scientific reports

Check for updates

OPEN

Analysis of factors influencing polycystic ovary syndrome in women of reproductive age based on directed acyclic graphs

Huixi Yu[1,2], Yu Guo[1,2], Yingni Ling[1], Yinhua Su[1]✉ & Zhongyu Li[1]✉

Polycystic ovary syndrome (PCOS) is a common gynecological endocrine disorder in women of reproductive age that seriously affects both their physical and mental health. The pathogenesis of PCOS is complex and not yet fully understood, and it is crucial to control for bias and analyze the risk factors for its development in order to provide a basis for developing preventive strategies. A case–control study design was used. Patients first diagnosed with PCOS from January 2024 to June 2024 at the First Affiliated Hospital of the University of South China, the Second Affiliated Hospital of the University of South China, and the Affiliated Nanhua Hospital of University of South China were selected as the case group (n = 210). Non-PCOS women attending during the same period were selected as the control group (n = 420). Information was collected using self-administered questionnaires, including the Pittsburgh Sleepiness Scale (PSQI), the Generalized Anxiety Disorder 7 (GAD-7) scale, and the Patient Health Questionnaire (PHQ-9). A directed acyclic graph was used for variable screening. Propensity score matching controlled for confounding variables, and multifactorial logistic regression analysis identified risk factors for PCOS. Multifactorial logistic regression analysis showed that obesity [*OR* = 4.088, 95% *CI* (2.580, 6.476), *P* < 0.001], alcohol consumption [*OR* = 2.305, 95% *CI* (1.320, 4.024), *P* = 0.003], family history of PCOS [*OR* = 6.468, 95% *CI* (1.986, 21.067), *P* = 0.002], low birth weight [*OR* = 0.637, 95% *CI* (0.438, 0.927), *P* = 0.018], and anxiety [*OR* = 4.905, 95% *CI* (2.768, 8.693), *P* < 0.001] were risk factors for PCOS development. BMI ≥ 25 kg/m^2, alcohol consumption, family history of PCOS, low birth weight, and anxiety are risk factors for the development of PCOS. Targeted measures should be implemented to address these factors, reducing the incidence of PCOS and promoting female reproductive endocrine health.

Keywords Polycystic ovary syndrome, Case–control study, Analysis of influencing factors, Directed acyclic graph, Propensity score matching

Polycystic ovarian syndrome (PCOS) is the most prevalent reproductive and endocrine disorder among women of reproductive age, with its global incidence on the rise. This condition imposes a healthcare cost in the billions annually, emerging as a major public health crisis that affects 5–18% of women over their lifetime[1–3]. PCOS is a major cause of anovulatory infertility and is associated with increased risks of gestational diabetes, hypertension, metabolic disorders, and endometrial cancer[4–6]. Early detection, diagnosis, and treatment of PCOS are essential. Hence, identifying the factors associated with PCOS onset is of critical importance. Numerous studies suggest that PCOS development stems from complex interactions between genetic, environmental, behavioral, and psychological factors, though its exact pathogenesis and risk factors remain incompletely understood[7–9]. Current research on PCOS risk factors remains inconsistent, with a notable paucity of epidemiological studies investigating the pathogenic roles of emotional and environmental toxins. Previous studies have employed a "single-factor followed by multifactorial" modeling approach, which likely overlooks confounding factors and covariates, while omitting key exposure variables, potentially leading to biased results[10,11].

Directed acyclic graphs (DAGs) are a theory-driven method for selecting independent variables in regression analysis. By constructing a causal relationship network based on theoretical assumptions, DAGs systematically identify variables that should be included in the model. This approach offers intuitive visualization of causal

[1]School of Nursing, University of South China, Hengyang 421001, Hunan, China. [2]Huixi Yu and Yu Guo contributed equally to this work. ✉email: 382373646@qq.comYinhua; lzhy1023@hotmail.com

pathways, enabling researchers to clarify direct and indirect effects among variables[12]. This methodology has been successfully applied in the identification of risk factors for premature physeal closure following distal femoral fractures and the analysis of factors associated with suicidal ideation in adolescents[13,14]. DAGs ensure robust variable screening aligned with causal hypotheses, bridging theoretical epidemiology and data-driven modeling for reproducible, clinically interpretable results.

In this study, we propose employing DAGs to elucidate the key confounders and mediating variables of PCOS, to control for bias and to investigate the direct risk factors linked to PCOS. This approach aims to inform primary prevention strategies, ultimately reducing PCOS incidence and advancing women's reproductive health.

Method

Participants

PCOS patients newly diagnosed at the First Affiliated Hospital of the University of South China, the Second Affiliated Hospital of the University of South China, and the Affiliated Nanhua Hospital of University of South China between January 2024 and June 2024 were recruited as the case group. The inclusion criteria for the case group were: (1) women aged 20–49; (2) meeting the 2003 Rotterdam PCOS diagnostic criteria, which include sporadic ovulation or anovulation, clinical manifestations of hyperandrogenism and/or biochemical hyperandrogenism, and ovarian polycystic changes (defined as > 12 follicles of 2–9 mm in one or both ovaries and/or ovarian volume > 10 mL); at least two of the three criteria must be met, excluding other causes of elevated androgen levels[15]; (3) providing informed consent and voluntarily cooperating with the investigation. Non-PCOS women attending the same hospitals during the study period were selected as the control group. The inclusion criteria for the control group were: (1) non-PCOS women matched 1:2 to the case group by age (± 3 years); (2) providing informed consent and voluntarily cooperating with the investigation; (3) absence of PCOS diagnosis as per the 2003 Rotterdam criteria following physician consultation, gynecological ultrasound, and a panel of sex hormone tests. The exclusion criteria for both case and control groups were: (1) psychiatric disorders impairing cooperation with the investigation; (2) conditions affecting ovulation, such as early-onset ovarian insufficiency or functional hypothalamic amenorrhea; (3) primary amenorrhea or menstrual abnormalities due to organic pathology; (4) conditions causing elevated androgen levels, such as congenital adrenal hyperplasia or androgen-secreting tumors; (5) recent use of hormonal drugs within the last three months; (6) pregnancy or lactation.

