Clinical Chemistry 51:8
1472–1479 (2005)

Endocrinology and Metabolism

Testosterone Measurement by Isotope-Dilution Liquid Chromatography–Tandem Mass Spectrometry: Validation of a Method for Routine Clinical Practice

Marion L. Cawood,[1*] Helen P. Field,[1] Clive G. Ford,[1] Scott Gillingwater,[2] Andrew Kicman,[3] David Cowan,[3] and Julian H. Barth[1]

Background: **Immunoassay is unsatisfactory for measuring the testosterone concentrations typically found in women. Bench-top tandem mass spectrometers are a viable alternative technology for measurements in the clinical laboratory.**

Methods: **We used stable-isotope dilution liquid chromatography–tandem mass spectrometry (ID/LC-MS/MS) to measure testosterone in plasma and serum. The sample volume was 50 μL in duplicate; preparation and analysis were carried out in a single tube, and a batch of 192 tubes was analyzed in 17.5 h.**

Results: **Intra- and interassay imprecision was <15% in the range 0.3–49 nmol/L. Recovery of testosterone added to samples at concentrations of 0.625–20 nmol/L was 96% (CV = 12%; n = 26). Six samples were serially diluted with double charcoal-stripped serum to demonstrate linearity. Correlation (r^2) with isotope-dilution gas chromatography–mass spectrometry for 20 pools of clinical samples (range, 0.5–38.5 nmol/L) was 0.99. Correlations with our extraction RIA were 0.97 for clinical samples from men (range, 8–46.3 nmol/L) and 0.66 for samples from women (range, 0.7–3.0 nmol/L), but were 0.35 for male samples containing <3 nmol/L testosterone and 0.77 for female samples containing >8 nmol/L. Various steroids added to double charcoal-stripped serum showed no interference at the retention time of the testosterone peak.**

Conclusions: **The ID/LC-MS/MS method has improved accuracy compared with immunoassay. The low sample volume and simplicity, rapidity, and robustness of the method make it suitable for use as a high-throughput assay in routine clinical biochemistry laboratories.**

Measurement of testosterone in serum/plasma is essential in the investigation of androgenic status and monitoring of stimulatory, suppressive, or replacement therapy in children and adults of both sexes. In the late 1960s and 1970s, RIA methods became available for this purpose. Early methods incorporated solvent extraction and chromatographic purification before measurement by competitive RIA. These methods were well validated but because of their complexity were confined to specialist units.

To address the rapidly increasing number of requests, high-throughput assays became essential and were first achieved through modification of existing methods by discarding the chromatographic separation and eventually the solvent extraction step. Commercial nonextraction, nonisotopic methods became widely available as manual techniques in the 1980s and 1990s and subsequently, with the further development of excess reagent nonequilibrium immunoassays, as fully automated immunoassay platforms,. This methodology has permitted the introduction of many complex analytes, including testosterone, to the repertoire of routine biochemistry laboratory tests, particularly because sample volume requirements are low and throughput high. However, the omission of preimmunoassay purification steps has led to problems of anomalous high or low results *(1–3)*. These may be attributable to interference by steroid or non-

[1] SAS Centre for Steroid Hormones (Leeds), Department of Clinical Biochemistry & Immunology, Leeds General Infirmary, Leeds, United Kingdom.
[2] Waters Corporation, Manchester, United Kingdom.
[3] Drug Control Centre, King's College, London, United Kingdom.
*Address correspondence to this author at: SAS Centre for Steroid Hormones (Leeds), Department of Clinical Biochemistry & Immunology, Leeds General Infirmary, Great George Street, Leeds LS1 3EX, UK. E-mail marion.cawood@leedsth.nhs.uk.

Received October 28, 2004; accepted May 9, 2005.
Previously published online at DOI: 10.1373/clinchem.2004.044503

steroid molecules, lack of antibody specificity, or interference by antibodies within the patient's own serum *(4)*.

During the years of increasing use of immunoassays in routine clinical chemistry, reference methods have been available for comparison. The reference method for steroids is isotope-dilution gas chromatography–mass spectrometry (ID/GC-MS)[4] *(5, 6)*. This technique is lengthy and requires relatively large volumes of serum. Moreover, because derivative formation and purification are necessary before measurements, the method is carried out only in specialist units. Recently, Starcevic et al. *(7)* described a method for the measurement of testosterone by liquid chromatography–tandem mass spectrometry (LC-MS/MS), which used a minimum of 2 mL of serum. We have further enhanced this technique to measure testosterone over the full range of concentrations required in clinical practice. Our technique requires only 50 μL of sample and can perform 11 analyses per hour. We demonstrated the accuracy of the method by comparing it with 2 ID/GC-MS methods and also assessed the imprecision, linearity, recovery, sensitivity, and specificity.

In a recent, extensive study of commercial testosterone immunoassays in routine use, Taieb et al. *(8)* demonstrated that the majority of these assays significantly overestimate testosterone concentrations in women but, in some cases, slightly underestimate testosterone in men. We therefore compared the results obtained by the ID/LC-MS/MS method for clinical samples, 110 from women and 111 from men, with those obtained by our in-house, solvent-extraction RIA.

