

Evaluating the diagnostic accuracy of androgen measurement in polycystic ovary syndrome: a systematic review and diagnostic meta-analysis to inform evidence-based guidelines

Asmamaw Demis Bizuneh¹, Anju E. Joham^{1,2}, Helena Teede^{1,2}, Aya Mousa¹, Arul Earnest³, James M. Hawley^{4,5}, Laura Smith⁶, Ricardo Azziz^{7,8}, Wiebke Arlt^{5,9,10}, and Chau Thien Tay^{1,2,*}

¹Monash Centre for Health Research and Implementation, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, VIC, Australia

²Department of Endocrinology and Diabetes, Monash Health, Clayton, Melbourne, VIC, Australia

³School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC, Australia

⁴Department of Clinical Biochemistry, Wythenshawe Hospital, Manchester University NHS Foundation Trust, Manchester, UK

⁵Medical Research Council, Laboratory of Medical Sciences, London, UK

⁶School of Medical, Indigenous and Health Sciences, Illawarra Health and Medical Research Institute, University of Wollongong, Wollongong, NSW, Australia

⁷Department of Obstetrics & Gynecology, and Medicine University of Alabama at Birmingham, Birmingham, AL, USA

⁸Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA

⁹Institute of Metabolism and Systems Research, University of Birmingham, Birmingham, UK

¹⁰Institute of Clinical Sciences, Imperial College London, London, UK

*Correspondence address. Monash Centre for Health Research and Implementation, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, VIC 3168, Australia and Department of Endocrinology and Diabetes, Monash Health, Clayton, Melbourne, VIC 3168, Australia. E-mail: jillian.tay@monash.edu

TABLE OF CONTENTS

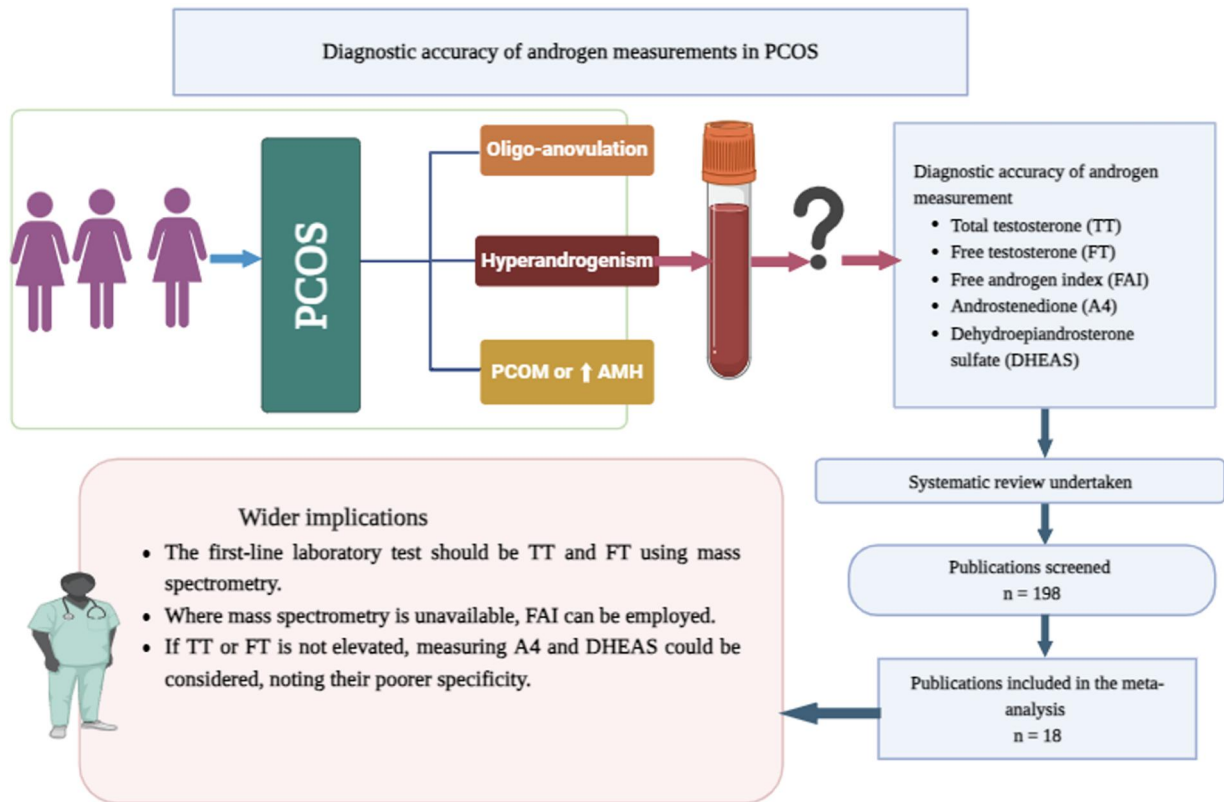
- Introduction
- Methods
 - Data sources and searches
 - Study selection
 - Quality appraisal of the selected studies
 - Data synthesis and analysis
- Results
 - Characteristics of included studies
 - Serum total testosterone
 - Free testosterone
 - Calculated free testosterone
 - Free androgen index
 - Serum androstenedione
 - Dehydroepiandrosterone sulfate
 - Bioavailable testosterone
 - Dihydrotestosterone
 - Salivary testosterone
- Discussion
 - Strength and limitations
- Conclusion

Received: May 26, 2024. Revised: July 30, 2024. Editorial decision: August 19, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

GRAPHICAL ABSTRACT



Health professionals should use FT and TT as the first-line laboratory test to assess biochemical hyperandrogenism in the diagnosis of PCOS.

FT: free testosterone, TT: total testosterone, AMH: anti-Müllerian hormone, PCOM: polycystic ovary morphology, PCOS: polycystic ovary syndrome.

ABSTRACT

BACKGROUND: Biochemical hyperandrogenism is a hallmark and diagnostic feature of polycystic ovary syndrome (PCOS). However, the most accurate androgen measurement for assessing biochemical hyperandrogenism in PCOS diagnosis remains uncertain.

OBJECTIVE AND RATIONALE: This systematic review aimed to assess different androgen measures [including total testosterone (TT), calculated free testosterone (cFT), free androgen index (FAI), androstenedione (A4), dehydroepiandrosterone sulfate (DHEAS), and dihydrotestosterone (DHT)] for accuracy in diagnosing biochemical hyperandrogenism in women with PCOS, to inform the 2023 International PCOS Evidence-based Guidelines.

SEARCH METHODS: To update evidence from the 2018 International PCOS Guidelines, a systematic search from 3 July 2017 to 23 June 2023 was conducted across Medline (Ovid), CINAHL, all EBM, EMBASE, and PsycInfo for articles evaluating androgens in the diagnosis of biochemical hyperandrogenism. The revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) was used to assess the risk of bias and applicability. A diagnostic test accuracy meta-analysis was performed using STATA 18 software. Summary sensitivity and specificity were calculated with 95% CIs using the bivariate model, while the hierarchical summary receiver operating characteristics (ROC) model was used to produce a summary ROC curve.

OUTCOMES: Of 23 studies reviewed, 18 were included in the meta-analysis, with data from 2857 participants (1650 with PCOS and 1207 controls). For diagnosing biochemical hyperandrogenism in PCOS, the pooled sensitivity, specificity, and AUC with 95% CI were for TT: 0.74 (0.63–0.82), 0.86 (0.77–0.91), and 0.87 (0.84–0.90); cFT: 0.89 (0.69–0.96), 0.83 (0.79–0.86), and 0.85 (0.81–0.88); FAI: 0.78 (0.70–0.83), 0.85 (0.76–0.90), and 0.87 (0.84–0.90); A4: 0.75 (0.60–0.86), 0.71 (0.51–0.85), and 0.80 (0.76–0.83); and DHEAS: 0.75 (0.61–0.85), 0.67 (0.48–0.81), and 0.77 (0.73–0.81), respectively. In subgroup analyses, liquid chromatography with tandem mass spectrometry (LC-MS/MS) had superior sensitivity for measuring cFT, FAI, A4, and DHEAS, and superior specificity for measuring TT, cFT, and FAI, compared to the direct immunoassay method.

WIDER IMPLICATIONS: Our results directly informed the 2023 International PCOS Guideline recommendations to use TT and FT as the first-line laboratory tests to assess biochemical hyperandrogenism in the diagnosis of PCOS. cFT should be assessed by equilibrium dialysis or ammonium sulfate precipitation, or calculated using FAI. If TT or cFT are not elevated, A4 and DHEAS could also be considered, noting their poorer specificity. Laboratories should utilize LC-MS/MS for androgen measurement given its high accuracy. Future studies should focus on establishing optimal normative cut-off values in large, unselected, and ethnically diverse cohorts of women.

REGISTRATION NUMBER: The review protocol was prepublished in the 2023 PCOS Guideline Technical Report (https://www.monash.edu/_data/assets/pdf_file/0010/3379591/TechnicalReport-2023.pdf).

Keywords: diagnostic accuracy / sensitivity / specificity / hyperandrogenism / hyperandrogenemia / PCOS / androgen / testosterone

Introduction

Polycystic ovary syndrome (PCOS) is a common endocrine disorder affecting women throughout their lifespan (Teede et al., 2010). The prevalence of PCOS varies depending on the population studied, but it is estimated to affect between 10% and 13% or ~140 million women globally (March et al., 2010; Teede et al., 2023). PCOS in adults is diagnosed according to the International Evidence-based Guideline Criteria based on two of either ovulatory dysfunction, clinical or biochemical hyperandrogenism, and/or polycystic ovary morphology (PCOM) on ultrasound or elevated anti-Müllerian hormone levels (Lizneva et al., 2016; Azziz, 2018; Teede et al., 2018). Hyperandrogenism, present in 75% of PCOS cases, is both a diagnostic criterion and a well-established contributor to the pathogenesis of PCOS (Rosenfield and Ehrmann, 2016; Ashraf et al., 2019). The detection of hyperandrogenism in women with PCOS can be accomplished clinically through assessment of hirsutism, acne or androgenic alopecia, or via biochemical testing of circulating androgen concentrations (Huang et al., 2010).

Despite its centrality in the diagnostic workup of PCOS, accurately diagnosing biochemical hyperandrogenism presents considerable challenges (Taieb et al., 2003; Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group, 2004). Liquid chromatography-tandem mass spectrometry (LC-MS/MS) is considered the reference standard for assessing steroids including androgens, but its use is limited by technical complexity and accessibility (Rosner and Vesper, 2010; Janse et al., 2011; Owen and Keevil, 2012; Keevil, 2013; Owen et al., 2016). While immunoassay methods are widely available, the results of total testosterone (TT) and calculated free androgen index (FAI) can be influenced by fluctuations in sex hormone-binding globulin (SHBG) levels, which are affected by age, weight, and medication use (Rosner et al., 2007). Free testosterone (FT) is also highly inaccurate by direct immunoassay methods due to its low serum concentrations (Miller et al., 2004). Furthermore, in addition to testosterone, the measurement of other androgens or androgen precursors such as androstenedione (A4), dehydroepiandrosterone (DHEA), dehydroepiandrosterone sulfate (DHEAS), and dihydrotestosterone (DHT) are also available. As each exhibits variation in potency, metabolism, and tissue-specific effects, it is difficult to determine which androgen(s) are responsible for hyperandrogenic features of PCOS and are best captured in the diagnosis (Taieb et al., 2003; Azziz et al., 2009; Barth et al., 2010; Keevil, 2013; Goodarzi et al., 2015; Keevil et al., 2018; Carmina and Longo, 2022). Lastly, there is a lack of consensus in reference ranges based on population characteristics such as ethnicity and age (Teede et al., 2010).

The lack of evidence regarding the diagnostic accuracy of specific androgen measurement methods poses challenges for clinicians in choosing the most reliable and accurate approach for assessing PCOS-related biochemical hyperandrogenism, yet this is a research priority identified by the 2023 International PCOS Guidelines (Teede et al., 2023). To address this critical gap in knowledge, this review aimed to collate and synthesize the available evidence to identify the most accurate androgen measurement for detecting biochemical hyperandrogenism among women with PCOS. Findings from this review directly informed recommendations of the 2023 International PCOS Guidelines.

