CLINICAL STUDY

Defining hyperandrogenism in polycystic ovary syndrome: measurement of testosterone and androstenedione by liquid chromatography–tandem mass spectrometry and analysis by receiver operator characteristic plots

Julian H Barth¹, Helen P Field¹, Ephia Yasmin² and Adam H Balen²

Departments of ¹Clinical Biochemistry and ²Reproductive Medicine, Leeds General Infirmary, Leeds LS1 3EX, UK

(Correspondence should be addressed to J H Barth; Email: julian.barth@leedsth.nhs.uk)

Abstract

Objective: Hyperandrogenism is one of the diagnostic criteria for the polycystic ovary syndrome (PCOS) despite no agreed definition of hyperandrogenism. In part, this is due to the quality of testosterone immunoassays. We have developed liquid chromatography–tandem mass spectrometry methods for analysing testosterone and androstenedione (Ad) to study their reference ranges and diagnostic utility in PCOS.

Design, setting and subjects: A consecutive series of 122 women attending a reproductive medicine clinic.

Methods: Blood samples were taken during the early follicular phase for measurement of LH, FSH, oestradiol, Ad, testosterone and sex hormone-binding globulin (SHBG). Retrospective case note analysis was used to determine the clinical features and ultrasound findings.

Results: The incidence of PCOS was 13.9%. The reference interval for testosterone was $1.8 \text{ nmol/l}$ and for Ad was $1.4–7.4 \text{ nmol/l}$. There were significant differences in total testosterone ($P<0.001$), Ad ($P<0.05$) and free androgen index (FAI; $P<0.0001$) between the women with and without PCOS. Diagnostic performance using receiver operator characteristic plots showed area under the curve (AUC) for FAI 0.81, testosterone 0.75 and Ad 0.66. The AUC for the LH:FSH ratio was 0.72.

Conclusions: Our analysis of a consecutive series of women attending a reproductive clinic has provided an appropriate series on which to construct reference ranges for key androgens in women. Secondly, it has allowed us to conclude that early follicular serum testosterone measured using tandem mass spectrometry, FAI and the LH:FSH ratio are valuable laboratory tests in the diagnosis of PCOS.

European Journal of Endocrinology

Introduction

All guidelines and consensus statements on the diagnosis of polycystic ovary syndrome (PCOS) use the terms clinical and biochemical hyperandrogenism in their definitions. Moreover, the terms hyperandrogenism and hyperandrogeaemia are both used to describe increased concentrations of serum androgens. This definition is unclear and creates a degree of uncertainty as shown by the majority of reported studies of women with PCOS, which have typically measured multiple androgens such as total and free testosterone, androstenedione (Ad), dehydroepiandrosterone (DHAS), sex hormone-binding globulin (SHBG) and the free androgen index (FAI). This has occurred despite the concern of the analytical accuracy of total testosterone measurement (1) and the variability of different formulae for assessing free testosterone (2). Moreover, there has been little clarity over the definition of the limits of hormone concentrations in ‘normal’ women.

The dependence on multiple androgen measurements may stem back to the days when RIA showed poor performance and there was a hope that multiple assays would give more information. However, the development of liquid chromatography–tandem mass spectrometry (LC–MS/MS) and its introduction into clinical practice should obviate this need (3). It is for this reason that we recently proposed that a single testosterone measurement taken in the early follicular phase should be used for the diagnosis of PCOS (4). We now have used this assay technology to define the reference intervals of testosterone and Ad in women without PCOS and assessed its diagnostic accuracy in women with PCOS.

Methods

Subjects

One hundred and twenty-two consecutive women attending the Reproductive Medicine Unit at the Leeds General Infirmary were identified from consecutive blood samples received in the endocrinology laboratory. The blood samples were not taken during the clinic visit.
but at a later time to coincide with days 1–5 of their next menstrual cycle. The diagnoses of the patients were made retrospectively by examination of the case records (EY). The following features were noted: menstrual history; presence of acne; hirsuties; body mass index (BMI); and early follicular phase hormones. A diagnosis of polycystic ovaries was based on the presence of 12 or more follicles measuring 2–9 mm and/or ovarian volume measuring >10 cm³ (Table 1).