Materials and Methods

REAGENTS AND EQUIPMENT

We purchased testosterone, epitestosterone, 5α-dihydrotestosterone, dihydroepiandrosterone, androstenedione, estrone, androsterone, 19-hydroxytestosterone, ethisterone, cyproterone acetate, ammonium acetate (SigmaUltra), and zinc sulfate heptahydrate from Sigma-Aldrich; HPLC gradient-grade methanol and water from Fisher Scientific; ACS reagent-grade ethanol from Hayman Ltd.; formic acid (Aristar) from BDH Laboratory Supplies; diethyl ether from Vickers Laboratories Ltd.; dideuterated testosterone {[1,2-^{2}H]testosterone (D_2T); chemical purity by HPLC >99%; mass purity, 99.7%; product code D-2962} from Qm_x Laboratories Ltd.; and double charcoal-stripped serum from SCIPAC Ltd. All reagents were used as supplied without further purification.

Eppendorf tubes (1.5-mL) were obtained from Sarstedt Ltd., the ABgene heat sealer and heat-sealing film were from VWR International Ltd., and autosampler racks were custom made by Springlab Ltd.

HPLC was carried out using XTerra 2.5 μm C_{18} columns [50 × 4.6 mm (i.d.)] obtained from Waters Ltd. The LC-MS/MS system consisted of a Waters 2795 sample preparation system connected to a Micromass Quattro *micro* API triple quadrupole mass spectrometer, both obtained from Waters Corporation. MassLynx and QuanLynx software (Ver. 4.0) were used. The settings on the instrument were optimized for maximum ion yield: capillary voltage was 0.8 kV, cone voltage 25 V, collision energy 20 eV; source temperature 140 °C, and desolvation temperature 450 °C. At these settings, testosterone and D_2T gave protonated molecular ions $[M+H]^+$ at *m/z* 289.1 and 291.1. The respective fragment ions used were *m/z* 96.7 and 98.7.

Quality-control samples at 4 testosterone concentrations were prepared in house from pooled clinical plasma/serum samples. Lyphochek quality-control material, routinely used in our immunoassays, was unsuitable for LC-MS/MS because of integration interference from additional peaks in the internal standard transition.

CALIBRATORS AND REAGENTS

Calibrators were prepared by weighing out testosterone (stated purity >99%, as measured by thin-layer chromatography) on a 6-place balance and dissolving in AnalR-grade ethanol to give a stock solution of 500 μmol/L. This solution was diluted in HPLC-grade water to give a series of solutions that were added to double charcoal-stripped serum to give calibrators at concentrations from 0.25 to 100 nmol/L. A zero calibrator was similarly prepared by the addition of HPLC-grade water to double charcoal-stripped serum. Aliquots of 300 μL were stored in Eppendorf tubes at −20 °C. D_2T was weighed out, dissolved, and serially diluted in HPLC-grade methanol to give a final internal standard working concentration of 2 nmol/L for the precipitation method of sample preparation. A second working dilution of 10 nmol/L in HPLC-grade water was prepared for the diethyl ether extraction method. Stock solutions were stored at −20 °C, and the working dilutions were stored at room temperature. Precipitating reagent was prepared by mixing 0.3 mol/L zinc sulfate in HPLC-grade water with internal standard solution (2 nmol/L) in proportions of 1:5 by volume and was stored at room temperature for up to 1 week.

SAMPLE PREPARATION

Zinc sulfate/methanol precipitation. Serum/plasma samples, calibrators, or quality-control material (50 μL) were pipetted in duplicate into Eppendorf tubes (1.5-mL). Precipitating reagent (100 μL) containing D_2T (0.167 pmol) was added, and the tubes were stoppered and thoroughly vortex-mixed on a mechanical mixer for 1 min. The tubes were then centrifuged at 15 000*g* for 10 min, after which the stoppers were removed and the tubes were loaded

[4] Nonstandard abbreviations: ID/GC-MS, isotope-dilution gas chromatography–mass spectrometry; LS-MS/MS, liquid chromatography–tandem mass spectrometry; D_2T, dideuterated testosterone; UKNEQAS, United Kingdom National External Quality Assessment Scheme; and CI, confidence interval.

into the autosampler racks and sealed with heat-sealing film. The racks were placed in the autosampler chamber and set at 10 °C.

Diethyl ether extraction. Serum/plasma samples, calibrators, or quality-control materials (50 μL) were pipetted in duplicate into disposable soda glass tubes (75 × 12 mm). Internal standard (20 μL of 10 nmol/L D_2T in HPLC-grade water) was added to each tube, and the contents were mixed. After allowing this mixture to sit for 30 min at room temperature, we extracted the testosterone into diethyl ether (2 mL) by vortex-mixing for 1 min. The aqueous layer was frozen in a dry ice–methanol bath, and the organic layer was decanted into clean disposable soda glass tubes (75 × 10 mm). After the solvent was removed under reduced pressure, aqueous methanol (150 μL of a 1:1 solution, by volume) was added to the dried residue and the tubes were vortex-mixed, left 15 min at room temperature, and vortex-mixed again before being transferred to Eppendorf tubes. The tubes were sealed with heat-sealing film as described above.

