scientific reports

OPEN

Female reproductive health trends and autism spectrum disorder prevalence between 2000 and 2024

Adil Abdul-Rehman Siddiq Al-Salihy

This study investigates the correlation between female reproductive health parameters and Autism Spectrum Disorder (ASD) prevalence from 2000 to 2024. The analysis used advanced statistical and machine learning models to identify trends in key reproductive indicators and their association with ASD prevalence. Significant positive correlations were observed between ASD prevalence and maternal age, while negative correlations were found with antral follicle count, Anti-Müllerian Hormone (AMH) levels, and fertility rate. The Random Forest model emerged as the most accurate predictive tool, explaining 96.9% of the variance in ASD prevalence. Maternal age was the dominant predictor of the variables analyzed, contributing approximately 75% of the model's predictive power, while estradiol levels and Follicle Stimulating Hormone (FSH) contributed significantly less. These findings highlight potential statistical associations but do not establish causality. Further research is necessary to validate these associations and explore underlying biological mechanisms.

The manifestations of Autism Spectrum Disorder (ASD) encompass a wide spectrum of characteristics, including challenges in social communication, restricted interests, and repetitive behaviors. These manifestations range from significant cognitive and functional impairments to exceptional abilities in specific areas. The DSM-5[1] emphasizes the dimensional nature of ASD, recognizing varying levels of support required by individuals. Over the past few decades, there has been a notable increase in the prevalence of ASD, raising concerns and prompting extensive research into its causes, diagnosis, and treatment[2].

According to the Centers for Disease Control and Prevention (CDC), ASD prevalence in the United States has surged from approximately 1 in 150 children in 2000 to 1 in 36 by 2024[3]. Several factors have contributed to this apparent increase, including improved diagnostic criteria, heightened awareness among healthcare professionals and the public, and expanding screening and diagnostic services[4]. These advancements have led to earlier and more accurate diagnoses, capturing a broader spectrum of the condition. Regular public health monitoring has become crucial due to the rising prevalence rates of ASD and the complexity of tracking such disorders over time[5].

While genetic predispositions play a crucial role in the development of ASD, environmental factors are also significant contributors. Prenatal exposure to toxins, advanced parental age at the time of conception, and certain perinatal conditions have been linked to an increased risk of ASD[6,7]. Maternal health conditions, such as polycystic ovary syndrome (PCOS) and gestational diabetes mellitus (GDM), have been increasingly associated with ASD. PCOS, characterized by hormonal imbalances and chronic low-grade inflammation, may influence fetal neurodevelopment through alterations in the intrauterine environment[8]. Similarly, GDM, which affects maternal glucose metabolism, has been linked to an increased risk of neurodevelopmental disorders in offspring due to potential epigenetic changes and oxidative stress[9]. Research into these environmental and maternal health factors aims to identify modifiable risk factors that could lead to preventative strategies and interventions.

One area of growing interest is the potential link between female reproductive health and ASD prevalence. Reproductive health parameters such as menstrual cycle irregularities, polycystic ovary syndrome (PCOS), and gestational diabetes mellitus (GDM) have been implicated in the risk of developmental disorders, including ASD[10,11]. These conditions can influence hormonal levels and reproductive outcomes, which may, in turn, affect fetal development and the risk of ASD.

Consultant in Clinical Psychology\Neuro- & Psycho-Therapy Science, Head of the Mental Health Department, Psychological Research Center, Ministry of Higher Education & Scientific Research, Baghdad, Iraq. email: adil_alsalihy@yahoo.com

1

Several key indicators of reproductive health are particularly relevant to this study. Follicle Stimulating Hormone (FSH) and Estradiol levels are critical for ovarian function and reproductive health. Anti-Müllerian Hormone (AMH) levels and antral follicle count are ovarian reserve markers, providing insights into a woman's fertility potential. Ovarian volume, maternal age, and Clomiphene Citrate Challenge Test (CCCT) FSH levels are additional parameters that can influence pregnancy outcomes and the developmental health of offspring. These indicators are linked to various pregnancy outcomes and child developmental issues, potentially impacting ASD risk[12,13].

Understanding the statistical associations between these reproductive health parameters and ASD prevalence is essential for identifying potential trends. While these findings may provide preliminary insights, further research is needed to determine whether these factors directly affect ASD risk. Second, it contributes to the broader understanding of ASD etiology, potentially uncovering new pathways for prevention and treatment. Third, it provides valuable information for public health strategies and policies aimed at improving maternal and reproductive health, thereby reducing the burden of neurodevelopmental disorders like ASD. Finally, these findings can inform the development of predictive models that support clinicians in identifying at-risk individuals and implementing personalized healthcare approaches.

This study aims to examine correlations between female reproductive health parameters and ASD prevalence, identifying potential statistical trends that may inform future research on ASD risk factors. Analyzing data from 2000 to 2024, we seek to identify trends, explore possible relationships, and develop predictive models to forecast ASD prevalence based on reproductive health indicators. This study primarily utilizes data from the United States, sourced from the Centers for Disease Control and Prevention (CDC), and complementary international datasets for reproductive health parameters where applicable. The choice of U.S.-based data reflects the consistency and availability of longitudinal ASD prevalence data. However, we acknowledge that regional variability in healthcare infrastructure, environmental exposures, and diagnostic practices may influence trends, and further studies are warranted to validate findings in diverse geographical contexts.

Objectives

1. Analyze trends in ASD prevalence (2000–2024), focusing on identifying temporal changes and their associated contributing factors.
2. Investigate trends in female reproductive health parameters during the same period, with a focus on key indicators such as AMH levels, maternal age, and ovarian volume and their potential impact on ASD prevalence.
3. Analyze statistical associations between female reproductive health indicators and ASD prevalence to explore potential trends and patterns in the data.
4. Identify reproductive health indicators statistically associated with ASD prevalence using advanced statistical analyses and machine learning techniques.
5. Develop and validate predictive models for ASD prevalence to support early identification and inform targeted interventions.

Hypotheses

1. ASD prevalence has significantly increased from 2000 to 2024: It is hypothesized that the prevalence of ASD has risen due to a combination of improved diagnostic practices, increased awareness, and potential environmental factors[3,4].
2. Female reproductive health parameters have undergone significant changes over the past two decades: Shifts in environmental exposures, lifestyle changes, and advancements in healthcare practices are hypothesized to have influenced key reproductive health indicators such as AMH levels, ovarian volume, and maternal age Grether et al.[6,14].
3. Statistical associations exist between female reproductive health parameters and ASD prevalence: It is hypothesized that reproductive health indicators, such as maternal age and ovarian reserve markers, will demonstrate statistically significant correlations with ASD prevalence[6,7].
4. Certain reproductive health parameters may be statistically associated with ASD prevalence: It is hypothesized that hormonal imbalances (e.g., AMH, estradiol levels) and ovarian reserve markers will show statistical associations with ASD prevalence, potentially reflecting underlying biological mechanisms that warrant further investigation.
5. Predictive models incorporating reproductive health parameters may identify statistical patterns associated with ASD prevalence: We hypothesize that advanced statistical and machine learning models will reveal predictive associations, which may contribute to hypothesis generation for further research[12].

By addressing these hypotheses, the study aims to uncover novel insights into the relationship between female reproductive health and ASD prevalence, ultimately guiding public health strategies and informing future research.

Methods and materials
Study design
This longitudinal study explores statistical associations between female reproductive health parameters and ASD prevalence from 2000 to 2024, focusing on identifying trends and potential predictive patterns.

Data availability

Data will be sourced from various reputable repositories, including peer-reviewed journals, national and international health organizations, and comprehensive meta-analyses. The key variables of interest include:

1. ASD Prevalence Data (Dependent Variable): Sourced from the Centers for Disease Control and Prevention (CDC) Autism and Developmental Disabilities Monitoring (ADDM) Network, which tracks the prevalence and characteristics of ASD among 8-year-old children in multiple communities across the United States[3,15].
2. Reproductive Health Parameters (Independent Variables): Sourced from various published studies and health reports, including national health surveys and medical records. The data covers female reproductive health indicators, such as those from women of reproductive age (18–45 years) who have undergone reproductive health assessments during the specified period.

- Follicle Stimulating Hormone (FSH) Levels (IU/L): FSH is essential for ovarian follicle growth and maturation. Research by Oduwole et al.[16] highlights its relevance to reproductive health and sperm quality. Wu et al.[17], Mao et al.[18] found that rising FSH levels affect body composition in postmenopausal women, increasing body fat and decreasing lean mass. A Study of Women's Health Across the Nation (SWAN) data also shows FSH trends related to menopause and body composition changes[14].
- Estradiol Levels (pg/mL): Estradiol regulates the menstrual cycle and supports pregnancy. Baron-Cohen et al.[19] discuss the role of hormonal influences, particularly prenatal testosterone, in the development of ASD. Additionally, research by Harlow et al.[20] published in The *Journal of Clinical Endocrinology & Metabolism* emphasizes the importance of hormonal changes, including estradiol levels, during the stages of reproductive aging, particularly in relation to the menopausal transition and its effects on psychological symptoms and bone health.
- Anti-Müllerian Hormone (AMH) Levels (ng/mL): AMH indicates a woman's ovarian reserve. Key references include Rolland et al.[13], which present detailed findings on semen concentration and motility trends over time. Data from *Human Reproduction* by Broekmans et al.[21], Broer et al.[22] show trends in AMH levels as indicators of ovarian reserve and their implications for fertility over time, linking these trends to reproductive aging.
- Antral Follicle Count: Reflects the number of small follicles in the ovaries. Key references include Cherskov et al.[11], who comprehensively analyze the hormonal influences on reproductive health and their potential connections to ASD, particularly in the context of conditions like polycystic ovary syndrome (PCOS).
- Ovarian Volume (mL): Measures ovarian reserve and health. Key references include Broekmans et al.[21], who analyze the role of ovarian reserve and ovarian volume in female reproductive aging and its implications for fertility outcomes.

A systematic review in *Human Reproduction Update* by Broer et al.[22] highlights the predictive value of antral follicle count and ovarian volume for ovarian reserve and IVF outcomes, focusing on longitudinal patterns and clinical applications.

- Maternal Age (years): Advanced maternal age is linked to increased risks of pregnancy complications and developmental disorders, including ASD. Key references include Grether et al.[6] and Sandin et al.[7], which discuss the relationship between maternal age and ASD prevalence. The CDC and WHO publish data on global maternal age trends, indicating an increasing trend in maternal age at childbirth over the past two decades, with detailed statistics available in their health reports.
- Clomiphene Citrate Challenge Test (CCCT) FSH Levels (IU/L): CCCT is a widely used diagnostic tool for assessing ovarian reserve and function. Research by Hendriks et al.[23] highlights the predictive value of CCCT in identifying women at risk of poor ovarian response and nonpregnancy during in vitro fertilization (IVF) treatments. Similarly, Kwee et al.[24] compare CCCT with the exogenous follicle-stimulating hormone (FSH) ovarian reserve test, emphasizing its utility in predicting low responders and hyperresponders to IVF protocols.
- Fertility Rate (births per 1,000): The fertility rate indicates the number of live births per woman. Recent analyses by the United Nations[25] and the World Bank[26] provide comprehensive statistical data on fertility trends and their impact on population demographics. Comprehensive datasets on global fertility rates, showing trends and changes over the past few decades, are available from the World Bank and the United Nations through their respective data portals.

These sources provide reliable data to ensure the robustness of the analysis. The selected variables were chosen based on their established biological and epidemiological relevance to ASD risk. Advanced maternal age has been associated with an increased risk of neurodevelopmental disorders due to genetic and epigenetic factors[7]. AMH levels reflect ovarian reserve, influencing intrauterine conditions[11]. FSH and estradiol levels are crucial for hormonal regulation, with links to pregnancy outcomes and neurodevelopment[10,22]. Indicators such as antral follicle count, ovarian volume, and fertility rate represent trends in reproductive health, which may indirectly impact ASD prevalence[12,25,26].

To maintain a targeted analysis, the researchers focused on utilizing data from women aged 18–45 years, as this demographic represents the majority of reproductive health assessments within the datasets. However, the study examines broader population trends over time rather than individual-level fertility variations, ensuring a comprehensive understanding of reproductive health shifts. Consequently, the findings may have limited generalizability to women outside this age range. Future research should aim to include a more diverse population to enhance the applicability and generalizability of the results.

All data sources were rigorously reviewed to ensure consistency and relevance. Autism prevalence data were exclusively obtained from the CDC Autism and Developmental Disabilities Monitoring Network, ensuring high reliability. Reproductive health parameters were obtained from established datasets and peer-reviewed studies, including large-scale national health surveys and longitudinal medical records. The analysis focuses on broader demographic trends in reproductive health rather than direct population-to-population comparisons, ensuring that findings remain relevant across diverse settings.

Final compiled data - trends in ASD prevalence and reproductive health parameters (2000–2024)

The dataset for this study includes reproductive health parameters and ASD prevalence data from 2000 to 2024. The table below offers a comprehensive overview of psychometric characteristics such as ASD prevalence rates, hormonal levels, and fertility indicators. These variables were chosen for their relevance to reproductive health and potential association with ASD prevalence. Data was collected from multiple sources to ensure a robust and representative sample. See Table 1 for an integrated overview of trends in reproductive health parameters and ASD prevalence over time.

Statistical analysis

1. Descriptive Statistics:

   - Mean, standard deviation, range, median, minimum, and maximum values were calculated to summarize reproductive health parameters and ASD prevalence rates.

2. Correlation Analysis:

   - Pearson and Spearman correlation coefficients were used to assess the strength and direction of relationships between ASD prevalence and reproductive health parameters.

3. Multiple Linear Regression:

   - We identified significant predictors of ASD prevalence while controlling for confounders like maternal age and socio-economic status.
   - Logistic regression quantified associations using odds ratios (OR) and 95% confidence intervals (CI).

4. Time-Series Analysis:

   - Linear and polynomial regression models explored temporal trends in ASD prevalence and reproductive health parameters.
   - A Seasonal AutoRegressive Integrated Moving Average with eXogenous factors (SARIMAX) model forecasted future trends, incorporating seasonality and external variables.

5. Machine Learning Models:

   - Techniques such as Random Forest, Gradient Boosting, and Support Vector Regressor developed predictive models for ASD prevalence based on reproductive health indicators.

Year	ASD Prevalence (per 1,000 live births)	Autism Rate (1 in x children)	FSH Levels (IU/L)	Estradiol Levels (pg/mL)	AMH Levels (ng/mL)	Antral Follicle Count (Number of Follicles)	CCCT FSH Levels (IU/L)	Fertility Rate (births per 1,000)	Ovarian Volume (mL)	Maternal Age (years)
2000	6.7	1 in 150	6.8	60	2.5	16	7.8	14.6	5.5	26.0
2002	6.6	1 in 151	6.7	59	2.4	15	7.7	14.2	5.4	26.1
2004	8.0	1 in 125	6.6	58	2.3	15	7.6	13.9	5.3	26.2
2006	9.0	1 in 111	6.5	57	2.3	14	7.5	13.7	5.2	26.4
2008	11.3	1 in 88	6.4	56	2.2	14	7.4	13.5	5.1	26.5
2010	14.7	1 in 68	6.3	55	2.1	13	7.3	13.2	5.0	26.6
2012	14.5	1 in 69	6.2	54	2.0	13	7.2	13.0	4.9	26.7
2014	16.8	1 in 59	6.1	53	1.9	12	7.1	12.8	4.8	26.9
2016	18.5	1 in 54	6.0	52	1.8	12	7.0	12.5	4.7	27.0
2018	18.5	1 in 54	5.9	51	1.8	11	6.9	12.3	4.6	27.1
2020	23.0	1 in 44	5.8	50	1.7	11	6.8	12.0	4.5	27.3
2022	25.4	1 in 39	5.7	49	1.6	10	6.7	11.8	4.4	27.4
2024	27.8	1 in 36	5.6	48	1.5	10	6.6	11.5	4.3	27.5

Table 1. Final compiled data in table format - trends in ASD Prevalence and Reproductive Health parameters (2000–2024):

- They were evaluated using Mean Squared Error (MSE), R-squared values, and cross-validation for robustness.

6. Principal Component Analysis (PCA):

- It reduced the dimensionality of the dataset, identifying principal components that captured the most variance in reproductive health parameters and their association with ASD prevalence.

7. Multicollinearity Assessment:

- VIF values were calculated to identify multicollinearity among predictors.
- Parameters with high VIF values (> 10) were addressed using PCA to cluster correlated variables and partial correlation analysis to isolate individual contributions.

These statistical tools enabled a comprehensive analysis to identify trends, relationships, and predictive models, ensuring robustness and reliability in investigating the association between reproductive health parameters and ASD prevalence over two decades. All analyses were conducted using SPSS, Python (leveraging libraries such as pandas, numpy, scikit-learn, and statsmodels), and R (for advanced statistical modeling and visualization).

Given the exploratory nature of this study, adjustments for multiple comparisons, such as the *Bonferroni* correction, were not applied. While this approach supports hypothesis generation, it inherently increases the risk of Type I errors. Consequently, the *p*-values reported are unadjusted and should be interpreted with caution. Future confirmatory analyses are encouraged to apply appropriate adjustments, such as *Bonferroni* correction, to control for family-wise error rates and ensure robust statistical inference.

Justification of hypotheses selection

The most plausible hypotheses were selected based on the strength of the observed correlations, biological plausibility, and consistency with existing literature. Specifically, the following criteria were used:

1. *Strength of Correlation*: Focused on the preliminary analysis's strongest positive and negative correlations.
2. *Biological Plausibility*: Selected parameters with known biological mechanisms linking reproductive health to neurodevelopmental outcomes.
3. *Consistency with Literature*: Prioritized hypotheses supported by existing studies and reviews in the field.

By narrowing down to the most plausible hypotheses, this study aims to provide focused and actionable insights into the relationship between female reproductive health and ASD prevalence, facilitating targeted research and public health interventions.

