Wang et al. BMC Pregnancy and Childbirth (2026) 26:452
https://doi.org/10.1186/s12884-026-08939-z

BMC Pregnancy and Childbirth

RESEARCH

Open Access

Integrating socioeconomic determinants into preterm birth risk prediction: evidence from Northwest China

Xiaoning Wang[1], Qin Yuan[1†] and Yongmei Yang[2*†]

Abstract

Background Preterm birth is a leading cause of neonatal mortality worldwide. Socioeconomic status (SES) is associated with preterm birth risk, yet existing studies in low-resource settings have rarely examined whether incorporating SES into clinical models can enhance prediction performance. This study aimed to develop and validate a SES-integrated risk prediction model for preterm birth in Northwest China.

Methods A retrospective cohort of 2,781 deliveries from county-level hospitals (January–November 2024) was analyzed. Logistic regression was used to construct a baseline model (clinical variables) and an extended model incorporating SES indicators (education, occupation, living environment). Predictive performance was evaluated using the area under the ROC curve (AUC) and calibration plots.

Results Including SES and healthcare accessibility variables markedly improved model discrimination (AUC increased from 0.688 to 0.854) and showed good calibration, indicating strong agreement between predicted and observed outcomes.

Conclusion Socioeconomic factors significantly enhance the accuracy and calibration of preterm birth prediction models. Integrating SES indicators into perinatal assessment can improve risk identification and inform equitable maternal health policies.

Keywords Preterm birth, Socioeconomic status (SES), Risk prediction, Logistic regression, Calibration, Maternal health

Introduction

Preterm birth (delivery before 37 weeks of gestation) remains a major global health challenge, with subcategories including extremely preterm (<28 weeks), very preterm (28–32 weeks), and moderate-to-late preterm (32–37 weeks) [1]. In 2019, about 900,000 children under five died from preterm-related complications, and it remained the leading cause of child mortality worldwide in 2022 [1, 2]. Globally, preterm birth affects ~11% of deliveries; in China, it ranges from 5 to 15%, totaling over 13 million infants annually [3]. Maternal age, pregnancy complications, and medical history are key determinants of preterm birth [4, 5]. Accurate risk prediction is critical for improving maternal and neonatal outcomes.

Socioeconomic status (SES)—a composite of education, income, and occupation—strongly influences maternal and child health [6]. Lower SES is consistently associated with poorer outcomes, including reduced antenatal care use [7, 8]. SDG 5 and 10 emphasize reducing social

†Qin Yuan and Yongmei Yang contributed equally to this work.

*Correspondence:
Yongmei Yang
yang8xspeed@hotmail.com
[1]Communication University of China, 100024 Beijing, China
[2]Xuanwu Hospital, Capital Medical University, Beijing 100053, China

and economic disparities to improve access to health resources [9, 10].

Measuring SES in resource-limited settings remains challenging. In Northwest China, agricultural dependence, limited industrial development, and restricted market access result in volatile income structures [11]. In the regions we studied, most participants were farmers with unstable or no fixed income, which made it particularly difficult to reliably collect data on household income or assets. According to the National Bureau of Statistics, the per capita regional GDP in this area was 72,690 RMB in 2024, substantially lower than the national average of 95,749 RMB. Despite national maternal health programs, regional disparities persist, particularly in western China. To address this, we incorporated education, occupation, and residential range into SES assessment, and additionally included indicators related to healthcare access and utilization, such as prenatal visit frequency and gestational age at first record. These variables were included as SES-related factors reflecting differences in access to and use of maternal healthcare services. Clinical indicators, such as high-risk factors and gestational hypertension, further reflect maternal care quality [12–14]. This framework captures local socioeconomic and healthcare conditions to better predict preterm birth risk.

Method

Data were collected from multiple county-level medical institutions in Northwest China between January and November 2024, including demographic characteristics, obstetric history, prenatal monitoring, pregnancy outcomes, and neonatal details. The data were obtained from the regional maternal and child health information cloud platform, which systematically records maternal health services across the area. Due to data privacy requirements, specific hospital names cannot be disclosed. All participating institutions are included according to the region's maternal healthcare service system, providing a representative sample of routine county-level obstetric care. Detailed inclusion and exclusion criteria and data cleaning steps were applied as follows: Pregnant women who delivered between 28 and 36 + 6 weeks of gestation and had established prenatal examination records within the study region were included in the preterm birth group. Only cases with complete prenatal and delivery information uploaded to the regional maternal and child health information system were eligible for retrospective analysis. Women were excluded if the fetus had major congenital malformations, chromosomal abnormalities, or if the pregnancy was terminated due to stillbirth. Records with incomplete information or loss to follow-up—such as missing key clinical indicators or unavailable outcome data—were also removed. The control group followed the same inclusion and exclusion criteria, with the exception that gestational age at delivery was required to be 37 weeks or more.Before analysis, raw data underwent a series of cleaning procedures. Cases missing essential variables, including gestational age, pregnancy outcomes, or critical perinatal indicators, were excluded. Logical checks were applied to identify implausible or inconsistent entries, such as abnormal gestational weight gain or values outside the physiologically plausible gestational age range. Duplicate records were reviewed and consolidated, and only data that met all eligibility criteria and retained complete clinical information were included in the final dataset.

After data cleaning, 2,781 valid cases were included. This study was approved by an ethics committee and obtained informed consent from all participants.Because this study employed an outcome-based design for predictive modeling, the analytic dataset included a higher proportion of preterm births than would be expected in a population-based cohort. This reflects the study's focus on comparing women with preterm and term deliveries for model development and performance evaluation, rather than estimating population-level incidence. Accordingly, the model is primarily intended to assess discriminatory performance between outcomes, and application to population settings may require recalibration using local prevalence data.

To illustrate the sample selection and grouping process, a flowchart is provided (Fig. 1), showing the screened

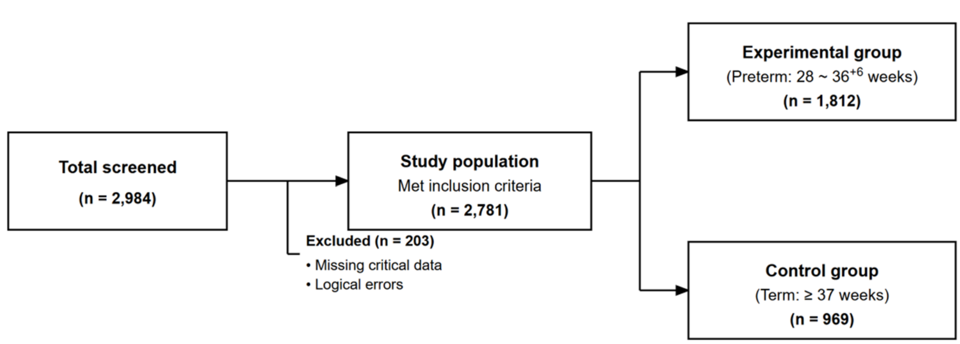

Fig. 1 Flow diagram of study sample selection

population, exclusion criteria, and the final numbers of preterm and term births.

