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.

Early risk stratification for antenatal depression-anxiety-sleep comorbidity: a three-factor nomogram based on PP-LCGM joint trajectories

Meng Zhao

zm0991zm@163.com

Xinjiang Medical University

ZhiXiong Li

Karamay Integrated Traditional Chinese and Western Medicine Hospital (People's Hospital of Karamay)

ZhiYue Zhang

Xinjiang Uygur Autonomous Region

QiQi Yang

Xinjiang Second Medical Colleg

JiaYing Tao

Xinjiang Medical University

Nan Zhang

Xinjiang Medical University

ShuPing You

Xinjiang Medical University

Mei Sun

Central South University

SiYuan Tang

Xinjiang Medical University

Research Article

Keywords: Antenatal depression, Antenatal anxiety, Sleep disturbance, Pregnancy, Joint trajectory, Machine learning, Nomogram, Risk stratification, Latent class growth modeling

Posted Date: May 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9755032/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

Antenatal depression, anxiety, and sleep disturbance are highly comorbid, yet early risk stratification tools remain underdeveloped. This study aimed to identify joint trajectories of these three symptoms from the second trimester (20 to 22 weeks) to the third trimester (36 to 38 weeks) through the parallel process latent class Growth model (PP-LCGM) and to develop a parsimonious early risk model using only baseline data (11 to 13 weeks).

Methods

A prospective longitudinal multicenter study enrolled 1,302 pregnant women in western China. PP-LCGM identified trajectory classes as the gold-standard outcome. A two-step evidence synthesis (Step 1: parallel Boruta shadow-feature classification and Bonferroni-corrected regression; union of selected predictors; Step 2: bootstrap optimism-corrected AUC comparison of nested candidate models) identified robust predictors from 18 candidates. Six machine learning (ML) algorithms were compared via cross-validation and external validation. The optimal model was translated into a nomogram with binary risk stratification. SHAP (SHapley Additive exPlanations) analysis quantified predictor contributions.

Results

PP-LCGM identified two trajectories: persistent comorbid (36.2% training, 33.5% test) and non-comorbid. The Self-Rating Anxiety Scale (SAS) (OR = 2.74), Self-Rating Depression Scale (SDS) (OR = 3.04), and Pittsburgh Sleep Quality Index (PSQI) (OR = 1.20) were selected as final predictors. The 3-factor logistic regression achieved optimism-corrected AUC = 0.915, with external validation AUC = 0.919 (95% CI: 0.890–0.945), sensitivity = 0.855 and specificity = 0.835 at the Youden-optimal threshold (0.316). DeLong tests showed no significant AUC difference between logistic regression and complex alternatives (P > 0.05). The nomogram demonstrated excellent calibration and positive net benefit. SHAP analysis confirmed the dominant contribution of the psychological triad and quantified PSQI's independent incremental value.

Conclusions

This study presents a clinically actionable early risk stratification tool for antenatal depression-anxiety-sleep comorbidity. The 3-factor model achieves optimal generalizability with minimal parameters. The nomogram enables immediate bedside risk assessment at the first antenatal visit.

Background

Antenatal depression, anxiety, and sleep disturbance constitute a clinically significant triad of interrelated conditions that substantially compromise maternal physical and mental health and exert adverse effects on short- and long-term maternal–infant outcomes[1–6]. The prevalence of antenatal depression is as high as 26.3%, commanding considerable attention[7], while the prevalence of antenatal anxiety reaches 21.1%, which is equally noteworthy [8]. Beyond their independent detrimental effects, the co-occurrence of depression and anxiety during pregnancy exerts synergistic adverse impacts on maternal–infant health[9], and poor sleep quality further amplifies this burden[10]. In clinical practice, these three conditions interact and reinforce one another during gestation, not only increasing the risk of postpartum depression but also heightening the complexity of disease identification and intervention, thereby posing a major challenge to antenatal mental health care [11–13]. This tripartite reinforcing relationship has been quantitatively supported by meta-

analytic evidence: poor sleep quality during pregnancy is significantly associated with both antenatal depression (OR = 3.33, 95% CI: 2.38–4.66) and postpartum depression (OR = 2.71, 95% CI: 1.90–3.87) [14]. Moreover, accumulating evidence indicates that depression, anxiety, and sleep disturbance share common biological mechanisms[15]. Taken together, existing studies consistently demonstrate that antenatal depression-anxiety-sleep (D-A-S) comorbidity is characterized by high prevalence, substantial clinical significance, and complex interdependence, underscoring the urgency of in-depth research in this field, particularly the development of precise predictive tools to enable early risk stratification and targeted intervention.

Trajectory modeling, as a methodological framework for analyzing disease progression dynamics and identifying high-risk populations, has been widely applied in perinatal mental health research [5, 11, 12]. Within this methodological domain, joint trajectory approaches—which simultaneously model multiple interrelated symptoms within the same cohort—offer unique advantages for capturing the inherent interdependence of comorbid states [16, 17]. However, the application of such joint trajectory methods in perinatal mental health remains markedly insufficient and methodologically constrained. Existing studies have predominantly employed unidimensional or bidimensional trajectory designs, overlooking the intrinsic associations and interactions among depression, anxiety, and sleep disturbance. This dimensional limitation has created three critical gaps in the current evidence base. First, joint trajectory studies encompassing all three comorbidity dimensions (depression, anxiety, and sleep disturbance) are scarce, restricting a comprehensive understanding of the dynamic evolution of this triad. Second, efficient risk prediction techniques for comorbid trajectories are lacking, impeding early identification and warning of high-risk populations. Third, there is insufficient understanding of the key determinants that dictate membership in triad comorbidity trajectory classes. These interrelated limitations severely constrain the effective implementation of precision management strategies for antenatal mental health.

Further in-depth analysis of the existing literature reveals that studies systematically exploring joint trajectories of depression, anxiety, and sleep quality within the same pregnant cohort are extremely limited, with each study addressing only a subset of the three dimensions. Gueron-Sela et al. [13] identified bivariate postpartum depression-anxiety trajectories linked to inadequate sleep during pregnancy in a longitudinal cohort of pregnant women followed through the postpartum period, though the focus was on infant sleep outcomes rather than maternal comorbidity patterns. Pan et al. [18] applied dual-trajectory modeling to investigate the co-development of insomnia and depression but did not include anxiety as a parallel process, precluding a comprehensive assessment of triadic interactions. Overall, these studies are confined to bivariate rather than trivariate comorbidity analysis, and none developed clinical prediction tools for risk stratification. The study by Jiang M et al.[19] represents the closest precedent to the PP-LCGM (parallel-process latent class growth modeling) approach used in the present study, having modeled joint trajectories of perinatal depression and anxiety; nevertheless, it included only two dimensions, omitted sleep quality, and did not develop clinical prediction tools. These limitations have resulted in an inadequate understanding of the population heterogeneity, dynamic evolution patterns, and key determinants of antenatal triad comorbidity, hindering the development of evidence-based early screening and intervention strategies.

Based on this background, the present study comprises three integrated components. First, we employ PP-LCGM to simultaneously model depression, anxiety, and sleep quality from the second to the third trimester, constructing them as joint trajectories to identify comorbid trajectories and elucidate their distributional characteristics and dynamic evolution. Second, we adopt a two-step evidence synthesis framework to systematically identify robust predictors of trajectory class membership: Step 1 employs parallel independent screening (Boruta shadow feature classification and Bonferroni-corrected regression) to identify candidate predictors; Step 2 uses bootstrap optimism-corrected nested model AUC comparison to determine the final parsimonious model. This framework effectively prevents overfitting while identifying robust baseline predictors, and subsequently elucidates the key determinants of D-A-S comorbidity trajectories through systematic comparison of six machine learning algorithms. Third, based on the above predictors, we develop an interpretable three-variable nomogram with binary risk stratification and an accompanying calculator to facilitate direct

clinical deployment, providing a theoretical foundation and practical tool for early screening, risk stratification, and targeted intervention.

Methods

Study Design and Participants

A prospective longitudinal multicenter study was conducted in western China between June 2023 and August 2024. Pregnant women undergoing regular antenatal check-ups at participating hospitals and community health centers were consecutively approached. A total of 1,500 potential participants were initially recruited; of these, 1,302 met inclusion criteria and provided written informed consent.

Inclusion criteria: normal language comprehension and expression abilities, and intention to reside in the city throughout pregnancy. Exclusion criteria: multiple gestation, history of psychiatric or personality disorders, intellectual impairment, substance or alcohol abuse, severe maternal infections, and fetal abnormalities.

Sample size justification. With 346 comorbid events in the training set (n = 911) and 3 final predictors, the events-per-variable (EPV) ratio was 115.7, substantially exceeding the minimum threshold of 10. The external validation set (n = 391; 140 events, 251 non-events) exceeded recommended thresholds for reliable performance estimation.

Data Collection and Follow-up

Data were collected via electronic questionnaires and antenatal medical records by uniformly trained obstetricians through face-to-face explanations, after which participants completed questionnaires independently. Assessments were conducted at four time points: T1 (11–13 weeks gestation), T2 (20–22 weeks), T3 (30–32 weeks), and T4 (36–38 weeks). Baseline assessment at T1 included depression (SDS), anxiety (SAS), sleep quality (PSQI), subjective well-being (OHQ-SF), and comprehensive sociodemographic and obstetric characteristics. Follow-ups at T2-T4 repeated the psychological and sleep assessments. Of 1,500 enrolled participants, 1,302 completed through T4 (attrition rate: 13.2%), with retention rates of 100%, 94.3%, 90.1%, and 86.8% at T1–T4, respectively. No significant baseline differences emerged between retained and attrited participants in age, education, or symptom severity.

Dataset Partitioning for Machine Learning

A geographic stratified partitioning strategy was employed to enhance generalizability. A total of 911 participants from central urban sites constituted the training set for model development. The remaining 391 participants from non-central sites (> 100 km away) formed the independent external validation set, as illustrated in Fig. 1.

