



What is the diagnostic accuracy of different clinical diagnostic criteria (Rotterdam, NIH, and Androgen Excess Society) for identifying polycystic ovary syndrome in women of reproductive age? : A Systematic Review

¹ Amanda Ezra Natasya Napitupulu, ² Jenary Immanuel Surbakti

¹ General Practitioner, Karubaga Regional General Hospital, Tolikara Regency, Highland Papua, Indonesia

² Obstetric and Gynaecology Specialist, Bagan Batu, Rokan Hilir Riau, Indonesia / Graduate of Faculty of Medicine, University of North Sumatera, Indonesia

Corresponding Email : amanda.napitupulu@yahoo.com

Article History :

Received date : 2026/01/21

Revised date : 2026/02/07

Accepted date : 2026/03/11

Published date : 2026/04/15



Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (BY NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).

E-ISSN :

ISSN 3048-1368



P-ISSN

ISSN 3048-1376



ABSTRACT

Introduction: Polycystic ovary syndrome (PCOS) is a complex endocrine disorder affecting women of reproductive age, with multiple diagnostic criteria currently in use including Rotterdam, NIH, and AES criteria. However, the diagnostic accuracy of these criteria remains variable across populations. This systematic review aimed to evaluate and compare the diagnostic accuracy of different clinical diagnostic criteria for identifying PCOS in reproductive-age women.

Methods: A systematic review of diagnostic accuracy studies was conducted. Studies were included if they evaluated at least one of the three specified diagnostic criteria (Rotterdam, NIH, or AES) against a reference standard in women of reproductive age (15-45

years). Diagnostic accuracy measures including sensitivity, specificity, and area under the ROC curve (AUC) were extracted. The quality of included studies was assessed using appropriate diagnostic accuracy assessment tools.

Results: Eighty studies encompassing diverse populations across North America, Europe, Asia, Africa, and the Middle East were included. The Rotterdam criteria demonstrated strong diagnostic utility with follicle number per ovary showing the highest accuracy (sensitivity 84%, specificity 91%, AUC 0.905). NIH criteria identified fewer women (27.1% prevalence) compared to Rotterdam (40%) and showed an AUC of 0.80 for AMH as a diagnostic marker. AES criteria yielded intermediate prevalence (29.3%) with AMH sensitivity of 84.4% and specificity of 72% (AUC 0.857). Anti-Müllerian hormone emerged as a promising biomarker with age-specific thresholds ranging from 5.7 ng/mL (20-27 years) to 3.72 ng/mL (35-40 years). Replacing PCOM with AMH in Rotterdam criteria improved diagnostic accuracy (AUC 0.934-0.97). Significant geographic and ethnic variations in optimal thresholds were observed.

Discussion: The Rotterdam criteria demonstrate superior sensitivity but may overdiagnose milder phenotypes, while NIH criteria identify metabolically high-risk women with greater specificity. AES criteria provide an intermediate approach emphasizing androgen excess. Age-stratified and population-specific thresholds are essential for optimal diagnostic accuracy. AMH shows promise as an objective alternative to ultrasound assessment of PCOM.

Conclusion: No single diagnostic criterion is universally optimal; the choice of criteria should be guided by clinical context, population characteristics, and available resources. Age-stratified

and population-specific thresholds, particularly for AMH and ultrasound parameters, are recommended to improve diagnostic accuracy.

Keywords: Polycystic ovary syndrome, diagnostic criteria, Rotterdam criteria, NIH criteria, AES criteria, anti-Müllerian hormone, diagnostic accuracy, systematic review

INTRODUCTION

Background

Polycystic ovary syndrome (PCOS) represents the most common endocrine disorder affecting women of reproductive age, with prevalence estimates ranging from 6% to 20% depending on the diagnostic criteria applied [1, 7]. The syndrome is characterized by a complex interplay of reproductive, metabolic, and psychological features, including oligo-ovulation or anovulation, clinical or biochemical hyperandrogenism, and polycystic ovarian morphology (PCOM) [2]. Beyond reproductive manifestations, PCOS is associated with significant long-term health consequences including type 2 diabetes mellitus, cardiovascular disease, endometrial cancer, and decreased quality of life [4, 20].

The diagnostic landscape for PCOS has evolved substantially since the condition was first described by Stein and Leventhal in 1935. Three major diagnostic frameworks have emerged over the past three decades, each reflecting different conceptualizations of the syndrome's essential features. The National Institutes of Health (NIH) criteria, established in 1990, require the presence of both chronic oligo-ovulation and clinical or biochemical hyperandrogenism, with exclusion of other etiologies [2, 8]. This approach represents the most stringent diagnostic standard, emphasizing androgen excess and ovulatory dysfunction as cardinal features.

The Rotterdam criteria, introduced in 2003 following a joint consensus workshop between the European Society for Human Reproduction and Embryology (ESHRE) and the American Society for Reproductive Medicine (ASRM), expanded the diagnostic framework by requiring at least two of three features: oligo-ovulation or anovulation, clinical or biochemical hyperandrogenism, and polycystic ovarian morphology on ultrasound [1, 3]. This broader definition increased the diagnosed prevalence of PCOS and introduced four distinct phenotypes, ranging from the classic NIH phenotype (hyperandrogenism plus oligo-ovulation) to milder presentations including ovulatory women with PCOM and hyperandrogenism [5, 6].

The Androgen Excess Society (AES) criteria, published in 2006, sought to reconcile these approaches by requiring the presence of hyperandrogenism (clinical or biochemical) as an essential feature, accompanied by either ovulatory dysfunction or polycystic ovarian morphology [2, 8]. This framework positions androgen excess as the central pathophysiological feature while acknowledging the heterogeneity of associated manifestations.

Problem Statement

Despite widespread clinical adoption of these diagnostic criteria, significant uncertainty persists regarding their comparative diagnostic accuracy. Healthcare providers face the challenge of selecting among three valid but discordant diagnostic frameworks, each potentially identifying different patient populations with varying implications for clinical management and prognosis [7, 9]. The absence of a true gold standard for PCOS diagnosis compounds this difficulty, as all current criteria represent expert consensus definitions rather than pathophysiologically validated entities [11].

Furthermore, emerging evidence suggests that the diagnostic performance of these criteria may vary substantially across different populations, age groups, and clinical settings [4, 48]. The original Rotterdam ultrasound thresholds for PCOM (≥ 12 follicles measuring 2-9 mm or ovarian volume $>10 \text{ cm}^3$) were established using older ultrasound technology and may not reflect contemporary imaging capabilities [11]. Similarly, biochemical thresholds for hyperandrogenism demonstrate significant assay-dependent and population-specific variation [55, 78].

The role of anti-Müllerian hormone (AMH) as a potential diagnostic biomarker has garnered considerable attention, with multiple studies suggesting that AMH may serve as a reliable surrogate for ultrasound assessment of PCOM [9, 12, 45]. However, the optimal AMH thresholds for diagnostic purposes remain debated, with substantial variation across studies related to assay methods, age distributions, and population characteristics [4, 32, 39].

Research Gap

Several critical gaps persist in the current literature regarding diagnostic accuracy of PCOS criteria. First, while individual studies have compared diagnostic criteria in specific populations, no comprehensive synthesis has systematically evaluated the comparative performance of all three

criteria across diverse settings [1, 23]. Second, the influence of demographic factors—particularly age and ethnicity—on diagnostic accuracy remains inadequately characterized, limiting the generalizability of existing thresholds [4, 41, 48]. Third, the potential for AMH to replace or supplement ultrasound assessment of PCOM has been examined in multiple studies but without consensus regarding optimal implementation strategies [9, 22, 27]. Fourth, geographic and ethnic variations in optimal diagnostic thresholds have been documented but not systematically integrated into clinical recommendations [25, 41, 59]. Fifth, the diagnostic accuracy of different criteria across specific PCOS phenotypes (A through D) requires further elucidation to guide phenotype-specific management approaches [6, 8, 57].

Research Question

What is the diagnostic accuracy of different clinical diagnostic criteria (Rotterdam, NIH, and Androgen Excess Society) for identifying polycystic ovary syndrome in women of reproductive age?

Objectives

The primary objective of this systematic review was to evaluate and compare the diagnostic accuracy of Rotterdam, NIH, and AES criteria for PCOS diagnosis in reproductive-age women. Secondary objectives included: (1) to assess the diagnostic performance of AMH as a potential biomarker for PCOS across different criteria and populations; (2) to identify optimal age-specific and population-specific thresholds for diagnostic markers; (3) to evaluate the accuracy of modified ultrasound criteria and alternative biochemical markers; and (4) to synthesize evidence regarding geographic and ethnic variations in diagnostic accuracy.

Hypotheses

We hypothesized that the Rotterdam criteria would demonstrate higher sensitivity but lower specificity compared to NIH and AES criteria, reflecting their broader phenotypic inclusion. We further hypothesized that AMH would show good diagnostic accuracy as a surrogate marker for PCOM, with optimal thresholds varying significantly by age and population. Finally, we anticipated that significant geographic and ethnic variations in diagnostic thresholds would necessitate population-specific calibration of diagnostic criteria.

Significance and Novelty

This systematic review provides several novel contributions to the PCOS diagnostic literature. To our knowledge, this represents the most comprehensive synthesis of diagnostic accuracy evidence across all three major PCOS criteria simultaneously, encompassing diverse global populations. The inclusion of studies from North America, Europe, Asia, Africa, and the Middle East enables unprecedented examination of geographic and ethnic variations in diagnostic performance. The systematic evaluation of AMH across different diagnostic frameworks provides evidence-based guidance for incorporating this biomarker into clinical practice. The identification of age-stratified and population-specific thresholds addresses critical gaps in the generalizability of existing diagnostic standards. Finally, this review synthesizes emerging evidence on novel biomarkers and combined diagnostic approaches, offering a roadmap for future diagnostic refinement.

METHODS

Protocol

The study strictly adhered to the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) 2020 guidelines to ensure methodological rigor and accuracy. This approach was chosen to enhance the precision and reliability of the conclusions drawn from the investigation.

Criteria for Eligibility

This systematic review aims to evaluate What is the diagnostic accuracy of different clinical diagnostic criteria (Rotterdam, NIH, and Androgen Excess Society) for identifying polycystic ovary syndrome in women of reproductive age?.

Screening

We screened in sources based on their abstracts that met these criteria:

- **Population Age Range:** Does the study include women of reproductive age (typically 15-45 years) as the primary study population?

- **Diagnostic Criteria Evaluated:** Does the study evaluate at least one of the three specified diagnostic criteria: Rotterdam criteria, NIH criteria, or Androgen Excess Society criteria?
- **Reference Standard:** Does the study have a clearly defined reference standard for PCOS diagnosis against which the diagnostic criteria are compared?
- **Study Groups:** Does the study include both women with PCOS and control women without PCOS?
- **Diagnostic Accuracy Outcomes:** Does the study report at least one diagnostic accuracy measure (sensitivity, specificity, positive predictive value, negative predictive value, likelihood ratios, or area under the ROC curve)?
- **Study Design:** Is the study design appropriate for diagnostic accuracy assessment (diagnostic accuracy study, cross-sectional study, cohort study, case-control study, systematic review, or meta-analysis)?
- **Study Focus:** Is the study focused on diagnostic accuracy rather than therapeutic interventions (i.e., is it NOT a randomized controlled trial evaluating treatments)?
- **Standard Criteria:** Does the study use the standard Rotterdam, NIH, or Androgen Excess Society criteria without significant modifications?
- **Sample Size:** Does the study include at least 10 participants per group (both PCOS and control groups), and is it NOT a case report or case series?

We considered all screening questions together and made a holistic judgement about whether to screen in each paper.

Search Strategy

The keywords used for this research based PICO :

Element	P (Population)	I (Intervention/Exposure)	C (Comparison/Context)	O (Outcome)
Keyword 1	Women of reproductive age	Diagnostic criteria	Diagnostic accuracy comparison	Diagnostic accuracy
Keyword 2	Non-menopausal women	Rotterdam criteria	Reference standard	Sensitivity and Specificity
Keyword 3	Women of childbearing age	NIH criteria	Comparative study	Predictive value
Keyword 4	Females (15-45 years)	AES-PCOS criteria	Criteria performance	Area under the curve (AUC)

The Boolean MeSH keywords inputted on databases for this research are: (*"Women of reproductive age" OR "Non-menopausal women" OR "Women of childbearing age" OR "Females (15-45 years)"*) AND (*"Diagnostic criteria" OR "Rotterdam criteria" OR "NIH criteria" OR "AES-PCOS criteria"*) AND (*"Diagnostic accuracy comparison" OR "Comparative study" OR "Criteria performance"*) AND (*"Diagnostic accuracy" OR "Sensitivity and Specificity" OR "Predictive value" OR "Area under the curve (AUC)"*)

Data extraction

- **Study Design:**

Extract study design and methodology details relevant to diagnostic accuracy assessment, including:

- Study type (cross-sectional, case-control, cohort, etc)
- Setting (clinic, hospital, population-based)
- Recruitment method and inclusion/exclusion criteria
- Sample size and power calculation (if reported)
- Study quality issues that might affect diagnostic accuracy results

- **Study Population:**

Extract characteristics of women studied for PCOS diagnostic accuracy, including:

- Age range and mean age (must be reproductive age)
- Population type (infertile women, symptomatic women, general population screening)
- BMI range and distribution
- Ethnicity/race if reported
- Exclusion of other conditions that might affect diagnosis
- Any population characteristics that might affect generalizability of diagnostic accuracy results

- **Diagnostic Criteria:**

Extract which specific PCOS diagnostic criteria were evaluated and compared, including:

- Rotterdam criteria (specify which components: oligo-ovulation, hyperandrogenism, polycystic ovaries)
- NIH criteria (specify which components: oligo-ovulation and hyperandrogenism)
- Androgen Excess Society (AES) criteria (specify which components)
- Any modifications to standard criteria (e.g., different ultrasound thresholds, alternative biochemical markers)
- Exact definitions used for each component (e.g., follicle count thresholds, testosterone levels)

- **Reference Standard:**

Extract what was used as the reference standard or gold standard for determining true PCOS diagnosis for diagnostic accuracy calculations:

- Which diagnostic criteria served as the reference (if comparing criteria against each other)
- Clinical expert consensus
- Long-term follow-up outcomes
- Combination of multiple criteria
- If no clear reference standard, note this limitation and how diagnostic accuracy was calculated

- **Diagnostic Accuracy:**

Extract diagnostic accuracy metrics for each diagnostic criterion evaluated (Rotterdam, NIH, AES), including:

- Sensitivity (%) with confidence intervals
- Specificity (%) with confidence intervals
- Area under ROC curve (AUC) with confidence intervals
- Positive predictive value (PPV) and negative predictive value (NPV)
- Likelihood ratios (positive and negative)
- Optimal cut-off thresholds (if applicable)
- Prevalence of PCOS according to each diagnostic criterion
- Direct comparisons between criteria performance

- **Additional Markers:**

Extract information about any additional diagnostic markers or modifications tested alongside or instead of standard criteria, including:

- Anti-Müllerian hormone (AMH) levels and cut-off thresholds
- Modified ultrasound criteria (different follicle counts, ovarian volume thresholds)
- Alternative biochemical markers
- Combined diagnostic approaches (e.g., AMH + clinical symptoms)
- Diagnostic accuracy results for these additional/modified approaches
- How these compare to standard Rotterdam/NIH/AES criteria

- **Key Findings:**

Extract the main conclusions about diagnostic accuracy of different PCOS criteria, including:

- Which diagnostic criteria performed best in this population
- Specific advantages or limitations of each criterion (Rotterdam, NIH, AES)
- Clinical implications for diagnosis in reproductive-age women
- Recommendations for clinical practice
- Any identified gaps or limitations in current diagnostic approaches
- Authors' conclusions about optimal diagnostic strategy

