Published in final edited form as:

Fertil Steril. 2014 April ; 101(4): 1135–1141.e2. doi:10.1016/j.fertnstert.2013.12.056.

Specificity and Predictive Value of Circulating Testosterone Assessed By Tandem Mass Spectrometry for the Diagnosis of Polycystic Ovary Syndrome by the NIH 1990 Criteria

W. Antoine Salameh, M.D.^{1,*}, Mildred M. Redor-Goldman, BS.¹, N.J. Clarke, Ph.D.¹, R. Mathur, M.D.^{2,3}, R. Azziz, M.D.², and R.E. Reitz, M.D.¹

¹Quest Diagnostics Nichols Institute, San Juan Capistrano, CA

²Dept. of Ob/Gyn and Center for Androgen Related Disorders, Cedars-Sinai Medical Center, Los Angeles, CA

³Department of Medicine, Cedars Sinai Medical Center, Los Angeles, CA

Abstract

Objective—To determine whether liquid chromatography–tandem mass spectrometry (LC-MS/MS) determination of total (TT) and free (FT) testosterone is more specific than extraction chromatography–radioimmunoassay (RIA) for distinguishing women with polycystic ovary syndrome (PCOS) from controls, and whether differing cut-off values should be used depending on the setting.

Design—Prospective case-control cohort study.

Setting—Tertiary care center and reference laboratory.

Interventions—Blood sampling.

Participants—100 PCOS patients and 100 controls.

Main Outcome Measure—Circulating TT measured by RIA and LC-MS/MS, and FT measured by equilibrium dialysis using the RIA or LC-MS/MS-generated TT values.

Results—Testosterone measurement by RIA and LC-MS/MS differentiated PCOS from controls similarly; detection of PCOS was higher for FT for both methods. TT values demonstrated greater overlap between PCOS and controls than did FT for both RIA (80% vs. 42% overlap) and LC-MS/MS (52% vs. 67% overlap). A lower cut-off value was better suited for the study of patients

© 2014 American Society for Reproductive Medicine. Published by Elsevier Inc. All rights reserved.

***Corresponding Author/Reprint Requests:** Wael, Antoine Salameh, M.D., Clarus Therapeutics, Inc., 555 Skokie Blvd., Suite 340, Northbrook, IL 60062.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Disclosure Statement: WA Salameh: salary and stock - Quest Diagnostics, MM Redor-Goldman: salary and stock - Quest Diagnostics, MP Caufield: stock and salary - Quest Diagnostics, NJ Clarke: salary and stock - Quest Diagnostics, R. Mathur: nothing to declare, R. Azziz: nothing to declare, RE Reitz: salary and stock in Quest Diagnostics.

seen in the clinical (referred) setting (35 ng/dL and 4.0 ng/dL for TT and FT, respectively) than in the screening of a general population (50 ng/dL and 5.0 ng/dL for TT and FT, respectively).

Conclusions—Extraction chromatography and RIA and LC-MS/MS measurements of testosterone have similar performance for differentiating PCOS from healthy controls; LC-MS/MS may be preferable given its relative ease of automation. Compared to FT, measurement of TT has relatively limited value for distinguishing PCOS from normal. Finally, different cut-off values should be considered depending on the clinical/investigative setting, with higher values used in the study of biased (e.g. clinical or referred) populations.

Keywords

Polycystic Ovary Syndrome; androgens; testosterone; radioimmunoassays; mass spectrometry

INTRODUCTION

Polycystic ovary syndrome (PCOS) is the most prevalent endocrinopathy in women of reproductive age. One of the hallmarks of PCOS is the presence of hyperandrogenism, either clinical or biochemical (1). The detection of hyperandrogenemia in women with PCOS is most often based on an accurate and sensitive measurement of total (TT) and free (FT) testosterone. Although frequently used, platform or direct assays for the determination of TT and FT are unreliable for the determination of levels below the normal male range, such as when used for the evaluation of hyperandrogenic women, children, and hypogonadal men (2-10). The Endocrine Society recently issued a position statement detailing the limitation of analogue assays for testosterone measurement and recommended, in most cases, that methods involving mass spectrometry (MS) or organic extraction, column chromatography, and radioimmunoassay (RIA) be used when testosterone levels are expected to be low (6).