[Insert the Fig. 1 here]

Figure 1 Participant enrollment and train/test split

Study Variables and Categorization

Eighteen candidate baseline predictors were extracted from electronic questionnaires and antenatal records and categorized into six domains: (1) Sociodemographic factors: age, educational background, family monthly income, income-expenditure balance; (2) Obstetric factors: parity, number of spontaneous/artificial abortions, pregnancy intention; (3) Behavioral factors: daily physical activity duration, weekly exercise frequency; (4) Personality/interpersonal factors:

self-assessed personality traits (introverted/extroverted/middle), social interaction proficiency; (5) Family factors: housing satisfaction, presence of sleep disruption factors at home; (6) Pregnancy-related physical and psychological factors: body aches, nausea frequency, fear of childbirth, well-being, and baseline SDS, SAS, and PSQI scores.

Standardized Assessment Scales

Four self-rating scales were administered at each time point. Scoring followed Chinese norms for adult females; internal consistency was evaluated using Cronbach's alpha.

Self-Rating Depression Scale (SDS) [20]

The 20-item SDS assesses depressive symptom severity. Raw scores are multiplied by 1.25 to obtain standard scores. Cutoff: greater than or equal to 53 indicates abnormal symptoms (mild: 53–62, moderate: 63–72, severe: greater than or equal to 72). Cronbach's alpha at T1–T4: 0.877–0.880.

Self-Rating Anxiety Scale (SAS) [21]

The 20-item SAS measures anxiety symptoms using the same scoring method (raw score times 1.25). Cutoff: greater than or equal to 50 indicates abnormal symptoms (mild: 50–59, moderate: 60–69, severe: greater than or equal to 70). Cronbach's alpha at T1–T4: 0.871–0.880.

Pittsburgh Sleep Quality Index (PSQI) [22]

The 19-item PSQI evaluates sleep quality across seven dimensions (each 0–3): subjective sleep quality, latency, duration, efficiency, disturbances, medication use, and daytime dysfunction. Total score range: 0–21, with > 5 indicating sleep problems. Cronbach's alpha at T1–T4: 0.892–0.895.

Oxford Happiness Questionnaire-Short Form (OHQ-SF) [23]

The 8-item OHQ-SF (6-point Likert scale, range: 8–48) assesses subjective well-being. The cohort was dichotomized at the training set median for subgroup analysis. Cronbach's alpha = 0.835.

Ethical Considerations

This study was approved by the Ethics Committee of Xinjiang Medical University (Approval Number: XJYKDXR20230303010). Written informed consent was obtained from all participants prior to enrollment. All participants were informed of their right to withdraw at any time. Their personal information was used only for this study and will remain strictly confidential. This research was supported by the Natural Science Foundation of Xinjiang Uygur Autonomous Region, China (Project number: 2022D01F32). During the preparation of this manuscript, the authors used ChatGPT (GPT-4, OpenAI) for language polishing and structural suggestions. All scientific content, data analysis, clinical interpretations, statistical computations, and final editorial decisions were developed and verified by the authors independently.

Parallel Process Latent Class Growth Modeling

PP-LCGM jointly modeled SDS (standard scores), SAS (standard scores), and PSQI (raw scores) across three time points (T2: 20–22, T3: 30–32, T4: 36–38 weeks). Models with 2–5 latent classes were fitted (within-class variances constrained to zero), with class selection based on AIC, BIC, aBIC, LMR-LRT, and entropy.

Machine Learning-Based Early Risk Stratification

To develop a clinically deployable early risk stratification tool, a four-phase analytical pipeline was implemented: (1) two-step evidence synthesis to identify robust baseline predictors; (2) systematic comparison of six machine learning algorithms; (3) development of an interpretable 3-variable nomogram; and (4) post-hoc explainability analysis using SHAP values. All predictive models were trained exclusively on baseline data (T1, 11–13 weeks gestation), with PP-LCGM trajectory classification (T2–T4) as the binary outcome. The sample was divided into a training set (n = 911) and an independent external test set (n = 391) based on geographical stratification. All statistical analyses employed two-sided tests at alpha = 0.05. Analytical tools included Mplus 8.3, R 4.4.1 (glmnet), and Python 3.10 (scikit-learn 1.3, XGBoost 2.0, LightGBM 4.1, CatBoost 1.2, SHAP 0.42, statsmodels 0.14).

Data Preprocessing

Continuous variables were standardized (Z-scores); categorical variables were one-hot encoded. Multicollinearity was assessed via variance inflation factor (VIF; values < 5 acceptable, greater than or equal to 10 severe) and condition number of the design matrix (threshold 30).

Two-Step Evidence Synthesis

A two-step evidence synthesis was implemented combining complementary validation perspectives.

Step 1: Parallel predictor screening. Two independent methods screened 18 candidate predictors simultaneously, with results combined by union:

Method A (machine-learning confirmation)

Boruta shadow-feature classification (n_estimators = 200, max_depth = 5, 200 iterations) compared each candidate predictor against randomized shadow features; predictors consistently outperforming noise were designated Confirmed.

Method B (statistical significance)

Multivariable logistic regression with all 18 predictors entered simultaneously; Bonferroni correction controlled family-wise error (adjusted alpha = 0.05/18 = 0.0028). Predictors with adjusted P < 0.0028 were designated Primary; 0.0028 less than or equal to adjusted P < 0.05 as Exploratory.

Union

Predictors selected by either Method A (Confirmed) or Method B (Primary/Exploratory) entered Step 2. This union reduces false negatives from any single screening method.

Step 2: Bootstrap nested-model validation. The union-derived candidate predictors were arranged in hierarchical nested models and evaluated by 1,000-resample optimism bootstrap [24]. The final model was the most parsimonious achieving the highest optimism-corrected AUC, with optimism-corrected to apparent AUC ratio greater than or equal to 0.90.

Systematic Machine Learning Comparison

Six ML algorithms (LR, RF, XGBoost, LightGBM, CatBoost, NN) were compared under identical conditions: 5-fold stratified cross-validation with independent external validation (n = 391), SMOTE applied to training folds only, and Bayesian hyperparameter optimization (Optuna, 100 iterations). Evaluation metrics: AUC, AUPRC, sensitivity, specificity, F1-score, Brier score. DeLong tests compared pairwise AUCs. The optimal threshold was determined by Youden's J on original imbalanced validation folds. A repeated 10 x 5-fold stratified cross-validation confirmed ranking stability.

Nomogram Development and Risk Stratification

The 3-factor logistic regression was translated into a nomogram using coefficient-to-points mapping (SAS: 20–80, SDS: 20–80, PSQI: 0–21). Binary risk stratification used a Youden-optimized threshold (median 0.316 from 1,000 bootstrap resamples; IQR: 0.300–0.333; 95% CI: 0.268–0.390). Low risk: predicted probability < 31.6%; High risk: greater than or equal to 31.6%. A web-based calculator was developed for bedside use (https://kp7tztyjmga.ok.kimi.link).

SHAP Analysis

SHAP values were computed using LinearExplainer (exact calculation for linear models) with the external validation set (n = 391) as background.

Common Method Bias

The expanded Harman single-factor test (17 self-report variables) assessed common method bias against the 50% variance threshold.

Results

PP-LCGM Trajectory Analysis

PP-LCGM was fitted to the training set (n = 911) with 1- to 5-class solutions. Model fit indices collectively supported the 2-class solution: the 2-class model achieved the highest entropy and a significant LMR-LRT (P < 0.0001), supporting selection of the two-class structure. The test set, strictly excluded from model selection, confirmed reproducibility with entropy = 0.924 and LMR-LRT P < 0.001; although not the highest entropy, other class solutions in the test set contained an extremely small subgroup (0.02%), violating the assumption of adequate class representation.For details, see Table 1.

Table 1
Fitting Indicators of the Piecewise Polynomial Latent Class Growth Model (PP-LCGM) for Solutions of Categories 1 to 5

Dataset	Classification	AIC	BIC	a-BIC	Entropy	LMR-LRT(p)	Class Proportions
Training Set	1 Class	50836.517	50908.735	50861.097	-	-	1.000
	2 Class	48562.734	48668.654	48598.785	0.865	< 0.0001	0.66/0.34
	3 Class	47888.174	48027.796	47935.696	0.836	0.2478	0.32/0.46/0.22
	4 Class	47455.298	47628.622	47514.291	0.864	0.0001	0.20/0.08/0.25/0.47
	5 Class	47215.984	47423.009	47286.447	0.823	0.2892	0.34/0.19/0.18/0.21/0.08
Test Set	1 Class	22006.132	22065.662	22018.068			
	2 Class	20886.254	20973.566	20903.761	0.924	< 0.0001	0.65/0.35
	3 Class	20552.862	20667.955	20575.939	0.948	< 0.0001	0.34/0.63/0.02
	4 Class	20217.813	20360.687	20246.460	0.911	< 0.0001	0.02/0.26/0.27/0.45
	5 Class	20055.951	20226.605	20090.168	0.913	0.0208	0.27/0.23/0.05/0.43/0.02

Footnotes: *Boldface indicates the selected model. AIC: Akaike information criterion; BIC: Bayesian information criterion; a-BIC: sample-size-adjusted Bayesian information criterion; LMR-LRT: Lo–Mendell–Rubin adjusted likelihood ratio test. The bold*

category is the final selection category.

The 2-class model identified persistent comorbid and non-comorbid trajectories; the persistent comorbid trajectory accounted for 36.2% in the training set and 33.5% in the test set, while the non-comorbid trajectory accounted for 63.8% in the training set and 66.5% in the test set. As shown in Fig. 2, the comorbid and non-comorbid groups exhibited distinctly different trajectory patterns from T2 to T4.

In the comorbid group, SDS scores remained above the clinical cutoff of 53 at all three time points (training set: 54.45-55.99-53.43; test set: 54.13-54.43-53.84). PSQI scores also remained above the cutoff of 5 (training set: 5.61-5.84-5.56; test set: 5.58-5.90-6.16). SAS scores fluctuated near the cutoff of 50: the training set peaked at T3 (50.64) then declined to 48.35 at T4; the test set was 50.87 at T2 then gradually decreased to 49.68 at T4. For details, please refer to Fig. 2.

[Insert the Fig. 2 here]

Figure 2. Estimated mean trajectories of comorbid depression-anxiety-sleep disturbance from T2 to T4.

Legend: Trajectories were derived from the parallel process latent class growth model (PP-LCGM) and represent class-specific estimated means. (A) Training set, comorbid group. (B) Training set, non-comorbid group. (C) Testing set, comorbid group (n = 136). (D) Testing set, non-comorbid group (n = 2). Dashed horizontal lines denote clinical cutoffs: PSQI ≥ 5 (green), SAS ≥ 50 (orange), and SDS ≥ 53 (blue). SDS, Self-Rating Depression Scale; SAS, Self-Rating Anxiety Scale; PSQI, Pittsburgh Sleep Quality Index.