Table 1. Article Search Strategy

Database	Keywords	Hits
Pubmed	<i>("Women of reproductive age" OR "Non-menopausal women" OR "Women of childbearing age" OR "Females (15-45 years)" AND ("Diagnostic criteria" OR "Rotterdam criteria" OR "NIH criteria" OR "AES-PCOS criteria") AND ("Diagnostic accuracy comparison" OR "Comparative study" OR "Criteria performance") AND ("Diagnostic accuracy" OR "Sensitivity and Specificity" OR "Predictive value" OR "Area under the curve (AUC)")</i>	6
Semantic Scholar	<i>("Women of reproductive age" OR "Non-menopausal women" OR "Women of childbearing age" OR "Females (15-45 years)" AND ("Diagnostic criteria" OR "Rotterdam criteria" OR "NIH criteria" OR "AES-PCOS criteria") AND ("Diagnostic accuracy comparison" OR "Comparative study" OR "Criteria performance") AND ("Diagnostic accuracy" OR "Sensitivity and Specificity" OR "Predictive value" OR "Area under the curve (AUC)")</i>	1
Springer	<i>("Women of reproductive age" OR "Non-menopausal women" OR "Women of childbearing age" OR "Females (15-45 years)" AND ("Diagnostic criteria" OR "Rotterdam criteria" OR "NIH criteria" OR "AES-PCOS criteria") AND ("Diagnostic accuracy comparison" OR "Comparative study" OR "Criteria performance") AND ("Diagnostic accuracy" OR "Sensitivity and Specificity" OR "Predictive value" OR "Area under the curve (AUC)")</i>	246
Google Scholar	<i>("Women of reproductive age" OR "Non-menopausal women" OR "Women of childbearing age" OR "Females (15-45 years)" AND ("Diagnostic criteria" OR "Rotterdam criteria" OR "NIH criteria" OR "AES-PCOS criteria") AND ("Diagnostic accuracy comparison" OR "Comparative study" OR "Criteria performance") AND ("Diagnostic accuracy" OR "Sensitivity and Specificity" OR "Predictive value" OR "Area under the curve (AUC)")</i>	2,890
Wiley Online Library	<i>("Women of reproductive age" OR "Non-menopausal women" OR "Women of childbearing age" OR "Females (15-45 years)" AND ("Diagnostic criteria" OR "Rotterdam criteria" OR "NIH criteria" OR "AES-PCOS criteria") AND ("Diagnostic accuracy comparison" OR "Comparative study" OR "Criteria performance") AND ("Diagnostic accuracy" OR "Sensitivity and Specificity" OR "Predictive value" OR "Area under the curve (AUC)")</i>	413

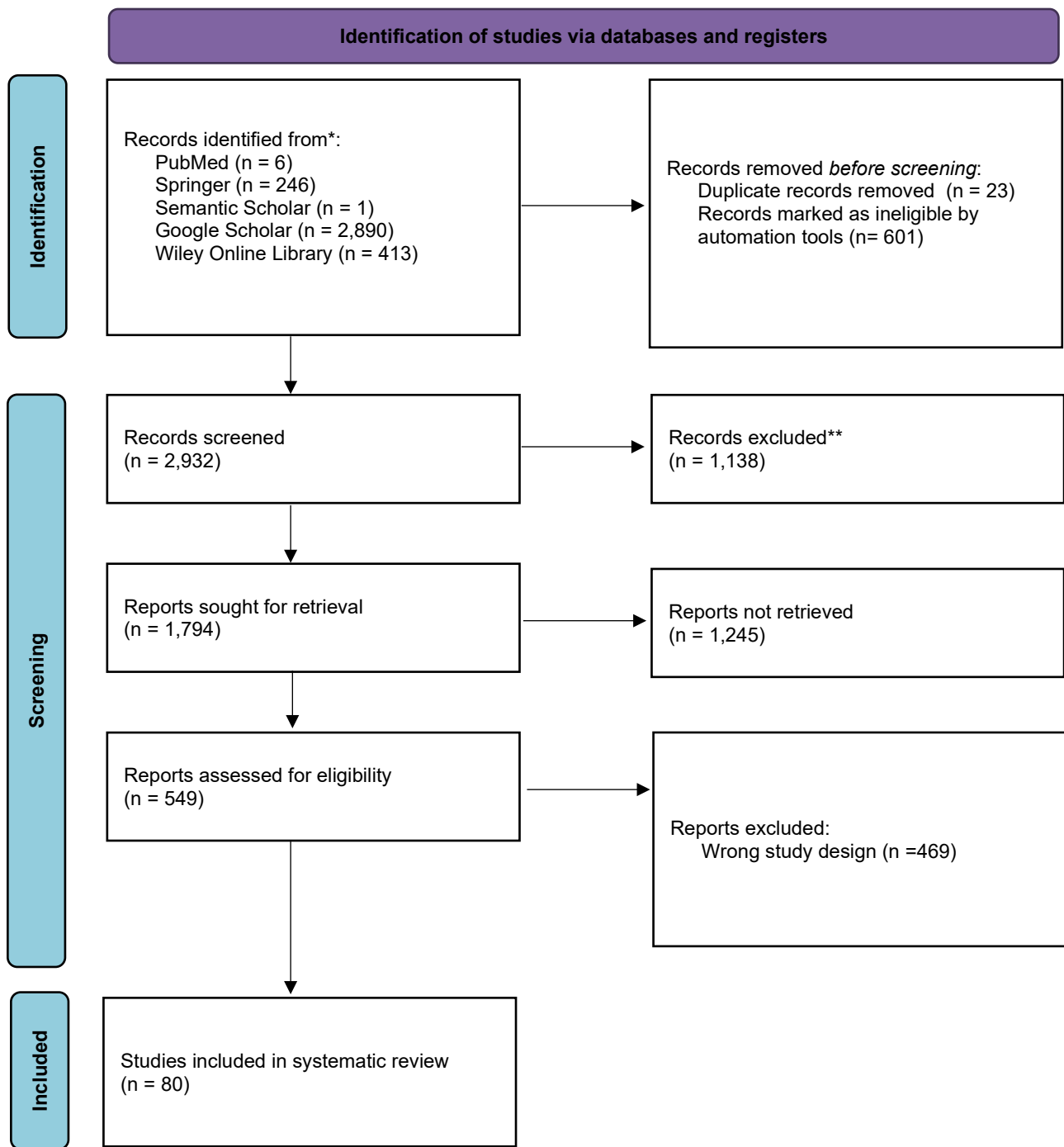


Figure 1. Article search flowchart

RESULTS

Characteristics of included studies

The included studies comprised a diverse array of research designs and populations examining diagnostic criteria for polycystic ovary syndrome (PCOS). The majority were case-control or cross-sectional studies, etc across multiple populations.

Study	Setting	Sample Size	Age Range	Population Type	Diagnostic Criteria Used
J. Pea et al., 2023 [1]	Database search	24 studies (3883 controls, 3859 PCOS)	Adult women	Not specified	Rotterdam [1]
J. Chan et al., 2019 [10]	Community-based	345 (206 Poss-AE, 139 Non-AE)	14-45 years	General population screening	Updated 2018 International Consortium guidelines [10]
L. Casadei et al., 2018 [2]	Clinical	140	21-35 years	Infertile women	NIH, Rotterdam, AE-PCOS [2]
M. Lujan et al., 2013 [11]	Hospital	168 (98 PCOS, 70 controls)	18-35 years	General population	NIH [11]

Study	Setting	Sample Size	Age Range	Population Type	Diagnostic Criteria Used
Mariane de Oliveira Gomes et al., 2025 [12]	Various	45 studies	Not mentioned	Symptomatic women	Rotterdam, NIH, AE-PCOS [12]
J. Pea et al., 2022 [5]	Database search	22 studies (6,337 participants)	Reproductive age	Not specified	Rotterdam [5]
Yumiko Tsukui et al., 2022 [13]	Literature review	15 trials	Up to 29 years	Symptomatic women	Rotterdam, NIH, AE-PCOS [13]
E. Khashchenko et al., 2020 [14]	Clinical/hospital	160 (130 PCOS, 30 controls)	15-17 years	Symptomatic adolescents	Rotterdam [14]
Abhimanyu Pandit et al., 2024 [15]	Tertiary care hospital	273 (100 PCOS, 173 controls)	Reproductive age	Symptomatic women	Rotterdam [15]
E. Hariton et al., 2020 [16]	Tertiary academic center	609 (364 PCOS, 245 controls)	25-40 years	Symptomatic women and controls	Rotterdam [16]

Study	Setting	Sample Size	Age Range	Population Type	Diagnostic Criteria Used
Nada Ahmed et al., 2019 [17]	Hospital	148 (79 PCOS, 69 controls)	18-38 years	Symptomatic women	Rotterdam [17]
Khansa Chaabouni et al., 2023 [18]	Clinical	130 (65 PCOS, 65 controls)	Reproductive age	Symptomatic women	Rotterdam [18]
I. Oueslati et al., 2022 [19]	Single center	50	Mean age 34.2 years	Women with morbid obesity	Rotterdam [19]
T. Piltonen et al., 2024 [20]	Population-based	1,803	Born 1985-1987	General population	Rotterdam [20]
Didier Dewailly et al., 2014 [21]	Clinical	893	Not mentioned	Infertile women	Rotterdam [21]
Alexandra Dietz de Loos et al., 2021 [22]	Multicenter	955 (455 cases, 500 controls)	25-45 years	Not specified	Rotterdam [22]

Study	Setting	Sample Size	Age Range	Population Type	Diagnostic Criteria Used
Mostafa Barghi et al., 2024 [23]	Various	202 studies	Not mentioned	Symptomatic women	NIH, Rotterdam, AE-PCOS [23]
Christina I. Cortes et al., 2021 [6]	Clinical	330 (227 PCOS, 103 controls)	<26, 26-34, 35-40 years	Not specified	Rotterdam [6]
A. D. D. Loos et al., 2019 [24]	Not mentioned	1,820 (discovery: 865; validation: 955)	25-45 years	Not specified	Rotterdam [24]
T. Ach et al., 2024 [25]	Clinical	92 PCOS	Adult females >18 years	Symptomatic women	Rotterdam [25]
Shahrzad Zadehmodares et al., 2015 [26]	Clinical/hospital	117 (60 PCOS, 57 controls)	20-40 years	Symptomatic women	Rotterdam [26]
R. Bell et al., 2021 [27]	Community-based	163	Mean age 32.5 years	General population	Rotterdam [27]

Study	Setting	Sample Size	Age Range	Population Type	Diagnostic Criteria Used
F. Ramezani Tehrani et al., 2021 [4]	Clinic and population-based	803 (303 PCOS, 500 controls)	20-40 years	Symptomatic women and controls	Rotterdam [4]
P. R. Kumar et al., 2024 [28]	Tertiary care center	281 (187 PCOS, 94 controls)	18-40 years	Symptomatic women	Rotterdam, NIH [28]
Prekshi Bansal et al., 2020 [29]	Clinical	120	≥25 years	Symptomatic women (acne)	Rotterdam [29]
Raghad Fadhil Taha et al., 2023 [30]	Hospital	200 (150 PCOS, 50 controls)	Not mentioned	Symptomatic women	Rotterdam [30]
Naveetha Lakshmi Narayanaswamy et al., 2023 [31]	Clinical	60 (30 PCOS, 30 controls)	18-40 years	Symptomatic women	Rotterdam [31]

Study	Setting	Sample Size	Age Range	Population Type	Diagnostic Criteria Used
N. Mahajan et al., 2019 [32]	Fertility center	367	Reproductive age	Infertile women	Rotterdam [32]
Lingling Jiang et al., 2023 [33]	Hospital	550 (450 PCOS, 100 controls)	20-40 years	Chinese women	Rotterdam [33]
H. W. Raymond Li et al., 2020 [34]	Clinic	751 (473 PCOS, 278 controls)	21-40 years	Symptomatic women	Rotterdam [34]
Z. Wang et al., 2022 [35]	Clinical/hospital	215 (132 PCOS, 83 controls)	Mean age 28-31 years	Infertile women with obesity	Rotterdam [35]
L. Lotz et al., 2025 [36]	Not mentioned	221 (136 PCOS, 85 controls)	18-24 years	Young adult women	Rotterdam [36]
Yang You et al., 2026 [37]	Hospital	454 (264 PCOS, 190 controls)	Reproductive age	Chinese women	Rotterdam [37]

Study	Setting	Sample Size	Age Range	Population Type	Diagnostic Criteria Used
Dr.Rajrani Sharma et al., 2017 [38]	Hospital	64 (32 PCOS, 32 controls)	21-35 years	Hospital-based	Rotterdam [38]
S. Ilidromiti et al., 2013 [39]	Literature review	10 studies	Not mentioned	Symptomatic women	Rotterdam [39]
S. Jonard et al., 2005 [40]	Clinical/hospital	211 (154 PCOS, 57 controls)	Not mentioned	Symptomatic women	NIH [40]
F. Vural et al., 2022 [8]	University clinic	401 (154 PCOS, 247 controls)	Mean age 25 years	Symptomatic women	Rotterdam, NIH, AES [8]
N. Köşüş et al., 2011 [41]	Clinical/hospital	316 (251 PCOS, 65 controls)	Not mentioned	Turkish women	AES [41]
N. Dhayat et al., 2018 [42]	University hospitals	107 (41 PCOS, 66 controls)	13-46 years	Symptomatic women	AE-PCOS [42]

Study	Setting	Sample Size	Age Range	Population Type	Diagnostic Criteria Used
S. Alhassan et al., 2023 [43]	Clinical/hospital	600 (300 PCOS, 300 controls)	Not mentioned	Sudanese women	Rotterdam [43]
D. Dewailly et al., 2012 [44]	Clinical/hospital	262	Not mentioned	Women with prior births	Rotterdam, AE-PCOS [44]
S. Şahmay et al., 2014 [3]	University hospital	606	Not mentioned	Symptomatic women	Rotterdam, NIH, AES [3]
P. Pigny et al., 2006 [45]	Clinical/hospital	169 (73 PCOS, 96 controls)	Not mentioned	Not specified	Rotterdam [45]
G. Bozdag et al., 2019 [7]	Population-based	392	Not mentioned	General population	NIH, Rotterdam, AE-PCOS [7]
M. Feichtinger et al., 2010 [46]	Literature review	10 studies	Not mentioned	Symptomatic women	Rotterdam [46]

Study	Setting	Sample Size	Age Range	Population Type	Diagnostic Criteria Used
T. Eilertsen et al., 2012 [47]	Clinical/hospital	262	Not mentioned	Women with prior births	Rotterdam, AE-PCOS [47]
Asima K. Ahmad et al., 2019 [48]	Tertiary academic center	1,001 (245 PCOS, 756 controls)	25-40 years	Symptomatic women	NIH [48]
Kulsoom Bahadur et al., 2023 [49]	Hospital	140	16-45 years	Symptomatic women	Rotterdam [49]
Yang Zhao et al., 2019 [9]	Various	29 studies	Not mentioned	Not specified	Rotterdam [9]
S. Alsamarai et al., 2009 [50]	Tertiary care hospital	Longitudinal: 26; Cross-sectional: 850	18-64 years	Symptomatic women and controls	NIH [50]
B. Joshi et al., 2014 [51]	Community-based	778 (600 completed)	15-24 years	General population	Rotterdam, AES [51]

Study	Setting	Sample Size	Age Range	Population Type	Diagnostic Criteria Used
L. Casadei et al., 2013 [52]	Clinical	59	Not mentioned	Infertile women	NIH [52]
V. Conoscenti et al., 2018 [53]	Hospital	26,711 identified; 150 reviewed	18-40 years	General population	Rotterdam [53]
Cong-shun Ma et al., 2011 [54]	Hospital	195	Not mentioned	Women with PCO	Rotterdam, Japan consensus [54]
R. Nadaraja et al., 2018 [55]	Hospital/clinical	312 (164 PCOS, 148 controls)	Not mentioned	Not specified	Rotterdam [55]
Moaz O Moursi et al., 2023 [56]	Infertility clinic	200 (100 PCOS, 100 controls)	18-44 years	Infertile women	Rotterdam [56]
Henri Sova et al., 2019 [57]	Infertility clinic and community	428 (319 PCOS, 109 controls)	18-39 years	Infertile women and controls	Rotterdam [57]

Study	Setting	Sample Size	Age Range	Population Type	Diagnostic Criteria Used
A. Königer et al., 2014 [58]	University hospital	128 (80 PCOS, 48 controls)	Not mentioned	German women	Rotterdam [58]
Ya-xiao Chen et al., 2008 [59]	Clinical/hospital	95 (69 PCOS, 26 controls)	Adolescents	Chinese adolescents	Rotterdam [59]
Türkan Gürsu et al., 2022 [60]	IVF Center	244	Not mentioned	Infertile women	Rotterdam [60]
Wanwan Yi et al., 2023 [61]	Hospital	213 (139 PCOS, 74 controls)	Not mentioned	Chinese women	Rotterdam [61]
Sikandar hayat Khan et al., 2024 [62]	Hospital	328	Not mentioned	Symptomatic women	Rotterdam [62]
T. Sathyapalan et al., 2017 [63]	Clinic/hospital	175 (110 PCOS, 65 controls)	Mean age 25.5 years	Caucasian women	Rotterdam [63]

Study	Setting	Sample Size	Age Range	Population Type	Diagnostic Criteria Used
M. Insenser et al., 2023 [64]	Clinical/hospital	140 (71 PCOS, 69 controls)	Premenopausal	Symptomatic women	Rotterdam [64]
H. T. Timur et al., 2023 [65]	University hospital	187 (93 PCOS, 94 controls)	21-35 years	Symptomatic women	Rotterdam [65]
Taif Raheem Hmoud et al., 2025 [66]	Hospital	120 (60 PCOS, 60 controls)	18-45 years	Symptomatic women	Rotterdam [66]
Linling Xie et al., 2021 [67]	Hospital	60 (30 PCOS, 30 controls)	Not mentioned	Chinese women	Rotterdam [67]
Bronwyn S. Bedrick et al., 2019 [68]	Not mentioned	101 (51 PCOS, 50 controls)	18-50 years	General population	Modified Rotterdam [68]
Jenna Salem et al., 2025 [69]	Outpatient clinic	3,951 reviewed	18-45 years	Symptomatic women	Rotterdam [69]

Study	Setting	Sample Size	Age Range	Population Type	Diagnostic Criteria Used
Justin Armstrong et al., 2022 [70]	University fertility institute	330 (227 PCOS, 103 controls)	18-40 years	Infertile women	Rotterdam [70]
G. Palomaki et al., 2020 [71]	Two medical centers	Not mentioned	18-39 years	Symptomatic women	NIH [71]
A. M. A. Khaleel et al., 2025 [72]	Hospital	100 (50 PCOS, 50 controls)	Not mentioned	Not specified	Rotterdam [72]
Zhen Wang et al., 2025 [73]	Hospital	426 (213 PCOS, 213 controls)	20-39 years	Chinese women	Rotterdam [73]
Maal Anwer Al-Hamzawi et al., 2025 [74]	Hospital	185 (95 PCOS, 90 controls)	≤40 years	Iraqi women	Rotterdam [74]
T. Yén et al., 2025 [75]	Hospital	600 (178 PCOS, 422 controls)	18-45 years	Infertile women	Rotterdam [75]

Study	Setting	Sample Size	Age Range	Population Type	Diagnostic Criteria Used
S. Abbas et al., 2024 [76]	Hospital	60 (30 PCOS, 30 controls)	20-45 years	Symptomatic women	Rotterdam [76]
C. N. Ndoua et al., 2017 [77]	Hospital	46 (23 PCOS, 23 controls)	<38 years	Infertile women	Rotterdam [77]
A. Bizuneh et al., 2024 [78]	Various	23 studies (2,857 participants)	Not mentioned	Not specified	Not specified [78]
Francisco J. Barrera et al., 2023 [79]	Multicenter	31 studies	Median age 29 years	Various	NIH, Rotterdam [79]
S. L. Fong et al., 2017 [80]	Clinical/population	997 (297 controls, 700 PCOS)	Not mentioned	General population	Rotterdam [80]

Most studies focused on women of reproductive age, though specific age ranges varied from adolescents (15-17 years) [14] to women up to 45 years [10, 22]. Study populations were predominantly symptomatic women presenting to clinical settings, though several included general population screening [10, 27] or community-based cohorts [20, 51].