Preterm birth was the dependent variable, and demographic, obstetric, and pregnancy-related indicators were independent variables for logistic regression. All candidate predictors were first included in a full multivariable logistic regression model in the training dataset. A bidirectional stepwise selection procedure based on the Akaike Information Criterion (AIC) was then applied, using the full model as the starting point, allowing variables to enter or be removed without forcing any variables to remain in the model. 80% of the sample was used for model training and 20% for validation [15]. Model performance was evaluated using ROC curves and AUC values [16]. A baseline model with clinical variables was first established, followed by an extended model incorporating SES-related factors. Changes in AUC were used to assess the contribution of SES to preterm birth prediction. Multicollinearity was assessed using variance inflation factors (VIFs). When potential model instability or separation was identified, Firth's penalized logistic regression was applied.

In addition, calibration analysis was performed using bootstrapped calibration plots to evaluate the agreement between predicted and obsrved probabilities, ensuring the model's reliability and goodness of fit [17].

Socioeconomic status (SES)–related factors were incorporated to examine their contribution to preterm birth prediction. SES was assessed using multiple indicators, including residence area [18], education level [19], occupation type [20], number of prenatal visits [21], gestational week at first record [22], high-risk pregnancy [23], gestational hypertension [24], antihypertensive treatment [25], and OGTT test [26] performed (Table 1). In the study region, most pregnant women are from agricultural or low-income households, and individual-level income data are difficult to collect reliably due to substantial missingness and reporting bias. Although high-risk pregnancy, gestational hypertension, antihypertensive treatment, and OGTT completion have clear clinical attributes, in this resource-limited setting these variables were included as SES-related factors reflecting differences in healthcare access, health literacy, continuity of prenatal care, and treatment uptake [27–29], and were therefore incorporated as SES-related predictors in the extended model.

Table 1 Definition and classification of socioeconomic and clinical variables

Variable	Categories / Measurement
Residence area	Urban community; Rural town/village
Education level	Junior high school or below; Technical secondary/high school; College or above
Occupation type	Farmer; Enterprise employee/civil servant; Self-employed/freelancer; Others
Number of prenatal visits	< 8 visits; ≥8 visits
Gestational week at first record (weeks)	Continuous variable (Mean ± SD)
High-risk pregnancy	No; Yes
Gestational hypertension	No; Yes
Antihypertensive treatment	No; Yes
OGTT test performed	No; Yes

Result

Data analyze

As shown in Table 2, the study included 2,781 parturients, of whom 969 were in the control group and 1,812 were in the experimental group. This proportion reflects the composition of the study sample rather than the population-level distribution.

Original clinical variables

The preterm birth group tended to be slightly older than the term group. Most participants were Uyghur. Regarding obstetric history, women in the preterm group had a higher proportion of multiple pregnancies. Pre-pregnancy weight and BMI were higher among preterm births. A history of cesarean section was more common in the preterm group, whereas gestational weight gain was greater in the term group.

SES-related variables

Most participants resided in villages or towns, with a smaller proportion living in urban communities. Women in the preterm group were more likely to have lower educational attainment and were predominantly farmers, whereas other occupations were less represented. Fewer prenatal visits and later gestational age at first record were observed in the preterm group. High-risk pregnancies and hypertensive disorders were more frequent among preterm births, along with a higher rate of antihypertensive treatment. Uptake of oral glucose tolerance testing was generally low, but relatively higher in the term group.

Modeling analysis

Univariate logistic regression analysis

The results of the univariate logistic regression analysis are shown in Table 3, indicating that multiple variables were significantly associated with the risk of preterm birth. With increasing maternal age, the risk of preterm birth rose — for each additional year of age, the risk increased by 1.06 times (95%CI: 1.039–1.076, $P < 0.001$). Significant differences were also observed among ethnic groups: Uyghur women had a 4.23-fold higher risk of preterm birth compared with women of other ethnicities (95%CI: 2.972–6.083, $P < 0.001$), whereas Han women

Table 2 Baseline Demographic and Clinical Characteristics of the Participants

Variable	Total (*N* = 2781)	Control Group (*N* = 969)	Experimental Group (*N* = 1812)
Age (years), Mean ± SD	29.0 ± 5.22	28.1 ± 4.60	29.5 ± 5.46
Ethnicity			
Uyghur	2598 (93.4%)	847 (87.4%)	1751 (96.6%)
Han	171 (6.1%)	114 (11.8%)	57 (3.1%)
Others	12 (0.4%)	8 (0.8%)	4 (0.2%)
Gravidity			
1–2 times	1883 (67.7%)	685 (70.7%)	1198 (66.1%)
>2 times	898 (32.3%)	284 (29.3%)	614 (33.9%)
Pre-pregnancy weight (kg), Mean ± SD	61.3 ± 11.3	59.5 ± 9.65	62.2 ± 12.0
(BMI), Mean ± SD	24.1 ± 4.32	23.0 ± 3.44	24.7 ± 4.62
History of cesarean section			
No	2322 (83.5%)	917 (94.6%)	1405 (77.5%)
Yes	459 (16.5%)	52 (5.4%)	407 (22.5%)
Gestational weight gain (kg), Mean ± SD	13.9 ± 5.53	14.7 ± 5.26	13.5 ± 5.63
Residence area			
Urban community	639 (23.0%)	294 (30.3%)	345 (19.0%)
Rural town/village	2142 (77.0%)	675 (69.7%)	1467 (81.0%)
Education level			
Junior high school or below	1567 (56.3%)	482 (49.7%)	1085 (59.9%)
College or above	597 (21.5%)	253 (26.1%)	344 (19.0%)
Technical secondary/high school	617 (22.2%)	234 (24.1%)	383 (21.1%)
Occupation type			
Farmer	1412 (50.8%)	518 (53.5%)	894 (49.3%)
Enterprise employee/civil servant	496 (17.8%)	129 (13.3%)	367 (20.3%)
Self-employed/freelancer	246 (8.8%)	164 (16.9%)	82 (4.5%)
Others	627 (22.5%)	158 (16.3%)	469 (25.9%)
Number of prenatal visits			
<8 visits	2167 (77.9%)	687 (70.9%)	1480 (81.7%)
≥8 visits	614 (22.1%)	282 (29.1%)	332 (18.3%)
Gestational week at first record (weeks), Mean ± SD	5.84 ± 2.00	6.10 ± 1.88	5.70 ± 2.04
High-risk pregnancy			
No	188 (6.8%)	182 (18.8%)	6 (0.3%)
Yes	2593 (93.2%)	787 (81.2%)	1806 (99.7%)
Gestational hypertension			
No	1919 (69.0%)	846 (87.3%)	1073 (59.2%)
Yes	862 (31.0%)	123 (12.7%)	739 (40.8%)
Antihypertensive treatment			
No	1856 (66.7%)	762 (78.6%)	1094 (60.4%)
Yes	925 (33.3%)	207 (21.4%)	718 (39.6%)
OGTT test performed			
No	2620 (94.2%)	849 (87.6%)	1771 (97.7%)
Yes	161 (5.8%)	120 (12.4%)	41 (2.3%)

showed a markedly lower risk (OR = 0.232, 95%CI: 0.160–0.333, $P < 0.001$).