Over the past 30 years, prior to the introduction of MS, the accurate measurement of testosterone and other sex steroids for clinical diagnostic purposes required the use of methods involving organic extraction, column chromatography, and RIA. The specificity of these methods relies on the completeness of the separation of closely related steroids in the chromatography step and the specificity of the primary antibody employed. The availability of these reference assays has declined recently with the development of automated direct immunoassays and is now limited to a few reference laboratories. These reference assays (i.e. those involving organic extraction, column chromatography, and RIA) have their own limitations as they are labor intensive, require a greater degree of technical skill, and are still subject to antibody supply and specificity. Of note, androgen assays using commercially available platform and kit assays provide results that vary widely from those obtained by reference methods (11) and a comparison of the results of commercial assays to reference assays cannot be made.

While TT levels in women with suspected PCOS can be accurately measured by extraction chromatography and RIA, with the use of equilibrium dialysis for FT determination, the quantification of steroids using mass-spectrometry-based methodologies has been proposed as the current gold standard for androgen assays (6; 12). The Endocrine Society position statement on testosterone assays called for further study validating the use of these newer

MS-based assay methods for the evaluation of women, children, and hypogonadal males (6), and a recent editorial by the Editor-In-Chief of the Journal of Clinical Endocrinology & Metabolism called for a ban on the use of any sex steroid assays other than those based on MS (12).

Distinct from the methodology used to detect the steroid or marker in question, is the question of what cut-off value for the detection of a diseased state will be used. Clinicians and researchers alike are frequently under the misconception that cut-off values are a fixed parameter, to be used uniformly regardless of what the population being studied or aim of the assay are. However, it should be noted that the cut-off values for detecting ‘disease’ greatly depends on the background prevalence of the ‘disease’ in the population studied, which in turn may depend on the aim of the study itself.

To select useful cut-off values, it will be important to determine what statistical method will be used (e.g. mean and standard deviation, percentile, cluster analysis, etc.), whether the cut-off value will vary according to various phenotypic factors such as age, body mass, or race, and how well the control population is identified (e.g. individuals from the general population vs. a those from a well-defined healthy population). In addition and frequently overlooked in establishing the cut-off value used to distinguish diseased from healthy is the setting, is the prevalence of the disorder in the population being studied. For example, the prevalence of PCOS in unselected populations of women approximates 10% (13-17), while the prevalence of PCOS among subjects seen in referred (biased or clinical) populations may be as high as 70% (18,19). Thus, it is likely that a different cut-off value will be needed depending on the setting, i.e. the expected prevalence of the disorder (e.g. PCOS) in the population being assessed.

In the present study we analyzed the predictive value of TT and equilibrium dialysis-derived FT, determined by measuring T using either serum extraction-chromatography-RIA or liquid chromatography tandem mass spectrometry (LC-MS/MS), in 100 well-phenotyped normal and 100 PCOS females. Furthermore, we modeled two different scenarios; one reflecting subjects seen in the clinical setting (a biased or referred population assumed to have a ~70% prevalence of PCOS) and the other reflecting subjects detected in a populational study (assumed to have a ~10% prevalence of PCOS).

MATERIAL AND METHODS

Subjects

PCOS patients—Serum samples from 100 women with PCOS were studied. The age range of the subjects was 14 to 46 years (28.2 ± 7.1 years) and their body mass index (BMI) ranged from 16.9 to 63.0 kg/m² (37.0 ± 9.1 kg/m²). Of the subjects studied, 78 were white, 17 were black, and the remainder was of other races. The presence of PCOS was defined according to the NIH 1990 criteria including: 1) clinical evidence of hyperandrogenism and/or hyperandrogenemia; 2) oligo-ovulation; and 3) exclusion of related disorders (e.g., congenital adrenal hyperplasia, thyroid dysfunction, hyperprolactinemia) (20). Ovulatory dysfunction was defined as menstrual cycles of greater than 45 days in length or less than 8 cycles per year, or by a luteal phase (cycle day 22-24) progesterone level of less than 4

mg/mL [12.72nmol/L] if cycles were less than 45 days in length. Hyperandrogenism in 76 of the 100 PCOS women was based on the presence of hirsutism, defined as a modified Ferriman-Gallwey hirsutism score of ≥ 6 (21); in 24 patients hyperandrogenism was based on prior results of TT, FT, and DHEAS assays, as previously described (21). For the original diagnosis, TT was measured by an in-house RIA method following serum extraction and chromatography; sex hormone binding globulin (SHBG) activity was measured by competitive binding using Sephadex G-25 with tritiated testosterone as the ligand and FT was calculated. DHEAS was measured by direct RIA using a commercially available kit (Diagnostic Products Corp., Los Angeles, CA) (11).

Controls—Serum samples from 100 healthy control women were studied. Controls were defined as healthy non-pregnant, non-hirsute, premenopausal, eumenorrheic women without personal or family history of hirsutism and/or endocrine disorders. Controls were recruited by responding to posted advertisements. The subjects ranged in age from 15 to 58 years (mean and SD of 33.2 ± 9.8 years) and in BMI from 16.6 to 58.7 kg/m^2 ($27.5 \pm 6.9 \text{ kg/m}^2$). Of the 100 controls studied, 53 were white and 47 were black.