Geographic representation was diverse, with studies conducted in North America, Europe, Asia, Africa, and the Middle East. Several studies specifically examined diagnostic criteria in ethnically diverse populations, including Chinese [33, 37, 54], Turkish [41], Sudanese [43], Iraqi

[74], Indian [32], and Tunisian [18] women. Body mass index (BMI) varied across studies, with some specifically focusing on women with obesity [19, 35].

Diagnostic accuracy of different PCOS criteria

Rotterdam criteria

The Rotterdam criteria (2003), which require the presence of at least two of three features—oligo-ovulation, hyperandrogenism, and polycystic ovarian morphology (PCOM)—were the most widely evaluated diagnostic standard across the included studies.

Multiple studies examined the diagnostic performance of the Rotterdam criteria as applied in various populations. In a large meta-analysis of adult women, the Rotterdam criteria demonstrated strong diagnostic utility when using ultrasound markers, with follicle number per ovary (FNPO) showing the highest accuracy (sensitivity 84%, specificity 91%, AUC 0.905) [1]. When comparing ultrasonographic markers within the Rotterdam framework, FNPO consistently outperformed ovarian volume (OV) and follicle number per single cross-section (FNPS) [1, 5].

The prevalence of PCOS diagnosed by Rotterdam criteria varied substantially across studies and populations. In an unselected population, prevalence ranged from 19.9% using an antral follicle count (AFC) threshold of 12 to 10.2% when the threshold was raised to 20 [7]. A community-based study in Mumbai identified a PCOS prevalence of 22.5% among adolescent and young women using Rotterdam criteria [51]. Regional differences in diagnostic accuracy were noted, with studies from North America showing lower diagnostic accuracy compared to Asian and European studies [1].

Age-stratified analysis revealed important considerations for the Rotterdam criteria. In young adult women aged 18-24 years, AMH at a cutoff of 3.2 ng/mL showed sensitivity of 85.0% and specificity of 68.2% for detecting PCOM [36]. For women aged 25-45 years, the same AMH cutoff demonstrated sensitivity of 82.4% and specificity of 89.8% [24]. These findings suggest that diagnostic thresholds may need adjustment based on age to maintain optimal accuracy [36, 48].

NIH criteria

The National Institutes of Health (NIH) criteria (1990), which require the presence of both chronic oligo-ovulation and clinical or biochemical hyperandrogenism, represent the most stringent diagnostic approach for PCOS.

In comparative studies, the NIH criteria generally identified fewer women with PCOS than the Rotterdam criteria. One study found PCOS prevalence of 27.1% using NIH criteria compared to 40% with Rotterdam criteria in the same population [2]. Another investigation reported a more restrictive case identification, with NIH criteria having an area under the ROC curve of 0.80 compared to 0.74 for Rotterdam criteria when evaluating AMH as a diagnostic marker [7].

When AMH was evaluated as a diagnostic marker under NIH criteria, the optimal threshold was 5.20 ng/mL with sensitivity of 79% and specificity of 80% (AUC 0.86) [2]. This was slightly higher than the optimal thresholds identified for Rotterdam or AE-PCOS criteria in the same study [2].

The NIH criteria's focus on hyperandrogenism and oligo-ovulation may provide advantages in identifying metabolically high-risk phenotypes. In one study, PCOS women with both biochemical and clinical hyperandrogenism showed a more pronounced metabolic phenotype and insulin resistance compared to those with clinical hyperandrogenism alone [29]. This suggests the NIH criteria may be particularly useful for identifying women at higher cardiometabolic risk [63].

Androgen Excess Society (AES) criteria

The Androgen Excess Society (AES) criteria (2006) require the presence of hyperandrogenism (clinical or biochemical) along with either ovulatory dysfunction or polycystic ovaries, emphasizing androgen excess as a cardinal feature of PCOS.

Studies comparing AES criteria to other diagnostic standards showed intermediate prevalence rates. One investigation found AES criteria identified 29.3% of women as having PCOS, compared to 27.1% for NIH and 40% for Rotterdam criteria [2]. When evaluating AMH as a diagnostic marker, the AES criteria yielded an optimal threshold of 4.85 ng/mL with sensitivity of 80% and specificity of 78% (AUC 0.85) [2].

In a cross-sectional study examining AMH diagnostic accuracy across different criteria, the AES criteria showed sensitivity of 84.4% and specificity of 72% when using AMH alone (AUC 0.857) [8]. This was comparable to but slightly lower than the Rotterdam criteria in the same population [8].

A population-based study found that AMH had poor to fair validity for diagnosing PCOS according to AE-PCOS criteria in an unselected group of women, with an age and BMI adjusted AUC of 0.71 [7]. However, diagnostic accuracy was substantially higher (AUC 0.85) in women with all features of the syndrome (Phenotype A) [7].

When AMH was combined with clinical features under AES criteria, sensitivity and specificity improved to 82% and 93.5%, respectively [3]. This suggests that the AES criteria may perform optimally when supplemented with additional biomarkers [3].

Direct comparison of criteria performance

Several studies directly compared the diagnostic accuracy of Rotterdam, NIH, and AES criteria using identical populations and testing conditions. A comprehensive comparative study found that when AMH was used as a diagnostic marker, the Rotterdam criteria demonstrated the highest sensitivity (83%) with perfect specificity (100%), compared to NIH criteria (sensitivity 83%, specificity 89%) and AES criteria (sensitivity 82%, specificity 93.5%) [3].

In another comparative analysis, all three criteria showed good diagnostic performance when supplemented with AMH levels. The combination of oligo/amenorrhea and/or hyperandrogenism with AMH yielded sensitivity and specificity of 83% and 100% for Rotterdam criteria, 83% and 89% for NIH criteria, and 82% and 93.5% for AES criteria [3].

A study examining AMH alone for diagnostic purposes found that the Rotterdam criteria had an AUC of 0.866, the AES criteria had an AUC of 0.857, and the NIH criteria had an AUC of 0.825 [8]. When AMH was used to replace PCOM in the Rotterdam criteria, diagnostic performance improved substantially to an AUC of 0.934 with sensitivity of 97.4% and specificity of 90.67% [8].

Population-specific considerations emerged as important factors in comparative assessments. In Turkish women, different ultrasound thresholds were optimal (follicle count ≥ 8 and

ovarian volume $\geq 6.43 \text{ cm}^3$) compared to Rotterdam consensus values [41]. For Chinese adolescents, optimal thresholds were even lower (mean ovarian volume 6.74 cm^3 and maximum ovarian volume 7.82 cm^3) [59]. These findings suggest that standardized criteria may require population-specific calibration [41, 59].

Role of anti-Müllerian hormone in PCOS diagnosis

Anti-Müllerian hormone (AMH) emerged as a promising additional or alternative biomarker across multiple studies examining PCOS diagnostic criteria. AMH levels were consistently elevated in women with PCOS compared to controls, with mean differences ranging from approximately 2-fold to 4-fold higher [45, 75].

AMH diagnostic thresholds

Optimal AMH cutoff values varied considerably across studies, influenced by assay methods, age groups, and population characteristics. A large meta-analysis of extracted data identified 4.7 ng/mL as an optimal cutoff, yielding sensitivity of 82.8% and specificity of 79.4% (AUC 0.87) [39]. However, individual studies reported a wide range of thresholds from 3.15 ng/mL [26] to 7.2 ng/mL [14].

Age-specific AMH thresholds showed consistent patterns of decline with advancing age. For women aged 20-27 years, the optimal cutoff was 5.7 ng/mL (95% CI: 5.48-6.19) [4]. This decreased to 4.55 ng/mL (95% CI: 4.52-4.64) for ages 27-35 years, and further to 3.72 ng/mL (95% CI: 3.55-3.80) for ages 35-40 years [4]. Similar age-stratified patterns were observed in other populations [65, 73].

Using standardized assays, particularly the Elecsys AMH Plus immunoassay, a cutoff of 3.2 ng/mL (23 pmol/L) demonstrated robust diagnostic performance with sensitivity of 88.6% and specificity of 84.6% (AUC 93.6%) in women aged 25-45 years [22]. This threshold remained consistent across PCOS phenotypes A-D [22]. Validation in young adult women aged 18-24 years confirmed the utility of this cutoff with sensitivity of 85.0% and specificity of 68.2% [36].

Population-specific variations in optimal AMH thresholds were notable. Indian women showed an optimal cutoff of 5.03 ng/mL (sensitivity 70.68%, specificity 79.91%) [32], while Chinese women demonstrated optimal performance at 4.64 ng/mL (sensitivity 81.6%, specificity

92.0%) [33]. Sudanese women had an even higher optimal threshold of 3.95 ng/mL with excellent diagnostic accuracy (AUC 0.999) [43]. For women with morbid obesity ($\text{BMI} \geq 40 \text{ kg/m}^2$), the optimal cutoff in those aged 35-45 years was substantially lower at 0.81 ng/mL (sensitivity 90%, specificity 71%, AUC 0.857) [19].

AMH performance across PCOS phenotypes

AMH diagnostic accuracy varied significantly across different PCOS phenotypes as defined by Rotterdam criteria. Women with the full phenotype (phenotype A: hyperandrogenism, oligo-anovulation, and PCOM) showed the highest AMH levels and best diagnostic performance [6, 60]. In one study, predicted AMH values were 9.88 ng/mL for phenotype A, 6.48 ng/mL for phenotype B (hyperandrogenism + PCOM), 6.32 ng/mL for phenotype C (hyperandrogenism + oligo-anovulation), and 5.56 ng/mL for phenotype D (oligo-anovulation + PCOM) [6].

AMH showed satisfactory diagnostic potential for phenotype A but not for other phenotypes when used alone [8]. However, when AMH was used to replace PCOM in the Rotterdam criteria, it achieved high sensitivity and specificity across all phenotypes, with phenotypes A and D diagnosed at 100% sensitivity and 94.5% specificity, and phenotype C at 96.15% sensitivity and 94.5% specificity [8].

A meta-analysis specific to adolescents and young women found AMH had overall specificity of 81% and sensitivity of 66.3% for PCOS diagnosis [13]. Cutoff values in this age group ranged from 6.1 to 7.25 ng/mL across studies, with an estimated optimal range of 6-7 ng/mL [13].

AMH as replacement for ultrasound criteria

A critical question addressed by multiple studies was whether AMH could replace ultrasound assessment of PCOM as a diagnostic criterion. When AMH replaced PCOM in the Rotterdam criteria, sensitivity increased from 76% to 93% and specificity increased from 86% to 99% (AUC improved from 0.88 to 0.97) [9]. This represented a statistically significant improvement in diagnostic efficacy [9].

Another investigation found that replacing PCOM with AMH in Rotterdam criteria yielded sensitivity of 97.4% and specificity of 90.67% (AUC 0.934) [8]. For individual phenotypes, this approach achieved sensitivity of 100% for phenotypes A and D, and 96.15% for phenotype C [8].

The substitution of AMH for PCOM affected the number of women classified as having PCOS. In a community-based sample, using AMH cutoffs resulted in reclassification of 5% of participants, with a net increase of six women identified as having PCOS [27]. This suggests that while overall population impact may be modest, individual-level diagnosis can change substantially [27].

Studies using cluster analysis to identify women with PCOM found that AMH had high correlation with follicle counts ($r=0.83-0.84$) and could effectively identify PCOM women within control populations [21]. After excluding PCOM women from controls using cluster analysis, an AMH threshold of 28 pmol/L achieved specificity of 97.5% and sensitivity of 84.2% (AUC 0.948) [21].

AMH combined with other markers

The combination of AMH with other diagnostic markers consistently improved diagnostic accuracy compared to AMH alone. When AMH was combined with free testosterone, the AUC increased from 0.938 to 0.948, with sensitivity rising from 81.6% to 86.1% and specificity decreasing slightly from 92.0% to 90.3% [33].

Combining AMH with clinical symptoms also enhanced diagnostic utility. The combination of oligo/amenorrhea and/or hyperandrogenism with AMH showed sensitivity of 83% and specificity of 100% for Rotterdam criteria, 83% and 89% for NIH criteria, and 82% and 93.5% for AES criteria [3]. This approach "markedly increased the clinical scope for PCOS diagnosis" [3].

Age-stratified AMH combined with waist-to-hip ratio (WHR) yielded an AUC of 0.834, while addition of LH/FSH ratio increased this to 0.869 [64]. Using both WHR and LH/FSH ratio with AMH achieved the highest AUC of 0.895, with positive predictive values up to 88% and negative predictive values up to 85% [64].

Multivariate logistic regression confirmed AMH as an independent predictor of PCOS alongside LH/FSH ratio, testosterone, and BMI [75]. A model using age, log ovarian volume,

follicle number, and testosterone distinguished PCOS subjects from controls with an ROC curve area of 0.90 [50].

Ultrasonographic criteria and modifications

Follicle number thresholds

The Rotterdam criteria originally specified 12 or more follicles measuring 2-9 mm in diameter as indicative of PCOM [11]. However, advances in ultrasound technology and more rigorous counting methodologies have prompted reassessment of these thresholds.

A diagnostic test accuracy study using newer ultrasound technology and a grid system approach concluded that a substantially higher threshold of 26 follicles for FNPO was required to distinguish PCOS women from healthy women (sensitivity 85%, specificity 94%, AUC 0.969) [11]. This represented more than double the original Rotterdam threshold and reflected improved follicle detection with modern equipment [11].

A systematic review and meta-analysis confirmed that FNPO was the most accurate diagnostic marker in adult women, with pooled sensitivity of 84% (CI: 81-87%) and specificity of 91% (CI: 86-94%) (AUC 0.905) [1]. When FNPO was compared to alternative ultrasonographic markers, it consistently outperformed both ovarian volume and follicle number per single cross-section [1, 5].

Studies using receiver operating characteristic curve analysis found that FNPO thresholds of 12-13 provided optimal balance between sensitivity and specificity across multiple populations [21, 48]. However, age-stratified analysis revealed declining optimal thresholds with advancing age: >15 for ages 25-<30 years, >14 for ages 30-<35 years, and >12 for ages 35-<40 years [48].

Population-specific variations in optimal follicle count thresholds were substantial. In Turkish women, a threshold of 8 follicles offered the best compromise between sensitivity and specificity (AUC 0.998) [41]. For Chinese adolescents, optimal thresholds were similarly adjusted [59]. North African women showed optimal performance at 18 follicles with sensitivity of 68.1% and specificity of 100% (AUC 0.955) [25].