Women with more than two pregnancies had a 1.25-fold higher risk (95%CI: 1.038–1.515, $P = 0.020$).Regarding physical indicators, both pre-pregnancy weight (OR = 1.027, 95%CI: 1.019–1.036, $P < 0.001$) and BMI (OR = 1.119, 95%CI: 1.093–1.146, $P < 0.001$) were positively associated with preterm birth risk, while greater gestational weight gain was linked to a lower risk (OR = 0.965, 95%CI: 0.950–0.980, $P < 0.001$).

Women with a history of cesarean section had a 4.79-fold higher risk than those without (95%CI: 3.473–6.754, $P < 0.001$).

In terms of socioeconomic factors, women living in rural areas had a 1.77-fold higher risk than those in urban communities (95%CI: 1.451–2.166, $P < 0.001$).

Table 3 Univariate Logistic Regression Analysis of Risk Factors for Preterm Birth

Characteristics	Sub groups	β (SE)	OR	%95CI	*P* value
Age	Age	0.055 (0.009)	1.057	1.039–1.076	$p<0.001$
Ethnicity	Other ethnicities				
	Uyghur	1.441 (0.182)	4.225	2.972–6.083	$p<0.001$
	Han	-1.460 (0.187)	0.232	0.160–0.333	$p<0.001$
Parity	1–2 times				
	> 2 times	0.225 (0.096)	1.253	1.038–1.515	0.020
Pre-pregnancy weight (kg)	Pre-pregnancy weight	0.027 (0.004)	1.027	1.019–1.036	$p<0.001$
BMI	BMI	0.112 (0.012)	1.119	1.093–1.146	$p<0.001$
History of Cesarean Section	No				
	Yes	1.566 (0.169)	4.786	3.473–6.754	$p<0.001$
Gestational weight gain (kg)	Gestational weight gain	-0.036 (0.008)	0.965	0.950–0.980	$p<0.001$
Residence	Community				
	Rural	0.573 (0.102)	1.773	1.451–2.166	$p<0.001$
Education level	College or above				
	Junior high school or below	-0.129 (0.107)	0.879	0.714–1.084	0.224
	Secondary or high school	0.359 (0.090)	1.431	1.201–1.706	$p<0.001$
Occupation	Others				
	Farmer	-0.195 (0.089)	0.823	0.690–0.980	0.028
	Enterprise staff or government official	0.459 (0.124)	1.583	1.245–2.025	$p<0.001$
	Self-employed or freelancer	-1.404 (0.162)	0.246	0.178–0.336	$p<0.001$
Number of antenatal visits	< 8 visits				
	≥ 8 visits	-0.643 (0.104)	0.525	0.429–0.644	$p<0.001$
Gestational week at first record	Gestational week	-0.099 (0.023)	0.906	0.866–0.947	$p<0.001$
High-risk pregnancy factors	No				
	Yes	4.784 (0.585)	119.53	45.159–485.416	$p<0.001$
Hypertensive disorders during pregnancy	No				
	Yes	1.491 (0.119)	4.442	3.532–5.633	$p<0.001$
Antihypertensive treatment	No				
	Yes	0.830 (0.102)	2.294	1.882–2.808	$p<0.001$
OGTT test conducted	No				
	Yes	-1.772 (0.205)	0.170	0.112–0.251	$p<0.001$

Those with secondary or high school education had a 1.43-fold higher risk (95%CI: 1.201–1.706, $P<0.001$), whereas the difference for women with lower education levels was not statistically significant.As for occupation, women working in government or enterprises had a 1.58-fold higher risk (95% CI: 1.245–2.025, $P<0.001$), while self-employed or freelance workers showed a significantly lower risk (OR = 0.246, 95%CI: 0.178–0.336, $P<0.001$).Antenatal care behaviors and pregnancy monitoring also had notable effects: women with ≥ 8 prenatal visits had a 47.5% lower risk (OR = 0.525, 95%CI: 0.429–0.644, $P<0.001$).An earlier gestational week at first record was associated with higher risk, while later registration reduced risk (OR = 0.906, 95%CI: 0.866–0.947, $P<0.001$).Women with high-risk pregnancy factors had an extremely elevated risk (OR = 119.53, 95%CI: 45.159–485.416, $P<0.001$).Those with hypertensive disorders during pregnancy had a 4.44-fold higher risk (95% CI: 3.532–5.633, $P<0.001$), and those receiving antihypertensive treatment had a 2.29-fold higher risk (95%CI: 1.882–2.808, $P<0.001$).Conversely, women who underwent OGTT testing had a significantly lower risk of preterm birth (OR = 0.170, 95%CI: 0.112–0.251, $P<0.001$).

To assess potential multicollinearity before multivariable modeling, we calculated variance inflation factors (VIFs) for all candidate predictors (threshold: VIF < 5). Based on the results, the ethnicity variable showed a VIF greater than 5 (Uyghur: 14.723, Han: 14.477) and demonstrated strong collinearity with blood pressure–related indicators. Given that ethnicity in this study population primarily reflects regional population structure and overlaps conceptually with residence and healthcare access, this variable was removed from the final model to avoid instability. Second, because gestational hypertension and antihypertensive treatment are inherently linked within the same clinical process, these two indicators were combined into a single composite variable with three levels: "no hypertension," "hypertension untreated," and "hypertension treated." This restructuring captured both disease status and management while reducing redundancy. All remaining variables showed acceptable VIF values and were retained in the final model.

Multivariate logistic regression analysis

Model with clinical variables only In this study, a logistic regression method was used to construct a predictive model for preterm birth. First, clinical variables including age, parity, pre-pregnancy weight, body mass index (BMI), history of cesarean section, and gestational weight gain were entered as independent variables to build the baseline model. The dataset was randomly divided into a training set and a validation set in an 8:2 ratio.

The results showed that each one-year increase in maternal age was associated with a 3.9% higher risk of preterm birth (OR = 1.039, 95%CI: 1.017–1.062, $P < 0.001$). Women with parity > 2 had a significantly lower risk (OR = 0.662, 95%CI: 0.520–0.842, $P < 0.001$). Higher pre-pregnancy weight decreased the risk (OR = 0.967, 95%CI: 0.950–0.983, $P < 0.001$), whereas each one-unit increase in BMI increased the risk by 18.2% (OR = 1.182, 95%CI: 1.128–1.240, $P < 0.001$). Women with a history of cesarean section had nearly a four-fold higher risk compared with those without such history (OR = 3.983, 95%CI: 2.850–5.689, $P < 0.001$). In contrast, greater gestational weight gain was associated with a lower risk (OR = 0.973, 95%CI: 0.957–0.989, $P = 0.001$).

These results are summarized in Table 4.

Model with clinical + SES variables After incorporating socioeconomic variables, the multivariate logistic regression analysis was further refined using Firth's penalized likelihood method to address small-sample bias and separation issues. The Firth-adjusted results of the extended model are presented in Table 5. The findings indicate that both clinical and socioeconomic characteristics significantly contributed to the risk of preterm birth.