A recent national study reported a 7.8% prevalence of PCOS among Chinese women of childbearing age[16]. The sample size for this study was determined using the ClinCalc.com Sample Size Calculator (https://clincalc.com/stats/samplesize.aspx), with an effect size of 0.95, a Type I error rate of 0.05, and a power of 95%. Based on the 1:2 matching criterion, a minimum of 157 subjects was required for the case group and 314 for the control group. Ultimately, this study included 210 subjects in the case group and 420 in the control group, for a total of 630 participants. The study adhered to the Declaration of Helsinki and STROBE guidelines for research reporting, and received approval from the Ethics Review Committee of the University of South China (Approval No. 2023NHHL054). All participants voluntarily consented to the study and provided signed informed consent.

Questionnaires

A General information questionnaire, developed by the research team after an extensive literature review and consultations with gynecologists, endocrinologists, and epidemiologists, was employed for data collection. The survey content primarily covered basic demographic information, menstrual history, lifestyle factors, family medical history, environmental exposures, and other relevant aspects. Demographic information included age, body mass index (BMI), occupation, education level, place of residence, and birth weight. Menstrual history encompassed age at menarche and menstrual cycle regularity. Lifestyle factors included smoking status, passive smoking, alcohol and tea consumption, dietary habits, and physical activity. Family medical history covered conditions such as PCOS, premature baldness, diabetes, infertility, and other diseases. Family medical history included details of PCOS, premature baldness, diabetes, infertility, and menstrual disorders. Environmental exposures primarily encompassed occupational and residential exposure histories. In this study, 20 study participants were randomly selected for pre-survey in December 2023 from the group that met the inclusion criteria, and the questionnaire was revised and improved based on the feedback obtained from the pre-survey. The complete content of the questionnaire is available in the supplementary material (2, 3, 4).

The Pittsburgh Sleep Quality Index (PSQI), developed by Buysse et al. at the Pittsburgh Medical Center in 1989, is a widely utilized tool for evaluating sleep quality over the past month. It has demonstrated good internal consistency, with a Cronbach's alpha of 0.84. The PSQI is frequently employed to assess sleep disturbances and quality over a 30-day period. The PSQI assesses 7 key dimensions: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleeping medications, and daytime dysfunction. It comprises 19 self-rated items scored on a 4-point Likert scale (0–3). The total PSQI score ranges from 0 to 21, with higher scores indicating poorer sleep quality. A total score above 7 indicates the presence of a potential sleep disorder[17].

The GAD-7 is a concise and effective tool for detecting generalized anxiety, developed by Dr. Kroenke et al. in the U.S. in 2005, based on the criteria from the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). This scale exhibits strong internal consistency, with a Cronbach's alpha of 0.92, and good test–retest reliability, with a coefficient of 0.83. The GAD-7 consists of seven items, each rated on a 4-point Likert scale (0–3), with response options ranging from "not at all" to "nearly every day." Scores range from 0 to 21, with higher scores indicating greater levels of anxiety. A score of 5 serves as the cut-off, with scores of ≥ 5 indicating the presence of anxiety symptoms[18].

Patients' Health Questionnaire (Patients' Health Questionnaire Depression Scale-9 item, PHQ-9) The PHQ-9 scale is a simple scale developed by Dr. KROENKE et al. in the United States in 1999 according to the Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition) (DSM-IV), It is a widely recognized and efficient

tool for detecting depression. The scale demonstrates strong internal consistency with a Cronbach's alpha of 0.86, and test–retest reliability of 0.84. The PHQ-9 includes 9 items, each rated on a 4-point Likert scale, with options ranging from "not at all" to "nearly every day.", Items are scored on a 4-point Likert scale (0–3), resulting in a total score ranging from 0 to 27, where higher scores reflect more severe depressive symptoms. A score of 5 serves as the cut-off, with scores of ≥ 5 indicating the presence of depressive symptoms[19].

Definition of relevant variables

Smoking: Based on WHO guidelines, smoking is defined as continuous or cumulative smoking for a period exceeding 6 months[20].

Passive smoking: Defined as the inhalation of secondhand smoke for 15 min or more on at least one day per week[21].

Alcohol consumption: Defined as drinking alcohol at least three times per week for a period exceeding six months.

Tea drinking: tea drinking is defined as the consumption of tea at least three times per week for a continuous period of more than six months.

Regular exercise: regular exercise is defined as engaging in physical activity at least three times per week, with each session lasting 30 min or more, sustained for a period of at least one year.

Body Mass Index (BMI): BMI was categorized based on the World Health Organization (WHO) guidelines into two groups: ≥ 25 kg/m^2 and < 25 kg/m^2[22].

Definitions variables are provided in Supplementary Table S1.

