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.

Association of maternal gestational exposures and child/family characteristics with non-PFO congenital heart disease in children: a multicenter retrospective case-control study

Haifan Wang
Hainan Women and Children's Medical Center
Wanxing Li
Hainan Medical University School of Pediatrics
Ailixiati Alifu
Hainan Women and Children's Medical Center
Shaojing Wu
Hainan Women and Children's Medical Center
Siyu Lin
Hainan Women and Children's Medical Center
Jiafei He
Hainan Women and Children's Medical Center
Shiting Lin
Hainan Women and Children's Medical Center
Renwei Chen
chenrenwei@hainmc.edu.cn
Hainan Women and Children's Medical Center

Article

Keywords: Congenital heart disease, Patent foramen ovale, Prenatal exposure, Machine learning, Nomogram

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10965108/v1

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

Additional Declarations: No competing interests reported.

Association of maternal gestational exposures and child/family characteristics with non-PFO congenital heart disease in children: a multicenter retrospective case-control study

Haifan Wang[†a],Wanxing Li[†b],Ailixiati Alifu[a],Shaojing Wu[c],Siyu Lin[a],Jiafei He[a],Shiting Lin[a],Renwei Chen[a,b]*

[a]Department of Cardiothoracic Surgery, Hainan Women and Children's Medical Center, Haikou,Hainan, China

[b]Hainan Medical University School of Pediatrics,Haikou,Hainan, China

[c]Department of Clinical Nutrition, Hainan Women and Children's Medical Center, Haikou,Hainan, China

†Haifan Wang and Wanxing Li contributed equally to this work

***Correspondence:**

Full name:Renwei Chen

Current professional contact address:No. 1, Changbei Road, Xiuying District, Haikou, Hainan, China

Tel: 13876804231

Email: chenrenwei@hainmc.edu.cn

ORCID

0000-0002-1322-5378

First Author:

Full name:Haifan Wang

Current professional contact address:No. 1, Changbei Road, Xiuying District, Haikou,

Hainan, China

Tel:15501853260

Email:hyffyu722@muhn.edu.cn

Full name:Wanxing Li

Current professional contact address:No. 1, Changbei Road, Xiuying District, Haikou, Hainan, China

Tel:18886979850

Email:18886979850@163.com

Second Author:

Full name:Ailixiati Alifu

Current professional contact address:No. 1, Changbei Road, Xiuying District, Haikou, Hainan, China

Tel:18886979850

Email:elxat7@163.com

The third Author:

Full name:Shaojing Wu

Current professional contact address:No. 1, Changbei Road, Xiuying District, Haikou, Hainan, China

Tel:18876163022

Email:wuxiaojing.cool@163.com

Fourth Author:

Full name:Siyu Lin

Current professional contact address:No. 1, Changbei Road, Xiuying District, Haikou, Hainan, China
Tel:18876810997
Email:598600035@qq.com
Fifth Author:
Full name:Jiafei He
Current professional contact address:No. 1, Changbei Road, Xiuying District, Haikou, Hainan, China
Tel:18876900057
Email:jiafei00617@163.com
The sixth Author:
Full name:Shiting Lin
Current professional contact address:No. 1, Changbei Road, Xiuying District, Haikou, Hainan, China
Tel:18689779565
Email:176004729@qq.com

Abstract

Background: Congenital heart disease (CHD) is a major cause of infant morbidity and mortality. Previous studies often combined patent foramen ovale (PFO) with structural cardiac malformations, potentially increasing phenotypic heterogeneity.

Objectives: To identify maternal gestational exposures and child/family characteristics associated with non-PFO CHD and develop a preliminary prediction model.

Methods: This multicenter retrospective case-control study included 208 children from Hainan Province: 130 with non-PFO CHD and 78 with PFO. Fifty-five candidate variables were assessed. Least absolute shrinkage and selection operator (LASSO) regression was used for feature selection, followed by five machine learning models. SHapley Additive exPlanations (SHAP) analysis was used to interpret feature importance, and a nomogram was constructed.

Results: LASSO selected eight features, of which seven had nonzero global SHAP importance: child age in months, maternal protein intake frequency during pregnancy, history of CHD in other offspring, folic acid supplementation, frequency of family conflicts during pregnancy, maternal preconception smoking, and first-trimester hazardous substance exposure. Linear discriminant analysis (LDA) showed the best test-set performance, with an area under the receiver operating characteristic curve (AUC) of 0.702 and a Brier score of 0.232. The nomogram achieved an AUC of 0.732, and decision curve analysis indicated potential net benefit across the evaluated threshold range.

Conclusions: Candidate predictors associated with non-PFO CHD were identified, and an internally evaluated nomogram was developed for preliminary risk stratification. External validation is required before clinical use.

Keywords: Congenital heart disease; Patent foramen ovale; Prenatal exposure; Machine learning; Nomogram

1. Introduction

Congenital heart disease (CHD) is the most prevalent birth defect and a leading cause of infant morbidity and mortality worldwide [1], with a global birth prevalence of approximately 9.4 per 1,000 live births [2]. In China, CHD has ranked first among all birth defects since 2005, reaching a perinatal prevalence of 17.32 per 1,000 in 2020 and accounting for an estimated 150,000–200,000 new cases annually [1]. Despite substantial advances in prenatal screening, surgical techniques, and interventional procedures, delayed diagnosis remains a persistent challenge, leading to missed opportunities for timely treatment and increased risks of complications and death. Consequently, elucidating the etiology of CHD and enabling early prevention are of paramount clinical and public health importance.

Embryonic cardiac development occurs predominantly during the first trimester (weeks 3–8), a critical window in which cardiac progenitor cells undergo rapid proliferation, differentiation, migration, and remodeling, rendering the fetal heart highly vulnerable to external disturbances. Maternal health conditions, lifestyle behaviors, medication use, and environmental exposures during pregnancy can all affect fetal cardiogenesis through the placenta [3]. Previous studies have reported associations between several maternal factors—including pregestational diabetes, obesity, smoking, alcohol consumption, air pollution, endocrine-disrupting chemicals, and nutritional status—and the risk of CHD [4-6]. However, most previous etiological investigations have pooled patent foramen ovale (PFO) together with structural cardiac malformations. PFO is a remnant interatrial communication that persists in a substantial proportion of the population and is clinically silent in most individuals [7-9]. Pooling PFO with structural CHD may increase etiologic heterogeneity and attenuate exposure–outcome

associations [10, 11]. Therefore, excluding PFO and focusing specifically on non-PFO structural CHD is essential for accurately identifying modifiable teratogenic factors and developing precise first-trimester prevention strategies. Therefore, analyzing non-PFO structural CHD separately may improve etiologic specificity when evaluating maternal exposures.

Hainan Province is characterized by a unique geographic environment, diverse ethnic composition, and distinctive dietary and lifestyle customs, which may be associated with region-specific maternal nutritional, environmental, and behavioral patterns [12]. Nevertheless, systematic data on gestational risk factors for non-PFO CHD in Hainan children remain scarce, and the lack of region-specific screening tools hampers effective local prenatal care and birth defect prevention efforts. To address this gap, we conducted a multicenter retrospective case-control study across multiple hospitals in Hainan Province, enrolling children diagnosed between 2024 and 2025. Participants were classified as either PFO (control group) or non-PFO structural CHD (case group). Comprehensive information was collected on maternal nutrition, medication use, tobacco and alcohol exposure, environmental contacts, psychological status during pregnancy, as well as child and family medical history. We employed least absolute shrinkage and selection operator (LASSO) regression to reduce dimensionality from 55 candidate variables, followed by five machine learning algorithms (logistic regression, random forest, boosted tree, linear discriminant analysis, and extreme gradient boosting) to identify core predictors. A nomogram was constructed based on the selected risk factors, and its discriminative ability and clinical utility were assessed using receiver operating characteristic curves, calibration plots, decision curve analysis, and clinical impact curves.

This study aimed to identify key local risk factors for non-PFO CHD in Hainan, fill the regional evidence gap, and develop a preliminary risk-stratification tool to support prenatal health education, high-risk population screening, and targeted preventive interventions. These findings may inform region-specific prenatal counseling and future externally validated risk-prediction research.

2. Materials and methods

2.1 Study population

This study was conducted in maternal and child health care institutions across 16 cities and counties in Hainan Province, including Wenchang, Qionghai, Wanning, Danzhou, Lingshui, Chengmai, Ding'an, Lingao, Qiongzhong, Wuzhishan, Changjiang, Tunchang, Baisha, Dongfang, and Ledong. Pediatric patients diagnosed with CHD at various municipal and county hospitals in Hainan Province between 2024 and 2025 were enrolled as participants. This study was approved by the Medical Ethics Committee of Hainan Women and Children's Medical Center on March 4, 2024. The ethics approval number was HNWCMC Ethics Review 2024 No. [43]. The study complies with the Declaration of Helsinki. Written informed consent was obtained from the guardians of all participants.

