Systematic Reviews

SYSTEMATIC REVIEW UPDATE

Open Access

The best ovarian reserve marker to predict ovarian response following controlled ovarian hyperstimulation: a systematic review and meta-analysis

Fateme Salemi[1], Sara Jambarsang[2], Amir Kheirkhah[3], Amin Salehi-Abargouei[4,5], Zahra Ahmadnia[3], Haniye Ali hosseini[3], Marzieh Lotfi[6,7*] and Saad Amer[8*]

Abstract

Background One of the most challenging aspects of treating patients facing primary ovarian insufficiency, especially those eligible for controlled ovarian hyperstimulation (COH), is the assessment of ovarian function and response to stimulatory protocols in terms of the number of oocytes retrieved. The lack of consistency between studies regarding the best parameter for response evaluation necessitates a comprehensive statistical analysis of the most commonly utilized ovarian reserve markers (ORM). This systematic review and meta-analysis aims to establish the optimal metric for assessing ovarian reserve among COH candidates.

Methods The PubMed/MEDLINE, Scopus, and ISI Web of Science databases were searched until July 2024, with no date or language limitations. The Newcastle–Ottawa scale was used to evaluate the validity of anti-Mullerian hormone (AMH), antral follicle count (AFC), follicle-stimulating hormone (FSH), and estradiol (E2) in patients receiving controlled ovarian hyperstimulation. Studies on the diagnostic accuracy of ovarian reserve markers in predicting ovarian response to controlled ovarian hyperstimulation in assisted reproduction technology (ART) candidates were reviewed. The diagnostic odds ratio (DOR) was determined using the Der Simonian-Laird random effects model meta-analysis to assess the likelihood of detecting low or high ovarian responses in COH candidates. Cochran's Q, and I-squared, were used to analyze between-study heterogeneity.

Results This systematic review and meta-analysis included 26 studies including 17 cohorts, 4 case controls, and 5 cross-sectional studies. AFC and AMH demonstrated significant diagnostic performance compared to FSH and E2 in poor and high response category. AMH slightly outperformed AMH and had the highest logarithm of DOR for detecting poor [2.68 (95% CI 1.90, 3.45)] and high ovarian response [2.76 (95% *CI* 1.57, 3.95)]. However, it showed a high between-study heterogeneity ($I^2 = 95.65$, $Q = 189.65$, $p < 0.05$).

Conclusions AFC and AMH were the most accurate predictors of poor and high ovarian response to controlled ovarian hyperstimulation. However, further research is needed to develop models assessing the combined impact of AMH and AFC on ovarian response prediction.

*Correspondence:
Marzieh Lotfi
marzeih.lotfi@gmail.com; m.lotfi@ssu.ac.ir
Saad Amer
Saad.Amer@nottingham.ac.uk
Full list of author information is available at the end of the article

Systematic review registration PROSPERO CRD42021245380.

Keywords Controlled ovarian hyperstimulation, Diagnostic accuracy, Ovarian reserve, Ovarian reserve marker, Ovarian response

Background

Over the last two decades, numerous ovarian reserve tests have been developed to assess oocyte reserve and quality, as well as to predict IVF (in vitro fertilization)/ICSI success in terms of oocyte yield and pregnancy. Many of these tests are now regularly performed on infertile patients undergoing ART [1, 2]. Finding out a patient's reproductive potential, ovarian reserve, and response to COH are the main objectives. A commonly used word, ovarian reserve describes the overall pool of follicles in the ovaries, including both dormant and actively developing follicles [3].

Premature ovarian insufficiency (POI) is a prevalent reproductive endocrine condition characterized by the loss of ovarian function prior to the age of 40 [4]. The rate at which a woman's ovarian reserve depletes is governed by a combination of genetic and environmental factors, and it usually diminishes permanently over time. The rate of follicular depletion accelerates around the age of 37–38 years [5]. Reduced ovarian reserves (ROR) is a term used to describe the rapid decline in ovarian follicular reserve in women in their early thirties, which is a known cause of infertility in this age group [6]. In most cases, ROR affects women in their late 30 s and early 40 s, but younger women may also be affected [7]. When compared to other women of the same age, these women tend to have lower fertility and a poor ovarian response (POR) to ovarian stimulation [8]. Young women should be educated about their future fertility potential and provided adequate advice on the best and most timely medical treatment to help them conceive [9]. Although age is the best predictor of poor oocyte quality, ovarian reserve markers (ORMs) are more commonly utilized as a surrogate for oocyte quantity [10].

Various ovarian reserve tests have been used to assess ovarian reserve and predict response to COH [11].

Basal FSH is one of the first parameters evaluated, and its increase is associated with insufficient ovarian response to ovarian stimulation. However, a normal FSH does not exclude inadequate response because its peak occurs relatively in the late phases of diminishing ovarian reserve [5]. Hence, basal FSH cannot be the single definitive test to identify poor responders [12]. E2 is a steroid sex hormone produced by ovarian follicles, the liver, the adrenal cortex, the breast, and adipose tissue [13]. It is often used to monitor ovarian reserve and detect hypoestrogenism and menopause in women with amenorrhea or menstrual dysfunction [14].

AFC and AMH are other ORMs that can be used separately or in tandem to assess ovarian response in ovarian stimulation protocols [15, 16]. AMH is a glycoprotein that belongs to the transforming growth factor-β superfamily and is one of the most important markers for detecting POR. It is exclusively produced by granulosa cells of small and large preantral as well as small antral follicles in women [17]. A decrease in AMH secretion has been identified as the first sign of diminished ovarian reserve [18]. AFC refers to the cumulative number of follicles observed via ultrasonography in both ovaries during the initial phase of the menstrual cycle, specifically days 2–4. Antral follicles are follicles that have a maximum mean diameter of 2–10 mm when measured in the two-dimensional plane [10]. AFC is widely considered to be the most dependable approach for evaluating the ovarian response to ovarian stimulation. Nevertheless, the outcome relies on the operator's expertise and the precision of the ultrasonogram [19].

Tests such as the clomiphene citrate challenge test (CCCT), gonadotropin-releasing hormone (GnRH) agonist, and inhibin B have limited predictive value because they cannot directly quantify ovarian reserve [20]. Given the high cost and inconvenience of frequent hospital visits, these provocative tests (CCCT and GnRH agonist) have been largely abandoned. The basal tests FSH, AFC, and AMH are now the most commonly used markers in clinical practice [21].

Literature reviews fail to predict the accuracy of each ORM in IVF outcoms quantitatively, and the previous systematic reviews and meta-analyses revealed controversial results. It was determined that several ORMs (AMH, FSH, AFC, and E2) were predictive of ovarian response in IVF patients [1]. Despite their noninvasive nature and cost-effectiveness, ORMs were not advised for evaluating the response to the initial IVF cycle with maximum ovarian stimulation [1]. Afterwards, Borer et al. tested the predictive power of AMH and AFC in IVF candidates. Both markers were equally sensitive (82%), but AFC's specificity (82%) was slightly higher than AMHs (76%) [22]. Another systematic review and meta-analysis investigated the potential of ORMs such as AMH, AFC, and FSH to predict ovarian hyperstimulation syndrome (OHSS) in women. By using AMH and AFC along with age, the predictive accuracy for identifying a high response was improved, as seen by an increased area under the receiver operating characteristic (ROC) curve (AUC) compared to using age alone [23].

La Marca et al. summarized AMH and AFC efficacy in detecting poor, normal, and high ovarian responses in patients receiving IVF after ovarian stimulation. They found that AFC and AMH, which are now the most accurate indicators of ovarian reserve, were suitable for creating personalized COH protocols. These sensitive indicators can be employed interchangeably in clinical settings to accurately predict the full range of ovarian response [16]. Also, in 2015, a different meta-analysis that evaluated the capacity of AMH to predict the success of blastocyst implantation and clinical pregnancy in individuals undergoing IVF/ICSI found out that while AMH can be beneficial in fertility consultations, it is unable to predict clinical pregnancy in those with reduced ovarian reserve [24].

However, the inconsistent data have made it unclear which variables are more effective at diagnosing ovarian reserve and predicting IVF success rates. Since the last systematic review in 2015, many studies have been conducted on this topic. Therefore, the goal of this meta-analysis is to enhance our understanding of the most reliable ovarian reserve marker (ORM) test for predicting IVF outcomes in patients with premature ovarian insufficiency (POI).

However, the inconsistent data made it unclear which of these variables is more effective at diagnosing ovarian reserve and forecasting IVF success rates. Many studies have been conducted since the previous systematic review on this topic, which was published in 2015. As a result, the goal of this meta-analysis is to improve our understanding of the best ORM test for predicting IVF outcomes in patients dealing with POI.

Methods

This meta-analysis was designed according to the recommended reporting items for systematic reviews and meta-analysis (PRISMA), statement [25]. The study protocol was registered in the International Prospective register of Systematic Reviews (PROSPERO) database (registration code: CRD42021245380). The PICO (population, intervention, comparison, outcome) framework was utilized to establish research questions and create a search strategy. The purpose of this systematic review and meta-analysis was to examine (C) ORMs and determine which ORM had the highest specificity and sensitivity to ovarian response (O) in IVF/ICSI candidates (P) receiving COH (I).

Search strategy and data gathering

Two researchers (F. S. and M. L.) separately conducted comprehensive searches of the literature using MeSH (Medical Subject Headings) and non-MeSH terms (Supplemental Table 1), in PubMed/MEDLINE, Scopus, and ISI Web of Science databases to find original papers published up to July 2024. After removing duplicates, all of the imported studies into EndNote software (version X7, USA) were thoroughly reviewed. Their bibliographies were then thoroughly examined to verify that no potentially relevant papers were overlooked.

Study design and eligibility criteria

Three authors (F. S., M. L., and S. O.) separately screened titles and abstracts of all the imported articles, selected those eligible for the meta-analysis, and discussed all controversies with the statistician (S. J.). Cohort, case–control, cross-sectional, and interventional studies that measured the diagnostic accuracy of ORMs to predict ovarian response to COH in ART candidates were included. During the screening, animal or in vitro studies, non-English papers, and case reports were dismissed. Following the screening process, we download the full text of the all the papers that were selected for full-text review. For research articles that were published on non-open-access journals, we emailed the corresponding author(s) and requested the full text.

The ovarian reserve was defined as the total ovarian follicular pool including both primordial and growing follicles [3]. Research without a gold standard, true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), sensitivity [TP/(TP + FN), specificity [TN/(FP + FN), and the AUC for a particular ORM were also disqualified, as were publications without the information required to compute the aforementioned parameters. Furthermore, articles that employed ORM for other purposes, such as predicting live birth, cycle cancellation, and the frequency of viable pregnancies rather than ovarian response, were excluded. Articles about women with iatrogenic ovarian insufficiency were eliminated. Also, studies that evaluated the value of less popular tests such as the LH:FSH ratio or the ovarian index, which are not commonly used in clinical practice, were dismissed.