Quality control

All researchers underwent standardized training during both the data collection and survey analysis phases. The questionnaire items were standardized according to strict adherence to the sample's inclusion and exclusion criteria, ensuring clarity and operationalization. Participants who were able to complete the questionnaire independently did so, while those unable to do so received assistance from the researchers. Upon completion, questionnaires were verified on site. Incomplete entries were filled out with assistance, while those with significant missing data were excluded. Data were double-entered and verified by two independent researchers to ensure accuracy and authenticity.

Variable screening

DAGs serve as a crucial theoretical framework for the analysis of disease causality in epidemiological research. They construct causal networks based on theoretical causal relationships to identify appropriate variables for inclusion in the model[23]. Employing DAGs to represent causal relationships between exposure and outcome is a critical prerequisite for regression analysis. This approach clarifies barrier points and connection points between variables, enabling in-depth variable screening to elucidate the underlying causal relationships. This study will employ DAGitty 3.1 software to construct a causal theoretical model using DAGs, aiming to identify independent and confounding variables, as well as connection and barrier points, thereby clarifying the factors influencing PCOS.

Statistical analysis

All statistical analyses were conducted on the dataset "original data.xlsx." Data were independently entered into an Excel spreadsheet by two individuals, and subsequently cross-checked by a second set of individuals to ensure accuracy and prevent entry errors. Data analysis was conducted using SPSS version 26.0, with normality of continuous variables assessed via the Shapiro–Wilk test. Normally distributed data were presented as mean ± standard deviation (± s), with between-group differences assessed using the t-test or ANOVA. Non-normally distributed data were expressed as median (interquartile range) M (P25, P75), and group differences were analyzed using non-parametric tests. Categorical data were described as proportions or rates, and comparisons were made using the chi-square test, corrected chi-square test, or Fisher's exact test as appropriate. DAGs were utilized for variable screening, combining variable relationships based on existing literature and clinical knowledge, variables are represented by circles, causal relationships are connected by arrows, and the entire graph does not allow any loops to be formed, ultimately providing a transparent and reproducible framework for causal analysis for PCOS risk factor identification and propensity score matching was applied to adjust for confounding variables. Multivariate logistic regression was employed to analyze factors influencing PCOS, with statistical significance set at $P < 0.05$.

An overview of the study process is presented in Fig. S1.

Result

Baseline characteristics

A total of 630 participants were included in the study, with 210 allocated to the case group and 420 to the control group. No statistically significant differences were observed between the two groups in terms of age, place of residence, occupation, smoking, passive smoking, tea consumption, dietary habits, regular physical activity, family history of premature baldness, diabetes, menstrual disorders, infertility, or occupational exposure to petrochemicals and metals ($P > 0.05$). However, statistically significant differences ($P < 0.05$) were identified between the two groups regarding education level, obesity, birth weight, age of menarche, alcohol consumption, sleep disturbances, family history of PCOS, proximity to petrochemical plants or garbage dumps within 300 m of residence, as well as anxiety and depression (Table 1).

Variable	Case group($n=210$)	Control group($n=420$)	χ^2\t	*P*
Age (years)	25.20 ± 3.24	24.86 ± 3.33	− 1.195[a]	0.193
Residential area			1.053	0.305
Urban (City)	115(54.8)	248(59.0)		
Rural (Countryside)	95(45.2)	172(41.0)		
Educational attainment			10.387	0.016
Below junior high school	29(13.8)	54(12.9)		
Middle or high school	70(33.3)	110(26.2)		
College (Post-secondary)	45(21.4)	141(33.6)		
Undergraduate degree and above	66(37.4)	115(27.4)		
Occupational type			0.323	0.851
Cognitively oriented	113(53.8)	236(56.2)		
Physically oriented	73(34.8)	138(32.9)		
Unemployed	24(11.4)	46(11.0)		
BMI			63.489	< 0.001
< 25 kg/m^2	78(37.1)	295(70.2)		
≥ 25 kg/m^2	132(62.9)	125(29.8)		
Birth weight (kg)	3.15 ± 0.50	3.39 ± 0.71	4.224[a]	< 0.001
Age at menarche	13.09 ± 1.29	13.63 ± 1.30	4.956[a]	< 0.001
Smoking			0.090	0.764
Yes	17(3.3)	16(3.8)		
No	203(96.7)	404(96.2)		
Passive smoking			1.422	0.233
Yes	85(40.5)	191(45.5)		
No	125(59.5)	229(54.5)		
Alcohol consumption			20.818	< 0.001
Yes	62(29.5)	60(14.3)		
No	145(69.0)	360(85.7)		
Tea drinking			0.606	0.436
Yes	65(31.0)	143(34.0)		
No	145(69.0)	277(66.0)		
Dietary preference			2.070	0.355
Meat-based diet	39(18.6)	60(14.3)		
Vegetarian diet	25(11.9)	57(13.6)		
Mixed diet (Meat and vegetables)	146(69.5)	303(72.1)		
Regular exercise			2.088	0.148
Yes	60(28.6)	144(34.3)		
No	150(71.4)	276(65.7)		
Sleep disturbance			70.213	< 0.001
Yes	114(54.3)	89(21.2)		
No	96(45.7)	331(78.8)		
Family history of PCOS			28.917	< 0.001
Yes	23(11.0)	6(1.4)		
No	187(89.0)	414(98.6)		
Family history of premature baldness			1.326	0.249
Yes	45(21.4)	74(17.6)		
No	165(78.6)	346(82.4)		
Family history of diabetes			2.858	0.091
Yes	37(17.6)	53(12.6)		
No	173(82.4)	367(87.4)		
Family history of menstrual disorders			0.864	0.353
Yes	20(9.5)	31(7.4)		
No	190(90.5)	389(92.6)		
Family history of infertility			2.814	0.093
Yes	17(8.1)	20(4.8)		
No	193(91.9)	400(95.2)		
Occupational exposure to petrochemicals and metals			2.044	0.153
Continued				