Among clinical factors, a higher body mass index (BMI) remained a significant risk factor (OR = 1.100, 95%CI: 1.039–1.164, $P < 0.001$). In contrast, greater gestational weight gain (OR = 0.952, 95%CI: 0.933–0.972, $P < 0.001$) and having ≥ 8 antenatal visits (OR = 0.578, 95%CI: 0.447–0.747, $P < 0.001$) were associated with reduced risk. Women with a history of cesarean section had an approximately 3.48-fold higher risk of preterm birth (OR = 3.475, 95%CI: 2.383–5.068, $P < 0.001$). Although pre-pregnancy weight showed a protective trend, the association did not reach statistical significance (OR = 0.980, 95%CI: 0.960–1.001, $P = 0.065$). Multiparity (> 2 births) was associated with lower risk (OR = 0.694, 95%CI: 0.519–0.928, $P = 0.014$).

Regarding socioeconomic factors, pregnant women residing in rural areas exhibited significantly higher risk (OR = 1.863, 95%CI: 1.392–2.495, $P < 0.001$). Conversely, farmers (OR = 0.338, 95%CI: 0.248–0.461, $P < 0.001$) and self-employed or freelance workers (OR = 0.212, 95%CI: 0.133–0.338, $P < 0.001$) had substantially lower risks compared with women in other occupations. No significant associations were observed for education level or employment in enterprises, institutions, or government agencies.

Delay in establishing prenatal records was protective, with each additional gestational week reducing the risk by approximately 8% (OR = 0.922, 95%CI: 0.871–0.976, $P = 0.005$). Women with high-risk pregnancy factors had markedly elevated risk (OR = 64.803, 95%CI: 22.215–189.037, $P < 0.001$). For pregnancy-induced hypertension status, untreated hypertension did not show a significant effect ($P = 0.287$), while treated hypertension was associated with a more than threefold increased risk (OR = 3.456, 95%CI: 2.603–4.590, $P < 0.001$).

Finally, women who underwent OGTT testing had a substantially lower risk of preterm birth (OR = 0.179, 95%CI: 0.106–0.301, $P < 0.001$).

Figure 2 shows a nomogram integrating clinical and socioeconomic factors to predict preterm birth risk. Points are assigned to each factor, and the total indicates the estimated probability, offering a practical tool for personalized prenatal assessment.

Model validation and performance

Using an 80:20 split for training and validation, the baseline model including only clinical variables achieved an AUC of 0.688 on the validation set (Fig. 3).The 95% confidence interval (CI) for the AUC was estimated using bootstrap resampling with 1,000 iterations (95%CI: 0.643–0.733). Incorporating SES-related variables in the extended model resulted in a notable improvement, with an AUC of 0.854 (Fig. 4). The 95% confidence interval

Table 4 Results of Logistic regression analysis for clinically relevant variables

Characteristics	Sub groups	β	SE	OR	95%CI	*P*-value
Age	Age	0.038	0.011	1.039	1.017–1.062	$p < 0.001$
Parity	1–2 times					
	> 2 times	-0.412	0.123	0.662	0.52–0.842	$p < 0.001$
Pre-pregnancy weight (kg)	Pre-pregnancy weight	-0.034	0.009	0.967	0.95–0.983	$p < 0.001$
BMI	BMI	0.167	0.024	1.182	1.128–1.24	$p < 0.001$
History of Cesarean Section	No					
	Yes	1.382	0.176	3.983	2.85–5.689	$p < 0.001$
Gestational weight gain (kg)	Gestational weight gain	-0.028	0.008	0.973	0.957–0.989	0.001

Table 5 Firth-adjusted logistic regression of preterm birth predictors

groups	Characteristics	Sub groups	β	SE	OR	95%CI	*P*-value
Clinical Variables	Age	Age	0.011	0.014	1.011	0.984–1.039	0.416
	Parity	1–2 times					
		> 2 times	-0.366	0.148	0.694	0.519–0.928	0.014
	Pre-pregnancy weight (kg)	Pre-pregnancy weight	-0.02	0.011	0.98	0.96–1.001	0.065
	BMI	BMI	0.095	0.029	1.1	1.039–1.164	< 0.001
	History of Cesarean Section	No					
		Yes	1.246	0.192	3.475	2.383–5.068	< 0.001
	Gestational weight gain (kg)	Gestational weight gain	-0.049	0.01	0.952	0.933–0.972	< 0.001
Clinical & SES Variables	Residence	Community					
		Rural	0.622	0.149	1.863	1.392–2.495	< 0.001
	Education level	College or above					
		Junior high school or below	0.309	0.167	1.362	0.983–1.888	0.065
		Secondary or high school	0.196	0.185	1.216	0.846–1.748	0.293
	Occupation	Others					
		Farmer	-1.085	0.158	0.338	0.248–0.461	< 0.001
		Enterprise staff or government official	0.199	0.216	1.221	0.799–1.866	0.359
		Self-employed or freelancer	-1.55	0.237	0.212	0.133–0.338	< 0.001
	Number of antenatal visits	< 8 visits					
		≥ 8 visits	-0.549	0.131	0.578	0.447–0.747	< 0.001
	Gestational week at first record	Gestational week	-0.081	0.029	0.922	0.871–0.976	0.005
	High-risk pregnancy factors	No					
		Yes	4.171	0.546	64.803	22.215-189.037	< 0.001
	Pregnancy-induced hypertension status	No hypertension					
		Hypertension untreated	-0.439	0.406	0.645	0.291–1.43	0.287
		Hypertension treated	1.24	0.145	3.456	2.603–4.59	< 0.001
	OGTT test conducted	No					
		Yes	-1.723	0.267	0.179	0.106–0.301	< 0.001

(CI) for the AUC was estimated using bootstrap resampling with 1,000 iterations (95%CI: 0.0.819–0.887). ROC curves were generated using the pROC and ggplot2 packages in R, allowing a visual comparison of predictive performance between the two models and highlighting the added value of SES factors in enhancing preterm birth risk prediction.To formally assess the improvement in discrimination, the difference in AUC between the two models was evaluated using bootstrap resampling. The extended model demonstrated a statistically significant increase in AUC compared with the baseline model, with a mean AUC difference of approximately 0.16 and a 95%CI of 0.124–0.209, indicating robust improvement in predictive performance. In addition to AUC performance, the classification metrics further demonstrated the improvement gained by including SES variables. For the baseline model, sensitivity, specificity, and balanced accuracy were 0.559, 0.758, and 0.659, respectively. Classification thresholds were determined using the Youden index on the validation set. In contrast, the extended model achieved a markedly higher sensitivity of 0.862, with specificity and balanced accuracy of 0.724 and 0.793. These results indicate that SES indicators substantially increased the model's ability to correctly identify preterm births, supporting their importance in risk prediction.