Study protocol

PCOS and control subjects were not taking any medications that could impact hormonal levels for at least 3 months prior to blood collection, and all underwent a history and physical examination. A fasting blood sample was obtained during the follicular phase (cycle days 3-8) of the menstrual cycle or, if oligo-amenorrheic, days 3-8 after a withdrawal bleed was induced with oral micronized progesterone. Serum samples were kept frozen at -80°C until thawed for analysis. All subjects were recruited either at the University of Alabama at Birmingham (UAB) or at Cedars-Sinai Medical Center (CSMC); the study was approved by the Institutional Review Boards for Human Protection of UAB and CSMC. Written consent was obtained from all subjects.

TT and FT Assays

Reagents—Testosterone was purchased from Sigma-Aldrich® Inc (St Louis, MO). Testosterone-2,2,4,6,6-d₅ was purchased from CDN Isotopes (Quebec, Canada). Each of these was 98 % pure. All other chemicals were of reagent grade or better and purchased from commercial sources.

Total Testosterone by High-Turbulence LC-MS/MS—This method was performed as previously described (22). Briefly preanalytical processing of samples included acidification, addition of internal standard, autosampling, and passing through a high-turbulence liquid chromatography (HTLC) column followed by passage through a C12 analytical HPLC column. TT was then quantitated using a Finnigan TSQ Quantum Ultra (ThermoFisher, San Jose, CA) tandem mass spectrometer. The sensitivity of the assay (LOQ), set at a coefficient of variation (CV) of 20%, was 0.3 ng/dL [0.01 nmol/L]. The CV for all reproducibility data was $<15\%$ and ranged from 7.6% to 10.8% for intra-assay and 9.8% to 13.4% for inter-assay variability.

Total Testosterone by Extraction Chromatography RIA—TT was determined using an independently validated extraction and chromatography RIA (Quest Diagnostics Nichols Institute, San Juan Capistrano, CA), which employs 0.5 mL of serum, as previously described (22). This RIA has a limit of quantitation of 2 ng/dL [0.069 nmol/L] at 20% CV and is linear to 1800 ng/dL [62.46 nmol/L] (linearity is extended by extraction of smaller volumes for levels above 200 ng/dL [6.94 nmol/L]). At the low level of 2.5 ng/dL [0.087 nmol/L], the intra-assay CV is 11.8 % while the inter-assay CV is 17.0 %.

Free Testosterone by Equilibrium Dialysis—Serum samples were diluted as previously described (22) in the presence of tritiated testosterone as the tracer. The diluted sample was placed on the sample side of the dialysis chamber and the same physiological salt buffer without tracer was placed on the other (buffer) side. The tracer was allowed to reach equilibrium with the endogenous testosterone and the binding proteins. Following incubation a sample from each side of the membrane was removed, and the amount of tracer present in each was determined. The FT concentration was then calculated from the percent of tracer that crossed to the other side of the chamber and the TT concentration as determined by one of the methods outlined above (22).

Statistical Analyses

For methods comparison between RIA and LC-MS/MS for TT or FT, Deming regression, an orthogonal regression algorithm, was used; unlike linear regression, Deming regression assumes there is uncertainty in both X and Y variables. Bland-Altman plots were used for bias analysis, which plots the difference between the two measurements on Y axis and the average of the two measurements on the X axis. The bias, or difference between x and y, is expressed as percentage of average of x and y in Bland-Altman plots ($\text{bias} = [x - y] / [0.5 \times (x + y)]$). These analyses were performed using a commercially available software package (Analyse-it™ Software, Ltd., Leeds, United Kingdom; www.analyse-it.com).

The association between PCOS and TT or FT was investigated using logistic regression for both LC-MS/MS and RIA measurements. PCOS was predicted against controls at different probability levels 0 through 1. With sensitivities and 1-specificities at these various probabilities, receiver operating characteristic (ROC) curves were generated and the area-under curve (AUC) for the ROC curves was calculated. Here $x = \text{RIA}$ and $y = \text{LC-MS/MS}$, or vis versa, depending on the variable used for the x axis.