Patients with a definitive diagnosis of CHD confirmed by echocardiography were included. Exclusion criteria were as follows: age older than 14 years, unclassifiable CHD subtypes, coexistence of severe infections or chronic genetic and metabolic disorders, missing core clinical variables, and invalid questionnaire data. A total of 208 children were included, consisting of 78 children with PFO assigned to the control group and 130 children with non-PFO CHD included in the case group. Non-PFO CHD encompassed patent ductus

arteriosus, atrial septal defect, ventricular septal defect, tetralogy of Fallot, and other structural cardiac abnormalities requiring clinical intervention. All diagnoses were verified by pediatric cardiologists in combination with echocardiographic findings.

2.2 Data collection and variable definition

Data were collected via standardized questionnaires, physical examinations, and electronic medical records. Non-PFO CHD status was defined as the binary outcome. A total of 55 candidate exposure variables were included, covering multiple dimensions: basic pediatric characteristics, parental sociodemographic features, parental health status, family medical history, maternal medication use during pregnancy, assisted reproductive technology, gestational nutritional supplementation, parental preconception alcohol and tobacco exposure, environmental exposure during pregnancy, maternal lifestyle during gestation, psychosocial factors in pregnancy, and maternal obstetric history. All categorical variables were coded following unified standardized protocols, while continuous variables were recorded as raw numerical values. Detailed information is presented in **Table 1**.

2.3 Statistical analysis

All data analyses and visualizations were performed using R software (v4.3.3). The tableone package (v0.13.2)[13] was adopted for descriptive statistics of baseline characteristics between the two groups. Categorical variables were presented as counts and percentages, while continuous variables were described as mean ± standard deviation or median (interquartile range), as appropriate. The least absolute shrinkage and selection operator (LASSO) regression with five-fold cross-validation was performed via the glmnet package (v4.1.4)[14] to reduce

dimensionality of high-dimensional variables. The model was set with a binary response variable and an alpha value of 1, and the optimal lambda value was determined according to the minimum cross-validated error to screen core predictive variables.

All enrolled samples were randomly divided into a training set and a test set at a ratio of 7:3. Based on the selected features, five machine learning models were established: logistic regression using the stats package (v4.3.3)[15], random forest and boosted tree using the randomForest package (v4.7.1.2)[16], linear discriminant analysis (LDA) using the MASS package (v7.3.60.0.1) (https://www.stats.ox.ac.uk/pub/MASS4/), and extreme gradient boosting (XGBoost) using the xgboost package (v1.7.11.1) (https://CRAN.R-project.org/package=xgboost). The area under the receiver operating characteristic curve (AUC), calibration curves generated by the regplot package (v1.1)[17], overall accuracy, sensitivity, specificity, F1-score and Brier score were comprehensively calculated to evaluate the discriminative ability and calibration performance of each model, and the optimal model was finally selected. SHapley Additive exPlanations (SHAP) analysis was implemented with the fastshap package (v0.1.1)[18] and shapviz package (v0.10.2) [19] to quantify the contribution of each feature to model prediction and interpret the effect of individual variables. A threshold of SHAP feature importance > 0 was adopted as the primary criterion for judgment.

Subsequently, a visual nomogram was constructed using the rms package (v6.8.1)[20] based on the core variables. The pROC package (v1.19.0.1)[21] was used to plot ROC curves for assessing model discrimination. Decision curve analysis was performed using the ggDCA package (v1.1)[22], and clinical impact curves were drawn via the rmda package (v1.6)[23].

Collectively, these analyses were applied to evaluate the potential clinical utility of the nomogram model.

3. Results

3.1 Baseline characteristics

A total of 208 children were enrolled in the study, consisting of 78 controls with patent foramen ovale and 130 cases with non-PFO CHD. Baseline characteristics of the two groups are comprehensively summarized in **Table 2**. Several descriptive between-group differences were observed. Children in the case group exhibited an older mean age (28.02 ± 30.70 months vs. 14.36 ± 20.90 months). Males accounted for a larger proportion in the case group (50.8% vs. 42.3%), and the proportions of Li and Miao ethnicities were relatively higher among cases. A positive history of other congenital cardiac malformations was more frequent in cases (8.5%) than in controls (1.3%).

For pre-conceptional parental exposures, higher maternal pre-pregnancy smoking rates were seen in the case group (8.5% vs. 1.3%), while paternal pre-pregnancy smoking was more common in controls (67.9% vs. 58.5%). Regarding gestational supplementation, folic acid use was less frequent among cases (81.5% vs. 88.5%), whereas the uptake of multivitamins and other nutritional supplements was higher in cases. More mothers of case children consumed protein multiple times per day during gestation (30.8% vs. 19.2%). Additionally, descriptive differences in gestational interpersonal conflict were also observed: 46.2% of case-group mothers reported rare quarrels during pregnancy versus 26.9% of controls; conversely, the proportion of mothers reporting almost no quarrels was substantially higher in the control group (53.8% vs. 33.1%).

No clear between-group differences were observed for the remaining variables, including birth weight, mode of delivery, preterm birth, comorbidities, parental demographics, anthropometric measurements, obstetric and family history, consanguinity, gestational passive smoking, environmental exposure, water source, pet raising, and prenatal examination adherence.

3.2 Feature selection by LASSO regression

LASSO regression analysis was conducted on all 55 clinical variables in the present study. The optimal λ value of 0.0409 was identified according to the minimum model error, and eight core feature variables were ultimately selected, namely child age in months, history of CHD in other offspring, gestational folic acid supplementation, maternal pre-conception smoking status, exposure to hazardous substances in the first trimester, frequency of protein intake during pregnancy, pet rearing during gestation, and frequency of family conflicts during pregnancy (**Figure 1A–B**).

3.3 Machine learning model assessment

The entire dataset was randomly split into a training set and a test set at a 7:3 ratio. Five machine learning models were established using the eight screened feature variables, followed by comprehensive performance evaluation. The LDA model yielded the optimal predictive performance. The AUC values of the LDA model were 0.721 in the training set (**Figure 2B**) and 0.702 in the test set (**Figure 2A**). Sensitivity, specificity, accuracy, F1-score and other composite evaluation metrics of the LDA model achieved favorable values in the test set (**Figure 2C**). The calibration curve demonstrated strong consistency between predicted

probabilities and actual event rates, with a Brier score of 0.2324, indicating reliable calibration performance of the LDA model (**Figure 2D**). The remaining four algorithms exhibited varying degrees of performance deterioration on the test set and possessed weaker generalization ability compared with the LDA model.

3.4 SHAP-based feature importance analysis

SHAP analysis was performed on the optimal linear discriminant analysis model to quantify the contribution of each variable to the prediction of non-PFO CHD. Among the eight LASSO-selected features, pet rearing during gestation was excluded after SHAP filtering (importance = 0), leaving seven core predictors with nonzero global importance. These factors ranked by contribution intensity from high to low were infant age in months, frequency of maternal protein intake during pregnancy, parental history of CHD in previous offspring, folic acid supplementation during pregnancy, frequency of quarrels during pregnancy, maternal smoking before pregnancy, and exposure to special substances in the first trimester (**Figure 3**).

3.5 Nomogram construction and validation

A nomogram for predicting the risk of non-PFO CHD was established based on the screened core predictive variables. The total score of 260 calculated from individual variable scores corresponded to a predicted disease probability of 0.865 for children (**Figure 4A**). The AUC of the nomogram was 0.732, indicating satisfactory discriminative power (**Figure 4B**). Decision curve analysis revealed that the net benefit of the model remained above zero across the risk threshold range of 0 to 0.8, which demonstrated promising value for clinical decision-making (**Figure 4C**). The clinical impact curve further verified that the actual disease prevalence

among high-risk individuals screened by the model increased gradually along with the rising risk threshold, suggesting favorable performance for population screening and reliable application in early identification of high-risk children (**Figure 4D**). These findings indicated that the established nomogram is a practical and reliable tool for clinical risk assessment of non-PFO CHD.

4. Discussion

CHD remains one of the leading causes of infant morbidity and mortality worldwide. However, most previous etiological studies have pooled PFO together with structural cardiac malformations, which may dilute true risk associations. Based on multicenter case-control data from Hainan Province, this study screened eight core features from 55 candidate variables using LASSO regression. After excluding variables with zero SHAP importance, seven predictors with nonzero global SHAP importance to non-PFO CHD were identified. The linear discriminant analysis (LDA) model achieved a test-set AUC of 0.702, and the constructed nomogram attained an AUC of 0.732, with decision curve analysis confirming a positive net benefit across the 0–0.8 risk threshold range, supporting preliminary risk stratification while indicating the need for external validation.

