CLINICAL STUDY

Assessment of androgen concentration in women: liquid chromatography–tandem mass spectrometry and extraction RIA show comparable results

Femi Janse1, Martinus J C Eijkemans1,2, Angelique J Goverde1, Eef G W M Lentjes3, Annemieke Hoek4, Cornelius B Lambalk5, Theresa E Hickey6, Bart C J M Fauser1 and Robert J Norman6

1Department of Reproductive Medicine and Gynecology, University Medical Center Utrecht, Room F05-126, PO Box 75500, 3508 GA Utrecht, The Netherlands, 2Julius Center for Health Sciences and Primary Care and 3Laboratory of Endocrinology, Department of Clinical Chemistry and Hematology, University Medical Center Utrecht, Utrecht, The Netherlands, 4Section of Reproductive Medicine, Department of Obstetrics and Gynecology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, 5Department of Reproductive Medicine, VU University Medical Center, Amsterdam, The Netherlands and 6Discipline of Obstetrics and Gynecology, School of Paediatrics and Reproductive Health, Robinson Institute, Research Centre for Reproductive Health, University of Adelaide, Adelaide, South Australia, Australia

(Correspondence should be addressed to F Janse; Email: f.janse@umcutrecht.nl)

Abstract

Objective: The measurement of serum testosterone in women is challenging due to lack of trueness, precision, and sensitivity of various available testosterone assays. Accurate assessment of testosterone in women is crucial especially in conditions associated with alleged over- or under-production of testosterone, such as in polycystic ovary syndrome (PCOS) or primary ovarian insufficiency (POI). The aim of this study was to measure and compare androgen concentrations in women with PCOS, POI, and female controls and to evaluate the performance of extraction RIA and liquid chromatography–tandem mass spectrometry (LC–MS/MS) in these women.

Design: Cross-sectional study.

Methods: Carefully phenotyped women with POI (n = 208) or PCOS (n = 200) and 45 healthy, regularly cyclic female controls were included. Method comparison analyses were performed for total testosterone, androstenedione (AD), and DHEA, as measured by LC–MS/MS and extraction RIA.

Results: All androgen levels were significantly elevated in women with PCOS compared with POI patients (P < 0.05) and controls (P < 0.05). Women with POI presented with similar androgen concentrations as controls, except for AD. Compared with measurements by extraction RIA, testosterone, DHEA, and AD concentrations measured by LC–MS/MS were systematically lower. However, using extraction RIA and LC–MS/MS, testosterone, DHEA, and AD measurements were shown to have good agreement as assessed by Bland–Altman analysis and intraclass correlation coefficient: 0.95 (95% confidence interval 0.94–0.91), 0.83 (0.79–0.86), and 0.96 (0.95–0.97) respectively.

Conclusions: LC–MS/MS, compared with a labor-intensive extraction RIA, shows good precision, sensitivity, and high accuracy for measuring female testosterone, DHEA, and AD concentrations under various clinical conditions. LC–MS/MS, therefore, represents a convenient and reliable assay for both clinical and research purposes, where androgen measurement in women is required.