Methods

This systematic review and meta-analysis are an update of the review undertaken for the 2018 International PCOS Guidelines (Teede et al., 2018) and is reported in line with the Preferred

Reporting Items for Systematic Reviews and Meta-Analyses evaluating Diagnostic Test Accuracy (PRISMA-DTA) (Campbell et al., 2015; McInnes et al., 2018; Salameh et al., 2020; Page et al., 2021). The protocol is publicly available online in the 2023 PCOS Guideline Technical Report (Mousa et al., 2023).

Data sources and searches

A systematic literature search was performed across Medline (Ovid), CINAHL, all EBM, EMBASE, and PsycInfo from 3 July 2017 until 23 June 2023. The search terms from the original review were used and consisted of a combination of keywords including polycystic ovary syndrome, anovulation, oligo-ovulation, hyperandrogenism, testosterone, sex hormone-binding globulin, free androgen index, androstenedione, dehydroepiandrosterone sulfate, sensitivity, specificity, etc. Details of the full search strategy are available in [Supplementary Table S1](#). A manual search of the reference lists of applicable reviews was also performed to identify additional studies that were not found in the databases.

Study selection

Women diagnosed with PCOS using the historical National Institutes of Health (NIH) criteria of 1990 (Zawadzki and Dunaif, 1992), the Rotterdam criteria of 2003 (Rotterdam ESHRE ASRM Sponsored PCOS Consensus Workshop Group, 2004), or the Androgen Excess and PCOS (AE-PCOS) diagnostic criteria of 2006 (Azziz et al., 2006) were included as the study population without any restriction on age, ethnicity, or weight. Any methods for assessment of androgens including TT, calculated or measured FT (cFT), FAI, A4, or serum DHEAS to diagnose biochemical hyperandrogenism were examined. The main outcomes of interest were the sensitivity and specificity of the test to detect biochemical hyperandrogenism in PCOS. We included randomized controlled trials, comparative studies, and prospective cohort studies published in the English language. Non-English language studies, case reports, case series, narrative reviews, letters, editorials/commentaries, and conference abstracts (posters or oral presentations) were excluded.

After removing duplicate studies, titles, and abstract screening, then full-text screening were performed in duplicate by four reviewers (A.D.B., L.S., T.R. and C.T.T.) using Covidence software. Where discrepancies occurred, another independent reviewer (A.E.J.) was consulted to resolve any conflicts. The data extracted included study information such as first author, country, publication year, sample size, number of cases and controls, and participant selection criteria. The participant characteristics included age and BMI, androgens assessed, methods or assay(s) used, and normative reference cut-off thresholds, true positives (TP), false positives (FP), false negatives (FN), true negatives (TN), sensitivity and specificity scores, and AUC values. Data extraction was performed by two reviewers (A.D.B. and T.R.) and cross-checked by C.T.T.

Quality appraisal of the selected studies

Two independent reviewers (A.D.B. and C.T.T.) assessed the risk of bias (RoB) and applicability of the included studies using the updated Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool (Whiting et al., 2011; Yang et al., 2021). The tool assesses four domains: patient selection, index test, reference standard, and flow and timing. All four domains assess RoB, while the first three assess applicability. Each domain was ranked as 'low risk', 'high risk', or 'unclear risk' (Whiting et al., 2011). Deek's funnel plot test was used to detect publication bias, with a slope coefficient of $P < 0.1$ considered suggestive of the

presence of significant asymmetry and small study bias (Deeks et al., 2005; van Enst et al., 2014).

Data synthesis and analysis

Data from the included studies were narratively described and presented in tables. A 2×2 contingency table was constructed using TP, FP, TN, and FN. Sensitivity and specificity were calculated with a 95% CI for each androgen measurement. Additionally, the AUC, diagnostic odds ratio (DOR), positive likelihood ratio (PLR), negative likelihood ratio (NLR), positive predictive value (PPV), and negative predictive value (NPV) were calculated. The bivariate and hierarchical summary receiver operating characteristics (HSROC) models were computed using the midas, metadta, and metandi modules of the STATA 18 software (Stata/SE Stata Corp, USA) to produce summary points of sensitivity and specificity and HSROC curves with confidence and prediction regions, respectively (Rutter and Gatsonis, 2001; Deeks et al., 2005; Reitsma et al., 2005; Harbord and Whiting, 2009; Macaskill et al., 2010; van Enst et al., 2014; StataCorp, 2023). Both models incorporate two levels to analyze data distribution: within-study variation using a 2×2 table and between-study variation. The HSROC model fits a curve based on sensitivity and specificity pairs at different thresholds, while the bivariate model estimates the correlation between sensitivity and specificity within each study and provides summary estimates across studies.

A sensitivity analysis was conducted to evaluate the potential impact of influential studies based on influential analysis and outlier detection output on sensitivity and specificity. A subgroup analysis was done by assay method (direct immunoassays, immunoassays after extraction, and LC-MS/MS) and PCOS diagnosis criteria (Rotterdam criteria vs NIH). For index tests with less than four studies, evidence synthesis was conducted without meta-analysis due to the prerequisite of a minimum of four studies for bivariate analysis. To explore sources of heterogeneity, a univariate meta-regression analysis was conducted to examine the assay method (direct immunoassays, immunoassays after extraction, and LC-MS/MS), the sample size (<150 vs ≥ 150), and the PCOS diagnostic criteria (Rotterdam vs NIH criteria). The accuracy of the index test can be measured by the AUC value, which is interpreted as excellent: 0.9–1.0, very good: 0.8–0.9, good: 0.7–0.8, sufficient: 0.6–0.7, bad: 0.5–0.6, and test not useful: <0.5 (Simundic, 2009). While Cochran's Q-statistic and I^2 statistics can assess heterogeneity, they are not sufficient for evaluating the threshold effect. To assess the threshold effect, coupled forest plots or HSROC plots were visually evaluated, and Spearman's correlation analysis between sensitivity and FP rate was performed. A correlation coefficient (r) of 0.6 or higher was considered indicative of a substantial threshold effect (Lee et al., 2015; Devillé et al., 2002).

Results

The database search yielded a total of 190 articles, which were combined with 8 articles identified in the 2018 International Guidelines systematic review (Teede et al., 2018) for inclusion in the current review. After removing 47 duplicates, 151 articles were screened by title and abstract, of which 59 full-text articles were retrieved. Finally, 23 articles met inclusion criteria and were included in qualitative synthesis, of which 18 articles were included in quantitative synthesis through meta-analysis, with five not included due to the absence of sensitivity and specificity reports (Fig. 1).

Characteristics of included studies

Of the 23 studies, 19 (Koskinen et al., 1996; Stener-Victorin et al., 2010; Bili et al., 2014; Villarroel et al., 2015; Rudnicka et al., 2016; Sathyapalan et al., 2017; Kannenberg et al., 2018; Luque-Ramirez et al., 2018; Nadaraja et al., 2018; Bansal et al., 2020; Cao et al., 2020; Grassi et al., 2020; Imran et al., 2020; Khashchenko et al., 2020; Chen et al., 2021; Beitel et al., 2022; Kumar et al., 2022; Xu et al., 2022; Yi et al., 2023) utilized the Rotterdam criteria and the remaining 4 studies (Escobar-Morreale et al., 2001; Hahn et al., 2007; Salameh et al., 2014; Diamandis et al., 2017) utilized the NIH criteria to diagnose PCOS. The 23 studies included 3751 participants (2225 with PCOS and 1526 controls) and the 18 studies included in the meta-analysis had a total of 2857 participants (1650 with PCOS and 1207 controls). The mean age spanned from 16 to 34.5 years. Mean BMI ranged from 22.4 to 37.6 kg/m² for those with PCOS, compared with 20.2 to 32.7 kg/m² for controls (Table 1). Out of the included studies, five studies (Koskinen et al., 1996; Hahn et al., 2007; Stener-Victorin et al., 2010; Nadaraja et al., 2018; Imran et al., 2020) employed clinical hyperandrogenism for PCOS diagnosis, while the remaining studies used both clinical and biochemical hyperandrogenism, following either the Rotterdam or NIH criteria, to assess the accuracy of index tests.

The results of QUADAS-2 are presented in Supplementary Table S2. More than half ($n = 14$) of the included studies had high RoB in the patient selection domain. All analyzed studies had unclear RoB for both the index test and reference standard domains due to inadequate description of whether the index test results were interpreted without knowledge of the reference standard results and vice versa. For the flow and timing domain, 20 studies had low RoB, while the remaining 3 studies had unclear RoB due to insufficient description of whether there was an appropriate interval between the index test and reference standard, and whether all patients received the same reference standard.

Serum total testosterone

Twenty-three studies examined serum TT for the diagnosis of biochemical hyperandrogenism. Twelve studies used direct immunoassays (Escobar-Morreale et al., 2001; Hahn et al., 2007; Bili et al., 2014; Villarroel et al., 2015; Rudnicka et al., 2016; Luque-Ramirez et al., 2018; Nadaraja et al., 2018; Bansal et al., 2020; Khashchenko et al., 2020; Beitel et al., 2022; Kumar et al., 2022; Yi et al., 2023), two used immunoassays after extraction (Koskinen et al., 1996; Imran et al., 2020), seven used LC-MS/MS (Stener-Victorin et al., 2010; Diamandis et al., 2017; Sathyapalan et al., 2017; Cao et al., 2020; Grassi et al., 2020; Chen et al., 2021; Xu et al., 2022), one study used both LC-MS/MS and immunoassays after extraction and chromatography (Salameh et al., 2014), and the remaining one study used LC-MS/MS and direct immunoassay (Kannenberg et al., 2018). The reported sensitivity ranged from 0.29 to 0.93, while specificity ranged from 0.60 to 1.00. The cut-off for TT varied across studies (median 1.74 nmol/L, IQR 1.08–2.37).

Ten studies (Escobar-Morreale et al., 2001; Stener-Victorin et al., 2010; Bili et al., 2014; Diamandis et al., 2017; Sathyapalan et al., 2017; Kannenberg et al., 2018; Luque-Ramirez et al., 2018; Khashchenko et al., 2020; Xu et al., 2022; Yi et al., 2023) were excluded from the meta-analyses due to the absence of sensitivity and specificity reports relating to TT. Pooled results from 13 studies with a total of 2182 participants (1256 with and 926 without PCOS) showed that serum TT had pooled sensitivity of 0.74 (95% CI: 0.63–0.82; $I^2 = 93.59\%$), specificity of 0.86 (95% CI: 0.77–0.91; $I^2 = 85.22\%$) (Table 2, Fig. 2, and Supplementary Table S3), AUC of