CHD is the leading birth defect in China, and structural cardiac malformations significantly increase the risk of infant mortality, long-term surgical intervention, and chronic cardiac dysfunction. However, most existing studies have analyzed PFO and non-PFO structural CHD cases together[24-26]. PFO is common in the general population and is often clinically silent, particularly in childhood[27]. Including PFO in CHD case groups substantially dilutes the true

285 associations between maternal exposures and structural cardiac anomalies[10, 11]. PFO and
286 non-PFO CHD differ fundamentally in embryonic pathogenesis—PFO results from failure of
287 postnatal foramen ovale closure, whereas structural cardiac malformations involve complex
288 morphogenetic abnormalities of cardiac septa, valves, and outflow tracts. The maternal
289 exposure profiles and pathogenic drivers of these two conditions are inherently heterogeneous,
290 and pooled analysis systematically underestimates the true effect sizes of environmental,
291 nutritional, and psychological exposures on teratogenic cardiac malformations[27-29]. By
292 separating PFO controls from non-PFO CHD cases, this study reduced phenotypic
293 heterogeneity in the case definition; however, this comparison did not eliminate confounding,
294 selection bias, or differences from unaffected children.
295 Ranked by SHAP contribution, the seven core predictive factors encompass disease follow-up
296 characteristics, genetic susceptibility, gestational nutrition, maternal psychological status, and
297 preconceptional and first-trimester teratogenic exposures. These factors span several domains;
298 however, interaction effects were not formally tested, and the observed associations may
299 reflect region-specific patterns in Hainan.
300 Child age in months was the predictor with the highest global SHAP importance. Child age
301 likely captured diagnostic timing and cumulative case ascertainment rather than prenatal
302 causation; therefore, it should be interpreted as a detection-related predictor. Child age is more
303 plausibly interpreted as a marker of cumulative diagnosis and referral timing than as a prenatal
304 etiologic exposure. Less symptomatic lesions, including some atrial or ventricular septal
305 defects, may remain unrecognized until later childhood or routine screening. Accordingly, the
306 observed age association should not be incorporated into a causal chain of maternal gestational

risk factors. Instead, it highlights the need to distinguish detection-related variables from potentially modifiable prenatal exposures when interpreting SHAP rankings and developing clinically meaningful prediction tools.

A history of CHD in other offspring may reflect inherited susceptibility, shared familial exposures, and intensified fetal cardiac surveillance. In this study, the case group had a positive rate of 8.5%, far higher than the 1.3% in controls. Genetically, pathogenic variants in key cardiac developmental transcription factors such as NKX2-5 and GATA4 can be vertically transmitted, with families carrying risk loci exhibiting significantly elevated recurrence risks. Environmentally, siblings within the same family share dietary patterns, living environments, and preconceptional tobacco and alcohol exposure habits, further amplifying risk on a background of genetic susceptibility. This history may inform preconception counseling and targeted fetal cardiac assessment, although its predictive value requires confirmation in larger cohorts.

Maternal high protein intake and folic acid supplementation during pregnancy were interpreted cautiously because the questionnaire did not quantify intake amount, source, timing, or total energy. In this study, mothers who consumed protein multiple times daily accounted for a significantly higher proportion in the case group (30.8% vs 19.2%). The present questionnaire did not quantify protein source, total intake, or co-exposure to contaminants; therefore, mechanisms involving heavy metals, organic pollutants, or MAPK signaling remain speculative[6, 30, 31]. Meanwhile, high protein intake may also indirectly reflect confounding effects of family economic status and prenatal health behaviors, which cannot be explained by a single "overnutrition" mechanism. The lower folic acid supplementation rate in the case

group is directionally consistent with a potentially protective association, but the direction cannot be inferred from global SHAP importance alone. Residual confounding, exposure timing, dosage, and recall bias should be considered before assigning a causal explanation. Future analyses should additionally examine supplement formulation, dietary source, dose-response patterns, and potential nonlinear effects. Prospective biomarker testing measuring maternal serum folate and homocysteine levels is needed to clarify the associations of nutritional factors.

Additionally, maternal psychological stress, preconceptional smoking, and first-trimester hazardous substance exposure are potentially modifiable maternal exposures. These exposures were associated with non-PFO CHD classification, but the retrospective design does not establish pathway-specific causality. Frequent interpersonal conflict may indicate maternal psychological stress, which has been associated with altered neuroendocrine and inflammatory signaling during pregnancy[32-34]. Maternal smoking has been associated with increased CHD risk, but the pathway-specific mechanism proposed here requires direct evidence[4, 35, 36]. First-trimester exposure to organic solvents, renovation pollutants, and pesticides has also been associated with congenital anomalies, although exposure-specific and phenotype-specific effects require further study[31]. These exposure categories are potentially amenable to prenatal health education, but their effects should be interpreted as associations rather than established causal pathways.

This study constructed five mainstream machine learning models—logistic regression, random forest, boosted tree, LDA, and XGBoost—with internal validation through 7:3 training-test set splitting. Discriminative and calibration performance were comprehensively evaluated using

AUC, calibration curves, Brier scores, sensitivity, and specificity. Among the evaluated models, LDA yielded the highest test-set AUC (0.702); however, the single random split, modest sample size, and absence of confidence intervals limit conclusions regarding comparative performance and generalizability. Tree-based models such as random forest and XGBoost performed well on the training set but showed marked performance degradation on the test set, suggesting possible overfitting in this small sample. The logistic model, as specified without interaction or nonlinear terms, may not capture more complex relationships among nutritional, psychological, and environmental exposures. Within this dataset, the LDA model balanced discrimination and computational simplicity; its suitability for primary-care deployment remains to be established.

Based on the retained predictors, the nomogram yielded an AUC of 0.732 in the available dataset. Decision and clinical impact curves suggested potential net benefit within the evaluated threshold range, but these analyses do not establish effectiveness for population screening. Moreover, because participants were children already classified as PFO or non-PFO CHD, the predicted probability represents discrimination within this selected clinical sample rather than absolute disease risk in the general maternal-child population. External validation and recalibration are therefore required before implementation in primary-care settings.

The main design feature of this study was the separate analysis of PFO controls and non-PFO structural CHD cases, which reduced phenotypic heterogeneity in the case definition but did not eliminate confounding or selection bias. Together with the seven retained predictors and the internally evaluated nomogram, the study provides preliminary evidence for classification within this Hainan clinical sample.

However, several limitations should be carefully considered. First, the retrospective case-control design cannot establish causal relationships, and recall bias may have affected exposure reporting—particularly for subjective variables such as nutritional supplementation and psychological status, whose reporting accuracy may be influenced by maternal retrospective cognition after the child's diagnosis. Second, although this study separated PFO from non-PFO CHD, etiological spectra may differ across non-PFO CHD subtypes (ventricular septal defect, atrial septal defect, tetralogy of Fallot, etc.), and pooled analysis may mask subtype-specific risk factors. Third, the study did not include biomarker data such as maternal serum folate, vitamin B_{12}, homocysteine, or heavy metals; nutritional exposure assessment relied solely on food frequency questionnaires, limiting precision. Fourth, the sample size (208 participants) remains modest for machine learning modeling, and statistical power for some variables may be constrained. In addition, using PFO rather than unaffected children as controls, performing only an internal split, and not reporting confidence intervals for performance metrics limit transportability and clinical interpretation. Future validation should include both unaffected controls and phenotype-specific CHD subgroups to determine whether the selected predictors are specific to structural CHD or primarily reflect patterns of clinical referral and ascertainment. Such evaluation should use prospectively recruited cohorts.

Future research should be based on prospective cohort designs, integrating multi-omics biomarker testing (metabolomics, genomics, epigenomics) with high-resolution fetal echocardiographic parameters to construct multimodal integrated risk prediction models. Subtype-stratified analyses in larger samples should be conducted to reveal exposure-risk profiles specific to different structural cardiac malformations. Prospective studies with

quantitative dietary assessment and biomarker measurements are needed to clarify the direction and potential causality of nutritional factors and provide high-quality evidence for developing regionally tailored prenatal care guidelines.

5. Conclusion

Among children with PFO or non-PFO CHD in this multicenter Hainan sample, seven candidate predictors were retained after LASSO and SHAP analysis. The LDA model and nomogram showed modest discrimination and potential net benefit for preliminary classification. Because the study used a retrospective case-control design, PFO controls, and internal validation only, the findings should not be interpreted as population risk estimates or causal effects. External validation in larger prospective cohorts including unaffected controls is required before clinical use.

Abbreviations

Abbreviation	Definition
CHD	Congenital heart disease
PFO	Patent foramen ovale
LASSO	Least absolute shrinkage and selection operator
LDA	Linear discriminant analysis
XGBoost	Extreme gradient boosting
SHAP	SHapley Additive exPlanations
AUC	Area under the receiver operating characteristic curve
ROC	Receiver operating characteristic
DCA	Decision curve analysis

CI	Confidence interval
SD	Standard deviation

Acknowledgements

We would like to express our sincere gratitude to all individuals and organizations who supported and assisted us throughout this research. Special thanks to the following authors:Renwei Chen,Ailixiati Alifu. In conclusion, we extend our thanks to everyone who has supported and assisted us along the way. Without your support, this research would not have been possible.

Author contributions

Haifan Wang,Wanxing Li: Conceptualization, Data curation, Validation, Visualization, Writing–original draft, Writing–review & editing. Ailixiati Alifu: Data curation, Validation, Visualization, Writing–review & editing. Shaojing Wu: Validation, Writing–review & editing. Siyu Lin: Visualization, Writing–review & editing. Jiafei He,Shiting Lin: Conceptualization, Supervision, Writing–review & editing. Renwei Chen: Conceptualization, Project administration, Supervision, Writing–review & editing. This work currently described has not been published, is not being considered for publication elsewhere, and its publication was approved by all authors.

Funding

The research reported in this project was generously supported by the Hainan Provincial Health Science and Technology Innovation Joint Project Fund (No.: WSJK2024QN054), National Natural Science Foundation of China(No.: 82660071)，the Hainan Clinical Research

Center for Children's Health and Diseases(No.: LCYXYJZX-202603), and the Excellent Talent Team of Hainan Province (No.: QRCBT202121).The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declaration of competing interest

The authors declare that they have no competing interests.

