Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Conditional Network of Demographic, Lifestyle, Reproductive, and Clinical Characteristics in Women with and Without Recorded Polycystic Ovary Syndrome

Zahra Amiri
Kerman University of Medical Sciences
Mahdi Tavakkoli
Kerman University of Medical Sciences
Ehsan Mosa Farkhani
Mashhad University of Medical Sciences
Neda Atashgahi
Mashhad University of Medical Sciences
Mitra Sotoude
Iranshahr University
Mohammadreza Balooch HasanKhani
m.balooch73@gmail.com

Kerman University of Medical Sciences

Research Article

Keywords: Polycystic ovary syndrome, Network analysis, Mixed graphical model, Reproductive health, Women's health

Posted Date: August 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10413484/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Polycystic ovary syndrome (PCOS) is a heterogeneous condition associated with reproductive, metabolic, lifestyle, and psychological characteristics. Conventional analyses commonly estimate associations between individual predictors and PCOS but do not describe the conditional dependence structure among all measured variables. This study characterized the conditional network of demographic, lifestyle, reproductive, clinical, and psychological variables in women with and without a recorded PCOS diagnosis.

Methods

This secondary analysis used data from the Sina Electronic Health Record (SinaEHR) system, affiliated with Mashhad University of Medical Sciences, Iran. The analysis included an outcome-enriched sample of 3,614 women aged 18–60 years: 1,807 women with a clinician-recorded PCOS diagnosis (ICD-10 code E28.2) and 1,807 randomly selected women without a recorded PCOS diagnosis. A mixed graphical model was used to estimate conditional dependencies among demographic, lifestyle, reproductive, metabolic, psychological, and clinical variables. Edge-weight accuracy and centrality stability were assessed using nonparametric and case-dropping bootstrap procedures.

Results

The estimated network contained a demographic–reproductive core centered on age, marital status, and number of childbirths, together with a diabetes–hypertension connection. PCOS had comparatively high strength centrality and its largest conditional-dependence weight was with place of residence; inverse conditional associations were estimated with age and physical activity. BMI, psychological distress, and smoking had relatively low strength in this selected-sample network. Bootstrap analyses suggested stable strength estimates, although the findings remain conditional on the outcome-enriched sampling design and the variables included in the model.

Conclusions

In this outcome-enriched case–control sample, recorded PCOS status was embedded in a wider conditional network of demographic, reproductive, lifestyle, and clinical characteristics. The analysis is exploratory and does not establish causal, temporal, preventive, or treatment relationships. Population-representative or appropriately weighted longitudinal analyses are needed to assess whether the observed network structure generalizes beyond the analyzed sample.

Introduction

Polycystic ovary syndrome (PCOS) is a heterogeneous, multisystem condition with reproductive, metabolic, cardiovascular, and psychological manifestations (1–4). Its clinical heterogeneity reflects interactions among endocrine, metabolic, genetic, inflammatory, developmental, and environmental factors (1, 4, 9–11). A recent global systematic review and meta-analysis estimated a pooled prevalence of approximately 12.1% among adult women using the Rotterdam criteria, with substantial geographic variation (2). Many affected women nevertheless experience delayed diagnosis, repeated healthcare consultations, limited information, and suboptimal care experiences (1, 5). Beyond ovulatory dysfunction, infertility, and menstrual irregularity, PCOS is associated with adverse pregnancy outcomes, cardiometabolic morbidity, anxiety, depression, and impaired health-related quality of life (1, 6–8).

The pathophysiology of PCOS involves genetic susceptibility, insulin resistance, hyperandrogenism, adiposity, lifestyle factors, and developmental and epigenetic mechanisms (9–12). These factors do not act independently and may contribute to distinct clinical phenotypes (1, 4, 17). Iranian population-based evidence also indicates a substantial burden of PCOS and cardiometabolic abnormalities, although estimates vary by diagnostic criteria, sampling frame, and study quality (13, 14). Population-specific analyses are therefore valuable, provided that their design and sampling limitations are explicitly considered.

Most epidemiological studies of PCOS use regression models to estimate adjusted associations between selected predictors and an outcome. These models are appropriate for specific etiologic or predictive questions, but they do not directly characterize the full conditional dependence structure among multiple measured variables (15, 16, 19). Network analysis offers a complementary, exploratory framework in which variables are represented as nodes and conditional dependencies as edges. This approach may be useful in PCOS because reproductive, metabolic, psychological, demographic, and lifestyle characteristics are interrelated (1, 17).

Network applications in PCOS research remain uncommon and have primarily focused on symptom-level relationships, including anxiety, depression, and sleep symptoms (18). To our knowledge, few studies have integrated a recorded PCOS diagnosis with demographic, lifestyle, reproductive, psychological, and clinical characteristics in a mixed graphical model. We therefore aimed to estimate and describe this conditional network in an outcome-enriched sample of Iranian women drawn from electronic health records, and to assess edge accuracy and centrality stability. Because the data were cross-sectional and sampled by PCOS status, the analysis was intended to describe conditional dependencies within the analyzed sample rather than causal pathways or a population-level network.

Methods

Study Design and Setting

This case–control study was a secondary analysis of electronic health records collected between March 1, 2017, and June 20, 2024, at healthcare facilities affiliated with Mashhad University of Medical Sciences, Iran. Data were extracted from the Sina Electronic Health Record (SinaEHR) system (20), which records structured demographic, behavioral, reproductive, and clinical information collected during routine healthcare encounters. Reporting was guided by the STROBE statement and published standards for cross-sectional network analyses (16).