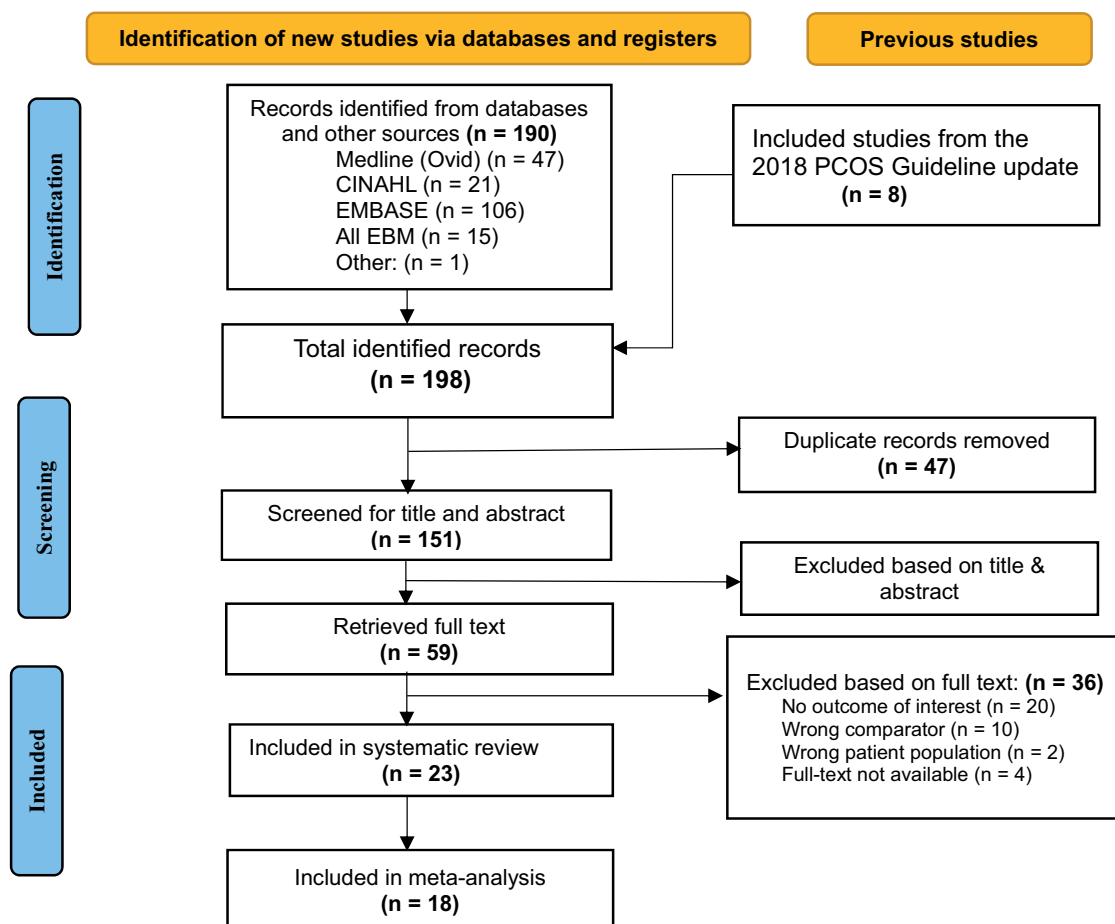


Figure 1. PRISMA flow diagram for the systematic reviews.

0.87 (95% CI: 0.84–0.90) (Fig. 3), DOR of 17 (95% CI: 10–26), PLR of 5.1 (95% CI: 3.4–7.7), NLR of 0.31 (95% CI: 0.23–0.42), PPV of 88%, and NPV of 30% (Table 2 and Supplementary Fig. S1). There was no threshold effect for TT ($\rho=0.43$, $P=0.14$) by Spearman's correlation, and this was confirmed on the HSROC curve (Fig. 3). Deek's funnel plot test showed no evidence of publication bias or small study effects for TT ($P=0.53$) (Supplementary Fig. S2).

A sensitivity analysis revealed that excluding the outlier study conducted by Salameh et al. (2014) resulted in an increase in sensitivity for TT to 0.77 (95% CI: 0.69–0.83) and a slight decrease in specificity to 0.82 (95% CI: 0.76–0.87). Notably, the AUC and the heterogeneity of the study remained unchanged and the CIs were similar with no compromise in precision.

In the subgroup analysis by assay methods, direct immunoassays had a sensitivity of 0.74 (95% CI: 0.60–0.85), a specificity of 0.78 (95% CI: 0.67–0.86), and an AUC of 0.82 (0.79–0.86), immunoassay after extraction had a sensitivity of 0.74 (95% CI: 0.46–0.91) and a specificity of 0.93 (95% CI: 0.80–0.98), while the LC-MS/MS method had a sensitivity of 0.71 (95% CI: 0.51–0.85), a specificity of 0.92 (95% CI: 0.81–0.97), and an AUC of 0.90 (0.87–0.92). In the subgroup analysis by diagnostic criteria, the Rotterdam criteria showed a sensitivity of 0.77 (95% CI: 0.68–0.84) and a specificity of 0.83 (95% CI: 0.74–0.90), while the NIH criteria showed a sensitivity of 0.51 (95% CI: 0.28–0.74) and a specificity of 0.94 (95% CI: 0.78–0.98) (Table 3). The univariate meta-regression analysis showed no statistically significant association between the assay method and sample size with sensitivity and specificity. However, PCOS diagnostic criteria may be a source of heterogeneity for sensitivity ($P < 0.05$) (Supplementary Fig. S3).

Free testosterone

Two studies assessed FT functionally measured using equilibrium dialysis (Salameh et al., 2014; Diamandis et al., 2017) but only one study reported sensitivity and specificity data and therefore not amenable to meta-analysis. In the study by Salameh et al. (2014), of 200 participants (100 with and 100 without PCOS), FT was measured functionally by equilibrium dialysis based on TT measured by LC-MS/MS at a threshold cut-off of ≥ 5 pg/ml and had a sensitivity of 0.52, specificity of 0.99, and AUC of 0.87. Diamandis et al. (2017) studied 85 participants (45 with and 40 without PCOS) and reported FT measured functionally by equilibrium dialysis based on TT measured by LC-MS/MS methods had an AUC of 0.91.

Calculated free testosterone

Seven studies assessed cFT, with four based on TT measured with direct immunoassays (Escobar-Morreale et al., 2001; Hahn et al., 2007; Luque-Ramirez et al., 2018; Nadaraja et al., 2018), one based on TT measured with immunoassay after extraction (Imran et al., 2020), and the remaining two based on TT measured with LC-MS/MS (Stener-Victorin et al., 2010; Grassi et al., 2020). The reported sensitivity ranged from 0.62 to 1.0, while specificity ranged from 0.80 to 0.89.

One study (Luque-Ramirez et al., 2018) was excluded from the cFT meta-analyses due to the absence of sensitivity and specificity reports. Pooled results from the remaining six studies with a total of 1035 participants (602 with and 433 without PCOS) showed that cFT had a pooled sensitivity of 0.89 (95% CI: 0.69–0.96; $I^2 = 92.57\%$) and specificity of 0.83 (95% CI: 0.79–0.86; $I^2 =$

Table 1. Summary of included study characteristics in the evidence synthesis.

Author, year	Country	Study design	Sample size	Mean age ($\pm$ SD)	Mean BMI ($\pm$ SD)	Diagnostic criteria	Assay method	Index tests
Koskinen <i>et al.</i> , 1996	Finland	Retrospective	PCOS: 54 Control: 29	PCOS: 27 $\pm$ 6 Control: 33 $\pm$ 5	PCOS: 27 $\pm$ 8 Control: 29 $\pm$ 10	Rotterdam	Immunoassay after extraction	TT, FAI, A4, DHT, SHBG
Escobar-Morreale <i>et al.</i> , 2001	Spain	Prospective	PCOS: 8 Control: 106	PCOS: 25 $\pm$ 5.4 Controls: 34.5 $\pm$ 9.4	PCOS: 25.5 $\pm$ 4.2 Control: 24.2 $\pm$ 3.1	NIH 1990	Direct immunoassay	TT, cFT, FAI, DHEAS, SHBG
Hahn <i>et al.</i> , 2007	Germany	Prospective	PCOS: 133 Control: 54	PCOS: 28 (15–45) ^a Control: 28 (20–41) ^a	PCOS: 30 (17–58) ^a Control: 21.3 (18–29) ^a	NIH 1990	Direct immunoassay	TT, cFT, FAI, A4, DHEAS, bio-T, SHBG
Stener-Victorin <i>et al.</i> , 2010	Sweden	Cross-sectional	PCOS: 74 Control: 31	PCOS: 30 $\pm$ 4.4 Controls: 27.8 $\pm$ 3.5	PCOS: 25.3 (22.1–33.3) ^a Control: 23.6 (21.0–27.2) ^a	Rotterdam	LC-MS/MS (TT, A4, DHT), direct immunoassay (SHBG)	TT, cFT, A4, DHT, SHBG
Bili <i>et al.</i> , 2014	Greece	Prospective case-control	PCOS: 43 Control: 40	PCOS: 28.9 $\pm$ 5 Control: 30.8 $\pm$ 4.3	PCOS: 24.9 $\pm$ 5.9 Control: 22.5 $\pm$ 3.7	Rotterdam	Direct immunoassay	TT, FAI, A4, DHEAS, SHBG
Salameh <i>et al.</i> , 2014	USA	Prospective case-control	PCOS: 100 Control: 100	PCOS: 28.2 $\pm$ 7.1 Control: 33.2 $\pm$ 9.8	PCOS: 37 $\pm$ 9.1 Control: 27.5 $\pm$ 6.9	NIH 1990	LC-MS/M and immunoassay after extraction and column chromatography (TT), equilibrium dialysis (FT)	TT, FT
Villarreal <i>et al.</i> , 2015	Chile	Cross-sectional	PCOS: 26 Control: 63	PCOS: 17.3 $\pm$ 1.9 Control: 16.6 $\pm$ 1.5	n/a	Rotterdam	Direct immunoassay	TT, FAI, A4, DHEAS, SHBG
Rudnicka <i>et al.</i> , 2016	Poland	Prospective	PCOS: 165 Control: 40	PCOS: 25 (23–25) ^a Control: 25 (24–27) ^a	PCOS: 23 (21–28) ^a Control: 22 (21.5–25) ^a	Rotterdam	Direct immunoassay	TT, FAI, A4, DHEAS, SHBG
Diamandis <i>et al.</i> , 2017	USA	Prospective case-control	PCOS: 45 Control: 40	PCOS: 27.9 $\pm$ 5.3 Control: 34.5 $\pm$ 8.6	PCOS: 37.6 $\pm$ 8.8 Control: 32.7 $\pm$ 7.3	NIH 1990	LC-MS/MS (TT), equilibrium dialysis (FT), Direct immunoassay (A4, DHEAS, and SHBG)	TT, FT, A4, DHEAS, SHBG
Sathyapalan <i>et al.</i> , 2017	UK	Prospective	PCOS: 110 Control: 65	PCOS: 25.5 $\pm$ 10 Control: 32 $\pm$ 12	PCOS: 33 (10.5) ^a Control: 25 (6.2) ^a	Rotterdam	LC-MS/MS (TT and A4), direct immunoassay (SHBG)	TT, FAI, A4, Salivary T, SHBG
Nadaraja <i>et al.</i> , 2018	Malaysia	Retrospective	PCOS: 164 Control: 148	PCOS: 30 (21–45) ^a Control: 31 (21–46) ^a	n/a	Rotterdam	Direct immunoassay	TT, cFT, FAI, bio-T, SHBG
Kannenber <i>et al.</i> , 2018	Germany	Prospective	PCOS: 43 Control: 41	PCOS: 32.2 $\pm$ 4.7 Control: 33 $\pm$ 4	PCOS: 27.7 $\pm$ 8.1 Control: 23.9 $\pm$ 3.8	Rotterdam	LC-MS/MS and direct immunoassay	TT, DHT
Luque-Ramirez <i>et al.</i> , 2018	Spain	Cross-sectional	PCOS: 263 Control: 73	PCOS: 27 $\pm$ 6 Control: 33 $\pm$ 8	PCOS: 30 $\pm$ 8 Control: 31 $\pm$ 7	Rotterdam	Direct immunoassay	TT, cFT, FAI, A4, DHEAS, SHBG
Cao <i>et al.</i> , 2020	China	Prospective	PCOS: 63 Control: 161	n/a	n/a	Rotterdam	LC-MS/MS	TT, A4, DHEAS, DHT

(continued)