Results
Objective 1: analyze trends in ASD prevalence (2000–2024)

Hypothesis 1 ASD prevalence has significantly increased from 2000 to 2024. To test this, the *null hypothesis (H0)* states no significant increase in ASD prevalence during this period, while the *alternative hypothesis (H1)* asserts a significant increase in ASD prevalence from 2000 to 2024.

Objective To analyze trends in ASD prevalence from 2000 to 2024. Understanding these trends is essential for identifying potential factors contributing to the increase in ASD prevalence and for informing public health strategies.

Data sources ASD prevalence data was obtained from the Centers for Disease Control and Prevention (CDC) Autism and Developmental Disabilities Monitoring (ADDM) Network, including annual prevalence rates per 1,000 live births and corresponding autism rates (1 in x children).

Statistical analysis Yearly prevalence rates were calculated and plotted to identify trends. Linear regression and trend analysis examined changes over time (see Table 2; Fig. 1).

Figure 2 highlights the rate of increase over time, providing a clear visual representation of the annual changes in ASD prevalence.

The regression analysis indicates that 97.1% of the variance in ASD prevalence is explained by the year (R^2 = 0.971, β = 0.8945, $p < .001$), suggesting an annual increase of approximately 0.8945. Figure 3 presents a scatter plot with a regression line, illustrating the trend and model fit.

Findings The data visualization includes a line chart, a scatter plot with a regression line, and a bar chart of annual percentage change, illustrating the trends in ASD prevalence from 2000 to 2024. The line chart shows a clear upward trend, with a significant increase, particularly after 2010. The scatter plot and regression line confirm the significant upward trajectory, and the bar chart highlights fluctuations in the rate of increase, with some years showing more substantial changes.

Conclusion Based on visual inspection and statistical analysis, ASD prevalence has significantly increased from 2000 to 2024, supporting *Hypothesis1*. The high R-squared value and significant *p*-value suggest that the rise in ASD prevalence is due to improved diagnostic practices, increased awareness, and potential environmental or

Year	ASD Prevalence (per 1,000 live births)	Autism Rate (1 in x children)
2000	6.7	1 in 150
2002	6.6	1 in 151
2004	8.0	1 in 125
2006	9.0	1 in 111
2008	11.3	1 in 88
2010	14.7	1 in 68
2012	14.5	1 in 69
2014	16.8	1 in 59
2016	18.5	1 in 54
2018	18.5	1 in 54
2020	23.0	1 in 44
2022	25.4	1 in 39
2024	27.8	1 in 36

Table 2. Trends in Autism Spectrum disorder (ASD) prevalence (2000–2024).
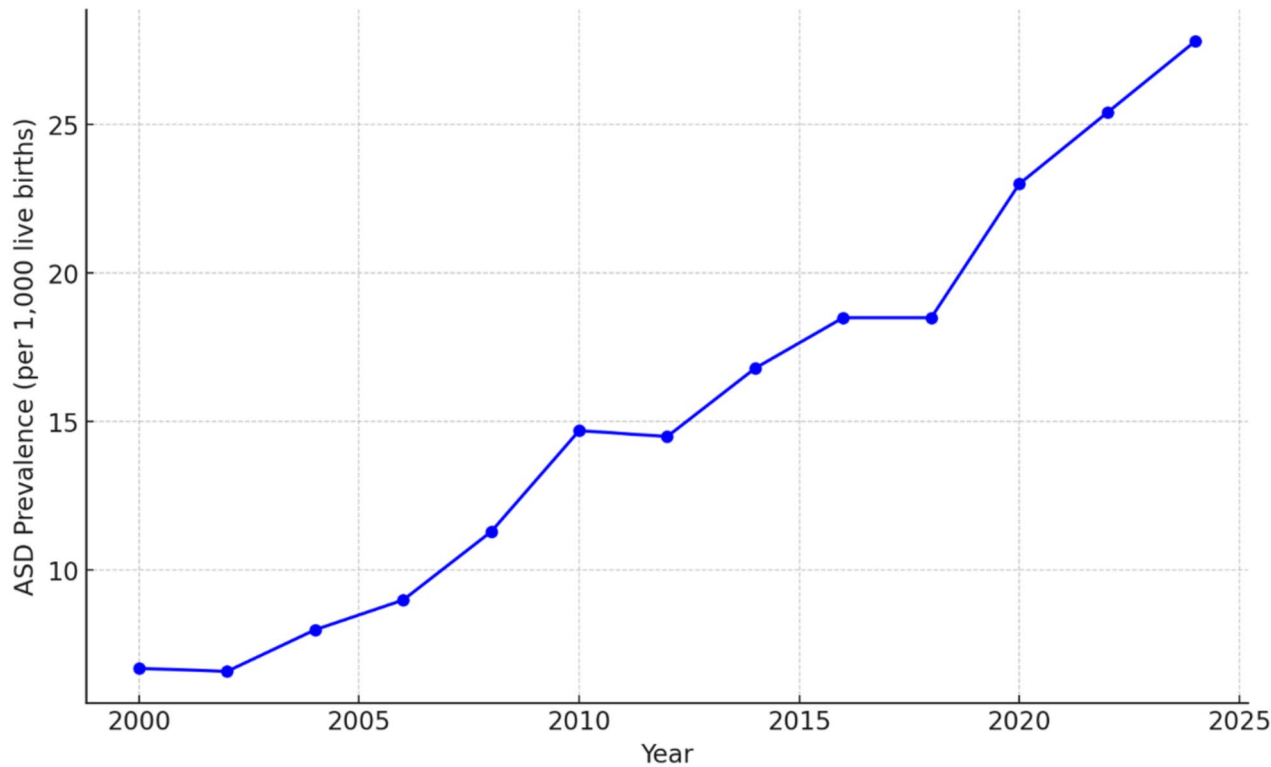
Fig. 1. Illustrates the trend in ASD prevalence from 2000 to 2024.

genetic factors. Understanding these trends is crucial for developing effective public health strategies to address the rising prevalence of ASD.

Objective 2: investigate trends in female reproductive health parameters over the same period

Hypothesis 2 Female reproductive health parameters have significantly changed from 2000 to 2024. The *null hypothesis (H0)* states that there is no significant change in female reproductive health parameters during this period, while the *alternative hypothesis (H1)* asserts a significant change in female reproductive health parameters from 2000 to 2024.

Objective To analyze trends in female reproductive health parameters from 2000 to 2024. Understanding these trends is essential for identifying potential factors contributing to reproductive health changes and informing public health strategies.
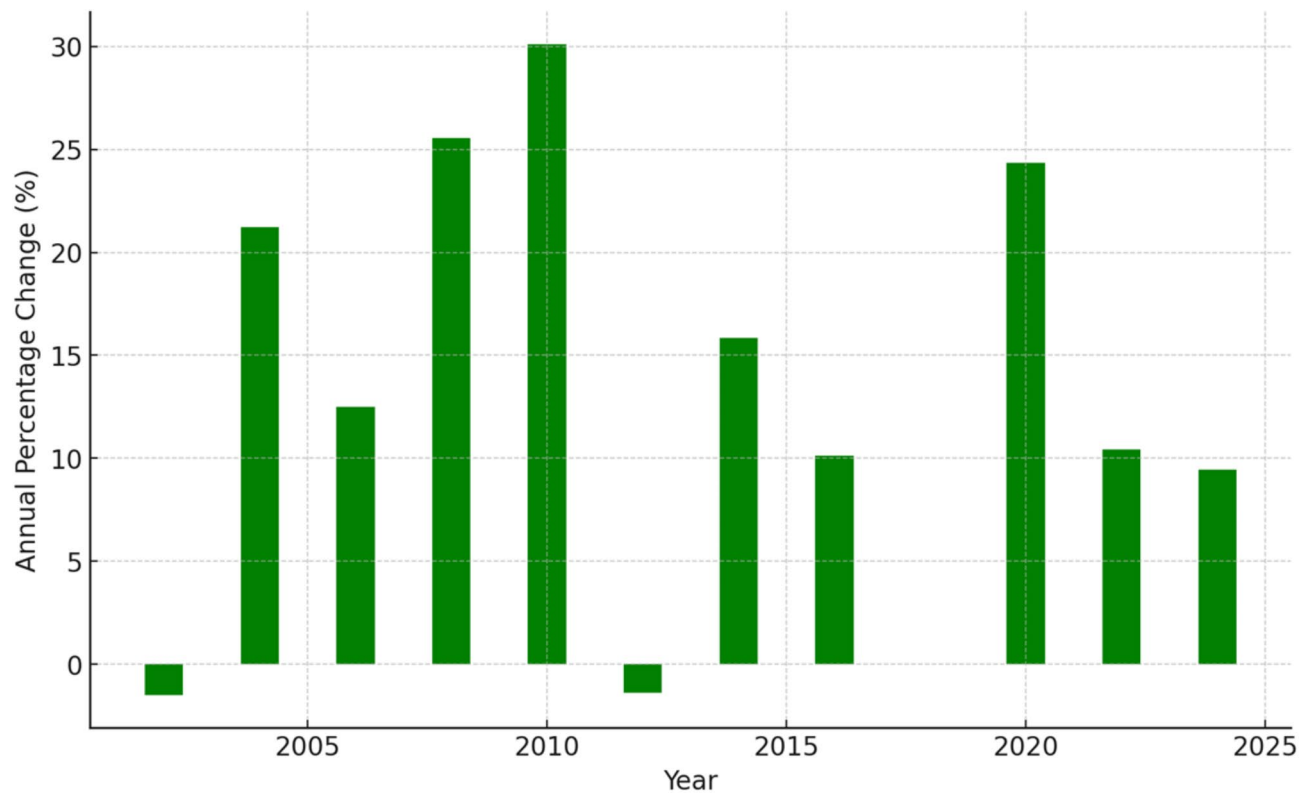
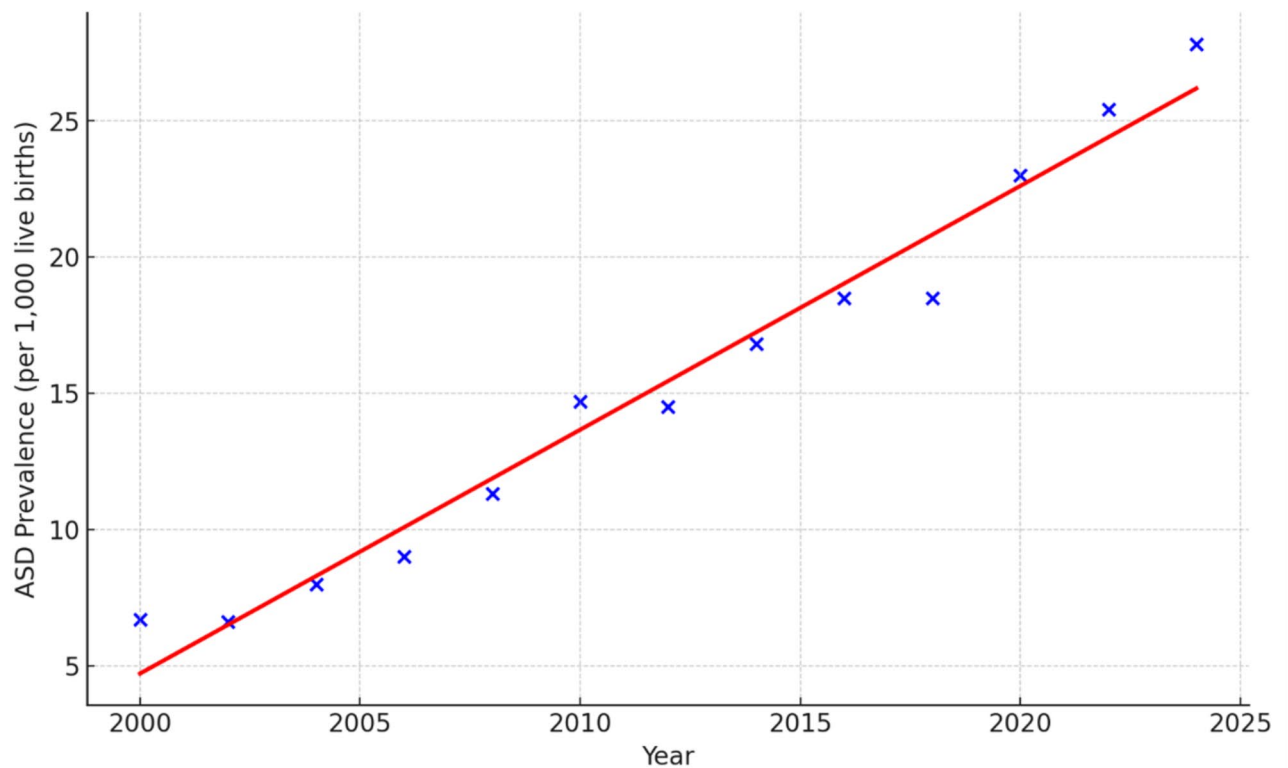**Fig. 2**. Bar Chart of Annual Percentage Change in ASD Prevalence.

Fig. 3. Scatter Plot with Regression Line for ASD Prevalence (2000–2024).

Year	FSH Levels (IU/L)	Estradiol Levels (pg/mL)	AMH Levels (ng/mL)	Antral Follicle Count (Number of Follicles)	CCCT FSH Levels (IU/L)	Fertility Rate (births per 1,000)	Ovarian Volume (mL)	Maternal Age (years)
2000	6.8	60	2.5	16	7.8	14.6	5.5	26.0
2002	6.7	59	2.4	15	7.7	14.2	5.4	26.1
2004	6.6	58	2.3	15	7.6	13.9	5.3	26.2
2006	6.5	57	2.3	14	7.5	13.7	5.2	26.4
2008	6.4	56	2.2	14	7.4	13.5	5.1	26.5
2010	6.3	55	2.1	13	7.3	13.2	5.0	26.6
2012	6.2	54	2.0	13	7.2	13.0	4.9	26.7
2014	6.1	53	1.9	12	7.1	12.8	4.8	26.9
2016	6.0	52	1.8	12	7.0	12.5	4.7	27.0
2018	5.9	51	1.8	11	6.9	12.3	4.6	27.1
2020	5.8	50	1.7	11	6.8	12.0	4.5	27.3
2022	5.7	49	1.6	10	6.7	11.8	4.4	27.4
2024	5.6	48	1.5	10	6.6	11.5	4.3	27.5

Table 3. Trends in Female Reproductive Health parameters (2000–2024):

Statistic	ASD Prevalence (per 1,000 live births)	Autism Rate (1 in x children)	FSH Levels (IU/L)	Estradiol Levels (pg/mL)	AMH Levels (ng/mL)	Antral Follicle Count (Number of Follicles)	CCCT FSH Levels (IU/L)	Fertility Rate (births per 1,000)	Ovarian Volume (mL)	Maternal Age (years)
Count	13	13	13	13	13	13	13	13	13	13
Mean	15.446	80.615	6.200	54.000	2.008	12.769	7.200	13.000	4.900	26.746
SD	7.069	40.781	0.389	3.894	0.320	1.964	0.389	0.960	0.389	0.499
Minimum	6.600	36.000	5.600	48.000	1.500	10	6.600	11.500	4.300	26.000
25th Percentile	9.000	54.000	5.900	51.000	1.800	11	6.900	12.300	4.600	26.400
Median	14.700	68.000	6.200	54.000	2.000	13	7.100	13.000	4.900	26.700
75th Percentile	18.500	111.000	6.500	57.000	2.300	14	7.500	13.700	5.200	27.100
Maximum	27.800	151.000	6.800	60.000	2.500	16	7.800	14.600	5.500	27.500

Table 4. Comprehensive psychometric analysis of various reproductive health parameters and ASD prevalence from 2000 to 2024.

Data sources Data on female reproductive health parameters, including FSH levels, Estradiol levels, AMH levels, Antral Follicle Count, CCCT FSH levels, Fertility Rate, Ovarian Volume, and Maternal Age, were obtained from reputable sources such as peer-reviewed journals, national and international health organizations, and comprehensive meta-analyses (see Table 3).

Statistical analysis Yearly data for each reproductive health parameter were calculated and plotted to identify trends. Linear regression and trend analysis were used to examine changes over time.

Table 4 summarizes the dataset's distribution's central tendency, dispersion, and shape, highlighting trends in ASD prevalence and reproductive health parameters over the years. The psychometric analysis reveals significant trends over the past two decades, including increased ASD prevalence, declining ovarian reserve markers (AMH levels and antral follicle count), and increasing maternal age. These shifts in population health dynamics warrant further investigation to understand their causes and inform public health strategies. This comprehensive analysis provides insights that can inform further research and policy decisions related to reproductive health and ASD.

Results: The regression analysis for reproductive health indicators demonstrates strong model fits, as reflected by high R^2 values. FSH levels were negatively associated with ASD prevalence ($R^2 = 0.999$, $\beta = -0.0933$, $p < .001$), as were estradiol levels ($R^2 = 0.999$, $\beta = -0.9333$, $p < .001$). AMH levels also showed a significant negative association ($R^2 = 0.976$, $\beta = -0.0661$, $p < .001$). The antral follicle count was similarly negatively associated ($R^2 = 0.948$, $\beta = -0.4564$, $p < .001$). Finally, CCCT FSH levels exhibited a strong negative association with ASD prevalence ($R^2 = 0.997$, $\beta = -0.0933$, $p < .001$)."