Baseline characteristics of the training cohort by comorbidity trajectory group are shown in Table 2.

Table 2
Baseline characteristics of the training cohort by comorbidity trajectory group

Variable	Overall (N = 911)	Comorbid Group (N = 330)	Non-Comorbid Group (N = 581)	Statistics	P-value
PSQI score, median [IQR]	5.00 [3.00, 7.00]	6.00 [4.00, 8.00]	4.00 [3.00, 6.00]	**7.23**	**< 0.001**
SAS score, median [IQR]	43.80 [38.80, 48.80]	48.80 [43.80, 52.50]	40.30 [36.30, 46.30]	**14.87**	**< 0.001**
SDS score, median [IQR]	51.00 [43.75, 56.25]	56.38 [52.56, 60.00]	46.25 [42.50, 52.50]	**16.28**	**< 0.001**
Age				3.76	0.052
≤ 29 years	515 (56.53%)	201 (60.91%)	314 (54.04%)		
> 30 years	396 (43.47%)	129 (39.09%)	267 (45.96%)		
Education background				**20.60**	**< 0.001**
Junior high school and below	168 (18.44%)	64 (19.39%)	104 (17.90%)		
High school or technical secondary	249 (27.33%)	117 (35.45%)	132 (22.72%)		
College or bachelor degree or above	494 (54.23%)	149 (45.15%)	345 (59.38%)		
Parity				**19.43**	**< 0.001**
= 1	609 (66.85%)	190 (57.58%)	419 (72.12%)		
≥ 2	302 (33.15%)	140 (42.42%)	162 (27.88%)		
Satisfaction with housing conditions				**8.74**	**0.003**
Satisfied	780 (85.62%)	267 (80.91%)	513 (88.30%)		
Moderately satisfied	131 (14.38%)	63 (19.09%)	68 (11.70%)		
Sleep disruption factors at home				**8.75**	**0.003**
Present	216 (23.71%)	97 (29.39%)	119 (20.48%)		
Absent	695 (76.29%)	233 (70.61%)	462 (79.52%)		
Family income and expenditure				1.15	0.563
Marginally insufficient	104 (11.42%)	34 (10.30%)	70 (12.05%)		
Balanced	666 (73.11%)	248 (75.15%)	418 (71.94%)		
Surplus	141 (15.48%)	48 (14.55%)	93 (16.01%)		
Fear of childbirth				1.02	0.313
Present	379 (41.60%)	145 (43.94%)	234 (40.28%)		

Variable	Overall (N = 911)	Comorbid Group (N = 330)	Non-Comorbid Group (N = 581)	Statistics	P-value
Absent	532 (58.40%)	185 (56.06%)	347 (59.72%)		

Footnotes: Data are presented as n (%) for categorical variables and median [IQR] for continuous variables. Bold text indicates statistically significant differences between the two groups (P < 0.05). Pearson's Chi-squared test was used for categorical variables; Wilcoxon rank sum test for continuous variables. Abbreviations: PSQI = Pittsburgh Sleep Quality Index; SAS = Self-Rating Anxiety Scale; SDS = Self-Rating Depression Scale; IQR = interquartile range

Common Method Bias Assessment

Harman's single-factor test showed the first unrotated factor explained 37.0% of variance for the five core self-report scales and 12.4% for all 17 self-report variables, both below the 50% threshold, indicating that common method bias is not a threatening concern.The results of the common method bias assessment are summarized in Table 3.

Table 3
Common method bias assessment using Harman's single-factor test

Variable set	Number of items	First unrotated factor (% variance)	Threshold	Conclusion
Five core self-report scales	5	37.0%	50%	Below threshold
All 17 self-report variables	17	12.4%	50%	Below threshold

Footnotes

Common method bias assessment using Harman’s single-factor test. Values below the 50% threshold indicate that common method bias is not a threatening concern.

Two-Step Evidence Synthesis and Predictive Model Development

Boruta Feature Selection and Bonferroni-Corrected Regression Results

All 18 predictors were entered simultaneously into both Step 1 models.

Step 1 Method A applied Boruta shadow-feature classification, comparing 18 candidate predictors against randomized noise over 200 iterations. Three variables were classified as **Confirmed**: SAS, SDS, and PSQI scores.

Step 1 Method B fitted a multivariable logistic regression with all 18 predictors entered simultaneously. At the Bonferroni-adjusted threshold (P < 0.0028), three predictors were designated **Primary**: SAS (OR = 3.04, 95% CI: 2.94–5.17, adjusted P < 0.001); SDS (OR = 3.04, 95% CI: 2.88–5.04, adjusted P < 0.001); and extroverted personality (OR = 0.30, 95% CI: 0.16–0.59, adjusted P = 0.007). Three additional predictors were **Exploratory** (P < 0.05): housing satisfaction, well-being, and middle personality.

Union of Step 1 results: SAS, SDS, PSQI (Boruta Confirmed) and SAS, SDS, extroverted personality (Bonferroni Primary/Exploratory) = **four candidate predictors** for Step 2: SAS, SDS, PSQI, and extroverted personality.

[Insert the Fig. 3 here]

Figure 3 Boruta feature selection results for 18 candidate baseline predictors.

Footnotes

Boruta algorithm (n_estimators = 200, max_depth = 5, 200 iterations) compared each candidate predictor against randomized shadow features. Green bars indicate Confirmed predictors (SAS, SDS, PSQI) that consistently outperformed the maximum shadow feature (Shadow Max, blue dashed line). Red bars indicate Rejected predictors (n = 12) that failed to exceed the shadow mean threshold. No predictors were classified as Tentative. The vertical dashed line represents the Shadow Mean threshold; features to the right of Shadow Max are confirmed as having genuine predictive importance beyond random noise.

[Insert the Fig. 4 here]

Figure 4 Forest plot of the Bonferroni-corrected multivariable regression results.

Footnotes

Odds ratios (OR) with 95% confidence intervals for the 18 candidate predictors tested in the multivariable logistic regression model after Bonferroni correction (adjusted significance threshold α = 0.0028). The dashed vertical line at OR = 1 represents no effect. Predictors with 95% CIs excluding 1 and passing the Bonferroni threshold are considered statistically significant. SAS = Self-Rating Anxiety Scale; SDS = Self-Rating Depression Scale; PSQI = Pittsburgh Sleep Quality Index.

Bootstrap Optimism-Corrected AUC Comparison of Nested Models

Step 2 evaluated three nested candidate models by 1,000-resample optimism bootstrap [24]. The 3-factor model (SAS + SDS + PSQI) achieved the highest optimism-corrected AUC of 0.915 (95% CI: 0.892–0.944), with EPV = 82.5. The 2-factor model (without PSQI) demonstrated a lower corrected AUC of 0.901 (delta-AUC = + 0.014), confirming PSQI's independent incremental value. The 4-factor model (adding extroverted personality) declined to 0.903 (delta-AUC = -0.012 vs. 3-factor), indicating its Bonferroni significance (P = 0.003) did not translate into independent predictive gain. These results support the selection of the 3-factor model as the final clinical model. Bootstrap optimistic-corrected AUC for nested candidate models is presented in Table 4.

Table 4
Bootstrap optimism-corrected AUC for nested candidate models

Model	Apparent AUC	Optimism	Corrected AUC	95% CI	EPV
2-factor (SAS + SDS)	0.904	0.001	0.901	0.882–0.942	110.0
3-factor (SAS + SDS +PSQI)	0.919	0.001	0.915	0.892–0.944	82.5
4-factor (+ Extroverted)	0.912	0.002	0.903	0.887–0.945	66.0

Footnotes: Bold indicates the final selected model. The 3-factor model achieved the highest optimism-corrected AUC. EPV = events per variable.

Systematic Comparison of Six Machine Learning Algorithms

Six ML algorithms (LR, RF, XGBoost, LightGBM, CatBoost, NN) were compared under identical conditions: 5-fold stratified cross-validation with independent external validation (n = 391), SMOTE applied to training folds only, and Bayesian hyperparameter optimization. On external validation, LightGBM achieved the numerically highest AUC (0.910), followed by XGBoost (0.907), Random Forest (0.907), Logistic Regression (0.903), CatBoost (0.902), and Neural Network (0.851). DeLong tests showed no statistically significant differences between Logistic Regression and any complex algorithm (all

P > 0.05, including vs. LightGBM P = 0.369). A repeated 10 x 5-fold cross-validation confirmed ranking stability. Given comparable performance, superior calibration, and direct nomogram translatability, Logistic Regression was selected as the optimal algorithm. The performance comparison of six machine learning algorithms on external validation is presented in Table 5.

Table 5
Performance comparison of six machine learning algorithms on external validation (n = 391)

Algorithm	AUC (95% CI)	AUPRC	Sensitivity	Specificity	F1-score	Brier score
Logistic Regression	0.903 (0.875–0.934)	0.840	0.794	0.842	0.754	0.124
Random Forest	0.907 (0.875–0.939)	0.848	0.779	0.862	0.758	0.121
XGBoost	0.907 (0.875–0.939)	0.824	0.847	0.804	0.758	0.136
LightGBM	0.910 (0.879–0.941)	0.836	0.863	0.800	0.764	0.153
Neural Network	0.851 (0.813–0.889)	0.755	0.710	0.812	0.681	0.184
CatBoost	0.902 (0.870–0.934)	0.841	0.824	0.842	0.771	0.123

Footnotes: All algorithms were evaluated under identical conditions: 5-fold stratified cross-validation with independent external validation, SMOTE applied to training folds only, and Bayesian hyperparameter optimization. DeLong tests showed no statistically significant AUC differences between Logistic Regression and any complex algorithm (all P > 0.05). AUC = area under the receiver operating characteristic curve; AUPRC = area under the precision-recall curve.

SHAP Interpretability Analysis

Following the selection of logistic regression as the optimal algorithm, SHAP analysis was conducted to quantify each predictor's contribution and enhance clinical interpretability.After selecting logistic regression as the optimal algorithm, we performed SHAP analyses to quantify the contribution of each predictor and to improve clinical interpretability. See Fig. 5 for details.