Table 1. Continued

Author, year	Country	Study design	Sample size	Mean age ($\pm$ SD)	Mean BMI ($\pm$ SD)	Diagnostic criteria	Assay method	Index tests
Imran et al., 2020	Iraq	Case-control	PCOS (hirsutism): 100 PCOS (no hirsutism): 30 Control: 70	PCOS (hirsutism): 24 $\pm$ 6 PCOS (no hirsutism): 26 $\pm$ 8 Control: 26 $\pm$ 7	PCOS (hirsutism): 32 $\pm$ 5.9 PCOS (no hirsutism): 27.4 $\pm$ 3.5 Control: 26.9 $\pm$ 5.7	Rotterdam	Immunoassay after extraction (TT and DHEAS), direct immunoassay (SHBG)	TT, cFT, DHEAS, SHBG
Khashchenko et al., 2020	Russia	Case-control	PCOS: 130 Controls: 30	PCOS: 16 (15-17) ^a Control: 16 (15-17) ^a	PCOS: 22.4 (19.9-27.2) ^a Control: 20.2 (18.4-21.8) ^a	Rotterdam	Direct immunoassay	TT, FAI, A4, DHEAS, SHBG
Bansal et al., 2020	India	Prospective	PCOS: 31 Control: 89	PCOS: 27.7 $\pm$ 4.9 Control: 30.4 $\pm$ 4.9	n/a	Rotterdam	Direct immunoassay	TT, FAI, DHEAS, SHBG
Chen et al., 2021	China	Prospective	PCOS: 102 Control: 41	PCOS: 29 $\pm$ 3.8 Control: 29.4 $\pm$ 4.2	PCOS: 24.3 $\pm$ 4.5 Controls: 20.3 $\pm$ 1.8	Rotterdam	LC-MS/MS (TT, A4, and DHEAS), direct immunoassay (SHBG)	TT, FAI, A4, DHEAS, SHBG
Grassi et al., 2020	Italy	Prospective	PCOS: 93 Controls: 24	PCOS: 23.7 $\pm$ 6.3 Control: 23.9 $\pm$ 5.8	PCOS: 23.5 (8.1) ^a Controls: 22.6 (5.4) ^a	Rotterdam	LC-MS/MS (TT and A4), direct immunoassay (SHBG)	TT, cFT, FAI, A4, SHBG
Beitl et al., 2022	Austria	Retrospective case-control	PCOS: 58 Control: 58	PCOS: 25.5 $\pm$ 4.7 Controls: 25.5 $\pm$ 4.7	PCOS: 26.3 $\pm$ 6.2 Controls: 26.4 $\pm$ 6.3	Rotterdam	Direct immunoassay	TT, A4
Kumar et al., 2022	India	Case-control	PCOS: 137 Control: 49	PCOS: 23.7 $\pm$ 4.8 Control: 26.2 $\pm$ 4.4	PCOS: 25.2 $\pm$ 5.11 Control: 22.6 $\pm$ 3.4	Rotterdam	DIA (TT, FAI, A4, DHEAS, and DHT), ELISA (SHBG)	TT, FAI, A4, DHEAS, DHT, SHBG
Xu et al., 2022	China	Prospective	PCOS: 114 Control: 100	PCOS: 24.2 $\pm$ 5 Control: 26.5 $\pm$ 3.3	PCOS: 24.8 $\pm$ 4.6 Control: 24 $\pm$ 3.6	Rotterdam	LC-MS/MS	TT, A4, DHT
Yi et al., 2023	China	Prospective	PCOS: 139 Control: 74	PCOS: 28.4 $\pm$ 4.4 Control: 31.9 $\pm$ 4.9	n/a	Rotterdam	Direct immunoassay	TT, FAI, A4, DHEAS, SHBG

BMI in kg/m²; LC-MS/MS, liquid chromatography-tandem mass spectrometry; PCOS, polycystic ovary syndrome; n/a, data not available; TT, total testosterone; cFT, calculated free testosterone; FAI, free androgen index; A4, androstenedione; DHEAS, dehydroepiandrosterone sulfate; DHT, dihydrotestosterone; SHBG, sex hormone binding globulin.

^a Median (range).

Table 2. Overall summary measures for androgen measurement using bivariate random effect model.

Index test	Studies	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	DOR (95% CI)	PLR (95% CI)	NLR (95% CI)
TT	13	0.74 (0.63–0.82)	0.86 (0.77–0.91)	0.87 (0.84–0.90)	17 (10–26.07)	5.1 (3.4–7.7)	0.31 (0.23–0.42)
cFT	6	0.89 (0.69–0.96)	0.83 (0.79–0.86)	0.85 (0.81–0.88)	38 (10–142)	5.2 (3.9–7.0)	0.14 (0.05–0.41)
FAI	11	0.78 (0.70–0.83)	0.85 (0.76–0.90)	0.87 (0.84–0.90)	19 (11–33)	5 (3.2–7.8)	0.27 (0.20–0.35)
A4	7	0.75 (0.60–0.86)	0.71 (0.51–0.85)	0.80 (0.76–0.83)	8 (3–19)	2.6 (1.5–4.7)	0.35 (0.21–0.57)
DHEAS	7	0.75 (0.61–0.85)	0.67 (0.48–0.81)	0.77 (0.73–0.81)	6 (4–8)	2.2 (1.5–3.3)	0.38 (0.29–0.51)

TT, total testosterone; cFT, calculated free testosterone; FAI, free androgen index; A4, androstenedione; DHEAS, dehydroepiandrosterone sulfate; DOR, diagnostic odds ratio; PLR, positive likelihood ratio; NLR, negative likelihood ratio.

49.46%) (Table 2, Fig. 2, and Supplementary Table S4), AUC of 0.85 (95% CI: 0.81–0.88) (Fig. 3), DOR 38 (95% CI: 10–142), PLR 5.2 (95% CI: 3.9–7.0), NLR of 0.14 (95% CI: 0.05–0.41), PPV of 87%, and NPV of 15% (Table 2 and Supplementary Fig. S1). There was no threshold effect for cFT ($\rho = -0.6$, $P = 0.2$) by Spearman's correlation, and this was confirmed on the HSROC curve (Fig. 3). Deek's funnel plot test showed no evidence of publication bias or small study effects for cFT ($P = 0.91$) (Supplementary Fig. S2).

A sensitivity analysis excluding the outlier study conducted by Imran *et al.* (2020) resulted in a decrease in sensitivity to 0.81 (95% CI: 0.72–0.88), whereas specificity remained the same with 0.83 (95% CI: 0.77–0.87). The heterogeneity of the study also reduced for both sensitivity ($I^2 = 83.53\%$) and specificity ($I^2 = 38.65\%$).

In the subgroup analysis by assay methods, the cFT based on direct immunoassay method had a sensitivity of 0.74 (95% CI: 0.64–0.82) and a specificity of 0.81 (95% CI: 0.75–0.85), while the LC-MS/MS method had a sensitivity of 0.88 (95% CI: 0.80–0.93) and a specificity of 0.87 (95% CI: 0.75–0.94). Subgroup analysis for cFT based on immunoassay after extraction was not performed as only a single study (Imran *et al.*, 2020) was available, which reported a sensitivity of 1.0, a specificity of 0.85, and an AUC of 1.0. In the subgroup analysis by diagnosis criteria, the Rotterdam criteria had a sensitivity of 0.92 (95% CI: 0.74–0.98) and a specificity of 0.84 (95% CI: 0.79–0.88), while the NIH criteria had a sensitivity of 0.76 (95% CI: 0.31–0.96) and a specificity of 0.81 (95% CI: 0.74–0.87) (Table 3). The univariate meta-regression analysis showed no statistically significant association between the assay method, but the sample size ($P < 0.001$) and PCOS diagnostic criteria ($P < 0.001$) may be a source of heterogeneity for specificity (Supplementary Fig. S3).

Free androgen index

Fifteen studies examined the FAI, including 11 studies, which calculated FAI from TT and SHBG measured with direct immunoassays (Koskinen *et al.*, 1996; Escobar-Morreale *et al.*, 2001; Hahn *et al.*, 2007; Bili *et al.*, 2014; Villarroel *et al.*, 2015; Rudnicka *et al.*, 2016; Luque-Ramirez *et al.*, 2018; Nadaraja *et al.*, 2018; Bansal *et al.*, 2020; Khashchenko *et al.*, 2020; Kumar *et al.*, 2022; Yi *et al.*, 2023), one study calculated FAI from TT and SHBG measured with immunoassay after extraction (Koskinen *et al.*, 1996), and the remaining three calculated FAI from TT measured with LC-MS/MS and SHBG measured with direct immunoassays (Sathyapalan *et al.*, 2017; Grassi *et al.*, 2020; Chen *et al.*, 2021). The reported sensitivity ranged from 0.58 to 0.91, while specificity ranged from 0.45 to 0.96. The threshold cut-off for FAI varied across studies (median 2.56, IQR 1.9–4.97).

Four studies were excluded from the meta-analyses due to the absence of sensitivity and specificity data (Koskinen *et al.*, 1996; Sathyapalan *et al.*, 2017; Luque-Ramirez *et al.*, 2018; Bansal *et al.*, 2020). Pooled results from 11 studies with 1809 participants (1140

with and 669 without PCOS) showed that FAI had a pooled sensitivity of 0.78 (95% CI: 0.70–0.83; $I^2 = 81.59\%$) and specificity of 0.85 (95% CI: 0.76–0.90; $I^2 = 86.48\%$) (Table 2, Fig. 2, and Supplementary Table S5), AUC of 0.87 (95% CI: 0.84–0.90) (Fig. 3), DOR of 19 (95% CI: 11–33), PLR of 5 (95% CI: 3.2–7.8), NLR of 0.27 (95% CI: 0.20–0.35), PPV of 89%, and NPV of 31% (Table 2 and Supplementary Fig. S1). There was no threshold effect for FAI ($\rho = 0.44$, $P = 0.17$) by Spearman's correlation and this was confirmed on the HSROC curve (Fig. 3). Deek's funnel plot test showed no evidence of publication bias or small study effects for FAI ($P = 0.16$) (Supplementary Fig. S2).

A sensitivity analysis revealed that excluding the outlier study conducted by Bili *et al.* (2014) resulted in a sensitivity of 0.76 (95% CI: 0.69–0.82), specificity of 0.87 (95% CI: 0.81–0.91), and AUC of 0.89 (95% CI: 0.85–0.91). Notably, the heterogeneity of the study remained unchanged.

In the subgroup analysis by assay methods, the direct immunoassay method had a sensitivity of 0.74 (95% CI: 0.68–0.79) and a specificity of 0.83 (95% CI: 0.73–0.90), while the LC-MS/MS method had a sensitivity of 0.89 (95% CI: 0.80–0.94) and a specificity of 0.89 (95% CI: 0.69–0.97). Subgroup analysis for FAI calculated from TT using immunoassay after extraction was not performed as only a single study was included and its sensitivity and specificity data were not reported. In the subgroup analysis by diagnosis criteria, the Rotterdam criteria had a sensitivity of 0.78 (95% CI: 0.71–0.84) and a specificity of 0.84 (95% CI: 0.74–0.91), while the NIH criteria had a sensitivity of 0.73 (95% CI: 0.51–0.87) and a specificity of 0.86 (95% CI: 0.64–0.95) (Table 3). The univariate meta-regression analysis showed no statistically significant association between the assay method (direct immunoassays vs LC-MS/MS), sample size or PCOS diagnosis criteria with sensitivity and specificity (Supplementary Fig. S3).

Serum androstenedione

Seventeen studies examined serum A4 in the diagnosis of biochemical hyperandrogenism, with 10 using direct immunoassays (Hahn *et al.*, 2007; Bili *et al.*, 2014; Villarroel *et al.*, 2015; Rudnicka *et al.*, 2016; Diamandis *et al.*, 2017; Luque-Ramirez *et al.*, 2018; Khashchenko *et al.*, 2020; Beitel *et al.*, 2022; Kumar *et al.*, 2022; Yi *et al.*, 2023), 1 study utilizing immunoassay after extraction (Koskinen *et al.*, 1996), and the remaining 6 using LC-MS/MS (Stener-Victorin *et al.*, 2010; Sathyapalan *et al.*, 2017; Cao *et al.*, 2020; Grassi *et al.*, 2020; Chen *et al.*, 2021; Xu *et al.*, 2022). The reported sensitivity ranged from 0.40 to 0.91, while specificity ranged from 0.25 to 0.91. The threshold cut-off for A4 varied across studies (median 5.22 nmol/L, IQR 4.16–13.79).