Calibration plots were generated to evaluate the agreement between predicted probabilities and observed outcomes for both models. As shown in Figs. 5 and 6, the baseline model (clinical variables only) demonstrated good calibration, with a mean absolute error of 0.009 and a 0.9 quantile of absolute error of 0.018. The extended model (clinical + SES variables) showed a similar level of calibration accuracy, also yielding a mean absolute error of 0.009 and a 0.9 quantile of absolute error of 0.018. Overall, both models exhibited satisfactory predictive consistency, indicating that the predicted probabilities were well aligned with the observed outcomes.

Discussion

This study, conducted among pregnant women in northwestern China, demonstrated that incorporating socioeconomic status (SES) variables markedly enhanced the

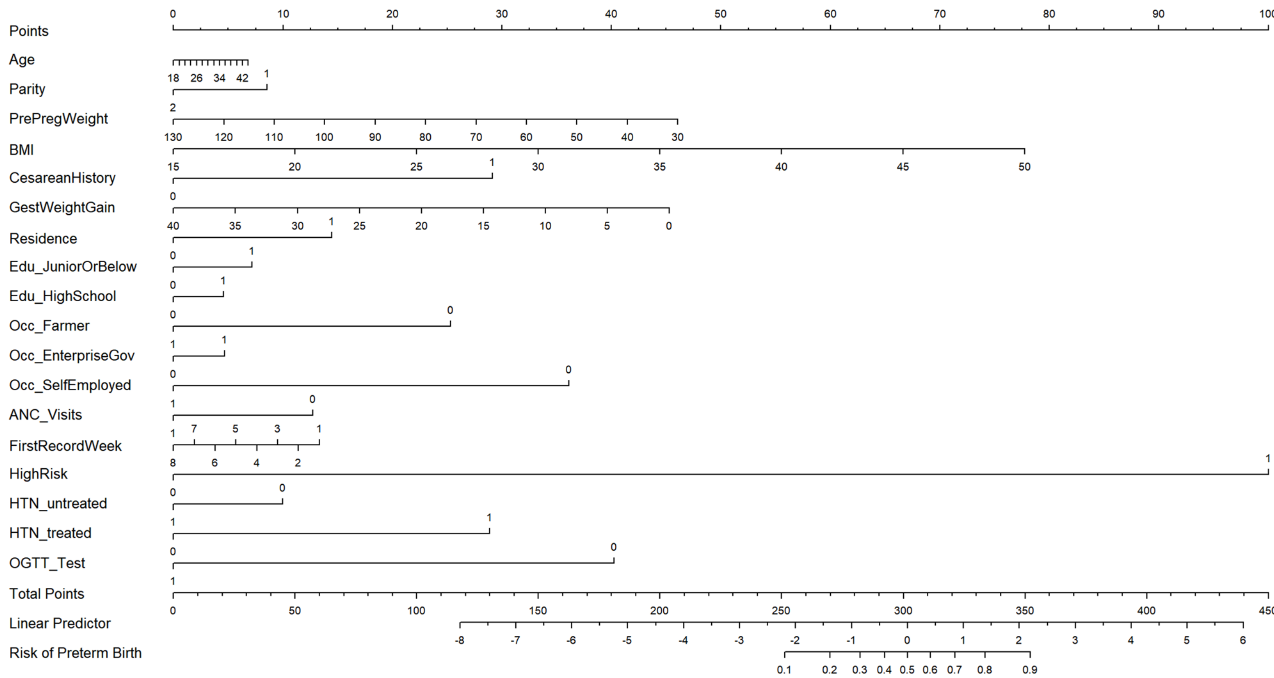

Fig. 2 Nomogram for predicting risk of preterm birth

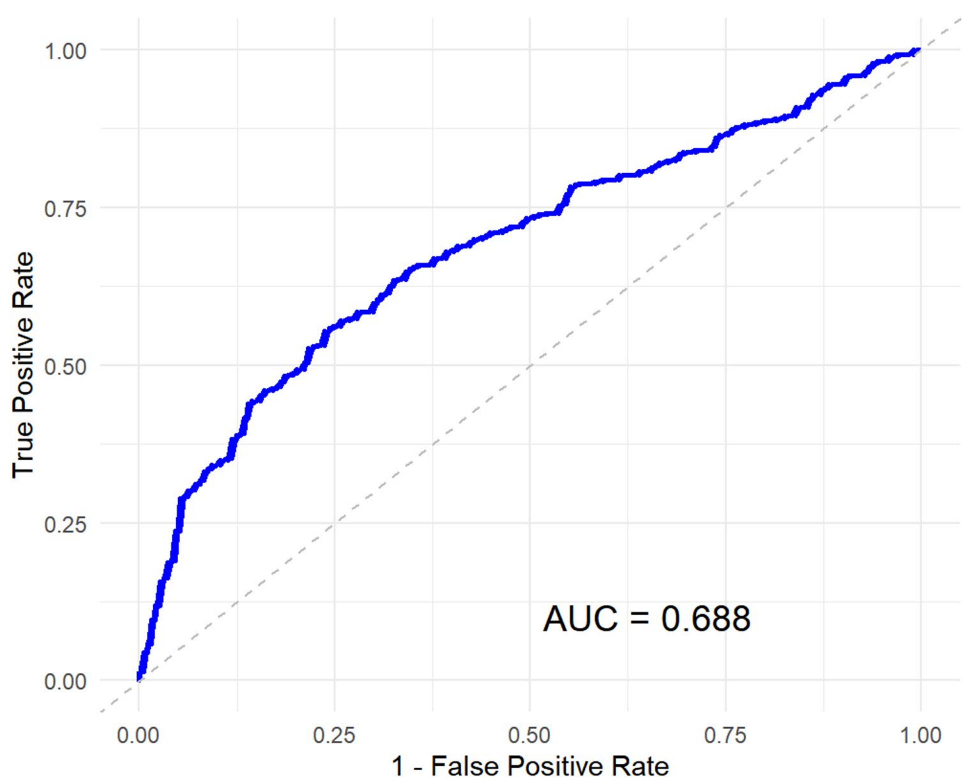

Fig. 3 ROC curve of the prediction model based on the original variables in the validation set

predictive performance of the preterm birth model compared with using clinical variables alone. These findings highlight the pivotal influence of social determinants on maternal health outcomes, particularly in low-resource or remote settings where traditional medical indicators alone may not fully capture risk heterogeneity.Conventional SES indicators, such as household income, were not available in this predominantly low-income,