Study Population, Case Ascertainment, and Control Selection

The source population comprised women aged 18–60 years registered in the SinaEHR system at healthcare centers affiliated with Mashhad University of Medical Sciences during the study period. Cases were identified by a clinician-recorded ICD-10 code for PCOS (E28.2). After records with missing values on any modeled variable were excluded, 1,807 cases remained. The analytical sample was then enriched to contain an equal number of cases and controls and therefore does not reproduce the PCOS prevalence in the source population.

Controls were randomly selected from women without a recorded PCOS diagnosis who attended the same healthcare centers during the same period for services not recorded as PCOS-related. Only women with complete data on all modeled variables were eligible. Eligible control records were assigned unique identifiers and randomly ordered using the RAND() function in Microsoft Excel (Microsoft Corp., Redmond, WA, USA); the first 1,807 records were selected. This procedure achieved a 1:1 case–control ratio but did not constitute individual matching.

Data Collection

Data were extracted from the SinaEHR system, an electronic health record database that routinely captures standardized demographic, lifestyle, reproductive, clinical, and psychological information during routine healthcare visits.

Demographic and Anthropometric Variables

Age was analyzed as a continuous variable in years (range, 18–60). Educational level was categorized as illiterate, elementary, diploma, or academic education. Employment status was categorized as employed, housewife, or unemployed; residence as city, city margin, metropolitan, or rural; and marital status as married or single. Body mass index (BMI; kg/m^2) was analyzed as a continuous variable in the mixed graphical model.

Lifestyle Factors

Lifestyle factors included physical activity, smoking, fast-food consumption, and cooking oil type. Physical activity was assessed by self-report and categorized according to the World Health Organization recommendations as meeting or not meeting the recommended level of physical activity (≥ 150 minutes of moderate-intensity or ≥ 75 minutes of vigorous-intensity aerobic activity per week), based on data recorded in the Sina Electronic Health Record (SinaEHR) system. Smoking status was determined from self-reported tobacco use during the preceding 12 months and classified as current smoker or non-smoker. Fast-food consumption was categorized into three levels (never, monthly, and weekly) according to participants' reported intake frequency. Cooking oil type was classified as liquid oil, mixed oil, or solid oil based on the predominant type of oil used for food preparation.

Reproductive Health Variables

Reproductive history was captured through the number of childbirths (0, 1, 2, 3, ≥ 4) and the number of prior abortions (0, 1, ≥ 2).

Medical Conditions

Diabetes (ICD-10 codes E10, E11) and hypertension (ICD-10 codes I10, I15) were identified from clinician-recorded diagnoses in the SinaEHR system.

Psychological Assessment

Symptoms of psychological distress (including anxiety and depressive symptoms) were assessed using the Kessler Psychological Distress Scale (K6), a brief self-report screening instrument recorded in the SinaEHR system. A total K6 score of ≥ 10 was used to classify participants as screening positive for psychological distress, consistent with the validated Persian version of the instrument(21).

Ascertainment of Recorded PCOS Status

Recorded PCOS status was ascertained from clinician-entered ICD-10 code E28.2 in SinaEHR. The available extract did not permit independent verification of whether every diagnosis fulfilled the 2023 International Evidence-based Guideline criteria (22), nor could women without a recorded code be confirmed to be free of undiagnosed PCOS. Accordingly, the modeled node represents recorded PCOS status rather than a uniformly re-adjudicated research diagnosis.

Statistical Analysis

Network analysis was conducted in R version 4.6.0 (R Foundation for Statistical Computing, Vienna, Austria) using the mgm, qgraph, and bootnet packages; readxl was used for data import. A pairwise mixed graphical model was fitted to 16 variables. Age and BMI were specified as Gaussian variables, and the remaining variables were specified as categorical with the number of levels defined by their observed

coding. The model estimates conditional dependencies within the joint distribution represented by the analyzed data.

The network was estimated using regularized model selection with the extended Bayesian information criterion (EBIC; γ = 0.25). Edge weights quantify conditional-dependence strength after conditioning on the other modeled variables. For edges involving multicategory variables, a single aggregate weight does not identify which category contrast drives the association, and its sign should not be interpreted as a simple positive or negative direction. A spring layout was used for visualization; node position is layout-dependent and was not treated as an inferential result.

Records with missing values on any modeled variable were excluded (complete-case analysis). The number of records excluded, the variable-specific amount of missingness, and comparisons between included and excluded records were not available in the analytical report and should be supplied to assess possible selection bias.

Node importance was summarized primarily using strength. Closeness and betweenness were also calculated but interpreted cautiously because these indices can be unstable and are sensitive to network structure. Edge-weight accuracy was assessed with 500 nonparametric bootstrap resamples, and centrality stability with 500 case-dropping bootstrap resamples. Correlation-stability (CS) coefficients were reported for the calculated centrality indices (23). These bootstrap procedures assess sampling variability within the selected analytical sample; they do not correct bias introduced by outcome-dependent sampling.

Ethical Considerations

This study was approved by the Ethics Committee of Mashhad University of Medical Sciences (IR.MUMS.FHMPM.REC.1402.234). All procedures were conducted in accordance with the Declaration of Helsinki and its subsequent amendments. The study was based on de-identified data extracted from the Sina Electronic Health Record (SinaEHR) system, and participant confidentiality was maintained throughout all stages of data management and analysis. Written informed consent had been obtained from all participants at the time of routine data collection and registration in the SinaEHR system.