ID/LC-MS/MS. Supernatant or reconstituted ether extract (40 μL) was injected on the column. The column temperature was held at 40 °C. The mobile phase was a mixture of HPLC-grade water (solvent A) and HPLC-grade methanol (solvent B), each containing ammonium acetate (2 mmol/L) and formic acid (1 mL/L). The initial solvent mixture used was 35% A-65% B for 3.06 min at 600 μL/min to elute the testosterone. The column was then washed with 5% A-95% B at 1000 μL/min for 0.5 min and then reequilibrated with 35% A-65% B at 1000 μL/min for 0.5 min and then for an additional 0.75 min at 600 μL/min. The retention time for testosterone and D_2T was 2.95 min, and the total run time was 4.75 min. Ion suppression was assessed by postcolumn infusion experiments *(9)*. Suppression occurred only between 0.6 and 1.2 min, well before the retention time for testosterone. Column eluate collected during the first 1.6 min was routinely diverted from the tandem mass spectrometer to waste. Results were obtained by automatic peak-area integration and correction for internal standard recovery using the QuanLynx software. The testosterone concentrations in the controls and unknowns were obtained by automatic interpolation of the corrected peak areas on the calibration curve prepared with each batch of samples. Because sample preparation is a manual procedure, samples, calibrators, and controls were assayed in duplicate as a precaution against operator error. The mean of duplicates was taken unless the imprecision suggested an error had occurred, in which case the sample was repeated.

Ether extraction was compared with zinc sulfate/methanol precipitation for 25 clinical samples over the concentration range <0.3 to 41 nmol/L. There was good agreement between the 2 methods of sample preparation (see *Results* and Fig. 1). All subsequent validation experiments were carried out using the precipitation method.

Three criteria were used to derive the functional limit of quantification of the assay: a signal-to-noise ratio ≥10, an achieved interassay imprecision (CV) ≤20%, and a testosterone concentration that was demonstrably different from zero (see *Results*). Intra- and interassay imprecision was assessed by use of pooled samples with a range of concentrations of testosterone. To assess the accuracy of the ID/LC-MS/MS method, we assayed 3 pools with known ID/GC-MS targets [supplied by Professor Linda Thienpont, Reference Laboratory for Analytical Chemistry, University of Ghent *(10)* and distributed by the United Kingdom National External Quality Assessment Scheme (UKNEQAS)] and 20 in-house pooled clinical samples previously assayed by ID/GC-MS *(11–13)*. We assessed recovery by use of samples from UKNEQAS recovery experiments and samples prepared in house by addition of 0.625, 1.25, 2.5, 5, 10, and 20 nmol/L testosterone to double charcoal-stripped serum and to samples from women and men. The dilutional linearity of the assay was demonstrated by serial dilution of samples with double charcoal-stripped serum. We assessed assay specificity by adding a range of steroids to double charcoal-stripped serum at concentrations of 1, 10, and 100 nmol/L.

We also assessed assay performance by comparing the results for clinical samples with our in-house solvent extraction RIA. Testosterone was extracted from the samples with diethyl ether, and an RIA was performed on the dried residue with tritiated testosterone and an antibody raised locally to testosterone-3-(*O*-carboxymethyloxime)–bovine serum albumin conjugate. Antibody bound and free fractions were separated by dextran-coated charcoal. The antibody shows a high cross-reaction with dihydrotestosterone (53%) and significant cross-reaction with 5α-androstan-3α,17β-diol (4.9%), 5α-androstan-3α-ol-11, 17-dione (2.5%), epitestosterone (2.2%), 5α-androstan-3α,11β-diol-17-one (1.7%), and androstenedione (1.4%). The interassay imprecision (CV) was between 7% and 10% over the concentration range (0.8–25 nmol/L) of the internal quality-control samples. The reference interval established for premenopausal women was <2.8 nmol/L, and that for men 18–60 years of age was 8–27 nmol/L.

Samples from women (RIA concentration range, 0.7–19.8 nmol/L; n = 110) and men (RIA range, 0.9–46.3 nmol/L; n = 111) that had been received from UK hospitals for routine testosterone measurement were assayed by RIA and LC-MS/MS.

ID/GC-MS. For GC-MS analysis, trideuterated testosterone ([16,16,17α-^{2}H]testosterone) was added to serum calibrants, samples, and quality controls. Urea (1 mL of a 4 mol/L solution) was then added to 1 mL of the serum, and the serum was left standing for 20 min to denature the binding proteins present. Solid-phase extraction was performed with Isolute® C_8 (200 mg) non–end-capped

cartridges (Kinesis). After elution with methanol, the extracts were dried under nitrogen. For GC-MS, testosterone was quantified as its bis-trimethylsilyl ether derivative *(11, 12)*, essentially as described previously *(13)*, except that 10 μL of derivatizing reagent and 10 μL of dodecane were used and the volume injected was 2 μL.

STATISTICS

All data were analyzed by the Analyse-it add-on package for Microsoft Excel (Ver. 1.6; Analyse-it UK).

Results

CALIBRATION CURVE

The ratios of analyte peak area to internal standard peak area were plotted against testosterone at 6 concentrations in the range 0.25–100 nmol/L. The line was calculated by Quanlynx software using least-squares linear regression. The response was linear between 0.25 and 100 nmol/L for 5 calibration curves according to the Clinical Laboratory Standards Institute (formerly NCCLS) EP-6 test for linearity. The equation for a representative calibration curve for the data was: $y = 0.31x + 0.16$ nmol/L (SE of the intercept, 0.10 nmol/L; SE of the slope, <0.01 nmol/L).

SAMPLE PREPARATION

There was good agreement between ether extraction and zinc sulfate/methanol precipitation for 25 clinical samples in the range <0.3 to 41 nmol/L (Fig. 1); the equation for the Passing–Bablok regression line for these data was: $y = 0.97x + 0.03$ nmol/L [95% confidence interval (CI) for the intercept, −0.05 to 0.09 nmol/L; 95% CI for the slope, 0.94–1.0 nmol/L]. We adopted zinc sulfate/methanol precipitation as our standard procedure because it is safer, easier, and quicker and can be accomplished within a single tube. All subsequent validation experiments were carried out using this standard procedure.