Variable	Case group($n=210$)	Control group($n=420$)	χ^2\t	*P*
Yes	25(11.9)	68(16.2)		
No	185(88.1)	352(83.8)		
Proximity to petrochemical plant or landfill (within 300 m)			5.432	0.020
Yes	32(15.2)	38(9.0)		
No	178(84.8)	382(91.0)		
Anxiety symptoms			50.709	< 0.001
Yes	77(8.1)	52(12.4)		
No	193(91.9)	368(87.6)		
Depression Symptoms			29.521	< 0.001
Yes	64(30.5)	53(12.6)		
No	146(69.5)	367(87.4)		

Table 1. Comparative analysis of baseline demographic and clinical data between the two groups. BMI, body mass index; [a] represents the t-value from the statistical analysis.

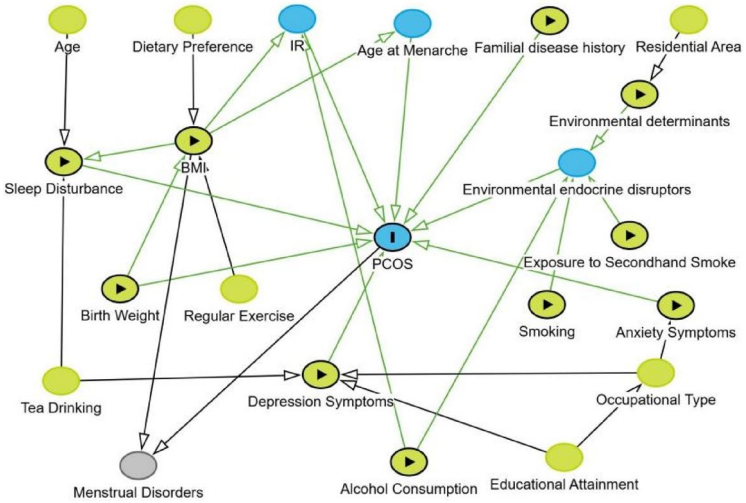

Fig. 1. Directed acyclic graph (DAG) model of PCOS in women of reproductive age.

Variable screening

Relevant variables were systematically identified and screened using the DAG methodology. In accordance with DAG theory, variables that could not be collected due to practical constraints were still represented in the DAG. As illustrated in Fig. 1, age at menarche, insulin resistance, and exposure to environmental endocrine disruptors were identified as mediating variables, whereas menstrual cycle disorders were classified as collider variables and therefore excluded from the model analysis.

Note: Variables designated in green with a positive sign represent exposure factors, while those in green without a sign denote confounders. Similarly, signed blue variables indicate outcome variables, whereas unsigned blue variables function as mediators between exposure and outcome. Lastly, grey variables are identified as collision variables.

Baseline data of the two groups after propensity score matching

Based on the DAG analysis, variables including age, occupation, regular exercise, education, place of residence, per capita family income, dietary habits, and tea drinking habits were identified as confounders. We consider statistical efficacy and confounding control of variables, adjusted for 1:1 nearest neighbor matching, with calipers set to 0.02. As a result, 210 participants were selected for both the PCOS group and the control group. After matching, there were no statistically significant differences ($P > 0.05$) between the two groups in terms of age, occupation, regular exercise, education, place of residence, per capita monthly family income, dietary habits, or tea drinking habits, indicating that the groups were well-balanced (Table 2).

Multifactorial analysis of risk factors for PCOS

Tests of covariance for statistically significant independent variables in the one-way analysis of variance showed that the variance inflation factors were all < 5, suggesting that there was no multicollinearity between the respective variables. A multifactorial logistic regression analysis was conducted using the presence of PCOS

Variable	Case group($n=210$)	Control group($n=210$)	χ^2\t	*P*
Age (years)	25.20 ± 3.24	25.08 ± 3.32	− 0.410[a]	0.682
Residential area			1.178	0.278
Urban (City)	115(54.8)	126(60.0)		
Rural (Countryside)	95(45.2)	84(40.0)		
Educational attainment			6.945	0.076
Below junior high school	29(13.8)	25(11.9)		
Middle or high school	70(33.3)	59(28.1)		
College (Post-secondary)	45(21.4)	69(32.9)		
Undergraduate degree and above	66(37.4)	57(27.1)		
Occupational type			0.649	0.723
Cognitively oriented	113(53.8)	113(53.8)		
Physically oriented	73(34.8)	68(32.4)		
Unemployed	24(11.4)	29(13.8)		
BMI			44.235	< 0.001
< 25 kg/m^2	78(37.1)	146(69.5)		
≥ 25 kg/m^2	132(62.9)	64(30.5)		
Birth weight (kg)	3.15 ± 0.50	3.38 ± 0.70	3.875[a]	< 0.001
Age at menarche	13.09 ± 1.29	13.68 ± 1.32	4.403[a]	< 0.001
Smoking			0.260	0.610
Yes	17(3.3)	9(4.3)		
No	203(96.7)	201(95.7)		
Exposure to secondhand smoke			0.010	0.921
Yes	85(40.5)	86(41.0)		
No	125(59.5)	124(59.0)		
Alcohol consumption			11.440	< 0.001
Yes	62(29.5)	33(15.7)		
No	145(69.0)	177(84.3)		
Tea drinking			0.011	0.916
Yes	65(31.0)	66(31.4)		
No	145(69.0)	144(68.6)		
Dietary preference			2.547	0.280
Meat-based diet	39(18.6)	30(14.3)		
Vegetarian diet	25(11.9)	34(16.2)		
Mixed diet (Meat and vegetables)	146(69.5)	146(69.5)		
Regular exercise			0.183	0.669
Yes	60(28.6)	64(30.5)		
No	150(71.4)	146(69.5)		
Sleep disturbance			5.959	0.015
Yes	114(54.3)	89(42.4)		
No	96(45.7)	121(57.6)		
Family history of PCOS			14.289	< 0.001
Yes	23(11.0)	4(1.9)		
No	187(89.0)	206(98.1)		
Family history of premature baldness			1.239	0.266
Yes	45(21.4)	36(17.1)		
No	165(78.6)	174(82.9)		
Family history of diabetes			2.725	0.099
Yes	37(17.6)	25(11.9)		
No	173(82.4)	185(88.1)		
Family history of menstrual disorders			2.165	0.141
Yes	20(9.5)	12(5.7)		
No	190(90.5)	198(94.3)		
Family history of infertility			4.419	0.036
Yes	17(8.1)	7(3.3)		
No	193(91,9)	203(96.7)		
Occupational exposure to petrochemicals and metals			1.597	0.206
Continued				