Ethics approval and consent to participate

This study was approved by the Medical Ethics Committee of Hainan Women and Children's Medical Center on March 4, 2024. The ethics approval number was HNWCMC Ethics Review 2024 No. [43]. The study complies with the Declaration of Helsinki. Written informed consent was obtained from the guardians of all participants.

Consent for publication

Not applicable.

Data availability

The datasets generated and/or analyzed during the current study are not publicly available because they contain clinical, questionnaire, and maternal–child information that may be subject to privacy and ethical restrictions. De-identified data supporting the findings of this study are available from the corresponding author upon reasonable request and subject to approval by the relevant ethics committee and institutional data-sharing requirements.

Declaration of Generative AI and AI-assisted technologies in the manuscript preparation process

Not applicable.

Author agreement

All authors have read and approved the manuscript, have directly participated in the work as appropriate, and agree to its submission to the American Journal of Preventive Cardiology. The manuscript is not under consideration for publication elsewhere.

References

1. Zhang Y, Wang J, Zhao J, Huang G, Liu K, Pan W, Sun L, Li J, Xu W, He C et al: Current status and challenges in prenatal and neonatal screening, diagnosis, and management of congenital heart disease in China. The Lancet Child & adolescent health 2023, 7(7):479-489.https://doi.org/10.1016/s2352-4642(23)00051-2.
2. O'Malley BR, Sholler GF, Smith J, Blue GM: The Evolving Landscape of CHD Genetics: A Contemporary Guide to Genetic Testing and Management. Journal of paediatrics and child health 2025.https://doi.org/10.1111/jpc.70176.

3. Elhassan YH, Alahmadi F, Albadawi EA, Albarakati A, Aljohany AH, Alzaman NS, Albadrani M: The Relationship Between Maternal Exposure to Endocrine-Disrupting Chemicals and the Incidence of Congenital Heart Diseases: A Systematic Review and Meta-Analysis. Metabolites 2024, 14(12).https://doi.org/10.3390/metabo14120709.

4. Suo Z, Rancz A, Jankó Á, Hegyi P, Klárik Z, Mayer Á, Vezér M, Ács N: Risk assessment of congenital heart defects based on maternal factors in early pregnancy-a systematic review and meta-analysis. Frontiers of medicine 2025.https://doi.org/10.1007/s11684-025-1170-2.

5. Salmeri N, Seidenari A, Cavoretto PI, Papale M, Candiani M, Farina A: Maternal prepregnancy weight as an independent risk factor for congenital heart defect: systematic review and meta-analysis stratified by subtype and severity of defect. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology 2024, 64(3):294-307.https://doi.org/10.1002/uog.27659.

6. Feng Y, Liu X, Zhang X, Zhao X, Chang H, Ouyang F, Yu Z, Gao Z, Zhang H: Global air pollution exposure and congenital anomalies: an updated systematic review and meta-analysis of epidemiological studies. International journal of environmental health research 2024, 34(5):2333-2352.https://doi.org/10.1080/09603123.2023.2246383.

7. Shah AH, Horlick EM, Kass M, Carroll JD, Krasuski RA: The pathophysiology of patent foramen ovale and its related complications. American heart journal 2024, 277:76-92.https://doi.org/10.1016/j.ahj.2024.08.001.

8. Saharan S, Vettukattil J, Bhat A, Amula V, Bansal M, Chowdhury D, Dyamenahalli U, Gupta SK, Das B, Susheel Kumar TK et al: Patent foramen ovale in children: Unique pediatric challenges and lessons learned from adult literature. Annals of pediatric cardiology 2022, 15(1):44-52.https://doi.org/10.4103/apc.apc_67_21.

9. Aoun J, Hatab T, Volpi J, Lin CH: Patent Foramen Ovale and Atrial Septal Defect. Cardiology clinics 2024, 42(3):417-431.https://doi.org/10.1016/j.ccl.2024.02.019.

10. Cho KK, Khanna S, Lo P, Cheng D, Roy D: Persistent pathology of the patent foramen ovale: a review of the literature. The Medical journal of Australia 2021, 215(2):89-93.https://doi.org/10.5694/mja2.51141.

11. Peng Z, Wei J, Huang X, Chen B, Liang L, Feng B, Song P, He J, Que T, Lan J et al: Epidemiology of Congenital Heart Defects in Perinatal Infants in Guangxi, China. International journal of general medicine 2024, 17:5381-5396.https://doi.org/10.2147/ijgm.S466195.

12. Chen X, Ding Y, Shi L, Wu D, Wang L, Chen F, Mo Y: Dietary patterns and gestational hypertension in nulliparous pregnant Chinese women: A CONSORT report. Medicine (Baltimore) 2020, 99(29):e20186.https://doi.org/10.1097/md.0000000000020186.

13. Panos A, Mavridis D: TableOne: an online web application and R package for summarising and visualising data. Evidence-based mental health 2020, 23(3):127-130.https://doi.org/10.1136/ebmental-2020-300162.

14. Engebretsen S, Bohlin J: Statistical predictions with glmnet. Clin Epigenetics 2019, 11(1):123.https://doi.org/10.1186/s13148-019-0730-1.

15. Goncharova IA, Nazarenko MS, Babushkina NP, Markov AV, Pecherina TB, Kashtalap VV, Tarasenko NV, Ponasenko AV, Barbarash OL, Puzyrev VP: [Genetic Predisposition to Early Myocardial Infarction]. Molekuliarnaia biologiia 2020, 54(2):224-232.https://doi.org/10.31857/s0026898420020044.

16. Wang Y, Zhuang H, Jiang XH, Zou RH, Wang HY, Fan ZN: Unveiling the key genes, environmental toxins, and drug exposures in modulating the severity of ulcerative colitis: a comprehensive analysis. Front Immunol 2023, 14:1162458.https://doi.org/10.3389/fimmu.2023.1162458.

17. Zhang Z, Cortese G, Combescure C, Marshall R, Lee M, Lim HJ, Haller B: Overview of model validation for survival regression model with competing risks using melanoma study data. Ann Transl Med 2018, 6(16):325.https://doi.org/10.21037/atm.2018.07.38.

18. Jethani N, Saporta A, Ranganath R: Don't be fooled: label leakage in explanation methods and the importance of their quantitative evaluation. Proceedings of machine learning research 2023, 206:8925-8953

19. Huang TF, Luo C, Guo LB, Liu HZ, Li JT, Lin QZ, Fan RL, Zhou WP, Li JD, Lin KC et al: Preoperative prediction of textbook outcome in intrahepatic cholangiocarcinoma by interpretable machine learning: A multicenter cohort study. World journal of gastroenterology 2025, 31(11):100911.https://doi.org/10.3748/wjg.v31.i11.100911.

20. Lin G, Gao Z, Wu S, Zheng J, Guo X, Zheng X, Chen R: scRNA-seq revealed high stemness epithelial malignant cell clusters and prognostic models of lung

adenocarcinoma. Sci Rep 2024, 14(1):3709.https://doi.org/10.1038/s41598-024-54135-4.

21. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M: pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011, 12:77

22. Chen Y, He J, Pan X: Prediction of risk factors for preoperative deep vein thrombosis in patients with pelvic fracture. Front Surg 2025, 12:1585460.https://doi.org/10.3389/fsurg.2025.1585460.

23. Liu C, He Y, Luo J: Application of Chest CT Imaging Feature Model in Distinguishing Squamous Cell Carcinoma and Adenocarcinoma of the Lung. Cancer Manag Res 2024, 16:547-557.https://doi.org/10.2147/cmar.S462951.

24. Ma K, He Q, Dou Z, Hou X, Li X, Zhao J, Rao C, Feng Z, Sun K, Chen X et al: Current treatment outcomes of congenital heart disease and future perspectives. The Lancet Child & adolescent health 2023, 7(7):490-501.https://doi.org/10.1016/s2352-4642(23)00076-7.

25. Marino BS, Cassedy A, Brown KL, Franklin R, Gaynor JW, Cvetkovic M, Laker S, Levinson K, MacGloin H, Mahony L et al: Long-Term Quality of Life in Congenital Heart Disease Surgical Survivors: Multicenter Retrospective Study of Surgical and ICU Explanatory Factors. Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies 2023, 24(5):391-398.https://doi.org/10.1097/pcc.0000000000003190.