Ovarian volume criteria

Ovarian volume (OV) served as an alternative or complementary ultrasonographic marker for PCOM in multiple diagnostic frameworks. The Rotterdam consensus specified an OV threshold of $>10 \text{ cm}^3$ [11], but subsequent studies questioned the optimality of this cutoff.

A prospective study examining OV as a diagnostic criterion found that lowering the threshold to 7 cm^3 offered the best compromise between specificity (91.2%) and sensitivity (67.5%) [40]. In comparison, the 10 cm^3 threshold yielded 98.2% specificity but only 45% sensitivity [40]. The study concluded that while follicle number remained the best diagnostic criterion (AUC 0.956), OV at 7 cm^3 provided good diagnostic accuracy [40].

Meta-analysis of multiple studies found that OV had pooled sensitivity of 81% (CI: 76-86%) and specificity of 81% (CI: 75-86%), with an AUC of 0.856 [1]. This was slightly inferior to FNPO but represented a robust alternative when total follicle counts could not be accurately obtained [1].

Population-specific OV thresholds varied considerably. Turkish women showed optimal performance at 6.43 cm^3 [41], while Chinese adolescents demonstrated best results at 6.74 cm^3 for mean OV and 7.82 cm^3 for maximum OV [59]. North African women had an optimal cutoff of 9.25 mL with sensitivity of 48.9% and specificity of 100% [25]. Iraqi women showed optimal performance at $>9.4 \text{ mL}$ (sensitivity 36.4%, specificity 86.7%) [74].

Age-dependent changes in ovarian volume were documented in longitudinal studies. A 7-15 year follow-up study found that ovarian volume decreased significantly over time in both PCOS and control women (PCOS: 15.2 ± 7.4 to $11.6 \pm 4.4 \text{ mL}$; controls: 7.1 ± 3.7 to $5.4 \pm 2.2 \text{ mL}$) [50]. This necessitated age-based criteria for defining PCOM, particularly in women over 40 years [50].

Follicle number per single cross-section

Follicle number per single cross-section (FNPS) represented a more practical but potentially less accurate alternative to complete ovarian follicle counting. Meta-analysis found FNPS had pooled sensitivity of 81% (CI: 70-89%) and specificity of 83% (CI: 75-88%), with an AUC of 0.870 [1].

A diagnostic threshold of 9 follicles for FNPS achieved sensitivity of 69% and specificity of 90% (AUC 0.880) [11]. While this represented good diagnostic performance, it was inferior to FNPO assessment of the entire ovary [11].

Studies comparing FNPS to FNPO consistently found that complete ovarian assessment provided superior diagnostic confidence and accuracy [1]. However, FNPS remained a valuable option in situations where comprehensive follicle counting was not feasible [1].

Biochemical markers of hyperandrogenism

Testosterone measurements

Total testosterone (TT) served as a primary biochemical marker of hyperandrogenism across diagnostic criteria. A systematic review found TT had pooled sensitivity of 0.74 and specificity of 0.86 (AUC 0.87) for diagnosing biochemical hyperandrogenism in PCOS [78]. However, diagnostic performance varied considerably by measurement method and population characteristics.

Optimal testosterone thresholds for identifying PCOS showed population-specific variation. In hyperandrogenic PCOS phenotypes, cutoffs ranged from 23.1 to 29.3 ng/dL across phenotypes A, B, and C [35]. For Chinese women, optimal testosterone cutoffs contributed to diagnostic models with high accuracy [61].

The method of testosterone measurement significantly impacted diagnostic accuracy. Studies using liquid chromatography-tandem mass spectrometry (LC-MS/MS) demonstrated superior sensitivity and specificity compared to direct immunoassay methods [14, 78]. International guidelines recommended LC-MS/MS for androgen measurement given its high accuracy [78].

Calculated free testosterone (cFT) showed improved diagnostic performance compared to total testosterone. A systematic review found cFT had pooled sensitivity of 0.89 and specificity of 0.83 (AUC 0.85) [78]. In comparative studies, cFT demonstrated higher area under the curve than total testosterone [62].

The free androgen index (FAI) emerged as a widely used and practical marker. ROC analysis showed FAI had strong diagnostic potential with AUC values ranging from 0.78 to 0.85 across studies [14, 55]. An optimal FAI cutoff of >2.75 achieved sensitivity of 75% and specificity

of 93% [14]. Calculated testosterone indices (FAI, cFT, calculated bioavailable testosterone) were more specific with higher positive likelihood ratios than measured androgen markers [55].

Bioavailable testosterone showed the highest AUC among testosterone-based measures in one comparative study (0.792), followed closely by free testosterone (0.791) and FAI (0.782) [62]. This suggested that accounting for sex hormone-binding globulin (SHBG) levels improved diagnostic accuracy for hyperandrogenism [62].

Androstenedione and DHEA-S

Androstenedione (A4) demonstrated promise as a diagnostic marker, particularly in specific populations. In women with infertility and obesity, A4 showed significant improvement in discriminatory value compared to age alone (AUC difference 0.14) [35]. The Youden index was highest for A4 (0.65) with sensitivity of 81.82% and specificity of 83.16% [61].

A systematic review found A4 had pooled sensitivity of 0.75 and specificity of 0.71 (AUC 0.80) [78]. While this was lower than testosterone-based measures, A4 could be considered if TT or cFT were not elevated [78]. Optimal A4 thresholds for PCOS phenotypes ranged from 11.45 to 23.2 ng/mL across studies [1, 14].

Dehydroepiandrosterone sulfate (DHEA-S) showed the lowest diagnostic accuracy among commonly measured androgens, with pooled sensitivity of 0.75 and specificity of 0.67 (AUC 0.77) [78]. DHEA-S levels were significantly elevated in PCOS women but with considerable overlap between groups [14, 29]. Guidelines recommended considering DHEA-S as a secondary marker if primary androgen measurements were not elevated [78].

The ratio of testosterone to dihydrotestosterone (T/DHT) and 11-oxygenated androgens did not emerge as promising biomarkers in women with obesity [35]. These measures showed poor diagnostic utility and were not recommended for routine clinical use in PCOS diagnosis [35].

Novel androgen markers

Dihydrotestosterone (DHT) showed excellent diagnostic accuracy in a case-control study, with serum levels significantly higher in PCOS patients (4256 ± 233 ng/L) compared to controls (3776 ± 186 ng/L) [66]. An optimal cutoff of approximately 408.3 ng/L yielded sensitivity of 90%

and specificity of 100% (AUC 0.93) [66]. However, this marker has not been widely evaluated and requires validation in larger populations [66].

Salivary testosterone (salT) emerged as a non-invasive alternative to serum measurements. In a comparative study, salT levels were significantly elevated in PCOS patients and showed at least equivalent predictive value to the free androgen index [63]. The combination of salT or FAI identified 100% of PCOS patients [63]. This suggested potential utility for salT measurement by LC-MS/MS as a complement to existing laboratory tests [63].

Urinary steroid profiling using gas chromatography-mass spectrometry identified specific metabolite patterns in PCOS. The best diagnostic combination included androstenediol, estriol, cortisol, and 20 β DH-cortisone, achieving sensitivity of 90.2% and specificity of 90.8% (AUC 0.961, PPV 86%, NPV 93.7%) [42]. This approach offered potential for diagnosis without reliance on ultrasound or multiple blood tests, though it requires specialized laboratory capabilities [42].

Additional clinical markers and combined approaches

LH/FSH ratio

The luteinizing hormone to follicle-stimulating hormone (LH/FSH) ratio has been proposed as a supportive biochemical marker for PCOS. Optimal LH/FSH ratio thresholds varied by diagnostic criteria, with values ≥ 1.0 providing significant diagnostic value in multiple populations [54].

For adolescents, an LH/FSH ratio > 1.23 achieved sensitivity of 63.2-78.2% and specificity of 84.4-93.7% [14]. In adult populations, the ratio correlated positively with AMH levels and contributed to multivariate diagnostic models [30, 64].

Studies comparing Rotterdam and Japanese consensus criteria found area under the ROC curve of 0.786 for Rotterdam versus 0.942 for Japanese criteria when using LH/FSH ratio as a diagnostic marker [54]. This suggested that the diagnostic utility of LH/FSH ratio may vary depending on how it is incorporated into broader diagnostic frameworks [54].

However, the LH/FSH ratio showed limitations as a standalone marker. It had lower discriminative power than AMH (AUC 0.932 versus 0.999 in one study) [43] and demonstrated

substantial sensitivity to timing and variability [75]. Current evidence supports its use as a supplementary rather than primary diagnostic criterion [75].

Self-reported symptoms

A telephone questionnaire using self-assessment of body hair and menstrual status demonstrated high predictive value for identifying women at risk for PCOS. The questionnaire achieved sensitivity of 88%, specificity of 77%, positive predictive value of 84%, and negative predictive value of 81% for predicting PCOS according to Rotterdam criteria [10].

For specific symptoms, hirsutism assessment (defined by modified Ferriman-Gallwey score ≥ 3 from upper lip and abdomen) showed sensitivity of 76% and specificity of 70% [68]. Report of any depilatory practices achieved sensitivity of 71% and specificity of 74% [68]. The combined sensitivity of these measures was 93%, though specificity decreased to 52% [68].

These findings suggested that simple screening tools based on self-reported symptoms could be valuable for identifying women who warrant further evaluation [10, 68]. Such approaches may be particularly useful in populations where complete diagnostic evaluation is not feasible or as initial screening in epidemiologic studies [68].

Electronic medical record-based algorithms

Attempts to develop algorithms using electronic medical record (EMR) data and ICD-9/ICD-10 codes showed limited accuracy. The use of PCOS diagnostic codes alone had very low positive predictive value (PPV=17%) [53]. This improved to 52% when requiring a PCOS code plus any one or more PCOS feature codes [53].

Individual PCOS features showed variable PPV, with alopecia having the highest (66%) and androgen excess the lowest (28%) [53]. These findings indicated that reliance on diagnostic codes alone was insufficient for accurate case identification [53]. Recommendations for improvement included requiring higher counts of PCOS features, using natural language processing of clinical notes, and limiting to reproductive or endocrine specialty clinics [53].

Machine learning and artificial intelligence approaches

A systematic review of 31 studies evaluating machine learning and artificial intelligence (ML/AI) for PCOS diagnosis found promising results but significant methodological limitations.

Diagnostic accuracy ranged from 89% to 100% across four studies, with sensitivity ranging from 41% to 100% across ten studies and specificity from 75% to 100% [79].

The most common AI techniques employed were support vector machines (42% of studies), K-nearest neighbor (26%), and regression models (23%) [79]. However, only 32% of studies employed standardized criteria (NIH, Rotterdam, or International PCOS classification) [79].

Circulating microRNAs

Novel biomarkers including circulating microRNAs showed promise for PCOS diagnosis. A validation study found that models based on miR-142-3p and miR-598-3p expression demonstrated good discrimination between women with and without PCOS (AUC 0.801) [64].

Diagnostic accuracy improved when miRNAs were combined with easily available measurements. The combination of miRNAs with waist-to-hip ratio (WHR) yielded AUC of 0.834, with LH/FSH ratio AUC of 0.869, and both WHR and LH/FSH ratio AUC of 0.895 [64]. Positive and negative predictive values reached as high as 88% and 85%, respectively [64].

DNA methylation patterns also showed diagnostic potential. A combination of three CpG methylation signatures (cg23464743, cg06834912, cg00103771) achieved sensitivity of 73.3% and specificity of 93.3% (AUC 0.8378) [67]. While promising, these biomarkers require validation in larger studies before clinical implementation [64].

Synthesis

Geographic and ethnic variations in diagnostic accuracy

Substantial geographic differences in diagnostic accuracy emerged across studies, suggesting that population-specific factors influence optimal diagnostic thresholds. Studies from North America showed lower diagnostic accuracy for follicle number per ovary compared to Asian studies (sensitivity differences) [1]. European studies demonstrated higher specificity and positive likelihood ratios for ovarian volume compared to North American studies [1].

These geographic variations appeared partly attributable to differences in age and BMI distributions. North American PCOS populations had higher mean BMI (30.29 ± 0.61 kg/m²) compared to European populations (27.62 ± 0.73 kg/m²) [5]. Body mass index independently

affected diagnostic markers, with higher BMI associated with different hormonal profiles and ultrasound parameters [1].

Ethnicity-specific diagnostic thresholds were documented for multiple parameters. Turkish women showed optimal follicle count threshold of 8 (compared to Rotterdam's 12) and ovarian volume threshold of 6.43 cm³ (compared to Rotterdam's 10 cm³) [41]. Chinese adolescents demonstrated even lower optimal thresholds for ovarian volume (6.74 cm³ for mean, 7.82 cm³ for maximum) [59]. North African women required 18 follicles for optimal sensitivity and specificity [25].

Age-specific variations in diagnostic markers were particularly pronounced for AMH. Optimal cutoffs declined progressively from 5.7 ng/mL in women aged 20-27 years to 3.72 ng/mL in women aged 35-40 years [4]. Similar age-related patterns were observed for follicle counts and ovarian volume [48, 50]. These findings indicate that applying universal thresholds across all ages and populations may result in misdiagnosis.

The mechanism underlying geographic and ethnic differences likely involves multiple factors beyond simple demographic variation. Genetic differences in ovarian reserve, hormone metabolism, and receptor sensitivity may contribute [80]. Environmental factors including diet, stress, and environmental exposures could also play roles [1]. However, technical factors such as ultrasound equipment quality and operator training should not be overlooked as potential contributors to observed variations [25].

Performance across PCOS phenotypes

Diagnostic accuracy varied substantially across different PCOS phenotypes defined by Rotterdam criteria. Phenotype A (hyperandrogenism + oligo-ovulation + PCOM), representing the most severe presentation, consistently showed the highest diagnostic accuracy for biomarkers. AMH levels were significantly higher in phenotype A (mean 9.88 ng/mL) compared to phenotypes B (6.48 ng/mL), C (6.32 ng/mL), and D (5.56 ng/mL) [6].

Studies examining phenotype-specific diagnostic accuracy found that AMH performed best for phenotype A, with satisfactory diagnostic potential, but showed poorer performance for other phenotypes when used alone [8]. In one investigation, AMH achieved AUC of 0.85 for phenotype

A but only fair validity for milder phenotypes [7]. This pattern suggests that diagnostic markers may be most reliable in women with full phenotypic expression of the syndrome.

The prevalence of different phenotypes varied across populations. In one study, phenotype A comprised 69.9% of PCOS cases, followed by phenotype B (28.6%), phenotype D (23.9%), and phenotype C (16.3%) [5]. Another investigation found phenotype distributions of 63.4% for phenotype D (oligo-ovulation + PCOM), with phenotype A being the second most common [10]. These prevalence differences may reflect population characteristics, ascertainment methods, or differences in diagnostic criteria application.

Phenotype-specific clinical implications were noted in multiple studies. Women with phenotype A (full PCOS) demonstrated higher insulin resistance and more pronounced metabolic derangements compared to other phenotypes [29, 63]. This suggests that different phenotypes may represent varying degrees of severity rather than distinct disease entities, with implications for both diagnosis and management [8].

The challenge of diagnosing milder phenotypes (particularly phenotype D, lacking hyperandrogenism) emerged as a consistent theme. These women may be missed by criteria emphasizing androgen excess (NIH, AES) but could still experience significant reproductive and metabolic consequences [51]. The Rotterdam criteria's inclusion of these milder phenotypes may offer advantages for early identification and intervention, though at the cost of lower diagnostic specificity [7].

Role of diagnostic marker combinations

Single biomarkers, while showing good diagnostic performance individually, consistently achieved higher accuracy when combined with complementary measures. The combination of AMH with clinical features (oligo/amenorrhea and/or hyperandrogenism) increased sensitivity to 83-97% and specificity to 89-100% across different diagnostic criteria [3, 8].

The mechanism by which marker combinations improve diagnostic accuracy appears to involve capturing different aspects of PCOS pathophysiology. AMH primarily reflects ovarian follicle excess, androgens indicate metabolic and ovarian dysfunction, and menstrual irregularity

suggests hypothalamic-pituitary-ovarian axis dysregulation [3]. By combining these complementary pathways, diagnostic models achieve more comprehensive phenotypic assessment.

Multivariate models incorporating age, BMI, and multiple hormonal parameters achieved the highest diagnostic accuracy. One study found that using age-adjusted AMH multiples of the median (MoM) combined with clinical parameters enabled patient-specific PCOS risk computation [71]. Another investigation demonstrated that four hormonal parameters (AMH, FAI, androstenedione, testosterone) combined yielded diagnostic accuracy of 90.2-91.6% [14].

The relative contribution of different markers varied by population and clinical context. In women with obesity, androstenedione showed greater discriminative value than testosterone or DHEA-S [35]. In adolescents, the combination of AMH and leptin concentrations provided superior diagnostic accuracy compared to other hormone combinations [14]. These findings suggest that optimal diagnostic strategies may require population-specific marker selection.