Variable	Case group($n=210$)	Control group($n=210$)	χ^2\t	*P*
Yes	25(11.9)	34(16.2)		
No	185(88.1)	176(83.8)		
Proximity to petrochemical plant or landfill (within 300 m)			0.995	0.319
Yes	32(15.2)	25(11.9)		
No	178(84.8)	185(88.1)		
Anxiety symptoms			36.618	< 0.001
Yes	77(8.1)	24(11.4)		
No	193(91.9)	186(88.6)		
Depression symptoms			1.434	0.231
Yes	64(30.5)	53(25.2)		
No	146(69.5)	157(74.8)		

Table 2. Baseline data of the case and control groups after propensity score matching. [a] represents the t-value from the statistical analysis.

Variable	B	SE	Waldχ^2	*P*	OR(95%*CI*)
Body mass Index(BMI)	1.408	0.235	35.967	< 0.001	4.088 (2.580–6.476)
Alcohol consumption	0.836	0.284	8.624	0.003	2.305 (1.320–4.024)
Sleep disturbance	0.291	0.306	0.907	0.341	1.338 (0.735–2.438)
Family history of PCOS	1.867	0.602	9.603	0.002	6.468 (1.986–21.067)
Family history of infertility	0.725	0.531	1.865	0.172	2.065 (0.729–5.847)
Birth weight (kg)	− 0.451	0.191	5.554	0.018	0.637 (0.438–0.927)
Anxiety symptoms	1.590	0.292	29.667	< 0.001	4.905 (2.768–8.693)
Smoking	− 0.260	0.640	0.165	0.685	0.771 (0.220–2.703)
Exposure to secondhand smoke	0.051	0.234	0.047	0.828	1.502 (0.666–1.633)
Occupational exposureto petrochemicals and metals	− 0.220	0.341	0.417	0.519	0.803 (0.412–1.565)
Proximity to petrochemical plant or landfill (within 300 m)	0.392	0.326	1.451	0.228	1.481 (0.782–2.804)
Family history of premature baldness	0.365	0.333	1.198	0.274	1.440 (0.750–2.766)
Family history of diabetes	0.283	0.371	0.579	0.447	1.327 (0.641–2.747)
Family history of menstrual disorders	0.640	0.437	2.145	0.143	1.897 (0.805–4.467)
Depression symptoms	− 0.028	0.336	0.007	0.933	0.972 (0.503–1.880)

Table 3. Multivariate logistic regression analysis of influencing factors of PCOS.

as the dependent variable, with exposure factors identified through the DAG as the independent variables. The analysis revealed that obesity, alcohol consumption, family history of PCOS, birth weight, and anxiety were significantly associated with PCOS. ($P<0.05$) , statistical analysis revealed no significant correlation between Sleep Disturbance, Family History of Infertility, Smoking, Exposure to Secondhand Smoke, Occupational Exposure to Petrochemicals and Metals, Proximity to Petrochemical Plant or Landfill (within 300 m), Family History of Premature Baldness, Family History of Diabetes, Family History of Menstrual Disorders, Depression Symptoms and PCOS (Table 3).

Discussion

The findings from this study demonstrate that a BMI ≥ 25 kg/m^2 is a significant risk factor for PCOS. Overweight and obesity have long been considered to be significantly associated with PCOS[24,25], a relationship that was reaffirmed in our study. Epidemiological studies have reported that over 60% of PCOS patients are overweight or obese[26]. In this study, 62.9% of the PCOS patients were classified as obese, a statistically significant difference compared to the control group. Obesity is a key contributor to insulin resistance, the central pathogenic mechanism of PCOS. Elevated insulin levels promote excessive androgen production in the ovaries, triggering symptoms such as hirsutism, acne, and menstrual irregularities[27]. In addition, aromatization (conversion of androgens to estrogens) in adipose tissue leads to elevated estrogen levels, and excess estrogen disrupts the normal regulation of the menstrual cycle and ovulation. Additionally, adipose tissue secretes pro-inflammatory cytokines, including TNF-α and IL-6. These inflammatory mediators intensify insulin resistance and contribute to ovarian follicular dysfunction. Collectively, these processes facilitate the development and progression of PCOS[28].