546 26. Amdani S, Conway J, George K, Martinez HR, Asante-Korang A, Goldberg CS,
547 Davies RR, Miyamoto SD, Hsu DT: Evaluation and Management of Chronic Heart
548 Failure in Children and Adolescents With Congenital Heart Disease: A Scientific
549 Statement From the American Heart Association. Circulation 2024,
550 150(2):e33-e50.https://doi.org/10.1161/cir.0000000000001245.
551 27. Blixenkrone-Møller E, Dannesbo S, Dehn AM, Pihl CA, Sillesen AS, Vøgg ROB,
552 Raja AA, Colan S, Mertens L, Vejlstrup N et al: Interatrial Communications:
553 Prevalence and Subtypes in 12,385 Newborns-a Copenhagen Baby Heart Study.
554 Pediatric cardiology 2025,
555 46(6):1570-1580.https://doi.org/10.1007/s00246-024-03571-0.
556 28. Rizzo S, De Gaspari M, Basso C, Fraccaro C, Thiene G: Patent foramen ovale: A
557 variant of normal or a true congenital heart disease? Cardiovascular pathology : the
558 official journal of the Society for Cardiovascular Pathology 2025,
559 76:107722.https://doi.org/10.1016/j.carpath.2025.107722.
560 29. Paolucci M, Vincenzi C, Romoli M, Amico G, Ceccherini I, Lattanzi S, Bersano A,
561 Longoni M, Sacco S, Vernieri F et al: The Genetic Landscape of Patent Foramen
562 Ovale: A Systematic Review. Genes (Basel) 2021,
563 12(12).https://doi.org/10.3390/genes12121953.
564 30. Liu JT, Zhou YW, Wang WD, Mao BH, Hu YG: Association of the levels of heavy
565 metals and trace elements during pregnancy with congenital heart defects in offspring:
566 a prospective cohort study. Zhongguo dang dai er ke za zhi = Chinese journal of

contemporary pediatrics 2022, 24(2):147-154.https://doi.org/10.7499/j.issn.1008-8830.2109053.

31. Wu L, Li N, Liu Y: Association Between Maternal Factors and Risk of Congenital Heart Disease in Offspring: A Systematic Review and Meta-Analysis. Maternal and child health journal 2023, 27(1):29-48.https://doi.org/10.1007/s10995-022-03538-8.

32. Gu J, Guan HB: Maternal psychological stress during pregnancy and risk of congenital heart disease in offspring: A systematic review and meta-analysis. J Affect Disord 2021, 291:32-38.https://doi.org/10.1016/j.jad.2021.05.002.

33. Jeličić L, Veselinović A, Ćirović M, Jakovljević V, Raičević S, Subotić M: Maternal Distress during Pregnancy and the Postpartum Period: Underlying Mechanisms and Child's Developmental Outcomes-A Narrative Review. Int J Mol Sci 2022, 23(22).https://doi.org/10.3390/ijms232213932.

34. Rakers F, Rupprecht S, Dreiling M, Bergmeier C, Witte OW, Schwab M: Transfer of maternal psychosocial stress to the fetus. Neuroscience and biobehavioral reviews 2017.https://doi.org/10.1016/j.neubiorev.2017.02.019.

35. Han J, Zhang Y, Liu Y, Liu J, Zhang Y, Wang K: Parental Smoking and the Risk of Birth Defects in Offspring in China: A Systematic Review and meta-Analysis. Birth defects research 2024, 116(12):e2422.https://doi.org/10.1002/bdr2.2422.

36. Zhong L, Yang X, Zhang H, Fu M, Cai J, Huang W, Huang Y: Nicotine Induces Fetal Cardiac Dysfunction by Promoting Cardiomyocyte Apoptosis Through Regulating the KCTD10-Notch Signaling. International heart journal 2025, 66(4):651-659.https://doi.org/10.1536/ihj.24-427.

Figures

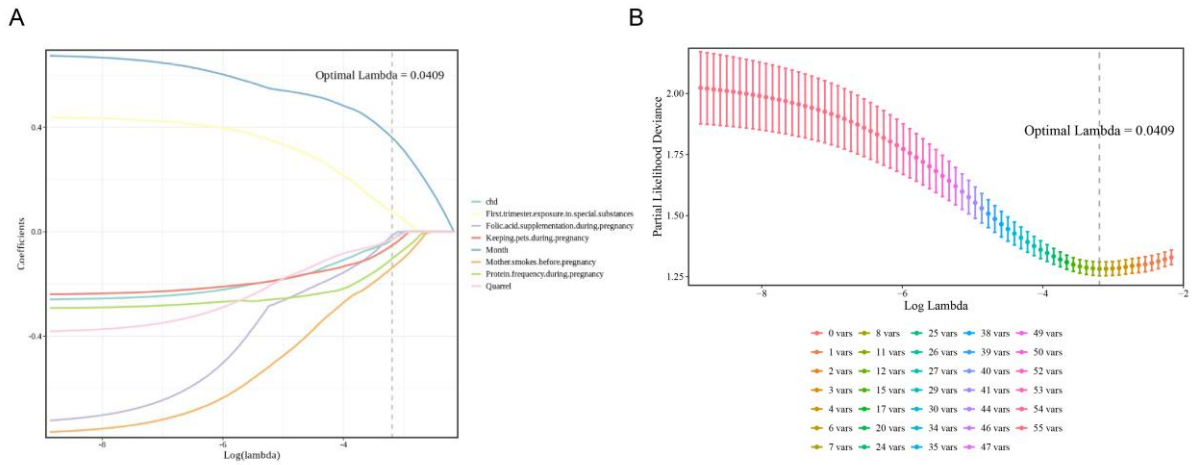


Figure 1. LASSO regression analysis for feature selection. **(A)** Coefficient profile plot showing the changes in regression coefficients of the candidate variables as the penalty parameter (λ) increased. **(B)** Five-fold cross-validation curve illustrating the selection of the optimal λ value based on the minimum cross-validation error. The optimal λ value was used to identify predictive features associated with non-patent foramen ovale congenital heart disease (non-PFO CHD).

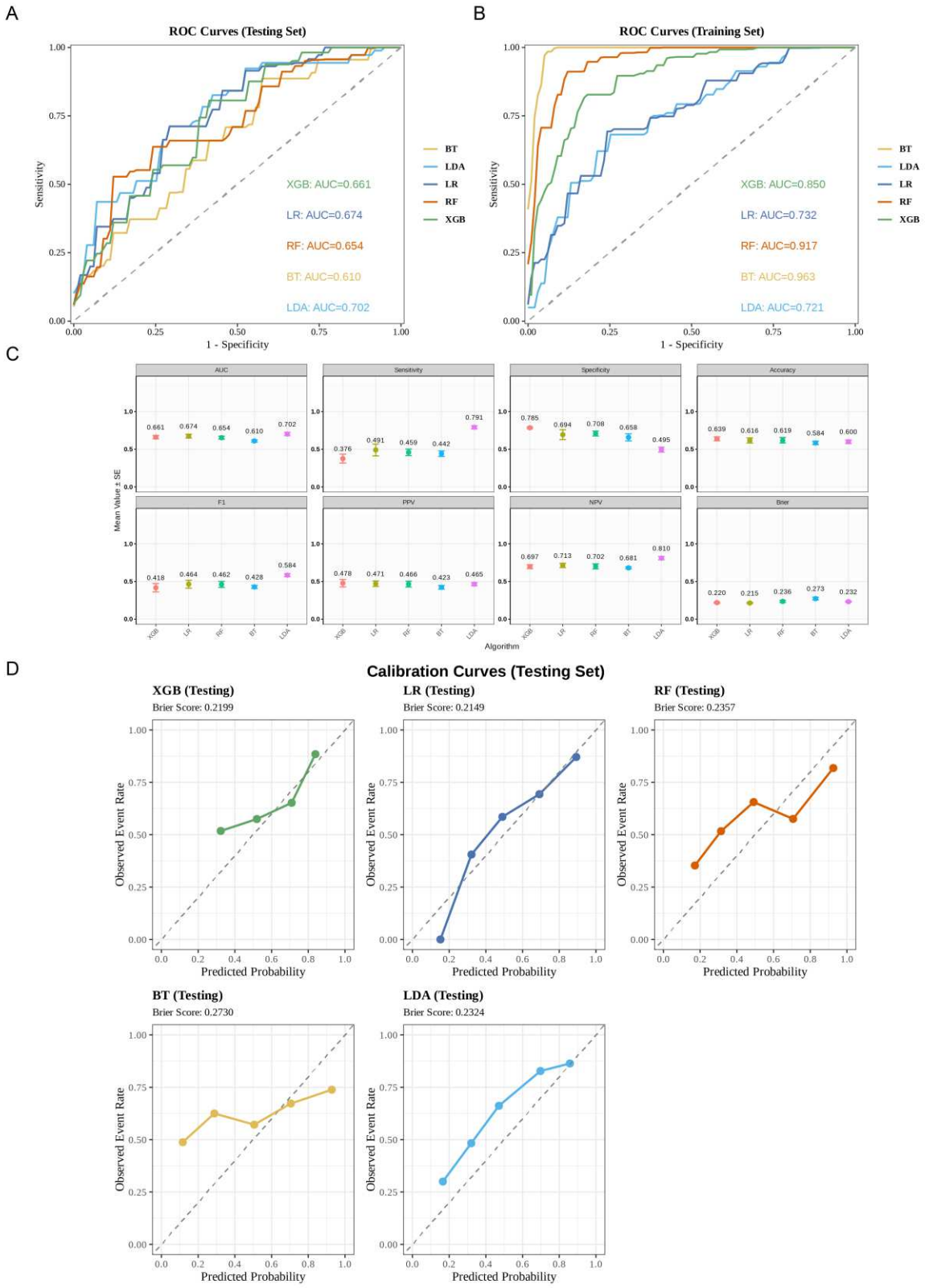


Figure 2. Predictive performance evaluation of machine learning models. **(A)** Receiver operating characteristic (ROC) curves for the five machine learning models in the test set. **(B)** Receiver operating characteristic (ROC) curves for the five machine learning models in the

training set. **(C)** Comparison of predictive performance metrics, including area under the ROC curve (AUC), accuracy, sensitivity, specificity, F1-score, positive predictive value, negative predictive value, and Brier score, among the five models in the test set. **(D)** Calibration curves and Brier scores for the five machine learning models in the test set.

