Biology* 2010 **121** 513–519. (doi:10.1016/j.jsbmb.2010.03.032)

5 Taieb J, Mathian B, Millot F, Patricot MC, Mathieu E, Queyrel N, Lacroix I, Somma-Delpero C & Boudou P. Testosterone measured by 10 immunoassays and by isotope-dilution gas chromatography–mass spectrometry in sera from 116 men, women, and children. *Clinical Chemistry* 2003 **49** 1381–1395. (doi:10.1373/49.8.1381)

6 Sikaris K, McLachlan RI, Kazlauskas R, de Kretser D, Holden CA & Handelsman DJ. Reproductive hormone reference intervals for healthy fertile young men: evaluation of automated platform assays. *Journal of Clinical Endocrinology and Metabolism* 2005 **90** 5928–5936. (doi:10.1210/jc.2005-0962)

7 Blair IA. Analysis of estrogens in serum and plasma from postmenopausal women: past present, and future. *Steroids* 2010 **75** 297–306. (doi:10.1016/j.steroids.2010.01.012)

8 Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder PJ, Swerdloff RS & Montori VM. Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. *Journal of Clinical Endocrinology and Metabolism* 2010 **95** 2536–2559. (doi:10.1210/jc.2009-2354)

9 Wartofsky L & Handelsman DJ. Standardization of hormonal assays for the 21st century. *Journal of Clinical Endocrinology and Metabolism* 2010 **95** 5141–5143. (doi:10.1210/jc.2010-2369)

10 Bhasin S, Zhang A, Coviello A, Jasuja R, Ulloor J, Singh R, Vesper H & Vasan RS. The impact of assay quality and reference ranges on clinical decision making in the diagnosis of androgen disorders. *Steroids* 2008 **73** 1311–1317. (doi:10.1016/j.steroids.2008.07.003)

11 Vesper HW, Bhasin S, Wang C, Tai SS, Dodge LA, Singh RJ, Nelson J, Ohorodnik S, Clarke NJ, Salameh WA, Parker CR Jr, Razdan R, Monsell EA & Myers GL. Interlaboratory comparison study of serum total testosterone (corrected) measurements performed by mass spectrometry methods. *Steroids* 2009 **74** 498–503. (doi:10.1016/j.steroids.2009.01.004)

12 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, McGovern PG, Cataldo NA, Gosman GG, Nestler JE, Myers ER, Santoro N, Eisenberg E, Zhang M & Zhang H. Total testosterone assays in