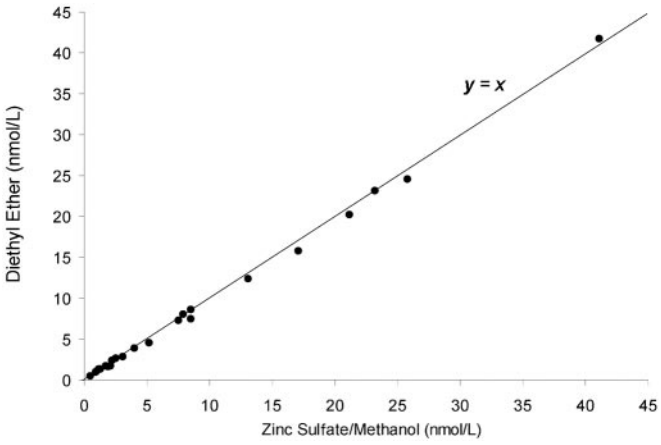

Fig. 1. Plot of testosterone measured in 25 clinical samples prepared by diethyl ether extraction compared with zinc sulfate/methanol precipitation.

The equation for the Passing–Bablok regression line is: $y = 0.97x + 0.03$ nmol/L (95% CI for the intercept, −0.05 to 0.09 nmol/L; 95% CI for the slope, 0.94–1.0). The *diagonal line* is the line of equality.

LC-MS/MS ION CHROMATOGRAMS

Shown in Fig. 2 are chromatograms of endogenous testosterone in clinical samples at 2 concentrations (traces A and B), the 0.25 nmol/L testosterone calibrator prepared in double charcoal-stripped serum (3.3 fmol injected; trace C), and a representative trace of D_2T as internal standard (45 fmol injected; trace D), all extracted from serum by the precipitation method.

LIMITS OF DETECTION AND QUANTIFICATION

The signal-to-noise ratio for the 0.25 nmol/L calibrator was between 9 and 13 over 10 consecutive batches. The mean (SD) response, i.e., the peak area for testosterone divided by the peak area of the internal standard D_2T, for the 0.25 nmol/L calibrator, for 10 batches was 0.0966 (0.0163) with a CV of 17%. Because there was no observable response recorded in the zero calibrator, the lowest response for the 0.25 nmol/L calibrator of 0.0477 (mean − 3 SD) is clearly distinct from zero. Both tests indicate that the lower limit of detection of the assay is <0.25 nmol/L; however, at lower concentrations, the signal-to-noise ratio was <10 and imprecision exceeded 20%, precluding quantification of concentrations <0.25 nmol/L. Assay results are quoted to one decimal place so that the functional limit of quantification is 0.3 nmol/L. The LC-MS/MS ion chromatogram for the calibrator was similar to that of the low-concentration clinical sample (Fig. 2); we therefore considered this designated limit of quantification appropriate for clinical samples.

STABILITY OF EXTRACTED SAMPLES

Four sets of calibrators and quality-control samples were extracted in duplicate. One set was assayed immediately; the remaining sets were left stoppered and stored at 4 °C

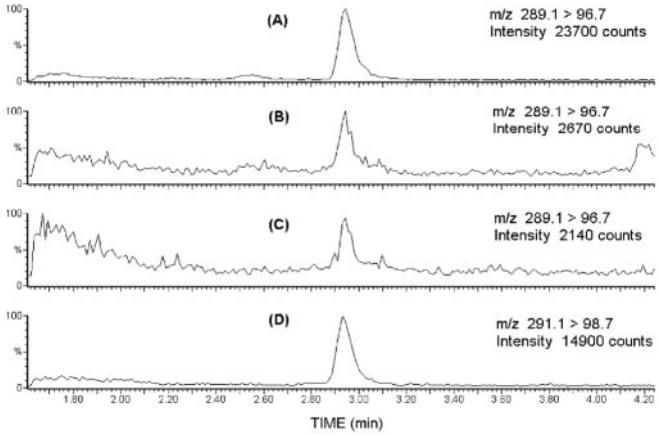

Fig. 2. LC-MS/MS ion chromatograms.

(*A*), transition *m/z* 289.1→96.7 for a clinical sample with a measured testosterone concentration of 4.5 nmol/. (*B*), transition *m/z* 289.1→96.7 for a clinical sample with a measured testosterone concentration of 0.4 nmol/L. (*C*), transition *m/z* 289.1→96.7 for the 0.25 nmol/L testosterone calibrator in double charcoal-stripped serum (3.3 fmol injected). (*D*), a representative trace of D_2T (transition *m/z* 291.1→98.7) as internal standard (45 fmol injected) in the low-concentration clinical sample. All samples were prepared by zinc sulfate/methanol precipitation.

Table 1. Imprecision.

A. Intraassay imprecision based on 10 replicates of serum pools at a range of concentrations measured within batch

Pool	n	Mean testosterone, nmol/L	CV, %
1	10	0.27	16
2	10	0.46	11
3	10	0.64	10
4	10	1.01	6.7
5	10	1.47	4.6
6	10	2.00	6.3
7	10	4.71	6.1
8	10	15.00	3.9
9	10	21.60	3.3
10	10	49.40	3.3

B. Interassay imprecision of 4 quality-control samples analyzed with each batch of samples

Low female	20	0.91	11
High female	20	3.02	6.2
Low male	20	8.59	5.6
High male	20	29.80	4.4

Table 2. Dilutional linearity.