Definitions

Reduced ovarian response (ROR) happens when assisted reproductive techniques (ART) fail to yield a successful fertility [26]. It is characterized by a state of hypergonadotropic hypogonadism associated with oligomenorrhea or amenorrhea [27]. A common procedure used before ART is called controlled ovarian hyperstimulation (COH) which uses exogenous gonadotrophins to stimulate multiple follicular maturation [28]. COH induces the growth of numerous follicles by administering external gonadotropins. At the same time, GnRH agonists or antagonists are given to prevent the release of natural gonadotropins and avoid early ovulation. Ovarian response to COH procedures

Table 1 Characteristics of selected studies for systematic review and meta-analysis

Author, year	Country	Study design	N	Age range	ORMs used	Ovarian response (OR)			Main Results
						Poor (POR)	Normal (NOR)	High (HOR)	
Dermolo, 2024	Ethiopia	Cohort	412	20-47	AMH, AFC	< 3 F	> 3 F	N/A	AMH (cut-off 0.71 ng/mL), outperformed AFC (cut-off 5.5 follicles), to predict POR in COH.
La Marca, 2023	Italy	Cohort	116	20-42	AMH	≤ 3 RO	4 – 15 RO	> 15 RO	AMH levels significantly correlated with the total number of RO after DuoStim, indicating good diagnostic performance in predicting OR. AMH cut-off for POR: 1.1 ng/ml.
Laqqan, 2023	Palestine	Cohort	500	20-39	AMH, AFC, age	FSH > 10 mIU/mL	FSH ≤ 10 mIU/mL	N/A	AMH was the only independent factor significantly related to ROR in women aged less than 40 years undergoing ICSI.
Laqqan, 2022	Palestine	Cohort	462	20-45	AFC, AMH, AFC	AMH <1 ng/ml	AMH = 1.0-3.5 ng/mL	AMH >3.5 ng/mL	AMH and AFC had the highest accuracy in predicting OR.
Sun, 2022	China	Cohort	2585	26-42	AMH, AFC	≤ 3 RO	4 – 15 RO	>15 RO	AMH and AFC were positively correlated with RO, while age and treatment regimen were negatively associated with RO. AMH cut-off for POR and HOR respectively: 2.23 mg/l, 3.56 mg/l. AFC cut-off for POR and HOR respectively: 8.5 and 11.5
Esteves, 2021	Brazil, Turkey and Vietnam	Cohort	9484	22-46	AMH, AFC	< 4 RO	4 – 14 RO	>14 RO	A strong agreement between AFC and AMH in classifying patients into POSEIDON groups. AMH and AFC cut-offs for POR were 1.27 ng/ml and 5 follicles respectively.
Laqqan, 2021	Palestine	CS	318	22-34	AMH, FSH, Age	< 4 RO < 3 F <14mm	4 – 14 RO	>14 RO	AMH was the best predictor of POR and HOR compared to FSH and age.

Table 1 (continued)

Author, year	Country	Study design	N	Age range	ORMs used	Ovarian response (OR)			Main Results
						Poor (POR)	Normal (NOR)	High (HOR)	
Izhar, 2021	Pakistan	Cohort	689	20-35	AFC, AMH	N/A	N/A	OHSS	AMH and AFC were reliable in predicting OHSS in PCOS and non-PCOS women undergoing COH.
Kasapoglu, 2021	Turkey	CS	126	27-36	AFC, AMH, FSH	< 10 RO	≥ 10 RO	N/A	AMH, AFC, and FSH failed to predict POR. Only small (2-5mm) AFC was predictive of POR.
Kheong Lee, 2020	Singapore	Cohort	263	30-38	AFC, AMH, E2	< 4 RO > 2 F >11mm	4 – 19 RO	>19 RO	AMH and AFC levels predicted both POR and HOR in women undergoing IVF.
Neves, 2020	Belgium, Spain, Italy, Germany	Cohort	219	18-43	AMH, FSH, AFC	≤ 3 RO	4 – 14 RO	≥ 15 RO	AMH and AFC were the best predictors of POR and HOR compared to age and FSH.
Peluso, 2021	Brazil	CS	188	27-36	AMH, FSH, AFC	3 F < 14mm	4 - 15 F	>15 F	None of the ORMs individually or combined showed a good predictive capacity for POR.
Baker, 2018	US	CS	160	21-45	AMH, AFC	≤ 4 RO	N/A	N/A	For prediction of POR, AMH significantly outperformed FSH. AMH cut-off level 0.93 ng/ml was a powerful indicator of POR.
Lerman, 2017	Germany, Norway	Cohort	212	18-45	AMH, AFC, FSH, age, cycle length	<6 RO	6 – 18 RO	>18 RO	AMH was the best marker that correlated with RO. AMH cut-off level 0.9 ng/ml, and 2.59 ng/ml, predicted POR and HOR respectively.
Zheng, 2017	China	Cohort	3983	21-46	AMH, FSH, age	≤ 3 RO	5 – 15 RO	> 15 RO	AMH cut-off level 0.8 ng/mL was predictive of POR. Although AMH was an accurate test to anticipate OR and RO following COH, it lacked the predictability for pregnancy results.

Table 1 (continued)

Author, year	Country	Study design	N	Age range	ORMs used	Ovarian response (OR)			Main Results
						Poor (POR)	Normal (NOR)	High (HOR)	
Li, 2016	China	Cohort	615	26-34	AMH, AFC, FSH, age	≤ 5 RO	6 – 15 RO	> 15 RO	AMH was a reliable predictor of POR, NOR and HOR. AMH cutoff 1.1 ng/ml, and 2.6 ng/ml was associated with POR and HOR respectively.
Aghssa, 2015	Iran	Cohort	105	28-37	AMH	≤4 RO	> 4 OR	OHSS	AMH cut-off 1.65 ng/ml was the best predictor of POR.
Aydın, 2015	Turkey	CS	50	24-39	AMH	≤ 4 RO	> 4 RO	N/A	AMH cut-off ≥1.9 ng/ml predicted NOR.
Martinez, 2013	Spain	CS	112	18-35	AMH, AFC, FSH	≤ 5 RO	N/A	N/A	AMH cut-off 2.3 ng/ml was predictive for POR.
Mehrafza, M. 2012	Iran	Cohort	101	21-46	AMH, LH, FSH, E2	≤ 3 RO	4 – 15 RO	≥ 16 RO	AMH (cutoff 0.85 ng/ml) was the best ORM for prediction of POR in comparison with FSH, LH, estradiol, and age.
Jayaprakasan, 2010	UK	CC	135	24-40	AMH, AFC	≤ 3 RO	> 3	N/A	AMH (cutoff 0.99 ng/ml) and AFC (cutoff 10mm) had similar predictive accuracy.
Nardo, 2009	UK	Cohort	165	26-36	AMH, AFC, FSH	> 3 F	N/A	> 20 RO	AMH (cut off: > 3.75 ng/mL and <1.0 ng/mL) outperformed AFC and FSH in predicting HOR.
Riggs, 2008	US	CC	123	24-44	AMH, AFC, Inhibin B	≤ 5 RO	6 – 15 RO	> 15	AMH outperformed FSH, LH, inhibin B, and estradiol in predicting OR to COH.
Tremellen, 2005	Australia	Cohort	238	18-46	AMH	≤ 4 RO	5 – 7 RO	≥ 8 RO	AMH cutoff 8.1 pmol/L, was better than FSH to estimate OR; but, to achieve the highest prognostic accuracy, it should be used in conjunction with FSH and E2.

Table 1 (continued)

Author, year	Country	Study design	N	Age range	ORMs used	Ovarian response (OR)			Main Results
						Poor (POR)	Normal (NOR)	High (HOR)	
Muttukrishna, 2005	UK	Cohort	108	21-46	AMH, AFC, Inhibin B	≤ 4 RO	≥ 5 RO	N/A	Inhibin B and AFC were the best ORMs for prediction of NOR. AMH (cutoff < 0.2 ng/ml) was the best predictor of POR. AFC showed the highest predictive accuracy to predict clinical pregnancy.
Durmusoglu, 2004	Turkey	CC	82	23-38	Day 7 AFC	≤ 3 RO	> 3 RO	N/A	Cycle day 7 AFC had the highest predictive value for POR.

AFC Antral follicle count, *AMH* Anti-Müllerian hormone, *ART* Assisted reproductive technology, *CC* Case control, *COH* Controlled ovarian hyperstimulation, *CS* Cross sectional, *DuoStim* Double stimulation, *E2* Estradiol, *FS* Follicle size, *FSH* Follicle stimulating hormone, *ICSI* Intracytoplasmic sperm injection, *IVF* In-vitro fertilization, *LH* Luteinizing hormone, *N/A* Not assessed, *ng* nanogram, *NOR* Normal ovarian response, *N* Number of cases, *NPV* Negative predictive value, *OHSS* Ovarian hyper-stimulation syndrome, *OR* Ovarian response, *ORM* Ovarian reserve marker, *PCOS* Polycystic ovarian syndrome, *pmol* Pico mol, *POSEIDON* Patient-Oriented Strategies Encompassing IndividualizeD Oocyte Number, *ROR* Reduced ovarian reserve, *RO* Retrieved oocytes, *UK* United Kingdom, *US* United States

is graded as normal, low, or high based on the number of recovered oocytes [29]. IVF oocyte retrieval rates are typically used to categorize ovarian response to COH regimens. The extraction of 8 to 15 oocytes is indicative of a normal ovarian response. To minimize the danger of OHSS and maximize the odds of successful fertilization and embryo development, this range is considered optimum (1). In most cases, a low ovarian response is indicated when fewer than five oocytes are recovered. People in this category, who are commonly known as "poor responders," may have a reduced ovarian reserve (2). Retrieving more than 15 oocytes is usually indicative of a high ovarian response, which increases the risk of OHSS, and needs to be managed carefully (3).

Data extraction and excluded literature

Three authors (F. S., M. L., and A. K.) extracted the required data from eligible articles separately, including first author and publication date, country of research, study design, sample size, population studied, age range, and ovarian reserve markers used. Two-by-two tables were also used to calculate sensitivity, specificity, TP, TN, FP, FN, AUC, and the optimal cut-off point for each ORM. The required data were estimated using the ROC curve for papers that did not include information on sensitivity or specificity in their text. The studies that reported their data per cycle instead of per patient were excluded unless cycle numbers matched the number of participants. The Microsoft Excel 2019 software was used for data extraction.

Quality assessment

For the quality assessment of the included studies, the QUADAS-2 tool, a widely recognized framework specifically designed to evaluate the methodological quality of diagnostic accuracy studies, was employed [30]. The QUADAS-2 tool examines four key domains: patient selection, index test, reference standard, and flow and timing, each evaluated for potential risk of bias and concerns regarding applicability of study's findings to the systematic review question or broader clinical settings and populations. The patient selection domain assesses how participants were chosen, focusing on the appropriateness of inclusion criteria and the risk of selection bias. The index test domain evaluates how the diagnostic test was conducted and interpreted, including the use of blinding and whether pre-specified thresholds were applied. The reference standard domain examines the objectivity and consistency of the outcome measure used to classify the target condition. The flow and timing domain reviews the sequence of participant assessments, including the timing between the index test and reference standard and any exclusions that might impact the study results.