women with polycystic ovary syndrome: precision and correlation with hirsutism. *Journal of Clinical Endocrinology and Metabolism* 2010 **95** 5305–5313. (doi:10.1210/jc.2010-1123)
13 Braunstein GD. Aromatase and gynecomastia. *Endocrine-Related Cancer* 1999 **6** 315–324. (doi:10.1677/erc.0.0060315)
14 Mellstrom D, Vandenput L, Mallmin H, Holmberg AH, Lorentzon M, Oden A, Johansson H, Orwoll ES, Labrie F, Karlsson MK, Ljunggren O & Ohlsson C. Older men with low serum estradiol and high serum SHBG have an increased risk of fractures. *Journal of Bone and Mineral Research* 2008 **23** 1552–1560. (doi:10.1359/jbmr.080518)
15 Gennari L, Khosla S & Bilezikian JP. Estrogen and fracture risk in men. *Journal of Bone and Mineral Research* 2008 **23** 1548–1551. (doi:10.1359/jbmr.0810c)
16 Callou de Sa EQ, de Sa FC, Silva RD, de Oliveira KC, Guedes AD, Feres F & Verreschi IT. Endogenous oestradiol but not testosterone is related to coronary artery disease in men sex hormones and coronary artery disease. *Clinical Endocrinology* 2011 **75** 177–183. (doi:10.1111/j.1365-2265.2011.04017.x)
17 Jeon GH, Kim SH, Yun SC, Chae HD, Kim CH & Kang BM. Association between serum estradiol level and coronary artery calcification in postmenopausal women. *Menopause* 2010 **17** 902–907. (doi:10.1097/gme.0b013e3181d76768)
18 Lee DM, O'Neill TW, Pye SR, Silman AJ, Finn JD, Pendleton N, Tajar A, Bartfai G, Casanueva F, Forti G, Giwercman A, Huhtaniemi IT, Kula K, Punab M, Boonen S, Vanderschueren D & Wu FC. The European Male Ageing Study (EMAS): design, methods and recruitment. *International Journal of Andrology* 2009 **32** 11–24. (doi:10.1111/j.1365-2605.2008.00879.x)
19 Labrie F, Bélanger A, Bélanger P, Bérubé R, Martel C, Cusan L, Gomez J, Candas B, Castiel I, Chaussade V, Deloche C & Leclaire J. Androgen glucuronides, instead of testosterone, as the new markers of androgenic activity in women. *Journal of Steroid Biochemistry and Molecular Biology* 2006 **99** 182–188. (doi:10.1016/j.jsbmb.2006.02.004)
20 Labrie F, Bélanger A, Bélager P, Bérubé R, Martel C, Cusan L, Gomez J, Candas B, Chaussade V, Castiel I, Deloche C & Leclaire J. Metabolism of DHEA in postmenopausal women following percutaneous administration. *Journal of Steroid Biochemistry and Molecular Biology* 2007 **103** 178–188. (doi:10.1016/j.jsbmb.2006.09.034)
21 Swanson C, Lorentzon M, Vanderput L, Labrie F, Rane A, Jakobsson J, Chouinard S, Bélanger A & Ohlsson C. Sex steroid levels and cortical bone size in young men are associated with uridine diphosphate glucuronyltransferase 2B7 polymorphism ($H^{268}Y$). *Journal of Clinical Endocrinology and Metabolism* 2007 **92** 3697–3704. (doi:10.1210/jc.2007-0359)
22 Bhasin S, Pencina M, Jasuja GK, Travison TG, Coviello A, Orwoll E, Wang PY, Nielson C, Wu F, Tajar A, Labrie F, Vesper H, Zhang A, Ulloor J, Singh R, D'Agostino R & Vasan RS. Reference ranges for testosterone in men generated using liquid chromatography tandem mass spectrometry in a community-based sample of healthy nonobese young men in the Framingham Heart Study and applied to three geographically distinct cohorts. *Journal of Clinical Endocrinology and Metabolism* 2011 **96** 2430–2439. (doi:10.1210/jc.2010-3012)
23 Vermeulen A, Verdonck L & Kaufman JM. A critical evaluation of simple methods for the estimation of free testosterone in serum. *Journal of Clinical Endocrinology and Metabolism* 1999 **84** 3666–3672. (doi:10.1210/jc.84.10.3666)
24 Bland JM & Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986 **1** 307–310. (doi:10.1016/S0140-6736(86)90837-8)
25 Kelly G. The influence function in the error in variables problem. *Annals of Statistics* 1984 **12** 87–100. (doi:10.1214/aos/1176346394)
26 Wu FC, Tajar A, Beynon JM, Pye SR, Silman AJ, Finn JD, O'Neill TW, Bartfai G, Casanueva FF, Forti G, Giwercman A, Han TS, Kula K, Lean ME, Pendleton N, Punab M, Boonen S, Vanderschueren D, Labrie F & Huhtaniemi IT. Identification of late-onset hypogonadism in middle-aged and elderly men. *New England Journal of Medicine* 2010 **363** 123–135. (doi:10.1056/NEJMoa0911101)
27 Khosla S, Amin S, Singh RJ, Atkinson EJ, Melton LJ III & Riggs BL. Comparison of sex steroid measurements in men by immunoassay versus mass spectroscopy and relationships with cortical and trabecular volumetric bone mineral density. *Osteoporosis International* 2008 **19** 1465–1471. (doi:10.1007/s00198-008-0591-5)
28 Chen Y, Zeleniuch-Jacquotte A, Arslan AA, Wojcik O, Toniolo P, Shore RE, Levitz M & Koenig KL. Endogenous hormones and coronary heart disease in postmenopausal women. *Atherosclerosis* 2011 **216** 414–419. (doi:10.1016/j.atherosclerosis.2011.01.053)
29 Ahmed B, Bairey Merz CN, Johnson BD, Bittner V, Berga SL, Braunstein GD, Hodgson TK, Smith K, Shaw L, Kelsey SF & Sopko G. Diabetes mellitus, hypothalamic hypoestrogenemia, and coronary artery disease in premenopausal women (from the National Heart, Lung, and Blood Institute sponsored WISE study). *American Journal of Cardiology* 2008 **102** 150–154. (doi:10.1016/j.amjcard.2008.03.029)
30 Steinberger E, Ayala C, Hsi B, Smith KD, Rodriguez-Rigau LJ, Weidman ER & Reimondo GG. Utilization of commercial laboratory results in management of hyperandrogenism in women. *Endocrine Practice* 1998 **4** 1–10.
31 Hsing AW, Stanczyk FZ, Belanger A, Schroeder P, Chang L, Falk RT & Fears TR. Reproducibility of serum sex steroid assays in men by RIA and mass spectrometry. *Cancer Epidemiology, Biomarkers & Prevention* 2007 **16** 1004–1008. (doi:10.1158/1055-9965.EPI-06-0792)

Received 6 December 2011
Revised version received 22 February 2012
Accepted 13 March 2012