	Testosterone,[a] nmol/L				
Clinical sample	Undiluted	1:2 dilution	1:4 dilution	1:8 dilution	1:16 dilution
Female 1 (with CAH)[b]	8.2	8.0	7.7	7.5	7.7
Female 2	3.1	2.8	3.0	3.0	–[c]
Female 3	2.1	2.3	2.2	2.4	–[c]
Male 1	28.9	27.9	30.2	30.1	31.2
Male 2	19.6	18.2	20.0	19.0	18.7
Male 3	10.5	10.9	11.2	11.4	11.3

[a] Testosterone concentrations in individual clinical samples diluted 1:2 to 1:16 with double charcoal-stripped serum.
[b] CAH, congenital adrenal hyperplasia.
[c] Testosterone concentration was below the limit of quantification.

for 1, 2, or 6 days and then assayed against freshly prepared calibration curves. The stored and fresh calibration curves could be superimposed. The quality-control samples were within the ranges established for the assay and showed no trend.

IMPRECISION

We assessed intraassay imprecision by assaying 10 replicates of plasma pools testosterone concentrations of 0.25–50 nmol/L and interassay imprecision by assaying quality-control samples at 4 concentrations with each batch of clinical samples. The results are summarized in Table 1.

DILUTION LINEARITY

Three samples from women (including 1 with uncontrolled congenital adrenal hyperplasia) and 3 from men were serially diluted with double charcoal-stripped serum. Testosterone concentrations at the different dilutions showed good agreement (Table 2), demonstrating the dilutional linearity of the method.

RECOVERY

Results for samples from UKNEQAS and for the in-house recovery experiments are shown in Table 3.

SPECIFICITY

For the transition *m/z* 289.1→96.7 and under the conditions specified in the testosterone method, estrone gave a peak with a retention time of 3.44 min and a peak area ~40% that of testosterone, and androstenedione gave a peak with a retention time of 2.54 min and an area ~2% of that of testosterone. Neither showed interference at 2.95 min, the retention time of testosterone. The remaining steroids tested, epitestosterone, dihydrotestosterone, dehydroepiandrosterone, androsterone, 19-hydroxytestosterone, ethisterone, and cyproterone acetate, showed no significant peaks above background noise.

ACCURACY

We assessed method accuracy by the results obtained by LC-MS/MS with the results obtained by 2 ID/GC-MS methods *(10–13)*. The values obtained for the UKNEQAS ID/GC-MS targeted pools are shown in Table 4, and those for the in-house pooled samples are shown in Fig. 3.

COMPARISON OF LC-MS/MS WITH IN-HOUSE EXTRACTION RIA

For the comparison of LC-MS/MS with our in-house extraction RIA, we plotted the LC-MS/MS values against the RIA results. Samples from women (concentration range 0.7–19.8 nmol/L; n = 110), showed a correlation (r^2) of 0.96 ($S_{y|x}$ = 0.93 nmol/L; $y = 1.04x - 0.91$ nmol/L; 95% CI for the slope, 0.99–1.07; 95% CI for the intercept, −1.16 to −0.66 nmol/L). The correlation for samples from men (concentration range, 0.9–46.3 nmol/L; n = 111) was similar (r^2 = 0.97) with a SE of 1.48 nmol/L ($y = 1.04x - 0.64$ nmol/L; 95% CI for the slope, 1.0–1.07; 95% CI for the intercept, −1.05 to −0.23 nmol/L). However, when the data were divided into concentration ranges of ≤3, 3.1–8.0, and >8 nmol/L and the results for men and women were plotted together, we found that in the range ≤3 nmol/L, some samples from men (n = 29) gave values that were more strongly negative, compared with the RIA results, than those from women (Fig. 4A). In addition, in the concentration range >8 nmol/L (n = 13), samples from women behaved similarly to samples from men at that concentration and were close to the line of equality (Fig. 4B).

Discussion

We report the development of an LC-MS/MS method for testosterone that is accurate and suitable for routine clinical practice. The importance of this development is powerfully confirmed by a recent report *(8)* and an editorial in this Journal on the use of immunoassays to

Table 3. Recovery.

A. UKNEQAS recovery experiment distributed in May 2004[a]

	ALTM[b] testosterone, nmol/L	ALTM recovery, %	ID/LC-MS/MS testosterone, nmol/L	ID/LC-MS/MS recovery, %
Female pool + added testosterone				
Base pool	1.2		0.8	
Base pool + 2.17 nmol/L	2.75	71.4	3.2	110.6
Base pool + 4.33 nmol/L	4.26	70.7	5.4	106.2
Male pool + added testosterone				
Base pool	10.4		11.5	
Base Pool + 10.4 nmol/L	21.0	101.9	21.7	98.1
Base pool + 20.1 nmol/L	30.5	100.0	31.6	100.0

B. Recovery of testosterone added to double charcoal-stripped plasma and samples from men and women

		Recovery, %						
Sample	Baseline testosterone, nmol/L	0.625 nmol/L[c]	1.25 nmol/L[c]	2.5 nmol/L[c]	5.0 nmol/L[c]	10 nmol/L[c]	20 nmol/L[c]	Mean recovery, %
DCSS	0	91	86	95	90	86	97	91
Female 1	1.7	90	111	96	88			96
Female 2	0.97	112	94	102	94			101
Female 3	–[d]		101	114	107			107
Male 1	13.7				94	91	85	90
Male 2	17.5				80	93	91	85
Male 3	11.5				74	73	95	81