To determine the best cutoff values to assess performance of TT and FT to classify phenotypic PCOS cases, we chose the following criteria: hormonal classification of a subject as PCOS was done only if TT and FT or just FT were elevated for any given cutoff. Cases with just high TT, but normal or low FT, were not classified as PCOS, as the elevation of TT in such cases might be due to elevated SHBG. In this analysis we also considered two different clinical situations: i) the evaluation of women with suspected androgen excess in the clinical setting, where a 70% prevalence of PCOS can be expected (18, 19); and ii) the evaluation of unselected women from a broader population, as in epidemiologic studies, where a 10% prevalence of PCOS may be expected (13-17).

RESULTS

Correlation between TT and FT results obtained with LC-MS/MS and extraction chromatography and RIA

All TT and FT values in both controls and women with PCOS were above the detection limits of the respective assays (not shown). Linear regression analysis showed high correlation between extraction chromatography RIA and LC-MS/MS for measurement of TT ($r^2 = 0.899$) and FT ($r^2 = 0.92$). Deming regression demonstrated a small but significant positive bias for TT measurement by RIA, with a Deming fit of $y = -7.37 + 0.96x$; LC-MS/MS values showed a constant negative intercept of -7.37 accentuated by a negative slope of 0.96 (Supplemental Fig. 1A). A similar trend was observed for FT, with a Deming fit of $y = -0.61 + 0.89x$ indicating a negligible negative intercept of -0.61 ; however, the slope of 0.89 indicated a greater proportional bias (Supplemental Fig. 1B). Overall, the TT and FT values calculated based on LC-MS/MS TT values were lower than the corresponding values by RIA. This positive bias for the RIA method is also visually observable for both PCOS and control subjects (Supplemental Figs. 2A & 2B).

There was considerable overlap in androgen values between PCOS and control subjects; for TT by RIA, 80% of values overlapped while for TT by LC-MS/MS, 67% overlapped. Corresponding figures for FT by RIA and LC-MS/MS were 42% and 52%, respectively. These figures indicate a greater degree of overlap in the measurement of TT than FT, independent of assay method.

The ROC curves for detecting PCOS using TT were very similar with the LC-MS/MS (AUC = 0.765) and RIA (AUC = 0.773) methods (Fig. 1A). The ROC curves for detecting PCOS were similar for FT as determined using TT measured by LC-MS/MS (AUC = 0.870) and RIA (AUC = 0.891) (Fig 1B). Note that the AUC values were higher for FT than for TT for both LC-MS/MS and RIA analysis. The highest AUC value was found when combining TT and FT values: 0.895 by LC-MS/MS and 0.901 by RIA (Fig 1C) —although these values were not statistically different from the AUC for FT alone.

Estimating cut-off values for TT and FT using LC-MS/MS for the assessment of two different clinical situations

We then calculated the positive and negative predictive values (PPV and NPV, respectively) for various cutoff values for TT alone (Table 1), FT alone (Table 2), and TT and FT combined (Table 3), as measured by the LC-MS/MS method. In this analysis we considered two different clinical situations: i) the evaluation of women with suspected androgen excess in the clinical setting, where a 70% prevalence of PCOS can be expected (18, 19); and ii) the evaluation of unselected women from a broader population, as in epidemiologic studies, where a 10% prevalence of PCOS may be expected (13-17). We then assessed the optimum combination of PPV and NPV values by calculating the incremental (percentage) value added, and considering a change in PPV to be twice the weight of a similar change (percentage) in NPV.

Firstly, these data highlighted the relative weakness of TT alone as a predictor of PCOS (Table 1). Secondly, the data indicated that the optimum cutoff values for both TT and FT,

resulting in the highest combined NPV and PPV values, were generally dependent on the background prevalence of PCOS expected. For example, and considering that a gain in PPV has twice the weight as a loss in NPV:

1. In the general population screening (10% prevalence) setting:

a. For TT (Table 1):

- i.** Increasing the cut-off value from 35 ng/dL to 40 ng/dL, changed the PPV from 47% to 54%, a 15% improvement, and it did not change the NPV (at 94%). Increasing the cut-off value further from 40 ng/dL to 45 ng/dL changed the PPV from 54% to 66%, an 18% improvement; it resulted in a decrease in NPV from 94% to 93%, a 1% reduction. Finally, increasing the cut-off value further from 45 ng/dL to 50 ng/dL increased PPV from 66% to 100%, a 51% improvement and did not change the NPV (at 93%).

b. For FT (Table 2):

- i.** Increasing the cut-off value from 3.3 ng/dL to 3.5 ng/dL increased the PPV from 38% to 47%, a 24% improvement; it did not decrease the NPV (at 97%). Increasing the cut-off value further, from 3.5 ng/dL to 4.0 ng/dL, increased PPV from 47% to 70%, a 50% improvement and resulted in a decrease in the NPV from 97% to 96%, a 1% reduction. Increasing the cut-off value from 4.0 ng/dL to 4.5 ng/dL increased PPV from 70% to 76%, a 8% improvement and resulted in a decrease in NPV from 96% to 95%, a 1% reduction. Finally, increasing the cut-off value further from 4.0 ng/dL to 4.5 ng/dL increased PPV from 76% to 85%, a 12% improvement; it did not change the NPV from 96% to 95%, a 1% reduction.