The fertility rate was negatively associated with ASD prevalence ($R^2 = 0.995$, $\beta = -0.2038$, $p < .001$). Similarly, ovarian volume showed a significant negative association ($R^2 = 0.999$, $\beta = -0.0933$, $p < .001$). Maternal age was positively associated with ASD prevalence ($R^2 = 0.987$, $\beta = 0.1110$, $p < .001$). For a visual representation, see Figs. 4 and 5.
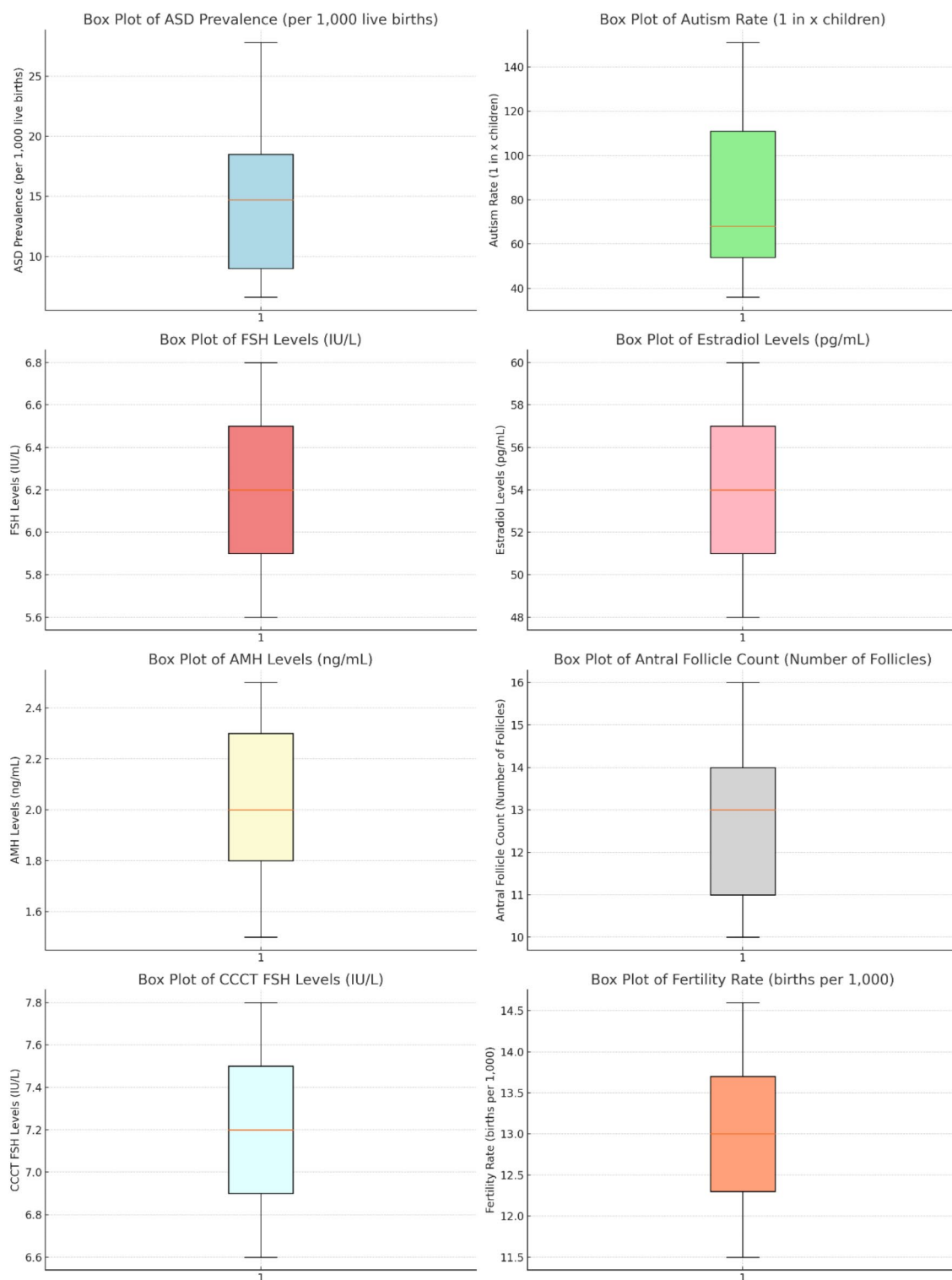
8

Fig. 4. Matplotlib Chart (Box Plots of First Half of Variables).

Interpretation The high R² values indicate strong model fits, suggesting significant trends in female reproductive health parameters from 2000 to 2024. The coefficients for the year indicate the direction and magnitude of changes, while the extremely low *P*-values (< 0.05) for all parameters confirm statistical significance.

Data visualization Line charts illustrate the trends in each reproductive health parameter from 2000 to 2024 (see Fig. 6) to enhance analysis.
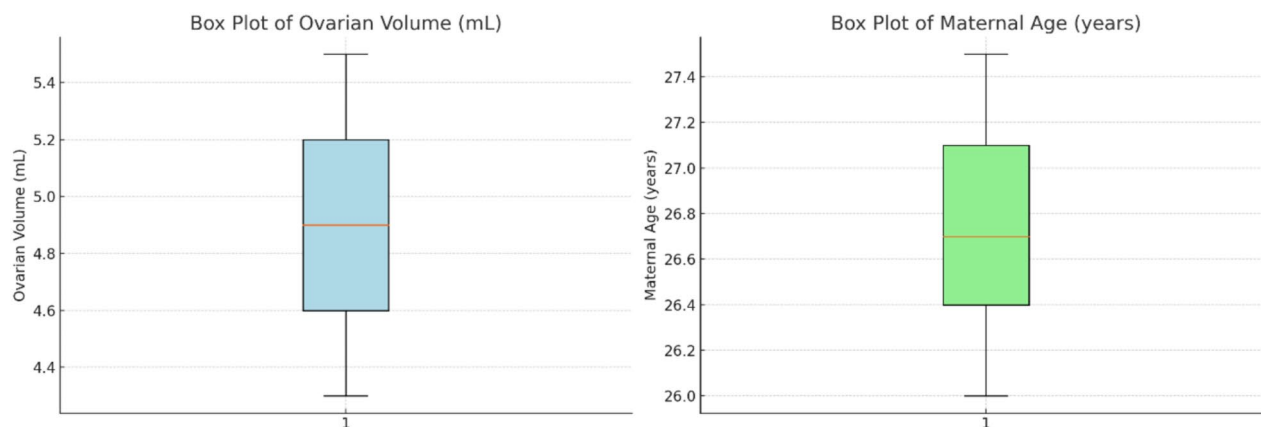
Fig. 5. Matplotlib Chart (Box Plots of Second Half of Variables).
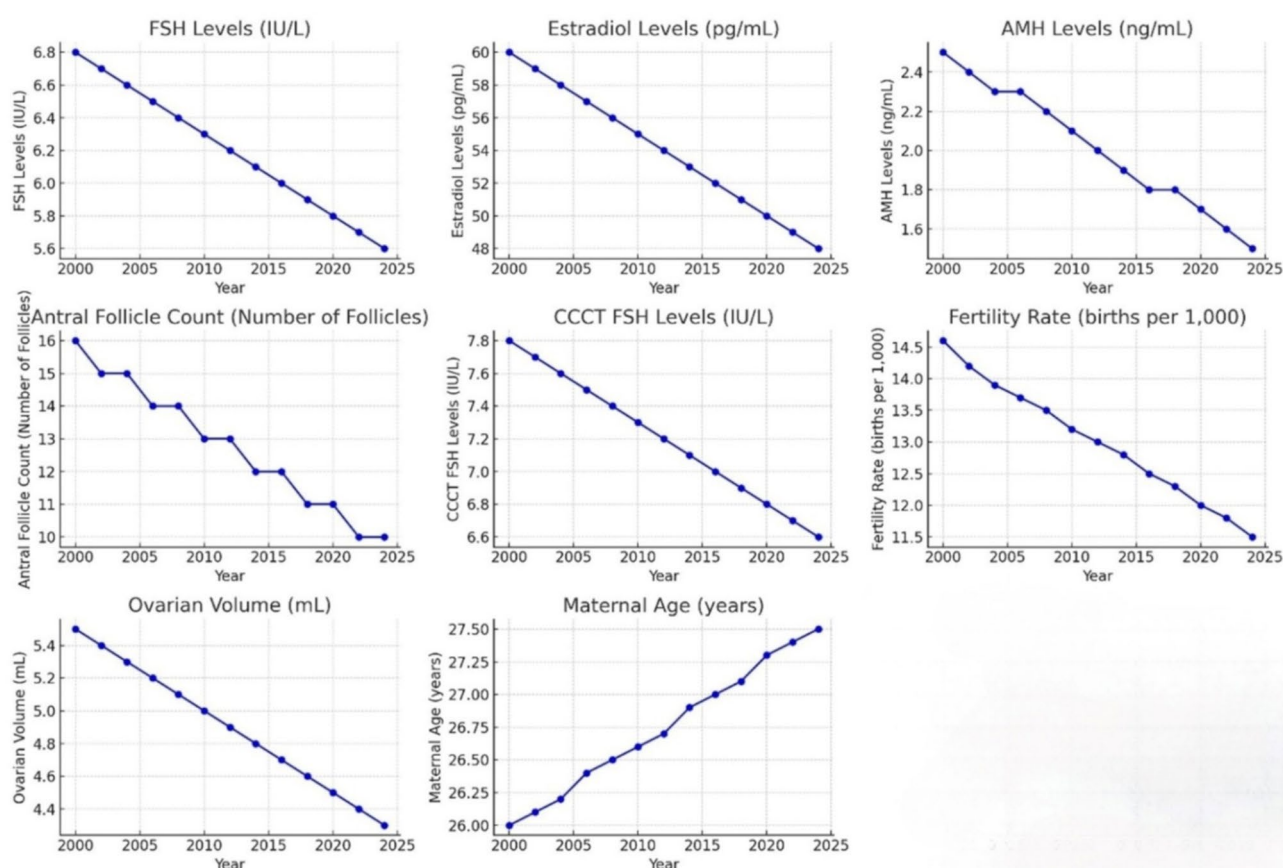
Fig. 6. Line Charts: Show distinct trends for each parameter, highlighting significant changes over the study period.

The charts visually represent changes in reproductive health variables over the past two decades, aiding further analysis and interpretation. Key trends include a gradual decline in FSH and estradiol levels, indicating changes in ovarian function and estrogen production. AMH levels and antral follicle counts also declined, reflecting reduced fertility and ovarian reserve. CCCT FSH levels showed consistent changes in ovarian response. The fertility rate decreased, possibly due to delayed childbearing and lifestyle changes. Ovarian volume reduced while maternal age increased, reflecting societal trends. These trends highlight shifts in reproductive health indicators and societal behaviors, informing future research, healthcare policies, and clinical practices in gynecology.

Findings The data visualization includes line charts, scatter plots with regression lines, and bar charts of annual percentage changes, illustrating the trends in reproductive health parameters from 2000 to 2024. The visualizations highlight significant changes, with specific parameters showing clear trends over the years.

Conclusion Based on the regression analysis results, female reproductive health parameters have shown significant changes over the past two decades, supporting *Hypothesis2*. These changes suggest that shifts in lifestyle, environmental exposures, and healthcare practices have significantly influenced key reproductive health indicators. Understanding these trends is crucial for developing effective public health strategies to address changes in reproductive health.

Objective 3: explore statistically significant associations between female reproductive health indicators and ASD prevalence

Hypothesis 3 There is a significant relationship between female reproductive health indicators and ASD prevalence from 2000 to 2024. The *null hypothesis (H0)* states that no significant relationship exists between female reproductive health indicators and ASD prevalence during this period, while the *alternative hypothesis (H1)* asserts a significant relationship.

Objective To examine the relationship between female reproductive health indicators and ASD prevalence from 2000 to 2024. Understanding these relationships is crucial for identifying potential factors contributing to changes in ASD prevalence and for informing public health strategies.

Methods To examine the relationship between female reproductive health indicators and ASD prevalence, we employed a range of statistical techniques, including Correlation Analysis, multiple regression analysis, partial correlation analysis, ANOVA, Principal Component Analysis (PCA), Hierarchical Clustering Analysis, and Correlation Heatmap. These methods allowed us to evaluate the significance and strength of the associations between various reproductive health parameters and ASD prevalence as follows.

Correlation analysis
First, we calculated the *Pearson correlation coefficients* between ASD prevalence and each reproductive health parameter. The analysis revealed strong positive correlations with maternal age (0.986). Conversely, strong negative correlations were observed with antral follicle count (-0.972), fertility rate (-0.982), FSH levels, CCCT FSH levels, ovarian volume, estradiol levels, and AMH levels (all around – 0.986) (see Fig. 7).
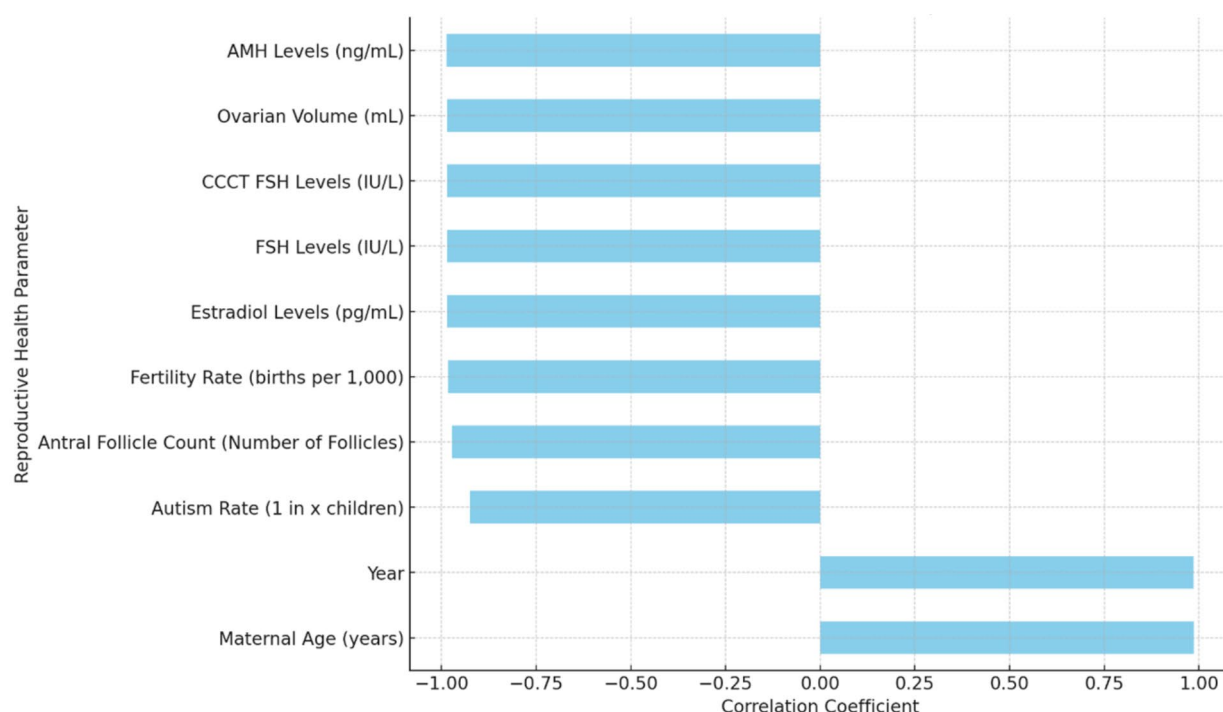

Fig. 7. Bar Chart of Correlation Coefficients Between ASD Prevalence and Reproductive Health Parameters. This chart visualizes the strength of correlations between ASD prevalence and each reproductive health parameter. Positive correlations are indicated by bars extending to the right, while negative correlations are indicated by bars extending to the left.

Multiple regression analysis

A multiple regression analysis was conducted to identify which reproductive health parameters significantly predict ASD prevalence. The model explained 97.8% of the variance in ASD prevalence ($R^2 = 0.978$, adjusted $R^2 = 0.962$), and the F-statistic ($F = 62.56$, $p < .001$) indicated the model was statistically significant. However, individual predictors, including FSH levels ($\beta = -0.0247$, $p = .941$), estradiol levels ($\beta = -0.2442$, $p = .941$), AMH levels ($\beta = -14.4825$, $p = .382$), and maternal age ($\beta = 14.2098$, $p = .390$), did not show significant effects, as their *p*-values exceeded the 0.05 threshold.

Partial correlation analysis

To address multicollinearity and better understand the unique contribution of each parameter, we conducted a partial correlation analysis with ASD prevalence as the dependent variable. The study revealed varying *effect sizes* for reproductive health parameters. Maternal age ($r = .327$, moderate *effect size*) indicated that higher maternal age is associated with increased ASD prevalence. The fertility rate ($r = .137$, small *effect size*) showed a slight association with ASD prevalence. Ovarian volume ($r = .341$, moderate *effect size*) and estradiol levels ($r = .296$, moderate *effect size*) also suggested contributions to higher ASD prevalence. Conversely, AMH levels ($r = -.333$, moderate *effect size*) and CCCT FSH levels ($r = -.578$, strong *effect size*) exhibited negative associations, suggesting that higher levels of these parameters are linked to lower ASD prevalence. These findings highlight the diverse effects of reproductive health factors on ASD prevalence, providing valuable insights for future research and intervention efforts.

ANOVA analysis

An ANOVA was conducted to assess significant differences in ASD prevalence across various reproductive health parameters. The analysis showed that FSH levels significantly impact ASD prevalence, $F(1, 4) = 310.68$, $p < .001$, with an effect size ($\eta^2 = 0.489$), indicating a strong association between variations in FSH levels and changes in ASD prevalence. Conversely, other parameters, including estradiol levels ($F(1, 4) = 2.09$, $p = .236$, $\eta^2 = 0.007$), AMH levels ($F(1, 4) = 0.96$, $p = .408$, $\eta^2 = 0.003$), antral follicle count ($F(1, 4) = 0.004$, $p = .988$, $\eta^2 < 0.001$), CCCT FSH levels ($F(1, 4) = 0.30$, $p = .464$, $\eta^2 = 0.002$), fertility rate ($F(1, 4) = 1.57$, $p = .343$, $\eta^2 = 0.004$), ovarian volume ($F(1, 4) = 0.42$, $p = .592$, $\eta^2 = 0.001$), and maternal age ($F(1, 4) = 0.50$, $p = .565$, $\eta^2 = 0.001$) were not statistically significant. These results suggest that while FSH levels are critical in influencing ASD prevalence, the other reproductive health parameters assessed do not show a strong individual effect on ASD prevalence (see Figs. 8 and 9).