[Insert the Fig. 5 here]

Figure 5 SHAP (SHapley Additive exPlanations) interpretability analysis for the 3-factor logistic regression model on external validation (n = 391)

Footnotes

SHAP summary dot plot showing the distribution of SHAP values for each predictor. Each dot represents one participant; dot color indicates the feature value (pink=high, blue = low), and horizontal position indicates the SHAP value (impact on model output). SAS and SDS show the widest SHAP value distributions, confirming their dominant predictive contributions. PSQI demonstrates a distinct independent incremental effect with a narrower but consistently positive SHAP distribution. The monotonic positive relationships support the clinical interpretation that elevated baseline symptoms in any domain increase the likelihood of persistent comorbidity.

Final Model Performance

On external validation (n = 391), the 3-factor model achieved an AUC of 0.919 (95% CI: 0.890–0.945), with sensitivity of 0.855, specificity of 0.835, PPV of 0.743, and NPV of 0.920 (Table 6). The model demonstrated excellent calibration (Brier score = 0.109) with intercept-corrected slope of 0.94 and negligible optimism (0.001), indicating that predicted probabilities closely matched observed outcomes (Fig. 6A). Decision curve analysis showed positive net benefit of the 3-factor nomogram across the clinically relevant threshold probability range (10%–60%), outperforming both treat-none and treat-all strategies (Fig. 6B). The Youden-optimized threshold of 0.316 (median from 1,000 bootstrap resamples; IQR: 0.300–0.333) was selected for binary risk stratification.

[Insert the Fig. 6 here]

Figure 6 Model performance evaluation on external validation set (n = 391)

Footnotes

(A) Calibration curve comparing predicted probabilities against observed proportions. The solid line represents the 3-factor logistic regression model with bootstrap 95% confidence interval. The dashed diagonal represents perfect calibration. Brier score = 0.109 indicates excellent calibration. The distribution histograms at the bottom show the density of predicted probabilities for non-comorbid and comorbid participants. (B) Decision curve analysis comparing the net benefit of the 3-factor nomogram against treat-none and treat-all strategies across threshold probabilities from 0% to 80%. The nomogram demonstrates positive net benefit across the clinically relevant range. The vertical dash-dot line marks the Youden-optimized threshold.

The final 3-factor logistic regression demonstrated strong discrimination on external validation (AUC = 0.919, 95% CI: 0.890–0.945). SAS (beta = 1.233, OR = 2.74, 95% CI: 2.67–4.41) and SDS (beta = 1.284, OR = 3.04, 95% CI: 2.80–4.67) were the strongest predictors, followed by PSQI (beta = 0.307, OR = 1.20, 95% CI: 1.01–1.42). At the Youden-optimized threshold of 0.316, sensitivity was 0.855, specificity 0.835, PPV 0.743, NPV 0.920, accuracy 84.1%, and Brier score 0.109. The model intercept was beta_0 = -2.156.The performance metrics of the final 3-factor logistic regression model are presented in Table 6.

Table 6
Performance of the final 3-factor logistic regression model

Metric	Training Set (5-fold CV)	External Validation (n = 391)
AUC	0.912 (0.884–0.940)	0.919 (0.890–0.945)
Sensitivity	0.821 (0.756–0.882)	0.855 (0.789–0.914)
Specificity	0.858 (0.815–0.898)	0.835 (0.790–0.878)
PPV	0.721 (0.658–0.783)	0.743 (0.674–0.810)
NPV	0.912 (0.876–0.945)	0.920 (0.884–0.953)
Accuracy (%)	83.5 (79.4–87.4)	84.1 (80.2–88.0)
Brier score	0.112 (0.096–0.130)	0.109 (0.092–0.126)
F1-score	0.769 (0.715–0.821)	0.783 (0.732–0.833)

Footnotes: Values are presented as optimism-corrected estimates (95% CI) from 1,000 bootstrap resamples for training set, and direct estimates (95% CI) for external validation. PPV = positive predictive value; NPV = negative predictive value.

Nomogram Development and Risk Stratification

Building upon the validated 3-factor model, a clinical nomogram was developed using raw scale scores (SAS: 20–80, SDS: 20–80, PSQI: 0–21) mapped to a point scale (total 0–123, predicted probability 2.3%–97.8%). See Fig. 6 for details. At the 0.316 threshold, 49.2% of the training cohort and 39.6% of the external validation cohort were classified as high risk, with PPVs of 69.0% and 72.3%, respectively. To facilitate clinical deployment, the nomogram was implemented as a freely accessible web-based calculator (https://kp7tztyjmga.ok.kimi.link), enabling immediate probability estimation and flexible threshold adjustment at the bedside.

[Insert the Fig. 7 here]

Figure 7 Clinical nomogram for predicting antenatal depression-anxiety-sleep comorbidity trajectory membership using raw scale scores at the first antenatal visit (11–13 weeks gestation)

Footnotes: To use: locate the participant's raw score on each predictor axis, draw vertically upward to read the corresponding points, sum the three point values to obtain Total Points, then draw vertically downward from the Total Points axis to read the predicted probability of persistent comorbid trajectory membership. Binary risk stratification: predicted probability < 31.6% = low risk; ≥31.6% = high risk).

Discussion

The present study tracked the co-development of depression, anxiety, and sleep disturbance throughout pregnancy and found that these symptoms cluster into stable, interrelated patterns that can be identified as early as the first trimester. For a substantial minority of women, the constellation of elevated depression, anxiety, and poor sleep persists across gestation; for the remainder, symptoms remain low and do not impair functioning. This finding aligns with accumulating evidence that perinatal psychological symptoms often co-occur and that their joint course is more informative than any single symptom trajectory considered alone [25]. Our observation that symptom configurations stabilise early—largely by the second trimester—suggests that the first antenatal visit may represent a critical, and perhaps narrow, window for risk identification and preventive intervention.

To address the current inadequate understanding of key determinants of triad comorbidity trajectories, this study employed a rigorous two-step evidence synthesis framework combining all-relevant feature selection (Boruta), Bonferroni-corrected significance testing, and bootstrap optimism-corrected discrimination indices. Among a broad set of candidate predictors, the most informative indicators of subsequent comorbidity risk were the woman's own reports of anxiety, depression, and sleep quality at the first antenatal visit. More distal variables such as demographic characteristics and routine obstetric factors showed comparatively weaker associations with long-term outcome. This does not imply that social context is irrelevant; rather, it suggests that self-reported psychological and sleep states measured at 11–13 weeks gestation may function as efficient, proximal summaries of the cumulative influence of more distant risk factors. In practical terms, a brief assessment of mood and sleep administered at the first encounter appears to capture most of the prognostic information needed for early risk stratification.

This finding carries both methodological and clinical significance. Methodologically, the two-step framework differs from traditional variable selection approaches that rely solely on statistical significance. By subjecting predictors to complementary algorithmic and inferential scrutiny, we demonstrate that three self-report indicators can achieve discriminative capacity comparable to far more complex assessment batteries. Clinically, this proximal-to-distal hierarchical relationship means that frontline health workers can identify high-risk pregnant women without awaiting detailed characterization or specialist referral; the requisite triage information is already available at the first clinical contact. This transformation shifts early pregnancy assessment from passive information gathering to an active, immediate risk-stratification opportunity.

The inclusion of sleep quality as an equally important contributing factor warrants particular emphasis. Historically, sleep disturbance during pregnancy has often been regarded as a secondary concomitant of mood symptoms rather than an independent clinical target. Our findings suggest that this framework is incomplete. Sleep quality predicts long-term comorbidity trajectories even after controlling for depression and anxiety severity, supporting the conceptualization of sleep disturbance as an independent and bidirectionally associated dimension of perinatal psychological distress[26]. Previous studies have recommended routine insomnia screening in the assessment of mental health disorders[9]. Therefore, neglecting the role of sleep in perinatal mental health assessment may overlook a modifiable pathway that sustains psychological distress throughout gestation.

This observation serves as a conceptual bridge between the trajectory findings and predictive contributions of the present study. The necessity of including sleep as an equal dimension—demonstrated by its independent predictive validity—simultaneously reinforces the value of joint trajectory modeling and explains why prediction tools limited to mood symptoms may underperform. Integrating sleep assessment with mood screening represents not merely an incremental improvement, but a necessary step toward a transdiagnostic, multi-domain assessment paradigm. Given that sleep-focused interventions during pregnancy may yield long-term benefits for both sleep and depressive symptoms[27], this reframing carries particularly urgent practical implications.

To directly address the current scarcity of efficient risk prediction techniques for comorbid trajectories, this study developed a parsimonious three-variable nomogram with binary risk categories that enables bedside risk calculation without specialized equipment or software. This simple three-factor assessment can distinguish women who will experience divergent symptom trajectories, raising a practical question: how should this information be used? Current international guidelines recommend universal mental health screening in early pregnancy[28, 29], yet implementation remains uneven, particularly in resource-limited settings.

The nomogram responds to this implementation gap by lowering the threshold for integrating psychological assessment into routine antenatal care. Unlike complex multivariable models requiring specialized software, this tool can be calculated at the bedside and classify pregnant women into distinct risk categories. High-risk women can be prioritized for closer monitoring, while low-risk women proceed with standard care. This stratified approach is more feasible than universal intensive follow-up and directs limited mental health resources toward those who need them most. It represents a shift from merely documenting risk to enabling risk-responsive care allocation—a translational step from statistical prediction to clinically feasible workflow.

It is important to emphasize what this study did not establish. We did not evaluate intervention effectiveness; the superiority of combined mood and sleep interventions over single-domain interventions remains a plausible hypothesis rather than an established fact. Randomized controlled trials of insomnia treatment during pregnancy have shown promising effects[27], yet whether risk-stratified interventions improve long-term outcomes remains to be validated. Moreover, the findings of this study cannot be interpreted causally: the observed predictive associations do not prove that early sleep disturbance causes later depression or anxiety, or vice versa. The directionality of these relationships and the potential role of shared biological or psychological antecedents await clarification by dedicated mechanistic studies.