Results

Study population

A total of 3,614 women were included in the final balanced analytical sample, consisting of 1,807 women with PCOS and 1,807 controls. Participants had a mean age of 34.92 ± 10.05 years and a mean BMI of 26.23 ± 5.20 kg/m². Most participants were married (85.7%) and housewives (69.0%), and nearly half lived in rural areas (45.5%). Additional demographic, lifestyle, reproductive, clinical, and psychological characteristics are summarized in Table 1.

Table 1
Descriptive characteristics of the study participants and network variables (N = 3614)

Characteristic	Category	N (%) / Mean ± SD
Age (years)	Mean ± SD	34.92 ± 10.05
	Range	18–60
BMI (kg/m²)	Mean ± SD	26.23 ± 5.20
	Range	15.10–48.60
Marital status	Single	516 (14.3)
	Married	3098 (85.7)
Educational level	Illiterate	244 (6.8)
	Elementary	1111 (30.7)
	Diploma	1832 (50.7)
	Academic	427 (11.8)
Occupation	Employed	618 (17.1)
	Housewife	2493 (69.0)
	Unemployed	503 (13.9)
Residence	City	674 (18.6)
	Margin city	662 (18.3)
	Metropolitan	634 (17.5)
	Rural	1644 (45.5)
Smoking	No	3381 (93.6)
	Yes	233 (6.4)
Physical activity	No	2672 (73.9)
	Yes	942 (26.1)
Fast food consumption	Low	1198 (33.1)
	Moderate	2183 (60.4)
	High	233 (6.4)
Oil type	Liquid	347 (9.6)
	Solid	1424 (39.4)

Characteristic	Category	N (%) / Mean ± SD
	Mixed	1843 (51.0)
Number of childbirths	0	871 (24.1)
	1	697 (19.3)
	2	1035 (28.6)
	3	591 (16.4)
	≥ 4	420 (11.6)
Number of abortions	0	2823 (78.1)
	1	561 (15.5)
	≥ 2	230 (6.4)
Diabetes	No	3472 (96.1)
	Yes	142 (3.9)
Hypertension	No	3382 (93.6)
	Yes	232 (6.4)
Psychological assessment	Normal (< 10)	2800 (77.5)
	Depression/Anxiety (≥ 10)	814 (22.5)
Polycystic ovary syndrome	No	1807 (50.0)
	Yes	1807 (50.0)

Network structure

The selected-sample network contained 49 of 120 possible pairwise edges after regularization (40.8%; Fig. 1). The complete weighted adjacency matrix is provided in Supplementary Table S1. The largest edge weights occurred among demographic–reproductive variables and between diabetes and hypertension. Because the sample was deliberately balanced on PCOS status, the network should be interpreted as the conditional-dependence structure of the analyzed sample rather than an estimate of the source-population network.

The largest edge weight connected marital status with number of childbirths (1.03), followed by residence with PCOS (0.70), diabetes with hypertension (0.68), age with number of childbirths (0.60), and marital status with number of abortions (0.57). Other comparatively large weights included age–hypertension (0.51), marital status–occupation (0.45), and education–occupation (0.37). Inverse conditional associations were estimated for the age–PCOS (− 0.32) and physical activity–PCOS (− 0.31) edges. The aggregate residence–PCOS weight does not indicate which residence category had higher or lower conditional probability of recorded PCOS.

Centrality profile

Strength centrality was highest for number of childbirths, marital status, and age (2.86, 2.77, and 2.65, respectively; Fig. 2). Occupation and PCOS also had comparatively high strength values (2.18 and 2.11). Among clinical variables, hypertension and diabetes had the highest strength values, whereas BMI, smoking, physical activity, and psychological distress had lower strength in this selected-sample network. Centrality is a descriptive property of the fitted network and should not be interpreted as causal importance or intervention priority.

Age had the highest betweenness value, followed by marital status and PCOS. Closeness was not interpreted because its case-dropping stability was poor. Betweenness was treated as secondary because of its lower stability and known sensitivity to small changes in estimated edges. PCOS ranked fifth in strength within the fitted network.

Domain-specific network patterns

The fitted network contained comparatively large connections among demographic and reproductive variables, particularly number of childbirths, marital status, and age. PCOS had its largest edge weight with residence, although the aggregate edge does not reveal the relevant category contrast. Diabetes and hypertension were also strongly connected. These patterns are descriptive and may reflect the study's sampling design, measurement choices, healthcare utilization, and recorded diagnoses as well as underlying clinical relationships.

Network stability

The robustness of the estimated network was examined using nonparametric and case-dropping bootstrap procedures. The nonparametric bootstrap showed relatively narrow confidence intervals around the strongest edge estimates, supporting the precision of the principal conditional associations, whereas weaker edges displayed greater sampling variability (Supplementary Figure S1).

In case-dropping bootstrap analyses, the CS coefficient was 0.75 for strength and expected influence, 0.517 for betweenness, and 0 for closeness (Supplementary Figure S2). These results suggest stable rank ordering for strength within the analyzed sample, but no inference about population representativeness follows from bootstrap stability. Expected influence should be interpreted only if the sign and coding of all contributing edge parameters are substantively defined; this requirement is not met by a simple aggregate weight for multicategory nodes.

Discussion

In this outcome-enriched case–control analysis of 3,614 women, the fitted network contained a demographic–reproductive core, a prominent diabetes–hypertension connection, and comparatively high strength for recorded PCOS status. The largest edge weight involving PCOS was with residence;

inverse conditional associations were estimated with age and physical activity. These findings describe conditional dependencies in the selected analytical sample and should not be interpreted as causal effects, population-level associations, or intervention targets.