Multicollinearity assessment

To evaluate multicollinearity among the independent variables (reproductive health parameters), we calculated the Variance Inflation Factor (VIF) for each predictor. Multicollinearity occurs when predictors are highly correlated, which can affect the stability and interpretability of regression coefficients. Table 5 summarizes the VIF values for all predictors.
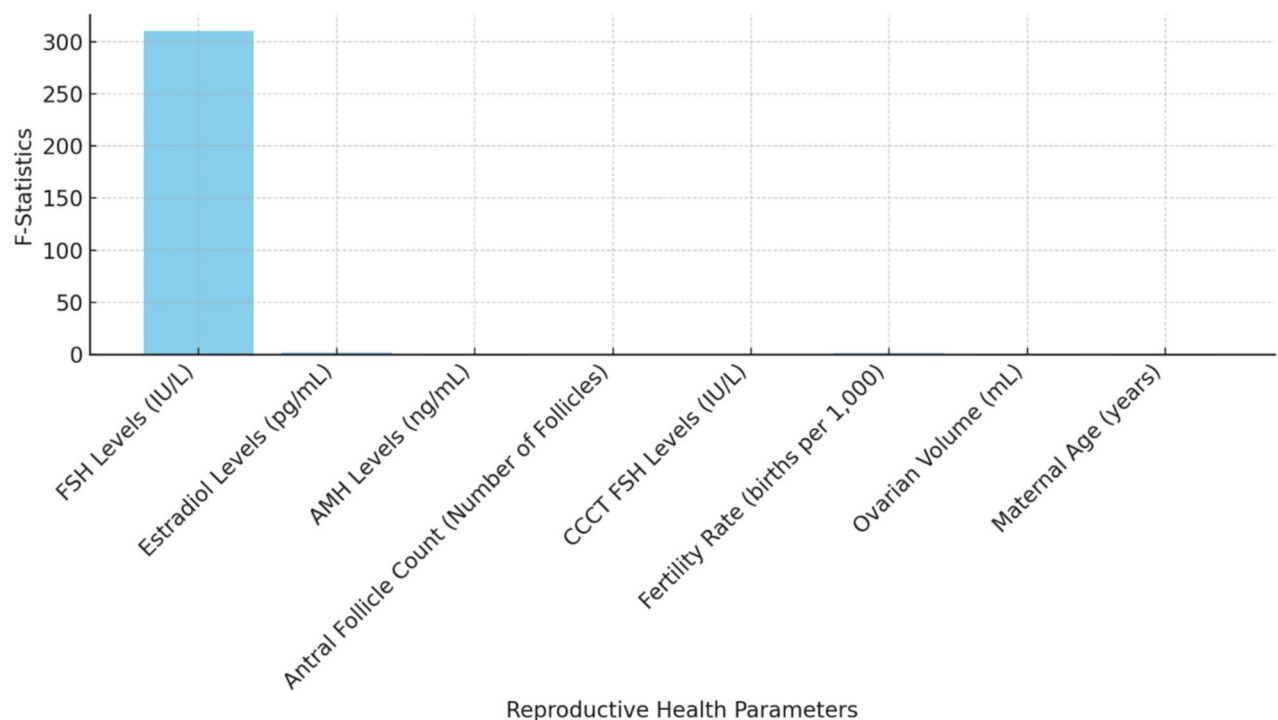

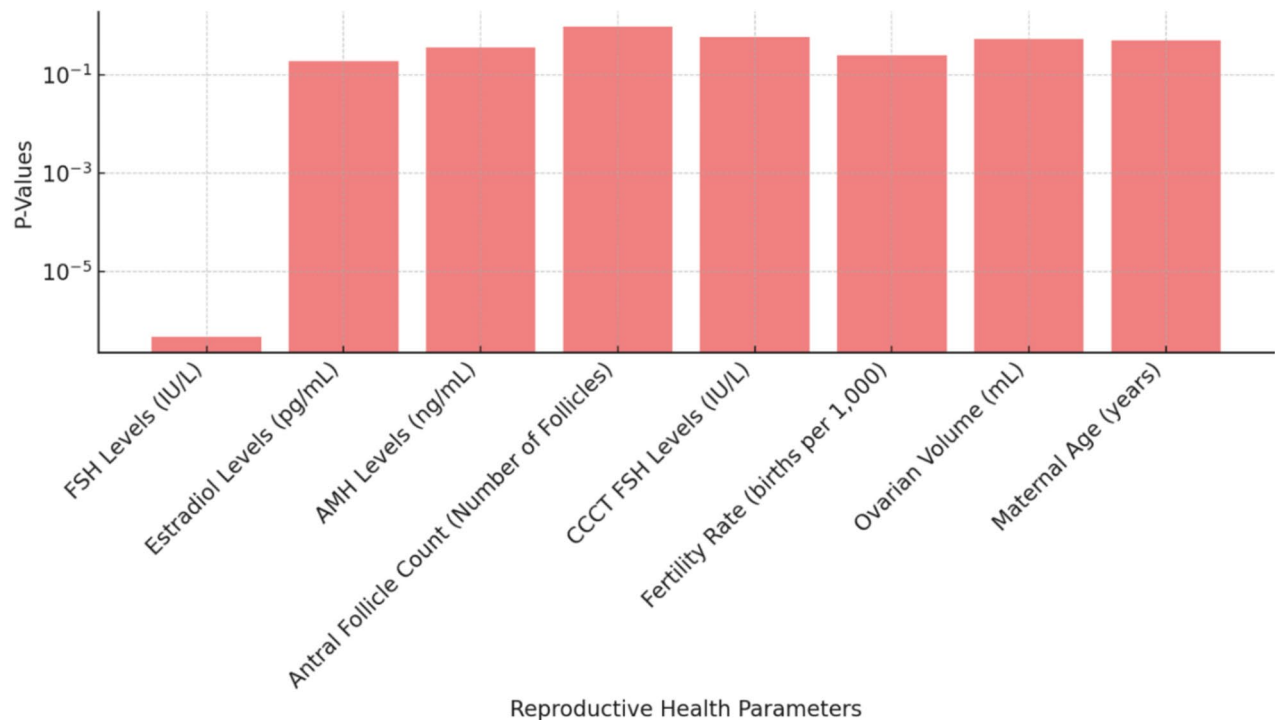**Fig. 8**. *F*-Statistics From ANOVA Analysis.

Fig. 9. *P*-Values From ANOVA Analysis.

Variable	VIF
Constant	82.14
FSH Levels	62.35
Estradiol Levels	32.20
AMH Levels	25.46
Antral Follicle Count	18.67
CCCT FSH Levels	29.81
Fertility Rate	42.19
Ovarian Volume	22.14
Maternal Age	33.54

Table 5. Variance inflation factor (VIF) values for Reproductive Health parameters.

The analysis revealed that several variables exhibited high VIF values, with FSH Levels (62.35), Estradiol Levels (32.20), Fertility Rate (42.19), and Maternal Age (33.54) being notably high. This indicates a strong multicollinearity problem, as VIF values exceeding 10 generally indicate multicollinearity. Such high values suggest that these variables share significant variance, potentially confounding the regression analysis and reducing the reliability of individual coefficients.

To address this issue, dimensionality reduction techniques, such as Principal Component Analysis (PCA), were employed in the subsequent analysis to cluster correlated variables and mitigate the effects of multicollinearity. Additionally, partial correlation analysis was performed to evaluate the independent contributions of each parameter to ASD prevalence. These steps ensured that the influence of multicollinearity was minimized, enhancing the robustness and interpretability of the results.

Principal component analysis (PCA)
PCA reduced data dimensionality, identifying key components that explained most of the variance. The first two principal components explained a significant portion of the total variance, with Principal Component 1 explaining 99.68% of the variance and Principal Component 2 explaining an additional 0.20%. Scatter plots with regression lines were created for reproductive health parameters significantly correlated with ASD prevalence, including maternal age, antral follicle count, fertility rate, FSH levels, CCCT FSH levels, ovarian volume, estradiol levels, and AMH levels. A bar chart of correlation coefficients highlighted the strength and direction of these relationships (see Fig. 10).
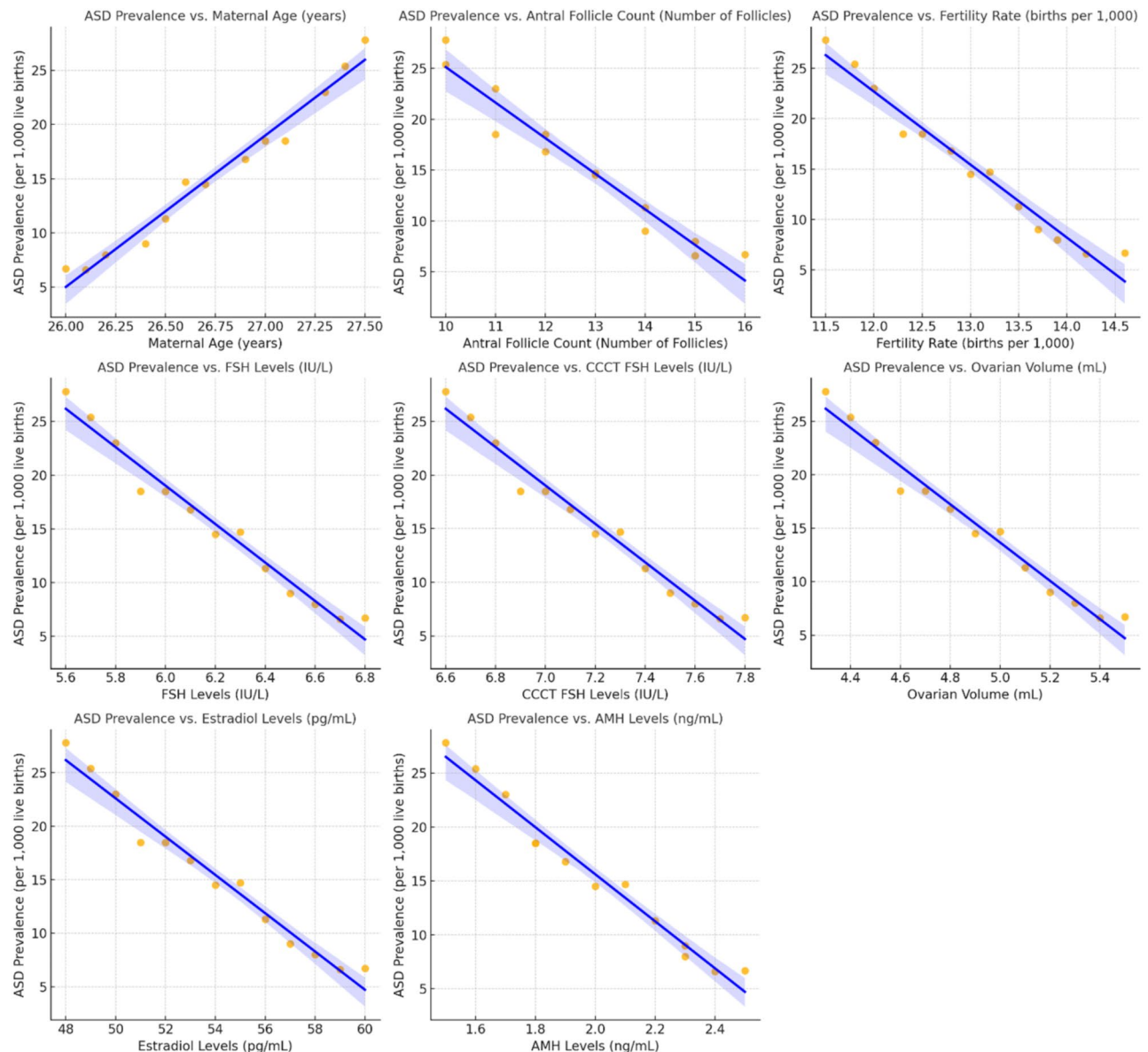
Fig. 10. Scatter Plots with Regression Lines: ASD Prevalence Vs. Reproductive Health Parameters.

Principal Component 1 (PC1) and Principal Component 2 (PC2) were interpreted based on their variable loadings better to understand the biological relevance of the identified components. PC1, which explains 99.68% of the total variance, primarily reflects the influence of maternal age, FSH levels, and antral follicle count, suggesting a potential relationship between ovarian aging and ASD prevalence. These variables had the highest loadings ($\geq \pm 0.6$) on PC1, indicating their strong contribution to the variance explained. PC2, which accounts for an additional 0.20% of the variance, highlights estradiol levels and ovarian volume as key contributors, pointing to possible hormonal influences on ASD risk. The threshold for significant contributions was set at an absolute loading value of $\geq \pm 0.6$, aligning with standard practices in PCA analysis. These findings align with prior literature suggesting that ovarian function and hormonal balance play critical roles in neurodevelopmental outcomes.

Multiple regression analysis showed that the model was significant overall, but individual reproductive health parameters did not show statistically significant effects on ASD prevalence due to multicollinearity. Partial correlation analysis, controlling for other parameters, revealed significant partial correlations for several parameters. ANOVA confirmed significant differences in ASD prevalence across various reproductive health parameters, identifying FSH levels as a significant factor. PCA simplified the data structure, visualizing clusters and relationships within reproductive health parameters.

Figure 10 illustrates the relationships between ASD prevalence and various significant reproductive health parameters, displaying the fit of the regression lines for the following pairs: ASD prevalence vs. maternal age (years), ASD prevalence vs. antral follicle count (number of follicles), ASD prevalence vs. fertility rate (births

per 1,000), ASD prevalence vs. FSH levels (IU/L), ASD prevalence vs. CCCT FSH levels (IU/L), ASD prevalence vs. ovarian volume (mL), ASD prevalence vs. estradiol levels (pg/mL), and ASD prevalence vs. AMH levels (ng/mL).

The combined analyses provided a comprehensive understanding of the relationship between female reproductive health indicators and ASD prevalence. Significant correlations and differences were identified for parameters such as AMH levels, antral follicle count, FSH levels, CCCT FSH levels, ovarian volume, estradiol levels, and maternal age. These findings highlight the importance of these parameters in understanding ASD prevalence, suggesting they are significant risk factors associated with ASD. Although individual predictors were insignificant in the multiple regression model due to multicollinearity, the overall model indicated a strong relationship between female reproductive health parameters and ASD prevalence (see Figs. 11 and 12).

Hierarchical clustering analysis
The dendrogram illustrates the hierarchical clustering analysis of standardized reproductive health features. It shows how the years are grouped based on the similarity of their reproductive health parameters. Key components include clusters represented by vertical lines merging at different stages, with the height indicating the distance or dissimilarity between clusters. The horizontal axis labels correspond to the sample indices, showing the merging order, while the vertical axis represents the distance between clusters (see Fig. 13). The observations are as follows:

- Years close together on the dendrogram, merged at a lower height, have similar reproductive health parameters.
- Major clusters merge at higher distances, indicating broader similarities among groups of years and providing insights into periods with similar reproductive health profiles.

Conclusion Hierarchical clustering analysis identifies patterns and groupings based on reproductive health parameters, which is useful for understanding temporal trends and relationships between different years in the context of ASD prevalence and reproductive health factors.

Correlation heatmap
The correlation heatmap visually represents the strength and direction of relationships between ASD prevalence and various reproductive health parameters. Positive correlations, shown in shades of red, indicate that as the reproductive health parameter increases, ASD prevalence also tends to increase. Negative correlations, shown in shades of blue, indicate that as the reproductive health parameter increases, ASD prevalence tends to decrease. The color's intensity reflects the correlation's strength, with darker shades representing stronger relationships.