Limitations

This study has several methodological limitations. First, all measurements were self-report questionnaires; sleep assessment was based on subjective reports rather than objective measures such as actigraphy or polysomnography. Second, the study was conducted in western China, and generalization of the findings to racially diverse populations requires validation. Third, attrition may introduce selection bias: if women with more severe symptoms were more likely to withdraw, the prevalence of persistent comorbidity may be underestimated. Fourth, the prediction model was developed

and preliminarily validated within the same multicenter cohort; fully independent external validation would enhance confidence in its transportability. Fifth, we did not evaluate implementation outcomes such as acceptability, time burden, or impact on referral patterns—factors that determine whether a prediction tool can be translated into sustainable clinical practice.

These limitations point to clear directions for future research. First, independent validation of the three-factor risk model in diverse populations is essential. Second, the incremental value of adding psychosocial risk indicators to the core mood-sleep triad should be examined, although our findings suggest diminishing returns once direct symptom reports are available. Third, objective sleep assessment methods should be employed to clarify whether subjective and objective measures provide complementary or redundant prognostic information. Fourth, intervention studies are urgently needed to determine whether high-risk women benefit from targeted preventive interventions—particularly those addressing both mood and sleep—and whether risk-stratified approaches are cost-effective.

Conclusions

In summary, this study demonstrates that depression, anxiety, and sleep disturbance during pregnancy follow stable and interrelated trajectories that can be identified as early as the first trimester. A brief assessment of these three domains at the first antenatal visit can accurately stratify women by risk of persistent comorbidity, providing a pragmatic foundation for early, targeted prevention. Incorporating sleep into routine perinatal mental health assessment, combined with joint trajectory methods and parsimonious risk-stratification tools, represents not merely an incremental improvement, but a necessary step toward a fundamental paradigm shift—from isolated single-symptom screening to an integrated, multi-dimensional approach to comorbidity risk identification and response during pregnancy.

Abbreviations

AIC
Akaike Information Criterion
aBIC
Sample-size adjusted Bayesian Information Criterion
AUC
Area under the receiver operating characteristic curve
AUPRC
Area under the precision-recall curve
BIC
Bayesian Information Criterion
CI
Confidence interval
HPA
Hypothalamic-pituitary-adrenal
LMR-LRT
Lo-Mendell-Rubin likelihood ratio test
LR
Logistic regression
ML
Machine learning
NN
Neural network
OHQ-SF

Oxford Happiness Questionnaire-Short Form
OR
Odds ratio
PP-LCGM
Parallel process latent class growth modeling
PSQI
Pittsburgh Sleep Quality Index
RF
Random forest
SAS
Self-Rating Anxiety Scale
SDS
Self-Rating Depression Scale
SHAP
SHapley Additive exPlanations
SMOTE
Synthetic Minority Over-sampling Technique
IQR
Interquartile range
EPV
Events per variable
PPV
Positive predictive value
NPV
Negative predictive value
CV
Cross-validation

Declarations

Ethics Approval and Consent to Participate

This study was approved by the Ethics Committee of Xinjiang Medical University (Approval Number: XJYKDXR20230303010). Written informed consent was obtained from all participants prior to enrollment. All participants were informed of their right to withdraw at any time. Their personal information was used only for this study and will remain strictly confidential.

Consent for Publication

Not applicable. This manuscript does not contain any individual person's data in any form (including individual details, images, or videos) that require consent for publication.

Availability of Data and Materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request. Due to the sensitive nature of mental health data in pregnant women and local ethical regulations, individual-level data cannot be publicly deposited. Aggregated data and analysis code are available upon reasonable request to the corresponding author.

Competing Interests

The authors declare that they have no competing interests.

Funding

This research was supported by the Natural Science Foundation of Xinjiang Uygur Autonomous Region, China (Project number: 2022D01F32), and the Innovation and Entrepreneurship Training Program for college students in Xinjiang Uygur Autonomous Region, China (Project number: S202410760140).

Authors' Contributions

Meng Zhao: Investigation, Writing — first draft, Visualization. Zhixiong Li: Investigation, Conceptualization. Zhiyue Zhang: Formal analysis. Qiqi Yang: Data curation. JiaYing Tao: Software. Nan Zhang: Writing — review & editing. Shuping You: Supervision. Mei Sun: Supervision. SiYuan Tang: Supervision. All authors have read and approved the final version of the manuscript and accept full responsibility for its content.

Acknowledgements

The authors would like to thank Karamay Maternal and Child Health Hospital and the community nursing workers for their strong support in data collection.

References

1. Mislu E, Kumsa H, Tadesse S, Arage MW, Susu B, Ayele M, Chane F. Sleep quality disparities in different pregnancy trimesters in low- and middle-income countries: a systematic review and meta-analysis. BMC Pregnancy Childbirth. 2024;24(1):627. 10.1186/s12884-024-06830-3. . PMID: 39354392; PMCID: PMC11446071.
2. Hagatulah N, Bränn E, Oberg AS, Valdimarsdóttir UA, Shen Q, Lu D. Perinatal depression and risk of mortality: nationwide, register based study in Sweden. BMJ. 2024;384:e075462. 10.1136/bmj-2023-075462. . PMID: 38199643; PMCID: PMC10777893.
3. Rogers A, Obst S, Teague SJ, Rossen L, Spry EA, Macdonald JA, Sunderland M, Olsson CA, Youssef G, Hutchinson D. Association Between Maternal Perinatal Depression and Anxiety and Child and Adolescent Development: A Meta-analysis. JAMA Pediatr. 2020;174(11):1082–92. PMID: 32926075; PMCID: PMC7490743.
4. Zhu J, Zheng T, Jin H, Wei M, Yu J, Ni J, Sun K, Zhang J. Early Life Plan Project. Sleep Disturbances in Early Gestation and the Risks of Hypertensive Disorders of Pregnancy: A Prospective Cohort Study. Am J Epidemiol. 2024;193(3):479–488. 10.1093/aje/kwad223. PMID: 37968336.
5. Miller ES, Saade GR, Simhan HN, Monk C, Haas DM, Silver RM, Mercer BM, Parry S, Wing DA, Reddy UM, Grobman WA. Trajectories of antenatal depression and adverse pregnancy outcomes. Am J Obstet Gynecol. 2022;226(1):108.e1-108.e9. doi: 10.1016/j.ajog.2021.07.007. Epub 2021 Jul 17. PMID: 34280383; PMCID: PMC8748269.
6. Astbury L, Pinnington DM, Milgrom J, Bei B. The longitudinal trajectory of depression and anxiety across the perinatal period. J Affect Disord. 2025;370:1–8. 10.1016/j.jad.2024.10.080. . Epub 2024 Oct 24. PMID: 39461375.
7. Roddy Mitchell A, Gordon H, Lindquist A, Walker SP, Homer CSE, Middleton A, Cluver CA, Tong S, Hastie R. Prevalence of Perinatal Depression in Low- and Middle-Income Countries: A Systematic Review and Meta-analysis. JAMA Psychiatry. 2023;80(5):425–31. PMID: 36884232; PMCID: PMC9996459.
8. Fawcett EJ, Fairbrother N, Cox ML, White IR, Fawcett JM. The Prevalence of Anxiety Disorders During Pregnancy and the Postpartum Period: A Multivariate Bayesian Meta-Analysis. J Clin Psychiatry. 2019;80(4):18r12527. 10.4088/JCP.18r12527. . PMID: 31347796; PMCID: PMC6839961.

9. Song Z, Zhang D, Yang R, Zhu P, Liu Y, Wang L, Wang L, Zheng R. Prenatal comorbid depression and anxiety based on a national longitudinal study in Chinese mainland: Relationship with adverse perinatal outcomes. J Affect Disord. 2025;385:119396. Epub 2025 May 13. PMID: 40374091.
10. Wang Y, Liu H, Zhang C, Li C, Xu JJ, Duan CC, Chen L, Liu ZW, Jin L, Lin XH, Zhang CJ, Zhang HQ, Yu JL, Li T, Dennis CL, Li H, Wu YT. Antepartum sleep quality, mental status, and postpartum depressive symptoms: a mediation analysis. BMC Psychiatry. 2022;22(1):521. 10.1186/s12888-022-04164-y. . PMID: 35918689; PMCID: PMC9344627.
11. Gao M, Hu J, Yang L, Ding N, Wei X, Li L, et al. Association of sleep quality during pregnancy with stress and depression: A prospective birth cohort study in China. BMC Pregnancy Childbirth. 2019;19:444. 10.1186/s12884-019-2583-1.
12. Huo L, Yang L, Song Z, Zhang Y, Huang X, Yang Y, Liu X, Zheng R, Zhang D. Trajectories of perinatal sleep quality from early pregnancy to six weeks postpartum and their risk factors: A multicenter longitudinal study in China. J Affect Disord. 2026;403:121289. Epub 2026 Jan 27. PMID: 41610893.
13. Gueron-Sela N, Shahar G, Volkovich E, Tikotzky L. Prenatal maternal sleep and trajectories of postpartum depression and anxiety symptoms. J Sleep Res. 2021;30(4):e13258. 10.1111/jsr.13258. . Epub 2020 Dec 15. PMID: 33321551.
14. Fu T, Wang C, Yan J, Zeng Q, Ma C. Relationship between antenatal sleep quality and depression in perinatal women: A comprehensive meta-analysis of observational studies. J Affect Disord. 2023;327:38–45. doi: 10.1016/j.jad.2023.01.125. Epub 2023 Feb 3. PMID: 36739002.
15. Peng C, Wang K, Wang J, Wassing R, Eickhoff SB, Tahmasian M, Chen J. Neural correlates of insomnia with depression and anxiety from a neuroimaging perspective: A systematic review. Sleep Med Rev. 2025;81:102093. 10.1016/j.smrv.2025.102093. . Epub 2025 May 2. PMID: 40349510.
16. Jiang M, Zheng H, Bao Z, Wu Z, Zheng X, Feng Y. Joint Developmental Trajectories of Perinatal Depression and Anxiety and Their Predictors: A Longitudinal Study. Healthc (Basel). 2025;13(11):1251. 10.3390/healthcare13111251. . PMID: 40508867; PMCID: PMC12154182.
17. Huang Y, Nie S, Yang P, Pei Y, Dong M, Cui N, Zhang X, Cao F. Exploring the developmental changes in and the relationship between maternal-offspring attachment and depression in perinatal women: A longitudinal study. J Affect Disord. 2026;399:121118. Epub 2026 Jan 2. PMID: 41485506.
18. Pan X, Chen Y, Chen C, Chen J, Wang J, Chen Y, Zhang W, Wu J, Liu W, Zou Z, Zhu L, Chen X. Dual trajectory of insomnia and depressive symptoms in women from early pregnancy to 6 months postpartum: a prospective cohort study. BMC Pregnancy Childbirth. 2025;25(1):582. 10.1186/s12884-025-07649-2. . PMID: 40382558; PMCID: PMC12085819.
19. Jiang M, Zheng H, Bao Z, Wu Z, Zheng X, Feng Y. Joint Developmental Trajectories of Perinatal Depression and Anxiety and Their Predictors: A Longitudinal Study. Healthc (Basel). 2025;13(11):1251. 10.3390/healthcare13111251. . PMID: 40508867; PMCID: PMC12154182.
20. ZUNG WW, A SELF-RATING DEPRESSION, SCALE. Arch Gen Psychiatry. 1965;12:63–70. 10.1001/archpsyc.1965.01720310065008. PMID: 14221692.
21. Zung WW. A rating instrument for anxiety disorders. Psychosomatics. 1971 Nov-Dec;12(6):371-9. 10.1016/S0033-3182(71)71479-0. PMID: 5172928.
22. Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989;28(2):193–213. 10.1016/0165-1781(89)90047-4. PMID: 2748771.
23. Hills P, Argyle M. The Oxford Happiness Questionnaire: a compact scale for the measurement of psychological well-being. Pers Individ Dif. 2002;33(7):1073–82. 10.1016/S0191-8869(01)00213-6.
24. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361–87. 10.1002/(SICI)1097-