The observed network structure also provides a rationale for modelling demographic characteristics such as marital status, place of residence, and occupation as network nodes rather than treating them solely as background covariates. Although Borsboom's original network theory conceptualized nodes primarily as symptoms, subsequent theoretical work has proposed that any variable may be represented as a node if it varies at the individual level and plausibly contributes to the development or maintenance of the condition under study, whereas variables reflecting only between-person differences (e.g., gender) are less appropriate for this purpose (24). Marital status, place of residence, and occupation satisfy these criteria because each may change over the life course and plausibly influence stress exposure, healthcare-seeking behavior, and access to diagnostic and reproductive services. Within this framework, the residence-PCOS edge represents a potentially meaningful conditional association rather than simply statistical adjustment for a background characteristic, although its causal significance requires confirmation in longitudinal or interventional studies. This interpretation is consistent with mixed graphical model studies that incorporated environmental exposures alongside clinical variables (25), as well as with Lunansky and colleagues' extension of symptom networks to include risk and protective factors as integral network nodes rather than external covariates (26).

Applications of network analysis to PCOS remain limited. A retrospective study of 1,068 women with PCOS examined the network structure of anxiety, depression, and sleep symptoms, identifying trouble relaxing and sad mood as the most central symptom nodes (27).Unlike that study, which focused exclusively on symptom-level relationships within women already diagnosed with PCOS, the present analysis positioned PCOS itself as one node within a broader demographic, reproductive, lifestyle, and clinical network. To our knowledge, this is among the first studies to examine a clinician-confirmed PCOS diagnosis within such a case–control mixed graphical model. Similar applications in other fields support this analytical framework. Isvoranu and colleagues integrated environmental risk factors, including cannabis use, developmental trauma, and urbanicity, with psychopathology measures using mixed graphical models (25), while Hoang and colleagues found that demographic characteristics occupied more central positions than dietary factors in a large Korean cohort(28).

Likewise, Chang and colleagues demonstrated that accounting for conditional relationships substantially altered the apparent centrality of psychiatric disorders compared with simple co-occurrence analyses (29). Collectively, these studies illustrate that mixed graphical models can reveal network structures that differ substantially from those identified using conventional association-based approaches, supporting their application to the complex demographic, reproductive, lifestyle, and clinical architecture of PCOS.

Within a mixed graphical model, an edge represents the association between two variables that remains after conditioning on all other variables in the network. Accordingly, the strong marital status-childbirth number and age-childbirth number edges indicate that reproductive history remains closely associated

with demographic characteristics after accounting for PCOS and all other variables included in the network. This interpretation is consistent with longitudinal evidence showing that women with PCOS have fewer children and a higher prevalence of nulliparity than women without PCOS (23% vs 18%), despite similar family-building aspirations, report infertility more frequently (51% vs 21% by 43 to 48 years of age), and are substantially more likely to undergo fertility treatment, including ovulation induction (22% vs 5%) and in vitro fertilization (17% vs 7%). In the same cohort, a later diagnosis of PCOS was associated with nearly twice the odds of advanced maternal age at first childbirth (adjusted OR 1.98, 95% CI 1.22 to 3.22) (30).Although the present case–control design cannot establish temporal ordering, these longitudinal findings provide a plausible clinical context for the close coupling of reproductive-history variables with demographic characteristics observed in the estimated network. This interpretation is further supported by a hospital-based case–control study from India, which reported marked differences in reproductive characteristics between women with PCOS and healthy controls and demonstrated that age, BMI, and marital status were associated with health-related quality of life among women with PCOS, with infertility emerging as the strongest predictor of impaired well-being (31). Together, these findings support the relevance of demographic and reproductive factors to the broader clinical context in which PCOS develops and is experienced.

The residence–PCOS edge had the largest aggregate weight among edges involving PCOS. Because residence had four categories, this weight indicates conditional dependence but does not identify which category contrast accounts for the edge. Residence may proxy healthcare access, referral pathways, socioeconomic context, or diagnostic opportunity, but these mechanisms were not measured. Moreover, outcome-dependent sampling and differential healthcare utilization may have influenced this edge. Category-specific parameters and conventional descriptive or regression analyses are needed before a directional interpretation can be made.

The comparatively large diabetes–hypertension edge is consistent with cardiometabolic clustering. However, this edge describes conditional dependence between two recorded diagnoses after conditioning on the other modeled variables; it does not show that the connection is specific to women with PCOS or that PCOS caused either condition. The 2023 International Evidence-based Guideline identifies dysglycemia and cardiovascular risk factors as important comorbidities in PCOS and recommends appropriate risk assessment (4, 22).

An inverse conditional association was estimated between age and recorded PCOS status. PCOS manifestations can change across the reproductive life course, but the present finding may also reflect diagnostic practices, cohort composition, healthcare use, or outcome-dependent sampling. Because the analysis was cross-sectional and diagnoses were not independently re-adjudicated, it cannot establish within-person phenotypic change or explain the observed age gradient (34).

An inverse conditional association was also estimated between physical activity and recorded PCOS status. This is compatible with prior evidence that physical activity is relevant to PCOS management (4, 35), but the cross-sectional edge cannot establish whether activity preceded diagnosis, changed after

diagnosis, or was affected by unmeasured confounding or reporting differences. The finding therefore supports hypothesis generation rather than a preventive or treatment-effect conclusion.