However, practical considerations limit the feasibility of highly complex diagnostic models. Requiring multiple blood draws, specialized assays, and advanced ultrasound examinations may not be feasible in many clinical settings [20]. The search for optimal diagnostic strategies must therefore balance accuracy against accessibility, cost, and clinical utility [76].

Age-related diagnostic considerations

Age emerged as a critical factor influencing the accuracy of all diagnostic criteria for PCOS. Both ovarian volume and follicle number decreased significantly with age in women with PCOS and controls, though the rate of decline was less pronounced in PCOS women [50]. This necessitated age-based criteria to maintain diagnostic accuracy across the reproductive lifespan.

For ultrasound parameters, optimal thresholds showed consistent patterns of decline with age. Follicle number per ovary thresholds decreased from >15 in women aged 25-<30 years to >14 in ages 30-<35 years and >12 in ages 35-<40 years [48]. Similarly, AMH cutoffs declined from 5.7 ng/mL in the youngest group to 3.72 ng/mL in women approaching 40 years [4].

The mechanism underlying age-related changes involves progressive loss of ovarian reserve affecting both PCOS and non-PCOS women. However, women with PCOS maintain relatively higher follicle counts and AMH levels even as absolute values decline [50]. This persistent

elevation creates a "moving target" for diagnosis, requiring age-stratified reference ranges to maintain optimal sensitivity and specificity.

Studies using fixed thresholds across all ages demonstrated reduced diagnostic accuracy, particularly in older reproductive-age women. In women aged 36-45 years, an AMH cutoff of 3.2 ng/mL showed lower sensitivity (77.8%) compared to women aged 25-35 years (88.5%), though specificity improved (90.1% vs 80.3%) [22]. This trade-off between sensitivity and specificity across age groups complicates the selection of universal diagnostic criteria.

The clinical implications of age-related diagnostic variation extend beyond technical considerations. Delayed diagnosis of PCOS in older reproductive-age women may result in missed opportunities for fertility preservation and early metabolic intervention [20]. Conversely, overdiagnosis in younger women using age-inappropriate thresholds could lead to unnecessary medicalization and psychological burden [4].

Proposed modifications and alternatives

Multiple studies proposed modifications to standard diagnostic criteria aimed at improving accuracy or accessibility. The most consistent recommendation involved incorporating AMH as an alternative to ultrasound assessment of PCOM. Studies found that replacing PCOM with age-appropriate AMH thresholds improved diagnostic accuracy (AUC increased from 0.88 to 0.97) while eliminating operator-dependent ultrasound variability [9, 9].

Specific AMH thresholds for replacing PCOM varied by assay and population but centered around 3.2-5.0 ng/mL for most reproductive-age women [22, 32, 33]. Age-stratified thresholds showed better performance than fixed cutoffs, with recommended values declining from approximately 5.7 ng/mL in the early reproductive years to 3.7 ng/mL approaching age 40 [4]. These thresholds required validation using standardized assays such as the Elecsys AMH Plus immunoassay [22].

Modified ultrasound criteria based on population-specific validation studies suggested lower thresholds than original Rotterdam recommendations for many populations. Turkish women showed optimal follicle count of 8 and ovarian volume of 6.43 cm³ [41]. Chinese adolescents demonstrated best performance at mean ovarian volume of 6.74 cm³ [59]. North African women

required 18 follicles for optimal accuracy [25]. These findings supported the development of ethnicity-adjusted reference ranges rather than universal standards.

Alternative biochemical markers beyond traditional testosterone measurements showed promise in specific contexts. Calculated free testosterone outperformed total testosterone in most comparative studies [78]. Salivary testosterone offered non-invasive assessment with equivalent predictive value to serum markers [63]. Urinary steroid profiling identified specific metabolite patterns with high diagnostic accuracy [42]. While these approaches required specialized laboratory capabilities, they offered potential advantages in particular clinical scenarios.

Combined diagnostic algorithms incorporating multiple complementary markers consistently outperformed single-criterion approaches. Models using AMH plus clinical symptoms (oligo/amenorrhea and/or hyperandrogenism) achieved sensitivity of 83-97% and specificity of 89-100% [3]. Addition of metabolic parameters such as waist-to-hip ratio further improved accuracy [64]. However, the optimal combination of markers likely varied by population characteristics and clinical context [14, 35].

Several studies proposed simplified screening approaches for resource-limited settings. Self-reported hirsutism and depilatory practices combined achieved 93% sensitivity for PCOS detection [68]. Telephone questionnaires based on menstrual patterns and hyperandrogenism symptoms showed 88% sensitivity and 77% specificity [10]. While these approaches had lower specificity than comprehensive evaluation, they offered valuable tools for initial screening and epidemiologic research [10, 68].

The international evidence-based guidelines increasingly recognized the role of AMH as a diagnostic criterion, with recent updates supporting its use as an alternative to ultrasound assessment of PCOM [1]. This represented a significant shift in diagnostic approach, reflecting accumulated evidence for AMH's reliability and accessibility compared to operator-dependent ultrasound examination [20].

DISCUSSION

Comparative Diagnostic Performance of PCOS Criteria

The findings of this systematic review reveal important distinctions in the diagnostic performance of the three major PCOS criteria. The Rotterdam criteria, requiring at least two of three features, demonstrated the highest sensitivity but potentially lower specificity compared to the more restrictive NIH and AES criteria [1, 3, 7]. This trade-off between sensitivity and specificity reflects fundamental differences in how each framework conceptualizes the syndrome and has important implications for clinical practice and research.

Rotterdam Criteria

The Rotterdam criteria were the most extensively evaluated diagnostic standard across the included studies, reflecting their widespread clinical adoption since 2003 [1, 5]. The diagnostic performance of Rotterdam criteria varied substantially depending on which component features were assessed and which thresholds were applied. Follicle number per ovary (FNPO) emerged as the most accurate ultrasonographic marker within the Rotterdam framework, with pooled sensitivity of 84% and specificity of 91% (AUC 0.905) in a large meta-analysis [1]. This superior performance compared to ovarian volume (AUC 0.856) and follicle number per single cross-section (AUC 0.870) suggests that complete ovarian follicle counting should remain the preferred ultrasonographic approach when feasible [1, 5].

The original Rotterdam threshold of ≥ 12 follicles for PCOM, established using conventional ultrasound technology, requires reconsideration in light of contemporary evidence. Studies employing high-resolution ultrasound and standardized counting methodologies have proposed substantially higher thresholds, with Lujan et al. recommending 26 follicles as optimal for distinguishing PCOS women from healthy controls (sensitivity 85%, specificity 94%, AUC 0.969) [11]. This more than doubling of the threshold reflects improved follicle visualization with modern equipment and more rigorous counting protocols, suggesting that historical thresholds may no longer be appropriate [11, 21].

Age-stratified analysis revealed that optimal follicle count thresholds decline progressively with advancing age, from >15 for women aged 25- <30 years to >12 for those aged 35- <40 years [48]. This age-dependent decline parallels the physiological depletion of ovarian reserve and necessitates dynamic diagnostic thresholds that account for chronological age [4, 48]. Failure to

incorporate age-appropriate thresholds may result in systematic underdiagnosis of PCOS in younger women and overdiagnosis in older reproductive-age women [4, 48].

The prevalence of PCOS diagnosed by Rotterdam criteria showed marked variation across studies and populations, ranging from 10.2% to 40% depending on the specific thresholds applied and population characteristics [2, 7, 51]. In an unselected population, Bozdag et al. demonstrated that prevalence varied from 19.9% using an antral follicle count threshold of 12 to 10.2% when the threshold was raised to 20 [7]. This threshold sensitivity underscores the importance of standardized, evidence-based cutoffs for ensuring consistent case identification across clinical and research settings.

NIH Criteria

The NIH criteria, requiring both oligo-ovulation and hyperandrogenism, consistently identified fewer women as having PCOS compared to Rotterdam criteria across all included studies [2, 3, 8]. In a direct comparative study, Casadei et al. found PCOS prevalence of 27.1% using NIH criteria versus 40% with Rotterdam criteria in the same population of young infertile women [2]. This more restrictive case identification reflects the NIH framework's emphasis on classic PCOS presentation and may offer advantages for identifying women at highest metabolic risk [29, 63].

The diagnostic performance of NIH criteria when evaluated using AMH as a marker showed an AUC of 0.80 compared to 0.74 for Rotterdam criteria in an unselected population [7]. However, in studies specifically designed to compare all three criteria, NIH criteria demonstrated intermediate performance, with AMH sensitivity of 83% and specificity of 89% (AUC 0.825) [3, 8]. The optimal AMH threshold for NIH criteria (5.20 ng/mL) was slightly higher than for Rotterdam (4.85 ng/mL) or AES criteria (4.85 ng/mL) in one study, potentially reflecting the more severe phenotype captured by NIH criteria [2].

The NIH criteria's requirement for both hyperandrogenism and oligo-ovulation may provide particular value for identifying women with adverse metabolic profiles. Bansal et al. demonstrated that PCOS women with both biochemical and clinical hyperandrogenism exhibited more pronounced metabolic phenotypes and insulin resistance compared to those with clinical hyperandrogenism alone [29]. Similarly, Sathyapalan et al. found that NIH-defined PCOS was

associated with more severe metabolic derangements [63]. These findings suggest that while NIH criteria may miss milder phenotypes (particularly those captured by Rotterdam's inclusion of PCOM without hyperandrogenism), they may be optimally suited for identifying women requiring intensive metabolic surveillance and intervention [29, 63].

Androgen Excess Society Criteria

The AES criteria, requiring hyperandrogenism as an essential feature accompanied by either ovulatory dysfunction or PCOM, represent an intermediate diagnostic approach between the broad Rotterdam framework and the restrictive NIH criteria [2, 8]. Studies comparing AES criteria to other diagnostic standards consistently demonstrated intermediate prevalence rates, with Casadei et al. reporting 29.3% prevalence compared to 27.1% for NIH and 40% for Rotterdam [2].

The diagnostic accuracy of AES criteria when evaluated using AMH showed comparable performance to Rotterdam criteria in some studies but slightly lower accuracy in others. Vural et al. found AES criteria had AMH sensitivity of 84.4% and specificity of 72% (AUC 0.857), compared to Rotterdam's AUC of 0.866 [8]. Sahmay et al. reported that when AMH was combined with clinical features, AES criteria achieved sensitivity of 82% and specificity of 93.5% [3]. This improvement with combined markers suggests that AES criteria may perform optimally when supplemented with additional biomarkers beyond those strictly required by the framework [3].

Population-based evaluation of AES criteria by Bozdag et al. revealed that AMH had poor to fair validity for diagnosing PCOS according to AES criteria in an unselected group of women (age and BMI adjusted AUC 0.71) [7]. However, diagnostic accuracy improved substantially (AUC 0.85) in women with all features of the syndrome (Phenotype A) [7]. This phenotype-dependent performance suggests that AES criteria, like NIH criteria, may be most accurate for identifying women with full phenotypic expression of PCOS [7].

The Emerging Role of Anti-Müllerian Hormone

Anti-Müllerian hormone emerged as the most extensively studied and promising diagnostic biomarker across the included literature, with consistent evidence supporting its utility as an adjunctive or alternative diagnostic marker for PCOS [9, 12, 23, 39]. AMH levels were consistently elevated 2-fold to 4-fold in women with PCOS compared to controls across diverse populations [45,

112

75], reflecting the underlying pathology of arrested follicular development and increased preantral and small antral follicle pool [21].

Optimal AMH Thresholds

The identification of optimal AMH thresholds for PCOS diagnosis proved complex, with substantial variation related to assay methods, age distributions, and population characteristics. A large meta-analysis by Iliodromiti et al. identified 4.7 ng/mL as an optimal cutoff, yielding sensitivity of 82.8% and specificity of 79.4% (AUC 0.87) [39]. However, individual studies reported a wide range of thresholds from 3.15 ng/mL [26] to 7.2 ng/mL [14], underscoring the importance of context-specific threshold selection.

The development of standardized assays, particularly the Elecsys AMH Plus immunoassay, has improved the comparability of AMH measurements across studies. Using this assay, Dietz de Loos et al. identified a cutoff of 3.2 ng/mL (23 pmol/L) demonstrating robust diagnostic performance with sensitivity of 88.6% and specificity of 84.6% (AUC 93.6%) in women aged 25-45 years [22]. This threshold remained consistent across PCOS phenotypes A through D and was subsequently validated in young adult women aged 18-24 years (sensitivity 85.0%, specificity 68.2%) [36]. The consistency of this threshold across validation studies supports its potential utility as a standardized diagnostic reference.

Age-Specific AMH Thresholds

Age emerged as the most critical determinant of optimal AMH thresholds, with consistent evidence demonstrating progressive decline in cutoffs with advancing age [4, 37, 65, 73]. Ramezani Tehrani et al. provided the most comprehensive age-stratified analysis, identifying optimal cutoffs of 5.7 ng/mL for women aged 20-27 years, declining to 4.55 ng/mL for ages 27-35 years, and further to 3.72 ng/mL for ages 35-40 years [4]. These age-specific thresholds reflect the physiological decline in ovarian reserve with age and ensure that diagnostic performance remains consistent across the reproductive lifespan [4, 50].

Similar age-stratified patterns were observed across diverse populations. Timur et al. confirmed age-dependent AMH thresholds in Turkish women [65], while Wang et al. established age-related cutoffs in Chinese populations [73]. The consistency of this age-dependent pattern

across populations suggests that age-stratified diagnostic approaches are essential regardless of geographic or ethnic context [4, 37, 65].

Population-Specific AMH Variations

Beyond age, significant population-specific variations in optimal AMH thresholds were documented, suggesting that ethnicity and geographic origin influence AMH levels independent of PCOS status. Indian women showed an optimal cutoff of 5.03 ng/mL (sensitivity 70.68%, specificity 79.91%) [32], while Chinese women demonstrated optimal performance at 4.64 ng/mL (sensitivity 81.6%, specificity 92.0%) [33]. Sudanese women had an even higher optimal threshold of 3.95 ng/mL with excellent diagnostic accuracy (AUC 0.999) [43]. These variations likely reflect genetic differences in ovarian reserve, follicle dynamics, and hormone metabolism [80].

Special populations require particularly nuanced AMH threshold selection. Women with morbid obesity (BMI ≥ 40 kg/m²) demonstrated substantially lower optimal AMH cutoffs of 0.81 ng/mL (sensitivity 90%, specificity 71%, AUC 0.857) in those aged 35-45 years [19]. This dramatic reduction may reflect obesity-related alterations in granulosa cell function, increased follicular atresia, or hemodilution effects [19]. Similarly, adolescents showed higher optimal thresholds ranging from 6.1 to 7.25 ng/mL, reflecting developmental stage-specific differences in ovarian physiology [13, 14].

AMH as Replacement for Ultrasound Assessment

A critical question addressed by multiple studies was whether AMH could reliably replace ultrasound assessment of PCOM as a diagnostic criterion. The evidence consistently supported this substitution, with studies demonstrating that replacing PCOM with AMH in Rotterdam criteria improved diagnostic accuracy [8, 9, 27]. Zhao et al. found that sensitivity increased from 76% to 93% and specificity from 86% to 99% (AUC improved from 0.88 to 0.97) when AMH replaced PCOM [9]. Vural et al. similarly reported sensitivity of 97.4% and specificity of 90.67% (AUC 0.934) with this substitution [8].

The mechanism underlying AMH's strong performance as a PCOM surrogate relates to its production by granulosa cells of preantral and small antral follicles, which are precisely the follicles whose excess characterizes PCOM [21, 45]. Cluster analysis studies confirmed high correlation

between AMH and follicle counts ($r=0.83-0.84$), supporting AMH as an objective biochemical reflection of the follicle excess that ultrasound visualizes [21, 80].

The substitution of AMH for ultrasound has important practical implications. Ultrasound assessment of PCOM requires specialized equipment, trained operators, and is subject to inter-observer variability [1, 11]. In contrast, AMH measurement using standardized automated assays offers objective, reproducible results that can be obtained from a simple blood draw [22, 32]. This accessibility may be particularly valuable in resource-limited settings where high-quality ultrasound is not readily available [20]. However, the population-specific and age-dependent variations in AMH thresholds necessitate careful validation before widespread implementation [4, 32].

AMH Combined with Other Markers

While AMH alone demonstrated good diagnostic performance, combinations with other markers consistently achieved superior accuracy [3, 33, 64]. Sahmay et al. demonstrated that combining AMH with clinical symptoms (oligo/amenorrhea and/or hyperandrogenism) increased sensitivity to 83-97% and specificity to 89-100% across different diagnostic criteria [3]. This combination approach leverages complementary pathophysiological pathways: AMH reflects ovarian follicle excess, clinical symptoms indicate hypothalamic-pituitary-ovarian axis dysregulation, and androgen measurements capture metabolic and ovarian dysfunction [3].