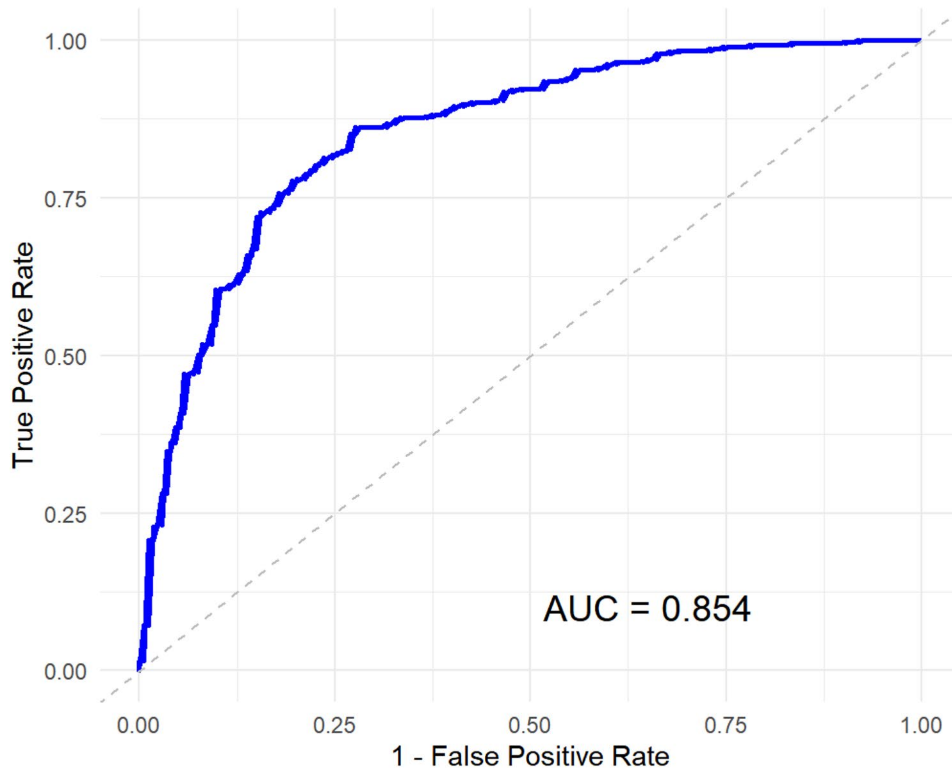

Fig. 4 ROC curves of the prediction models based on original variables and SES indicators in the validation set

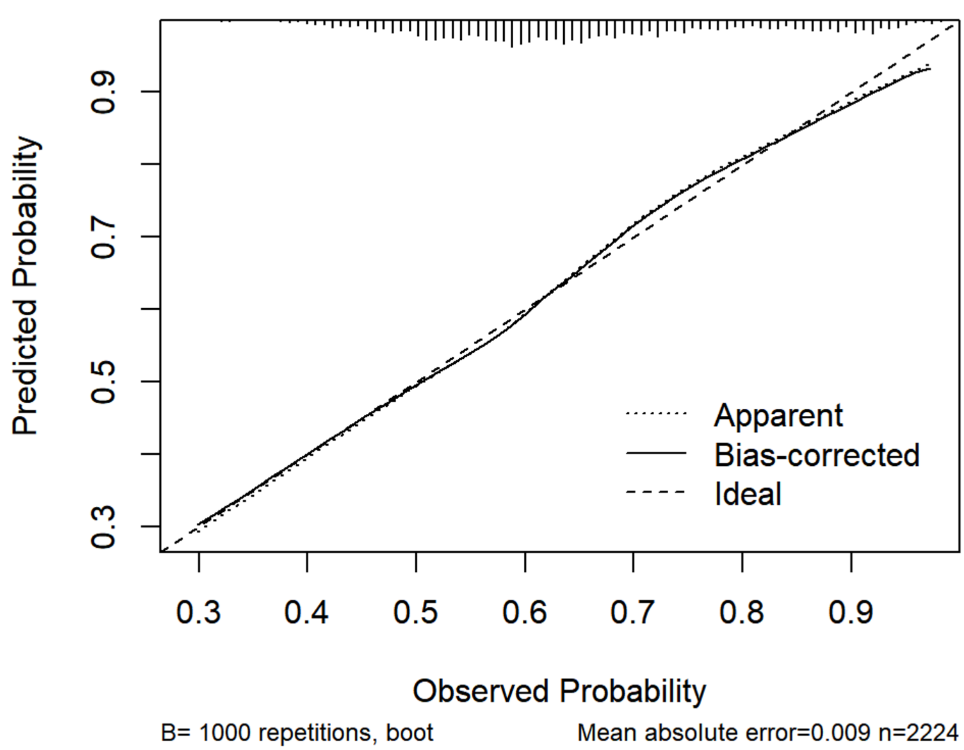

Fig. 5 Baseline model calibration

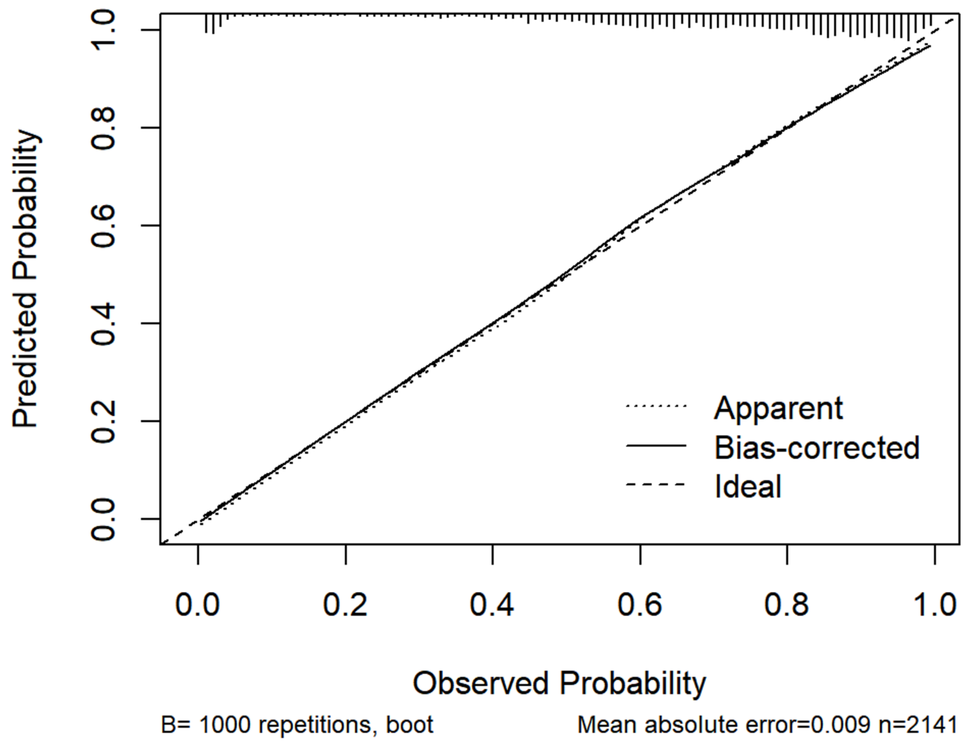

Fig. 6 Extended model calibration

agricultural population. Clinical variables—high-risk pregnancy, gestational hypertension, antihypertensive treatment, and OGTT completion—were used as SES-related proxies, reflecting healthcare access, health literacy, continuity of prenatal care, and treatment adherence, providing additional predictive information beyond standard clinical factors [27–29].

Consistent with Patel et al. (2020) [10], this study shows that SES can be effectively characterized through non-traditional indicators such as place of residence, ethnicity, and healthcare accessibility (e.g., number of antenatal visits and gestational age at registration). Similar to evidence from the Demographic and Health Surveys (DHS) conducted in several Sub-Saharan African countries [30], healthcare utilization indicators were strongly correlated with SES components like household wealth and education, further underscoring the predictive value of SES-related factors in preterm birth.