Statistical analysis

To assess each ORM, the "metafor" R package of the R program (R-4.4.0 for macOS) was employed for meta-analysis. The logarithm of diagnostic odds ratio (Ln DOR) with 95% confidence intervals for each ORM was measured by a random-effects model. DOR evaluates that how much the odds of having a poor or high ovarian response increase when each ORM is positive. The Ln DOR was utilized to enhance normality and variance stabilization and facilitate clearer interpretation of the pooled results on the original DOR scale. Two separate meta-analyses were performed to evaluate each ORM's efficacy in predicting poor and high ovarian responses. The Ln DOR combines sensitivity and specificity to evaluate the diagnostic efficacy of each ORM. The DOR compares the odds of a poor or high response. An Ln DOR greater than 1 indicates the test can discriminate between cases, with higher values reflecting better accuracy. An Ln DOR of 1 suggests no discriminatory ability of the ORM. Together with AUC, Ln DOR provides a fuller view of any of the ORMs' performance, with an Ln DOR of 3.0, signaling strong diagnostic potential. The meta-analysis included forest plots that visually displayed the Ln DOR for ORMs in predicting poor and high ovarian response. These plots illustrated the differences in diagnostic accuracy observed across the selected studies.

Furthermore, this meta-analysis aggregated the findings of several studies that evaluated the diagnostic accuracy of ORMs using receiver operating characteristic (ROC) curve analysis. These curves were used to plot the true positive rate (sensitivity) against the false-positive rate (1-specificity) for each study, providing a visual summary of the diagnostic accuracy across all included studies. We calculated the average area under the ROC curve (AUC) that measures how well any ORM diagnostic test distinguishes between poor and high responders. It comes from the ROC curve, which plots sensitivity versus 1-specificity at different thresholds. An AUC of 1.0 means perfect discrimination between cases, while 0.9–1.0 indicates excellent accuracy. An AUC of 0.8–0.9 shows good reliability with some overlap, 0.7–0.8 is acceptable but with limitations, and 0.5 means the test is no better than chance (similar to an $Ln\ DOR = 1$). Clinically, these AUC values help determine the effectiveness of ORMs with an AUC of 0.85, indicating good utility for patient outcomes. Also, we calculated the degree of heterogeneity between studies, using metrics such as Tau,

Tau-squared, I-squared (I^2), and Cochran's Q statistic. The Begg's funnel plot was used in this meta-analysis to visually identify publication bias.

Subgroup analyses were performed in both poor and high response category based on study design, study country, and sample size to find out whether any of these variables account for potential differences in diagnostic accuracy. These subgroup analyses helped to identify any variations in diagnostic performance across different study characteristics. For the subgroup meta-analysis of sample size, we used the median sample size of the included studies as a cut-off point to categorize studies into two groups: "low" and "high" sample size. This data-driven dichotomization ensures that the studies are evenly distributed across the two categories, allowing us to analyze the impact of sample size on the diagnostic accuracy of ovarian reserve markers (ORMs). Studies with sample sizes below the median value were classified as "low sample size," while those above the median were categorized as "high sample size." We performed mixed-effects model meta-analysis separately for both groups to assess whether sample size influenced the diagnostic outcomes of ORMs in predicting ovarian response to COH. The median value was selected based on the distribution of sample sizes across the studies, ensuring an objective and balanced categorization. This approach avoided any potential biases that may occur using arbitrarily defined thresholds.

Results

After evaluating the titles and abstracts of all the imported articles, 65 studies were selected for full-text review and data extraction. Then, 39 papers were dismissed due to the following reasons: 16 studies lacked the required data for meta-analysis [31–46], 6 articles assessed the efficacy of uncommon ORMs (ovarian index, repeated antral follicular counts, elevated day 3 FSH:LH ratio, repeated clomiphene citrate challenge test) [47–52], 4 papers assessed POI patients in the absence of COH setting [53–56] and 7 publications examined the accuracy of ORM in the diagnosis of other outcomes such as pregnancy and OHSS without evaluating the ovarian response [55, 57–62], 3 studies considered the number of cycles instead of patients in assessing ORMs [63–65], and 3 papers only assessed the ORMs' predictability for normal response instead of high or low response [66–68].

Eventually, 26 relevant articles with 21,584 participants were selected for this systematic review and meta-analysis (Fig. 1). From these, 13 studies evaluated ORMs' accuracy in predicting poor and high ovarian response [69–81] following COH, and one article only assessed high response [82]. We found 12 studies that estimated the predictive accuracy of ORMs during COH only in women with poor ovarian response [83–94]. Only one study calculated its data per treatment cycle instead of the per participant [84]; however, it attributed cycles to each participant since total cycles and participants almost matched.

Study characteristics

Selected studies for this systematic review and meta-analysis were published from 2005 until 2024. This meta-analysis evaluated 4 case controls [75, 76, 84, 85], 17 cohorts [69–71, 74, 78–82, 87–94], and 5 cross-sectional studies [72, 73, 77, 83, 86] (Table 1).

Assessment of risk of *bias* in eligible studies

The studies reviewed using the QUADAS-2 framework reveal varying levels of bias across different domains, particularly in patient selection and the index test interpretation. While many studies demonstrated low risk of bias in terms of reference standards and flow and timing, a significant number had moderate to high risk due to issues like unclear selection criteria and lack of blinding in test interpretation. However, the majority of studies showed low applicability concerns, indicating that their findings are relevant to clinical practice. Despite some methodological limitations, the overall body of research provides a credible foundation for understanding diagnostic tests in predicting ovarian response and OHSS, with caution needed when generalizing results to broader populations (Table 2).

Meta-analysis

In women characterized by either poor or high response to ovarian stimulation, both AMH and AFC demonstrated superior diagnostic performance compared to FSH and E2 though AFC appeared slightly more effective. Similarly, in high response category, in contrast, FSH and E2 exhibited lower diagnostic accuracy (Table 3). According to the forest plots of ORMs in both response categories, both AMH and AFC demonstrated narrower confidence intervals and more consistent effect sizes compared to FSH and E2. In the high response group, particularly, AFC showed marginally higher diagnostic performance than AMH (Figs. 2 and 3). The ROC curve analysis further substantiated these findings; both AMH and AFC demonstrated higher area under the curve (AUC) values, indicating superior diagnostic accuracy. However, slight variations were observed between the two markers, particularly in the high response group, where AFC demonstrated marginally higher performance (Fig. 4).

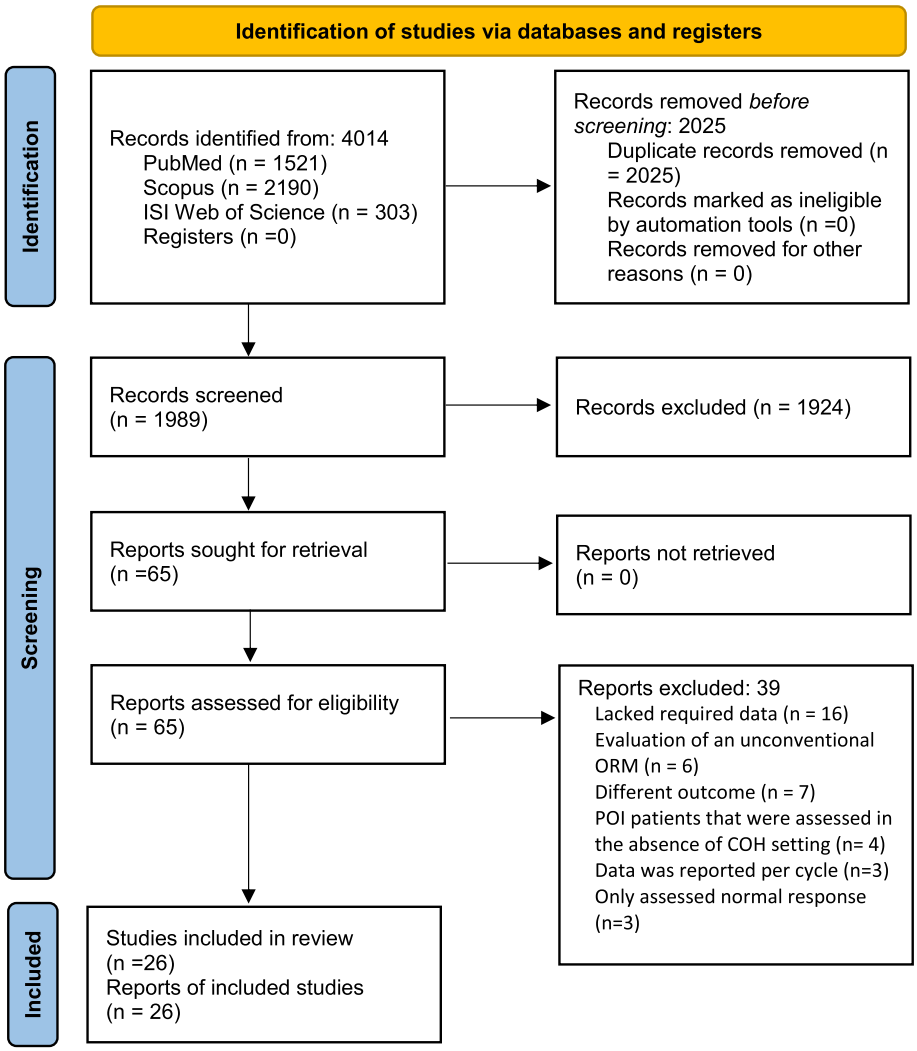


Fig. 1 Flowchart of the study selection process

The analysis of heterogeneity within the poor response group revealed substantial variability across studies, as indicated by a significant Cochran's Q statistic and its associated *p*-values. High I^2 percentages further underscored the heterogeneity, suggesting that differences in study outcomes might influence the meta-analysis results. The T^2 and T-values additionally highlighted that this variability was not merely due to random error but pointed to actual differences in effect sizes across studies. Unlike the poor response group, the high response group showed lower levels of variability among the studies, reflected by reduced I^2 values. The Cochran's Q statistic supported the consistency of the results across studies, suggesting robustness in the findings for this population (Table 3). Funnel plots also illustrate the potential for publication bias across the included studies in both response categories (Fig. 5).