European Journal of Endocrinology 165 925–933

Introduction

The measurement of serum testosterone has proven to be a great challenge due to lack of precision and sensitivity of various commercially available testosterone assays, resulting in limited utility (1). Testosterone assays were originally developed to measure testosterone concentrations in the normal male range (2). However, women exhibit testosterone levels well below this range, and therefore, reliable measurement of female testosterone concentrations is particularly problematic due to trueness and precision problems (3). In women of reproductive age, 25% of testosterone biosynthesis occurs in the ovarian theca cells. Another 25% is produced by the adrenal gland, and the remaining 50% of circulating testosterone derives from peripheral conversion of testosterone precursors (DHEA and androstenedione (AD)). DHEA and its sulfate metabolite (DHEAS) are exclusively produced by the adrenal glands, whereas AD is produced by both the adrenal and the ovary (4, 5). Most circulating testosterone is biologically inactive as it binds to serum proteins, primarily sex hormone-binding protein (SHBG) and albumin (2, 6).
The most widely used methods for measuring serum testosterone are the simple, relatively inexpensive RIA and chemiluminescence immunoassay, performed directly in serum. Although the immunoassays show average good precision, they often show more bias. This is especially true for the lower range, where they can be subject to increased interference and overestimation of steroid concentrations compared with other assays (7, 8). Using extraction and chromatography methods preceding RIA has the advantage of removing interfering proteins and cross-reacting steroids. Although extraction RIA seems preferable over direct assays, it is infrequently used in clinical practice because proper validation is lacking and extraction is labor intensive and time consuming (1, 9, 10). A commonly used practical approach to estimate bioactive testosterone is the calculation of the free androgen index (FAI). While FAI was shown to correlate quite well with physical separation measures of female free testosterone (not male), FAI is highly dependent on the quality of testosterone and SHBG assay measurements (2).

In order to overcome these difficulties, tandem mass spectrometry (MS/MS) preceded by gas or liquid chromatography (GC–MS/MS and LC–MS/MS) assays for steroid measurement is emerging (11). LC–MS/MS has some advantages over the use of immunoassays, especially compared with those performed on platforms. LC–MS/MS shows equal or better precision, and these assays do not suffer from interferences due to chromatographic separation and mass spectrometry analysis. Interferences can cause serious bias at low testosterone concentrations, which may be encountered in postmenopausal women and children. However, the high costs of mass spectrometry equipment may be a possible drawback for routine clinical use. Although LC–MS/MS assays show higher precision and better accuracy than immunoassays, there is still a difference in performance (sensitivity, specificity, and trueness) among LC–MS/MS methods and thus there is a need for standardization (12).

Reliable assessment of testosterone in women is crucial for the diagnosis of conditions associated with alleged overproduction of testosterone such as in polycystic ovary syndrome (PCOS), or conditions that may incur androgen deficiency, such as postmenopause, oophorectomy, or primary ovarian insufficiency (POI) (13–15). However, analysis of serum testosterone by commercial assays fails to identify biochemical hyperandrogenism in 20–40% of women with PCOS, while hyperandrogenism is one of the criteria on which the diagnosis is based (16, 17). Moreover, female androgen insufficiency remains controversial due to the lack of reliable serum testosterone assays to define this condition (10). Clinical guidelines concerning indications for androgen therapy in case of hypoandrogenism cannot be established (18). These examples illustrate the need for a reliable testosterone measurement, which is easily available at reasonable cost for both clinical and research purposes.

The current study was aimed at evaluating the measurement of testosterone and other androgens by LC–MS/MS in comparison to measurement by extraction RIA in women with normal and presumably elevated and decreased androgens.

Materials and methods

Subjects

In 2004, a study involving standardized phenotyping and follow-up was initiated in the University Medical Center Utrecht, The Netherlands, in which all women suspected of menstrual cycle disturbance and referred to our outpatient clinic participated. Details of the systematic evaluation, approved by the local ethics committee, have been described previously (19, 20). For the current study, 169 women with POI and 200 women affected by PCOS, who visited the outpatient clinic until July 2009, were included. POI was defined as secondary amenorrhea for at least 4 months in women younger than 40 years, along with repeated FSH concentrations exceeding 40 IU/l (21). PCOS was diagnosed according to the Rotterdam consensus criteria if at least two of the following criteria were present: i) oligo/anovulation, ii) clinical and/or biochemical hyperandrogenism (testosterone > 2.0 nmol/l, assessed by our in-house extraction RIA), and/or iii) polycystic ovaries on ultrasonography (17). Additionally, 39 women with POI visiting four other centers (participating in the Dutch POI Consortium (20, 22)) were also included in the current study.

As controls, 45 women visiting the preconceptional clinic prior to starting IVF/ICSI treatment because of severe male infertility between October 2006 and July 2008 were included. Severe male infertility was defined as semen analysis with volume × concentration × motility below 2.0 million. All women were healthy and experienced regular menstrual cycles. Exclusion criteria were PCOS, previous poor response after ovarian hyperstimulation, age > 38 years, early follicular FSH concentrations > 12 U/l, and thyroid dysfunction. All included study subjects gave written informed consent.