BMI had comparatively low strength in the fitted network despite extensive evidence that excess adiposity can amplify reproductive and metabolic manifestations of PCOS (34, 36, 37). Low network strength does not imply that BMI is clinically unimportant. It may reflect conditioning on correlated variables, the selected case–control distribution, the measurement scale, regularization, or indirect associations. The absence of a retained BMI–PCOS edge should therefore be reported as a model-specific result, not as evidence of no epidemiological association.

Psychological distress and smoking also had comparatively low strength. This should not be interpreted as evidence that psychological health or smoking is unimportant in PCOS. Centrality depends on the modeled variables, their coding, measurement quality, and sampling design. The binary K6 classification may additionally conceal symptom-level variation that network studies of anxiety and depression are designed to capture (18, 21).

Methodological considerations

The mixed graphical model allowed continuous and categorical variables to be modeled together, while regularization and bootstrap procedures quantified selected aspects of model uncertainty (23, 39). Nevertheless, edges are conditional associations rather than causal relationships. Furthermore, the 1:1 selection of cases and controls is outcome-dependent sampling. Without a design-based justification, inverse-probability weighting, or a sensitivity analysis using the available source cohort, the resulting network cannot be assumed to equal the source-population conditional-dependence structure.

Strengths and limitations

Strengths include use of a large electronic health-record dataset, explicit reporting of a balanced analytical sample, mixed-variable modeling, and bootstrap assessment of edge accuracy and centrality stability. Important limitations remain. First, the cross-sectional design precludes temporal and causal inference. Second, outcome-dependent 1:1 sampling may alter the joint distribution and network structure; therefore, the network is sample-specific unless supported by weighted or population-based sensitivity analyses. Third, PCOS was identified from clinician-recorded ICD-10 codes rather than uniform research re-adjudication, and controls may have included women with undiagnosed PCOS. Fourth, the numbers screened and excluded, variable-specific missingness, and characteristics of excluded records were not reported. Fifth, self-reported physical activity, smoking, diet-related variables, and psychological distress are susceptible to measurement error. Sixth, several categorical variables were coarsely coded, and aggregate MGM edge weights for multicategory nodes do not identify category-specific direction. Finally, all data came from one regional health-record system, limiting generalizability.

Clinical and public health implications

The results generate hypotheses about how recorded PCOS status co-occurs conditionally with demographic, reproductive, lifestyle, and clinical characteristics in this healthcare sample. They support continued attention to cardiometabolic assessment and physical activity in accordance with existing PCOS guidelines, but they do not independently demonstrate the effectiveness of an intervention or identify residence as a causal target. The residence finding requires category-specific analysis and investigation of healthcare access and diagnostic pathways.

Conclusion

In this outcome-enriched electronic health-record sample, recorded PCOS status formed part of a wider conditional network of demographic, reproductive, lifestyle, and clinical characteristics. The largest PCOS-related edge weight involved residence, and inverse conditional associations were estimated with age and physical activity; however, the cross-sectional case–control design and aggregate coding limit directional and causal interpretation. Population-representative or appropriately weighted longitudinal analyses, category-specific estimates, and transparent reporting of missing data are needed before the network can inform etiologic or clinical priorities.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Mashhad University of Medical Sciences (IR.MUMS.FHMPM.REC.1402.234). All procedures were conducted in accordance with the Declaration of Helsinki and its subsequent amendments. The study was based on de-identified data extracted from the Sina Electronic Health Record (SinaEHR) system, and participant confidentiality was maintained throughout all stages of data management and analysis. Written informed consent had been obtained from all participants at the time of routine data collection and registration in the SinaEHR system.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contribution

ZA and MT contributed equally to this work. ZA and MT conceived and designed the study, performed the statistical analyses, interpreted the data, and drafted the manuscript. EMF contributed to the study methodology, interpretation of the findings, and critically revised the manuscript. NA contributed to the study design, interpretation of the results, and critically revised the manuscript. MS supervised the study, contributed to the interpretation of the findings, and critically revised the manuscript. MRBHK supervised the project, contributed to the study design and interpretation of the results, and critically revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to acknowledge the use of ChatGPT for paraphrasing sentences and checking grammar. After using this tool/service, the authors reviewed and edited the content as necessary and take full responsibility for the final content of the publication.

Availability of data and materials

The datasets analyzed during the current study are not publicly available due to institutional and ethical restrictions on the Sina Electronic Health Record (SinaEHR) system but are available from the corresponding author on reasonable request and with permission of Mashhad University of Medical Sciences.

References

1. Stener-Victorin E, Teede H, Norman RJ, Legro R, Goodarzi MO, Dokras A, et al. Polycystic ovary syndrome. Nat Reviews Disease Primers. 2024;10(1):27.
2. Neven ACH, Forslund M, Ranasinha S, Sethi P, Dhungana RR, Mousa A, et al. Prevalence of polycystic ovary syndrome: a global and regional systematic review and meta-analysis. Hum Reprod Update. 2026;32(3):277–312.
3. Teede HJ, Khomami MB, Morman R, Laven JSE, Joham AE, Costello MF, et al. Polyendocrine metabolic ovarian syndrome, the new name for polycystic ovary syndrome: a multistep global consensus process. Lancet. 2026;407(10545):2329–39.
4. Teede HJ, Tay CT, Laven J, Dokras A, Moran LJ, Piltonen TT, et al. Recommendations from the 2023 International Evidence-based Guideline for the Assessment and Management of Polycystic Ovary Syndrome. Fertil Steril. 2023;120(4):767–93.
5. Sydora BC, Wilke MS, McPherson M, Chambers S, Ghosh M, Vine DF. Challenges in diagnosis and health care in polycystic ovary syndrome in Canada: a patient view to improve health care. BMC Womens Health. 2023;23(1):569.