0258(19960229)15:4%3C361::AID-SIM168%3E3.0.CO;2-4.

25. Falah-Hassani K, Shiri R, Dennis CL. The prevalence of antenatal and postnatal co-morbid anxiety and depression: a meta-analysis. Psychol Med. 2017;47(12):2041–53. Epub 2017 Apr 17. PMID: 28414017.
26. Freeman D, Sheaves B, Waite F, Harvey AG, Harrison PJ. Sleep disturbance and psychiatric disorders. Lancet Psychiatry. 2020;7(7):628–637. 10.1016/S2215-0366(20)30136-X. PMID: 32563308.
27. Felder JN, Epel ES, Neuhaus J, Krystal AD, Prather AA. Efficacy of Digital Cognitive Behavioral Therapy for the Treatment of Insomnia Symptoms Among Pregnant Women: A Randomized Clinical Trial. JAMA Psychiatry. 2020;77(5):484–492. doi: 10.1001/jamapsychiatry.2019.4491. Erratum in: JAMA Psychiatry. 2020;77(7):768. 10.1001/jamapsychiatry.2020.0724. PMID: 31968068; PMCID: PMC6990703.
28. National Institute for Health and Care Excellence. Antenatal and Postnatal Mental Health: Clinical Management and Service Guidance (Clinical Guideline CG192). London: NICE. 2014. Available from: https://www.nice.org.uk/guidance/cg192
29. American College of Obstetricians and Gynecologists. ACOG Committee Opinion No. 757: Screening for perinatal depression. Obstet Gynecol. 2018;132(5):e208-e212. 2018;132(5):e208-e212.

Figures

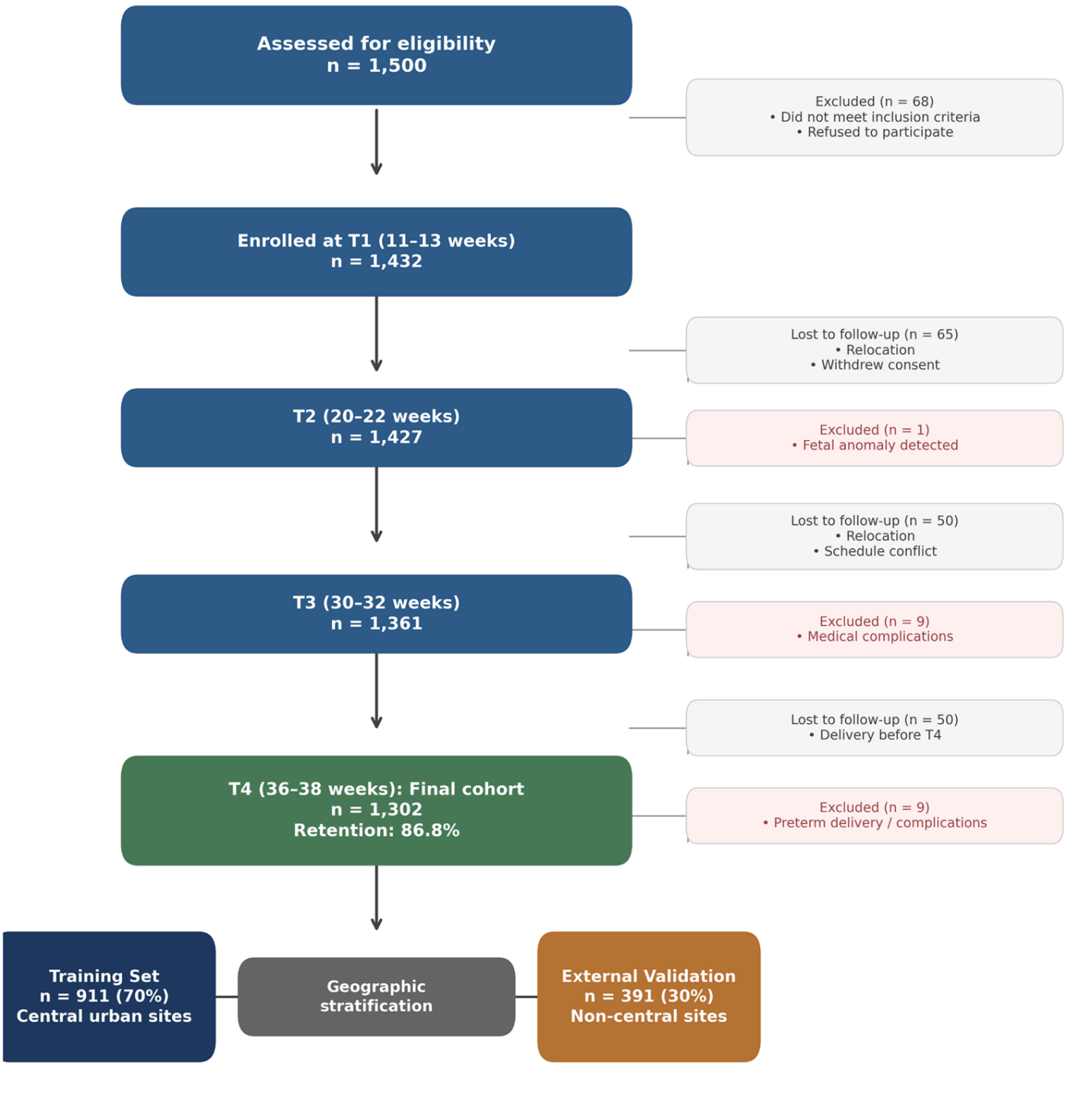


Figure 1

Participant enrollment and train/test split

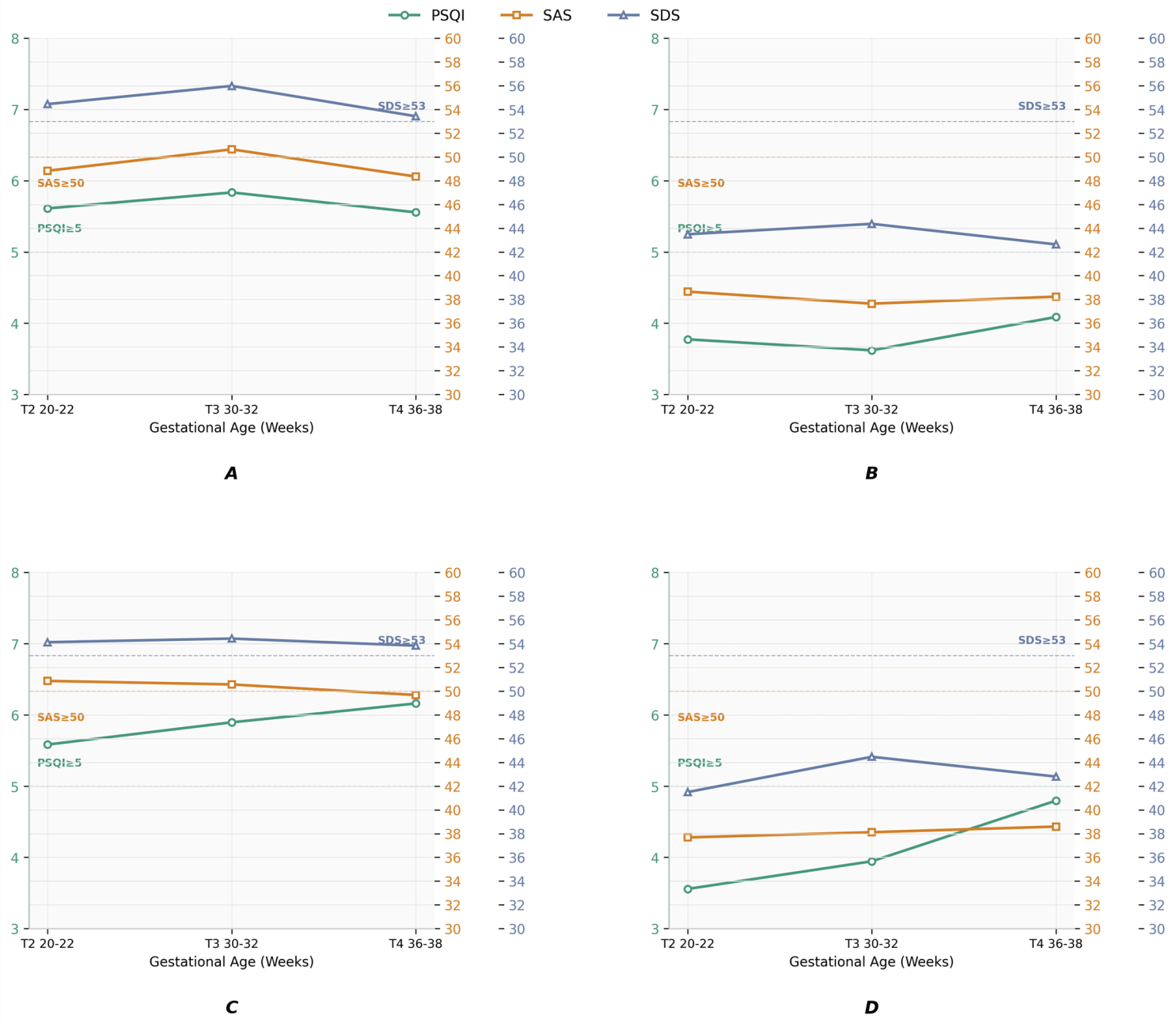


Figure 2

Estimated mean trajectories of comorbid depression-anxiety-sleep disturbance from T2 to T4.