Measurements

Serum samples were collected during the early follicular phase of the menstrual cycle in women with regular cycles. In case of a cycle disturbance, samples were drawn at a random cycle day and a progesterone concentration was measured concomitantly to ensure preovulatory state. None of the women had used any hormonal medication for at least 2 weeks. After collection and aliquoting, one fresh sample was analyzed immediately while the remaining aliquots were stored at −20 °C. In Utrecht, the fresh testosterone measurement was performed using an in-house
Serum for women with POI and PCOS: estradiol (E\textsubscript{2}), following measurements were performed on fresh intercept:

K

95% confidence limit of the slope: 1.04–1.20 and of the 1.04 to 0.91) (23). Furthermore, the following measurements were performed in duplicate. Results were similar across the complete standard range (eight replicates) and between −2.0 and +1.6% in the range 0.173–3.47 nmol/l. Precision for testosterone ranged between 1.5 and 7.5% across the whole range and was 1.5–9.3% between 0.173 and 3.47 nmol/l. Results were similar for all analytes measured. The following inter-assay CVs were identified: in controls: 5.6% (testosterone), 8.4% (DHEA), and 6.4% (AD); in PCOS: 4.3% (testosterone), 5.9% (DHEA), and 4.5% (AD). The limit of detection was determined by repeated measures of the three lowest standards used in the assay. A precision dose profile was established and statistical significance between the standards shown by clear separation of the doses. The run time was 10 min per sample.

FAI was calculated for both testosterone measurements: testosterone/SHBG×100 (24) using the SHBG measurement performed in Utrecht.

**Statistical analysis**

Median values and range for each individual group of women were compared using one-way analysis of covariance and Bonferroni post-hoc tests for significance, correcting for age and body mass index (BMI; possible confounders) (25). Method comparison analyses were performed by constructing scatter plots and Bland–Altman plots to evaluate the extent to which testosterone, DHEA, and AD measurements agreed in both methods. Bland–Altman plots show the percentage difference between two corresponding measurements plotted against the measurements’ mean.

For all Bland–Altman plots, the mean percentage difference between the methods (mean bias) with
associated 95% confidence intervals (CIs) and limits of agreement (±2 S.D.) were calculated. Intraclass correlation coefficients (ICC) with 95% CIs were calculated for testosterone, DHEA, and AD. Large ICC (maximum 1) indicates that variation between both assays is equal to zero (perfect agreement) (26, 27). Because androgen concentrations were not normally distributed, logarithmic transformation of variables was applied before execution of analyses. When testosterone or DHEA concentrations were undetectable, 0.5×LLOD was used (28).

To further compare the performance of extraction RIA and LC–MS/MS for the measurement of testosterone, DHEA, and AD in this population, correlations between various androgens and E2 were assessed. To this end, linear regression analyses were performed, where testosterone (RIA) and testosterone (LC–MS/MS) were entered separately and similarly repeated, entering the variables in the opposite order. Univariate standardized β coefficients with associated s.e.m. are reported. Larger univariate standardized β coefficients represent stronger associations with the steroids. Univariate standardized β coefficients of both testosterone assays were compared using a bootstrapping procedure with 2000 replications, because both testosterone measurements were performed in identical samples. 95% CIs were calculated to identify significant differences between the correlations with androgens and E2 vs the testosterone measurements by RIA and LC–MS/MS assays. Identical analyses were performed for DHEA (RIA) and DHEA (LC–MS/MS), and for AD (RIA) and AD (LC–MS/MS). Statistical analyses were performed using SPSS 15.0 (SPSS, Inc., Chicago, IL, USA) and R version 2.9.0. (http://www.r-project.org/).