6. Infante-Cano M, García-Muñoz C, Matias-Soto J, Pineda-Escobar S, Villar-Alises O, Martinez-Calderon J. The prevalence and risk of anxiety and depression in polycystic ovary syndrome: an overview of systematic reviews with meta-analysis. Arch Women Ment Health. 2025;28(3):475–89.
7. Bahri Khomami M, Shorakae S, Hashemi S, Harrison CL, Piltonen TT, Romualdi D, et al. Systematic review and meta-analysis of pregnancy outcomes in women with polycystic ovary syndrome. Nat Commun. 2024;15(1):5591.
8. Guan C, Zahid S, Minhas AS, Ouyang P, Vaught A, Baker VL, et al. Polycystic ovary syndrome: a risk-enhancing factor for cardiovascular disease. Fertil Steril. 2022;117(5):924–35.
9. Zhao Y, Pang J, Fang X, Yan Z, Yang H, Deng Q, et al. Causal relationships between modifiable risk factors and polycystic ovary syndrome: a comprehensive Mendelian randomization study. Front Endocrinol (Lausanne). 2024;15:1348368.
10. Gautam R, Maan P, Jyoti A, Kumar A, Malhotra N, Arora T. The Role of Lifestyle Interventions in PCOS Management: A Systematic Review. Nutrients. 2025;17(2).
11. Stener-Victorin E, Deng Q. Epigenetic inheritance of polycystic ovary syndrome - challenges and opportunities for treatment. Nat Rev Endocrinol. 2021;17(9):521–33.
12. Prosperi S, Chiarelli F. Insulin resistance, metabolic syndrome and polycystic ovaries: an intriguing conundrum. Front Endocrinol (Lausanne). 2025;16:1669716.
13. Farhadi-Azar M, Behboudi-Gandevani S, Rahmati M, Mahboobifard F, Khalili Pouya E, Ramezani Tehrani F, et al. The Prevalence of Polycystic Ovary Syndrome, Its Phenotypes and Cardio-Metabolic Features in a Community Sample of Iranian Population: Tehran Lipid and Glucose Study. Front Endocrinol (Lausanne). 2022;13:825528.
14. Jalilian A, Kiani F, Sayehmiri F, Sayehmiri K, Khodaee Z, Akbari M. Prevalence of polycystic ovary syndrome and its associated complications in Iranian women: A meta-analysis. Iran J Reprod Med. 2015;13(10):591–604.
15. Borsboom D, Deserno MK, Rhemtulla M, Epskamp S, Fried EI, McNally RJ, et al. Network analysis of multivariate data in psychological science. Nat Reviews Methods Primers. 2021;1(1):58.
16. Burger J, Isvoranu AM, Lunansky G, Haslbeck JMB, Epskamp S, Hoekstra RHA, et al. Reporting standards for psychological network analyses in cross-sectional data. Psychol Methods. 2023;28(4):806–24.
17. Gao X, Zhao S, Du Y, Yang Z, Tian Y, Zhao J, et al. Data-driven subtypes of polycystic ovary syndrome and their association with clinical outcomes. Nat Med. 2025;31(12):4214–24.
18. Li S, Zhang Y, Tang X, Deng M, Luo H, Chen Y. Mapping the network structure of anxiety, depression, and sleep symptoms in patients with polycystic ovary syndrome. Front Psychiatry. 2026;17:1738355.
19. Park J, Choi J, Choi JY. Network Analysis in Systems Epidemiology. J Prev Med Public Health. 2021;54(4):259–64.
20. Norouzi M, Amiri Z, Farkhani EM, Hoseini SJ, Khosravi Asl T. Advancing Healthcare Infrastructure: The Features of Iran's Sina Electronic Health Record System. Epidemiol Health Syst J.

2024;11(2):48–54.

21. Hajebi A, Motevalian A, Amin-Esmaeili M, Rahimi-Movaghar A, Sharifi V, Hoseini L, et al. Adaptation and validation of short scales for assessment of psychological distress in Iran: The Persian K10 and K6. Int J Methods Psychiatr Res. 2018;27(3):e1726.
22. Teede HJ, Tay CT, Laven JJE, Dokras A, Moran LJ, Piltonen TT, et al. Recommendations From the 2023 International Evidence-based Guideline for the Assessment and Management of Polycystic Ovary Syndrome. J Clin Endocrinol Metab. 2023;108(10):2447–69.
23. Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: A tutorial paper. Behav Res Methods. 2018;50(1):195–212.
24. Jones PJ, Heeren A, McNally RJ. Commentary: A network theory of mental disorders. Front Psychol. 2017;8:1305.
25. Isvoranu AM, Borsboom D, van Os J, Guloksuz S. A Network Approach to Environmental Impact in Psychotic Disorder: Brief Theoretical Framework. Schizophr Bull. 2016;42(4):870–3.
26. Lunansky G, van Borkulo CD, Haslbeck JMB, van der Linden MA, Garay CJ, Etchevers MJ, et al. The Mental Health Ecosystem: Extending Symptom Networks With Risk and Protective Factors. Front Psychiatry. 2021;12:640658.
27. Szczuko M, Kikut J, Szczuko U, Szydłowska I, Nawrocka-Rutkowska J, Ziętek M et al. Nutrition Strategy and Life Style in Polycystic Ovary Syndrome-Narrative Review. Nutrients. 2021;13(7).
28. Hoang T, Lee J, Kim J. Network Analysis of Demographics, Dietary Intake, and Comorbidity Interactions. Nutrients. 2021;13(10):3563.
29. Chang Y, Huang SS, Hsu WY, Liu YC. Mapping psychiatric comorbidity network: A pilot multi-method weighted network analysis with a focus on key disorders. Compr Psychiatry. 2025;141:152608.
30. Forslund M, Teede H, Melin J, Tay CT, Loxton D, Joham AE. Fertility and age at childbirth in polycystic ovary syndrome: results from a longitudinal population-based cohort study. Am J Obstet Gynecol. 2025;232(6):545.e1-.e10.
31. Tabassum F, Jyoti C, Sinha HH, Dhar K, Akhtar MS. Impact of polycystic ovary syndrome on quality of life of women in correlation to age, basal metabolic index, education and marriage. PLoS ONE. 2021;16(3):e0247486.
32. Talukder A, Tithi TA, Rumi MH, Muyeed A. Polycystic ovary syndrome and psychological distress: Urban-rural comparison among reproductive-aged women in Bangladesh. PLOS Glob Public Health. 2025;5(12):e0004937.
33. Pililis S, Lampsas S, Kountouri A, Pliouta L, Korakas E, Livadas S et al. The Cardiometabolic Risk in Women with Polycystic Ovarian Syndrome (PCOS): From Pathophysiology to Diagnosis and Treatment. Med (Kaunas). 2024;60(10).
34. Louwers YV, Laven JSE. Characteristics of polycystic ovary syndrome throughout life. Ther Adv Reprod Health. 2020;14:2633494120911038.