From a clinical perspective, integrating SES factors into predictive models facilitates the earlier identification of high-risk pregnancies and supports timely preventive interventions or closer monitoring. Embedding such models into electronic medical record (EMR) systems could enable automated risk stratification and personalized care planning.From a public health perspective, these findings suggest that improving access to prenatal care among low-SES women—through early registration incentives, community-based outreach, and better allocation of healthcare resources—may substantially reduce preterm birth risk. Moreover, this framework can guide the development of regionally adapted prediction tools that reflect local socioeconomic realities.

Furthermore, several counterintuitive associations observed in the univariate analysis—such as the higher risk among women with lower education and the apparent protective effect of being a farmer—were largely attenuated after multivariable adjustment. These crude associations are likely influenced by confounding from SES-related and healthcare utilization factors, including antenatal visit frequency, residence, and high-risk pregnancy status. Education often acts as a proxy for broader SES and health-seeking behaviors [27, 28, 31], whereas occupational categories may capture lifestyle patterns, environmental exposures, and rural–urban differences that influence perinatal outcomes [32, 33]. The multivariable Firth-adjusted estimates therefore provide a more reliable assessment of each factor's independent effect on preterm birth risk.

A similar pattern was observed for pre-pregnancy weight and BMI. Although descriptive statistics showed higher averages among preterm births, the adjusted model identified a protective association with higher pre-pregnancy weight. This apparent discrepancy likely reflects the influence of covariates such as gestational weight gain, SES, and healthcare utilization, which reduce confounding in the multivariate analysis. In settings characterized by lower baseline nutrition or higher physical workload, lower maternal weight may reflect limited nutritional reserves or socioeconomic constraints associated with elevated preterm birth risk, whereas relatively higher pre-pregnancy weight may indicate better nutritional status. Similar context-dependent associations have been reported in previous studies [34–36], highlighting the importance of interpreting anthropometric measures within an adjusted statistical framework.

Future research should aim to incorporate more granular SES measures, such as income, household assets, social support, and environmental exposures, to refine predictive accuracy and clarify the mechanisms linking SES to preterm birth. Previous systematic reviews have shown that distinct SES components (e.g., maternal physical condition, housing quality, and environmental behaviors) exert varying mediating effects on preterm birth [31], suggesting that refining SES-based models could further enhance their precision and clinical utility.

Despite these strengths, this study has certain limitations. The data were derived from several county-level medical institutions in northwestern China, which may limit the generalizability of the findings to other regions. Additionally, some conventional SES indicators—such as income and household assets—were unavailable, potentially affecting the comprehensiveness of SES assessment. In addition, preterm birth was analyzed as a binary outcome rather than by clinical subtype. Although information on spontaneous and medically indicated preterm birth was available, subtype-specific analyses were beyond the scope of the present study and may have revealed heterogeneous associations with socioeconomic factors. In addition, the proportion of preterm births reflects an outcome-based analytic sample constructed for predictive modeling, in which women with preterm and term deliveries were compared, rather than a population-based cohort. This design may influence the calibration of absolute risk estimates and the direct transferability of probability thresholds, but does not affect the model's primary goal of assessing and comparing discriminative performance. Future work incorporating reliable regional prevalence data will allow recalibration through weighting or prevalence adjustment to enhance clinical applicability.

Specifically, the SES indicators used in this study primarily captured educational attainment, occupational category, residential context, and access to prenatal healthcare services. Other important dimensions, including household income, accumulated assets, long-term economic stability, and neighborhood-level socioeconomic conditions, were not directly measured. As a result, SES may have been incompletely characterized, and some degree of misclassification cannot be ruled out, which should be considered when interpreting the findings. Moreover, although the study design and variable selection were intended to minimize time-related bias, it is not possible to completely exclude the influence of gestational progression on certain predictors. These variables were incorporated primarily for risk prediction rather than causal inference and may contribute to improved discriminative performance (e.g., AUC). Accordingly, the findings should be interpreted within a predictive modeling framework, and causal interpretations should be made with caution.

Conclusion

It is increasingly recognized that integrating socioeconomic factors with clinical risk factors plays a crucial role in improving the accuracy of preterm birth risk prediction. The findings of this study indicate that models relying solely on clinical variables have limited predictive performance, whereas incorporating socioeconomic characteristics significantly enhances model discrimination. This result suggests that socioeconomic status (SES) is not only an important determinant of health outcomes but also a key factor influencing preterm birth risk.In practice, effectively integrating socioeconomic information into routine clinical management remains challenging, particularly in regions where data accessibility is limited. Future research should focus on developing context-appropriate SES indicator systems applicable to diverse populations and regions, and further validate their associations with maternal and infant health outcomes. Such efforts will provide a stronger evidence base for the precise prediction and targeted intervention of preterm birth.

Acknowledgements

The authors would like to thank the participating hospitals and staff in Northwest China for their assistance and support during data collection.

Authors' contributions

XNW was responsible for the study design and overall direction of the research. QY conducted data processing and statistical analysis. YYM provided overall supervision, data support, and coordination of the manuscript. All authors read and approved the final version of the manuscript.

Funding

This work was supported by the Fundamental Research Funds for the Central Universities (CUC25GT23) and by Professor Xiaoning Wang of the Communication University of China.

Data availability

The datasets used and/or analyzed during the present study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

This retrospective study was approved by the Ethics Committee of Xuanwu Hospital, Capital Medical University, Beijing, China (Approval No. Clinical Research and Review [2024] No.379-001). All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional research committee and with the Declaration of Helsinki. The requirement for informed consent was waived by the Ethics Committee due to the retrospective nature of the study and the use of anonymized clinical data.

Competing interests
The authors declare no competing interests.

Received: 30 October 2025 / Accepted: 9 March 2026
Published online: 16 March 2026