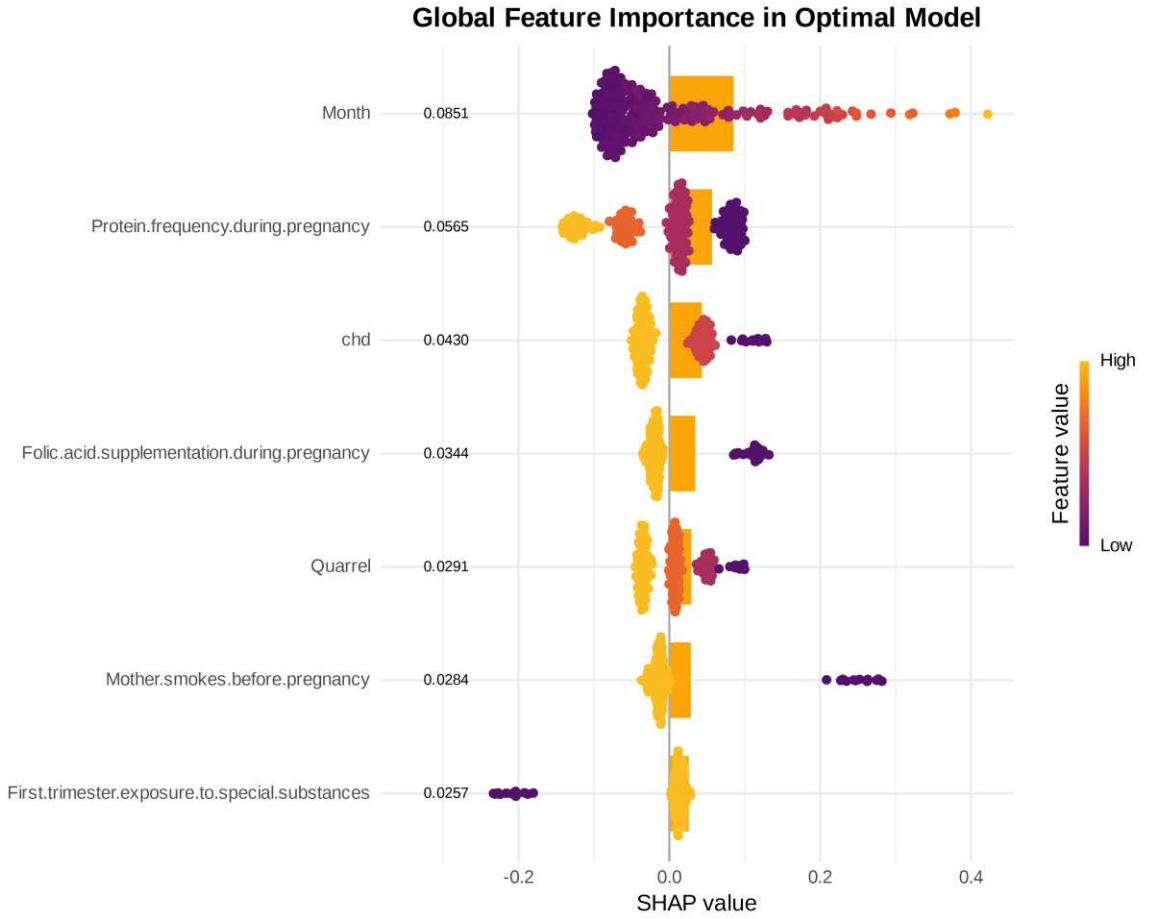
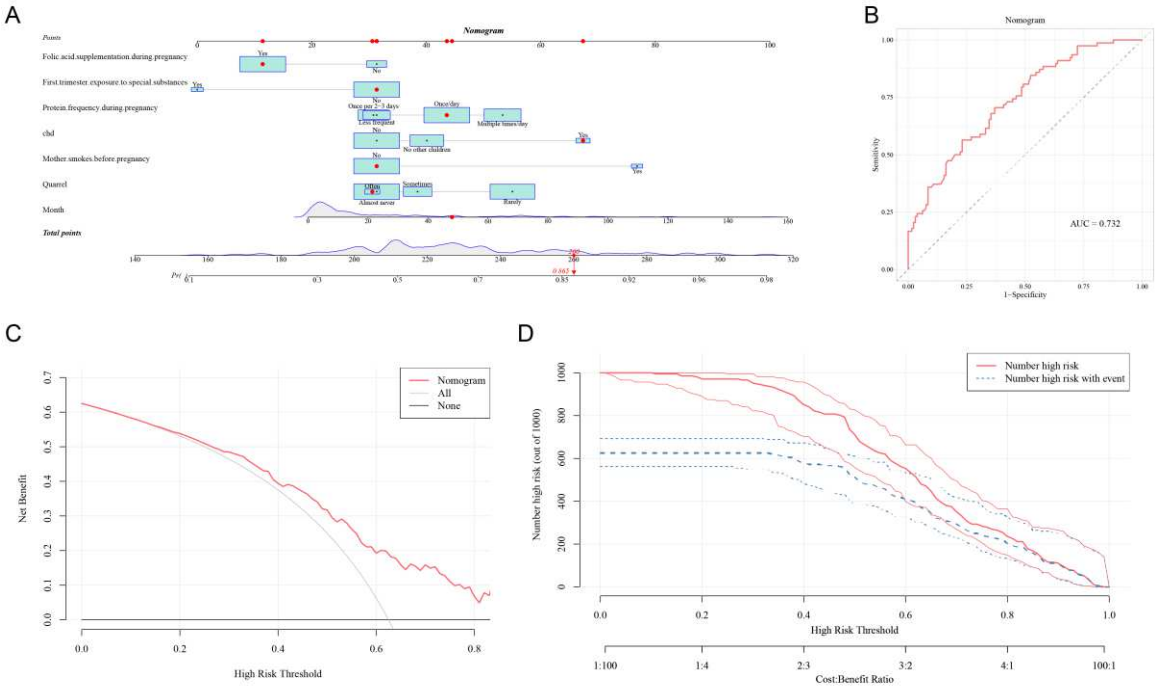


Figure 3. SHAP-based feature importance analysis of predictors for non-PFO CHD. SHapley Additive exPlanations (SHAP) analysis showing the global importance ranking of the seven retained predictors in the optimal linear discriminant analysis model. Variables with nonzero SHAP importance were considered core predictors, including child age in months, maternal protein intake frequency during pregnancy, history of congenital heart disease (CHD) in other offspring, folic acid supplementation during pregnancy, frequency of family conflicts during

612 pregnancy, maternal pre-conception smoking status, and first-trimester hazardous substance
613 exposure.



614

615 **Figure 4**. Construction and evaluation of the nomogram model for non-PFO CHD risk
616 prediction. **(A)** Nomogram developed based on the selected core predictors for estimating the
617 probability of non-patent foramen ovale congenital heart disease (non-PFO CHD). **(B)**
618 Receiver operating characteristic (ROC) curve demonstrating the discriminative performance
619 of the nomogram model. **(C)** Decision curve analysis (DCA) evaluating the potential clinical
620 net benefit of the nomogram across different risk thresholds. **(D)** Clinical impact curve
621 illustrating the estimated number of high-risk individuals and true positive cases across varying
622 threshold probabilities.

623

624

625 **Tables**

626 **Table 1**. Variables and coding scheme.

Variable	Variable name	Coding description	Category
Child sex	Gender	1=male, 2=female	Child characteristics
Child ethnicity	Race	1=Han, 2=Hui, 3=Li, 4=Miao, 5=other	Child characteristics
Gestational age (months)	Month	Continuous value (months)	Child characteristics
Birth weight	Weight	Continuous value (g)	Child characteristics
Mode of delivery	Delivery.method	1=vaginal delivery, 2=cesarean section	Child characteristics
Premature birth	Premature	1=yes, 2=no	Child characteristics

Variable	Variable name	Coding description	Category
Comorbidities (severe acute/chronic infections, inherited metabolic or chronic wasting diseases)	Comorbidities	1=present, 2=absent	Child characteristics
Other congenital heart disease history in children	chd	1=present, 2=present in other children, 3=absent	Child characteristics
Paternal ethnicity	Father.race	1=Han, 2=Hui, 3=Li, 4=Miao, 5=other	Parental sociodemographic characteristics

Variable	Variable name	Coding description	Category
Maternal ethnicity	Maternal.race	1=Han, 2=Hui, 3=Li, 4=Miao, 5=other	Parental sociodemographic characteristics
Paternal age	Father.age	Continuous value (years)	Parental sociodemographic characteristics
Maternal age	Maternal.age	Continuous value (years)	Parental sociodemographic characteristics