Legend: Trajectories were derived from the parallel process latent class growth model (PP-LCGM) and represent class-specific estimated means. (A) Training set, comorbid group. (B) Training set, non-comorbid group. (C) Testing set, comorbid group (n=136). (D) Testing set, non-comorbid group (n=2). Dashed horizontal lines denote clinical cutoffs: PSQI ≥ 5 (green), SAS ≥ 50 (orange), and SDS ≥ 53 (blue). SDS, Self-Rating Depression Scale; SAS, Self-Rating Anxiety Scale; PSQI, Pittsburgh Sleep Quality Index.

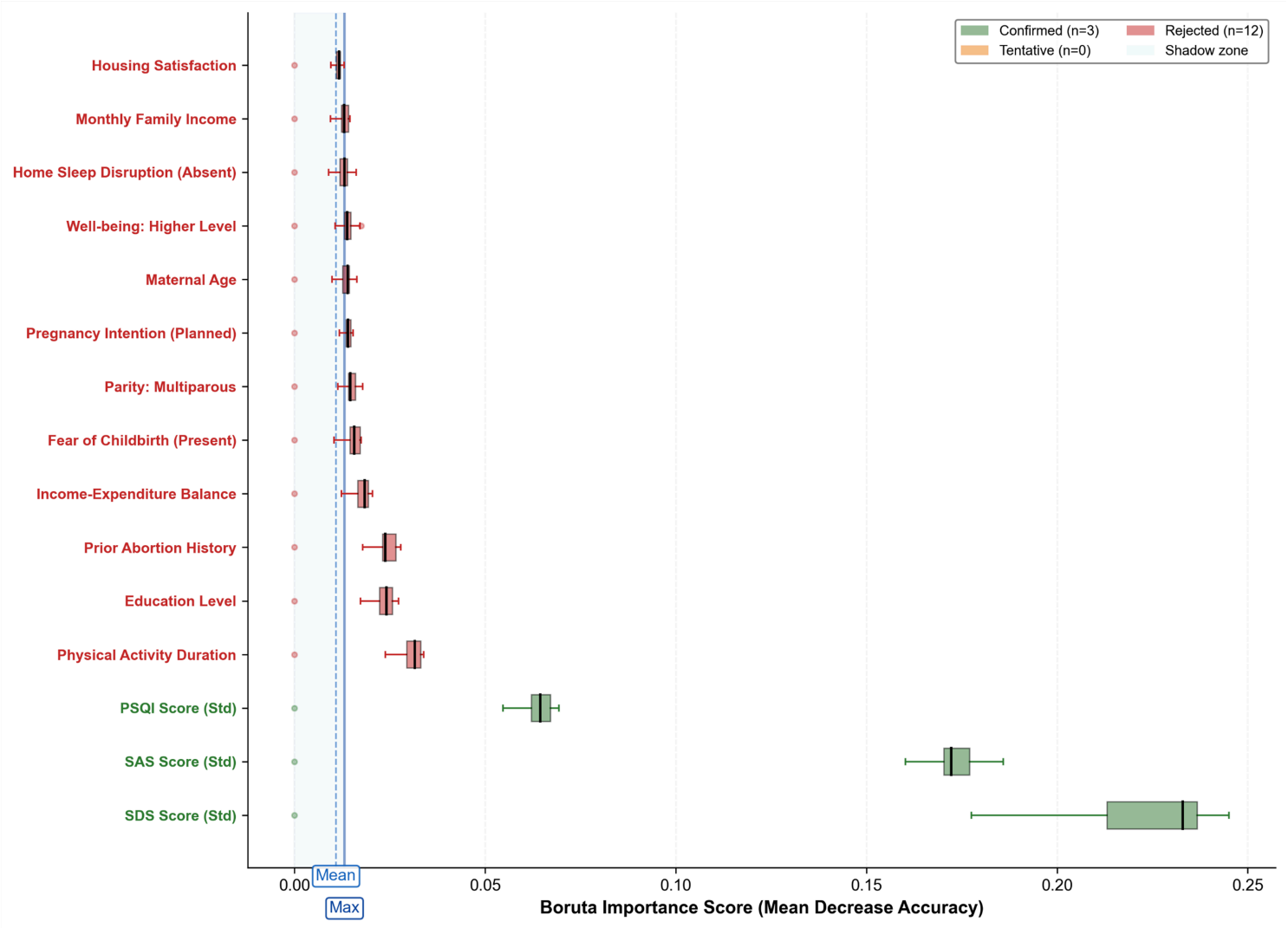


Figure 3

Boruta feature selection results for 18 candidate baseline predictors.

Footnotes: Boruta algorithm (n_estimators=200, max_depth=5, 200 iterations) compared each candidate predictor against randomized shadow features. Green bars indicate Confirmed predictors (SAS, SDS, PSQI) that consistently outperformed the maximum shadow feature (Shadow Max, blue dashed line). Red bars indicate Rejected predictors (n=12) that failed to exceed the shadow mean threshold. No predictors were classified as Tentative. The vertical dashed line represents the Shadow Mean threshold; features to the right of Shadow Max are confirmed as having genuine predictive importance beyond random noise.

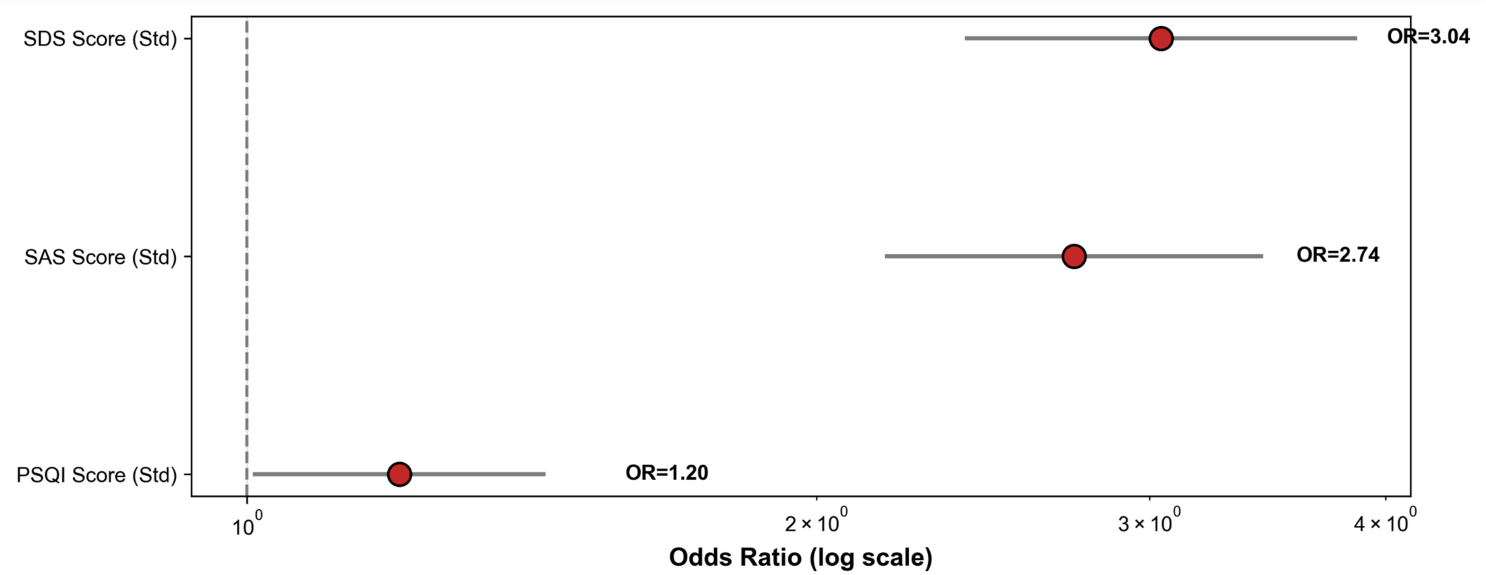


Figure 4

Forest plot of the Bonferroni-corrected multivariable regression results.

Footnotes: Odds ratios (OR) with 95% confidence intervals for the 18 candidate predictors tested in the multivariable logistic regression model after Bonferroni correction (adjusted significance threshold α = 0.0028). The dashed vertical line at OR = 1 represents no effect. Predictors with 95% CIs excluding 1 and passing the Bonferroni threshold are considered statistically significant. SAS = Self-Rating Anxiety Scale; SDS = Self-Rating Depression Scale; PSQI = Pittsburgh Sleep Quality Index.