35. Parmar SK, Whooten R. Efficacy of physical activity in polycystic ovary syndrome treatment. Curr Opin Endocrinol Diabetes Obes. 2024;31(6):216–21.
36. Amiri M, Hatoum S, Hopkins D, Buyalos RP, Ezeh U, Pace LA, et al. The Association Between Obesity and Polycystic Ovary Syndrome: An Epidemiologic Study of Observational Data. J Clin Endocrinol Metab. 2024;109(10):2640–57.
37. Lim SS, Norman RJ, Davies MJ, Moran LJ. The effect of obesity on polycystic ovary syndrome: a systematic review and meta-analysis. Obes Rev. 2013;14(2):95–109.
38. Pau CT, Keefe CC, Welt CK. Cigarette smoking, nicotine levels and increased risk for metabolic syndrome in women with polycystic ovary syndrome. Gynecol Endocrinol. 2013;29(6):551–5.
39. Haslbeck J, Waldorp L. mgm: Estimating Time-Varying Mixed Graphical Models in High-Dimensional Data. J Stat Softw. 2020;93.

Figures

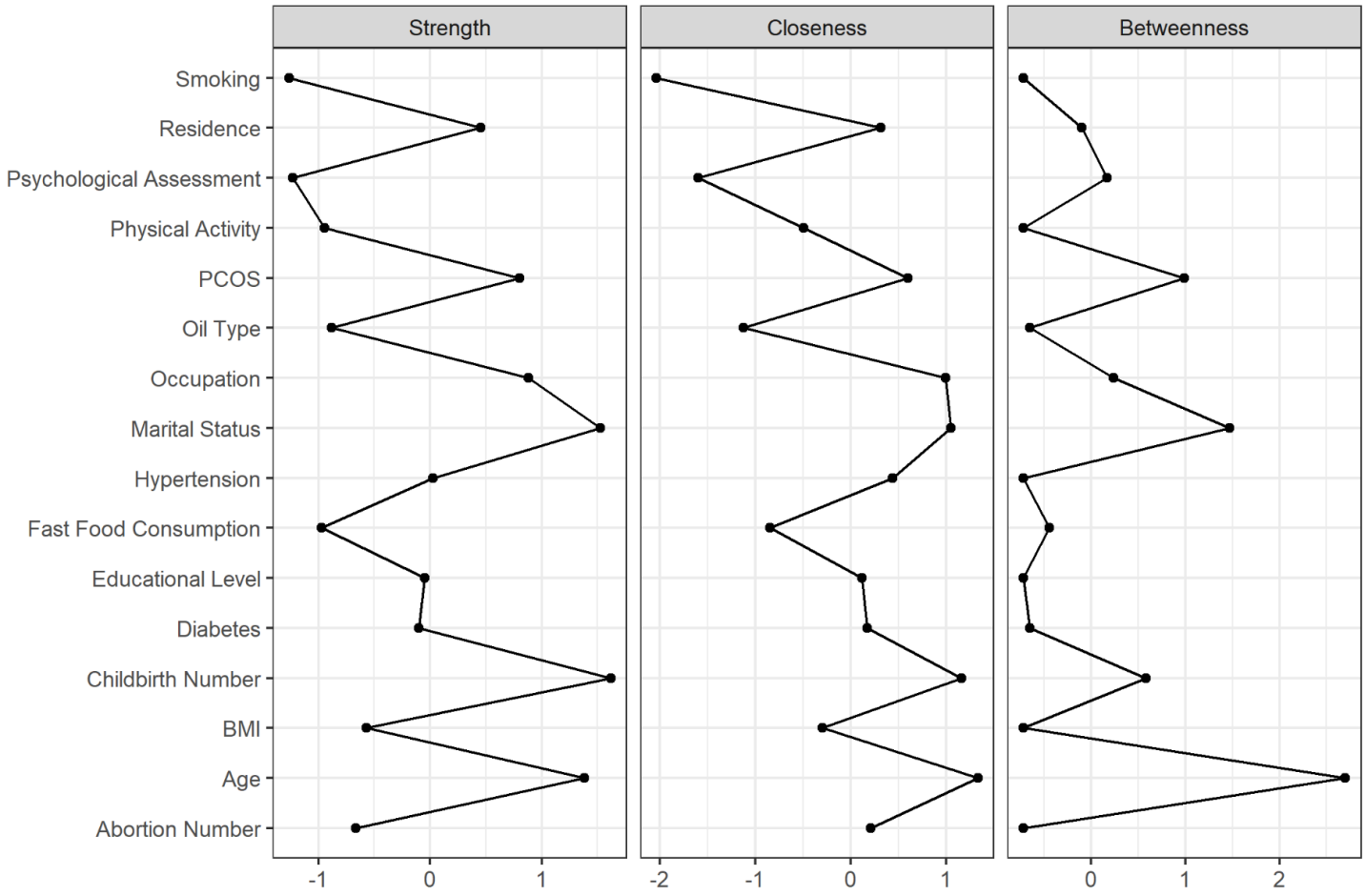


Figure 1

Strength centrality of demographic, lifestyle, reproductive, and clinical factors associated with polycystic ovary syndrome.

Note. Higher strength values indicate greater overall connectivity of each node within the estimated network.

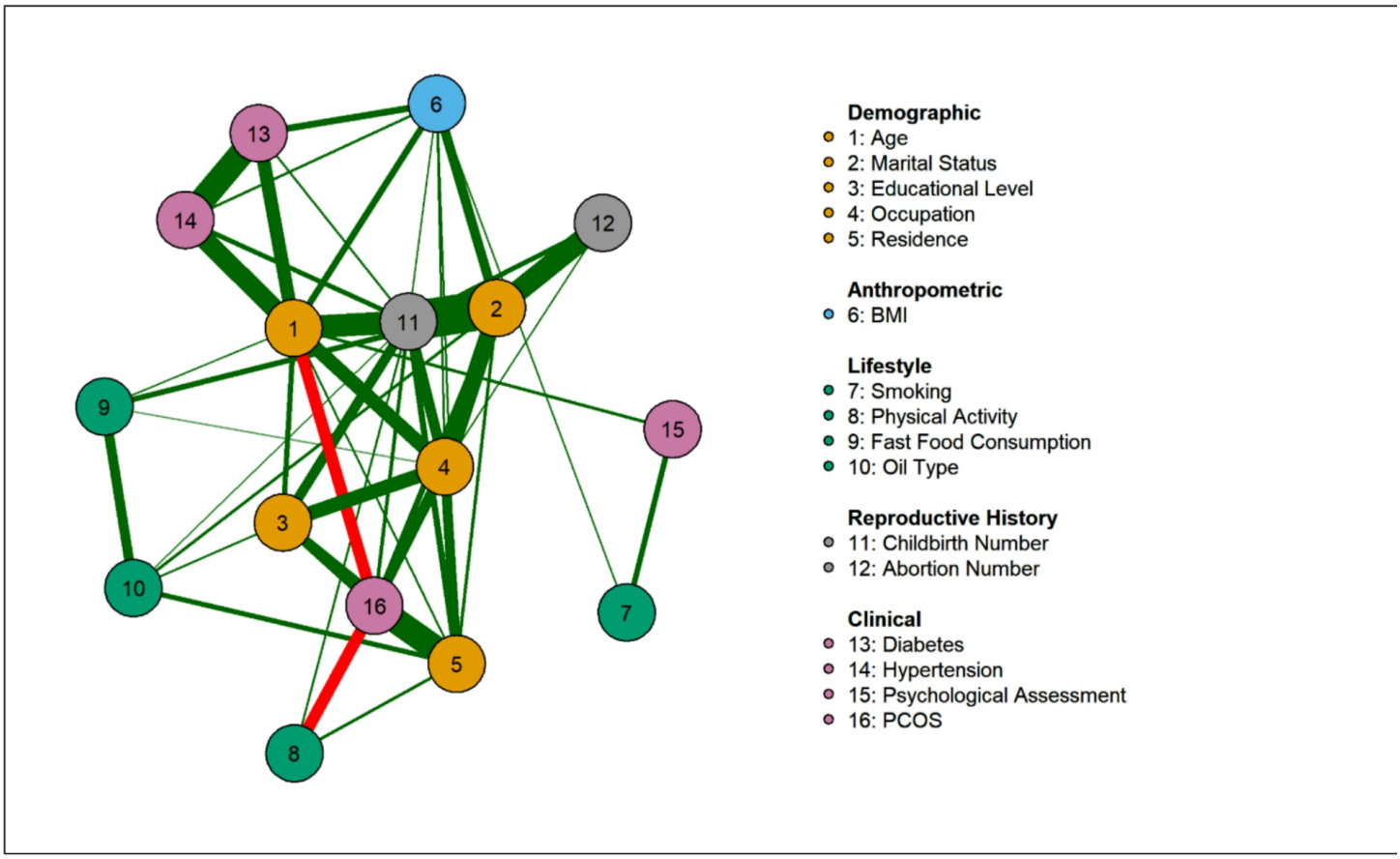


Figure 2

Network architecture of demographic, lifestyle, reproductive, and clinical factors associated with polycystic ovary syndrome.

Note. Nodes represent study variables, and edges represent regularized conditional associations estimated using a mixed graphical model. Edge thickness is proportional to association strength. Green and red edges indicate positive and negative associations, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary.docx