Variable	Variable name	Coding description	Category
Paternal height	Father.height	Continuous value (cm)	Parental sociodemographic characteristics
Paternal weight	Father.weight	Continuous value (kg)	Parental sociodemographic characteristics
Maternal height	Maternal.height	Continuous value (cm)	Parental sociodemographic characteristics

Variable	Variable name	Coding description	Category
Maternal weight	Maternal.weight	Continuous value (kg)	Parental sociodemographic characteristics
Maternal education level	Maternal.education	1=illiterate, 2=primary school, 3=junior high school, 4=senior high school, 5=technical school, 6=secondary specialized school, 7=junior college, 8=bachelor's degree, 9=postgraduate or above	Parental sociodemographic characteristics

Variable	Variable name	Coding description	Category
Paternal education level	Father.education	1=illiterate, 2=primary school, 3=junior high school, 4=senior high school, 5=technical school, 6=secondary specialized school, 7=junior college, 8=bachelor's degree, 9=postgraduate or above	Parental sociodemographic characteristics
Monthly household income per capita (yuan)	Income	1=<1000, 2=1000-3000, 3=3000-5000, 4=5000-10000, 5=10000-20000, 6=>20000	Parental sociodemographic characteristics
Maternal medical history	Maternal.medical.history	0=present, 1=absent	Parental health status
Paternal medical history	Father.medical.history	0=present, 1=absent	Parental health status

Variable	Variable name	Coding description	Category
Paternal family history of congenital heart disease	Father.family.congenital.heart.disease	1=present, 2=absent	Family history
Maternal family history of congenital heart disease	Maternal.family.congenital.heart.disease	1=present, 2=absent	Family history
Consanguineous marriage	Consanguineous.marriage	1=yes, 2=no	Family history
Maternal medication use during pregnancy	Maternal.medication.during.pregnancy	1=yes, 2=no	Medication during pregnancy
Assisted reproductive technology	Artificial.assisted.reproductive.technology	1=yes, 2=no	Assisted reproduction

Variable	Variable name	Coding description	Category
Folic acid supplementation during pregnancy	Folic.acid.supplementation.during.pregnancy	1=yes, 0=no	Nutritional supplementation during pregnancy
Multivitamin supplementation during pregnancy	Multivitamin.supplementation.during.pregnancy	1=yes, 0=no	Nutritional supplementation during pregnancy
Calcium supplementation during pregnancy	Calcium.supplements.during.pregnancy	1=yes, 0=no	Nutritional supplementation during pregnancy

Variable	Variable name	Coding description	Category
Iron supplementation during pregnancy	Iron.supplements.during.pregnancy	1=yes, 0=no	Nutritional supplementation during pregnancy
Other nutrient supplementation during pregnancy	Supplement.other.nutrients.during.pregnancy	1=yes, 0=no	Nutritional supplementation during pregnancy
Paternal smoking before pregnancy	Father.smoked.before.pregnancy	1=yes, 2=no	Parental smoking and drinking before pregnancy

Variable	Variable name	Coding description	Category
Paternal drinking before pregnancy	Father.drinking.before.pregnancy	1=yes, 2=no	Parental smoking and drinking before pregnancy
Maternal smoking before pregnancy	Mother.smokes.before.pregnancy	1=yes, 2=no	Parental smoking and drinking before pregnancy
Maternal drinking before pregnancy	Mother.drinking.before.pregnancy	1=yes, 2=no	Parental smoking and drinking before pregnancy

Variable	Variable name	Coding description	Category
Exposure to special substances in first trimester	First.trimester.exposure.to.special.substances	1=yes, 2=no	Environmental exposure during pregnancy
Indoor home decoration during pregnancy	Interior.decorating.during.pregnancy	1=yes, 2=no	Environmental exposure during pregnancy
Presence of pollution-related factory	Pollution.related.plant	1=yes, 2=no	Environmental exposure during pregnancy

Variable	Variable name	Coding description	Category
Presence of gas station	gas.station	1=yes, 2=no	Environmental exposure during pregnancy
Presence of garbage station	Garbage.station	1=yes, 2=no	Environmental exposure during pregnancy
Presence of viaduct or highway	Viaduct.or.highway	1=yes, 2=no	Environmental exposure during pregnancy

Variable	Variable name	Coding description	Category
Presence of main traffic road	Main.traffic.road	1=yes, 2=no	Environmental exposure during pregnancy
Protein intake frequency during pregnancy	Protein.frequency.during.pregnancy	1=multiple times/day, 2=once/day, 3=once/2-3 days, 4=less often	Lifestyle during pregnancy
Fruit and vegetable intake frequency during pregnancy	Frequency.of.fruits.and.vegetables.during.pregnancy	1=multiple times/day, 2=once/day, 3=once/2-3 days, 4=less often	Lifestyle during pregnancy
Passive smoking during pregnancy	Passive.smoking.during.pregnancy	1=yes, 2=no	Lifestyle during pregnancy

Variable	Variable name	Coding description	Category
Source of drinking water	Source.of.drinking.water	1=tap water, 2=well water, 3=spring water, 4=river water, 5=lake water, 6=other	Lifestyle during pregnancy
Keeping pets during pregnancy	Keeping.pets.during.pregnancy	1=yes, 2=no	Lifestyle during pregnancy
Raising poultry during pregnancy	Raising.poultry.during.pregnancy	1=yes, 2=no	Lifestyle during pregnancy
Irritability level during pregnancy	Irritability.level	1=very low, 2=low, 3=moderate, 4=high, 5=very high (ordinal)	Psychosocial factors during pregnancy
Emotional dimension of personality	Emotional.dimension	1=optimistic, 2=pessimistic	Psychosocial factors during pregnancy

Variable	Variable name	Coding description	Category
Social dimension of personality	Social.dimension	1=introverted, 2=extroverted	Psychosocial factors during pregnancy
Frequency of quarrels during pregnancy	Quarrel	1=often, 2=sometimes, 3=rarely, 4=almost never	Psychosocial factors during pregnancy
Psychological attack during pregnancy	Psychological.attack.during.pregnancy	1=yes, 2=no	Psychosocial factors during pregnancy
Regular prenatal examination	Regular.prenatal.examination	1=yes, 2=no	Psychosocial factors during pregnancy
Non-patent foramen ovale (outcome)	Non_PFO	1=yes, 0=no	Outcome variable

Variable	Variable name	Coding description	Category
Parity (number of deliveries)	Fertility.history	Continuous value (number of deliveries)	Maternal obstetric history

Notes:

Continuous variables are presented as raw values with units indicated.

All categorical variables are dummy-coded as shown.

The outcome variable Non_PFO indicates the absence of patent foramen ovale (coded as 1) or presence (coded as 0).

Table 2. Baseline characteristics of participants stratified by non-PFO status.

Characteristic	Level	Non-PFO (−) (n = 78)	Non-PFO (+) (n = 130)
Child characteristics			
Gestational age, months, mean (SD)	–	14.36 (20.90)	28.02 (30.70)

Characteristic	Level	Non-PFO (−) (n = 78)	Non-PFO (+) (n = 130)
Birth weight, g, mean (SD)	–	3151.41 (759.14)	3037.06 (597.45)
Sex, n (%)	Female	45 (57.7)	64 (49.2)
	Male	33 (42.3)	66 (50.8)
Ethnicity, n (%)	Han	45 (57.7)	73 (56.2)
	Hui	29 (37.2)	41 (31.5)
	Li	3 (3.8)	11 (8.5)
	Miao	1 (1.3)	3 (2.3)
	Other	0 (0.0)	2 (1.5)
Mode of delivery, n (%)	Vaginal delivery	50 (64.1)	83 (63.8)

Characteristic	Level	Non-PFO (−) (n = 78)	Non-PFO (+) (n = 130)
	Cesarean section	28 (35.9)	47 (36.2)
Premature birth, n (%)	No	61 (78.2)	101 (77.7)
	Yes	17 (21.8)	29 (22.3)
Comorbidities, n (%)	No	74 (94.9)	124 (95.4)
	Yes	4 (5.1)	6 (4.6)
Other congenital heart disease history, n (%)	Absent	51 (65.4)	76 (58.5)
	Present in other children	26 (33.3)	43 (33.1)
	Present	1 (1.3)	11 (8.5)
Parental sociodemographics			

Characteristic	Level	Non-PFO (−) (n = 78)	Non-PFO (+) (n = 130)
Paternal age, years, mean (SD)	–	33.56 (6.21)	32.58 (6.13)
Maternal age, years, mean (SD)	–	30.47 (5.08)	29.92 (5.51)
Paternal ethnicity, n (%)	Han	47 (60.3)	78 (60.0)
	Hui	28 (35.9)	37 (28.5)
	Li	3 (3.8)	5 (3.8)
	Miao	0 (0.0)	8 (6.2)
	Other	0 (0.0)	2 (1.5)
Maternal ethnicity, n (%)	Han	45 (57.7)	77 (59.2)
	Li	30 (38.5)	49 (37.7)