Ten studies were excluded from the meta-analyses due to the absence of sensitivity and specificity data (Koskinen *et al.*, 1996; Stener-Victorin *et al.*, 2010; Bili *et al.*, 2014; Villarroel *et al.*, 2015; Diamandis *et al.*, 2017; Sathyapalan *et al.*, 2017; ; Luque-Ramirez *et al.*, 2018; Khashchenko *et al.*, 2020; Beitel *et al.*, 2022; Xu *et al.*

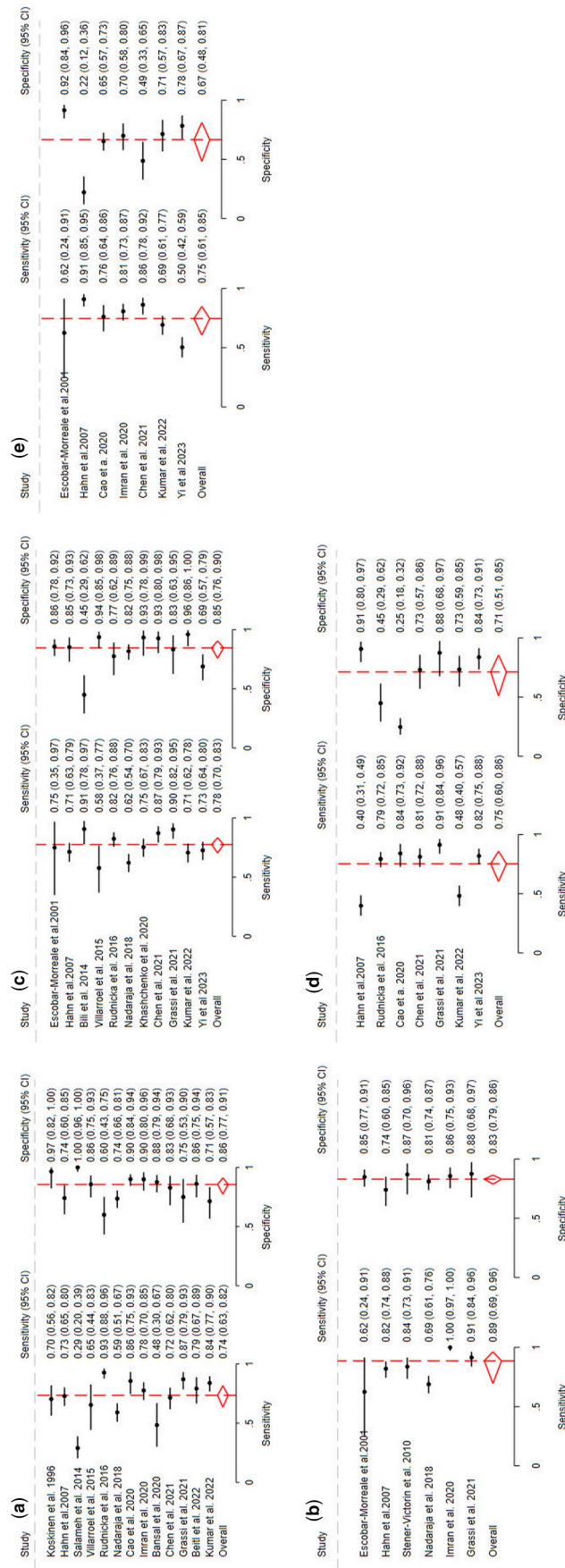


Figure 2. Forest plots of sensitivity and specificity with 95% CI for androgens. (a) Total testosterone (TT), (b) calculated free testosterone (cFT), (c) free androgen index (FAI), (d) androstenedione (A4), and (e) dehydroepiandrosterone sulfate (DHEAS).

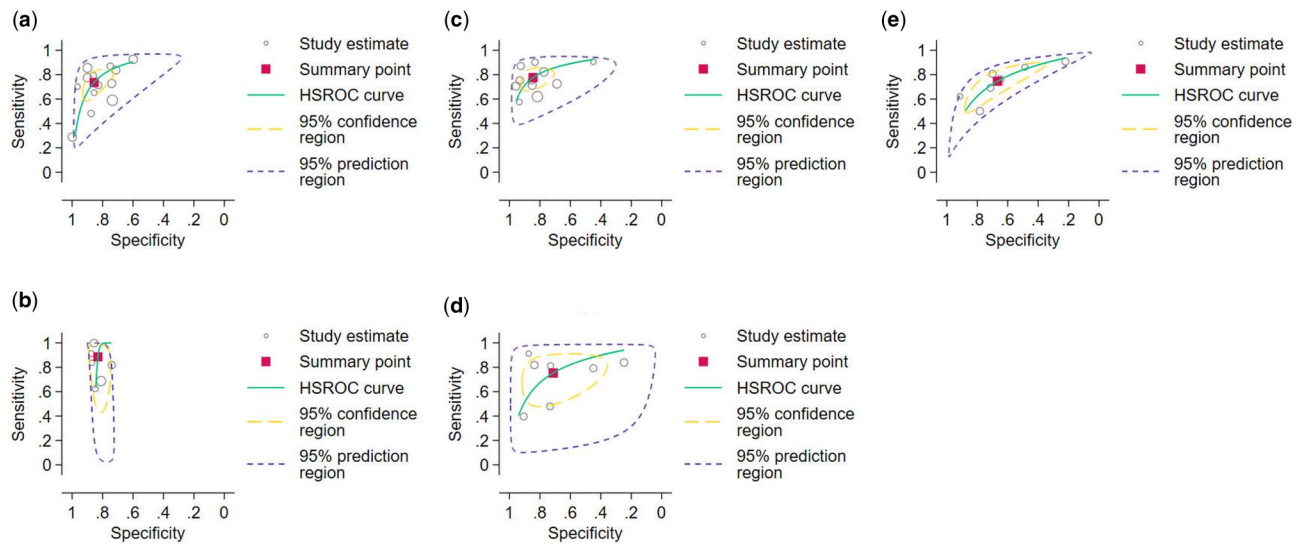


Figure 3. Hierarchical summary receiver operating characteristics (HSROC) summary curve that represents the performance of a test across different thresholds. (a) Total testosterone (TT), (b) calculated free testosterone (cFT), (c) free androgen index (FAI), (d) androstenedione (A4), and (e) dehydroepiandrosterone sulfate (DHEAS). Circles represent estimates of individual primary studies, squares represent the summary points of sensitivity and specificity. The AUC is 0.87 (95% CI: 0.84–0.90) for TT, 0.85 (95% CI: 0.81–0.88) for cFT, 0.87 (95% CI: 0.84–0.90) for FAI, 0.80 (95% CI: 0.76–0.83) for A4, and 0.77 (95% CI: 0.73–0.81) for DHEAS.

Table 3. Subgroup analysis summary measures for androgen measurement by assay methods.

Index test	Variable	Category	Studies	Sensitivity (95% CI)	Specificity (95% CI)
TT	Method of assay	Direct immunoassays	7	0.74 (0.60–0.85)	0.78 (0.67–0.86)
		LC-MS/MS	4	0.71 (0.51–0.85)	0.92 (0.83–0.96)
	Diagnostic criteria	Immunoassays after extraction	2	0.74 (0.46–0.91)	0.93 (0.80–0.98)
		NIH 1990	2	0.51 (0.28–0.74)	0.94 (0.78–0.98)
cFT	Method of assay	Direct immunoassays	3	0.74 (0.64–0.82)	0.81 (0.75–0.85)
		LC-MS/MS	2	0.88 (0.80–0.93)	0.87 (0.75–0.94)
	Diagnostic criteria	NIH 1990	2	0.76 (0.31–0.96)	0.81 (0.74–0.87)
		Rotterdam	4	0.92 (0.74–0.98)	0.84 (0.79–0.88)
FAI	Method of assay	Direct immunoassays	9	0.74 (0.68–0.79)	0.83 (0.73–0.90)
		LC-MS/MS	2	0.89 (0.80–0.94)	0.89 (0.69–0.97)
	Diagnostic criteria	NIH 1990	2	0.73 (0.51–0.87)	0.86 (0.64–0.95)
		Rotterdam	9	0.78 (0.71–0.84)	0.84 (0.74–0.91)
A4	Method of assay	Direct immunoassays	4	0.64 (0.48–0.78)	0.77 (0.53–0.91)
		LC-MS/MS	3	0.86 (0.73–0.94)	0.63 (0.32–0.86)
DHEAS	Method of assay	Direct immunoassays	4	0.67 (0.48–0.82)	0.70 (0.46–0.86)
		LC-MS/MS	2	0.82 (0.59–0.93)	0.57 (0.25–0.84)
	Diagnostic criteria	NIH 1990	2	0.77 (0.47–0.93)	0.64 (0.29–0.88)
		Rotterdam	5	0.74 (0.57–0.85)	0.68 (0.46–0.84)

TT, total testosterone; cFT, calculated free testosterone; FAI, free androgen index; A4, androstenedione; DHEAS, dehydroepiandrosterone sulfate; LC-MS/MS, liquid chromatography-tandem mass spectrometry.

2022). The pooled results of seven studies with 1275 participants (832 with and 443 without PCOS) showed that serum A4 had a pooled sensitivity of 0.75 (95% CI: 0.60–0.86; $I^2 = 95.86\%$) and specificity of 0.71 (95% CI: 0.51–0.85; $I^2 = 96.40\%$) (Table 2, Fig. 2, and Supplementary Table S6), AUC of 0.80 (95% CI: 0.76–0.83) (Fig. 3), DOR of 8 (95% CI: 3–19), PLR of 2.6 (95% CI: 1.5–4.7), and NLR of 0.35 (95% CI: 0.21–0.57), PPV of 83%, and NPV of 39% (Table 2 and Supplementary Fig. S1). There was no threshold effect for A4 ($\rho = 0.21$, $P = 0.64$) by Spearman's correlation and this was confirmed on the HSROC curve (Fig. 3). Deek's funnel plot test showed no evidence of publication bias or small study effects for A4 ($P = 0.36$) (Supplementary Fig. S2). There was no identified outlier study, so sensitivity analysis was not performed.

In the subgroup analysis by assay method, the direct immunoassay method had a sensitivity of 0.64 (95% CI: 0.48–0.78) and a specificity of 0.77 (95% CI: 0.53–0.91), while the LC-MS/MS

method had a sensitivity of 0.86 (95% CI: 0.73–0.94) and a specificity of 0.63 (95% CI: 0.32–0.86). Subgroup analysis for A4 using immunoassay after extraction was not done due to the absence of sensitivity and specificity data. Of the included studies for meta-analyses, only one study used the NIH criteria precluding subgroup analysis by diagnostic criteria (Table 3). The univariate meta-regression analysis showed no statistically significant association between the assay method (direct immunoassays vs LC-MS/MS) or sample size with sensitivity and specificity. However, PCOS diagnostic criteria may be a source of heterogeneity for sensitivity ($P < 0.05$) (Supplementary Fig. S3).

Dehydroepiandrosterone sulfate

Fourteen studies examined serum DHEAS in the diagnosis of biochemical hyperandrogenism, with 11 using direct immunoassays (Escobar-Morreale et al., 2001; Hahn et al., 2007; Bili et al., 2014;

Villarroel et al., 2015; Rudnicka et al., 2016; Diamandis et al., 2017; Luque-Ramirez et al., 2018; Bansal et al., 2020; Khashchenko et al., 2020; Kumar et al., 2022; Yi et al., 2023), 1 study using immunoassay after extraction (Imran et al., 2020), and 2 using LC-MS/MS (Cao et al., 2020; Chen et al., 2021). The reported sensitivity ranged from 0.37 to 0.91, while specificity ranged from 0.22 to 1.00. The threshold cut-off for DHEAS varied across studies (median 6.53 $\mu\text{mol/L}$, IQR 5.11–8.11).