Our research indicates that alcohol consumption is consistent with PCOS. Our results are consistent with the findings reported by Angelis et al.[29], Alcohol intake disrupts the body's capacity to regulate blood glucose, resulting in fluctuations and abnormal insulin secretion. This disruption may exacerbate disease-related

symptoms by promoting additional hormonal imbalances, as PCOS is closely linked with insulin resistance[30]. Furthermore, alcoholic beverages are known to contain specific environmental toxins, such as perfluorinated and polyfluoroalkyl substances (PFAS). Recent studies have demonstrated that exposure to PFAS may significantly elevate the risk of developing PCOS[31].

The familial aggregation of PCOS indicates a significant genetic component. Women with a familial history of PCOS are at a higher risk of developing the condition which is generally consistent with previous research[32,33]Prior studies have identified numerous genetic variants linked to PCOS, many of which influence androgen secretion, insulin regulation, and follicular development. For instance, variants in genes related to gonadotropin secretion (Follicle-Stimulating Hormone Receptor, Luteinizing Hormone/Choriogonadotropin Receptor, Follicle-Stimulating Hormone Beta Subunit) and androgen biosynthesis (DENN Domain Containing Protein 1A) have been correlated with an elevated risk of developing PCOS[34]. Moreover, women with a history of PCOS can cause fetal androgen exposure during pregnancy, contributing to the reprogramming of the fetal ovaries, which alters gene pathways and mitochondrial function at birth. This process may lead to the development of a lean phenotype of PCOS later in life[35].

The findings of this study indicate that low birth weight serves as a significant risk factor for PCOS. This is consistent with previous studies[36–38]. Intrauterine growth restriction (IUGR) typically results in low birth weight due to the fetus's insufficient supply of nutrition and oxygen. This unfavorable environment results in what is known as "fetal programming," where the fetus adapts to limited resources by modifying its metabolic and endocrine pathways, and fetal programming induces enduring alterations in hormone regulation, including those involved in glucose metabolism and ovarian function, thereby elevating the risk of developing PCOS[39]. Furthermore, infants born with low birth weight often undergo rapid catch-up growth following delivery. This rapid catch-up growth, combined with increased visceral adiposity, heightens the risk of insulin resistance and the emergence of metabolic syndrome, potentially resulting in the development of PCOS[40].

Our findings further emphasize the significant association between anxiety and PCOS. Anxiety and chronic stress activate the hypothalamic–pituitary–adrenal (HPA) axis, resulting in heightened secretion of cortisol and other stress-related hormones. Prolonged elevation of cortisol disrupts the balance of estrogen and progesterone while elevating androgen levels, thereby exacerbating the symptoms of PCOS[41]. Additionally, anxiety can contribute to unhealthy lifestyle choices, such as emotional eating and sedentary behavior. These behaviors may result in obesity and endocrine disruption, thereby establishing a vicious cycle that further facilitates the development of PCOS[42].This study has several limitations. Firstly, the case–control study design necessitated reliance on patients' self-reported data, which may introduce recall bias and consequently affect the study's validity. Therefore, high-quality prospective cohort studies are essential for further validation in the future. Secondly, the sample population was exclusively derived from Hengyang City, and the sample size was relatively small, which may introduce selection bias and limit the external validity of the findings. Therefore, future research should involve a more extensive multicenter study to gain a comprehensive understanding of PCOS across diverse populations and environments, thereby enhancing the generalizability of the results.

statistical analysis revealed no significant correlation between Sleep Disturbance, Family History of Infertility, Smoking, Exposure to Secondhand Smoke, Occupational Exposure to Petrochemicals and Metals, Proximity to Petrochemical Plant or Landfill (within 300 m), Family History of Premature Baldness, Family History of Diabetes, Family History of Menstrual Disorders, Depression Symptoms and PCOS. These null findings appear inconsistent with some previous research, and several factors may help explain the discrepancies.

First, the limited statistical power of our study, due to an insufficient sample size, may have restricted our ability to detect modest but clinically meaningful effect sizes. Furthermore, exposure–response relationships regarding occupational hazards could not be quantified, particularly given the absence of biomonitoring data on specific chemical concentrations and exposure durations. Potential recall bias in self-reported family history may have led to non-differential misclassification and attenuation of associations. In addition, unmeasured gene–environment interactions may contribute to disease susceptibility, as suggested by emerging evidence on epigenetic mechanisms in PCOS pathogenesis. Finally, certain environmental factors may exert context-dependent effects that manifest only in the presence of specific genetic polymorphisms, such as androgen receptor variants. Collectively, these limitations highlight the complexity of disentangling environmental and genetic contributions to PCOS.

Conclusion

In conclusion, a diagnosis of PCOS is likely associated with several factors, including obesity, alcohol consumption, family history of PCOS, low birth weight, and anxiety. Identifying these factors is crucial for informing management of PCOS, healthcare professionals can develop intervention programs targeting obesity, alcohol consumption, and anxiety factors to reduce the incidence of PCOS and promote women's physical and mental health. It is suggested that a multicenter, large-sample longitudinal study could be conducted in the future to further explore the influencing factors of PCOS in order to guide disease prevention.

Data availability

Due to privacy/ethical restrictions, the raw dataset ("original data.xlsx") cannot be made publicly available. Requests for access may be directed to the corresponding author.