Characteristic	Level	Non-PFO (−) (n = 78)	Non-PFO (+) (n = 130)
	Miao	3 (3.8)	2 (1.5)
	Other	0 (0.0)	2 (1.5)
Paternal height, cm, mean (SD)	–	168.49 (5.40)	167.04 (15.60)
Paternal weight, kg, mean (SD)	–	68.40 (18.09)	70.97 (22.75)
Maternal height, cm, mean (SD)	–	156.72 (6.81)	156.46 (6.27)
Maternal weight, kg, mean (SD)	–	59.50 (19.01)	60.71 (21.45)
Maternal education, n (%)	Illiteracy	2 (2.6)	0 (0.0)
	Primary school	3 (3.8)	11 (8.5)
	Junior high school	39 (50.0)	57 (43.8)

Characteristic	Level	Non-PFO (−) (n = 78)	Non-PFO (+) (n = 130)
	High school	5 (6.4)	11 (8.5)
	Technical school	1 (1.3)	5 (3.8)
	Secondary specialized school	11 (14.1)	7 (5.4)
	Junior college	8 (10.3)	21 (16.2)
	Bachelor's degree	9 (11.5)	15 (11.5)
	Postgraduate or above	0 (0.0)	3 (2.3)
Paternal education, n (%)	Illiteracy	2 (2.6)	0 (0.0)
	Primary school	5 (6.4)	11 (8.5)
	Junior high school	40 (51.3)	53 (40.8)

Characteristic	Level	Non-PFO (−) (n = 78)	Non-PFO (+) (n = 130)
	High school	5 (6.4)	15 (11.5)
	Technical school	4 (5.1)	8 (6.2)
	Secondary specialized school	5 (6.4)	14 (10.8)
	Junior college	9 (11.5)	16 (12.3)
	Bachelor's degree	8 (10.3)	11 (8.5)
	Postgraduate or above	0 (0.0)	2 (1.5)
Monthly household income per capita, yuan, n (%)	<1000	12 (15.4)	21 (16.2)
	1000–3000	35 (44.9)	52 (40.0)
	3000–5000	20 (25.6)	39 (30.0)

Characteristic	Level	Non-PFO (−) (n = 78)	Non-PFO (+) (n = 130)
	5000–10000	8 (10.3)	13 (10.0)
	10000–20000	1 (1.3)	3 (2.3)
	>20000	2 (2.6)	2 (1.5)
Parity (number of deliveries), mean (SD)	–	1.95 (0.94)	1.86 (0.87)
Parental health and family history			
Maternal medical history (any disease), n (%)	No	58 (74.4)	100 (76.9)
	Yes	20 (25.6)	30 (23.1)
Paternal medical history (any disease), n (%)	No	67 (85.9)	111 (85.4)
	Yes	11 (14.1)	19 (14.6)

Characteristic	Level	Non-PFO (−) (n = 78)	Non-PFO (+) (n = 130)
Paternal family history of CHD, n (%)	No	77 (98.7)	129 (99.2)
	Yes	1 (1.3)	1 (0.8)
Maternal family history of CHD, n (%)	No	78 (100.0)	127 (97.7)
	Yes	0 (0.0)	3 (2.3)
Consanguineous marriage, n (%)	No	76 (97.4)	128 (98.5)
	Yes	2 (2.6)	2 (1.5)
Medication and assisted reproduction			
Maternal medication during pregnancy, n (%)	No	63 (80.8)	104 (80.0)
	Yes	15 (19.2)	26 (20.0)

Characteristic	Level	Non-PFO (−) (n = 78)	Non-PFO (+) (n = 130)
Assisted reproductive technology, n (%)	No	72 (92.3)	115 (88.5)
	Yes	6 (7.7)	15 (11.5)
Nutritional supplementation during pregnancy			
Folic acid supplementation, n (%)	No	9 (11.5)	24 (18.5)
	Yes	69 (88.5)	106 (81.5)
Multivitamin supplementation, n (%)	No	66 (84.6)	101 (77.7)
	Yes	12 (15.4)	29 (22.3)
Calcium supplementation, n (%)	No	40 (51.3)	74 (56.9)
	Yes	38 (48.7)	56 (43.1)

Characteristic	Level	Non-PFO (−) (n = 78)	Non-PFO (+) (n = 130)
Iron supplementation, n (%)	No	49 (62.8)	88 (67.7)
	Yes	29 (37.2)	42 (32.3)
Other nutrient supplementation, n (%)	No	70 (89.7)	107 (82.3)
	Yes	8 (10.3)	23 (17.7)
Parental pre-pregnancy lifestyle			
Paternal smoking before pregnancy, n (%)	No	25 (32.1)	54 (41.5)
	Yes	53 (67.9)	76 (58.5)
Paternal drinking before pregnancy, n (%)	No	44 (56.4)	82 (63.1)
	Yes	34 (43.6)	48 (36.9)

Characteristic	Level	Non-PFO (−) (n = 78)	Non-PFO (+) (n = 130)
Maternal smoking before pregnancy, n (%)	No	77 (98.7)	119 (91.5)
	Yes	1 (1.3)	11 (8.5)
Maternal drinking before pregnancy, n (%)	No	70 (89.7)	120 (92.3)
	Yes	8 (10.3)	10 (7.7)
Environmental exposure during pregnancy			
First-trimester exposure to special substances, n (%)	No	70 (89.7)	125 (96.2)
	Yes	8 (10.3)	5 (3.8)
Indoor home decoration, n (%)	No	75 (96.2)	127 (97.7)
	Yes	3 (3.8)	3 (2.3)

Characteristic	Level	Non-PFO (−) (n = 78)	Non-PFO (+) (n = 130)
Presence of pollution-related factory, n (%)	No	78 (100.0)	129 (99.2)
	Yes	0 (0.0)	1 (0.8)
Presence of gas station, n (%)	No	76 (97.4)	125 (96.2)
	Yes	2 (2.6)	5 (3.8)
Presence of garbage station, n (%)	No	74 (94.9)	121 (93.1)
	Yes	4 (5.1)	9 (6.9)
Presence of viaduct or highway, n (%)	No	76 (97.4)	128 (98.5)
	Yes	2 (2.6)	2 (1.5)
Presence of main traffic road, n (%)	No	65 (83.3)	108 (83.1)

Characteristic	Level	Non-PFO (−) (n = 78)	Non-PFO (+) (n = 130)
	Yes	13 (16.7)	22 (16.9)
Lifestyle and psychosocial factors during pregnancy			
Protein intake frequency, n (%)	Multiple times/day	15 (19.2)	40 (30.8)
	Once/day	31 (39.7)	54 (41.5)
	Once/2–3 days	19 (24.4)	19 (14.6)
	Less often	13 (16.7)	17 (13.1)
Fruit/vegetable intake frequency, n (%)	Multiple times/day	27 (34.6)	50 (38.5)
	Once/day	34 (43.6)	56 (43.1)
	Once/2–3 days	6 (7.7)	16 (12.3)

Characteristic	Level	Non-PFO (−) (n = 78)	Non-PFO (+) (n = 130)
	Less often	11 (14.1)	8 (6.2)
Passive smoking during pregnancy, n (%)	No	66 (84.6)	109 (83.8)
	Yes	12 (15.4)	21 (16.2)
Source of drinking water, n (%)	Tap water	61 (78.2)	97 (74.6)
	Well water	15 (19.2)	27 (20.8)
	Spring water	1 (1.3)	3 (2.3)
	River water	0 (0.0)	1 (0.8)
	Other	1 (1.3)	2 (1.5)
Keeping pets during pregnancy, n (%)	No	75 (96.2)	118 (90.8)

Characteristic	Level	Non-PFO (−) (n = 78)	Non-PFO (+) (n = 130)
	Yes	3 (3.8)	12 (9.2)
Raising poultry during pregnancy, n (%)	No	66 (84.6)	104 (80.0)
	Yes	12 (15.4)	26 (20.0)
Irritability level, n (%)	Very low	34 (43.6)	52 (40.0)
	Low	25 (32.1)	36 (27.7)
	Moderate	12 (15.4)	28 (21.5)
	High	3 (3.8)	9 (6.9)
	Very high	4 (5.1)	5 (3.8)
Emotional dimension, n (%)	Optimistic	70 (89.7)	115 (88.5)

Characteristic	Level	Non-PFO (−) (n = 78)	Non-PFO (+) (n = 130)
	Pessimistic	8 (10.3)	15 (11.5)
Social dimension, n (%)	Extroverted	58 (74.4)	84 (64.6)
	Introverted	20 (25.6)	46 (35.4)
Frequency of quarrels, n (%)	Almost never	42 (53.8)	43 (33.1)
	Rarely	21 (26.9)	60 (46.2)
	Sometimes	12 (15.4)	21 (16.2)
	Often	3 (3.8)	6 (4.6)
Psychological attack during pregnancy, n (%)	No	73 (93.6)	118 (90.8)
	Yes	5 (6.4)	12 (9.2)

Characteristic	Level	Non-PFO (−) (n = 78)	Non-PFO (+) (n = 130)
Regular prenatal examination, n (%)	No	10 (12.8)	17 (13.1)
	Yes	68 (87.2)	113 (86.9)

Notes: Non-PFO (−) indicates absence of the outcome (i.e., patent foramen ovale or other non-PFO conditions), and Non-PFO (+) indicates presence of non-PFO. Data are presented as mean (standard deviation) for continuous variables and as number (percentage) for categorical variables. Percentages may not sum to 100 due to rounding. CHD, congenital heart disease.