Subgroup *meta*-analysis

Study design

The subgroup analysis for the poor response category demonstrated notable differences between cohort and cross-sectional study designs. In cohort studies, both AMH and AFC markers exhibited strong diagnostic performance compared to FSH and E2. This indicates that cohort studies, which track outcomes over time, may provide more reliable indicators of ovarian reserve. Conversely, the cross-sectional subgroup showed a reduced performance of these markers, especially for FSH, suggesting that the nature of cross-sectional studies might not capture the dynamic aspects of ovarian reserve as effectively. Among the ORMs in the poor response category, AMH's diagnostic accuracy was most influenced by study design, showing greater reliability in cohort studies than in cross-sectional

Table 2 Risk of bias and applicability concerns for included studies based on the QUADAS-2 assessment form

Author, year	Risk of bias*				Concerns regarding applicability**		
	Patient selection	Index text	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Case control							
Laqqan, 2021	Moderate	Unclear	Low	Unclear	Low	Low	Low
Jayaprakasan, 2008	Low	Low	Low	Low	Low	Low	Low
Riggs, 2008	Moderate	Low	Low	Low	Low	Low	Low
Durmusoglu, 2004	Moderate	Low	Low	Moderate	Low	Low	Low
Cross sectional							
Peluso, 2020	Low	Low	Low	Low	Low	Low	Low
Aghssa, 2015	Low	Low	Low	Moderate	Low	Low	Low
Aydin, 2015	Moderate	Moderate	Low	Low	Low	Low	Low
Martı´nez, 2013	Low	Low	Low	Low	Low	Low	Low
Mehraffza, 2012	Moderate	Moderate	Low	Moderate	Low	Low	Low
Baker, 2018	Low	Low	Low	Low	Low	Low	Low
Cohort							
Dermolo, 2024	High	Unclear	Unclear	Low	Low	Low	Low
La Marca, 2023	High	Moderate	Moderate	Low	Low	Low	Low
Laqqan, 2023	Low	Low	Low	Low	Low	Low	Low
Laqqan, 2022	Low	Low	Low	Low	Low	Low	Low
Sun, 2022	Moderate	Unclear	Low	Unclear	Low	Low	Low
Esteves, 2021	Low	Low	Low	Low	Low	Low	Low
Kasapoglu, 2021	Low	Unclear	Low	Unclear	Low	Low	Low
Izhar, 2021	Low	Low	Low	Low	Low	Low	Low
Neves, 2020	Moderate	Low	Low	Low	Low	Low	Low
Lee, 2020	Low	Low	Low	Low	Low	Low	Low
Zheng, 2017	Low	Low	Low	Low	Low	Low	Low
Lerman, 2017	Moderate	Low	Low	Low	Low	Low	Low
Li, 2016	Low	Low	Low	Low	Low	Low	Low
Nardo, 2009	Low	Low	Low	Low	Low	Low	Low
Tremellen, 2005	Moderate	Low	Low	Low	Low	Low	Low
Muttukrishna, 2005	Moderate	Low	Low	Moderate	Low	Low	Low

*Risk of bias refers to the likelihood that the design, conduct, or analysis of a study introduces systematic errors, leading to inaccurate or misleading results. In the QUADAS-2 framework, the risk of bias is evaluated across four domains (patient selection, index text, reference standard, flow and timing) each addressing specific potential biases in the study

**Concerns regarding applicability address the extent to which the study findings are relevant to the systematic review question and can be generalized to other clinical settings or populations. These concerns are also evaluated across the same four domains

analyses. In the high response category, cohort studies again resulted in higher diagnostic accuracy for both AMH and AFC compared to cross-sectional studies, with AFC performing particularly well in the cohort subgroup. Therefore, study design significantly impacts diagnostic outcomes, with AMH's accuracy being notably affected, as it was more consistent in cohort studies (Table 4).

Study country

Due to most subgroups containing only a single study, it was not possible to draw definitive conclusions from the subgroup meta-analysis based on the study country. Therefore, only countries that evaluated the diagnostic accuracy of ORMs to predict either poor or high ovarian response at least twice were included in this subgroup meta-analysis. The goal of this analysis was to determine whether the high rate of heterogeneity among studies was attributable to the study country.

In the poor response group, the USA conducted two studies on AMH, which showed statistically nonsignificant results (*Ln DOR*: 1.95, $p=0.13$) and high heterogeneity (I^2: 92.31%, $p<0.05$). China assessed AMH in three studies and found significant diagnostic accuracy (*Ln*

Table 3 Diagnostic accuracy of ORMs for prediction of ovarian response, and between study heterogeneity for each ORM

			Diagnostic Accuracy				Heterogeneity			
	Number of studies	Number of participants	AUC (pAUC*)	Ln DOR (95%CI)	SE	*P* value#	I^2(%)	T	T^2 (SE)	Cochran's Q (*p*-value)
Poor Response										
AMH	22	18745	0.828 (0.738)	2.37 (1.84, 2.89)	0.26	< 0.0001	93.29	1.12	1.26 (0.47)	178.09 (< 0.0001)
AFC	16	15065	0.853 (0.787)	2.68 (1.90, 3.45)	0.39	< 0.0001	95.65	1.50	2.25 (0.91)	189.65 (< 0.0001)
FSH	10	4079	0.633 (0.595)	1.13 (0.64, 1.61)	0.24	<0.0001	70.20	0.60	0.36 (0.27	34.29 (< 0.0001)
E2	4	495	0.606 (0.593)	0.80 (-0.21, 1.83)	0.52	0.12	65.57	0.81	0.66 (0.88)	9.83 (0.0200)
High Response										
AMH	14	9152	0.839 (0.747)	2.48 (2.00, 2.96)	0.24	<0.0001	90.60	0.78	0.61 (0.31)	75.97 (< 0.0001)
AFC	8	4217	0.898 (0.854)	2.76 (1.57, 3.95)	0.60	<0.0001	92.71	1.50	2.27 (1.47)	41.01 (< 0.0001)
FSH	7	3726	0.702 (0.689)	1.16 (0.42, 1.89)	0.37	0.0019	84.75	0.87	0.77 (0.56)	24.02 (0.0005)
E2	3	440	0.578 (0.569)	0.39 (-0.07, 0.87)	0.24	0.0991	0.00	0	9 (0.18)	1.31 (0.5174)

Abbreviations: *AFC* Antral follicular count, *AMH* Anti, Mullerian hormone, *AUC* Area under curve, *CI* Confidence interval, *DOR* Diagnostic odds ratio, *E2* Estradiol, *FSH* Follicle stimulating hormone, *Ln* Logarithm, *ORM* Ovarian reserve marker, *pAUC* Partial AUC, *SE* Standard error, *T* Tau, T^2 Tau squared

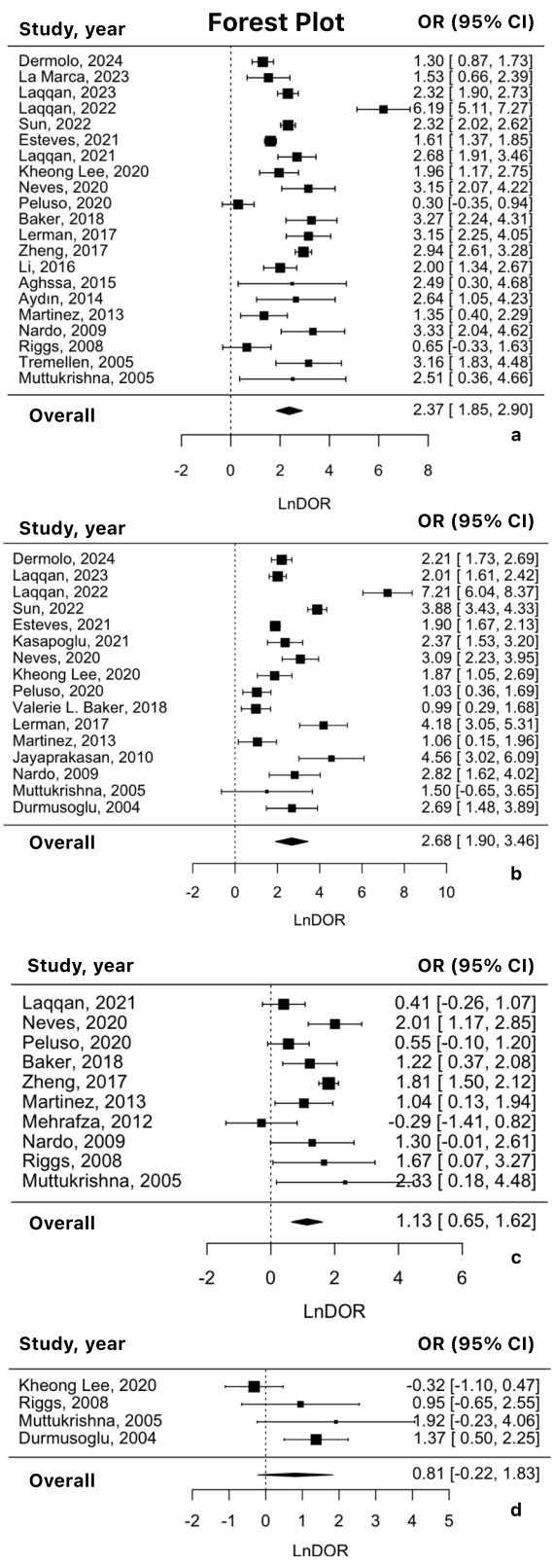
* Restricted to observed false positive rates and normalized

P-value for Ln DOR

DOR: 2.47, $p<0.05$) despite considerable heterogeneity (I^2: 80.79%, $p<0.05$), suggesting variability in study results. Similarly, Palestine evaluated AMH accuracy for poor response prediction, demonstrating meaningful accuracy (*Ln DOR*: 3.69, $p<0.05$) while displaying high heterogeneity (I^2: 96.95%, $p<0.05$). The UK measured AMH predictability for poor response, with AFC showing meaningful overall diagnostic accuracy (*Ln DOR*: 3.07, $p<0.05$) and no considerable heterogeneity (I^2: 67.43%, $p<0.05$). In the high response category, China assessed AMH in three studies, yielding significant findings (*Ln DOR*: 1.71, $p<0.05$) and high between-study heterogeneity (I^2: 84.75%, $p<0.05$). However, according to two studies conducted in Iran, AMH had significant diagnostic precision for high response prediction (*Ln DOR*: 2.43, $p<0.05$) with low variability among studies (I^2: 0.00%, $p=0.47$). These findings suggest that country-specific factors, including population characteristics and study methodologies, can contribute to the heterogeneity observed in the meta-analysis.

Sample size

In the subgroup meta-analysis based on sample size, studies were categorized as small or large sample size based on the median sample size of all included studies (Table 5). Studies with fewer participants than the median were categorized as "small sample size," while those with more participants were classified as "large sample size." This data-driven approach ensured a balanced comparison between groups. In the poor ovarian response category, studies with "high" sample sizes demonstrated significant diagnostic accuracy for AMH, AFC, and FSH, with strong statistical significance ($p<0.0001$). In contrast, studies with small sample sizes showed less consistent results, particularly affecting the diagnostic accuracy of FSH, suggesting that smaller sample sizes may introduce variability in the outcomes. In the high ovarian response category, AMH and AFC retained significant diagnostic performance across both small and large sample size studies, but the results were more consistent in the "large" sample size group ($p<0.0001$). Conversely, small studies indicate a lack of data for AMH and AFC and a negative influence on FSH results. This suggests that in the poor response category, smaller studies may lead to less reliable conclusions, particularly affecting the outcomes for FSH. For the high response category, AMH retains its significance in large studies ($p<0.0001$) but shows a substantial drop in small studies, signifying the impact of sample size. AFC's results are robust in large studies but not reported for small studies, again indicating sample size influence. FSH in large studies shows a non-significant trend, further compromised in small studies through negative results, suggesting less reliable outputs. Across both response categories, AMH and AFC are notably affected by sample size, with larger studies providing more consistent conclusions. Similarly, FSH shows variability across both large and small studies, indicating that smaller sample sizes might lead to inconclusive or skewed data. Overall, this subgroup analysis highlights the importance of sufficient sample size in obtaining reliable meta-analysis outcomes, emphasizing that smaller studies can contribute to misleading heterogeneity and less definitive results (Table 5).