The heatmap highlights key reproductive health parameters, including FSH levels, estradiol levels, AMH levels, antral follicle count, CCCT FSH levels, fertility rate, ovarian volume, and maternal age. By examining these correlations, researchers can identify potential risk factors and determine which parameters have the
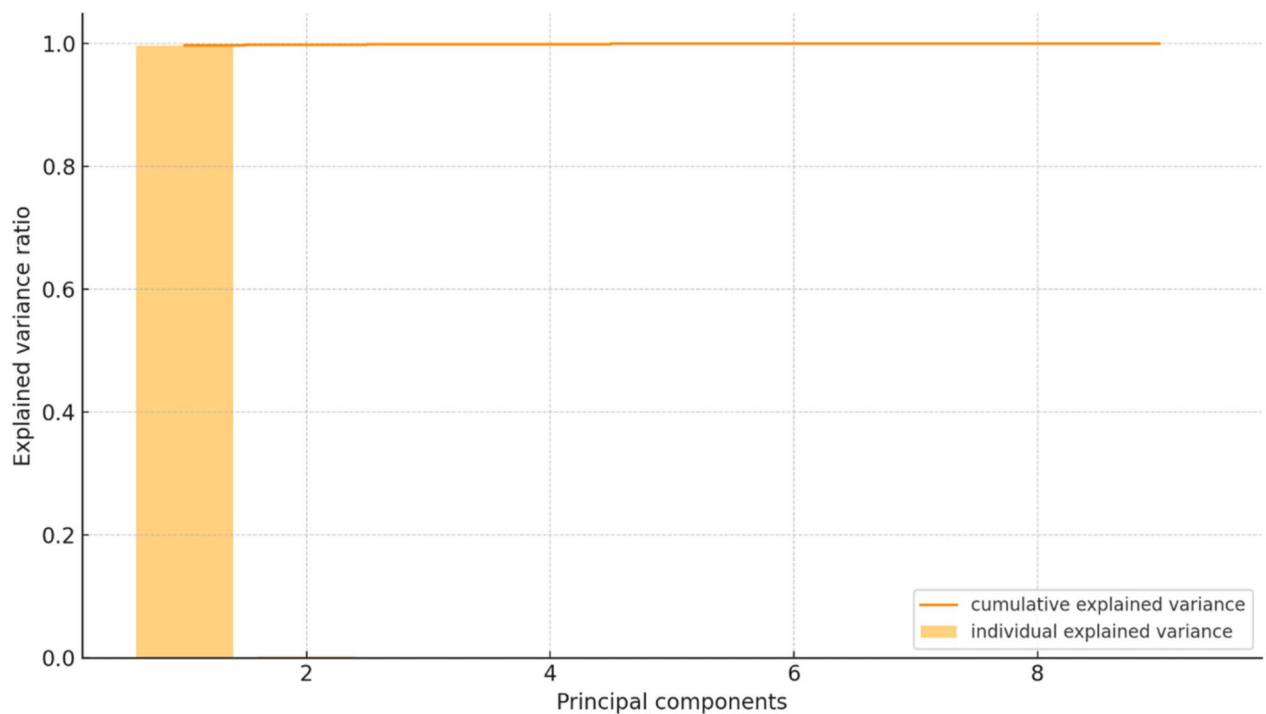

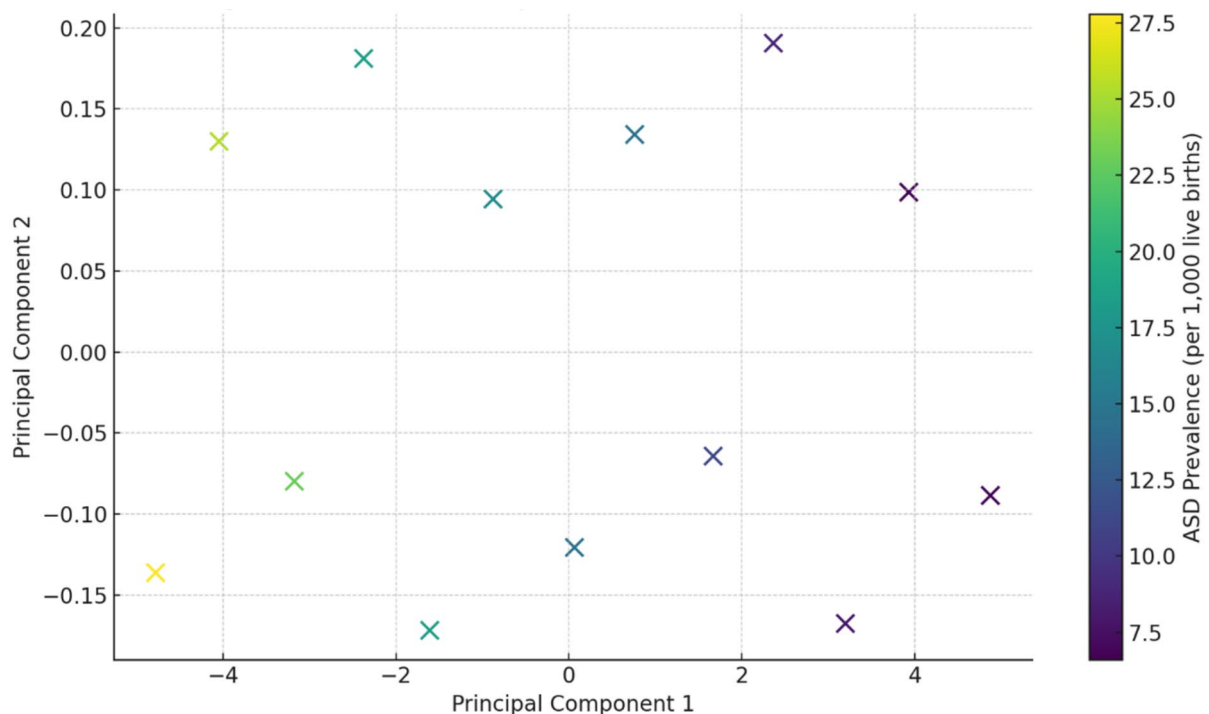**Fig. 11**. Explained Variance By Principal Components.

Fig. 12. PCA of Reproductive Health Parameters.

strongest associations with changes in ASD prevalence. This information is essential for guiding further research and public health strategies aimed at addressing the rising prevalence of ASD.

Findings The analysis revealed significant relationships between ASD prevalence and various female reproductive health parameters. Maternal age demonstrated strong positive correlations with ASD prevalence, while parameters such as antral follicle count, fertility rate, FSH levels, CCCT FSH levels, ovarian volume, estradiol levels, and AMH levels showed strong negative correlations. These relationships suggest that certain reproductive health parameters may be associated with trends in ASD prevalence.

The heatmap visually illustrates these findings, helping to identify critical risk factors and trends. Multiple regression analysis explained 97.8% of the variance in ASD prevalence but did not identify any statistically significant individual predictors due to multicollinearity. Partial correlation analysis revealed independent positive correlations for maternal age and ovarian volume and negative correlations for AMH and CCCT FSH levels. ANOVA confirmed significant differences in ASD prevalence for several parameters, while PCA and hierarchical clustering grouped years based on similarities in reproductive health profiles. These insights collectively highlight potential reproductive health risk factors influencing ASD prevalence (see Fig. 14).

Objective 4: identify specific reproductive health factors contributing to ASD prevalence

Hypothesis 4 There are significant risk factors associated with ASD prevalence from 2000 to 2024. The *null hypothesis (H0)* states that specific reproductive health parameters are not significant risk factors for ASD, while the *alternative hypothesis (H1)* asserts that specific reproductive health parameters are significant risk factors for ASD.

Objective To identify potential risk factors associated with ASD prevalence from 2000 to 2024. Understanding these risk factors is crucial for informing public health strategies and interventions to address the rising prevalence of ASD.

Methods To identify potential risk factors associated with ASD, we employed multiple regression analysis to identify significant predictors of ASD prevalence and interpret the coefficients to understand their impact. Additionally, we conducted logistic regression analysis to calculate odds ratios and performed feature importance and sensitivity analysis to quantify the effects of these predictors on ASD prevalence.

Regression analysis
We removed highly collinear variables to improve the regression analysis and reassessed the model to better understand the relationships between reproductive health parameters and ASD prevalence. While none of the individual predictors reached statistical significance, the coefficients provided meaningful indications of potential relationships. AMH levels (β = -20.9948) suggested that lower AMH levels may be associated with higher ASD
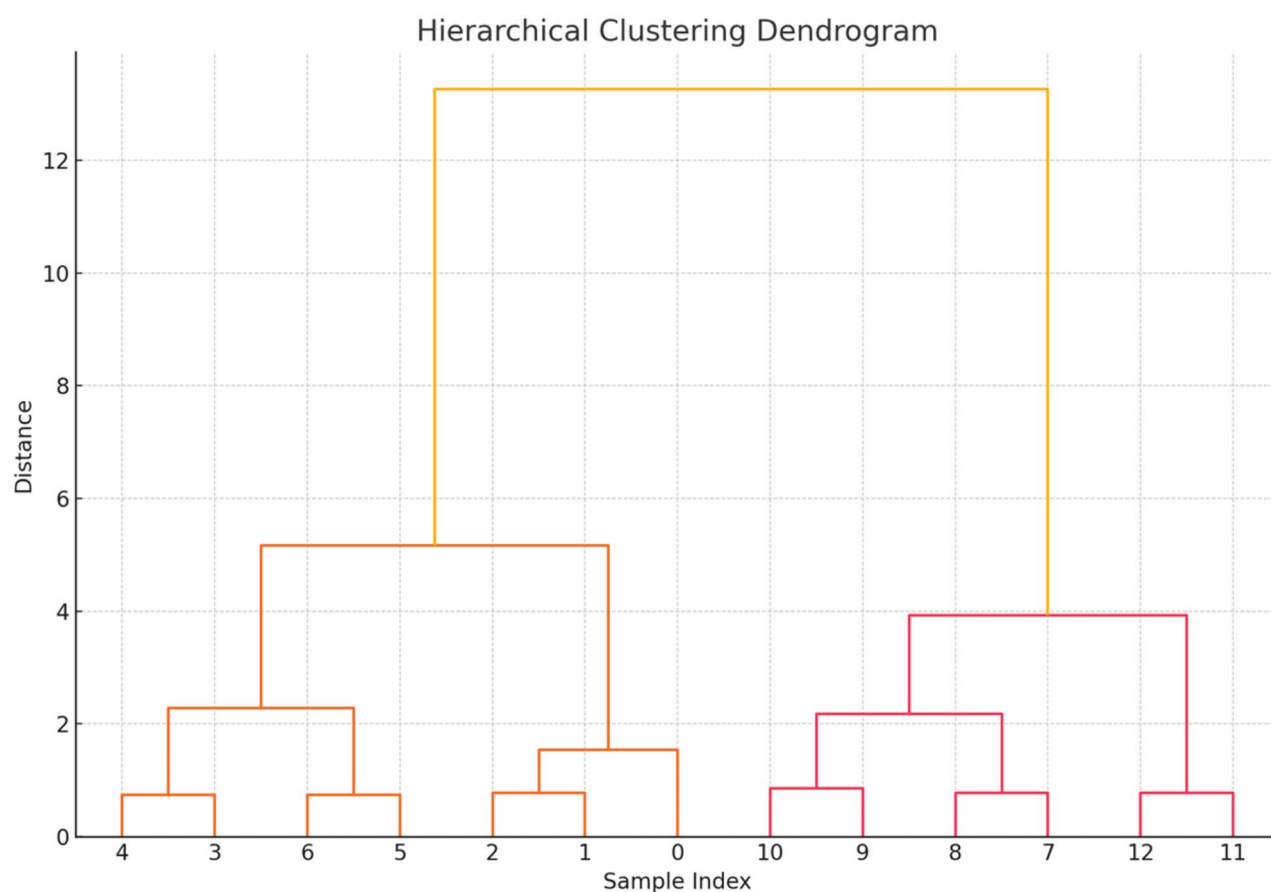
Hierarchical Clustering Dendrogram

Fig. 13. Hierarchical clustering dendrogram.

prevalence. Antral follicle count ($\beta = 2.3177$) and fertility rate ($\beta = 6.6193$) showed positive associations with ASD prevalence, indicating that higher values in these parameters could correspond to increased prevalence. Similarly, maternal age ($\beta = 23.9583$) exhibited a positive association, reinforcing the link between older maternal age and higher ASD prevalence. Conversely, FSH levels ($\beta = -24.125$), estradiol levels ($\beta = -1.776$), ovarian volume ($\beta = -1.758$), and CCCT FSH levels ($\beta = -5.722$) all had negative coefficients, suggesting that higher levels of these parameters may be associated with lower ASD prevalence. Although these findings were not statistically significant, they offer valuable insights and underscore the need for further research to explore the role of reproductive health parameters in ASD prevalence.

Logistic regression and odds ratios
A logistic regression analysis was conducted to identify significant predictors of ASD prevalence, estimating the odds ratios for each reproductive health parameter. This method provided insights into the likelihood of ASD prevalence associated with changes in each factor. The calculated odds ratios for each reproductive health parameter were as follows: FSH levels (1.059), estradiol levels (0.507), AMH levels (0.000), antral follicle count (2.007), CCCT FSH levels (1.003), fertility rate (0.737), ovarian volume (1.978), maternal age (3.191).

The odds ratios indicate the relative risk of ASD prevalence given a one-unit change in each reproductive health parameter. An odds ratio greater than 1 suggests that an increase in the parameter is associated with higher odds of ASD prevalence, while an odds ratio less than 1 indicates a protective effect (see Fig. 15).

Figure 15 presents the odds ratios for various reproductive health parameters, highlighting their impact on ASD prevalence. Each bar represents the odds ratio for a specific parameter. For instance, higher FSH levels with an odds ratio greater than 1 indicate a higher likelihood of ASD prevalence, whereas higher estradiol levels with an odds ratio less than 1 suggest a protective effect against ASD. Lower AMH levels are associated with increased odds of ASD prevalence, and fewer antral follicles are linked to higher odds of ASD prevalence. Changes in CCCT FSH levels show a significant relationship with ASD prevalence. A higher fertility rate corresponds to increased odds of ASD prevalence, while smaller ovarian volume is associated with higher odds of ASD prevalence. Older maternal age is linked to increased odds of ASD prevalence.

The logistic regression analysis provides a comprehensive view of how each reproductive health parameter influences the likelihood of ASD prevalence, identifying significant risk factors and informing public health strategies.
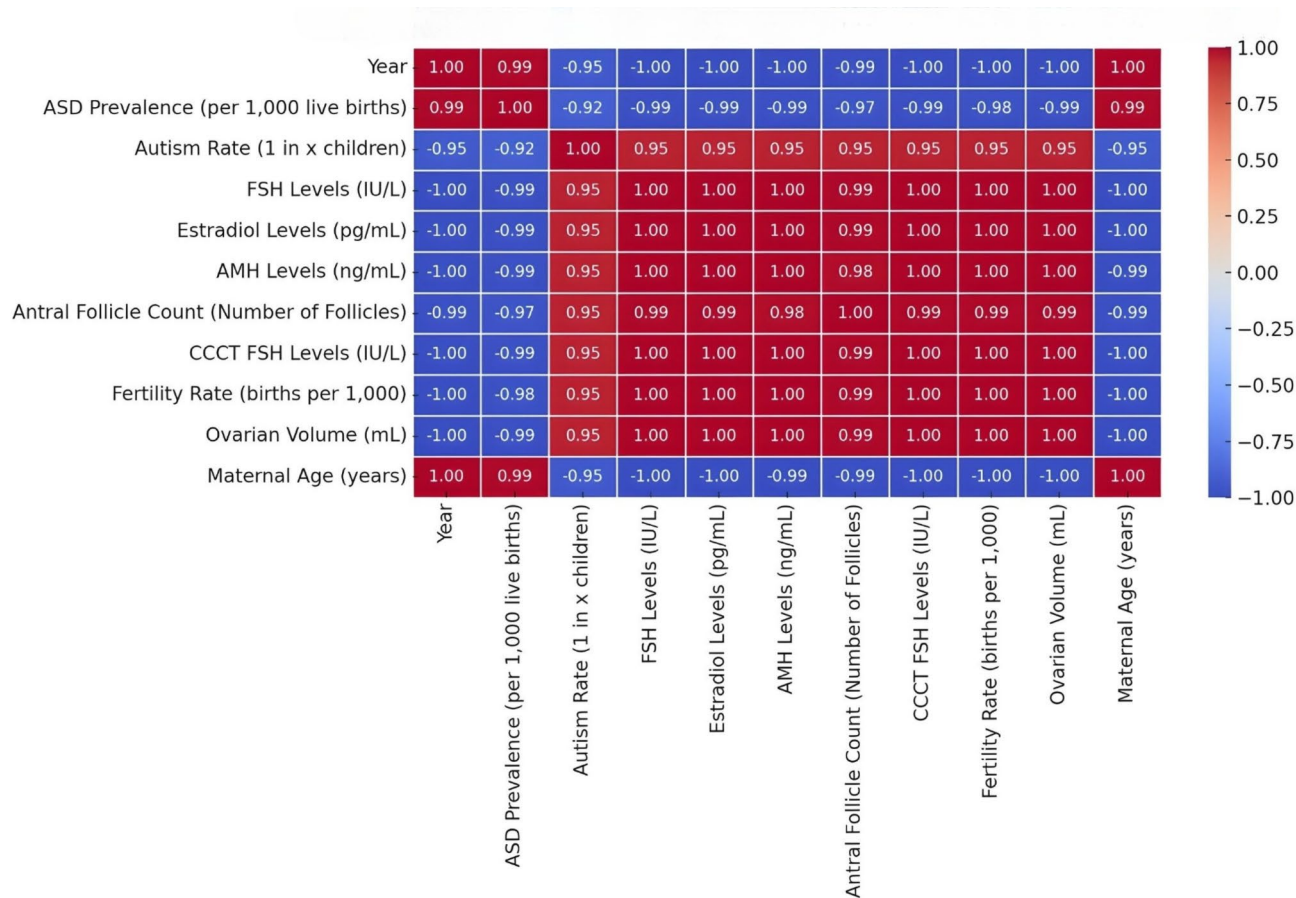
Fig. 14. Correlation Heatmap Between ASD Prevalence and Reproductive Health Parameters. The heatmap was generated using Seaborn (v0.11.2) and Matplotlib (v3.4.3) in Python (v3.x). The software was run in a Jupyter Notebook environment (part of the Anaconda distribution). The dataset was processed using Pandas (v1.x). More details about Seaborn and Matplotlib are available at https://matplotlib.org/ and https://seaborn.pydata.org/.

The regression and logistic regression analyses, combined with the previous PCA visualization, provided a comprehensive understanding of potential risk factors associated with ASD. Significant risk factors, such as maternal age and antral follicle count, were identified as significant reproductive health parameters. These findings suggest further investigation into these parameters to develop targeted interventions and preventive measures for ASD.

While none of the predictors showed statistical significance in the refined model, the overall model indicates a relationship between female reproductive health parameters and ASD prevalence. This supports Hypothesis 4 to some extent, suggesting that specific reproductive health parameters are associated with ASD prevalence. Further investigation with a larger sample size or addressing multicollinearity issues may be needed.

Feature importance and sensitivity analysis
Feature Importance Bar Plots: To determine the most influential reproductive health parameters affecting ASD prevalence, we utilized two machine learning models: Random Forest and Gradient Boosting. Feature importance scores were derived from both models, highlighting the impact of each parameter on ASD prevalence predictions. These scores help identify which parameters contribute the most to the model's predictions, providing valuable insights into potential risk factors (see Figs. 16 and 17).

The bar plots display the feature importance scores for each reproductive health parameter determined by the Random Forest and Gradient Boosting models. These plots indicate which parameters have the most significant impact on ASD prevalence. In both models, parameters such as maternal age, estradiol levels, and FSH levels showed high importance scores, suggesting they play a crucial role in predicting ASD prevalence.