2. In the clinical (70% prevalence) setting:

a. For TT (Table 1):

- i.** Increasing the cut-off value from 33 ng/dL to 35 ng/dL increased the PPV from 93% to 95%, a 2% improvement; it decreased the NPV from 45% to 44%, a 1% reduction. Alternatively, increasing the cut-off value from 35 ng/dL to 40 ng/dL increased the PPV from 95% to 96%, a 1% improvement; it resulted in a decrease in NPV from 44% to 42%, a 2% reduction.

b. For FT (Table 2):

- i.** Increasing the cut-off value from 3.3 ng/dL to 3.5 ng/dL increased the PPV from 93% to 95%, a 2% improvement; it decreased the NPV from 58% to 57%, a 2% reduction.

Increasing the cut-off value further from 3.5 ng/dL to 4.0 ng/dL increased PPV from 95% to 98%, a 3% improvement; it resulted in a decrease in NPV from 57% to 54%, a 5% reduction.

- ii. Alternatively, increasing the cut-off value further from 4.0 ng/dL to 4.5 ng/dL increased PPV from 98% to 99%, a 1% improvement; however, it resulted in a decrease in NPV from 54% to 50%, a 8% reduction.

Thus, cut-off values of 50 ng/dL and 5.0 ng/dL appear to be optimum for TT measured by LC-MS/MS, and for the derived FT, respectively, in the screening of a general population for PCOS (with an expected ~10% prevalence of the disorder). Alternatively, cut-off values of 35 ng/dL and 4.0 ng/dL for TT and FT, respectively, are optimum in the clinical arena (where a ~70% prevalence of PCOS can be expected). A similar analysis using TT and FT combined demonstrated the same optimum cut-off values (Table 3).

DISCUSSION

For clinical and investigational purposes in the evaluation of androgen excess disorders, testosterone assays must be accurate, reproducible, and sensitive. There is a growing concern about the accuracy of clinically reported testosterone results, particularly in the lower concentrations, such as those seen in women including those with androgen excess disorders (6, 13). In the present study we undertook an analysis of the predictive value of TT and equilibrium dialysis-derived FT, determined by measuring TT using either serum extraction-chromatography-RIA or LC-MS/MS, in 100 well-phenotyped normal and 100 PCOS females. We also modeled two different clinical scenarios; one reflecting subjects seen in the clinical setting and the other reflecting subjects detected in a population study.

Overall the TT and FT values calculated based on LC-MS/MS TT values were slightly lower than the corresponding values by extraction chromatography and RIA, for both PCOS and control subjects. The amplitude of the difference is provided in the Deming fit, which for total T reveals – 7.37 ng/dl higher levels for RIA and a proportional bias of 4% that will increase the difference as the values increase. At low values it is insignificant but at higher values, say 50 ng/dl, it adds 2 ng/dl to the 7 ng higher values of the RIA. At the extremes (very low or very high) these differences are not clinically significant, but at the cusp of the cutoff they do matter if one was to adopt a cutoff that is method independent. The formula provided in the first paragraph of the ‘Results’ section (Deming fit of $y = -7.37 + 0.96x$) allows for extrapolating one value from one method to the other (i.e. LC-MS/MS values showed a constant negative intercept of –7.37 accentuated by a negative slope of 0.96 relative to RIA). Nonetheless, all values in both controls and PCOS women were above the detection limits, regardless of assay type. Although the lower limit of detection for these methods might be of relevance in measuring TT and FT in women with post-surgical or natural menopause, or suffering from androgen deficiency, it does not appear to limit the evaluation of PCOS or healthy women.

Distinguishing between PCOS and healthy controls can be complicated. The two assay types produced similar degrees of overlap between PCOS and healthy controls for both TT and FT, although the degree of overlap was almost twice as great for TT as for FT. There was also no difference in the ROC curves for the detection of PCOS by either assay method, although the ROC curve was much higher for FT than TT. Overall, the value of TT alone to distinguish PCOS from normal women is relatively limited, consistent with our prior data using a different chromatography and RIA assay for TT and a competitive binding assay for calculating FT (23, 24).