Seven studies were excluded from the meta-analyses due to the absence of sensitivity and specificity data (Bili et al., 2014; Villarroel et al., 2015; Rudnicka et al., 2016; Diamandis et al., 2017; Luque-Ramirez et al., 2018; Bansal et al., 2020; Khashchenko et al., 2020). The pooled results from seven studies with 1267 participants (712 with and 555 without PCOS) showed that serum DHEAS had a pooled sensitivity of 0.75 (95% CI: 0.61–0.85; $I^2 = 93.31\%$) and specificity of 0.67 (95% CI: 0.48–0.81; $I^2 = 93.34\%$) (Table 2, Fig. 2, and Supplementary Table S7), AUC of 0.77 (95% CI: 0.73–0.81) (Fig. 3), DOR of 6 (95% CI: 4–8), PLR of 2.2 (95% CI: 1.5–3.3), NLR of 0.38 (95% CI: 0.29–0.51), PPV of 74%, and NPV of 33% (Table 2 and Supplementary Fig. S1). There was a threshold effect for DHEAS ($\rho = 0.93$, $P = 0.006$) by Spearman's correlation and this was confirmed by the presence of a typical 'shoulder arm' distribution on the HSROC curve (Fig. 3). Deek's funnel plot test showed no evidence of publication bias or small study effects for DHEAS ($P = 0.34$) (Supplementary Fig. S2). There was no identified outlier study, so sensitivity analysis was not performed.

In the subgroup analysis by assay methods, the direct immunoassay method had a sensitivity of 0.67 (95% CI: 0.48–0.82) and a specificity of 0.70 (95% CI: 0.46–0.86), while the LC-MS/MS method had a sensitivity of 0.82 (95% CI: 0.59–0.93) and a specificity of 0.57 (95% CI: 0.25–0.84). Subgroup analysis for immunoassay after extraction was not done because only a single study (Imran et al., 2020) was included, which reported a sensitivity of 0.81, a specificity of 0.70, and an AUC of 0.84. In the subgroup analysis by diagnosis criteria, the Rotterdam criteria had a sensitivity of 0.74 (95% CI: 0.57–0.85) and a specificity of 0.68 (95% CI: 0.46–0.84), while the NIH criteria had a sensitivity of 0.77 (95% CI: 0.47–0.93) and a specificity of 0.64 (95% CI: 0.29–0.88) (Table 3). The univariate meta-regression analysis showed no statistically significant association between the assay method (direct immunoassays vs LC-MS/MS), sample size, or PCOS diagnosis criteria with sensitivity and specificity (Supplementary Fig. S3).

Bioavailable testosterone

Two studies reported bioavailable testosterone calculated from direct immunoassay TT, SHBG, and albumin in the diagnosis of biochemical hyperandrogenism in PCOS. The first study by Hahn et al. (2007) of 187 participants (133 with and 54 without PCOS) found that calculated bioavailable testosterone at a threshold of 0.78 nmol/L had a sensitivity of 0.76, a specificity of 0.83, and AUC of 0.85. Nadaraja et al. (2018) studied 312 participants (164 with and 148 without PCOS) and reported calculated bioavailable testosterone had a sensitivity of 0.64, a specificity of 0.83, and AUC of 0.81.

Dihydrotestosterone

Six studies assessed DHT with one study using direct immunoassay (Kumar et al., 2022), one study using immunoassay after extraction (Koskinen et al., 1996), three using LC-MS/MS (Stener-Victorin et al., 2010; Cao et al., 2020; Xu et al., 2022), and the last study using both gas chromatography-mass spectrometry (GC-MS) and direct immunoassay (Kannenberg et al., 2018). Only two studies reported sensitivity and specificity data, and therefore meta-analysis was not performed. In the first study by Cao et al.

(2020) of 224 participants (63 with and 161 without PCOS), DHT levels measured by GC-MS/MS method at a threshold cut-off of ≤ 0.41 ng/ml had a sensitivity of 0.76, specificity of 0.70, and AUC of 0.81. Kannenberg et al. (2018) studied the use of DHT levels measured by GC-MS and direct immunoassay in the diagnosis of PCOS in a cohort of 43 women with and 41 without PCOS. The AUC for GC-MS was 0.73, while the AUC for direct immunoassay was 0.84. Kumar et al. (2022) also investigated the use of DHT levels measured by direct immunoassay among 186 participants (137 with and 49 without PCOS) and found that DHT levels had a sensitivity of 0.77 and specificity of 0.89 at a threshold cut-off of 382 pg/ml.

Salivary testosterone

A single study by Sathyapalan et al. (2017) examined the use of salivary testosterone levels as a diagnostic androgen marker for biochemical hyperandrogenism. Measured with LC-MS/MS among 175 participants (110 with and 65 without PCOS), salivary testosterone had an AUC of 0.76. Additionally, the combination of salivary testosterone and FAI has identified 100% of classic PCOS phenotype patients in the study, employing threshold cut-offs of 19.9 pmol/L.

Discussion

This systematic review and diagnostic test accuracy meta-analysis is the most comprehensive and contemporary synthesis of evidence on the available peer-reviewed literature on the various diagnostic accuracy of androgen biomarkers in the diagnosis of biochemical hyperandrogenism in PCOS to inform the International PCOS Guidelines (Teede et al., 2023). Where the 2018 International PCOS Guidelines only included eight studies in the systematic review, here we include 23 studies, of which 18 were meta-analyzed, significantly increasing the robustness of the results. Our results indicate that cFT, FAI, and TT have the highest sensitivity and specificity in diagnosing biochemical hyperandrogenism in PCOS. In turn, DHEAS and A4 were less accurate and had lower specificity, albeit the results of these were based on a smaller number of participants. FT functionally measured by equilibrium dialysis is the gold standard, while calculated bioavailable testosterone, DHT, and salivary testosterone show potential for detecting biochemical hyperandrogenism. However, there were too few studies available with these measures, precluding quantitative assessment of their diagnostic accuracy.

Hyperandrogenism is the most common feature of PCOS, impacting ~75% of women with the condition (Azziz et al., 2004; Alexiou et al., 2017). Currently, the routinely measured androgenic steroids in women with PCOS often includes testosterone, DHEAS, and A4. The ovaries are the main source of androgen production in premenopausal women, producing not only testosterone but also A4. The adrenal glands predominantly produce DHEA, DHEAS, and lesser amounts of A4 and testosterone. Adrenal androgens are mostly precursors of active androgens, which are taken up by the ovary and other peripheral tissues such as skin and adipose tissue, where they are activated dependent on local steroidogenic enzyme expression. Women with adrenal insufficiency and intact ovarian function have near total androgen deficiency (Arlt et al., 1999; Gurnell et al., 2008), illustrating the importance of adrenal androgen precursor production for the female androgen pool.

The most potent androgen, DHT, is primarily produced in peripheral tissues through the action of 5 α -reductase, which catalyzes its production through conversion of testosterone

(Marchetti and Barth, 2013). There is currently no clear evidence regarding the most accurate androgenic steroid to diagnose biochemical hyperandrogenism in women with PCOS. A previous systematic review in 2021 (Wang and Li, 2021) demonstrated that FAI has moderate diagnostic accuracy for PCOS and could serve as a potential biomarker to differentiate PCOS from controls but other types of androgens were not assessed. Our study provides a comprehensive assessment of the diagnostic accuracy of a wider range of markers including TT, functional FT, cFT, FAI, A4, DHEAS, bioavailable testosterone, DHT, and salivary testosterone in the diagnosis of PCOS. Our finding directly informed the 2023 International PCOS Guidelines, enabling clinicians to make informed decisions regarding biochemical hyperandrogenism testing in clinical settings.

Testosterone is the most commonly used androgen marker in diagnosing biochemical hyperandrogenism in clinical practice. It can be found in bound form with SHBG and albumin, measured as TT, and a small fraction is in the unbound form, measured or calculated as FT. Currently, radioimmunoassay and chemiluminescence immunoassays are the most commonly employed methods for measuring TT due to convenience and cost, but they are prone to interference, and accuracy may vary depending on the specific immunoassay method used (Rosner et al., 2007). Mass spectrometry, specifically LC-MS/MS, offers higher specificity and improved accuracy. As shown in our meta-analysis, serum TT had a high degree of diagnostic accuracy for biochemical hyperandrogenism in PCOS, with a pooled sensitivity of 0.74, specificity of 0.86, and AUC of 0.87. Our subgroup analysis also showed that LC-MS/MS and immunoassays after extraction had higher performance over direct immunoassay method.

FT, the biologically active form of testosterone unbound to SHBG or albumin (Rosner et al., 2007), is a more sensitive indicator of hyperandrogenemia than TT and should ideally be measured functionally using equilibrium dialysis or ammonium sulfate precipitation techniques (Goodman et al., 2015). However, functional measurement of FT can be resource-intensive and less practical for routine clinical use. In our current meta-analysis, only two studies reported FT measured by equilibrium dialysis and precluding the possibilities of quantitative synthesis, the narrative review nonetheless suggested high specificity for FT. The limited data on functionally measured FT using equilibrium dialysis in our review underscores the need for future research to explore this method and its diagnostic accuracy in the diagnosis of biochemical hyperandrogenism in PCOS. Alternatively, cFT can be estimated from high-quality TT and SHBG measurements, using various formulas, most commonly the Vermeulen equation (Vermeulen et al., 1999). Our study found that cFT is a reliable biochemical marker for diagnosing biochemical hyperandrogenism in PCOS, with a pooled sensitivity of 0.89, a specificity of 0.83, and an AUC of 0.85.

In the absence of cFT or functional FT measurement by equilibrium dialysis or ammonium sulfate precipitation techniques, FAI, derived from TT to SHBG ratio, is a simple and potentially valuable tool for diagnosing biochemical hyperandrogenism in PCOS (Mueller et al., 2006; Wang and Li, 2021). The accuracy of FAI is contingent upon the assay method used to measure TT and SHBG and can be inaccurate if the concentration of SHBG is <30 nmol/L (Keevil et al., 2018). In this meta-analysis, FAI was found to be a reliable biomarker in the diagnosis of PCOS with a pooled sensitivity of 0.78, specificity of 0.85, and AUC of 0.87. FAI calculated from measured TT by LC-MS/MS resulted in superior sensitivity and specificity compared to direct immunoassays, confirming the importance of using LC-MS/MS.

A4 is a direct precursor of testosterone, which is not only secreted by the ovaries and adrenals, but can also be produced in peripheral tissues from its precursor DHEA (O'Reilly et al., 2014). Unlike TT, A4 is not influenced by low levels of SHBG and may be a reliable marker of biochemical hyperandrogenism in PCOS (O'Reilly et al., 2014). Serum DHEAS is the most reliable measure of adrenal androgen production (Labrie et al., 2001). High concentrations of DHEAS have been reported in 33% of non-classic PCOS patients (Carmina and Longo, 2022), with rates of 20% among White and 33% among Black patients with PCOS when adjusted for age and race normative values (Kumar et al., 2005). Research showed that there is variability in the value of DHEAS as a marker for PCOS depending on age, with more value in pre- than in post-menopausal women (Kumar et al., 2005; Elhassan et al., 2018), as it reaches maximum circulating concentrations between 20 and 30 years of age followed by a steady gradual decline. Our study highlighted the need to consider A4 and DHEAS as second-line diagnostic androgen markers for PCOS, but A4 and DHEAS bring additional costs and risk of overdiagnosis due to poorer specificity than TT, cFT, or FAI. Therefore, the 2023 International PCOS Guidelines considers A4 and DHEAS not as routine diagnostic biomarkers, instead they may be considered in cases with negative first-line diagnostic biomarkers and a strong index of clinical suspicion (Teede et al., 2023).

In light of the available research, it is worth noting that while bioavailable testosterone, DHT, and salivary testosterone have shown promise as potential diagnostic markers for biochemical hyperandrogenism in PCOS, their clinical applicability remains uncertain due to limited research and variability in measurement methods. Furthermore, DHT circulates in much lower concentrations than testosterone in women, which further limits its utility for clinical use. More robust validation in larger cohorts is required to establish their reliability before routine clinical use. While our systematic review did not specifically address the use of active 11-oxygenated androgens, their potential role also warrants further investigation.