[a] Recoveries are shown for the all-laboratory trimmed mean and our ID/LC-MS/MS method.
[b] ALTM, all-laboratory trimmed mean; DCSS, double charcoal-stripped serum.
[c] Concentration of testosterone added.
[d] Insufficient sample volume for baseline assay of female 3. Baseline for recovery estimation taken as the sample with 0.625 nmol/L testosterone added (1.95 nmol/L).

measure testosterone in women, which suggested that such assays were "[no] better than a guess" *(14)*. As a tertiary referral center for steroid hormone analyses, we have maintained an in-house solvent extraction RIA for more than 20 years so that we can verify results for samples from women found to contain high testosterone concentrations by direct immunoassay performed in other hospitals *(15)*. To introduce the ID/LC-MS/MS method into routine service, it was necessary to run the 2 assays in parallel to compare performance. Our data for samples from women with testosterone concentrations ≤3 nmol/L by RIA are similar to those reported previously *(8, 16, 17)* for direct immunoassays, in that extraction RIA also overestimates the testosterone concentration in the majority of samples. However, the results for samples from men in this low concentration range, although small in number (n = 29), are of interest. The distribution appears to be bimodal, with some samples approaching the line of equality and others showing values that are more strongly negative, when compared with RIA results, than those for most samples from women (Fig. 4A). This finding suggests that the problem is not caused solely by an inherent matrix effect in female serum, as might be inferred from previous publications on the topic *(8, 17)*. Male hypogonadism can be either primary or secondary, and the secondary causes include both pituitary failure and antiandrogen therapy, such as that used for prostate cancer. The different causes may introduce different interferences, which could explain the bimodal distribution of our results. However, because many testosterone requests come from outside our hospital, the unavailability of complete clinical details makes it impossible to conclude whether these pathophysiologic causes are relevant. The overestimation of testosterone in low-concentration samples from some men is important because the efficacy of suppressive treatment in prostate cancer is monitored by testosterone concentration, usually measured in an immunoassay.

Our limited experience (n = 13) of clinical samples from women with testosterone concentrations >8 nmol/L indicates that at these high concentrations, correlation with the RIA is better than at low concentrations (Fig. 4B). These results support the view that the main problem with immunoassay is poor specificity, which manifests itself most at low testosterone concentrations, irrespective of gender.

Table 4. Comparison of ID/LC-MS/MS results with ID/GC-MS target values for pools distributed by UKNEQAS.

	Testosterone, nmol/L			
		LC-MS/MS		LC-MS/MS as percentage of GC-MS target
Pool	GC-MS target	Individual values	Mean	
286	0.7361	0.76;0.73;0.74;0.92	0.79	107
288	1.225	1.26;1.36;1.09;1.13	1.22	99
293	2.72	2.84;2.53;2.67;2.95	2.75	101

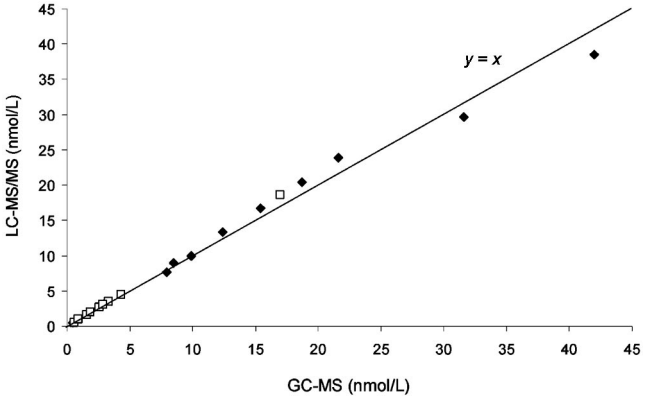

Fig. 3. Comparison of testosterone measured by LC-MS/MS compared with GC-MS values.

Each *point* represents a single pool of samples from women (□) or men (♦), as described in the *Materials and Methods*. Equation for the regression line for all samples: $y = 0.96x + 0.65$ nmol/L ($r^2 = 0.99$; SE of the slope, 0.02; SE of the intercept, 0.36 nmol/L; SE of r^2, 1.17). If the 2 highest concentration pools are omitted, r^2 becomes 1.0 with a SE of 0.3 nmol/L, and the equation for the line becomes: $y = 1.09x - 0.16$ nmol/L (SE of the slope, 0.01; SE of the intercept, 0.11 nmol/L). The *diagonal line* is the line of equality.