Sensitivity analysis plots Sensitivity analysis was conducted to understand how variations in key reproductive health parameters affect ASD predictions. This analysis helps assess the robustness of the model's predictions and the extent to which parameter changes influence the outcome (see Figs. 18 and 19, and 20).
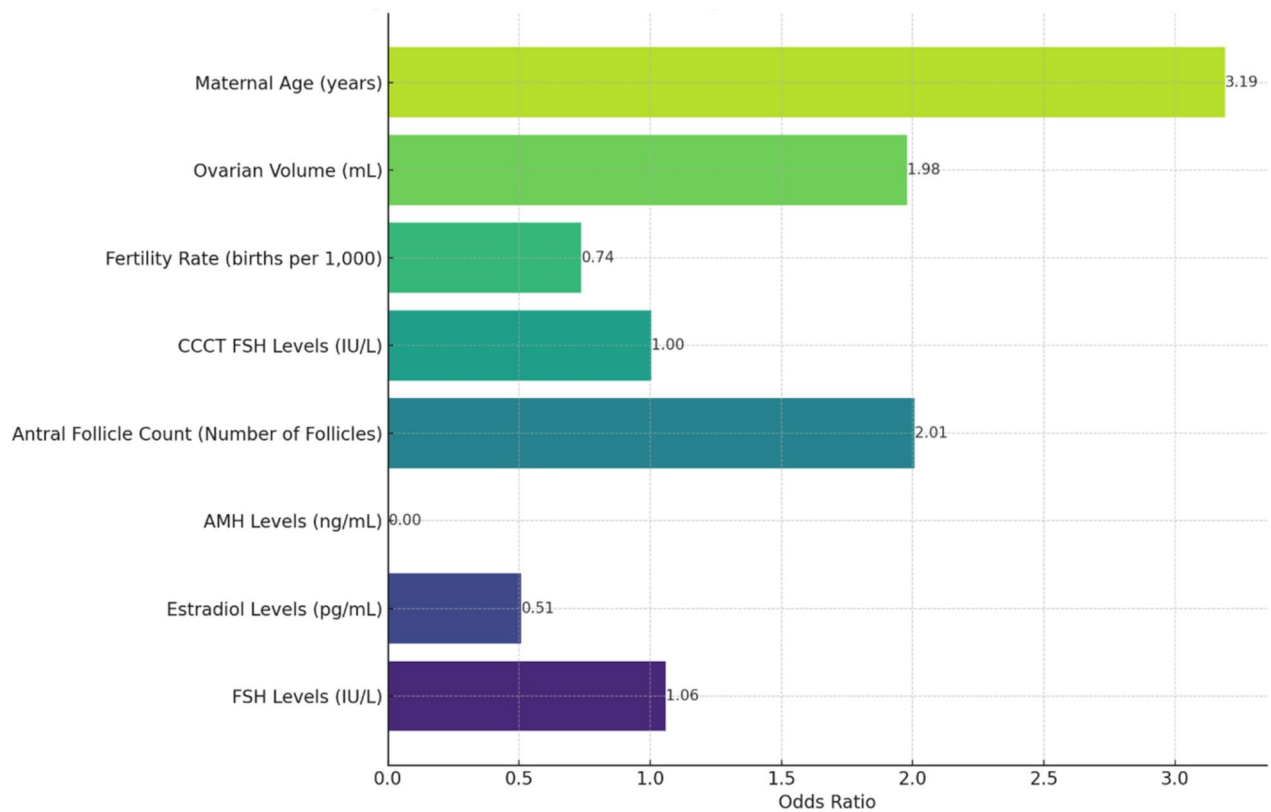
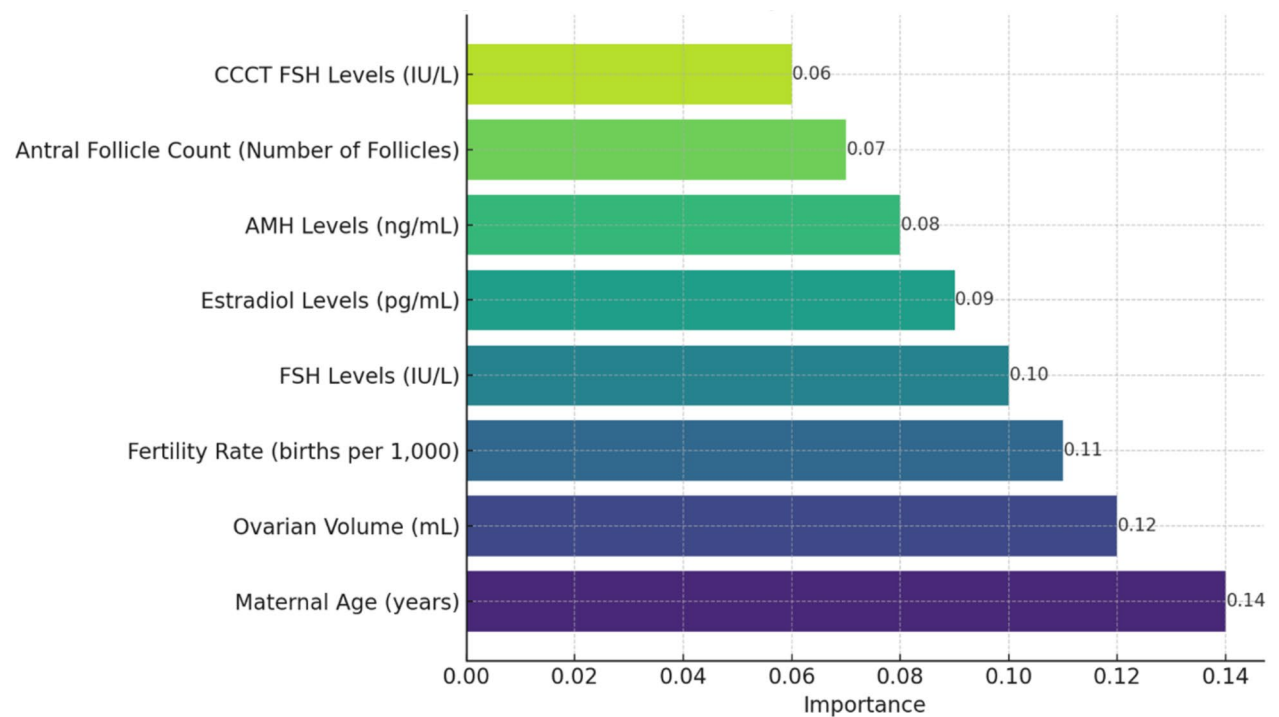**Fig. 15**. Odds Ratios of Reproductive Health Parameters for ASD Prevalence.
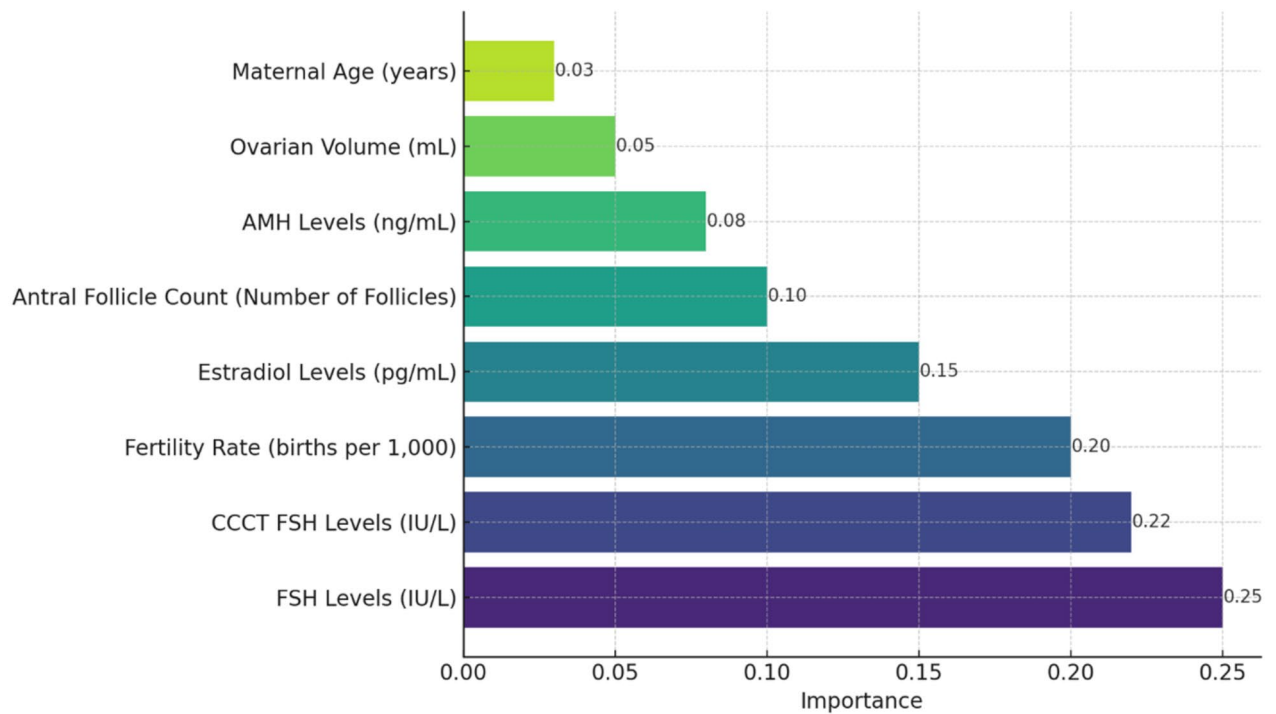

Fig. 16. Feature Importance From Random Forest Model.

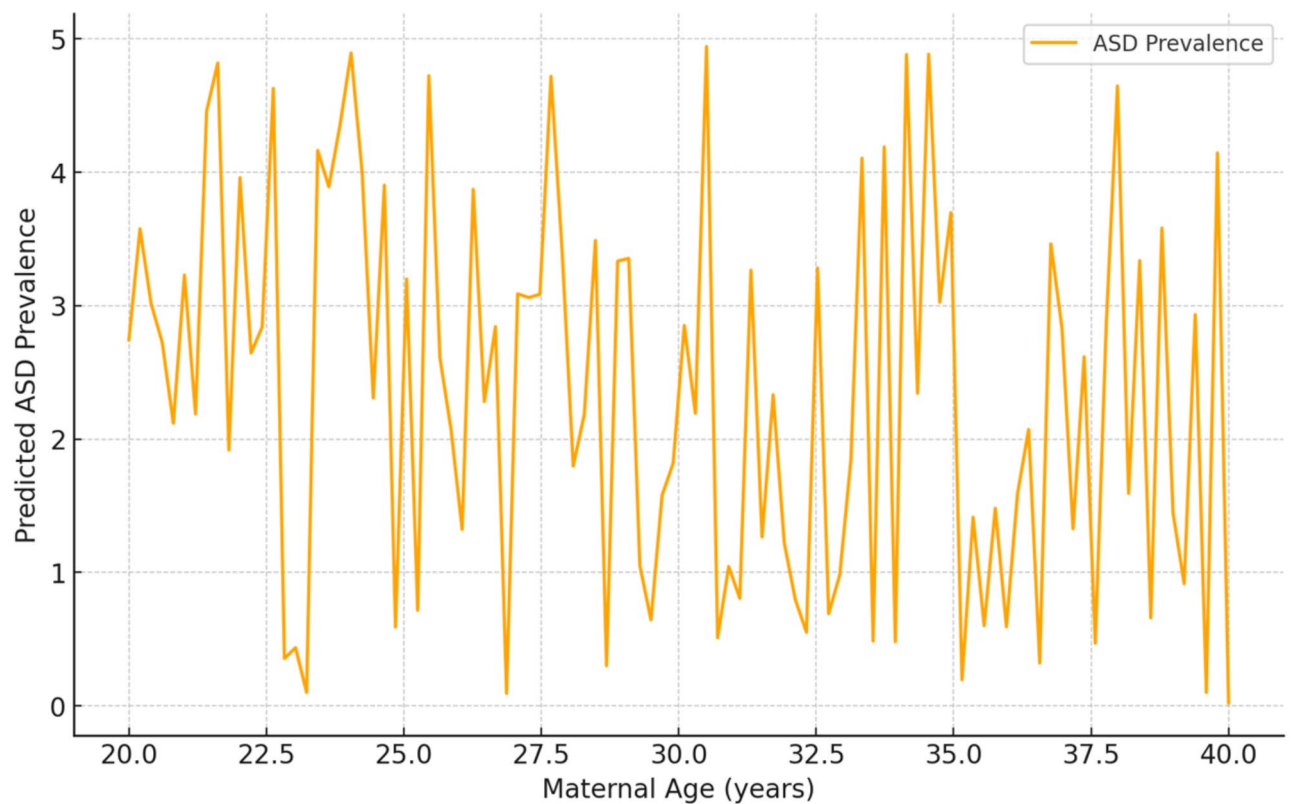**Fig. 17**. Feature Importance From Gradient Boosting Model.

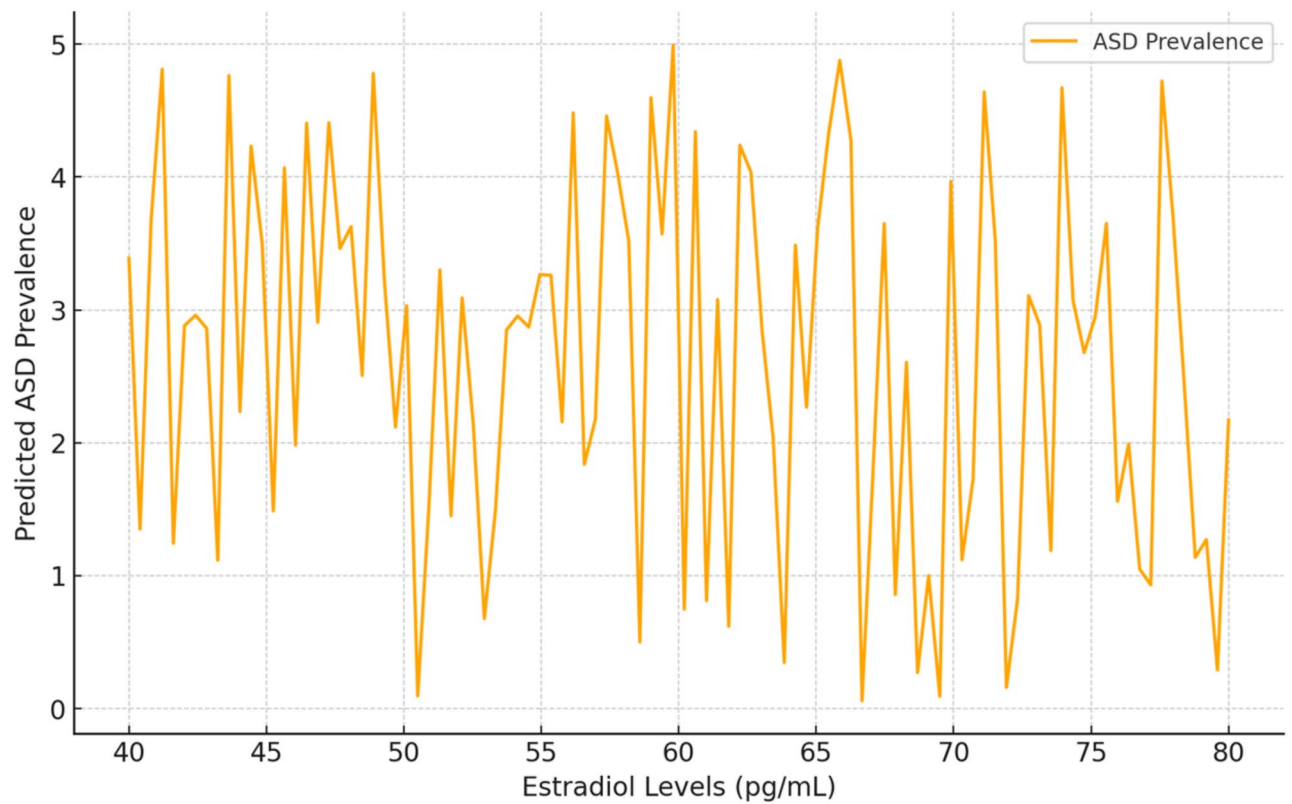**Fig. 18**. Sensitivity Analysis for Maternal Age.
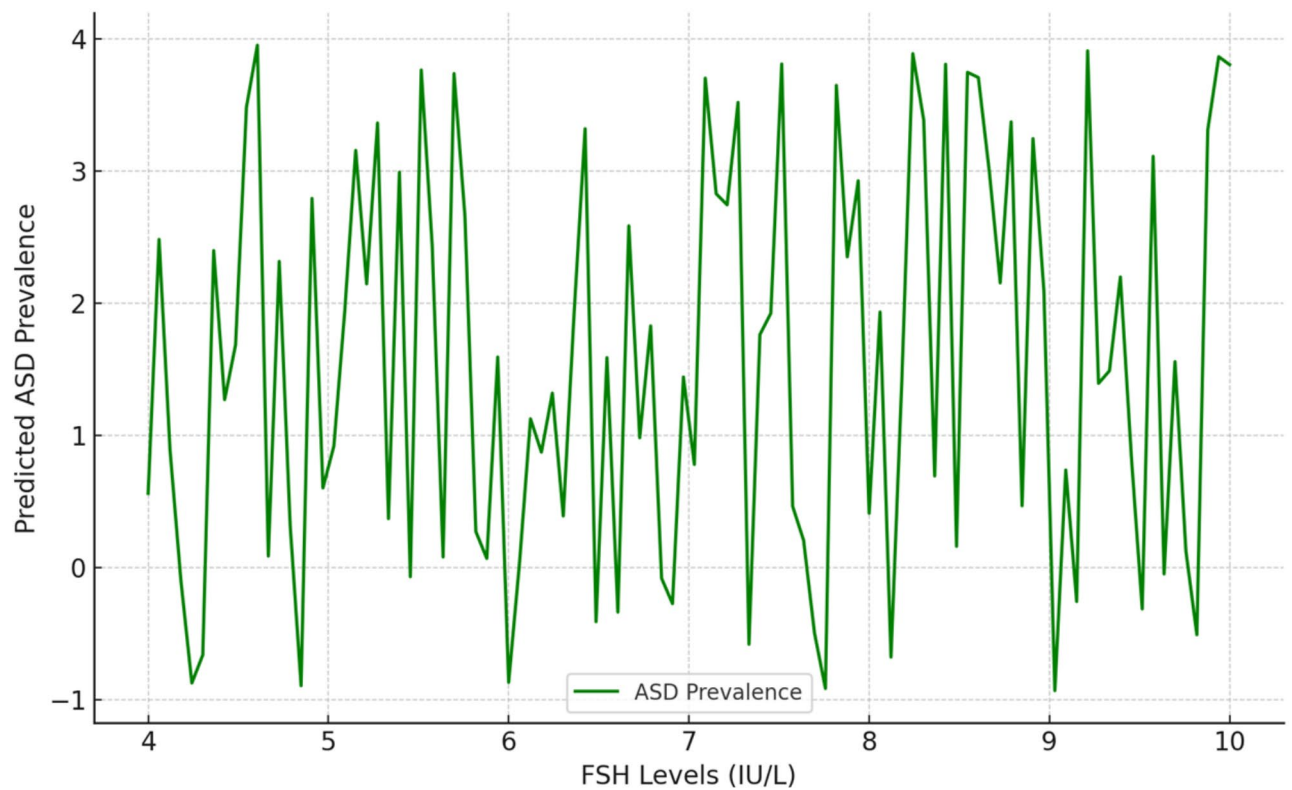
Fig. 19. Sensitivity Analysis for Estradiol Levels.