The combination of AMH with androgen markers proved particularly effective. Jiang et al. found that adding free testosterone to AMH increased AUC from 0.938 to 0.948, with sensitivity rising from 81.6% to 86.1% [33]. Multivariate models incorporating multiple hormonal parameters achieved the highest accuracy, with Insenser et al. reporting AUC of 0.895 when AMH was combined with waist-to-hip ratio and LH/FSH ratio [64]. These findings suggest that while single biomarkers offer convenience, comprehensive diagnostic assessment should incorporate multiple complementary measures when feasible [33, 64].

Ultrasonographic Criteria: Need for Revision

Follicle Number Thresholds

The original Rotterdam threshold of 12 or more follicles measuring 2-9 mm as indicative of PCOM requires substantial revision based on contemporary evidence [1, 11]. Advances in

ultrasound technology, particularly high-frequency transducers and three-dimensional imaging, have dramatically improved follicle visualization and counting accuracy [11, 21]. Studies using modern equipment and standardized counting protocols consistently recommend higher thresholds. Lujan et al.'s prospective diagnostic accuracy study using a grid system approach concluded that 26 follicles optimally distinguished PCOS women from healthy controls (sensitivity 85%, specificity 94%) [11]. This threshold, more than double the original Rotterdam value, reflects the improved sensitivity of contemporary ultrasound for detecting small follicles [11].

Meta-analytic confirmation by Pea et al. established FNPO as the most accurate ultrasonographic marker, with pooled sensitivity of 84% and specificity of 91% [1]. The superiority of FNPO over ovarian volume and FNPS likely reflects the direct relationship between follicle count and the underlying ovarian pathology of PCOS [1, 5]. Complete ovarian assessment, while more time-consuming than simplified approaches, provides superior diagnostic confidence and should remain the preferred approach when feasible [1].

Age-specific follicle count thresholds are essential for maintaining diagnostic accuracy across the reproductive lifespan. Ahmad et al. demonstrated that optimal thresholds decline progressively from >15 for ages 25-<30 years to >12 for ages 35-<40 years [48]. This age-dependent pattern parallels physiological follicle depletion and ensures that diagnostic performance remains consistent as women age [48, 50]. Failure to apply age-appropriate thresholds may result in systematic misclassification, particularly in older reproductive-age women [48].

Population-specific variations in optimal follicle count thresholds were substantial and clinically significant. Turkish women showed optimal performance with a threshold of 8 follicles (AUC 0.998), substantially lower than Rotterdam's 12 [41]. North African women required 18 follicles for optimal accuracy [25], while Chinese adolescents demonstrated best performance at lower thresholds [59]. These variations likely reflect genetic differences in ovarian morphology, follicle dynamics, and perhaps environmental factors [25, 41]. The implication is clear: universal follicle count thresholds applied across all populations will inevitably result in differential diagnostic accuracy [1, 25].

Ovarian Volume Criteria

Ovarian volume served as an alternative or complementary ultrasonographic marker in multiple diagnostic frameworks. The original Rotterdam threshold of $>10 \text{ cm}^3$ was questioned by Jonard et al., who found that lowering the threshold to 7 cm^3 offered the best compromise between specificity (91.2%) and sensitivity (67.5%) [40]. The 10 cm^3 threshold yielded excellent specificity (98.2%) but unacceptably low sensitivity (45%), potentially missing many women with PCOS [40]. Meta-analysis confirmed OV's pooled sensitivity of 81% and specificity of 81% (AUC 0.856), representing a robust alternative when follicle counting cannot be accurately performed [1].

Population-specific OV thresholds demonstrated similar variation to follicle counts. Turkish women showed optimal performance at 6.43 cm^3 [41], Chinese adolescents at 6.74 cm^3 for mean OV and 7.82 cm^3 for maximum OV [59], North African women at 9.25 mL [25], and Iraqi women at $>9.4 \text{ mL}$ [74]. These variations underscore the importance of population-specific validation for all ultrasonographic diagnostic parameters [25, 41, 59].

Age-dependent changes in ovarian volume were documented in longitudinal studies. Alsamarai et al. found that ovarian volume decreased significantly over 7-15 years of follow-up in both PCOS and control women, though PCOS women maintained relatively higher volumes throughout [50]. This age-related decline necessitates age-based criteria, particularly in women approaching 40 years where the distinction between PCOS and age-appropriate ovarian changes becomes more subtle [50].

Geographic and Ethnic Variations

The substantial geographic and ethnic variations documented across multiple diagnostic parameters represent one of the most important findings of this review. Studies from North America showed lower diagnostic accuracy for follicle number per ovary compared to Asian studies [1], while European studies demonstrated higher specificity and positive likelihood ratios for ovarian volume compared to North American studies [1]. These variations likely reflect a complex interplay of genetic, environmental, and technical factors [1, 25].

Differences in body composition contribute significantly to geographic variations in diagnostic accuracy. North American PCOS populations had higher mean BMI ($30.29 \pm 0.61 \text{ kg/m}^2$) compared to European populations ($27.62 \pm 0.73 \text{ kg/m}^2$) [5]. BMI independently affects diagnostic

markers through multiple mechanisms: obesity alters hormone binding proteins and free hormone concentrations [55], affects granulosa cell function [19], and may influence ultrasound visualization [1]. These BMI-related effects compound geographic variations and necessitate population-specific calibration of diagnostic thresholds [1, 5].

Genetic differences likely play fundamental roles in determining optimal diagnostic thresholds. Studies of Turkish [41], Chinese [59], North African [25], Indian [32], and Sudanese [43] women revealed threshold variations that cannot be fully explained by technical factors or BMI differences alone. These genetic influences may affect ovarian reserve [80], hormone metabolism [78], receptor sensitivity [62], and the phenotypic expression of PCOS [6, 57]. The clinical implication is that diagnostic criteria developed in Western European populations cannot be assumed to perform optimally in other ethnic groups without validation [1, 25].

Environmental factors including diet, stress, and environmental exposures may also contribute to geographic variations [1]. However, disentangling genetic from environmental influences requires carefully designed studies comparing migrant populations with their countries of origin—research that remains largely absent from the literature [1].

Technical factors should not be overlooked as contributors to observed geographic variations. Differences in ultrasound equipment quality, operator training, and counting protocols may systematically affect reported follicle counts and ovarian volumes [11, 25]. Similarly, variations in hormone assay methods, calibration standards, and reference ranges may influence biochemical marker performance [78]. While these technical factors do not diminish the importance of population-specific thresholds, they underscore the need for standardized methodologies in future research [1, 78].

Phenotype-Specific Diagnostic Performance

The diagnostic accuracy of all evaluated markers varied substantially across different PCOS phenotypes as defined by Rotterdam criteria [6, 8, 57]. Phenotype A (hyperandrogenism + oligo-ovulation + PCOM), representing the most severe and complete expression of the syndrome, consistently showed the highest diagnostic accuracy for all biomarkers [6, 8]. Cortes et al. found that AMH levels were significantly higher in phenotype A (mean 9.88 ng/mL) compared to

phenotypes B (6.48 ng/mL), C (6.32 ng/mL), and D (5.56 ng/mL) [6]. This gradient across phenotypes suggests that diagnostic markers are most reliable when the full syndrome is expressed and become progressively less discriminative as phenotypic expression becomes more limited [6, 8].

Bozdag et al. specifically examined phenotype-dependent AMH performance and found satisfactory diagnostic potential for phenotype A but only fair validity for other phenotypes [7]. This pattern has important clinical implications: women with milder phenotypes (particularly phenotype D, lacking hyperandrogenism) may be most difficult to diagnose accurately, yet these are precisely the women who may benefit from early identification and intervention [7, 51]. The Rotterdam criteria's inclusion of these milder phenotypes offers advantages for early case identification, but at the cost of lower diagnostic specificity and greater diagnostic uncertainty [7].

Phenotype-specific clinical implications extend beyond diagnostic accuracy. Women with phenotype A demonstrate higher insulin resistance and more pronounced metabolic derangements compared to other phenotypes [29, 63]. This suggests that different phenotypes may represent varying degrees of severity along a continuum rather than distinct disease entities [8]. The diagnostic approach should therefore consider not only whether PCOS is present, but which phenotypic expression characterizes the individual patient, as this informs prognosis and management [6, 8, 57].

The challenge of diagnosing milder phenotypes emerged as a consistent theme across studies. These women may be missed entirely by criteria emphasizing androgen excess (NIH, AES) but could still experience significant reproductive and metabolic consequences [51, 57]. The optimal diagnostic strategy may therefore depend on clinical context: research studies requiring high specificity may prefer NIH or AES criteria, while clinical settings emphasizing early detection may favor the more inclusive Rotterdam framework [7, 8].

Biochemical Markers of Hyperandrogenism

Testosterone and Androgen Indices

Testosterone measurements remain central to the biochemical diagnosis of hyperandrogenism across all diagnostic criteria [78]. However, the method of testosterone

measurement significantly impacts diagnostic accuracy. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) demonstrated superior sensitivity and specificity compared to direct immunoassay methods, leading international guidelines to recommend LC-MS/MS for androgen measurement whenever possible [78]. The improved accuracy of LC-MS/MS reflects its ability to overcome the cross-reactivity and matrix effects that plague immunoassays, particularly at the low testosterone concentrations typical of women [78].

Calculated free testosterone (cFT) consistently outperformed total testosterone as a diagnostic marker across multiple studies [55, 62, 78]. A systematic review by Bizuneh et al. found cFT had pooled sensitivity of 0.89 and specificity of 0.83 (AUC 0.85), compared to total testosterone's 0.74 sensitivity and 0.86 specificity (AUC 0.87) [78]. The superior performance of cFT reflects its accounting for sex hormone-binding globulin (SHBG) levels, which are frequently decreased in PCOS due to hyperinsulinemia, thereby increasing the biologically active free hormone fraction [55, 62].

The free androgen index (FAI), calculated as total testosterone divided by SHBG multiplied by 100, emerged as a practical and widely available marker with strong diagnostic potential. Khashchenko et al. found an optimal FAI cutoff of >2.75 achieved sensitivity of 75% and specificity of 93% in adolescents [14]. Nadaraja et al. reported ROC analysis showing FAI AUC values ranging from 0.78 to 0.85 across populations [55]. The FAI's advantage lies in its calculation from routinely available clinical assays, requiring no specialized equipment or additional blood draws [55].

Bioavailable testosterone, which includes free plus albumin-bound fractions, showed the highest AUC among testosterone-based measures in one comparative study (0.792), followed closely by free testosterone (0.791) and FAI (0.782) [62]. This consistent pattern across androgen indices reinforces the importance of accounting for SHBG when assessing androgen status in PCOS [55, 62, 78].

Androstenedione and Adrenal Androgens

Androstenedione demonstrated promise as a diagnostic marker, particularly in specific populations and clinical contexts. Wang et al. found that androstenedione showed significant

improvement in discriminatory value compared to age alone in women with infertility and obesity, with Youden index of 0.65 (sensitivity 81.82%, specificity 83.16%) [35, 61]. This suggests that androstenedione may be particularly valuable in populations where testosterone measurements are less discriminative [35].

A systematic review by Bizuneh et al. found androstenedione had pooled sensitivity of 0.75 and specificity of 0.71 (AUC 0.80), lower than testosterone-based measures but still representing clinically useful diagnostic information [78]. Guidelines suggest that androstenedione can be considered as a secondary marker if total or free testosterone are not elevated, potentially capturing women with adrenal androgen excess who might otherwise be missed [78].

Dehydroepiandrosterone sulfate (DHEA-S) showed the lowest diagnostic accuracy among commonly measured androgens, with pooled sensitivity of 0.75 and specificity of 0.67 (AUC 0.77) [78]. While DHEA-S levels were significantly elevated in PCOS women, considerable overlap between groups limited diagnostic utility [14, 29]. DHEA-S measurement may be most valuable for identifying adrenal contributions to hyperandrogenism rather than for primary diagnosis [78].

Novel Androgen Markers

Emerging evidence supports the potential utility of novel androgen markers in specific contexts. Dihydrotestosterone (DHT) showed excellent diagnostic accuracy in a case-control study by Hmoud et al., with serum levels significantly higher in PCOS patients (4256 ± 233 ng/L) compared to controls (3776 ± 186 ng/L) [66]. An optimal cutoff of approximately 408.3 ng/L yielded sensitivity of 90% and specificity of 100% (AUC 0.93) [66]. While promising, this marker requires validation in larger, more diverse populations before clinical implementation [66].

Salivary testosterone (salT) emerged as a non-invasive alternative to serum measurements with at least equivalent predictive value to the free androgen index [63]. Sathyapalan et al. demonstrated that salT levels were significantly elevated in PCOS patients and that the combination of salT or FAI identified 100% of PCOS patients [63]. Salivary measurement offers advantages of convenience, non-invasiveness, and reflects the free (biologically active) hormone fraction [63]. However, standardization of collection protocols and assay methods requires further development [63].

Urinary steroid profiling using gas chromatography-mass spectrometry identified specific metabolite patterns with high diagnostic accuracy in a pilot study by Dhayat et al. [42]. The best diagnostic combination included androstanediol, estriol, cortisol, and 20 β DH-cortisone, achieving sensitivity of 90.2% and specificity of 90.8% (AUC 0.961) [42]. This approach offers potential for comprehensive diagnosis without reliance on ultrasound or multiple blood tests, though the requirement for specialized laboratory capabilities limits immediate clinical applicability [42].

Combined Diagnostic Approaches and Novel Technologies

Multivariate Models and Algorithms

Multivariate diagnostic models consistently outperformed single-marker approaches across multiple studies [3, 14, 33, 64]. The superiority of combined approaches reflects the multifactorial nature of PCOS pathophysiology, with different markers capturing distinct aspects of the syndrome [3]. Optimal models typically incorporated markers reflecting ovarian function (AMH, follicle count), androgen status (testosterone, FAI, androstenedione), and hypothalamic-pituitary-ovarian axis function (LH/FSH ratio, menstrual history) [14, 33, 64].

Khashchenko et al. demonstrated that four hormonal parameters (AMH, FAI, androstenedione, testosterone) combined yielded diagnostic accuracy of 90.2-91.6% in adolescents [14]. Insenser et al. found that adding waist-to-hip ratio and LH/FSH ratio to AMH increased AUC to 0.895, with positive predictive values up to 88% [64]. These incremental improvements suggest that while single markers offer convenience, comprehensive diagnostic assessment should incorporate multiple complementary measures when clinically feasible [14, 64].

Age-adjusted approaches using multiples of the median (MoM) showed particular promise for improving diagnostic accuracy across the reproductive lifespan [34, 71]. Palomaki et al. demonstrated that adjusting AMH levels for age and BMI enabled patient-specific PCOS risk computation with improved discrimination compared to fixed thresholds [71]. This personalized approach recognizes that diagnostic marker distributions shift with age and body composition, and that optimal interpretation requires accounting for these factors [34, 71].

Machine Learning and Artificial Intelligence

The application of machine learning and artificial intelligence to PCOS diagnosis represents an emerging frontier with promising initial results [79]. Barrera et al.'s systematic review of 31 studies found diagnostic accuracy ranging from 89% to 100% across four studies, with sensitivity from 41% to 100% and specificity from 75% to 100% [79]. The most common AI techniques included support vector machines (42% of studies), K-nearest neighbor (26%), and regression models (23%) [79].

However, significant methodological limitations temper enthusiasm for immediate clinical implementation. Only 32% of studies employed standardized diagnostic criteria (NIH, Rotterdam, or International PCOS classification) as reference standards, raising concerns about the validity of derived algorithms [79]. Most studies lacked external validation, and few compared AI performance to clinician judgment [79]. While machine learning approaches hold promise for integrating complex multi-parametric data into diagnostic decisions, rigorous validation studies using standardized outcome definitions are needed before clinical adoption [79].

Circulating MicroRNAs and Epigenetic Markers

Novel biomarkers including circulating microRNAs showed promise for PCOS diagnosis in early-phase studies [64, 67]. Insenser et al. validated models based on miR-142-3p and miR-598-3p expression demonstrating good discrimination between women with and without PCOS (AUC 0.801) [64]. Diagnostic accuracy improved when miRNAs were combined with easily available clinical measurements, achieving AUC of 0.895 with positive predictive values up to 88% [64].

DNA methylation patterns also showed diagnostic potential. Xie et al. identified a combination of three CpG methylation signatures (cg23464743, cg06834912, cg00103771) achieving sensitivity of 73.3% and specificity of 93.3% (AUC 0.8378) [67]. These epigenetic markers reflect the interaction between genetic predisposition and environmental exposures, potentially capturing aspects of PCOS pathophysiology not accessible through conventional hormone measurements [67].

While these novel biomarkers offer exciting possibilities for improved diagnostic accuracy and pathophysiological insights, they remain investigational. Validation in larger, diverse

populations, standardization of laboratory methods, and demonstration of incremental value over existing diagnostic approaches are required before clinical implementation [64, 67].