**Results**

Median concentrations of all measurements are shown per patient group in Table 1. In women with PCOS, all hormone concentrations were significantly higher compared with those in POI patients and in controls, with exception of SHBG. Androgens (except AD), E2, and SHBG concentrations were similar for women with POI and controls.

The scatter and Bland–Altman plots for testosterone, DHEA, and AD measurements are depicted in Figs 1–3. All scatter plots show high correlation between extraction RIA and LC–MS/MS with correlation coefficients of 0.960 for testosterone and AD and 0.830 for DHEA. The ICC for testosterone, DHEA, and AD were 0.95 (95% CI 0.94–0.96), 0.83 (0.79–0.86), and 0.96 (0.95–0.97), respectively, indicating good agreement between both extraction RIA and LC–MS/MS assays. It becomes evident from all Bland–Altman plots that the average percent difference (i.e. mean bias) in serum testosterone, DHEA, and AD levels between LC–MS/MS and extraction RIA were −31.5, −7.1, and −20.9% respectively. This indicates that LC–MS/MS yields significantly lower serum testosterone and AD results than extraction RIA, which may possibly be explained by interfering substances in the latter assay. In 24 (5.3%) of testosterone, 22 (6.0%) of DHEA, and 16 (4.4%) of AD, the differences between the methods were beyond the limits of agreement (mean difference ±2 S.D.), with the majority of these outliers occurring in the lower concentration ranges.

<table>
<thead>
<tr>
<th>Table 1 Baseline characteristics for each patient group. Results are shown as median (range). Analysis of covariance was carried out on logarithmically transformed variables to study the differences between patient groups.</th>
<th>Controls (n=45)</th>
<th>POI (n=208)</th>
<th>PCOS (n=200)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>33.1 (20.3–37.8)</td>
<td>37.6 (15.8–62.9)</td>
<td>29.3 (17.9–41.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>23.4 (17.1–36.9)</td>
<td>23.4 (17.1–37.3)</td>
<td>24.2 (16.2–55.9)</td>
<td>0.003</td>
</tr>
<tr>
<td>Testosterone RIA (nmol/l)</td>
<td>1.20 (0.61–3.90)</td>
<td>1.10 (0.29–3.00)</td>
<td>2.00 (0.64–6.30)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Testosterone LC–MS/MS (nmol/l)*</td>
<td>0.87 (0.31–2.78)</td>
<td>0.76 (0.09–2.30)</td>
<td>1.55 (0.34–5.48)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SHBG RIA (nmol/l)</td>
<td>60 (12–199)</td>
<td>55 (17–201)</td>
<td>47 (7–191)</td>
<td>0.04*</td>
</tr>
<tr>
<td>DHEA RIA (nmol/l)</td>
<td>11.0 (1.10–43.0)</td>
<td>18.0 (2.00–67.0)</td>
<td>15.8 (4.34–65.2)</td>
<td>NS</td>
</tr>
<tr>
<td>DHEA LC–MS/MS (nmol/l)^a</td>
<td>16.0 (8.74–58.9)</td>
<td>12.1 (4.34–43.0)</td>
<td>15.8 (4.34–65.2)</td>
<td>NS</td>
</tr>
<tr>
<td>DHEAS RIA (µmol/l)</td>
<td>N/A</td>
<td>3.90 (0.41–12.0)</td>
<td>4.80 (0.77–14.0)</td>
<td>NS</td>
</tr>
<tr>
<td>AD RIA (nmol/l)</td>
<td>N/A</td>
<td>2.90 (0.80–8.90)</td>
<td>7.20 (2.40–22.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AD LC–MS/MS (nmol/l)</td>
<td>3.29 (1.31–10.9)</td>
<td>2.23 (0.33–8.46)</td>
<td>5.87 (1.81–15.9)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FAI RIA</td>
<td>1.83 (0.75–15.8)</td>
<td>1.76 (0.29–8.24)</td>
<td>4.65 (0.68–26.2)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FAI LC–MS/MS</td>
<td>1.29 (0.48–10.7)</td>
<td>0.56 (0.02–2.11)</td>
<td>3.36 (0.60–21.0)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

All P values are adjusted for age and BMI. Italic values in P column indicate that all groups were significantly different from each other; bold values in P column indicate that PCOS was significantly different from controls and POI; * indicates that PCOS was significantly different from controls. N/A, not available; NS, not significant; LLOD, lower limit of detection.