In establishing cut-off values to distinguish normal from abnormal it is critical to consider the background frequency of the disorder in the population being assessed. While a complete discussion of all the possible permutations for cut-off values exceeds the aim of the present narrative, we can note the following: for those clinicians who are evaluating patients for PCOS in their practices, their population will be enriched due to patient self-referral (25). Hence, the prevalence of PCOS in their patient population will be higher and the cut-off values used for detection will then be lower (i.e. in deference to a having a lower false positive rate). Alternatively, if the clinician (or researcher) was attempting to detect disease in the general or unselected population (i.e. if screening asymptomatic subjects or when attempting to determine the prevalence of disease in unselected individuals) the cut-off values chosen would be higher (i.e. in consideration of a higher false positive rate).

Assays to measure TT and FT for the diagnosis of PCOS may be used in the clinical setting, where the population being studied is enriched with patients present with complaints associated with PCOS such as irregular menstruation and/or unwanted hair growth. In this setting it is not uncommon to observe a 70% prevalence of PCOS in the subgroup of women being evaluated (18, 19). Alternatively, when studies are performed in unselected women or for the general screening of a population, for example to establish the populational prevalence of the disorder, the populational rate for PCOS generally ranges between 7% and 10% (13-17). In the clinical setting, where the prevalence and PPV is higher, lower cut-off values for TT, or preferably FT, should be used (i.e. because the false-positive rate will be lower). Alternatively, when used for the detection of PCOS in an unselected population, where the prevalence and PPV is significantly lower, the use of higher cut-off values is preferred (i.e. when the false-positive rate is higher)..

Total T and FT measurements by extraction chromatography and RIA and by LC-MS/MS performed similarly in our hands in their ability to provide similar AUC in their respective ROC curves. However, many factors favor the use of LC-MS/MS for TT measurement over that of extraction chromatography and RIA, including: 1) the lack of standardization of RIAs, 2) the decreasing availability of extraction chromatography and RIA in reference laboratories, 3) the reliance of RIAs on radioisotopes, 4) significant batch to batch tracer variation, 5) the labor intensive nature of RIA methods, and 6) the inability to automate and reduce turnaround time significantly.

With the improvements in tandem mass spectrometers in terms of sensitivity and robustness that have occurred in recent years, we undertook the development of an LC-MS/MS assay that is highly sensitive and accurate for measuring TT over a wide dynamic range

encompassing the problematic low values that are a vexing problem in the platform assays. The validation of this assay has been previously published (22). Briefly, the assay relies on an on-line extraction, HTLC, and MS detection. It requires no extraction and with the use of the Aria TX-4 system is capable of producing a testosterone result every 1.15 min. It has a dynamic measuring range from 0.3 to 2,000 ng/dL. The accuracy of FT measurements by either equilibrium dialysis or calculation is entirely dependent on the accuracy of the TT measurement. Consequently we used this LC-MS/MS assay for TT to re-establish the reference ranges for free testosterone (by tracer equilibrium dialysis).

The diagnosis of a hyperandrogenic woman depends to a great degree on the accuracy of the androgen measurements. Employing methodology that is accurate, reproducible and sensitive is key, particularly in this group of subjects where androgen levels, while elevated for women, are lower than in men. While chromatography and RIA and LC-MS/MS are generally similar in performance, LC-MS/MS may be preferable as it is the procedure most amenable to automation. Furthermore, compared to FT, measurement of TT has relatively limited value for distinguishing PCOS from normal. Based on the scenarios considered above, the population being evaluated needs to be clearly identified, as normative ranges can be impacted to a significant degree by the expected prevalence of the disorder in the population being studied. Further large scale trials are warranted to determine normative TT and FT ranges in these subpopulations of women.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank the Medical Information and Steroids Operations staff of Quest Diagnostics Nichols Institute, San Juan Capistrano, CA; and Veneta Kirilova at Cedars-Sinai Medical Center.

Grant Support: This study was supported in part by NIH grants R01-HD29364 and an endowment from the Helping Hand of Los Angeles, Inc. (to RA)

REFERENCES

1. Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Futterweit W, Janssen OE, Legro RS, Norman RJ, Taylor AE, Witchel SF, Task Force on the Phenotype of the Polycystic Ovary Syndrome of The Androgen Excess and PCOS Society. The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report. *Fertil Steril.* 2009; 91:456–488. [PubMed: 18950759]
2. Cawood ML, Field HP, Ford CG, Gillingwater S, Kicman A, Cowan D, Barth JH. Testosterone measurement by isotope-dilution liquid chromatography-tandem mass spectrometry: validation of a method for routine clinical practice. *Clin Chem.* 2005; 51:1472–1479. [PubMed: 15961553]
3. Chang YC, Li CM, Li LA, Jong SB, Liao PC, Chang LW. Quantitative measurement of male steroid hormones using automated on-line solid phase extraction-liquid chromatography-tandem mass spectrometry and comparison with radioimmunoassay. *Analyst.* 2003; 128:363–368. [PubMed: 12741642]
4. Herold DA, Fitzgerald RL. Immunoassays for testosterone in women: better than a guess? *Clin Chem.* 2003; 49:1250–1251. [PubMed: 12881438]
5. Matsumoto AM, Bremner WJ. Serum testosterone assays--accuracy matters. *J Clin Endocrinol Metab.* 2004; 89:520–524. [PubMed: 14764756]

6. Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H. Position statement: Utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. *J Clin Endocrinol Metab.* 2007; 92:405–413. [PubMed: 17090633]
7. Stanczyk FZ. Measurement of androgens in women. *Semin Reprod Med.* 2006; 24:78–85. [PubMed: 16633981]
8. 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. *Clin Chem.* 2003; 49:1381–1395. [PubMed: 12881456]
9. 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. *J Clin Endocrinol Metab.* 2004; 89:534–543. [PubMed: 14764758]
10. Wheeler MJ. Measurement of androgens. *Methods Mol Biol.* 2006; 324:197–211. [PubMed: 16761380]
11. Boots LR, Potter S, Potter HD, Azziz R. Measurement of total serum testosterone levels using commercially available kits: High degree of between-kit variability. *Fertil Steril.* 1998; 69:286–92. [PubMed: 9496343]
12. Handelsman DJ, Wartofsky L. Requirement for mass spectrometry sex steroid assays in the Journal of Clinical Endocrinology and Metabolism. *J Clin Endocrinol Metab.* 2013; 98:3971–3. [PubMed: 24098015]
13. Azziz R, Woods KS, Reyna R, Key TJ, Knochenhauer ES, Yildiz BO. The prevalence and features of the polycystic ovary syndrome in an unselected population. *J Clin Endocrinol Metab.* 2004; 89:2745–2749. [PubMed: 15181052]
14. Asuncion M, Calvo RM, San Millan JL, Sancho J, Avila S, Escobar-Morreale HF. A prospective study of the prevalence of the polycystic ovary syndrome in unselected Caucasian women from Spain. *J Clin Endocrinol Metab.* 2000; 85:2434–2438. [PubMed: 10902790]
15. Diamanti-Kandarakis E, Kouli CR, Bergiele AT, Filandra FA, Tsianateli TC, Spina GG, Zapanti ED, Bartzis MI. A survey of the polycystic ovary syndrome in the Greek island of Lesbos: hormonal and metabolic profile. *J Clin Endocrinol Metab.* 1999; 84:4006–4011. [PubMed: 10566641]
16. Yildiz BO, Azziz R. The adrenal and polycystic ovary syndrome. *Rev Endocr Metab Disord.* 2007; 8:331–342. [PubMed: 17932770]
17. March WA, Moore VM, Willson KJ, Phillips DI, Norman RJ, Davies MJ. The prevalence of polycystic ovary syndrome in a community sample assessed under contrasting diagnostic criteria. *Hum Reprod.* 2010; 25:544–551. [PubMed: 19910321]
18. Azziz R, Sanchez LA, Knochenhauer ES, Moran C, Lazenby J, Stephens KC, Taylor K, Boots LR. Androgen excess in women: experience with over 1000 consecutive patients. *J Clin Endocrinol Metab.* 2004; 89:453–462. [PubMed: 14764747]
19. Carmina E, Rosato F, Jannì A, Rizzo M, Longo RA. Extensive clinical experience: relative prevalence of different androgen excess disorders in 950 women referred because of clinical hyperandrogenism. *J Clin Endocrinol Metab.* 2006; 91:2–6. [PubMed: 16263820]
20. Zawadzki, JK.; Dunaif, A. Diagnostic criteria for polycystic ovary syndrome: towards a rational approach. Dunaif A GJHFMGe. , editor. Blackwell Scientific; Boston: 1992. p. 377-384.
21. Knochenhauer ES, Key TJ, Kahsar-Miller M, Waggoner W, Boots LR, Azziz R. Prevalence of the polycystic ovary syndrome in unselected black and white women of the southeastern United States: a prospective study. *J Clin Endocrinol Metab.* 1998; 83:3078–3082. [PubMed: 9745406]
22. Salameh WA, Redor-Goldman MM, Clarke NJ, Reitz RE, Caulfield MP. Validation of a total testosterone assay using high-turbulence liquid chromatography tandem mass spectrometry: Total and free testosterone reference ranges. *Steroids.* 2009; 75:169–175. [PubMed: 19925815]
23. Chang WY, Knochenhauer ES, Bartolucci AA, Azziz R. Phenotypic spectrum of polycystic ovary syndrome: clinical and biochemical characterization of the three major clinical subgroups. *Fertil Steril.* 2005; 83:1717–1723. [PubMed: 15950641]