◀ **Fig. 2** Forest plot representing the meta-analysis of the Ln DOR of AMH (**a**), AFC (**b**), FSH (**c**), and estradiol (**d**) for predicting poor ovarian response after COH. Ln DOR, logarithm of diagnostic odds ratio; OR, odds ratio; CI, confidence interval

Discussion

Through this meta-analysis of the most frequently employed ovarian reserve markers (ORMs), we identified the optimal parameters for accurately determining whether the oocyte yield was inadequate or excessive. In summary, AFC emerged as marginally superior to AMH as the best indicator of ovarian response to COH in both groups. Both FSH and estradiol showed almost similar ability to identify cases of poor and excessive ovarian response, with FSH slightly outperforming estradiol in both groups.

Consistent with our research, Rosen et al. observed that among many indicators including AMH, AFC, FSH, and E2, only AMH and AFC could accurately characterize the histological pattern of diminished ovarian reserve. Although AMH was more economical, AFC was more accurate and less invasive [95]. Studies that examined multiple ovarian reserve indicators and POI progression found that AMH has the highest diagnostic sensitivity for predicting the severity of POI. Furthermore, the combination of AMH and AFC proved to be the most reliable method for identifying POI in its early stages [96]. According to a recent analysis, AMH is one of the greatest functional ovarian reserve measures since it shows the number of developing follicles capable of ovulation. Furthermore, it can predict poor or high response in COH candidates and indicate menopausal state [97].

However, further research is required to determine the internal and environmental elements that interact with the synthesis and physiology of AMH, in order to enhance its effectiveness in therapeutic applications [97]. While age affects IVF outcomes in terms of ovarian response to GnRH protocols, AMH and AFC provide a more precise assessment of oocyte retrieval in COH. After evaluating the interaction of AFC and AMH, Keane et al. reported that both AFC and AMH demonstrated a positive relationship with oocyte number and, when combined, offered the highest accuracy for predicting ovarian response in IVF patients [98].

AMH is a member of the transforming growth factor β (TGF-β) family that acts via a specific receptor called AMH receptor type II [99]. During human embryogenesis, AMH signaling pathways play an important role in the maturation of GnRH neurons, as their abnormalities may lead to infertility later in life [100]. Anti-Müllerian hormone (AMH) is specifically released by

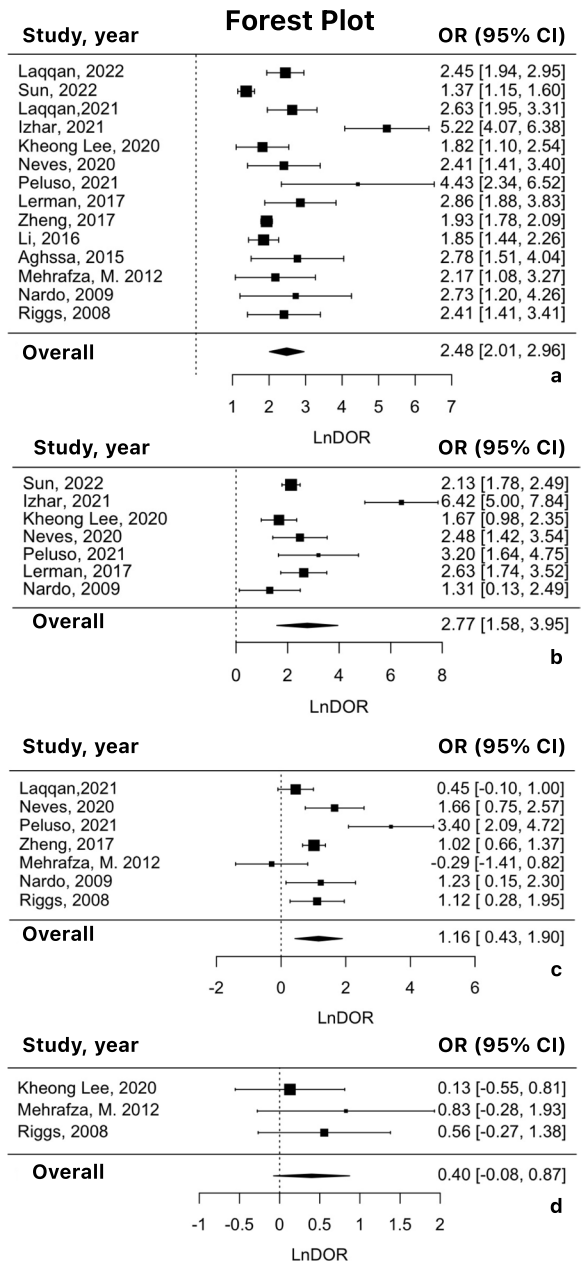


Fig. 3 Forest plot representing the meta-analysis of the Ln DOR of AMH (**a**), AFC (**b**), FSH (**c**), and estradiol (**d**) for predicting high ovarian response after COH. Ln DOR, logarithm of diagnostic odds ratio; OR, odds ratio; CI, confidence interval

developing follicles and can be used to accurately determine if the ovarian reserve is depleted. This hormone remains unaffected by the hypothalamic–pituitary–gonadal axis and has only minimal and insignificant changes during each ovarian cycle. During menopause, serum AMH levels progressively decline until they are undetectable [101]. AMH levels gradually rise as dormant follicles mature into growing follicles, eventually reaching the preantral and antral stages. Antral and preantral follicles secrete AMH, predominantly hindering the maturation of primordial follicles [101]. AMH also inhibits intraovarian pathways implicated in follicular atresia, as seen by its absence in follicles undergoing degeneration [102]. Hayes et al. found that while in vivo AMH inhibits follicular growth and ovulation in mouse models, it also prevents follicular degeneration by coordinating FSH function and follicular development through miR-181a and miR-181b miRNAs. Overall, AMH might have therapeutic benefits in mouse models through enhancing oocyte retrieval [103].

FSH has traditionally been the ovarian reserve biomarker of choice. Since the late 1980s, it has indicated hypothalamic-pituitary–gonadal axis functioning. The World Health Organization classified ovarian dysfunction based on serum FSH and estradiol levels [104]. Follicular impairment due to FSH receptor failure and steroid-cell autoantibodies produce gradual follicular pool degeneration in POI, which justifies the use of FSH in identifying these individuals [105]. However, FSH on day 3 of the menstrual cycle exhibited both inter- and intra-cyclic variations, which questioned its status as the preferred ORM, thus shifting the focus to other tests [106]. Similar to this study, all the mentioned AMH physiology and its association with antral follicles make AMH and AFC more eligible ORMs than FSH to predict the outcome in patients suffering from PCOS, POI, and other fertility-related disorders [107].

To our knowledge, this is the first systematic review and meta-analysis that has assessed all available information on ovarian reserve markers (ORMs) and how well they predict ovarian response in IVF patients. One of the primary strengths is the robust comparison of multiple markers across distinct subgroups classified by ovarian response, study design, and country, providing a comprehensive view of diagnostic efficacy. Much of the prior literature has concentrated on how ORMs relate to specific IVF outcomes. This work adds a new dimension to clinical assessment tools by quantifying the diagnostic performance of these indicators in predicting ovarian response. Clinicians can optimize medication dosages and minimize risks like ovarian hyperstimulation syndrome (OHSS) by tailoring COH procedures to individual patients when they know which ORMs (e.g., AMH, AFC) are the best predictors of ovarian response. By early intervention and consideration of alternate treatments like oocyte donation or modified stimulation protocols, patients whose ovarian responses are expected to be poor can be identified. By decreasing the need for many cycles and the costs that come with them, accurate prediction of ovarian response can lead to more efficient use of medical resources. In addition, reliable ovarian

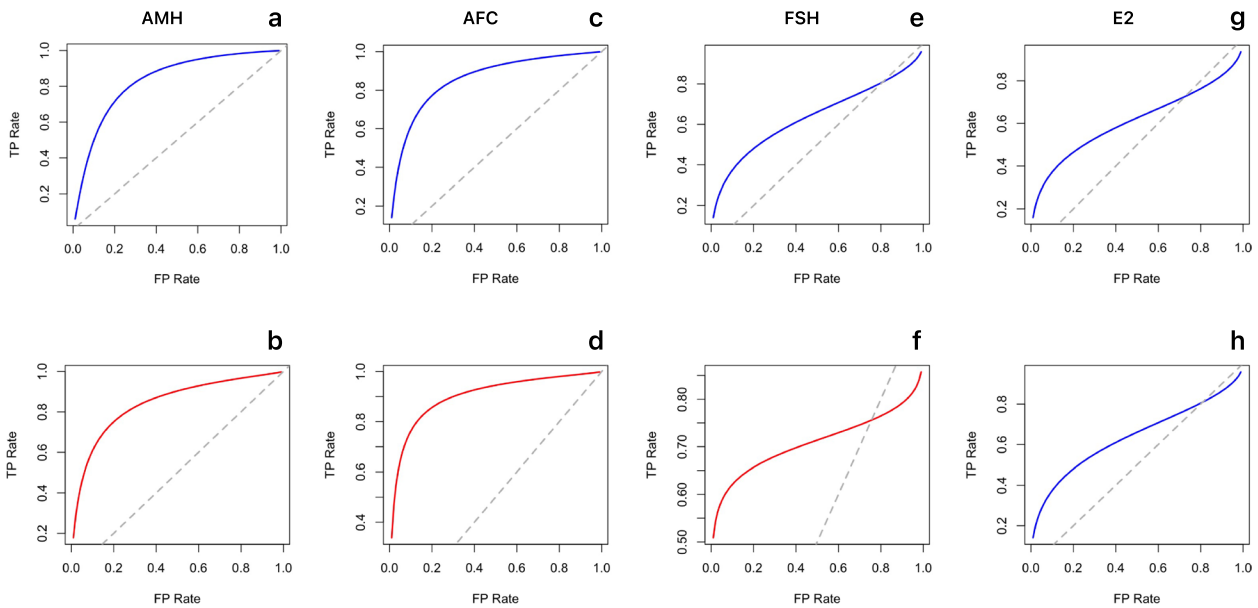


Fig. 4 Summary of the receiver operating characteristic (ROC) curve for AMH (**a**, **b**), AFC (**c**, **d**), FSH (**e**, **f**), and estradiol (**g**, **h**)