Age-Related Diagnostic Considerations

Age emerged as a critical factor influencing the accuracy of all diagnostic criteria and markers for PCOS [4, 48, 50]. Both ovarian volume and follicle number decreased significantly with age in women with PCOS and controls, though the rate of decline was less pronounced in PCOS women [50]. This age-related change creates a "moving target" for diagnosis, requiring age-stratified reference ranges to maintain optimal sensitivity and specificity across the reproductive lifespan [4, 48].

For ultrasound parameters, optimal thresholds showed consistent patterns of decline with age. Ahmad et al. demonstrated that follicle number per ovary thresholds decreased from >15 in women aged 25-<30 years to >14 in ages 30-<35 years and >12 in ages 35-<40 years [48]. Similarly, ovarian volume thresholds required downward adjustment with advancing age [50]. Studies using fixed thresholds across all ages demonstrated reduced diagnostic accuracy, particularly in older reproductive-age women where PCOS may be underdiagnosed [48, 50].

AMH thresholds showed even more pronounced age-dependent variation [4, 37, 65]. The comprehensive analysis by Ramezani Tehrani et al. established that optimal cutoffs declined from 5.7 ng/mL in women aged 20-27 years to 3.72 ng/mL in those aged 35-40 years [4]. This threefold variation underscores the impossibility of a single universal AMH cutoff serving all reproductive-age women [4]. Studies applying a fixed AMH threshold of 3.2 ng/mL across all ages demonstrated lower sensitivity in younger women and lower specificity in older women [22, 36].

The mechanism underlying age-related changes involves progressive loss of ovarian reserve affecting both PCOS and non-PCOS women [50]. However, women with PCOS maintain relatively higher follicle counts and AMH levels even as absolute values decline, creating a persistent but diminishing separation from normal women [50]. This pattern suggests that age-stratified approaches are essential, but also that diagnostic confidence may decrease in older reproductive-age women as the separation between PCOS and normal narrows [48, 50].

The clinical implications of age-related diagnostic variation extend beyond technical considerations. Delayed diagnosis of PCOS in older reproductive-age women may result in missed opportunities for fertility preservation and early metabolic intervention [20]. Conversely, overdiagnosis in younger women using age-inappropriate thresholds could lead to unnecessary medicalization and psychological burden [4]. Age-appropriate diagnostic thresholds are therefore essential for optimal clinical care [4, 48].

Practical and Clinical Implications

The findings of this systematic review have several important implications for clinical practice. First, no single diagnostic criterion can be recommended as universally optimal; rather, the choice of criteria should be guided by clinical context, population characteristics, and the purpose of diagnosis [1, 7]. In research settings requiring high specificity, NIH or AES criteria may be preferred [2, 7]. In clinical settings emphasizing early detection and comprehensive case identification, Rotterdam criteria offer advantages [1, 3]. In resource-limited settings, simplified approaches based on self-reported symptoms may provide acceptable diagnostic accuracy for initial screening [10, 68].

Second, age-stratified diagnostic thresholds are essential for all diagnostic markers [4, 48]. Clinicians should interpret AMH levels, follicle counts, and ovarian volume in the context of the patient's chronological age, using reference ranges specific to the patient's age group [4, 48]. Laboratory and ultrasound reports should ideally provide age-specific interpretation rather than applying universal thresholds [48].

Third, population-specific validation of diagnostic thresholds is necessary for optimal accuracy [1, 25, 41]. Diagnostic criteria and thresholds developed in Western European populations cannot be assumed to perform optimally in other ethnic groups without validation [1, 25]. Multicenter studies establishing population-specific reference ranges should be prioritized in regions where such data are lacking [1].

Fourth, AMH shows promise as an objective alternative to ultrasound assessment of PCOM, particularly when age-appropriate thresholds are applied [9, 22]. The substitution of AMH for ultrasound may improve diagnostic accessibility in resource-limited settings and reduce operator-

dependent variability [20]. However, population-specific validation and assay standardization remain essential prerequisites for widespread implementation [22, 32].

Fifth, combined diagnostic approaches using multiple complementary markers achieve superior accuracy to single markers [3, 14, 33]. Whenever feasible, diagnostic assessment should incorporate clinical features (menstrual history, signs of hyperandrogenism), biochemical markers (AMH, androgens), and ultrasound findings in an integrated manner [3]. The incremental value of each additional marker should be considered in light of clinical context and resource availability [14].

Sixth, phenotype-specific diagnosis and management are essential [6, 8, 57]. Simply diagnosing PCOS is insufficient; clinicians should characterize which phenotypic expression the patient manifests, as this informs prognosis and guides management [6, 8]. Women with phenotype A (full PCOS) require particularly intensive metabolic surveillance, while those with phenotype D may have different reproductive implications [6, 57].

CONCLUSION AND RECOMMENDATIONS

Summary of Key Findings

This systematic review of 80 studies provides comprehensive evidence regarding the diagnostic accuracy of Rotterdam, NIH, and AES criteria for PCOS in reproductive-age women. The Rotterdam criteria demonstrate the highest sensitivity but lowest specificity, identifying the broadest spectrum of PCOS phenotypes including milder presentations without hyperandrogenism. The NIH criteria offer greater specificity for classic PCOS, identifying fewer women but potentially those at highest metabolic risk. The AES criteria provide an intermediate approach emphasizing androgen excess as the cardinal feature.

Anti-Müllerian hormone emerged as the most promising diagnostic biomarker, with consistent evidence supporting its utility as both an adjunctive marker and potential replacement for ultrasound assessment of PCOM. Age-specific AMH thresholds are essential, with optimal cutoffs declining from approximately 5.7 ng/mL in early reproductive years to 3.7 ng/mL approaching age

40. Population-specific variations in AMH thresholds necessitate local validation before widespread implementation.

Ultrasonographic criteria require substantial revision based on contemporary evidence. The original Rotterdam threshold of 12 follicles for PCOM is inadequate with modern ultrasound technology; thresholds of 26 follicles or higher may be more appropriate. Age-specific and population-specific follicle count thresholds are essential for maintaining diagnostic accuracy across diverse populations and reproductive stages.

Geographic and ethnic variations in optimal diagnostic thresholds are substantial and clinically significant. Turkish, Chinese, North African, Indian, and Sudanese populations demonstrate optimal thresholds that differ markedly from those established in Western European populations, underscoring the need for population-specific diagnostic standards.

Combined diagnostic approaches incorporating multiple complementary markers consistently outperform single-marker strategies. Models integrating AMH, androgens, clinical features, and anthropometric measurements achieve the highest diagnostic accuracy, suggesting that comprehensive assessment remains superior to simplified approaches when resources permit.

Clinical Recommendations

Based on the evidence synthesized in this review, we offer the following recommendations for clinical practice:

1. The choice of diagnostic criteria should be guided by clinical context. For comprehensive case identification in general practice or fertility clinics, Rotterdam criteria are appropriate. For research studies requiring high specificity or for identifying women at highest metabolic risk, NIH or AES criteria may be preferred.
2. Age-stratified diagnostic thresholds should be applied for all markers, including AMH, follicle counts, and ovarian volume. Fixed thresholds applied across all reproductive ages result in systematic misclassification.

3. Population-specific validation of diagnostic thresholds should be undertaken before applying criteria developed in Western populations to other ethnic groups. Multicenter studies establishing regional reference ranges should be prioritized.
4. AMH measurement using standardized assays can be considered as an alternative to ultrasound assessment of PCOM when high-quality ultrasound is unavailable or when operator-dependent variability is a concern. Age-appropriate AMH thresholds must be applied.
5. Combined diagnostic approaches incorporating multiple complementary markers should be used whenever feasible. Single-marker strategies, while convenient, sacrifice diagnostic accuracy.
6. Phenotypic characterization beyond simple diagnosis is essential for optimal management. Clinicians should determine which Rotterdam phenotype (A-D) characterizes each patient, as this informs prognosis and guides management.
7. Simplified screening approaches based on self-reported symptoms (menstrual irregularity, hirsutism, depilatory practices) may be valuable for initial case identification in resource-limited settings or epidemiologic research.

Future Research Directions

This review identifies several priorities for future research. First, prospective cohort studies using long-term outcomes as reference standards are needed to determine which diagnostic approaches best predict clinically meaningful endpoints including infertility, metabolic disease, cardiovascular events, and quality of life.

Second, standardized methodologies for diagnostic accuracy studies in PCOS should be developed and implemented, including consensus on optimal ultrasound protocols, hormone assay methods, and threshold selection approaches.

Third, multi-ethnic collaborative studies to establish population-specific reference ranges for all diagnostic markers should be prioritized, particularly in underrepresented regions.

Fourth, validation studies of AMH as a replacement for ultrasound PCOM should be expanded to diverse populations with standardized assays, establishing age-specific and population-specific thresholds.

Fifth, machine learning and artificial intelligence approaches to integrating complex multi-parametric data should be rigorously validated against standardized outcome definitions in diverse populations.

Sixth, novel biomarkers including microRNAs, epigenetic markers, and steroid metabolites require validation in large, diverse cohorts with demonstration of incremental value over existing diagnostic approaches.

Seventh, implementation research evaluating how best to translate evidence-based diagnostic recommendations into clinical practice across diverse healthcare settings is needed to ensure that advances in diagnostic accuracy translate into improved patient outcomes.

In conclusion, the diagnostic accuracy of PCOS criteria varies substantially across populations, age groups, and clinical contexts. Optimal diagnosis requires an individualized approach that considers the patient's age, ethnicity, clinical presentation, and the purpose of diagnosis. The integration of AMH as a diagnostic marker, the application of age-specific and population-specific thresholds, and the use of combined diagnostic approaches offer the potential for improved diagnostic accuracy and, ultimately, better outcomes for women with PCOS.

REFERENCES

1. J. Pea, Jahnay Bryan, Cynthia Wan, et al (2023) Ultrasonographic criteria in the diagnosis of polycystic ovary syndrome: a systematic review and diagnostic meta-analysis. Human Reproduction Update. <https://doi.org/10.1093/humupd/dmad027>
2. L. Casadei, Francesca Fanisio, R. Sorge, et al (2018) The diagnosis of PCOS in young infertile women according to different diagnostic criteria: the role of serum anti-Müllerian hormone. Archives of Gynecology and Obstetrics. <https://doi.org/10.1007/s00404-018-4803-8>

3. S. Şahmay, Y. Aydın, M. Oncul, L. Senturk (2014) Diagnosis of Polycystic Ovary Syndrome: AMH in combination with clinical symptoms. Journal of Assisted Reproduction and Genetics. <https://doi.org/10.1007/s10815-013-0149-0>
4. F. Ramezani Tehrani, M. Rahmati, Fatemeh Mahboobifard, et al (2021) Age-specific cut-off levels of anti-Müllerian hormone can be used as diagnostic markers for polycystic ovary syndrome. Reproductive Biology and Endocrinology. <https://doi.org/10.1186/s12958-021-00755-8>
5. J. Pea, Jahnay Bryan, Alexis L Oldfield, et al (2022) LBMON234 Ultrasonographic Criteria In The Diagnosis Of Polycystic Ovary Syndrome: A Systematic Review And Meta Analysis. Journal of the Endocrine Society. <https://doi.org/10.1210/jendso/bvac150.1342>
6. Christina I. Cortes, Justin Armstrong, K. Hawkins, A. Younis (2021) ANTI-MULLERIAN HORMONE (AMH) LEVELS IS EFFECTIVE IN DETECTION AND DIAGNOSIS OF VARIOUS POLYCYSTIC OVARIAN SYNDROME (PCOS) PHENOTYPES. Fertility and Sterility. <https://doi.org/10.1016/j.fertnstert.2021.07.337>
7. G. Bozdag, S. Mumuşoğlu, Z. Y. Coşkun, et al (2019) Anti-Müllerian hormone as a diagnostic tool for PCOS under different diagnostic criteria in an unselected population. Reproductive Biomedicine Online. <https://doi.org/10.1016/J.RBMO.2019.04.002>
8. F. Vural, B. Vural, Eyupcan Kardaş, et al (2022) The diagnostic performance of antimüllerian hormone for polycystic ovarian syndrome and polycystic ovarian morphology. Archives of Gynecology and Obstetrics. <https://doi.org/10.1007/s00404-022-06874-w>
9. Yang Zhao, Yinlong Zhao, Chunpeng Wang, et al (2019) DIAGNOSTIC VALUE OF ANTI-MÜLLERIAN HORMONE AS A BIOMARKER FOR POLYCYSTIC OVARY SYNDROME: A META-ANALYSIS UPDATE. Endocrine Practice. <https://doi.org/10.4158/EP-2019-0098>
10. J. Chan, M. Pall, U. Ezech, et al (2019) Screening for androgen excess in women: accuracy of self-reported excess body hair growth and menstrual dysfunction. Fertility and Sterility. <https://doi.org/10.1016/j.fertnstert.2019.07.247>

11. M. Lujan, Brittany Y. Jarrett, E. Brooks, et al (2013) Updated ultrasound criteria for polycystic ovary syndrome: reliable thresholds for elevated follicle population and ovarian volume. *Human Reproduction*. <https://doi.org/10.1093/humrep/det062>
12. Mariane de Oliveira Gomes, Juliane de Oliveira Gomes, L. F. Ananias, et al (2025) ANTI-MÜLLERIAN HORMONE AS A DIAGNOSTIC MARKER OF POLYCYSTIC OVARY SYNDROME: A SYSTEMATIC REVIEW WITH META-ANALYSIS. *American Journal of Obstetrics and Gynecology*. <https://doi.org/10.1016/j.ajog.2025.01.044>
13. Yumiko Tsukui, Yoshikazu Kitahara, Y. Hasegawa, et al (2022) Anti-Müllerian hormone levels in the diagnosis of adolescent polycystic ovarian syndrome: a systematic review and meta-analysis. *Endocrine journal*. <https://doi.org/10.1507/endocrj.ej22-0081>
14. E. Khashchenko, E. Uvarova, M. Vysokikh, et al (2020) The Relevant Hormonal Levels and Diagnostic Features of Polycystic Ovary Syndrome in Adolescents. *Journal of Clinical Medicine*. <https://doi.org/10.3390/jcm9061831>
15. Abhimanyu Pandit, Bikram Bhardwaj, Souvik Nandy, Rony Chakravarty (2024) Study of Anti-mullerian Hormone as a Diagnostic Criteria for Polycystic Ovarian Syndrome. *Medical Journal of Dr DY Patil Vidyapeeth*. https://doi.org/10.4103/mjdrdypu.mjdrdypu_557_23
16. E. Hariton, J. Morris, C. Kao, et al (2020) AN AMH-BASED SCREENING TOOL IS BOTH SENSITIVE AND SPECIFIC FOR PREDICTING A DIAGNOSIS OF PCOS BY ROTTERDAM CRITERIA. <https://doi.org/10.1016/j.fertnstert.2020.08.137>
17. Nada Ahmed, Asma A. Batarfi, O. Bajouh, S. Bakhashab (2019) Serum Anti-Müllerian Hormone in the Diagnosis of Polycystic Ovary Syndrome in Association with Clinical Symptoms. *Diagnostics*. <https://doi.org/10.3390/diagnostics9040136>
18. Khansa Chaabouni, Rihab Makhlouf, Fatma Khanfir, et al (2023) The anti-Müllerian hormone: A useful tool for diagnosis of Polycystic Ovary Syndrome. *Annales de Biologie Clinique*. <https://doi.org/10.1684/abc.2023.1830>
19. I. Oueslati, M. Hammami, S. Boukriba, et al (2022) Anti Mullerian hormone as a diagnostic tool for polycystic ovary syndrome in women of reproductive age with morbid obesity. *Hormone Molecular Biology and Clinical Investigation*. <https://doi.org/10.1515/hmbci-2021-0078>

20. T. Piltonen, Deirdre Allegranza, M. Hund, et al (2024) Validation of an Anti-Müllerian Hormone Cutoff for Polycystic Ovarian Morphology in the Diagnosis of Polycystic Ovary Syndrome in the HARMONIA Study: Protocol for a Prospective, Noninterventional Study. JMIR Research Protocols. <https://doi.org/10.2196/48854>
21. Didier Dewailly, M. Alebić, Alain Duhamel, Nataša Stojanović (2014) Using cluster analysis to identify a homogeneous subpopulation of women with polycystic ovarian morphology in a population of non-hyperandrogenic women with regular menstrual cycles. Human Reproduction. <https://doi.org/10.1093/humrep/deu242>
22. Alexandra Dietz de Loos, M. Hund, K. Buck, et al (2021) Antimüllerian hormone to determine polycystic ovarian morphology. Fertility and Sterility. <https://doi.org/10.1016/j.fertnstert.2021.05.094>
23. Mostafa Barghi, Zahra Heidari, F. Haghighatdoost, et al (2024) New insights into the relationship between Anti-Mullerian Hormone and Polycystic ovary syndrome and its diagnostic accuracy: An updated and extended meta-analysis using marginal beta-binomial model. American Journal of Obstetrics and Gynecology. <https://doi.org/10.1016/j.ajog.2024.10.004>
24. A. D. D. Loos, M. Hund, K. Buck, et al (2019) Establishing an anti-müllerian hormone (AMH) cut-off to determine polycystic ovarian morphology (PCOM) supporting diagnosis of polycystic ovarian syndrome (PCOS): The APHRODITE study. Endocrine Abstracts. <https://doi.org/10.1016/j.fertnstert.2019.07.1116>
25. T. Ach, A. Guesmi, Maha Kalboussi, et al (2024) Validation of the follicular and ovarian thresholds by an 18-MHz ultrasound imaging in polycystic ovary syndrome: a pilot cutoff for North African patients. Therapeutic Advances in Reproductive Health. <https://doi.org/10.1177/26334941241270372>
26. Shahrzad Zadehmodarres, Z. Heidar, Zahra Razzaghi, et al (2015) Anti-mullerian hormon level and polycystic ovarian syndrome diagnosis. Iranian Journal of Reproductive Medicine
27. R. Bell, R. Islam, M. Skiba, et al (2021) Substituting serum anti-Müllerian hormone for polycystic ovary morphology increases the number of women diagnosed with polycystic ovary