24. Huang A, Brennan K, Azziz R. Prevalence of hyperandrogenemia in the polycystic ovary syndrome diagnosed by the National Institutes of Health 1990 criteria. *Fertil Steril*. 2010; 93:1938–1941. [PubMed: 19249030]
25. Ezeh U, Yildiz BO, Azziz R. Referral bias in defining the phenotype and prevalence of obesity in polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2013; 98:E1088–96. [PubMed: 23539721]

Figure 1.

Receiver-operator characteristics (ROC) curve for detecting polycystic ovary syndrome (PCOS) using total testosterone (TT) measured by LC-MS/MS or RIA (**Fig. 1A**), free testosterone (FT) measured by LC-MS/MS or RIA (**Fig. 1B**), or both TT and FT (**Fig. 1C**).

Table 1

Sensitivity, Specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV) of Total Testosterone (TT) Cutoff Values Determined by the LC-MS/MS Method

		TT Cutoff (ng/dL)													
		4	10	15	18	20	25	28	30	33	35	40	45	50	
Sensitivity, %		98	93	86	80	76	66	59	56	52	48	43	35	29	
Specificity, %		0	12	34	46	60	80	85	85	91	94	96	98	100	
10% Prevalence															
PPV, %		10	10	13	14	17	27	30	29	39	47	54	66	100	
NPV, %		0	94	96	95	96	95	95	95	94	94	94	93	93	
70% Prevalence															
PPV, %		70	71	75	78	82	89	90	90	93	95	96	97	100	
NPV, %		0	42	51	50	52	50	47	45	45	44	42	39	38	

In our calculations we considered two difference clinical situations: i) the evaluation of women with suspected androgen excess in a referral clinic, where a 70% prevalence of PCOS can be expected; and ii) the evaluation of unselected women from a broader population, as in epidemiologic studies, where a 10% prevalence of PCOS is expected.

The shading indicates the level(s) that have the highest combined PPV and NPV values useful for epidemiologic studies (dotted grey) and for evaluation of patients in the clinical setting (solid grey).

Table 2

Sensitivity, Specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV) of Free Testosterone (FT) cutoff values determined by the LC-MS/MS method

	FT Cutoff (pg/mL)											
	0.2	1	1.5	1.8	2	2.5	3	3.3	3.5	4	4.5	5
Sensitivity, %	100	96	95	91	87	80	75	73	71	64	58	52
Specificity, %	0	21	42	56	63	71	84	87	91	97	98	99
10% Prevalence												
PPV, %	10	12	15	19	21	23	34	38	47	70	76	85
NPV, %	N/A	98	99	98	98	97	97	97	97	96	95	95
70% Prevalence												
PPV, %	70	74	79	83	85	87	92	93	95	98	99	99
NPV, %	N/A	69	78	73	68	60	59	58	57	54	50	47

The shading indicates the level(s) that have the highest combined PPV and NPV values useful for epidemiologic studies (dotted grey) and for evaluation of patients in the clinical setting (solid grey).

Table 3

Sensitivity, Specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV) of Free Testosterone (FT) and Total Testosterone (TT) Cutoff Values Combined Determined by the LC-MS/MS Method

		Cut-Off Value															
TT (ng/dL)	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30
FT (pg/mL)	3	3.3	3.5	4	3	3.3	3.5	4	3	3.3	3.5	4	3	3.3	3.5	4	4
Sensitivity, %	75	74	73	66	75	74	73	66	74	73	71	65	70	64			
Specificity, %	81	82	86	88	83	84	89	93	83	84	89	93	91	94			
10% Prevalence																	
PPV, %	30	31	37	38	33	34	40	51	33	34	42	51	46	54			
NPV, %	97	97	97	96	97	97	97	96	97	97	97	96	96	96			
2×PPV+NPV	157	159	171	172	163	165	177	198	163	165	181	198	188	204			
70% Prevalence																	
PPV, %	90	94	92	93	91	92	93	96	91	91	94	96	95	96			
NPV, %	58	57	58	53	59	58	62	54	58	57	57	53	57	53			
2×PPV+NPV	238	245	242	239	241	242	248	246	240	239	245	245	247	245			

The shading indicates the level(s) that have the highest combined PPV and NPV values useful for epidemiologic studies (dotted grey) and for evaluation of patients in the clinical setting (solid grey).