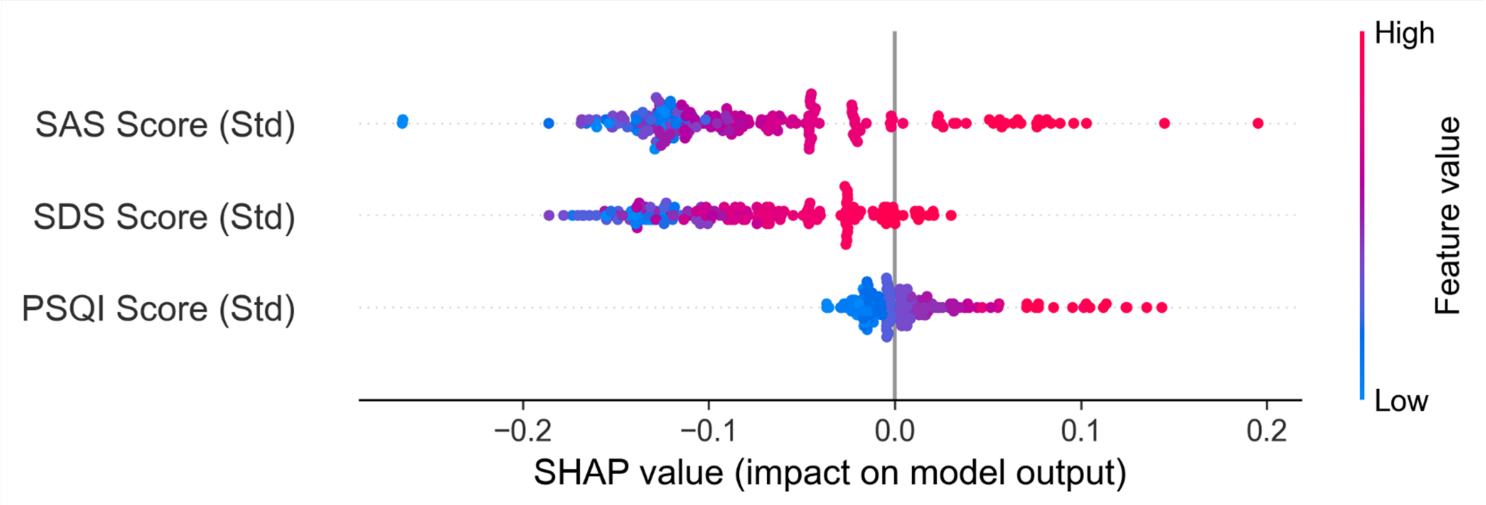


Figure 5

SHAP (SHapley Additive exPlanations) interpretability analysis for the 3-factor logistic regression model on external validation (n=391)

Footnotes: SHAP summary dot plot showing the distribution of SHAP values for each predictor. Each dot represents one participant; dot color indicates the feature value (pink=high, blue=low), and horizontal position indicates the SHAP value (impact on model output). SAS and SDS show the widest SHAP value distributions, confirming their dominant predictive contributions. PSQI demonstrates a distinct independent incremental effect with a narrower but consistently positive

SHAP distribution. The monotonic positive relationships support the clinical interpretation that elevated baseline symptoms in any domain increase the likelihood of persistent comorbidity.

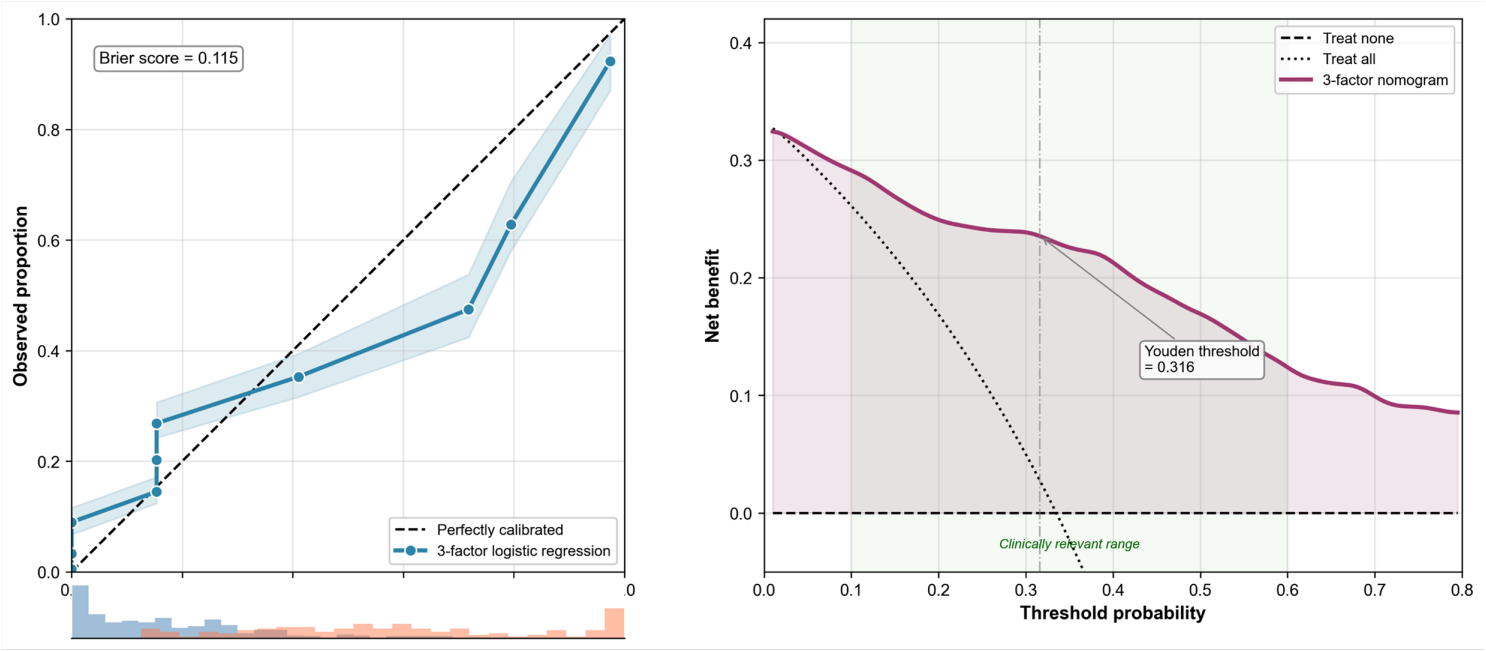


Figure 6

Model performance evaluation on external validation set (n = 391)

Footnotes: (A) Calibration curve comparing predicted probabilities against observed proportions. The solid line represents the 3-factor logistic regression model with bootstrap 95% confidence interval. The dashed diagonal represents perfect calibration. Brier score=0.109 indicates excellent calibration. The distribution histograms at the bottom show the density of predicted probabilities for non-comorbid and comorbid participants. (B) Decision curve analysis comparing the net benefit of the 3-factor nomogram against treat-none and treat-all strategies across threshold probabilities from 0% to 80%. The nomogram demonstrates positive net benefit across the clinically relevant range. The vertical dash-dot line marks the Youden-optimized threshold.

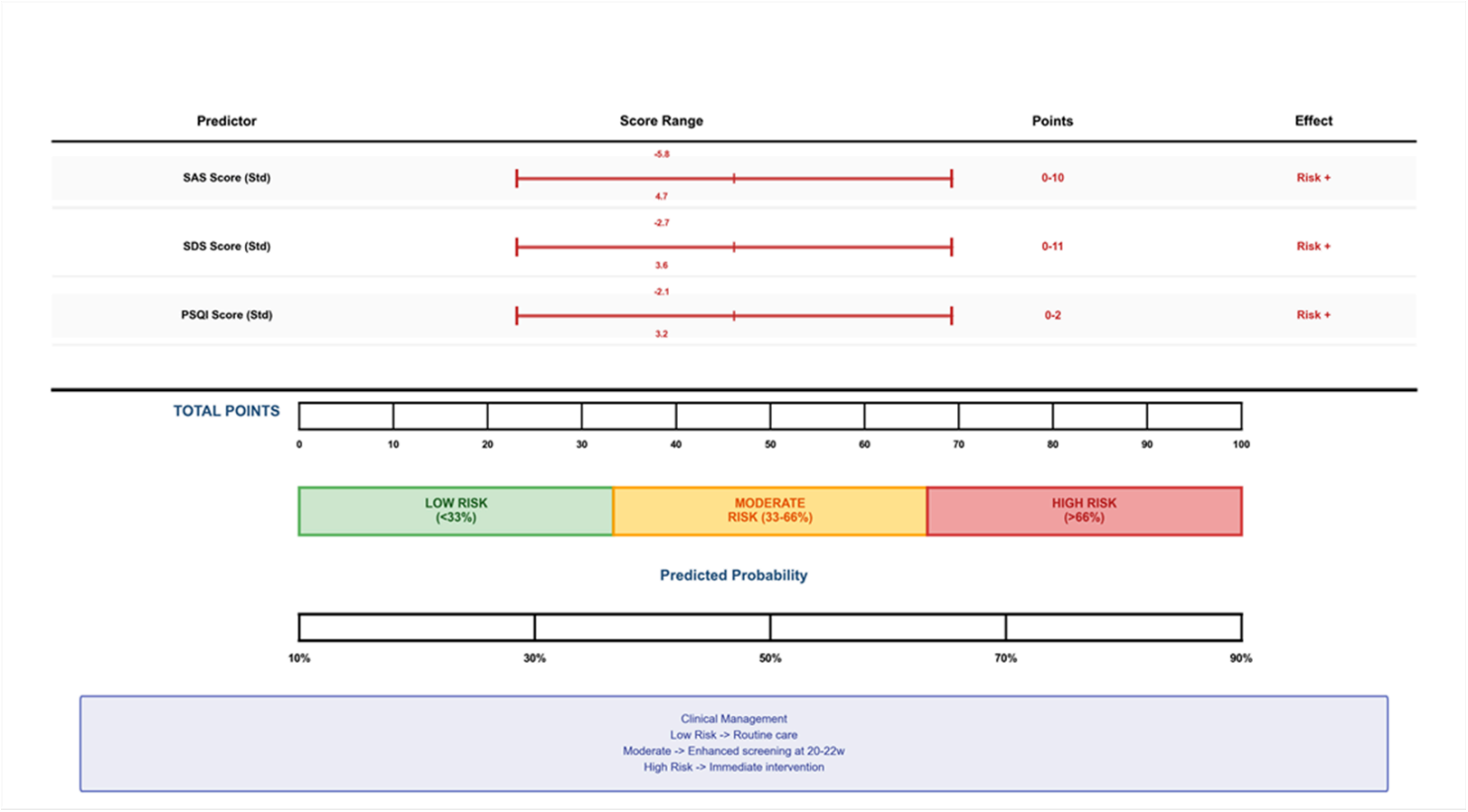


Figure 7

Clinical nomogram for predicting antenatal depression-anxiety-sleep comorbidity trajectory membership using raw scale scores at the first antenatal visit (11–13 weeks gestation)

Footnotes: To use: locate the participant's raw score on each predictor axis, draw vertically upward to read the corresponding points, sum the three point values to obtain Total Points, then draw vertically downward from the Total Points axis to read the predicted probability of persistent comorbid trajectory membership. Binary risk stratification: predicted probability <31.6% = low risk; ≥31.6% = high risk).