*Measurements below LLOD: in POI, n=1 (0.4%).

| n=195. |

^aMeasurements below LLOD: in controls, n=15 (33%); POI, n=58 (28%); and PCOS, n=14 (7%).

www.eje-online.org
In conclusion, the Bland–Altman plots suggest that the two methods are in good agreement with testosterone, DHEA, and AD.

Finally, correlations between testosterone, DHEA, and AD vs related steroids (i.e. other androgens and E$_2$) were calculated to identify whether the extraction RIA or the LC–MS/MS assays showed a difference in these associations (Supplementary Tables 1–3, see section on supplementary data given at the end of this article). In general, most correlations were similar among the assays. The identification of some small significant differences indicates that the strength of the association increased when related steroids were measured employing the identical assay with which the original measurement was performed.

**Discussion**

In this study, the performances of LC–MS/MS and an in-house extraction RIA were compared for a broad range of androgen concentrations encountered in women of reproductive age, including the clinical conditions POI and PCOS. LC–MS/MS and extraction RIA demonstrated good agreement with the measurement of testosterone, DHEA, and AD, although LC–MS/MS gave significantly lower concentrations when compared with extraction RIA. Furthermore, similar correlations were identified in comparison with the association for each assay with other steroids. Therefore, LC–MS/MS, compared with a labor-intensive extraction RIA, shows good precision, sensitivity, and high accuracy for the measurement of androgens in women with PCOS, POI, and regularly cycling female controls.

Androgen concentrations for women with PCOS were significantly higher compared with female controls and women affected by POI, which confirms other recent studies applying LC–MS/MS (26, 27). However, this is the first study to measure androgens with LC–MS/MS in women with POI. In this group, testosterone and DHEA concentrations were similar as female controls, while AD concentrations were significantly lower. In the existing literature, it is debated whether POI is associated with decreased androgen concentrations (14, 29–33), and only one study presented with results similar to ours (34). As previous studies only applied direct RIA, chemiluminescence immunoassays, and extraction RIA, conflicting findings most probably resulted from lower sensitivity, lack of specificity, or lower accuracy of these assays (10). While decline of androgen synthesis may result from menopausal status, this has also been associated with increasing age (30, 35). Therefore, the results of the current study support the concept that young women with POI encounter normal testosterone and DHEA concentrations, in contrast with women who experience menopause at a regular age. Currently, there is no explanation for the significantly lower AD concentrations encountered in women with POI. This finding needs further investigation.

MS after GC or LC is a relatively new technique for steroid sex hormone measurement and data on its applicability, particularly in the female sex hormone concentration range, are scarce. The LC–MS/MS assay in this study showed excellent precision and accuracy down to at least 0.173 nmol/l (precision $-0.2%$; accuracy 1.5% for authentic standards). This allows for a great deal of confidence in the reliability of an assay, which is completely specific for testosterone or other nominated steroid analytes. A study comparing LC–MS/MS with direct and extraction RIAs in normal and hypogonadal male subjects by Bland–Altman plots showed systematically lower results by LC–MS/MS (36). Studies on female and pediatric samples confirmed this finding, while also identifying that LC–MS/MS showed greater accuracy in lower androgen concentrations.
than direct assays (37, 38). This study included serum samples of women with PCOS, POI, and normally cyclic women. We identified good agreement between LC–MS/MS and the in-house extraction RIA (even for lower ranges encountered in POI) and slightly lower hormone concentrations reported after LC–MS/MS. Previous studies on the comparison of LC–MS with extraction RIA identified overestimation of extraction RIA between 20 and 80% in testosterone concentrations below 3.47 nmol/l (1 ng/dl), similar to our findings: between −0.2 and 63% (36). We are cautious to generalize the agreement between extracted RIA and LC–MS/MS in the current study, because only two centers participated in the current study, but the results are in accordance with previous studies (36, 37, 39), including a study on women with PCOS, which confirms that well-chosen RIA may offer equal precision and accuracy like LC–MS/MS (40).