References

1. World Health Organization. 2023. Preterm birth.
2. Bradley E, Lawn JE, Mwansa JK, et al. Global epidemiology of preterm birth and drivers for change. Nat Reviews Disease Primers. 2025;11(1):41. https://doi.org/10.1038/s41572-025-00892-5.
3. Ohuma E, Møller A-B, Bradley E, et al. National, regional, and worldwide estimates of preterm birth in 2020, with trends from 2010: A systematic analysis. Lancet. 2023;402(10409):1261–71. https://doi.org/10.1016/S0140-6736(23)00878-4.
4. Koullali B, Oudijk MA, Nijhuis JG, et al. Risk factors for preterm birth: An umbrella review of meta-analyses of observational studies. BMC Med. 2023;21:494. https://doi.org/10.1186/s12916-023-03171-4.
5. Huang S, Guo Y, Xu X, et al. Gestational diabetes complicated with preterm birth: A retrospective cohort study. BMC Pregnancy Childbirth. 2024;24:631. https://doi.org/10.1186/s12884-024-06810-7.
6. American Psychological Association. Socioeconomic status. 2019.
7. United Nations. Sustainable Development Goal 5: Achieve gender equality and empower all women and girls. 2025.
8. United Nations. Sustainable Development Goal 10: Reduce inequality within and among countries. 2025.
9. Systematic Reviews. Utilization of maternal healthcare services in low- and middle-income countries: A systematic review and meta-analysis. 2025.
10. Patel AB, Bann CM, Garces AL, et al. Development of the Global Network for Women's and Children's Health Research's socioeconomic status index for use in the network's sites in low and lower middle-income countries. Reproductive Health. 2020;17(Suppl 3):193. https://doi.org/10.1186/s12978-020-01034-2.
11. Xia L, Wu Z. A study on livelihood capital of residents in Hotan Prefecture under the sustainable livelihood framework. Agric Technol. 2019;39(10):153–5. https://doi.org/10.19754/j.nyyjs.20190530075. (in Chinese).
12. Vousden N, Lawley E, Seed PT, et al. Ethnic disparities in maternal health outcomes: A systematic review and meta-analysis. EClinicalMedicine. 2022;44:101283. https://doi.org/10.1016/j.eclinm.2021.101283.
13. Ekholuenetale M, Nzoputam CI, Barrow A. Effects of socioeconomic factors and booking time on the WHO recommended eight antenatal care contacts in Liberia. PLOS Global Public Health. 2022;2(2):e0000136. https://doi.org/10.1371/journal.pgph.0000136.
14. Baten A, Biswas RK, Kendal E, et al. Utilization of maternal healthcare services in low- and middle-income countries: A systematic review and meta-analysis. Syst Reviews. 2025;14(1):88. https://doi.org/10.1186/s13643-025-02832-0.
15. Wang Q, Chen D, Wei Y, et al. Study on preterm birth risk prediction model. Prev Med. 2024;36(8):663–8. (in Chinese).
16. Pepe MS. The statistical evaluation of medical tests for classification and prediction. Oxford University Press; 2003. https://doi.org/10.1093/oso/9780198509844.001.0001.
17. Dejene EA, Endris BS, Geremew AB, et al. Development and validation of a simplified risk prediction model for preterm birth: A prospective cohort study in rural Ethiopia. BMC Pregnancy Childbirth. 2024;24:112. https://doi.org/10.1038/s41598-024-55627-z.
18. Pediatric and Child Health Journal. Neighborhood child opportunity and preterm birth rates by race and ethnicity. JAMA Netw Open. 2024;7(9):e2823537. https://doi.org/10.1001/jamanetworkopen.2024.2823537.
19. Zhao Y, Li X, Wang H, et al. Maternal education level and risk of preterm birth: A population-based cohort study in China. BMC Pregnancy Childbirth. 2024;24:512. https://doi.org/10.1186/s12884-024-06602-z.
20. Adane HA, Iles R, Boyle JA, Collie A. Maternal occupational risk factors and preterm birth: A systematic review and meta-analysis. PLoS ONE. 2023;18(7):e0283752. https://doi.org/10.1371/journal.pone.0283752.
21. Lima dos Santos SP, Calloway EE, Chertok IRA, et al. Development and testing of the comprehensive prenatal care index: Relationship with preterm birth and small for gestational age across racial and ethnic groups. J Midwifery Women's Health. 2024;69(6):917–28. https://doi.org/10.1111/jmwh.13707.
22. Papageorghiou AT, Ohuma EO, Altman DG, et al. International standards for fetal growth based on serial ultrasound measurements: The Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project. Lancet. 2014;384(9946):869–79. https://doi.org/10.1016/S0140-6736(14)61490-2.
23. Perejón D, Ríos-González S, Cuesta M, et al. Hypertension subtypes and adverse maternal and perinatal outcomes: Associations with preterm birth. BMC Pregnancy Childbirth. 2024;24:754. https://doi.org/10.1186/s12884-024-06754-y.
24. Watson HA, Carter J, Seed PT, Tribe RM, Shennan AH. The QUiPP app: A safe alternative to a treat-all policy for threatened preterm labour. PLoS Med. 2019;16(9):e1002904. https://doi.org/10.1371/journal.pmed.1002904.
25. Locatelli A, Santucci A, Gervasi MT, et al. Antihypertensive therapy during pregnancy: Prescription patterns and safety considerations. Front Pharmacol. 2024;15:1370797. https://doi.org/10.3389/fphar.2024.1370797.
26. Liang R, Zhang X, Liu Y, et al. Outcome of pregnancy oral glucose tolerance test and associations with preterm birth: A population-based study. BMC Pregnancy Childbirth. 2024;24:631. https://doi.org/10.1186/s12884-024-06810-7.
27. Blumenshine P, Egerter S, Barclay CJ, Cubbin C, Braveman PA. Socioeconomic disparities in adverse birth outcomes: A systematic review. Am J Prev Med. 2010;39(3):263–72. https://doi.org/10.1016/j.amepre.2010.05.012.
28. Adler NE, Newman K. Socioeconomic disparities in health: Pathways and policies. Health Aff. 2002;21(2):60–76. https://doi.org/10.1377/hlthaff.21.2.60.
29. Kieffer EC, Sinco BR, Pirsch C, et al. The role of health literacy and socioeconomic factors in gestational diabetes screening: Determinants of OGTT completion. Diabetes Care. 2013;36(5):1219–25. https://doi.org/10.2337/dc12-1413.
30. Tekeba B, Techane MA, Workneh BS, et al. Determinants of preterm birth among reproductive-age women in sub-Saharan Africa: Evidence from Demographic and Health Survey data (2019–2022). PLoS ONE. 2024;19(6):e0305810. https://doi.org/10.1371/journal.pone.0305810.
31. Balbierz A, Kawachi I, Goin DE, Denney JT, Sohn MA. Mediators of socioeconomic inequalities in preterm birth: A systematic review. BMC Public Health. 2022;22:2201. https://doi.org/10.1186/s12889-022-14661-5.
32. Målqvist M, Persson LÅ, Selling KE, Thomsen S. Ethnic disparities in neonatal mortality in rural northern India: A population-based study. Int J Equity Health. 2007;6:4. https://doi.org/10.1186/1475-9276-6-4.
33. Kroll ME, Green J, Beral V, Reeves GK. Social inequalities in risk of preterm birth: Retrospective cohort study. BMJ. 2012;345:e8232. https://doi.org/10.1136/bmj.e8232.
34. Villar J, Cheikh Ismail L, Victora CG, et al. Maternal and neonatal anthropometric indicators and risk of preterm birth: A multicountry prospective cohort study. Lancet Global Health. 2014;2(8):e437–46. https://doi.org/10.1016/S2214-109X(14)70200-4.
35. Kramer MS, McLean FH, Boyd ME, et al. Maternal anthropometry and pregnancy outcomes. JAMA. 2002;288(6):749–57. https://doi.org/10.1001/jama.288.6.749.
36. Li Z, Zhang Y, Liu L, et al. Maternal prepregnancy body mass index and risk of preterm birth: A population-based study in China. BMC Pregnancy Childbirth. 2015;15:128. https://doi.org/10.1186/s12884-015-0562-5.

Publisher's Note