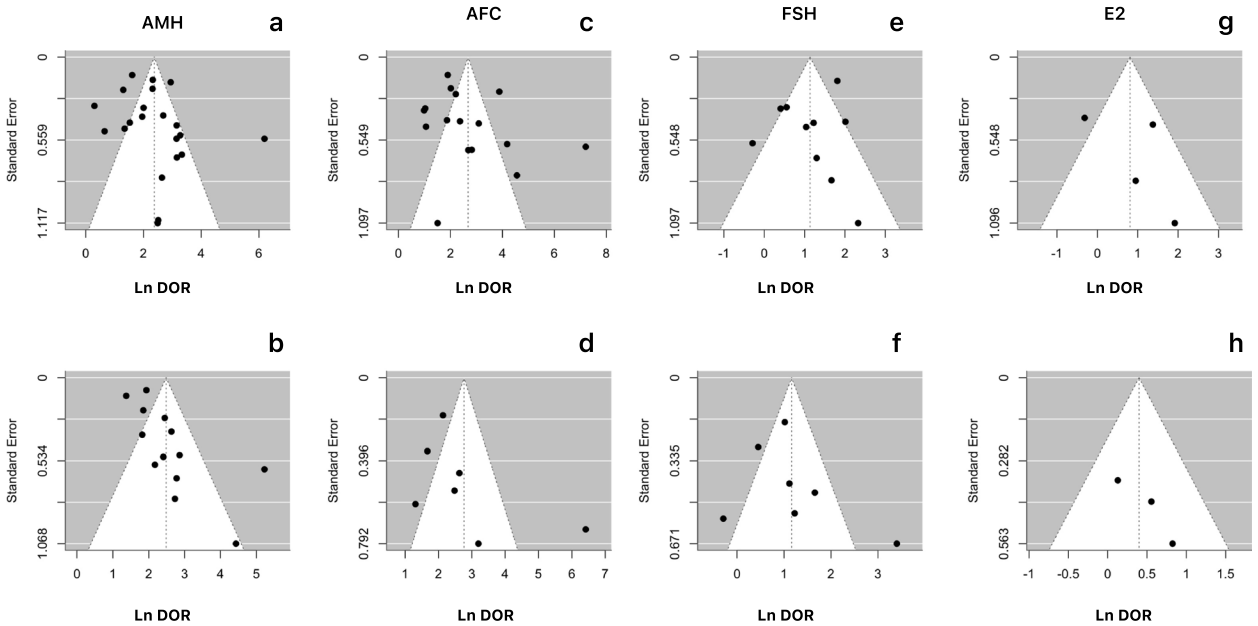


Fig. 5 Funnel plot showing the publication bias for AMH (**a**, **b**), AFC (**c**, **d**), FSH (**e**, **f**), and estradiol (**g**, **h**)

response prediction helps control patient expectations and decreases the emotional and financial strain associated with IVF therapy.

Since 2006, comprehensive reviews and meta-analyses have been performed to assess the efficacy of several ORMs in predicting ovarian response to hyperstimulation [1, 16, 22–24]. Incorporating the most current studies that analyzed multiple ORMs in IVF candidates strengthened this hypothesis. There were, however, certain caveats to our research. Due to a lack of relevant studies, this meta-analysis did not include other ovarian reserve tests including LH and inhibin B. Depending on the study, it may have been necessary to manually compute the sensitivity and specificity of an ovarian reserve marker using the ROC curve. Selected studies used a

Table 4 Subgroup meta-analysis results based on study design

	Number of studies	Number of participants	Ln DOR (95%CI)	SE	*P*-value*	I^2(%)	*T*	T^2 (SE)	Cochran's Q (*P*-value)
Poor response									
Cohort									
AMH	15	17772	2.60 (2.01, 3.20)	0.30	< 0.0001	93.86	1.07	1.14 (0.51)	128.55 (< 0.0001)
AFC	10	14240	2.61 (1.86, 3.35)	0.37	< 0.0001	74.65	0.75	0.57 (0.53)	18.05 (0.00)
FSH	5	3315	1.40 (0.52, 2.28)	0.44	<0.0001	77.22	0.82	0.68 (0.70)	13.95 (0.00)
E2	2	290	0.56 (-1.57, 2.70)	1.09	0.60	72.67	1.34	1.81 (3.52)	3.65 (0.05)
Cross Sectional									
AMH	5	723	1.99 (0.88, 3.10)	0.56	0.0004	86.20	1.15	1.33 (1.13)	35.06 (< 0.0001)
AFC	4	584	1.01 (0.59, 1.44)	0.21	<0.0001	0.00	0	0 (0.14)	0.01 (0.99)
FSH	4	679	0.71 (0.34, 1.08)	0.19	0.0002	0.00	0	0 (0.11)	2.91 (0.40)
High response									
Cohort									
AMH	11	8639	2.40 (1.84, 2.96)	0.28	< 0.0001	93.02	0.84	0.70 (0.39)	64.75 (< 0.0001)
AFC	7	4083	2.71 (1.32, 4.09)	0.70	< 0.0001	94.53	1.65	2.74 (1.89)	39.61 (< 0.0001)
FSH	3	3112	0.89 (-0.25, 2.04)	0.585	0.1262	73.05	0.86	0.74 (1.02)	7.30 (0.02)
Cross Sectional									
AMH	2	401	3.24 (1.57, 4.92)	0.85	0.00	60.99	0.99	0.98 (2.28)	2.56 (0.10)
FSH	2	235	1.86 (-1.02, 4.75)	1.47	0.20	93.92	2.02	4.08 (6.15)	16.4567 (< 0.0001)

Abbreviations: *AFC* Antral follicular count, *AMH* Anti, Mullerian hormone, *CI* Confidence interval, *DOR* Diagnostic odds ratio, *AUC* Area under curve, *E2* Estradiol, *FSH* Follicle stimulating hormone, *Ln* Logarithm, *ORM* Ovarian reserve marker, *pAUC* Partial AUC, *SE* Standard error, *T* Tau, T^2 Tau squared

* *P*-value for Ln DOR

variety of AMH tests, which could increase the risk of bias in interpreting results. Also, disparities in research methods, sample populations, and assessment tools might account for this variability. Only studies published in English were included, which may introduce language bias and exclude significant research published in other languages. Some studies did not report sensitivity or specificity, necessitating the use of ROC curves for approximation, which may include estimation errors. The funnel plot analysis revealed potential publication bias, implying that papers yielding positive outcomes may be overrepresented. Furthermore, differences in ORM definitions and measures between studies may impact the accuracy and generalizability of the pooled values. The significant heterogeneity observed, especially in the poor response group, indicates variability in study outcomes that could affect the overall conclusions. Additionally, subgroup analyses based on study design and country revealed that diagnostic performance could be heavily influenced by these factors, with cohort studies generally providing more reliable indicators than cross-sectional studies. The potential for publication bias, as hinted at by funnel plots, also poses a limitation, possibly skewing the results towards more favorable outcomes reported in published studies. To address these limitations and build on the current findings, future studies should aim to standardize methodologies and reporting to reduce heterogeneity. Larger sample sizes in studies are crucial to provide more precise effect estimates and reduce the impact of random variations. Additionally, longitudinal cohort studies should be prioritized over cross-sectional designs to capture the dynamic aspects of ovarian reserve more accurately. Collaborative efforts across different countries can help mitigate the impact of country-specific factors and contribute to more generalizable conclusions. Finally, robust strategies to minimize publication bias, such as pre-registering study protocols and including unpublished data, can enhance the reliability and validity of future meta-analyses.

Conclusions

In summary, this systematic review and meta-analysis provide valuable insights into enhancing clinical management and evidence-based practices in IVF. AMH and AFC show the highest predictive accuracy for ovarian response in patients with low ovarian reserve, outperforming other markers. These insights can help optimize COH protocols and be beneficial for POI patients considering oocyte cryopreservation. Additionally, investigating

Table 5 Subgroup meta-analysis results based on sample size

	Median sample size *	Number of studies		Number of participants		Ln DOR (95%CI)		SE		*P*-value**		I^2(%)	T	T^2 (SE)	Cochran's Q (*P*-value)
		Large	Small	Large	Small	Large	Small	Large	Small	Large	Small				
Poor response															
AMH	219	8	14	16873	1872	2.58 (1.84, 3.31)	0.43 (-1.50, 0.63)	0.37	0.54	< 0.0001	0.424	93.44	1.13	1.29 (0.49)	174.08 (< 0.0001)
AFC	200	6	10	13656	1409	3.22 (2.19, 4.25)	-1.14 (-2.63, 0.34)	0.52	0.76	< 0.0001	0.13	95.15	1.43	2.06 (0.87)	175.07 (< 0.0001
FSH	162.5	4	6	3404	675	1.21 (0.55, 1.87)	-0.19 (-1.22, 0.83)	0.33	0.52	0.0003	0.7	77.91	0.64	0.41 (0.32)	31.33 (0.0001)
E2		2	2	333	162	0.01 (-0.86, 0.88)	1.46 (0.15, 2.78)	0.44	0.66	0.98	0.02	17.12	0.32	0.10 (0.62)	2.13 (0.34)
High response															
AMH	241	7	7	8219	983	2.33 (1.70, 2.96)	-0.37 (-0.61, 1.35)	0.32	0.50	< 0.0001	0.45	91.11	0.79	0.63 (0.33)	65.20 (<0 .0001)
AFC	219	4	4	3552	665	3.26 (1.34, 5.17)	-0.86 (-3.43, 1.69)	0.97	0.30	0.0008	0.50	92.97	0.62	2.65 (1.86)	40.85 (< 0.0001)
FSH	188	2	5	3060	666	1.01 (-0.17, 2.21)	-0.29 (-1.34, 1.93)	0.60	0.83	0.09	0.72	87.55	1.00	01.01 (0.77)	23.26 (0.0003)
E2	123	1	2	227	213	0.12 (-0.55, 0.81)	0.52 (-0.42, 1.47)	0.34	0.48	0.71	2.27	0.00	0.00	0 (0.34)	0.14 (0.70)

Abbreviations: *AFC* Antral follicular count, *AMH* Anti, Mullerian hormone, *CI* Confidence interval, *DOR* Diagnostic odds ratio, *AUC* Area under curve, *E2* Estradiol, *FSH* Follicle stimulating hormone, *Ln* Logarithm, *MSS* Median sample size, *ORM* Ovarian reserve marker, *pAUC* Partial AUC, *SE* standard error, T Tau, T^2 Tau squared. Large and small represent sample size

* Studies classified as 'small sample size' include those with fewer participants than the median sample size, while 'large sample size' refers to studies with more participants than the median"

** *P*-value for Ln DOR

the effectiveness of various hormonal and genetic indicators in predicting egg quality and quantity can reduce treatment failures and complications. Future research should focus on distinguishing ovarian response from the likelihood of successful pregnancy by creating sensitive assays to measure oocyte quality.

Abbreviations

IVF	In vitro fertilization
ART	Assisted reproductive technology
ROR	Reduced ovarian reserves
POR	Poor ovarian response
ORM	Ovarian reserve marker
COH	Controlled ovarian hyperstimulation
FSH	Follicle-stimulating hormone
LH	Luteinizing hormone
E2	Estradiol
AFC	Antral follicular count
AMH	Anti-Mullerian hormone
CCCT	Clomiphene citrate challenge test
GnRH	Gonadotropin hormone-releasing hormone
ROC	Receiver operating characteristic
AUC	Area under curve
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
PROSPERO	International Prospective Register of Systematic Reviews
TP	True positive
TN	True negative
FP	False positive
FN	False negative
NOS	Newcastle-Ottawa scale
POI	Premature ovarian insufficiency

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13643-024-02684-0.

Additional file 1: Supplementary Table 1. Search strategies used for each database to find relevant studies.

Acknowledgements
Not applicable

Authors' contributions
FS conceptualized and designed the study, drafted and critically reviewed the manuscript, and approved the final manuscript as submitted. ML conceptualized and designed the study, critically reviewed the manuscript, and approved the final manuscript as submitted. ASA designed the study, critically reviewed the manuscript, and approved the final manuscript as submitted. SJ conceptualized the study, designed the study, critically reviewed the manuscript, and approved the final manuscript as submitted. AK, ZA, and HA drafted the manuscript and approved the final manuscript as submitted. SA drafted and critically reviewed the manuscript and approved the final manuscript as submitted. All authors read and approved the final manuscript.