Received: 21 November 2024; Accepted: 18 September 2025
Published online: 23 October 2025

References

1. Joham, A. E. et al. Polycystic ovary syndrome. *Lancet Diabetes Endocrinol.* **10**, 668–680. https://doi.org/10.1016/S2213-8587(22)00163-2 (2022).
2. Walters, K. A. et al. Key signalling pathways underlying the aetiology of polycystic ovary syndrome. *J. Endocrinol.* **255**, R1–R26. https://doi.org/10.1530/joe-22-0059 (2022).
3. Stener-Victorin, E. et al. Animal models to understand the etiology and pathophysiology of polycystic ovary syndrome. *Endocr. Rev.* https://doi.org/10.1210/endrev/bnaa010 (2020).
4. Balen, A. H. et al. The management of anovulatory infertility in women with polycystic ovary syndrome: An analysis of the evidence to support the development of global WHO guidance. *Hum. Reprod. Update* **22**, 687–708. https://doi.org/10.1093/humupd/dmw025 (2016).
5. van Baal, L. & Tan, S. Polycystic ovary syndrome as a gender-specific cardiometabolic risk factor. *Innere Med.* https://doi.org/10.1007/s00108-023-01529-7 (2023).
6. Helvaci, N. & Yildiz, B. O. The impact of ageing and menopause in women with polycystic ovary syndrome. *Clin. Endocrinol.* **97**, 371–382. https://doi.org/10.1111/cen.14558 (2022).
7. Barber, T. M. & Franks, S. Obesity and polycystic ovary syndrome. *Clin. Endocrinol.* **95**, 531–541. https://doi.org/10.1111/cen.14421 (2021).
8. Rao, V. S. et al. A global survey of ethnic indian women living with polycystic ovary syndrome: Co-morbidities, concerns, diagnosis experiences, quality of life, and use of treatment methods. *Int. J. Environ. Res. Public Health* https://doi.org/10.3390/ijerph192315850 (2022).
9. Basnet, J., Rezq, S., Huffman, A. M., Cardozo, L. L. Y. & Romero, D. G. High-fat diet exacerbates androgen-mediated obesity and white adipose tissue hypertrophy in a mouse model of polycystic ovary syndrome. *Faseb J.* https://doi.org/10.1096/fasebj.2022.36.S1.R3899 (2022).
10. Mishra, G., Mohanty, S. K. & Biswal, P. K. A cross-sectional analysis of prevalence of PCOS and risk factors associated with it among young women at a tertiary care hospital. *J. Cardiovasc. Dis. Res.* **14**, 2133–2136. https://doi.org/10.31838/jcdr.2023.14.02.275 (2023).
11. Zhang, J. et al. Environmental risk factors for women with polycystic ovary syndrome in china: A population-based case-control study. *J. Biol. Regul. Homeost. Agents* **28**, 203–211 (2014).
12. Tennant, P. W. G. et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: Review and recommendations. *Int. J. Epidemiol.* **50**, 620–632. https://doi.org/10.1093/ije/dyaa213 (2021).
13. Peng, X., Tang, T., Wu, M., Tan, L. & Pan, Y. Network analysis of risk and protective factors for suicidal ideation in adolescents. *Chil. Youth Serv. Rev.* https://doi.org/10.1016/j.childyouth.2024.107458 (2024).
14. Koivisto, S.-T., Laaksonen, T., Helenius, I., Vasara, H. & Stenroos, A. Epidemiology and risk factors for premature physeal closure in distal femur fractures. *Acta Orthop.* **94**, 348–353. https://doi.org/10.2340/17453674.2023.13654 (2023).
15. Azziz, R. Diagnosis of polycystic ovarian syndrome: The rotterdam criteria are premature. *J. Clin. Endocrinol. Metab.* **91**, 781–785. https://doi.org/10.1210/jc.2005-2153 (2006).
16. Yang, R. et al. Changes in the prevalence of polycystic ovary syndrome in China over the past decade. *Lancet Reg. Health West Pac.* **25**, 100494. https://doi.org/10.1016/j.lanwpc.2022.100494 (2022).
17. Buysse, D. J., Reynolds, C. F. 3rd., Monk, T. H., Berman, S. R. & Kupfer, D. J. The pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Res.* **28**, 193–213. https://doi.org/10.1016/0165-1781(89)90047-4 (1989).
18. Spitzer, R. L., Kroenke, K., Williams, J. B. & Löwe, B. A brief measure for assessing generalized anxiety disorder: The GAD-7. *Arch. Intern. Med.* **166**, 1092–1097. https://doi.org/10.1001/archinte.166.10.1092 (2006).
19. Kroenke, K., Spitzer, R. L. & Williams, J. B. The PHQ-9: Validity of a brief depression severity measure. *J. Gen. Intern. Med.* **16**, 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x (2001).
20. Rigotti, N. A., Kruse, G. R., Livingstone-Banks, J. & Hartmann-Boyce, J. Treatment of tobacco smoking: A review. *JAMA* **327**, 566–577. https://doi.org/10.1001/jama.2022.0395 (2022).
21. Toyama, N. et al. Associations between sleep bruxism, sleep quality, and exposure to secondhand smoke in Japanese young adults: A cross-sectional study. *Sleep Med.* **68**, 57–62. https://doi.org/10.1016/j.sleep.2019.09.003 (2020).
22. Nakao, Y. M. et al. Risks and benefits of oral anticoagulants for stroke prophylaxis in atrial fibrillation according to body mass index: Nationwide cohort study of primary care records in England. *EClinicalMedicine* **54**, 101709. https://doi.org/10.1016/j.eclinm.2022.101709 (2022).
23. Digitale, J. C., Martin, J. N. & Glymour, M. M. Tutorial on directed acyclic graphs. *J. Clin. Epidemiol.* **142**, 264–267. https://doi.org/10.1016/j.jclinepi.2021.08.001 (2022).
24. Mohapatra, I. & Samantaray, S. R. BMI and polycystic ovary syndrome: Demographic trends in weight and health. *Cureus* **16**, 55439. https://doi.org/10.7759/cureus.55439 (2024).
25. Barber, T. M., Hanson, P., Weickert, M. O. & Franks, S. Obesity and polycystic ovary syndrome: Implications for pathogenesis and novel management strategies. *Clin. Med. Insights Reprod. Health* https://doi.org/10.1177/1179558119874042 (2019).
26. Pirotta, S. et al. Obesity and the risk of infertility, gestational diabetes, and type 2 diabetes in polycystic ovary syndrome. *Semin. Reprod. Med.* **38**, 342–351. https://doi.org/10.1055/s-0041-1726866 (2020).
27. Zhao, H., Zhang, J., Cheng, X., Nie, X. & He, B. Insulin resistance in polycystic ovary syndrome across various tissues: An updated review of pathogenesis, evaluation, and treatment. *J. Ovarian. Res.* **16**, 9. https://doi.org/10.1186/s13048-022-01091-0 (2023).
28. Rudnicka, E. et al. Chronic low grade inflammation in pathogenesis of PCOS. *Int. J. Mol. Sci.* https://doi.org/10.3390/ijms22073789 (2021).
29. de Angelis, C. et al. Smoke, alcohol and drug addiction and female fertility. *Reprod. Biol. Endocrinol.* **18**, 21. https://doi.org/10.1186/s12958-020-0567-7 (2020).
30. Miyagi, S. et al. Moderate alcohol consumption is associated with impaired insulin secretion and fasting glucose in non-obese non-diabetic men. *J. Diabetes Investig.* **12**, 869–876. https://doi.org/10.1111/jdi.13402 (2021).
31. Zhan, W. et al. Environmental exposure to emerging alternatives of per- and polyfluoroalkyl substances and polycystic ovarian syndrome in women diagnosed with infertility: A mixture analysis. *Environ. Health Perspect.* **131**, 57001. https://doi.org/10.1289/ehp11814 (2023).
32. Boldis, B. V. et al. Comorbidities in women with polycystic ovary syndrome: A sibling study. *BMC Womens Health* **24**, 221. https://doi.org/10.1186/s12905-024-03028-9 (2024).
33. Helvaci, N. & Yildiz, B. O. Polycystic ovary syndrome as a metabolic disease. *Nat. Rev. Endocrinol.* **21**, 230–244. https://doi.org/10.1038/s41574-024-01057-w (2025).
34. Liu, M. et al. Expression of PCOS candidate genes in bovine fetal and adult ovarian somatic cells. *Reprod. Fertil.* **3**, 273–286. https://doi.org/10.1530/raf-22-0068 (2022).
35. Barsky, M. et al. Fetal programming of polycystic ovary syndrome: Effects of androgen exposure on prenatal ovarian development. *J. Steroid Biochem. Mol. Biol.* https://doi.org/10.1016/j.jsbmb.2021.105830 (2021).
36. Liu, D. et al. Fetal genome predicted birth weight and polycystic ovary syndrome in later life: A Mendelian randomization study. *Front. Endocrinol.* **14**, 1140499. https://doi.org/10.3389/fendo.2023.1140499 (2023).
37. Stracquadanio, M. & Ciotta, L. Low birth-weight is a PCOS risk factor for Southern-Italian women. *Gynecol. Endocrinol.* **33**, 373–377. https://doi.org/10.1080/09513590.2017.1283487 (2017).