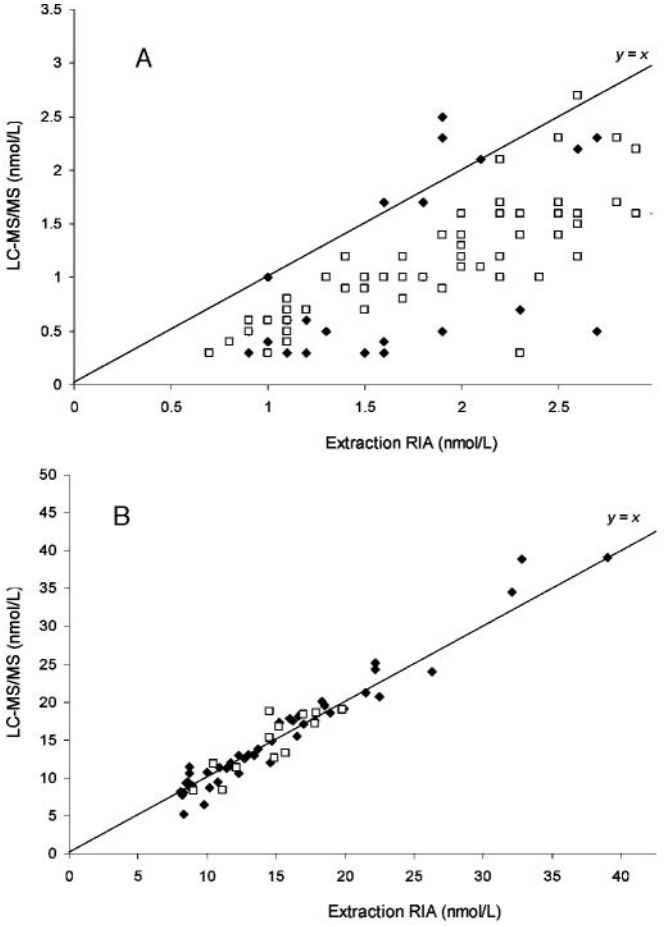

Fig. 4. Comparison of testosterone measured by LC-MS/MS with our in-house extraction RIA for clinical samples from women (□) and men (♦).

(*A*), testosterone concentrations ≤3 nmol/L by RIA. The bimodal distribution for samples from men can be clearly seen and is further demonstrated by the differences in correlation: $r^2 = 0.66$ (SE = 0.3 nmol/L) for women (n = 67), and $r^2 = 0.35$ (SE of 0.6 nmol/L) for men (n = 29). The equations for the regression lines are similar; for women: $y = 0.69x - 0.15$ nmol/L (95% CI for the slope, 0.58–0.84; 95% CI for the intercept, −0.42 to 0.08 nmol/L); for men: $y = 0.74x - 0.26$ nmol/L (95% CI for the slope, 0.43–1.2; 95% CI for the intercept, −1.06 to 0.31 nmol/L). (*B*), testosterone concentrations >8 nmol/L by RIA. Samples from men show good agreement between LC-MS/MS and RIA ($r^2 = 0.97$; SE = 1.72 nmol/L; $y = 1.02x - 0.2$ nmol/L; 95% CI of the slope, 0.97–1.07; 95% CI of the intercept, −1.14 to 0.78 nmol/L). Samples from women were limited in number (n = 13) but showed a better correlation with the RIA than at low concentrations ($r^2 = 0.77$; SE = 2.13 nmol/L; $y = 1.12x - 1.7$ nmol/L; 95% CI of the slope, 0.46–1.46; 95% CI of the intercept, −6.81 to 5.36 nmol/L).

We have now assayed more than 2000 clinical samples by ID/LC-MS/MS. During this time, 6 samples have had similar chromatographic patterns, showing significant additional peaks in the region of the endogenous testosterone peak. These peaks interfered with the integration of the testosterone peak, making it impossible to obtain an accurate measurement of the concentration. We have been unable to explain this interference in terms of sample collection and/or storage conditions or drug treatment, although it is noteworthy that all of these samples had been referred by a single external laboratory. We were able to remove the interference by substituting the ether extraction method of sample preparation for the zinc sulfate/methanol precipitation we use routinely.

Measurement of testosterone over a wide concentration range is required to aid clinical decisions. Very low values are needed for patients with cancer of the prostate, for the investigation of hypoandrogenic states in women, and also for pediatric samples. Values at the top of the female reference interval are required in female hyperandrogenism; values at the bottom of the male range are needed to test for hypogonadism and at the top of the male range for androgen insensitivity and monitoring testosterone replacement. The precision profile for this ID/LC-MS/MS assay is acceptable, with a CV <7% across the range 1–50 nmol/L and a functional limit of quantification of 0.25 nmol/L with a CV of 17%. Although lower testosterone concentrations are found in some clinical conditions, a reported value <0.3 nmol/L is adequate for most clinical purposes.

We compared ID/LC-MS/MS with an ID/GC-MS assay in a single blind manner *(11–13)*. Ideally, single collections from patients would have been used for this exercise. Unfortunately, the ID/GC-MS method requires more serum than is available from a single clinical sample, and it was necessary to prepare pools of samples from men and women; the number of samples pooled, however, was kept to a minimum. The results for the 20 pools (range, 0.5–38.5) prepared in house were very encouraging (Fig. 3). The volumes of the in-house pools were insufficient to allow duplicate measurements by ID/GC-MS, making the agreement remarkable. However, the results obtained by ID/LC-MS/MS were slightly different from those obtained by ID/GC-MS, being slightly higher for pools containing 5–25 nmol/L testosterone and lower for the 2 highest pools, a finding that is of interest and requires further elucidation. It is possible that the differ-

ence may be attributable to different calibration protocols. The ID/LC-MS/MS results were quantified in exactly the same way as for a routine batch, i.e., by interpolation of the results on a single calibration curve with a range of 0.25–100 nmol/L. The ID/GC-MS method splits the calibration curve into a series of narrow concentration ranges appropriate to the sample being assessed. This method is potentially more accurate but would add considerably to the time required to report a routine batch of clinical samples, and the difference achieved is unlikely to be of clinical significance. The 3 ID/GC-MS targeted pools distributed by UKNEQAS also gave mean values close to the targets (Table 4), indicating good accuracy, although the number of pools was small and the concentration range narrow. However, the precision for these pools is not as good as would have been expected from the imprecision data shown in Table 1.