The ultrasound examinations were performed transvaginally in the early follicular phase of the menstrual cycle in menstruating women or after a withdrawal bleed in oligomenorrhoeic and amenorrhoeic women. Follicles were counted and the ovarian volumes were measured using the formula of the prolate ellipsoid (0.5×length×width×thickness). In non-ellipsoid ovaries, the ovary was outlined manually in the scan and automated calculation applied (5).

The diagnosis of PCOS was based on the Rotterdam (Joint European Society for Human Reproduction and Endocrinology (ESHRE)/American Society for Reproductive Medicine (ASRM)) consensus (6) criteria. The Rotterdam criteria require two of the following three features for the diagnosis of PCOS: oligo-ovulation or amenorrhoea; clinical or biochemical evidence of hyperandrogenism; polycystic ovaries by ultrasound examination as well as exclusion of other causes. However, in order to be able to develop hormonal data on which to diagnose PCOS, the testosterone data were not used to define subjects.

**Hormone assays**

Testosterone was measured in 50 μl serum by LC–MS/MS as previously described (3), after extraction with 1 ml HPLC grade methyl tert-butyl ether (MTBE, supplied by Fisher Scientific, Loughborough, UK). The transition 291>111 was used for dideuterated testosterone. Recovery of testosterone was assessed using external quality assessment (UKNEQAS) samples and ranged from 101 to 105%. Inter-assay precision was 9.34% at 0.79 nmol/l, 5.72% at 2.76 nmol/l and 4.4% at 5.88 nmol/l (Δ = 10).

Ad was measured in 100 μl serum by LC–MS/MS after extraction with 1 ml MTBE. Calibrators were prepared by weighing out Ad (stated purity ≥98%, Sigma) on a six-place balance and dissolving in Analar grade ethanol (Hayman Ltd, Witham, UK) to give a stock solution of 5 mmol/l. This stock solution was diluted in HPLC grade water (Fisher Scientific) to give a series of solutions that were added to aqueous methanol, 1:1 solution by volume, to make seven calibrator solutions ranging from 0.5 to 50 mmol/l. Calibrators were stored at 4 °C before use. The response of the calibrators was linear between 0.5 and 50 nmol/l. Serum samples, calibrators or quality control samples (Lyphochek Immunoassay plus control levels 1, 2 and 3, Bio-Rad Laboratories) were pipetted into washed disposable glass tubes (12×75 mm). Internal standard (50 μl of 3 mmol/l heptadeuterated Ad in HPLC grade water (d7Ad, supplied by C/D/N Isotopes Inc., Point-Claire, Quebec, Canada)) was added to each tube, and the contents were mixed for 30 s.

Ad was extracted into 1 ml MTBE by vortex mixing for 5 min. After the aqueous layer was frozen in a dry ice–methanol bath, the organic layer was decanted into a washed disposable glass tube (10×75 mm). Solvent was removed using a vacuum oven at 35–40 °C; then, the dried residue was reconstituted in 100 μl aqueous methanol (1:1 solution by volume). After 5 min at room temperature, the extracts were vortex mixed and then transferred to plastic 1.5 ml microcentrifuge tubes (Sarstedt, Leicester, UK). The autosampler plates were sealed as for the testosterone assay. The equipment, HPLC reagents and LC column used for the Ad assay were as described for the testosterone assay (3). Thirty microlitres of the extract were injected into the column, held at 40 °C. The initial mobile phase conditions were a 50:50% mixture of HPLC grade water (solvent A) and HPLC grade methanol (solvent B), each containing ammonium acetate (2 mmol/l, SigmaUltra, Sigma) and formic acid (1 ml/l, BDH Aristar grade, VWR).