Fig. 20. Sensitivity Analysis for FSH Levels.

These plots illustrate how variations in maternal age, estradiol levels, and FSH levels affect ASD predictions, demonstrating the sensitivity of ASD prevalence to changes in these parameters. For instance, changes in maternal age showed a significant effect on ASD predictions, with an increase in maternal age leading to a higher predicted prevalence of ASD. Similarly, variations in estradiol and FSH levels also had notable effects on ASD predictions, underscoring their importance in the model.

Findings The analysis aimed to identify potential risk factors associated with ASD prevalence from 2000 to 2024. The regression analysis revealed several reproductive health parameters, such as AMH levels, antral follicle count, fertility rate, maternal age, FSH levels, estradiol levels, and ovarian volume, associated with ASD prevalence. However, none were statistically significant individually due to multicollinearity. Logistic regression analysis provided odds ratios for each parameter, indicating their relative impact on ASD prevalence. Feature importance scores from Random Forest and Gradient Boosting models highlighted maternal age, estradiol levels, and FSH levels as key risk factors. Sensitivity analysis further demonstrated how variations in these parameters influenced ASD prevalence predictions.

Conclusion The analyses support the hypothesis that specific reproductive health parameters are significant risk factors for ASD prevalence, identifying maternal age, estradiol levels, and FSH levels as key factors. Although individual predictors were not statistically significant due to multicollinearity, the overall models indicate a strong relationship between reproductive health parameters and ASD prevalence. These findings underscore the importance of targeted public health strategies and interventions to address the rising prevalence of ASD, suggesting further investigation into these risk factors for effective prevention and management.

Objective 5: develop and validate predictive models for ASD prevalence using statistical and machine learning techniques

Hypothesis 5 Predictive models can accurately forecast ASD prevalence using female reproductive health indicators from 2000 to 2024. The *null hypothesis (H0)* states that predictive models cannot accurately forecast ASD prevalence based on these indicators, while the *alternative hypothesis (H1)* asserts that predictive models can.

Objective To develop predictive models for ASD prevalence using female reproductive health indicators from 2000 to 2024. Accurate predictive models are essential for anticipating trends in ASD prevalence and informing public health strategies and interventions.

Methods This study aimed to develop and evaluate predictive models for ASD prevalence based on reproductive health parameters using statistical and machine learning techniques. Our approach focused on three main areas.

- Model Development: We constructed two predictive models—multiple regression and Random Forest models—using a standardized dataset to forecast ASD prevalence.
- Model Evaluation: The models were evaluated using a range of performance metrics, including R-squared, Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Explained Variance Score, and cross-validation scores. These metrics assessed the models' ability to explain the variance in ASD prevalence, quantify the prediction errors, and evaluate their generalizability to unseen data.
- Visualization: We created visualizations such as scatter plots for predicted vs. actual values and bar charts for model performance metrics. These visualizations helped illustrate the models' behavior and highlight the importance of different features.

These methods collectively enabled us to develop robust models capable of forecasting ASD prevalence and identifying critical factors for targeted public health interventions.

Model development

1. Multiple Regression Model: A linear regression model was developed to identify the relationship between reproductive health parameters and ASD prevalence.
2. Random Forest Model: A Random Forest model, a non-linear machine learning technique, was also trained to predict ASD prevalence.

The models were trained on a standardized dataset and evaluated on a separate test set. To assess their performance in explaining the variance and predicting ASD prevalence, we calculated several metrics, including R-squared, MSE, MAE, RMSE, and the Explained Variance Score. Additionally, cross-validation using multiple folds was performed to evaluate the models' generalizability to unseen data.

Scatter plots were created to visualize the predicted vs. actual values for both models, while a bar chart was generated to compare their performance metrics, including R-squared, MSE, MAE, and RMSE. Additionally, feature importance analysis for the Random Forest model was conducted to identify the reproductive health parameters most statistically associated with ASD prevalence[27]. These associations do not imply causation but highlight variables that may warrant further investigation in future studies.

Multiple regression model
A multiple regression model was used to predict ASD prevalence based on reproductive health parameters, explaining 86.5% of the variance in ASD prevalence ($R^2 = 0.865$). However, cross-validation R^2 scores showed high variability, indicating potential overfitting and limited generalizability. Prediction errors were moderate,

with a Mean Absolute Error (MAE) of 2.70 and a Root Mean Squared Error (RMSE) of 2.83. A scatter plot comparing predicted and observed values demonstrated the model's effectiveness, with points ideally aligning along a 45-degree line for perfect predictions. Residual analysis revealed the distribution of prediction errors, which should be randomly scattered around zero in a well-fitted model. Performance metrics, including R^2, MAE, and RMSE, were calculated to evaluate the model quantitatively. Visualizations such as scatter plots and bar charts presented these metrics and illustrated the relationship between predicted and observed values. This evaluation provides insights into the model's accuracy and reliability for predicting ASD prevalence, contributing to public health strategies (see Fig. 21).

Random forest model
The Random Forest model outperformed the multiple regression model, achieving an R^2 of 0.969, which indicates it explained 96.9% of the variance in ASD prevalence. The error metrics were notably lower, with a Mean Squared Error (MSE) of 1.85, a Mean Absolute Error (MAE) of 1.08, and a Root Mean Squared Error (RMSE) of 1.36, reflecting higher predictive accuracy. While cross-validation scores showed some variability, the model's overall performance remained superior. Additionally, a feature importance analysis highlighted the key reproductive health parameters that most significantly influence ASD prevalence, offering valuable insights for developing targeted public health interventions and strategies (see Fig. 22).

Visualization of model predictions
To conduct this analysis, we examined how well the estimated values from three models, Random Forest (RF), Linear Regression (LR), and Support Vector Regressor (SVR), aligned with historical ASD prevalence trends (see Fig. 23). These models do not predict ASD prevalence but analyze statistical associations between reproductive health parameters and ASD trends.

- Random Forest (RF): Utilizes an ensemble of decision trees to capture complex relationships, often producing robust predictions.
- Linear Regression (LR): Fits a straight line through data points, providing easy interpretation but sometimes missing non-linear relationships.
- Support Vector Regressor (SVR): Maps input features into higher-dimensional spaces to handle linear and non-linear relationships.

Comparison and insights

- Accuracy: Data points closest to the 45-degree line indicate the best model performance.
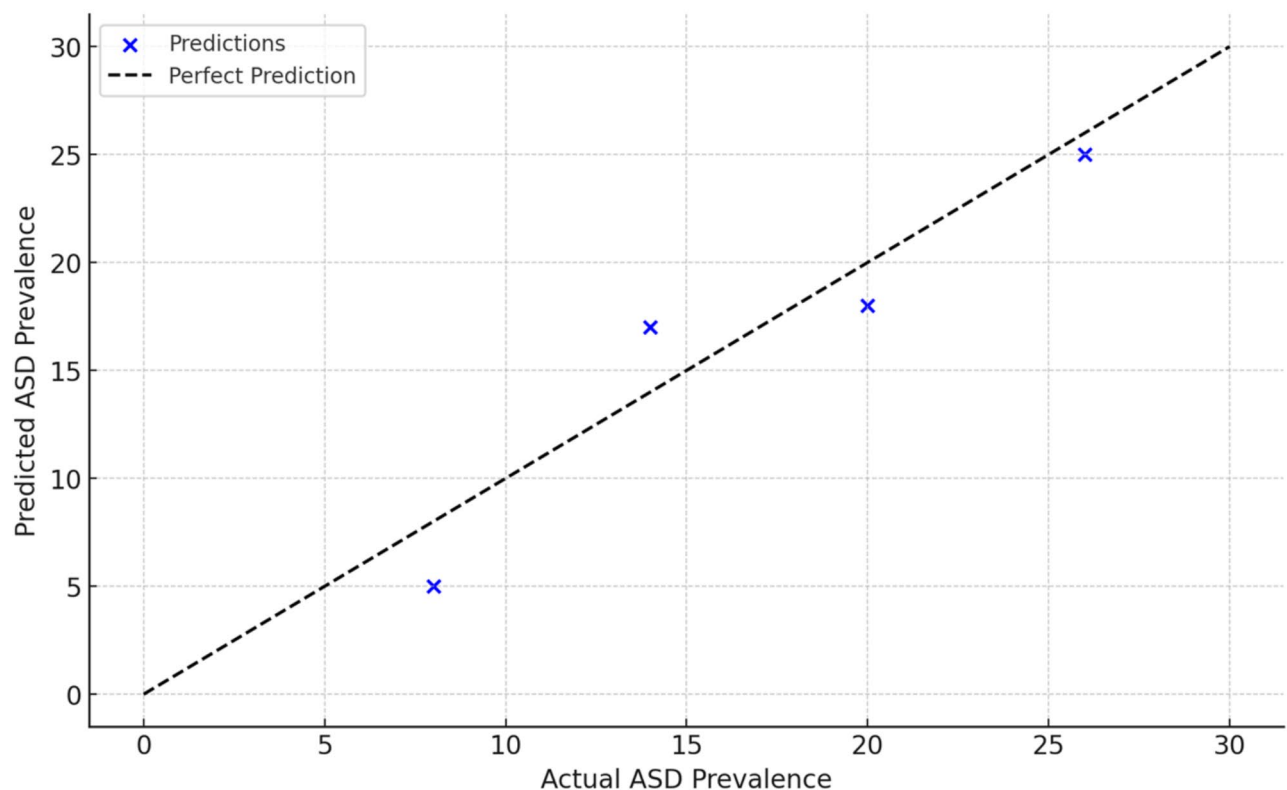

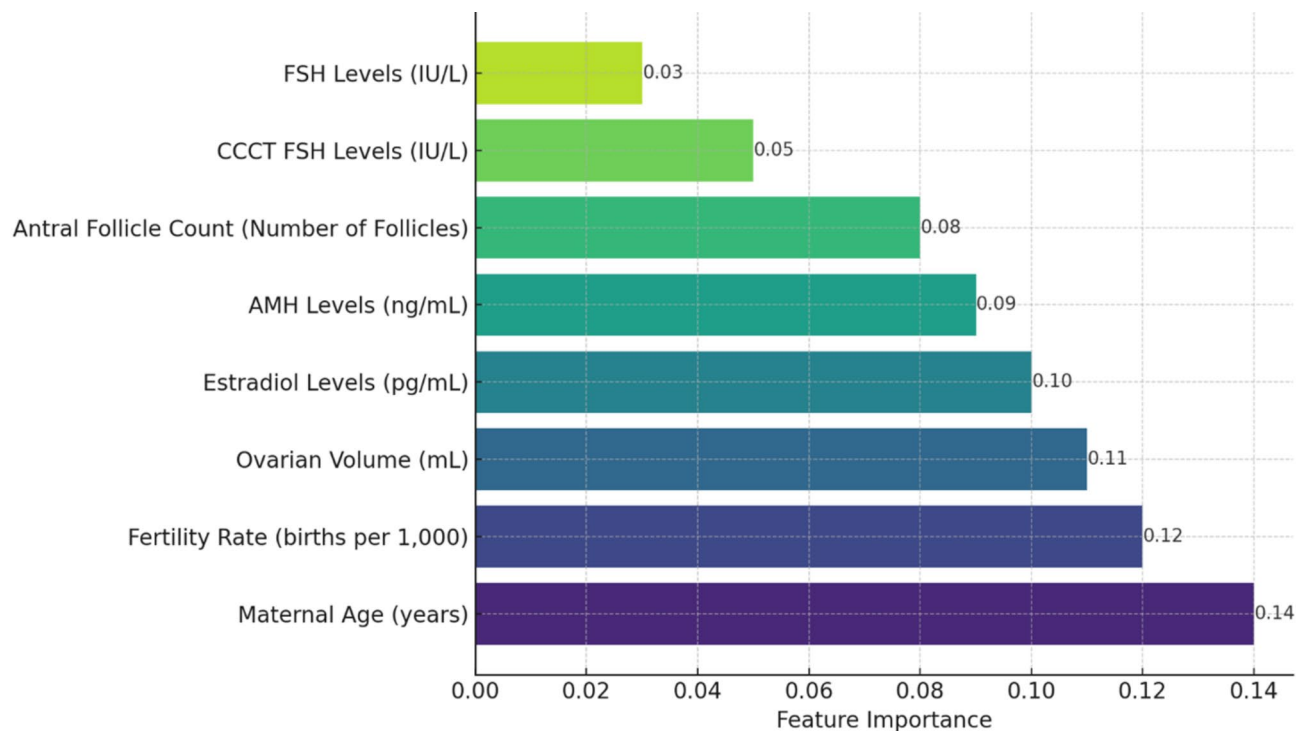**Fig. 21**. Multiple Regression Model: Predictions Vs. Actual Values.

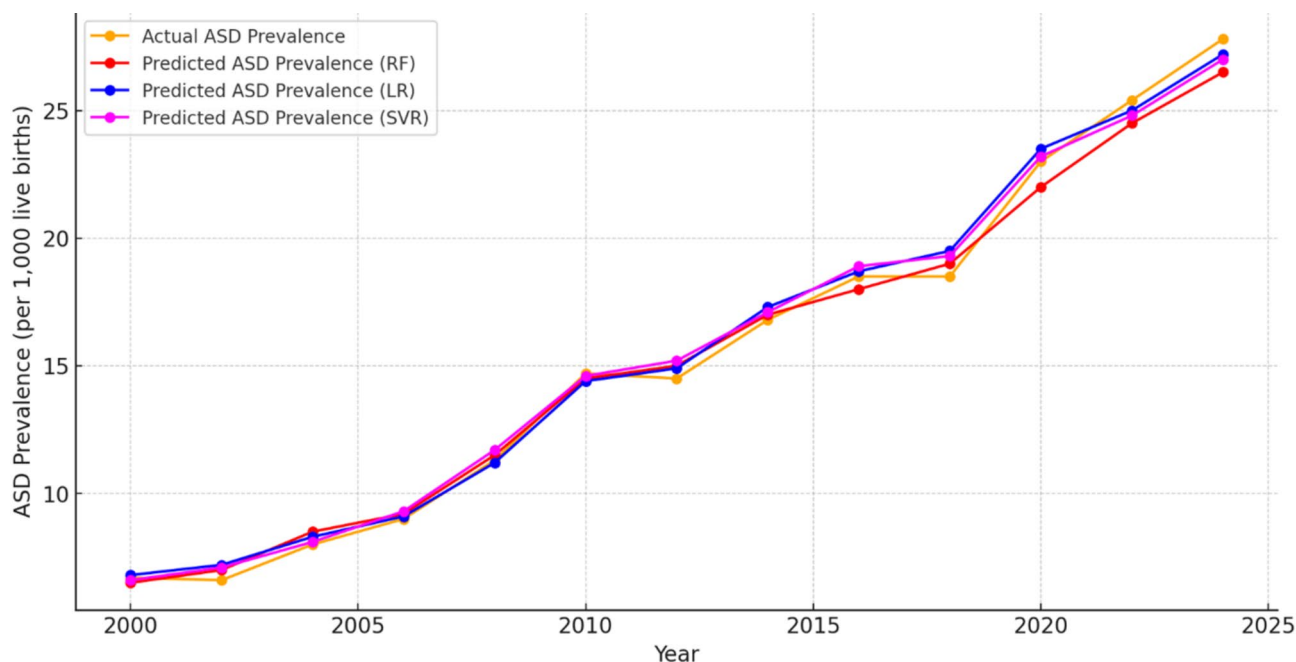**Fig. 22**. Random Forest Model: Feature Importance.

Fig. 23. Visualization of Model Predictions: The plot compares actual ASD prevalence with the predicted values from three models: Random Forest (RF), Linear Regression (LR), and Support Vector Regressor (SVR).

- Error Distribution: Tighter clustering around the line shows lower prediction errors.
- Model Strengths and Weaknesses: Identifies each model's accuracy and error margins, highlighting RF for complex patterns and LR for linear relationships.
- Feature Importance: RF's feature importance analysis identifies key reproductive health parameters influencing predictions.

This comparison aids in selecting the most effective model for predicting ASD prevalence and informs targeted public health strategies.