Strength and limitations

This systematic review and diagnostic meta-analysis is the most comprehensive to date on the topic, yet it has some limitations. First, only studies published in English were included, and we acknowledge that some non-English studies may have been missed. Second, substantial heterogeneity between the included studies was observed, but sensitivity analysis and meta-regression showed similar results, supporting the overall findings. Third, different studies have reported different cut-off values for each androgen measurement, with some studies reporting cut-offs that maximize sensitivity and specificity, but these cut-offs have not been validated in subsequent studies. Additionally, a potential limitation is that some studies used the Rotterdam criteria for PCOS diagnosis, which do not necessarily include patients with hyperandrogenism, potentially affecting the diagnostic accuracy assessment of biochemical hyperandrogenism. Understanding and defining the diagnostic accuracy of biochemical hyperandrogenism in PCOS in different subgroups (e.g. race-ethnic differences) and life stages (e.g. adolescents and post-menopause) is a priority identified by the 2023 International PCOS Guidelines research roadmap (Teede et al. 2023). Lastly, most studies included assessed PCOS subjects evaluated in the clinic, which is prone to referral bias, whereby participants generally have higher androgen levels than women with PCOS from unselected populations (Ezeh et al., 2013). Notwithstanding these limitations, this review comprehensively synthesized the current evidence on the predictive value of androgen markers to assess

biochemical hyperandrogenism in the diagnosis of PCOS. The review was reported in accordance with the PRISMA-DTA guidelines and incorporated meta-regression, sensitivity analysis and subgroup analysis to explore heterogeneity. Robust, multi-country studies with large sample sizes were included, improving the generalizability of the findings. The quality of studies was evaluated using the updated QUADAS-2 tool, demonstrating high-quality data with high completeness. Finally, our findings have directly informed the 2023 International PCOS Guidelines (Teede et al., 2023) and have highlighted key gaps in the evidence to guide future research in this area.

Conclusion

The findings of this review informed the 2023 International PCOS Guidelines, which now recommends that health professionals should use TT, FT, and FAI as first-line laboratory tests to assess biochemical hyperandrogenism in the diagnosis of PCOS. If they are not elevated, and clinical suspicion is high, health professionals could consider measuring A4 and DHEAS, noting their poorer specificity. Importantly, highly accurate LC-MS/MS should be utilized for assessing biochemical hyperandrogenism for TT, DHEAS, and A4, while FT should be assessed by calculation, functionally by equilibrium dialysis or ammonium sulfate precipitation. Direct androgen immunoassays have limited accuracy and results that are not consistent with the clinical presentation should be rechecked by an LC-MS/MS method. Future studies should explore the diagnostic value of other androgens and establish the optimal normative cut-off values of androgen measurements in large and ethnically diverse cohorts of unselected women to ensure wide representation and generalizability of the results to various contexts and population groups.

Supplementary data

Supplementary data are available at *Human Reproduction Update* online.

Data availability

The authors confirm that the data underlying this article are available in the article and its [online supplementary material](#).

Acknowledgements

We would like to acknowledge Thais Rocha (T.R.) for her assistance with the initial title screening process of this systematic review and diagnostic test meta-analysis.

Authors' roles

A.D.B.: search and screening of articles, data extraction, study design, data analysis, interpretation of results, and manuscript preparation. A.E.J.: supervision of data analysis, interpretation of results, and manuscript preparation. A.M.: study design, methodology, and critical revision of the manuscript. H.T.: funding, Guideline leadership, and manuscript preparation. A.E.: supervision of data analysis, and manuscript preparation. J.M.H.: information review/commentary and editing the draft manuscript. L.S.: title and abstract screening. R.A.: study design, supervision of data analysis, interpretation of results, and manuscript writing. W.A.: study design, supervision of data analysis, and manuscript preparation. C.T.T.: search and screening of articles, data

extraction, study design, supervision, interpretation of results, and manuscript preparation. All authors were involved in the data preparation and manuscript writing, and approved the final version of the manuscript.

Funding

This review was conducted to inform recommendations of the International Evidence-based Guideline in PCOS, which was funded by the Australian National Health and Medical Research Council (NHMRC) through the Centre for Research Excellence for Women's Health in Reproductive Life (CRE-WHiRL; APP1171592) and the CRE in Polycystic Ovary Syndrome (CRE-PCOS) (APP1078444), led by Monash University, Australia, and partner societies including The American Society for Reproductive Medicine, the Endocrine Society, the European Society of Endocrinology, and the European Society of Human Reproduction and Embryology. The funders were not involved in any way to influence this work. A.D.B. is supported by a PhD scholarship funded by the Monash Graduate Scholarship (MGS) and the Monash International Tuition Scholarship (MITS). H.T. and A.M. are supported by fellowships provided by the Australian NHMRC, and C.T.T. is supported by CRE-WHiRL funded by the NHMRC. W.A. received funding from the Wellcome Trust (WT209492/Z/17/Z) and the MRC (MC_UP_1605/15). J.M.H. is an MRC LMS Chain-Florey Clinical Research Training Fellow.

Conflict of interest

Dr R.A. discloses his relationships with the following entities: grants/contracts from Ferring Pharmaceuticals and Foundation for Research and Education Excellence, royalties from Wolters Kluwer Health, John Hopkins University Press, Springer, and McGraw Hill; consulting fees from Rani Therapeutics, Core Access Surgical Technologies, Fortress Biotech, May Health, Acacia Bio, and Spruce Biosciences; honoraria and conference support from the Davidson-Mestman Course and Merck; conference support from Style Paul Oration; and stock holdings in Martin Imaging and Arora Forge. Additionally, he serves on the Data Safety Monitoring Board for research studies supported by The First Affiliated Hospital of Guangzhou Medical University and the University of Michigan. He also holds leadership positions in the American Society for Reproductive Medicine, the Endocrine Society, and the Journal of Clinical Endocrinology and Metabolism. However, the author affirms that these relationships did not influence the design, conduct, or reporting of this study. The remaining authors declare no financial or personal conflicts of interest that could have impacted the outcomes of this research.

References

- Alexiou E, Hatzigelaki E, Pergialiotis V, Chrelias C, Kassanos D, Siristatidis C, Kyrkou G, Kreatsa M, Trakakis E. Hyperandrogenemia in women with polycystic ovary syndrome: prevalence, characteristics and association with body mass index. *Horm Mol Biol Clin Investig* 2017;**29**:105–111.
- Arlt W, Callies F, van Vlijmen JC, Koehler I, Reincke M, Bidlingmaier M, Huebler D, Oettel M, Ernst M, Schulte HM et al. Dehydroepiandrosterone replacement in women with adrenal insufficiency. *N Engl J Med* 1999;**341**:1013–1020.

- Ashraf S, Nabi M, Rasool S, Rashid F, Amin S. Hyperandrogenism in polycystic ovarian syndrome and role of CYP gene variants: a review. *Egypt J Med Hum Genet* 2019;**20**:1–8.
- Azziz R. Polycystic ovary syndrome. *Obstet Gynecol* 2018;**132**:321–336.
- Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Futterweit W, Janssen OE, Legro RS, Norman RJ, Taylor AE et al.; Androgen Excess Society. Criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: an androgen excess society guideline. *J Clin Endocrinol Metab* 2006;**91**:4237–4245.
- Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Futterweit W, Janssen OE, Legro RS, Norman RJ, Taylor AE et al.; 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.
- Azziz R, Sanchez L, Knochenhauer E, Moran C, Lazenby J, Stephens K, Taylor K, Boots L. Androgen excess in women: experience with over 1000 consecutive patients. *J Clin Endocrinol Metab* 2004;**89**:453–462.
- Bansal P, Sardana K, Arora P, Khurana A, Garga UC, Sharma L. A prospective study of anti-mullerian hormone and other ovarian and adrenal hormones in adult female acne. *Dermatol Ther* 2020;**33**:e13974.
- Barth JH, Field HP, Yasmin E, Balen AH. Defining hyperandrogenism in polycystic ovary syndrome: measurement of testosterone and androstenedione by liquid chromatography-tandem mass spectrometry and analysis by receiver operator characteristic plots. *Eur J Endocrinol* 2010;**162**:611–615.
- Beitl K, Dewailly D, Seemann R, Hager M, Bünker J, Mayrhofer D, Holzer I, Ott J. Polycystic ovary syndrome phenotype D versus functional hypothalamic amenorrhea with polycystic ovarian morphology: a retrospective study about a frequent differential diagnosis. *Front Endocrinol (Lausanne)* 2022;**13**:904706.
- Bili E, Dampala K, Iakovou I, Tsolakidis D, Giannakou A, Tarlatzis BC. The combination of ovarian volume and outline has better diagnostic accuracy than prostate-specific antigen (PSA) concentrations in women with polycystic ovarian syndrome (PCOs). *Eur J Obstet Gynecol Reprod Biol* 2014;**179**:32–35.
- Campbell JM, Klugar M, Ding S, Carmody DP, Hakonsen SJ, Jadotte YT, White S, Munn Z. Diagnostic test accuracy: methods for systematic review and meta-analysis. *Int J Evid Based Healthc* 2015;**13**:154–162.
- Cao Z, Lu Y, Cong Y, Liu Y, Li Y, Wang H, Zhang Q, Huang W, Liu J, Dong Y et al. Simultaneous quantitation of four androgens and 17-hydroxyprogesterone in polycystic ovarian syndrome patients by LC-MS/MS. *J Clin Lab Anal* 2020;**34**:e23539.
- Carmina E, Longo RA. Increased prevalence of elevated DHEAS in PCOS women with non-classic (B or C) phenotypes: a retrospective analysis in patients aged 20 to 29 years. *Cells* 2022;**11**:3255.
- Chen F, Chen M, Zhang W, Yin H, Chen G, Huang Q, Yang X, Chen L, Lin C, Yin G. Comparison of the efficacy of different androgens measured by LC-MS/MS in representing hyperandrogenemia and an evaluation of adrenal-origin androgens with a dexamethasone suppression test in patients with PCOS. *J Ovarian Res* 2021;**14**:32–39.
- Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 2005;**58**:882–893.
- Devillé WL, Buntinx F, Bouter LM, Montori VM, De Vet HC, Van der Windt DA, Bezemer PD. Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol* 2002;**2**:9–13.
- Diamandis EP, Stanczyk FZ, Wheeler S, Mathew A, Stengelin M, Nikolenko G, Glezer EN, Brown MD, Zheng Y, Chen Y-H et al. Serum complexed and free prostate-specific antigen (PSA) for the diagnosis of the polycystic ovarian syndrome (PCOS). *Clin Chem Lab Med* 2017;**55**:1789–1797.
- Elhassan YS, Idkowiak J, Smith K, Asia M, Gleeson H, Webster R, Arlt W, O'Reilly MW. Causes, patterns, and severity of androgen excess in 1205 consecutively recruited women. *J Clin Endocrinol Metab* 2018;**103**:1214–1223.
- Escobar-Morreale HF, Asuncion M, Calvo RM, Sancho J, San MJL. Receiver operating characteristic analysis of the performance of basal serum hormone profiles for the diagnosis of polycystic ovary syndrome in epidemiological studies. *Eur J Endocrinol* 2001;**145**:619–624.
- 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–E1096.
- Goodarzi MO, Carmina E, Azziz R. DHEA, DHEAS and PCOS. *J Steroid Biochem Mol Biol* 2015;**145**:213–225.
- Goodman NF, Cobin RH, Futterweit W, Glueck JS, Legro RS, Carmina E; Androgen Excess and PCOS Society (AES). American Association of Clinical Endocrinologists, American College of Endocrinology, and androgen excess and PCOS society disease state clinical review: guide to the best practices in the evaluation and treatment of polycystic ovary syndrome—part 1. *Endocr Pract* 2015;**21**:1291–1300.
- Grassi G, Polledri E, Fustinoni S, Chiodini I, Ceriotti F, D'Agostino S, Filippi F, Somigliana E, Mantovani G, Arosio M et al. Hyperandrogenism by liquid chromatography tandem mass spectrometry in PCOS: focus on testosterone and androstenedione. *J Clin Med* 2020;**10**:119.
- Gurnell EM, Hunt PJ, Curran SE, Conway CL, Pullenayegum EM, Huppert FA, Compston JE, Herbert J, Chatterjee VK. Long-term DHEA replacement in primary adrenal insufficiency: a randomized, controlled trial. *J Clin Endocrinol Metab* 2008;**93**:400–409.
- Hahn S, Kuehnelt W, Tan S, Kramer K, Schmidt M, Roesler S, Kimmig R, Mann K, Janssen OE. Diagnostic value of calculated testosterone indices in the assessment of polycystic ovary syndrome. *Clin Chem Lab Med* 2007;**45**:202–207.
- Harbord RM, Whiting P. Metandi: meta-analysis of diagnostic accuracy using hierarchical logistic regression. *Stata J* 2009;**9**:211–229.
- 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.
- Imran HJ, Dhaher SA, Mansour AA. Testosterone or dehydroepiandrosterone sulfate as a biomarker for hirsutism in women with polycystic ovary syndrome. *Biomed Pharmacol J* 2020;**13**:1815–1823.
- Janse F, Eijkemans MJ, Goverde AJ, Lentjes EG, Hoek A, Lambalk CB, Hickey TE, Fauser BC, Norman RJ. Assessment of androgen concentration in women: liquid chromatography-tandem mass spectrometry and extraction RIA show comparable results. *Eur J Endocrinol* 2011;**165**:925–933.
- Kannenberg F, Fobker M, Schulte E, Pierściński G, Kelsch R, Zitzmann M, Nofer J-R, Schüring AN. The simultaneous measurement of serum testosterone and 5 α -dihydrotestosterone by gas chromatography-mass spectrometry (GC-MS). *Clin Chim Acta* 2018;**476**:15–24.
- Keevil BG. Novel liquid chromatography tandem mass spectrometry (LC-MS/MS) methods for measuring steroids. *Best Pract Res Clin Endocrinol Metab* 2013;**27**:663–674.