### Table 1 Patient characteristics.

<table>
<thead>
<tr>
<th></th>
<th>Non-PCOS (n = 107)</th>
<th>PCOS (n = 17)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (s.d.)</td>
<td>Median (IQR)</td>
<td>Mean (s.d.)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>33.2 (5.43)</td>
<td>33.0 (29.0–37.0)</td>
<td>31.35 (4.85)</td>
</tr>
<tr>
<td>BMI</td>
<td>24.26 (3.97)</td>
<td>24.00 (21.00–26.00)</td>
<td>27.04 (5.74)</td>
</tr>
<tr>
<td>Testosterone (nmol/l)</td>
<td>0.66 (0.40)</td>
<td>0.56 (0.40–0.80)</td>
<td>1.01 (0.51)</td>
</tr>
<tr>
<td>Androstenedione (nmol/l)</td>
<td>3.56 (1.56)</td>
<td>3.25 (2.55–4.52)</td>
<td>4.32 (1.59)</td>
</tr>
<tr>
<td>SHBG (nmol/l)</td>
<td>59.2 (26.5)</td>
<td>52.2 (41.6–73.2)</td>
<td>50.3 (29.2)</td>
</tr>
<tr>
<td>FAI</td>
<td>1.22 (0.83)</td>
<td>1.01 (0.66–1.56)</td>
<td>2.39 (1.23)</td>
</tr>
<tr>
<td>Oestradiol (pmol/l)</td>
<td>188 (104)</td>
<td>163 (126–219)</td>
<td>222 (79)</td>
</tr>
<tr>
<td>LH (IU/l)</td>
<td>5.39 (4.89)</td>
<td>4.70 (3.6–6.0)</td>
<td>6.38 (3.7)</td>
</tr>
<tr>
<td>FSH (IU/l)</td>
<td>7.51 (6.87)</td>
<td>6.40 (5.2–7.8)</td>
<td>5.24 (1.63)</td>
</tr>
<tr>
<td>LH/FSH ratio</td>
<td>0.77 (0.34)</td>
<td>0.69 (0.53–0.98)</td>
<td>1.20 (0.66)</td>
</tr>
</tbody>
</table>

One hundred and twenty-two consecutive patients; 17 defined as having PCOS using the Rotterdam consensus criteria (see Methods section). Significance was tested using Mann–Whitney U test.
Lutterworth, Leicestershire, UK). The flow rate was 0.6 ml/min. A linear gradient was run to 5% A:95% B over 5 min; then, the column was washed with 95% B at 1 ml/min and then reequilibrated with 50% A:50% B. Column eluate was diverted from the mass spectrometer during the first 2 min of the run and then later during the column washing stage. The retention time for Ad and d7Ad was 4.8 min, and the total run time was 7 min. The MS/MS settings in positive ion mode for Ad were capillary voltage 0.8 kV, cone voltage 25 V and collision energy 20 eV. The multiple reaction monitoring transitions used for Ad and d7Ad were m/z 287 → 97 and 294 → 100 respectively. Ion suppression was assessed by post-column infusion and occurred only between 0.8 and 1.2 min. The limit of quantification of Ad was 0.5 nmol/l using the definition of signal:noise ratio $R \geq 10$ and coefficient of variation (CV) $\leq 20\%$. The recovery of Ad ranged from 87 to 111%. Intra-assay imprecision for Ad in patient pools was 6.4% at 1.1 nmol/l, 4.3% at 2.3 nmol/l, 4.5% at 5.2 nmol/l and 5.7% at 10.3 nmol/l. Inter-assay imprecision for Ad in quality control material was 8.8% at 2.9 nmol/l, 7.3% at 8.7 nmol/l and 5.4% at 21.5 nmol/l. Extracts were stable for 10 days at 4 °C.