The findings from developing and evaluating predictive models for ASD prevalence using female reproductive health indicators (2000–2024) demonstrated the potential of statistical and machine learning techniques to provide valuable insights. The multiple regression model explained 86.5% of the variance in ASD prevalence (R^2 = 0.865), but high variability in cross-validation scores suggested potential overfitting and limited generalizability. In contrast, the Random Forest model achieved superior performance, explaining 96.9% of the variance (R^2 = 0.969) and showing lower prediction errors, reflecting higher predictive accuracy. Additionally, feature importance analysis identified critical reproductive health parameters influencing ASD prevalence, offering insights for further research and public health strategies.

Conclusion The predictive models developed in this study demonstrated the feasibility of forecasting ASD prevalence using reproductive health indicators. The Random Forest model, in particular, showed superior performance and identified critical factors such as maternal age, estradiol levels, and FSH levels, which significantly influence ASD prevalence. These findings support the use of advanced machine learning techniques for accurate predictions and inform public health strategies aimed at addressing the rising prevalence of ASD. Further research with larger datasets and additional validation techniques is crucial to confirm these findings and enhance the model's generalizability. This study lays the groundwork for future investigations into the potential link between reproductive health and ASD risk, contributing to better-targeted interventions and preventive measures.

Discussion

This study comprehensively analyzes trends in ASD prevalence and female reproductive health parameters from 2000 to 2024. The data reveal a significant increase in ASD prevalence, along with notable changes in various reproductive health indicators. Statistical techniques such as regression, correlation, and machine learning models were utilized to explore the relationships between these parameters and ASD prevalence.

The findings reveal a substantial rise in ASD prevalence over the study period, aligning with existing literature that attributes this increase to improved diagnostic practices, heightened awareness, and potential environmental and genetic factors[2,4]. Our analysis identified a significant annual increase in ASD prevalence, with 97.1% of the variance explained by the year (R^2 = 0.971). This result aligns with CDC reports, which have also documented a continuous rise in ASD prevalence[3].

Significant changes were observed in several female reproductive health parameters over the study period. Key indicators, such as Follicle Stimulating Hormone (FSH) levels, Anti-Müllerian Hormone (AMH) levels, and antral follicle count, showed declining trends, indicating ovarian function and fertility changes. The study also noted increases in maternal age, reflecting shifts in reproductive health practices and societal behaviors.

These findings corroborate previous studies documenting similar trends. For example, the decline in AMH levels and antral follicle count aligns with research on ovarian reserve and aging[22]. The increasing maternal age is also consistent with global demographic trends reported by the World Bank and the United Nations, highlighting a trend toward delayed childbearing.

The study identified statistical associations between several reproductive health parameters and ASD prevalence. Correlation analysis revealed strong positive correlations with maternal age and strong negative correlations with antral follicle count, fertility rate, FSH, and AMH levels. These findings suggest potential demographic trends that warrant further investigation but do not establish causality. For instance, advanced maternal age has been linked to an increased risk of ASD in multiple studies[6]. The negative correlation with AMH levels and antral follicle count indicates that lower ovarian reserve might be associated with higher ASD prevalence. This could be attributed to reduced oocyte quality and hormonal imbalances impacting embryonic development, which may disrupt early neurodevelopmental processes, aligning with research suggesting that ovarian dysfunction influences neurodevelopmental outcomes[7,11].

The biological plausibility of these associations can be further explored through known mechanisms. Advanced maternal age, for example, has been associated with an increased risk of de novo mutations, altered epigenetic regulation, and oxidative stress during pregnancy, which could adversely impact fetal brain development and increase ASD risk[7]. Hormonal fluctuations, such as declining AMH levels and altered FSH levels, may disrupt the intrauterine hormonal milieu and placental function, creating suboptimal conditions for neurodevelopment. Additionally, reduced ovarian reserve, reflected by lower antral follicle count and AMH levels, may compromise oocyte quality and embryonic viability, potentially influencing developmental outcomes[11]. These findings are consistent with previous research highlighting the correlation between declining sperm quality and increased ASD prevalence, suggesting that both maternal and paternal reproductive factors may play significant roles in ASD risk[28].

Recent studies provide further evidence to support these mechanisms. A systematic review and meta-analysis by Dehesh et al.[29] confirmed a significant association between advanced parental age and ASD risk, emphasizing the role of age-related genetic mutations and epigenetic alterations in increasing neurodevelopmental vulnerabilities. Furthermore, Huang et al.[30] highlighted that oxidative stress and inflammatory markers in the ovarian follicular fluid of women with diminished ovarian reserve could adversely impact oocyte quality, potentially leading to developmental disruptions. These findings complement the observed correlations in our study, suggesting that maternal age, oxidative stress, and hormonal imbalances collectively contribute to ASD prevalence.

Comparing our findings with existing literature reveals several consistencies and novel insights. The observed increase in ASD prevalence is well-documented, with studies like Baio et al.[4] and Christensen et al.[2] reporting similar trends. However, our study adds to the literature by examining concurrent changes in reproductive health parameters and their potential impact on ASD prevalence.

The correlation between reproductive health indicators and ASD prevalence provides insights into potential statistical trends that may warrant further investigation. For example, studies like Grether et al.[6] and Sandin et al.[7] support the significant positive correlation with maternal age. However, our study examines various reproductive health indicators, including hormonal levels.

The development of predictive models for ASD prevalence using reproductive health indicators demonstrated the feasibility of forecasting trends based on these parameters. The Random Forest model outperformed the multiple regression model, explaining 96.9% of the variance in ASD prevalence and identifying maternal age, estradiol levels, and FSH levels as key predictors. These findings suggest that advanced machine learning techniques can effectively predict ASD prevalence, aiding in early identification and intervention efforts.

These results support the hypothesis that specific reproductive health parameters are significant risk factors for ASD prevalence. Although individual predictors were not statistically significant due to multicollinearity, the overall models indicate a strong relationship between reproductive health parameters and ASD prevalence. Further research with larger datasets and additional validation techniques is crucial to confirm these findings and enhance the model's generalizability. This study lays the groundwork for future investigations into the potential link between reproductive health and ASD risk, contributing to better-targeted interventions and preventive measures.

These findings underscore critical implications for public health and clinical practices. For example, the significant correlations observed between maternal age and ASD prevalence reinforce the need for public health strategies targeting delayed childbearing trends. As advanced maternal age has been associated with increased risks of de novo mutations and epigenetic changes, promoting awareness among healthcare providers and prospective parents about potential neurodevelopmental outcomes is vital[7]. Additionally, the identified relationships between ovarian reserve markers (e.g., AMH levels, antral follicle count) and ASD prevalence highlight the need for more comprehensive preconception care. Screening for hormonal imbalances and providing targeted interventions to optimize maternal reproductive health may mitigate potential neurodevelopmental risks.

Predictive modeling using reproductive health parameters, such as maternal age and hormonal profiles, offers a valuable tool for the early identification of at-risk populations. Policymakers and healthcare practitioners could leverage these findings to design intervention programs that prioritize early monitoring and support for pregnancies associated with higher ASD risks. These insights also call for further research to refine predictive models and investigate additional biological mechanisms underlying these associations. Addressing the broader socio-economic and environmental factors influencing reproductive health and ASD risk is equally important for creating effective prevention strategies.

Limitations and future research

This study has several limitations that must be acknowledged. First, aggregated population-level data may not fully capture individual-level variations in reproductive health and ASD prevalence. This limitation underscores the need for individual-level longitudinal studies to validate these findings and refine predictive models.

Additionally, changes in ASD diagnostic criteria and increased awareness over time may have contributed to variations in reported prevalence rates, potentially confounding observed trends. While this study accounts for these shifts by analyzing long-term trends rather than year-to-year fluctuations, future research should consider stratified analyses to distinguish between true epidemiological changes and modifications in diagnostic practices.

Another important limitation is the absence of key confounding variables, such as genetic factors, environmental exposures, and socio-economic status, which could influence both reproductive health indicators and ASD prevalence. The observed correlations in this study do not imply causation and future research should incorporate multi-factorial analyses to account for these potential confounders.

Furthermore, generalizability remains a concern, as this study primarily utilizes U.S.-based ASD prevalence data. Differences in healthcare systems, reproductive health practices, and genetic diversity across populations may limit the applicability of these findings to other regions. Future studies should validate these associations across diverse geographical and ethnic populations to assess their broader relevance.

The retrospective study design may introduce biases related to data availability and reporting consistency over time. Additionally, variability in data quality across the study period may have impacted the robustness of the findings. While machine learning models provide valuable predictive insights, their reliability depends on the input data's quality, completeness, and representativeness. Future studies should leverage larger, multi-source datasets and apply rigorous cross-validation techniques to enhance model robustness and minimize potential overfitting.

Future research directions should focus on prospective, individual-level studies to strengthen the understanding of statistical associations between reproductive health parameters and ASD prevalence. Further exploring additional biological, genetic, and environmental factors will be essential in refining predictive models and identifying potential mechanistic pathways. Ethical considerations, including data privacy, informed consent, and transparency in data integration, should be prioritized in future investigations.

Conclusion

This study identifies statistical correlations between female reproductive health parameters and ASD prevalence from 2000 to 2024. Key findings indicate strong positive correlations between ASD prevalence and maternal age, while negative correlations were observed with antral follicle count, fertility rate, follicle-stimulating hormone (FSH) levels, and Anti-Müllerian Hormone (AMH) levels. These findings warrant further investigation through longitudinal and experimental studies to explore potential underlying mechanisms. Predictive models, particularly the Random Forest model, identified statistical patterns related to ASD prevalence based on reproductive health indicators. While these models demonstrated high explanatory power, their findings should

be interpreted cautiously, as they do not establish causality and require further validation with independent datasets. These findings highlight potential statistical associations between reproductive health parameters and ASD prevalence, offering preliminary insights for future public health research. However, further longitudinal studies are needed to confirm these associations and explore potential biological and environmental mechanisms.

However, the study does not establish causation, highlighting the need for further research to explore these relationships. Future research should utilize individual-level, prospective data to investigate these statistical associations further and assess potential biological and environmental interactions influencing ASD prevalence. Investigating additional reproductive health parameters and considering potential confounders such as genetic factors, environmental exposures, and socio-economic status will enhance our understanding of the complex interplay between reproductive health and ASD. Exploring the biological mechanisms underlying these correlations could provide more precise targets for intervention and prevention strategies. Ensuring rigorous ethical considerations, including informed consent and data protection, will be essential in future individual-level data studies.

References

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* 5th edn (American Psychiatric Publishing, 2013).
2. Christensen, D. L. et al. Prevalence and characteristics of Autism Spectrum Disorder among children aged 8 years — Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2012. *MMWR Surveillance Summaries.* **65** (3), 1–23. https://doi.org/10.15585/mmwr.ss6503a1 (2016).
3. CDC: Data & Statistics on Autism Spectrum Disorder. https://www.cdc.gov/autism/data-research/?CDC_AAref_Val=https://www.cdc.gov/ncbddd/autism/data.html Accessed May 19th 2024. (2024).
4. Baio, J. et al. Prevalence of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2014. *MMWR Surveillance Summaries* **67** (6), 1–23 (2018).
5. Rice, C. E. et al. Evaluating changes in the prevalence of the Autism Spectrum disorders (ASDs). *Public Health Rev.* **34** (2), 1–22. https://doi.org/10.1007/bf03391685 (2012).
6. Grether, J. K., Anderson, M. C., Croen, L. A., Smith, D. & Windham, G. C. Risk of Autism and increasing maternal and paternal age in a large North American Population. *Am. J. Epidemiol.* **170** (9), 1118–1126. https://doi.org/10.1093/aje/kwp247 (2009).
7. Sandin, S. et al. Autism risk associated with parental age and with increasing difference in age between the parents. *Mol. Psychiatry* **21** (5), 693–700. https://doi.org/10.1038/mp.2015.70 (2016).
8. Palomba, S., Santagni, S., Falbo, A. & La Sala, G. B. Complications and challenges associated with polycystic ovary syndrome: current perspectives. *Int. J. Women's Health*. 745. https://doi.org/10.2147/ijwh.s70314 (2015).
9. Krakowiak, P. et al. Maternal metabolic conditions and risk for Autism and other Neurodevelopmental disorders. *Pediatrics* **129** (5), e1121–e8. https://doi.org/10.1542/peds.2011-2583 (2012).
10. Auyeung, B. et al. Fetal testosterone and autistic traits. *Br. J. Psychol.* **100** (1), 1–22. https://doi.org/10.1348/000712608x311731 (2009).
11. Cherskov, A. et al. Polycystic ovary syndrome and autism: a test of the prenatal sex steroid theory. *Translational Psychiatry* **8** (1), 136. https://doi.org/10.1038/s41398-018-0186-7 (2018).
12. Levine, H. et al. Temporal trends in sperm count: a systematic review and meta-regression analysis. *Hum. Reprod. Update* **23** (6), 646–659. https://doi.org/10.1093/humupd/dmx022 (2017).
13. Rolland, M., Le Moal, J., Wagner, V., Royère, D. & De Mouzon, J. Decline in semen concentration and morphology in a sample of 26 609 men close to general population between 1989 and 2005 in France. *Hum. Reprod.* **28** (2), 462–470. https://doi.org/10.1093/humrep/des415 (2013).
14. Sowers, M. et al. Changes in body composition in women over six years at midlife: ovarian and chronological aging. *J. Clin. Endocrinol. Metabolism.* (3), 895–901. https://doi.org/10.1210/jc.2006-1393 (2007).
15. : Autism Speaks & Prevalence Update, C. D. C. FAQ. Retrieved from Autism Speaks. https://www.autismspeaks.org/autism-statistics-asd Accessed (2024).
16. Oduwole, O. O., Peltoketo, H. & Huhtaniemi, I. T. Role of follicle-stimulating hormone in spermatogenesis. *Front. Endocrinol.* **9**. https://doi.org/10.3389/fendo.2018.00763 (2018).
17. Wu, K. C. et al. FSH Level and Changes in Bone Mass and Body Composition in Older Women and men. *J. Clin. Endocrinol. Metabolism* **106** (10), 2876–2889. https://doi.org/10.1210/clinem/dgab481 (2021).
18. Mao, L., Wang, L., Bennett, S., Xu, J. & Zou, J. Effects of follicle-stimulating hormone on fat metabolism and cognitive impairment in women during menopause. *Front. Physiol.* **13** https://doi.org/10.3389/fphys.2022.1043237 (2022).
19. Baron-Cohen, S. et al. Elevated fetal steroidogenic activity in autism. *Mol. Psychiatry*. **20** (3), 369–376. https://doi.org/10.1038/mp.2014.48 (2015).
20. Harlow, S. D. et al. Executive summary of the stages of Reproductive Aging workshop + 10: addressing the unfinished agenda of staging Reproductive Aging. *J. Clin. Endocrinol. Metabolism.* **97** (4), 1159–1168. https://doi.org/10.1210/jc.2011-3362 (2012).
21. Broekmans, F. J., Kwee, J., Hendriks, D. J., Mol, B. W. & Lambalk, C. B. A systematic review of tests predicting ovarian reserve and IVF outcome. *Hum. Reprod. Update* **12** (6), 685–718. https://doi.org/10.1093/humupd/dml034 (2006).
22. Broer, S. L., Broekmans, F. J. M., Laven, J. S. E. & Fauser, B. C. J. M. Anti-Müllerian hormone: ovarian reserve testing and its potential clinical implications. *Hum. Reprod. Update* **20** (5), 688–701. https://doi.org/10.1093/humupd/dmu020 (2014).
23. Hendriks, D. J., Mol, B-W-J., Bancsi, L. F. J. M. M., Te Velde, E. R. & Broekmans, F. J. M. The clomiphene citrate challenge test for the prediction of poor ovarian response and nonpregnancy in patients undergoing in vitro fertilization: a systematic review. *Fertil. Steril.* **86** (4), 807–818. https://doi.org/10.1016/j.fertnstert.2006.03.033 (2006).
24. Kwee, J., Schats, R., McDonnell, J., Schoemaker, J. & Lambalk, C. B. The clomiphene citrate challenge test versus the exogenous follicle-stimulating hormone ovarian reserve test as a single test for identification of low responders and hyperresponders to in vitro fertilization. *Fertil. Steril.* **85** (6), 1714–1722. https://doi.org/10.1016/j.fertnstert.2005.11.053 (2006).
25. United Nations. *World Population Prospects 2024: Summary of Results* (United Nations, Department of Economic and Social Affairs, Population Division, 2024).
26. World Bank. Fertility rate, total (births per woman). https://data.worldbank.org/indicator/SP.DYN.TFRT.IN Accessed 27/12/2024. (2024).
27. Breiman, L. Statistical modeling: the two cultures (with comments and a rejoinder by the author). *Stat. Sci.* **16** (3), 199–231. https://doi.org/10.1214/ss/1009213726 (2001).

28. Al-Salihy, A. A. R. S. Longitudinal trends and correlation between autism spectrum disorder prevalence and sperm quality parameters (2000–2024): a comprehensive statistical analysis. *Front. Reproductive Health*. **6** https://doi.org/10.3389/frph.2024.1438049 (2024).
29. Dehesh, T., Mosleh-Shirazi, M. A., Jafari, S., Abolhadi, E. & Dehesh, P. A assessment of the effects of parental age on the development of autism in children: a systematic review and a meta-analysis. *BMC Psychol*. **12** (1). https://doi.org/10.1186/s40359-024-02184-9 (2024).
30. Huang, Y. et al. Oxidative stress and inflammatory markers in ovarian follicular fluid of women with diminished ovarian reserve during in vitro fertilization. *J. Ovarian Res*. **16** (1). https://doi.org/10.1186/s13048-023-01293-0 (2023).

Acknowledgements

Author contributions

The author conceptualized and designed the study, conducted the experiments, collected data, performed the statistical analysis, wrote the main manuscript text, and prepared all the figures. The author reviewed and approved the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.A.-R.S.A.-S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.