Funding
None.

Data availability
The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Hematology, Oncology and Stem Cell Transplantation Research Center, Hematology and Cell Therapy, Research Institute for Oncology, Tehran University of Medical Sciences, Tehran, Iran. [2]Departments of Biostatistics and Epidemiology, Center for Healthcare Data Modeling, School of Public Health, Shahid Sadoughi University of Medical Sciences, Yazd 8915173160, Iran. [3]Department of Medical Genetics, School of Medicine, Shahid Sadoughi University of Medical Sciences, Yazd, Iran. [4]Nutrition and Food Security Research Center, Shahid Sadoughi University of Medical Sciences, Yazd, Iran. [5]Department of Nutrition, School of Public Health, Shahid Sadoughi University of Medical Sciences, Yazd, Iran. [6]Genetic and Environmental Adventure Research Center,, school of Abarkouh paramedicine, shahid sadoughi university of medical science, Yazd, Iran. [7]Abortion Research Center, Yazd Reproductive Sciences Institute, Shahid Sadoughi University of Medical Sciences, Yazd, Iran. [8]School of Medicine, Academic Unit of Translational Medical Sciences, University of Nottingham, Nottingham, UK.

Received: 22 October 2023 Accepted: 14 October 2024
Published online: 18 December 2024

References

1. Broekmans FJ, et al. A systematic review of tests predicting ovarian reserve and IVF outcome. Hum Reprod Update. 2006;12(6):685–718.
2. Wang X, et al. Evaluation of ovarian reserve tests and age in the prediction of poor ovarian response to controlled ovarian stimulation-a real-world data analysis of 89,002 patients. Front Endocrinol (Lausanne). 2021;12: 702061.
3. Medicine, P.C.o.t.A.S.f.R. Diagnostic evaluation of the infertile male: a committee opinion. Fertility and sterility. 2015;103(3):e18–25.
4. Chon SJ, Umair Z, Yoon M-S. Premature ovarian insufficiency: past, present, and future. Frontiers in cell and developmental biology. 2021;9: 672890.
5. Jirge PR. Poor ovarian reserve. J Hum Reprod Sci. 2016;9(2):63–9.
6. Man L, et al. Ovarian reserve disorders, can we prevent them? A review. Int J Mol Sci. 2022;23(23):15426.
7. Jirge PR. Poor ovarian reserve. Journal of human reproductive sciences. 2016;9(2):63.
8. Medicine, P.C.o.t.A.S.f.R. Fertility and sterility. Fertility and sterility. 2015;103(3):e9–17.
9. Koert, E., et al., *Women's perceptions of fertility assessment and counselling 6 years after attending a Fertility Assessment and Counselling Clinic in Denmark*. Hum Reprod Open, 2020. 2020(4): p. hoaa036.
10. Tal R, Seifer DB. Ovarian reserve testing: a user's guide. Am J Obstet Gynecol. 2017;217(2):129–40.
11. Jirge PR. Ovarian reserve tests. Journal of human reproductive sciences. 2011;4(3):108.
12. Galey-Fontaine J, et al. Age and ovarian reserve are distinct predictive factors of cycle outcome in low responders. Reprod Biomed Online. 2005;10(1):94–9.
13. Ryan KJ. Biochemistry of aromatase: significance to female reproductive physiology. Can Res. 1982;42(8 Supplement):3342s–4s.
14. Lee T-H, et al. Serum anti-Müllerian hormone and estradiol levels as predictors of ovarian hyperstimulation syndrome in assisted reproduction technology cycles. Hum Reprod. 2008;23(1):160–7.
15. Aflatoonian A, et al. Effects of intraovarian injection of autologous platelet-rich plasma on ovarian rejuvenation in poor responders and women with primary ovarian insufficiency. Reprod Sci. 2021;28(7):2050–9.
16. La Marca A, Sunkara SK. Individualization of controlled ovarian stimulation in IVF using ovarian reserve markers: from theory to practice. Hum Reprod Update. 2014;20(1):124–40.
17. La Marca A, et al. Anti-Müllerian hormone (AMH) as a predictive marker in assisted reproductive technology (ART). Hum Reprod Update. 2010;16(2):113–30.

18. Jamil Z, et al. Anti-Mullerian hormone: above and beyond conventional ovarian reserve markers. Dis Markers. 2016;2016(1):5246217.
19. La Marca A, et al. Serum anti-Mullerian hormone throughout the human menstrual cycle. Hum Reprod. 2006;21(12):3103–7.
20. Jayaprakasan K, et al. Establishing the intercycle variability of three-dimensional ultrasonographic predictors of ovarian reserve. Fertil Steril. 2008;90(6):2126–32.
21. Sharara FI, Scott RT Jr, Seifer DB. The detection of diminished ovarian reserve in infertile women. Am J Obstet Gynecol. 1998;179(3):804–12.
22. Broer S, et al. AMH and AFC as predictors of excessive response in controlled ovarian hyperstimulation: a meta-analysis. Hum Reprod Update. 2011;17(1):46–54.
23. Broer S.L, et al. Prediction of an excessive response in in vitro fetilization from patient characteristics and ovarian reserve test and comparison in subgroups: an individual patient data meta-analysis. Fertility and sterility. 2013;100(2):420-429. e7.
24. Tal, R., et al., *AntiMüllerian hormone as predictor of implantation and clinical pregnancy after assisted conception: a systematic review and meta-analysis.* Fertility and sterility, 2015. 103(1): p. 119–130. e3.
25. Moher D, et al. PRISMA statement. Epidemiology. 2011;22(1):128.
26. Cohen J, Chabbert-Buffet N, Darai E. Diminished ovarian reserve, premature ovarian failure, poor ovarian responder—a plea for universal definitions. J Assist Reprod Genet. 2015;32(12):1709–12.
27. Gunning M, et al. Premature ovarian insufficiency. Female Reprod Dysfunction. 2020:287–307.
28. Altmäe S, et al. Genetic predictors of controlled ovarian hyperstimulation: where do we stand today? Hum Reprod Update. 2011;17(6):813–28.
29. Cohlen BJ, et al. Controlled ovarian hyperstimulation and intrauterine insemination for treating male subfertility: a controlled study. Human reproduction (Oxford, England). 1998;13(6):1553–8.
30. Whiting PF, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36.
31. Amer Abed, F., R. Ezzat Maroof, and U.M.A. Al-Nakkash, *Comparing the diagnostic accuracy of anti-Mullerian hormone and follicle stimulating hormone in detecting premature ovarian failure in Iraqi women by ROC analysis.* Rep Biochem Mol Biol, 2019. 8(2): p. 126–131.
32. Lunding SA, et al. AMH as predictor of premature ovarian insufficiency: a longitudinal study of 120 Turner syndrome patients. J Clin Endocrinol Metab. 2015;100(7):E1030–8.
33. Alipour F, et al. Comparison of specificity and sensitivity of AMH and FSH in diagnosis of premature ovarian failure. Dis Markers. 2015;2015: 585604.
34. Reichman DE, Goldschlag D, Rosenwaks Z. Value of antimüllerian hormone as a prognostic indicator of in vitro fertilization outcome. Fertil Steril. 2014;101(4):1012-1018.e1.
35. Kallio S, et al. Anti-Mullerian hormone as a predictor of follicular reserve in ovarian insufficiency: special emphasis on FSH-resistant ovaries. Hum Reprod. 2012;27(3):854–60.
36. Nikbakht R, Borhani M, Sarvestani GY. Evaluation of ovarian reserve by measurement of the serum levels of anti-mullerian hormone and follicle-stimulating hormone in intracytoplasmic sperm injection cycles. International Journal of Fertility and Sterility. 2010;4(2):57–60.
37. Kaya C, Pabuccu R, Satiroglu H. Serum antimullerian hormone concentrations on day 3 of the in vitro fertilization stimulation cycle are predictive of the fertilization, implantation, and pregnancy in polycystic ovary syndrome patients undergoing assisted reproduction. Fertil Steril. 2010;94(6):2202–7.
38. Aslan K, et al. Which is more predictive ovarian sensitivity marker if there is discordance between serum anti-Müllerian hormone levels and antral follicle count? A retrospective analysis J Obstet Gynaecol. 2022;42(5):1461–6.
39. Barrière P, et al. Added value of anti-Müllerian hormone serum concentration in assisted reproduction clinical practice using highly purified human menopausal gonadotropin (HP-hMG). J Gynecol Obstet Hum Reprod. 2022;51(2): 102289.
40. Heidary Z, et al. The association of AMH level with the number and quality of oocytes in women undergoing IVF/ICSI: a single-center study. Journal of Reproduction and Infertility. 2024;25(1):38–45.
41. Jabarpour M, et al. High anti-Müllerian hormone strongly correlates with reproductive outcomes in women undergoing assisted reproduction. Clin Lab. 2023;69(12).
42. Permadi W, et al. Correlation of anti-Mullerian hormone level and antral follicle count with oocyte number in a fixed-dose controlled ovarian hyperstimulation of patients of in vitro fertilization program. International Journal of Fertility and Sterility. 2021;15(1):40–3.
43. Sinha S, Sharan A, Sinha S. Anti-Mullerian hormone as a marker of ovarian reserve and function. Cureus. 2022;14(9): e29214.
44. Sun PP, et al. The relationship between serum FSH level and ovarian response during controlled ovarian stimulation. Ginekol Pol. 2023;94(6):470–5.
45. Vagios S, et al. Response to ovulation induction treatments in women with polycystic ovary syndrome as a function of serum anti-Müllerian hormone levels. J Assist Reprod Genet. 2021;38(7):1827–33.
46. Zajicek M, et al. Evaluation of ovarian reserve in young females with non-iatrogenic ovarian insufficiency to establish criteria for ovarian tissue cryopreservation. Reprod Biomed Online. 2023;47(1):102–9.
47. El-Mazny A, Abou-Salem N. Anti-Müllerian hormone and antral follicle count for prediction of ovarian stimulation response in polycystic ovary syndrome. Gynecol Endocrinol. 2013;29(9):826–9.
48. Bancsi L, et al. Impact of repeated antral follicle counts on the prediction of poor ovarian response in women undergoing in vitro fertilization. Fertil Steril. 2004;81(1):35–41.
49. Mukherjee T, et al. An elevated day three follicle-stimulating hormone:luteinizing hormone ratio (FSH:LH) in the presence of a normal day 3 FSH predicts a poor response to controlled ovarian hyperstimulation. Fertil Steril. 1996;65(3):588–93.
50. Hendriks DJ, et al. Repeated clomiphene citrate challenge testing in the prediction of outcome in IVF: a comparison with basal markers for ovarian reserve. Hum Reprod. 2005;20(1):163–9.
51. Xu X, et al. Development and validation of a prediction model for unexpected poor ovarian response during IVF/ICSI. Front Endocrinol. 2024;15:1340329.
52. Yan Y, et al. Clinical features and management of suboptimal ovarian response during in vitro fertilization and embryo transfer: analysis based on a retrospective cohort study. Front Endocrinol. 2022;13:938926.
53. Jacobs MH, et al. A multicentre evaluation of the Elecsys ® anti-Müllerian hormone immunoassay for prediction of antral follicle count. Reprod Biomed Online. 2019;38(5):845–52.
54. Sahmay S, et al. Elevated LH levels draw a stronger distinction than AMH in premature ovarian insufficiency. Climacteric. 2014;17(2):197–203.
55. Fouda F, et al. Anti-mullerian hormone level is a reliable predictor for cycle cancellation in ICSI. Middle East Fertility Society Journal. 2010;15(3):194–9.
56. Rashad NM, et al. Anti-Müllerian hormone: predictor of premature ovarian insufficiency in Egyptian women with autoimmune thyroiditis. Middle East Fertility Society Journal. 2018;23(4):286–91.
57. Saikumar P, et al. Anti Mullerian hormone: a potential marker for recruited non growing follicle of ovarian pool in women with polycystic ovarian syndrome. J Clin Diagn Res. 2013;7(9):1866–9.
58. La Marca A, et al. Anti-Müllerian hormone-based prediction model for a live birth in assisted reproduction. Reprod Biomed Online. 2011;22(4):341–9.
59. Gleicher N, Weghofer A, Barad DH. Anti-Müllerian hormone (AMH) defines, independent of age, low versus good live-birth chances in women with severely diminished ovarian reserve. Fertil Steril. 2010;94(7):2824–7.
60. Burks HR, et al. Can highly sensitive antimüllerian hormone testing predict failed response to ovarian stimulation? Fertil Steril. 2015;104(3):643–8.
61. Lee TH, et al. Serum anti-Mullerian hormone and estradiol levels as predictors of ovarian hyperstimulation syndrome in assisted reproduction technology cycles. Hum Reprod. 2008;23(1):160–7.
62. Aydin T, et al. The association between anti-Müllerian hormone and IVF-ICSI outcome in poor responder patients performing long protocol. Clin Exp Obstet Gynecol. 2015;42(5):663–5.