When examining the performance of extraction RIA and LC–MS/MS, it is important to include a clinical validation of the androgens measured by each assay. Ideally, correlation analyses should be performed with clear clinical endpoints, such as degree of hirsutism or presence of the female androgen deficiency syndrome. However, until now, clear clinical endpoints related to androgen concentrations cannot irrefutably be identified (40–42). Therefore, the associations of androgens, measured by both assays, were compared with other physiologically related steroids (androgens and E2). Our results indicate that for the two assays evaluated in this study, the correlations were similar and both assays showed equal performance.

The major strength of this study is the large sample size of well-phenotyped women with PCOS and POI, representing the broad spectrum of female androgen concentrations during reproductive life. Laboratory personnel at both sites were blinded for study design and identity of the samples. But this study may also have potential weaknesses. While all extraction RIA assessments were done in fresh samples, frozen samples were shipped to Australia before LC–MS/MS processing. However, previous studies on long-term stored samples identified that steroids are very stable after being isolated in the serum (43). Furthermore, early follicular androgen concentrations in regularly cyclic controls may underrepresent androgen levels present during the middle third of the cycle (18). Finally, the group of female controls was relatively small, which may have concealed potential differences in androgen concentrations compared with those in women with POI.

Although our results show that LC–MS/MS is a reliable alternative for measuring steroid sex hormones in women, its introduction into clinical or research practice may depend on several factors: costs of the LC–MS equipment, LC–MS expertise in laboratories, and future RIA development. The LC–MS is now introduced in many laboratories, but the methods designed on this analyzer should meet the same requirements (defined by the FDA or European Directive) like the commercial immunoassays. This means that the validation of an LC–MS/MS procedure, like the validation of extraction RIAs, has to be much more extensive than for commercial immunoassays, because LC–MS assays are mostly in-house-developed methods unlike already validated kits. Furthermore, inappropriate sample preparation and analyte detection may greatly influence the performance of LC–MS/MS and extraction RIA procedures, thus emphasizing the importance of validation. Extraction RIA may gain increased specificity if more specific antibodies are selected, and the method for the displacement of bound testosterone from its binding proteins is optimized (36). If this occurs, new comparison studies
With the demonstration that LC–MS/MS, compared with a labor-intensive RIA, exhibits good precision, sensitivity, and high accuracy for measuring female androgen concentrations, a convenient assay has become available for clinical and research purposes. Reproducibility of the LC–MS/MS between different LC–MS/MS laboratories needs to be assessed to meet the performance criteria as is customary for new assays, before this assay will be adopted universally. The next phase in the introduction of LC–MS/MS should aim at obtaining highly true and precise reference values. These reference values could then be applied using other assays with same accuracy for normo-, hypo-, and hyperandrogenic women.

**Supplementary data**

This is linked to the online version of the paper at http://dx.doi.org/10.1530/EJE-11-0482.

**Declaration of interest**

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

**Funding**

This work was supported by an unrestricted research grant of National Health and Medical Research Committee (NHMRC), Australia, to T E Hickey and R J Norman.

**Acknowledgements**

We would like to thank Andrew Dinan at CPR Pharma Pty Ltd, Adelaide, for his assistance in the measurement of androgens by LC–MS/MS. We would like to thank Dr Yvonne M van Kasteren (Medical Center Alkmaar) and Dr Peter A van Dop (Catharina Hospital, Eindhoven) for the inclusion of two and seven patients with POI.

**References**


5 Burger HG. Androgen production in women. *Fertility and Sterility* 2002 77 3–5. (doi:10.1016/S0015-0282(02)02985-0)


Received 31 May 2011
Revised version received 28 September 2011
Accepted 3 October 2011