Gonadotrophins and oestradiol (OE2) were measured by chemiluminescence on an Advia Centaur (Siemens Medical Solutions, Camberley, UK) with typical between-batch CV: OE2 18% at 52 pmol/l, FSH 7.0% at 6.8 IU/l, LH 6.0% at 2.5 IU/l. SHBG was measured by chemiluminescence on Siemens Immulite 2000 with CV 7.4% at 78 nmol/l. FAI was calculated using the formula: (testosterone/SHBG) × 100.

**Statistical analysis**

Statistical tests were performed using the Analyse-it version 2.08 add-in package for Microsoft Excel (www.analyse-it.com).

**Results**

The incidence of PCOS in our consecutive cohort was 13.9% (17/122). The patient characteristics and endocrinology are tabulated in Table 1. There is a significant difference in total testosterone and Ad concentrations and the FAI and LH:FSH ratio between the women with and without PCOS (Fig. 1).

The distribution of all androgens was non-Gaussian using the Anderson–Darling A2 test. Data were log-transformed before calculations were performed. The reference intervals for women without PCOS were: testosterone <1.8 nmol/l, Ad 1.4–7.4 nmol/l, SHBG 22–129 nmol/l and FAI 0.30–3.36. There were trivial differences when the ten women with BMI >30 were excluded.

Analysis by receiver operator characteristic (ROC) was used to determine the diagnostic performance of the androgens. The AUC for each androgen was: FAI 0.81, testosterone 0.75, Ad 0.66 and SHBG 0.63. There was a significant difference between the AUC for FAI and SHBG ($P < 0.005$), but not between the other variables. The AUC for LH:FSH ratio was 0.72 and LH 0.62 (LH:FSH ratio > LH, $P < 0.01$).

**Discussion**

There have been several attempts over the past years to define PCOS. Each of these definitions uses the term hyperandrogenism and hyperandrogenaemia, yet no definition is given for these terms despite the enormous subjective and physiological variations (6–8). We have previously proposed that a single blood sample for serum testosterone should be taken in the morning during the early follicular phase as the diagnostic criterion for hyperandrogenaemia (4). In this study, we have evaluated the diagnostic utility of testosterone, Ad and gonadotrophins in such early follicular samples using state of the art technology to determine the best laboratory strategy to diagnose PCOS.

The introduction of LC–MS/MS into routine practice has permitted precise and accurate measurements of serum androgens for the first time. In particular, it means that testosterone can be accurately measured in the range <5 nmol/l that is typically seen in women. Furthermore, it has been used to show that the inaccuracy of testosterone immunoassay is in part due to cross-reactivity with DHEAS (9, 10). This represents a significant proportion of the testosterone measured with immunoassays and has made it necessary to define the upper reference value found in
normal women. The situation for Ad is further beset by a lack of international standard preparations to compare methods. The development of Ad methods by LC–MS/MS increases the chance of harmonisation between different laboratories. The use of LC–MS/MS to measure serum steroids can then be used to standardise a key diagnostic criterion on which to define the term ‘hyperandrogenism’ in the definition of PCOS and make clinical decisions.

We have chosen to study a consecutive series of women attending a Reproductive Medicine Unit rather than a series of randomly recruited women. We believe that this removes recruitment bias, as the subjects represent a group of women presenting with the appropriate gynaecological and dermatological symptoms that might lead to the consideration of PCOS as a diagnosis. This group would therefore appear to be ideal for testing the diagnostic performance of testosterone and Ad measured by LC–MS/MS. The case mix of consecutive patients attending a reproductive medicine clinic might have been expected to show a rather high proportion of subjects with PCOS, but our population had an incidence of only 13.9%. This is greater than unselected women in previous studies but similar to the Goodarzi et al. (11) study and significantly lower than a series of normal women. The situation for Ad is further beset by a lack of international standard preparations to compare methods. The development of Ad methods by LC–MS/MS increases the chance of harmonisation between different laboratories. The use of LC–MS/MS to measure serum steroids can then be used to standardise a key diagnostic criterion on which to define the term ‘hyperandrogenism’ in the definition of PCOS and make clinical decisions.