63. Asada Y, et al. Performance of anti-Müllerian hormone (AMH) levels measured by Beckman Coulter Access AMH assay to predict oocyte yield following controlled ovarian stimulation for in vitro fertilization. Reproductive Medicine and Biology. 2019;18(3):273–7.
64. Bosch E, et al. Prediction of ovarian response using the automated Elecsys anti-Müllerian hormone assay in gonadotrophin-releasing hormone antagonist cycles. Reprod Biomed Online. 2023;46(2):295–301.
65. Peralta S, et al. Antral follicle count measured at down-regulation as predictor of ovarian response and cumulative live birth: single center analysis including 2731 long agonist IVF cycles. Gynecol Endocrinol. 2022;38(12):1079–86.
66. Hochberg A, et al. Significance of serum AMH and antral follicle count discrepancy for the prediction of ovarian stimulation response in Poseidon criteria patients. J Assist Reprod Genet. 2024;41(3):717–26.
67. Jabir Edan B, Mahdi Kadhim H, Resheed Behayaa H. Sensitivity and specificity of measuring anti-Müllerian hormone and follicle stimulating hormone levels in predicting response to ovarian stimulation in infertile women. J Babol Univ Med Sci. 2024;26:0.
68. Razafintsalama-Bourdet M, et al. Random antral follicle count performed on any day of the menstrual cycle has the same predictive value as AMH for good ovarian response in IVF cycles. J Gynecol Obstet Hum Reproduct. 2022;51(1):102233.
69. Zheng H, et al. Ovarian response prediction in controlled ovarian stimulation for IVF using anti-Mullerian hormone in Chinese women: a retrospective cohort study. Medicine (Baltimore). 2017;96(13): e6495.
70. Lerman T, et al. Ovarian response to 150 µg corifollitropin alfa in a GnRH-antagonist multiple-dose protocol: a prospective cohort study. Reprod Biomed Online. 2017;34(5):534–40.
71. Li R, et al. Anti-Müllerian hormone for prediction of ovarian response in Chinese infertile women undergoing IVF/ICSI cycles: a prospective, multi-centre, observational study. Reprod Biomed Online. 2016;33(4):506–12.
72. Aghssa MM, et al. Optimal cutoff value of basal anti-mullerian hormone in iranian infertile women for prediction of ovarian hyper-stimulation syndrome and poor response to stimulation female fertility. Reprod Health. 2015;12:1–8.
73. Mehrafza M, et al. Anti-müllerian hormone as a predictor of IVF treatment. Tehran University Medical Journal. 2012;70(6):371–7.
74. Nardo LG, et al. Circulating basal anti-Müllerian hormone levels as predictor of ovarian response in women undergoing ovarian stimulation for in vitro fertilization. Fertil Steril. 2009;92(5):1586–93.
75. Riggs RM, et al. Assessment of ovarian reserve with anti-Mullerian hormone: a comparison of the predictive value of anti-Mullerian hormone, follicle-stimulating hormone, inhibin B, and age. Am J Obstet Gynecol. 2008;199(2):202.e1-8.
76. Laqqan MM, Yassin MM. Predictive factors of ovarian response to GnRH antagonist stimulation protocol: AMH and age are potential candidates. Middle East Fertility Society Journal. 2021;26(1):1–11.
77. Peluso C, et al. Are ovarian reserve tests reliable in predicting ovarian response? Results from a prospective, cross-sectional, single-center analysis. Gynecol Endocrinol. 2021;37(4):358–66.
78. Neves AR, et al. The performance of the Elecsys® anti-Müllerian hormone assay in predicting extremes of ovarian response to corifollitropin alfa. Reprod Biomed Online. 2020;41(1):29–36.
79. Lee RWK, et al. Ovarian biomarkers predict controlled ovarian stimulation for in vitro fertilisation treatment in Singapore. Singapore Med J. 2020;61(9):463–8.
80. Laqqan MM, Yassin MM. Anti-Mullerian hormone and antral follicle count predict ovarian response in women less than 45 years following GnRH antagonist multiple-dose protocol. Asian Pacific Journal of Reproduction. 2022;11(5):208–16.
81. Sun X, et al. Comparison of the predictive capability of antral follicle count vs. the anti-Müllerian hormone for ovarian response in infertile women. Front Endocrinol. 2022;13:862733.
82. Izhar R, et al. Antral follicle count and anti-Müllerian hormone level as predictors of ovarian hyperstimulation syndrome in women with polycystic ovarian syndrome undergoing controlled ovarian stimulation. Journal of Ultrasonography. 2021;21(86): e200.
83. Aydın GA, et al. Assessment of the relationship of basal serum anti-mullerian hormone levels with oocyte quality and pregnancy outcomes in patients undergoing ICSI. Iranian Journal of Reproductive Medicine. 2015;13(4):231–6.
84. Durmusoglu F, et al. Combining cycle day 7 follicle count with the basal antral follicle count improves the prediction of ovarian response. Fertil Steril. 2004;81(4):1073–8.
85. Jayaprakasan K, et al. A prospective, comparative analysis of anti-Müllerian hormone, inhibin-B, and three-dimensional ultrasound determinants of ovarian reserve in the prediction of poor response to controlled ovarian stimulation. Fertil Steril. 2010;93(3):855–64.
86. Martínez F, et al. Is AMH useful to reduce low ovarian response to GnRH antagonist protocol in oocyte donors? Gynecol Endocrinol. 2013;29(8):754–7.
87. Baker VL, et al. Multicenter evaluation of the access AMH antimüllerian hormone assay for the prediction of antral follicle count and poor ovarian response to controlled ovarian stimulation. Fertil Steril. 2018;110(3):506-513.e3.
88. Muttukrishna, S., et al., *Antral follicle count, anti-mullerian hormone and inhibin B: predictors of ovarian response in assisted reproductive technology?* BJOG: An International Journal of Obstetrics and Gynaecology, 2005. 112(10): p. 1384–1390.
89. Kasapoglu I, et al. Are all antral follicles the same? Size of antral follicles as a key predictor for response to controlled ovarian stimulation. J Obstet Gynaecol. 2022;42(3):461–6.
90. Tremellen KP, et al. Anti-mullerian hormone as a marker of ovarian reserve. Aust N Z J Obstet Gynaecol. 2005;45(1):20–4.
91. Dermolo M, Ansa M, Siferih M. Ovarian response to controlled stimulation and its predictors in a limited-resource setting. BMC Women's Health. 2024;24(1):279.
92. Esteves SC, et al. Antral follicle count and anti-Müllerian hormone to classify low-prognosis women under the POSEIDON criteria: a classification agreement study of over 9000 patients. Hum Reprod. 2021;36(6):1530–41.
93. La Marca A, et al. Predicting the total number of retrieved oocytes following double ovarian stimulation (DuoStim). Hum Reprod. 2023;38(9):1784–8.
94. Laqqan MM, Yassin MM. Investigation of the predictive factors of diminished ovarian reserve in women aged less than 40 years and undergoing ICSI cycle. Reprod Sci. 2023;30(3):873–82.
95. Rosen MP, et al. A characterization of the relationship of ovarian reserve markers with age. Fertil Steril. 2012;97(1):238–43.
96. Jiao X, et al. Ovarian reserve markers in premature ovarian insufficiency: within different clinical stages and different etiologies. Front Endocrinol. 2021;12:601752–601752.
97. Moolhuijsen LME, Visser JA. Anti-Müllerian hormone and ovarian reserve: update on assessing ovarian function. J Clin Endocrinol Metab. 2020;105(11):3361–73.
98. Keane K, et al. Specific ranges of anti-Mullerian hormone and antral follicle count correlate to provide a prognostic indicator for IVF outcome. Reprod Biol. 2017;17(1):51–9.
99. Hart KN, et al. Structure of AMH bound to AMHR2 provides insight into a unique signaling pair in the TGF-β family. Proc Natl Acad Sci. 2021;118(26): e2104809118.
100. Malone SA, et al. Defective AMH signaling disrupts GnRH neuron development and function and contributes to hypogonadotropic hypogonadism. Elife. 2019;8:e47198.
101. Visser JA, et al. Anti-Müllerian hormone: an ovarian reserve marker in primary ovarian insufficiency. Nat Rev Endocrinol. 2012;8(6):331–41.
102. Visser JA, et al. Increased oocyte degeneration and follicular atresia during the estrous cycle in anti-Müllerian hormone null mice. Endocrinology. 2007;148(5):2301–8.
103. Hayes E, et al. Intra-cellular mechanism of anti-Müllerian hormone (AMH) in regulation of follicular development. Mol Cell Endocrinol. 2016;433:56–65.
104. Broer SL, et al. Anti-Müllerian hormone: ovarian reserve testing and its potential clinical implications. Hum Reprod Update. 2014;20(5):688–701.
105. Welt CK. Primary ovarian insufficiency: a more accurate term for premature ovarian failure. Clin Endocrinol. 2008;68(4):499–509.

106. Bukman A, Heineman M. Ovarian reserve testing and the use of prognostic models in patients with subfertility. Hum Reprod Update. 2001;7(6):581–90.
107. Sakaguchi K, et al. Relationships between the antral follicle count, steroidogenesis, and secretion of follicle-stimulating hormone and anti-Müllerian hormone during follicular growth in cattle. Reprod Biol Endocrinol. 2019;17(1):88.

Publisher's Note