38. Mumm, H., Kamper-Jørgensen, M., Nybo Andersen, A. M., Glintborg, D. & Andersen, M. Birth weight and polycystic ovary syndrome in adult life: A register-based study on 523,757 Danish women born 1973-1991. *Fertil. Steril.* **99**, 777–782. https://doi.org/10.1016/j.fertnstert.2012.11.004 (2013).
39. Beetch, M. & Alejandro, E. U. Placental mTOR signaling and sexual dimorphism in metabolic health across the lifespan of offspring. *Children* https://doi.org/10.3390/children8110970 (2021).
40. Diaz, M. et al. Circulating GDF15 concentrations in girls with low birth weight: Effects of prolonged metformin treatment. *Pediatr. Res.* **93**, 964–968. https://doi.org/10.1038/s41390-022-02175-9 (2023).
41. Moreno-Fernandez, R. D. et al. Social avoidance and altered hypothalamic-pituitary-adrenal axis in a mouse model of anxious depression: The role of LPA1 receptor. *Behav. Brain Res.* https://doi.org/10.1016/j.bbr.2023.114681 (2023).
42. Peng, H., Jing, L., Liu, Y., Tang, Y. & Wang, H. Physical activity, body anxiety, self-discipline, and emotional eating among Chinese women in the workforce. *Soc. Behav. Personal.* https://doi.org/10.2224/sbp.13060 (2024).

Author contributions
Study design was done by Z.Y.L and Y.H.S; Acquisition of datat was done by H.X.Y and Y.G; Analysis or interpretation of data; and Drafting of the manuscript was done by H.X.Y. Preparation and revision of the manuscript was shared by all authors. All authors consented to the publication of this study.

Funding
This work was supported by the National Natural Science Foundation of China (No. 82272383) and Hunan Provincial Postgraduate Scientific Research and Innovation Project(No. 20240088) .

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-20980-0.

Correspondence and requests for materials should be addressed to Y.S. or Z.L.

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.