- syndrome: a community-based cross-sectional study. Human Reproduction. <https://doi.org/10.1093/humrep/deab232>
28. P. R. Kumar, Radhakrishna Telagareddy, D. K. Dash, Debasish Patro (2024) Diagnostic Utility of Various Hormones across Different Polycystic Ovary Syndrome Phenotypes: A Cross-sectional Study. Journal of Human Reproductive Sciences. https://doi.org/10.4103/jhrs.jhrs_152_24
29. Prekshi Bansal, K. Sardana, P. Arora, et al (2020) A prospective study of anti-mullerian hormone and other ovarian and adrenal hormones in adult female acne. Dermatologic Therapy. <https://doi.org/10.1111/dth.13974>
30. Raghad Fadhil Taha, Mushtak T. S. Al-Ouqaili, Sulaiman Abdullah (2023) The Association of Anti-Mullerian Hormone and Infertility Hormonal Imbalance with Polycystic Ovarian Syndrome Among Iraqi Patients. Pakistan Journal of Biological Sciences. <https://doi.org/10.3923/pjbs.2023.241.248>
31. Naveetha Lakshmi Narayanaswamy, Freethi Ramanathan, Moonishaa Thiyagarajan Manjuladevi (2023) Diagnostic utility of serum anti-Mullerian hormone levels in south Indian women with polycystic ovarian syndrome. BioMedicine. <https://doi.org/10.51248/v43i4.2312>
32. N. Mahajan, J. Kaur (2019) Establishing an Anti-Müllerian Hormone Cutoff for Diagnosis of Polycystic Ovarian Syndrome in Women of Reproductive Age-Bearing Indian Ethnicity Using the Automated Anti-Müllerian Hormone Assay. Journal of Human Reproductive Sciences. https://doi.org/10.4103/jhrs.JHRS_149_18
33. Lingling Jiang, X. Ruan, Yanqiu Li, et al (2023) Diagnostic value of anti-Müllerian hormone combined with androgen-levels in Chinese patients with polycystic ovary syndrome. Gynecological Endocrinology. <https://doi.org/10.1080/09513590.2023.2206927>
34. H. W. Raymond Li, Yilei He, Rong Li, et al (2020) SUN-021 Age-Specific Multiples of the Median Level of Serum Anti-Mullerian Hormone Is a Potential Marker for Diagnosis of Polycystic Ovary Syndrome. Journal of the Endocrine Society. <https://doi.org/10.1210/jendso/bvaa046.1228>
35. Z. Wang, A. van der Veen, H. Groen, et al (2022) P-590 Discriminatory value of steroid hormones on polycystic ovary syndrome in women with infertility and obesity. Human Reproduction. <https://doi.org/10.1093/humrep/deac107.544>

36. L. Lotz, R. Dittrich, J. Sillman, et al (2025) P-638 Anti-Müllerian hormone to determine polycystic ovarian morphology in young adult women 18-24 years of age. Human Reproduction. <https://doi.org/10.1093/humrep/deaf097.944>
37. Yang You, Zaixin Guo, Xinyu Hong, et al (2026) Age- and BMI-stratified assessment of serum anti-Müllerian hormone as a biomarker for polycystic ovary syndrome diagnosis in Chinese women. BMC Endocrine Disorders. <https://doi.org/10.1186/s12902-025-02136-3>
38. Dr.Rajrani Sharma (2017) Correlation of Anti-Mullerian Hormone with clinical, hormonal and ultrasonographic parameters in PCOS and normo-ovulatory women. <https://doi.org/10.9790/0853-1602085156>
39. S. Iliodromiti, T. Kelsey, R. Anderson, S. Nelson (2013) Can anti-Mullerian hormone predict the diagnosis of polycystic ovary syndrome? A systematic review and meta-analysis of extracted data. Journal of Clinical Endocrinology and Metabolism. <https://doi.org/10.1210/jc.2013-1393>
40. S. Jonard, Y. Robert, D. Dewailly (2005) Revisiting the ovarian volume as a diagnostic criterion for polycystic ovaries. Human Reproduction. <https://doi.org/10.1093/HUMREP/DEI159>
41. N. Köşüş, A. Köşüş, N. Turhan, Z. Kamalak (2011) Do threshold values of ovarian volume and follicle number for diagnosing polycystic ovarian syndrome in Turkish women differ from western countries? European Journal of Obstetrics, Gynecology, and Reproductive Biology. <https://doi.org/10.1016/j.ejogrb.2010.10.007>
42. N. Dhayat, N. Marti, Z. Kollmann, et al (2018) Urinary steroid profiling in women hints at a diagnostic signature of the polycystic ovary syndrome: A pilot study considering neglected steroid metabolites. PLoS ONE. <https://doi.org/10.1371/journal.pone.0203903>
43. S. Alhassan, Abdelgadir Elmugadam, G. Elfadil, et al (2023) Diagnostic Performance of Anti-Müllerian Hormone, Luteinizing Hormone to Follicle-Stimulating Hormone Ratio, Testosterone, and Prolactin to Predict Polycystic Ovary Syndrome Among Sudanese Women. International Journal of Women's Health. <https://doi.org/10.2147/IJWH.S403347>

44. D. Dewailly (2012) Faculty Opinions recommendation of Anti-Mullerian hormone in the diagnosis of polycystic ovary syndrome: can morphologic description be replaced? <https://doi.org/10.3410/f.717950032.793455346>
45. P. Pigny, S. Jonard, Y. Robert, D. Dewailly (2006) Serum anti-Mullerian hormone as a surrogate for antral follicle count for definition of the polycystic ovary syndrome. *Journal of Clinical Endocrinology and Metabolism*. <https://doi.org/10.1210/JC.2005-2076>
46. M. Feichtinger (2010) News-Screen Assistierte Reproduktion. *Journal für Gynäkologische Endokrinologie/Schweiz*. <https://doi.org/10.1007/s41975-017-0017-x>
47. T. Eilertsen, E. Vanky, S. Carlsen (2012) Anti-Mullerian hormone in the diagnosis of polycystic ovary syndrome: can morphologic description be replaced? *Human Reproduction*. <https://doi.org/10.1093/humrep/des213>
48. Asima K. Ahmad, M. Quinn, C. Kao, et al (2019) Improved diagnostic performance for the diagnosis of polycystic ovary syndrome using age-stratified criteria. *Fertility and Sterility*. <https://doi.org/10.1016/j.fertnstert.2018.11.044>
49. Kulsoom Bahadur, Aamir Ijaz, S. Baqai, Naveed Asif (2023) Diagnostic Accuracy of Anti-Mullerian Hormone for Polycystic Ovary Syndrome. *Pakistan Armed Forces Medical Journal*. <https://doi.org/10.51253/pafmj.v73i2.4249>
50. S. Alsamarai, J. Adams, M. K. Murphy, et al (2009) Criteria for polycystic ovarian morphology in polycystic ovary syndrome as a function of age. *Journal of Clinical Endocrinology and Metabolism*. <https://doi.org/10.1210/jc.2009-0839>
51. B. Joshi, Srabani Mukherjee, Anushree Patil, et al (2014) A cross-sectional study of polycystic ovarian syndrome among adolescent and young girls in Mumbai, India. *Indian journal of endocrinology and metabolism*. <https://doi.org/10.4103/2230-8210.131162>
52. L. Casadei, A. Madrigale, F. Puca, et al (2013) The role of serum anti-Müllerian hormone (AMH) in the hormonal diagnosis of polycystic ovary syndrome. *Gynecological Endocrinology*. <https://doi.org/10.3109/09513590.2013.777415>

53. V. Conoscenti, A. Kho, G. Hayes (2018) Using the electronic medical record (EMR) for identifying women with polycystic ovary syndrome (PCOS). *Fertility and Sterility*. <https://doi.org/10.1016/j.fertnstert.2018.07.360>
54. Cong-shun Ma, Yun Lin, Chun-hui Zhang, et al (2011) [Preliminary study on the value of ratio of serum luteinizing hormone/follicle-stimulating hormone in diagnosis of polycystic ovarian syndrome among women with polycystic ovary]. *Zhonghua fu chan ke za zhi*. <https://doi.org/10.3760/CMA.J.ISSN.0529-567X.2011.03.004>
55. R. Nadaraja, P. Sthaneshwar, N. Razali (2018) Establishing the cut off values of androgen markers in the assessment of polycystic ovarian syndrome. *The Malaysian journal of pathology*
56. Moaz O Moursi, Haya Salem, A. Ibrahim, et al (2023) The role of anti-Müllerian hormone and other correlates in patients with polycystic ovary syndrome. *Gynecological Endocrinology*. <https://doi.org/10.1080/09513590.2023.2247098>
57. Henri Sova, L. Unkila-Kallio, A. Tiitinen, et al (2019) Hormone profiling, including anti-Müllerian hormone (AMH), for the diagnosis of polycystic ovary syndrome (PCOS) and characterization of PCOS phenotypes. *Gynecological Endocrinology*. <https://doi.org/10.1080/09513590.2018.1559807>
58. A. Köninger, L. Koch, P. Edimiris, et al (2014) Anti-Müllerian Hormone: an indicator for the severity of polycystic ovarian syndrome. *Archives of Gynecology and Obstetrics*. <https://doi.org/10.1007/s00404-014-3317-2>
59. Ya-xiao Chen, Dongzi Yang, Lin Li, Xiaoli Chen (2008) The role of ovarian volume as a diagnostic criterion for Chinese adolescents with polycystic ovary syndrome. *Journal of Pediatric and Adolescent Gynecology*. <https://doi.org/10.1016/j.jpbg.2008.01.081>
60. Türkan Gürsu, A. Eraslan, B. Angun (2022) Comparison of body mass index, anti-müllerian hormone and insulin resistance parameters among different phenotypes of polycystic ovarian syndrome. *Gynecology and Obstetrics Clinical Medicine*. <https://doi.org/10.1016/j.gocm.2022.10.002>

61. Wanwan Yi, Mengyu Zhang, Xueyu Yuan, et al (2023) A model combining testosterone, Androstenedione and free testosterone index improved the diagnostic efficiency of polycystic ovary syndrome. *Endocrine Practice*. <https://doi.org/10.1016/j.eprac.2023.05.007>
62. Sikandar hayat Khan, Uzma Urooj, Zeeshan Ali Quereshi, et al (2024) Albumin and Sex Hormone Binding Globulin based Testosterone Indices in Polycystic Ovarian Syndrome. *Pakistan Armed Forces Medical Journal*. <https://doi.org/10.51253/pafmj.v74i6.8627>
63. T. Sathyapalan, Ahmed Al-Qaissi, E. Kilpatrick, et al (2017) Salivary testosterone measurement in women with and without polycystic ovary syndrome. *Scientific Reports*. <https://doi.org/10.1038/s41598-017-03945-w>
64. M. Insenser, A. Quintero, S. de Lope, et al (2023) Validation of circulating microRNAs miR-142-3p and miR-598-3p in women with polycystic ovary syndrome as potential diagnostic markers. *Human Reproduction*. <https://doi.org/10.1093/humrep/dead043>
65. H. T. Timur, D. Çımrın, Ozlem Gursoy Doruk, O. Dogan (2023) Determining the age group-based cut-off values of serum anti-Mullerian hormone concentrations to diagnose polycystic ovary syndrome. *Current Medical Research and Opinion*. <https://doi.org/10.1080/03007995.2023.2204768>
66. Taif Raheem Hmoud, Murtadha M. Jawad, Redha Dawud Abd Alredha (2025) Hormonal profiles and the diagnostic utility of serum dihydrotestosterone in polycystic ovary syndrome: a comparative case-control study. *European Journal of Translational and Clinical Medicine*. <https://doi.org/10.31373/ejtcmm/208403>
67. Linling Xie, Xiaotao Jiang, Yi Chen, et al (2021) 3 CpG methylation biomarkers for the diagnosis of polycystic ovary syndrome (PCOS) in blood. *Combinatorial chemistry & high throughput screening*. <https://doi.org/10.2174/1386207321666210602170054>
68. Bronwyn S. Bedrick, A. Eskew, J. Chavarro, E. Jungheim (2019) Self-Administered Questionnaire to Screen for Polycystic Ovarian Syndrome. *Women's Health Reports*. <https://doi.org/10.1089/whr.2020.0073>

69. Jenna Salem, Robyn Vettese, Jennifer M. Yamamoto, et al (2025) Development of a case definition for polycystic ovary syndrome using administrative health data: a validation study. *BMJ Open*. <https://doi.org/10.1136/bmjopen-2024-097971>
70. Justin Armstrong, Christina Cortes, K. Hawkins, Abdelmoneim Younis (2022) Serum Anti-Mullerian Hormone (AMH) Levels are Effective in Predicting the Diagnosis of Four Polycystic Ovarian Syndrome (PCOS) Phenotypes. *Gynecology & Reproductive Health*. <https://doi.org/10.33425/2639-9342.1203>
71. G. Palomaki, B. Kalra, T. Kumar, et al (2020) Adjusting anti-Müllerian hormone levels for age and body mass index improves detection of polycystic ovary syndrome. *Fertility and Sterility*. <https://doi.org/10.1016/j.fertnstert.2019.12.012>
72. A. M. A. Khaleel, Amr Ali Sharaf El-Din, M. Mohammed, L. K. El Sayed (2025) Anti-Mullerian Hormone (AMH) Variability in Cases of Polycystic Ovary Syndrome (PCOS) Phenotypes. *Benha Journal of Applied Sciences*. <https://doi.org/10.21608/bjas.2025.361960.1595>
73. Zhen Wang, Xiaojing Teng, Yanfei Liu, et al (2025) Establishment of age-related AMH screening cutoffs in Chinese women with PCOS: a retrospective study using propensity score matching analysis. *BMC Endocrine Disorders*. <https://doi.org/10.1186/s12902-025-01975-4>
74. Maal Anwer Al-Hamzawi (2025) Evaluating Ovarian Volume as a Diagnostic Marker for Polycystic Ovary Syndrome: A Comparative Study in Iraqi Women. *Medical Science Journal for Advance Research*. <https://doi.org/10.46966/msjar.v6i1.251>
75. T. Yén, Tang Duc Cuong, Nguyen Thi Van Anh, Tran Thi Chi Mai (2025) Value of serum anti-mullerian hormone in the diagnosis of polycystic ovary syndrome. *Tap chi nghien cuu y hoc*. <https://doi.org/10.52852/tencyh.v196i11e17.3741>
76. S. Abbas, Asma Hayat, Rabiya Shabir, et al (2024) Anti-Mullarian Hormone As A Diagnostic Marker of Polycystic Ovary Syndrome. *Pakistan Armed Forces Medical Journal*. <https://doi.org/10.51253/pafmj.v74i1.9166>
77. C. N. Ndoua, G. Ngah, V. M. Koh, et al (2017) Correlation between the anti-Müllerian hormone and endovaginal ultrasound in the predictivity of polycystic ovary syndrome at Chracerh. <https://doi.org/10.18203/2320-1770.ijrcog20175224>

78. A. Bizuneh, A. Joham, H. Teede, et al (2024) Evaluating the diagnostic accuracy of androgen measurement in polycystic ovary syndrome: a systematic review and diagnostic meta-analysis to inform evidence-based guidelines. Human Reproduction Update. <https://doi.org/10.1093/humupd/dmae028>
79. Francisco J. Barrera, Ethan Brown, Amanda Rojo, et al (2023) Application of machine learning and artificial intelligence in the diagnosis and classification of polycystic ovarian syndrome: a systematic review. Frontiers in Endocrinology. <https://doi.org/10.3389/fendo.2023.1106625>
80. S. L. Fong, S. L. Fong, J. Laven, et al (2017) Polycystic ovarian morphology and the diagnosis of polycystic ovary syndrome: redefining threshold levels for follicle count and serum anti-Müllerian hormone using cluster analysis. Human Reproduction. <https://doi.org/10.1093/humrep/dex226>