- Keevil BG, Adaway J, Fiers T, Moghetti P, Kaufman JM. The free androgen index is inaccurate in women when the SHBG concentration is low. *Clin Endocrinol (Oxf)* 2018;**88**:706–710.
- Khashchenko E, Uvarova E, Vysokikh M, Ivanets T, Krechetova L, Tarasova N, Sukhanova I, Mamedova F, Borovikov P, Balashov I et al. The relevant hormonal levels and diagnostic features of polycystic ovary syndrome in adolescents. *J Clin Med* 2020;**9**:1831.
- Koskinen P, Penttilä T-A, Anttila L, Erkkola R, Irjala K. Optimal use of hormone determinations in the biochemical diagnosis of the polycystic ovary syndrome. *Fertil Steril* 1996;**65**:517–522.
- Kumar A, Woods KS, Bartolucci AA, Azziz R. Prevalence of adrenal androgen excess in patients with the polycystic ovary syndrome (PCOS). *Clin Endocrinol (Oxf)* 2005;**62**:644–649.
- Kumar H, Halder A, Sharma M, Jain M, Kalsi AK. Dihydrotestosterone—a potential biomarker of hyperandrogenaemia in polycystic ovary syndrome: a case-control study from North India. *J Clin Diagn Res* 2022;**16**:9–14.
- Labrie F, Luu-The V, Labrie C, Simard J. DHEA and its transformation into androgens and estrogens in peripheral target tissues: intracrinology. *Front Neuroendocrinol* 2001;**22**:185–212.
- Lee J, Kim KW, Choi SH, Huh J, Park SH. Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: a practical review for clinical researchers—part II. Statistical methods of meta-analysis. *Korean J Radiol* 2015;**16**:1188–1196.
- Lizneva D, Suturina L, Walker W, Brakta S, Gavrilova-Jordan L, Azziz R. Criteria, prevalence, and phenotypes of polycystic ovary syndrome. *Fertil Steril* 2016;**106**:6–15.
- Luque-Ramirez M, Jimenez-Mendiguchia L, Garcia-Cano A, Fernandez-Duran E, de Dios Rosa V, Nattero-Chavez L, Ortiz-Flores AE, Escobar-Morreale HF. Certified testosterone immunoassays for hyperandrogenaemia. *Eur J Clin Invest* 2018;**48**:e13029.
- Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0*. The Cochrane Collaboration, 2010. Available from: <http://srdta.cochrane.org/>.
- 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.
- Marchetti PM, Barth JH. Clinical biochemistry of dihydrotestosterone. *Ann Clin Biochem* 2013;**50**:95–107.
- McInnes MDF, Moher D, Thoms BD, McGrath TA, Bossuyt PM, Clifford T, Cohen JF, Deeks JJ, Gatsonis C, Hooft L et al.; the PRISMA-DTA Group. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA* 2018;**319**:388–396.
- Miller KK, Rosner W, Lee H, Hier J, Sesmilo G, Schoenfeld D, Neubauer G, Klubanski A. Measurement of free testosterone in normal women and women with androgen deficiency: comparison of methods. *J Clin Endocrinol Metab* 2004;**89**:525–533.
- Mousa A, Tay C, Teede H. Technical Report for the 2023 International Evidence-based Guideline for the Assessment and Management of Polycystic Ovary Syndrome. Monash University, 2023. https://bridges.monash.edu/articles/report/Technical_Report_for_the_2023_International_Evidence-based_Guideline_for_the_Assessment_and_Management_of_Polycystic_Ovary_Syndrome/23625288/1 (5 Sep 2023, date last accessed).
- Mueller A, Dittrich R, Cupisti S, Beckmann M, Binder H. Is it necessary to measure free testosterone to assess hyperandrogenemia in women? The role of calculated free and bioavailable testosterone. *Exp Clin Endocrinol Diabetes* 2006;**114**:182–187.
- Nadaraja RND, Sthaneshwar P, Razali N. Establishing the cut off values of androgen markers in the assessment of polycystic ovarian syndrome. *Malays J Pathol* 2018;**40**:33–39.
- O'Reilly MW, Taylor AE, Crabtree NJ, Hughes BA, Capper F, Crowley RK, Stewart PM, Tomlinson JW, Arlt W. Hyperandrogenemia predicts metabolic phenotype in polycystic ovary syndrome: the utility of serum androstenedione. *J Clin Endocrinol Metab* 2014;**99**:1027–1036.
- Owen LJ, Keevil BG. Testosterone measurement by liquid chromatography tandem mass spectrometry: the importance of internal standard choice. *Ann Clin Biochem* 2012;**49**:600–602.
- Owen LJ, Wu FC, Büttler RM, Keevil BG. A direct assay for the routine measurement of testosterone, androstenedione, dihydrotestosterone and dehydroepiandrosterone by liquid chromatography tandem mass spectrometry. *Ann Clin Biochem* 2016;**53**:580–587.
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Int J Surg* 2021;**88**:105906.
- Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;**58**:982–990.
- Rosenfield RL, Ehrmann DA. The pathogenesis of polycystic ovary syndrome (PCOS): the hypothesis of PCOS as functional ovarian hyperandrogenism revisited. *Endocr Rev* 2016;**37**:467–520.
- Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H. Limitations, and pitfalls in measuring testosterone: an endocrine society position statement. *J Clin Endocrinol Metab* 2007;**92**:405–413.
- Rosner W, Vesper H; Pediatric Endocrine Society. Toward excellence in testosterone testing: a consensus statement. *J Clin Endocrinol Metab* 2010;**95**:4542–4548.
- Rotterdam ESHRE ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod* 2004;**19**:41–47.
- Rudnicka E, Radowski S, Suchta K. Prostate specific antigen (PSA) in diagnosis of polycystic ovarian syndrome—a new insight. *Gynecol Endocrinol* 2016;**32**:931–935.
- Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001;**20**:2865–2884.
- Salameh JP, Bossuyt PM, McGrath TA, Thoms BD, Hyde CJ, Macaskill P, Deeks JJ, Leeftang M, Korevaar DA, Whiting P et al. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. *BMJ* 2020;**370**:m2632.
- Salameh WA, Redor-Goldman MM, Clarke NJ, Mathur R, Azziz R, Reitz RE. Specificity and predictive value of circulating testosterone assessed by tandem mass spectrometry for the diagnosis of polycystic ovary syndrome by the National Institutes of Health 1990 criteria. *Fertil Steril* 2014;**101**:1135–1141.e2.
- Sathyapalan T, Al-Qaissi A, Kilpatrick ES, Dargham SR, Adaway J, Keevil B, Atkin SL. Salivary testosterone measurement in women with and without polycystic ovary syndrome. *Sci Rep* 2017;**7**:3589.
- Simundic AM. Measures of diagnostic accuracy: basic definitions. *EJIFCC* 2009;**19**:203–211.
- StataCorp L. *Stata Statistical Software: Release 18* (2023). College Station, TX: StataCorp LP, 2023.
- Stener-Victorin E, Holm G, Labrie F, Nilsson L, Janson PO, Ohlsson C. Are there any sensitive and specific sex steroid markers for polycystic ovary syndrome? *J Clin Endocrinol Metab* 2010;**95**:810–819.
- 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.

- Teede H, Deeks A, Moran L. Polycystic ovary syndrome: a complex condition with psychological, reproductive and metabolic manifestations that impacts on health across the lifespan. *BMC Med* 2010;**8**:41–10.
- Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, Piltonen T, Norman RJ; International PCOS Network. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Hum Reprod* 2018;**33**:1602–1618.
- Teede HJ, Tay CT, Laven JJE, Dokras A, Moran LJ, Piltonen TT, Costello MF, Boivin J, Redman LM, Boyle JA et al. Recommendations from the 2023 international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *J Clin Endocrinol Metab* 2023;**108**:2447–2469.
- van Enst WA, Ochodo E, Scholten RJ, Hooft L, Leeftang MM. Investigation of publication bias in meta-analyses of diagnostic test accuracy: a meta-epidemiological study. *BMC Med Res Methodol* 2014;**14**:70.
- Vermeulen A, Verdonck L, Kaufman JM. A critical evaluation of simple methods for the estimation of free testosterone in serum. *J Clin Endocrinol Metab* 1999;**84**:3666–3672.
- Villarreal C, López P, Merino PM, Iñiguez G, Sir-Petermann T, Codner E. Hirsutism and oligomenorrhea are appropriate screening criteria for polycystic ovary syndrome in adolescents. *Gynecol Endocrinol* 2015;**31**:625–629.
- Wang L, Li J. The value of serum-free androgen index in the diagnosis of polycystic ovary syndrome: a systematic review and meta-analysis. *J Obstet Gynaecol Res* 2021;**47**:1221–1231.
- Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeftang MM, Sterne JA, Bossuyt PM; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;**155**:529–536.
- Xu S, Liu Y, Xue K, Liu X, Jia G, Zeng Y, Chen Y. Diagnostic value of total testosterone and free androgen index measured by LC-MS/MS for PCOS and insulin resistance. *J Clin Lab Anal* 2022;**36**:e24739.
- Yang B, Mallett S, Takwoingi Y, Davenport CF, Hyde CJ, Whiting PF, Deeks JJ, Leeftang MMG, Bossuyt PMM, Brazzelli MG et al.; QUADAS-C Group. QUADAS-C: a tool for assessing risk of bias in comparative diagnostic accuracy studies. *Ann Intern Med* 2021;**174**:1592–1599.
- Yi W, Zhang M, Yuan X, Shi L, Yuan X, Sun M, Liu J, Cai H, Lv Z. A model combining testosterone, androstenedione and free testosterone index improved the diagnostic efficiency of polycystic ovary syndrome. *Endocr Pract* 2023;**29**:629–636.
- Zawadzki JK, Dunaif A. Diagnostic criteria for polycystic ovary syndrome: towards a rational approach. In: Dunaif A, Givens JR, Haseltine F (eds). *Polycystic Ovary Syndrome*. Boston: Blackwell Scientific, 1992:377–384.