In summary, the ID/LC-MS/MS analytical technique is robust with good recovery and linearity and a low limit of detection. Moreover, the accuracy was confirmed by comparison with 2 ID/GC-MS methods. The validation experiments indicate that the ID/LC-MS/MS method is superior in performance to our in-house solvent extraction RIA and to the commercial immunoassays in general use. Furthermore, improved LC-MS/MS technology means that testosterone can be measured in 50 μL of serum rather than the previously reported amounts of 1 mL for GC-MS *(16)* or 2 mL for LC-MS/MS *(17)*. Individual samples can be analyzed in duplicate over a period of 11 min, which permits a batch of 192 tubes (75 clinical samples, calibrators, and controls) to be measured in duplicate during an overnight run. Because of the low sample volume requirement and the minimal sample preparation and high sample throughput, LC-MS/MS is a suitable technique for a routine clinical biochemistry laboratory analyzing >8000 samples for testosterone per year.

We would like to thank Dr Jonathan Middle, UKNEQAS organizer for steroid hormones, for providing extra aliquots of the GC-MS targeted pools and for giving permission to use the recovery experiment pools as part of our assay validation. We would also like to thank Drs. A.A.A. Ismail and M. Magera for invaluable discussions at the inception of this project and April Fairhurst, Annie Fleming, Karolina Ossowska, and Chris Walker for expert technical assistance.

References

1. Weber TH, Kapyaho KI, Tanner P. Endogenous interference in immunoassays in clinical chemistry. A review. Scand J Clin Lab Invest 1990;50(Suppl 201):77–82.
2. Marks V. False-positive immunoassay results: a multicenter survey of erroneous immunoassay results from assays of 74 analytes in 10 donors from 66 laboratories in seven countries. Clin Chem 2002;48:2008–16.
3. Torjesen PA, Bjoro T. Antibodies against [^{125}I]testosterone in patient's serum: a problem for the laboratory and the patient [Letter]. Clin Chem 1996;42:2047–8.
4. Ismail AAA, Walker PL, Cawood ML, Barth JH. Interference in immunoassay is an underestimated problem. Ann Clin Biochem 2002;39:366–73.
5. Sickman L. Determination of steroid hormones by the use of isotope dilution-mass spectrometry: a definitive method in clinical chemistry. J Steroid Biochem 1979;11:117–23.
6. Lawson AM, Gaskell SJ, Holm M. International Federation of Clinical Chemistry (IFCC), Office for Reference Methods and Materials (ORMM). Methodological aspects on quantitative mass spectrometry used for accuracy control in clinical chemistry. J Clin Chem Clin Biochem 1985;23:433–41.
7. Starcevic B, Distefano E, Wang C, Catlin DH. Liquid chromatography-tandem mass spectrometry assay for human serum testosterone and trideuterated testosterone. J Chromatog B Analyt Technol Biomed Life Sci 2003;792:197–204.
8. Taieb J, Mathian B, Millot F, Patricot M-C, Mathieu E, Queyrel N, et al. Testosterone measured by 10 immunoassays and by isotope-dilution gas chromatography-mass spectrometry in sera from 116 men, women and children. Clin Chem 2003;49:1381–95.
9. Annesley TM. Ion suppression in mass spectrometry [Review]. Clin Chem 2003;49:1041–4.
10. Thienpoint LM, Van Nieuwenhove B, Stockl D, Reinauer H, De Leenheer AP. Determination of reference method values by isotope dilution-gas chromatography/mass spectrometry: a five years' experience of two European Reference Laboratories. Eur J Clin Chem Clin Biochem 1996;34:853–60.
11. Donike M, Zimmermann J. Preparation of trimethylsilyl, triethylsilyl and tert-butyldimethylsilyl enol ethers from ketosteroids for investigations by gas-chromatography and mass-spectrometry. J Chromatogr 1980;202:483–6.
12. Opfermann G, Schanzer W. Trimethylsilylation—aspects for derivatisation. In: Schanzer W, Geyer H, Gotzmann A, Mareck-Engelke U, eds. Recent advances in doping analysis. Proceedings of the 14th Cologne Workshop on Dope Analysis, 17th to 22nd March 1996, Vol. 4. Koln: Sport und Buch Strauss, 1997:247–52.
13. Cowan DA, Kicman AT, Walker CJ, Wheeler MJ. Effect of administration of human chorionic gonadotrophin on criteria used to assess testosterone administration in athletes. J Endocrinol 1991;131:147–54.
14. Herold DA, Fitzgerald RL. Immunoassays for testosterone in women: Better than a guess? [Editorial]. Clin Chem 2003;49:1250–1.
15. Wheeler MJ, Shaikh M, Jennings RD. An evaluation of 13 commercial kits for the measurement of testosterone in serum and plasma. Ann Clin Biochem 1986;23:303–9.
16. Fitzgerald RL, Herold DA. Serum total testosterone: immunoassay compared with negative chemical ionization gas chromatography-mass spectrometry. Clin Chem 1996;42:749–55.
17. Wang C, Catlin DH, Demers LM, Starcevic B, Swerdloff RS. Measurement of total testosterone in adult men: comparison of current laboratory methods versus liquid chromatography-tandem mass spectrometry. J Clin Endocrinol Metab 2004;89:534–43.