Table 2 Published studies of the incidence of polycystic ovary syndrome (PCOS).

<table>
<thead>
<tr>
<th>Location</th>
<th>Recruitment</th>
<th>Number</th>
<th>PCOS (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leeds, UK</td>
<td>Consecutive attenders at reproductive medicine clinic</td>
<td>122</td>
<td>13.9</td>
</tr>
<tr>
<td>US (21)</td>
<td>Consecutive attenders at employment medical</td>
<td>400 (608 total cohort)</td>
<td>6.6</td>
</tr>
<tr>
<td>Lesbos (22)</td>
<td>Open unspecified invitation</td>
<td>192</td>
<td>6.8</td>
</tr>
<tr>
<td>Madrid (23)</td>
<td>Consecutive blood donors</td>
<td>154</td>
<td>6.5</td>
</tr>
<tr>
<td>Oxford (15)</td>
<td>Open unspecified invitation</td>
<td>230</td>
<td>8+</td>
</tr>
<tr>
<td>Mexican–Americans (11)</td>
<td>Unselected consecutive subjects in MACAD study (FH Ischaemic heart disease)</td>
<td>156</td>
<td>13</td>
</tr>
<tr>
<td>Madrid (24)</td>
<td>Consecutive obesity clinic attendees</td>
<td>113</td>
<td>28</td>
</tr>
</tbody>
</table>

All studies cited in this table (except this study) have used the NIH criteria for diagnosing PCOS rather than the Rotterdam (2003) criteria. *When these data were reanalysed using the Rotterdam criteria, the incidence of PCOS rose to 26% (see Discussion section).*

There is much debate in the literature regarding the diagnostic utility of LH and the LH:FSH ratio in PCOS. The Rotterdam consensus criteria proposed that LH had no role in the diagnosis (6). This is supported by two studies: firstly, Cho et al. (17) found that the LH:FSH ratio has poor within-person reproducibility in women with or without proven PCOS; secondly, Escobar-Morreale et al. (18) using consecutive series of women showed that LH and FSH had practically no diagnostic utility, although samples were taken at random times within the menstrual cycle. These two studies contrast with our findings where LH/FSH is significantly elevated in samples taken during the early follicular phase. This may be due to different endocrine pathologies in slim and obese women with PCOS. In slim women, episodic pituitary hypersecretion of LH is the cause of hyperandrogenism, whereas in the obese women, elevated LH is dependent upon hyperinsulinaemia. This would be supported by the studies of obese women with PCOS who do have raised LH:FSH ratios (19, 20).

ROC curves are used to assess sensitivity and specificity across the dynamic range of an analyte concentration. However, these two variables are quite dependent upon the case mix. Therefore, it is necessary to have a patient series that reflects the population under study and not to pick patients and controls in artificial proportions. ROC curves have been used to evaluate diagnostic tests in PCOS, but the only study we have identified that has used an appropriate series of consecutive patients is that of Escobar-Morreale et al. (17). They measured testosterone and SHBG using the Immulite methods, and identified SHBG and FAI as the best assays with AUC 0.875 and 0.867. Their population had a similar BMI range to our population.

Our analysis of a consecutive series of women attending a reproductive clinic has provided an appropriate series on which to construct a reference range for androgens in women. Secondly, it has allowed us to conclude that early follicular serum testosterone measured using tandem MS, FAI and LH:FSH ratio are valuable laboratory tests in the diagnosis of PCOS.

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 research did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.
Author contribution statement

J H Barth designed the study. H P Field developed the LC–MS/MS methods and clinical information was collected by E Yasmin and A H Balen. All authors have reviewed and approved the final report.

References