Ding *et al.*
European Journal of Medical Research (2025) 30:99
https://doi.org/10.1186/s40001-025-02361-5

European Journal
of Medical Research

RESEARCH

Open Access

Constructing a logistic regression-based prediction model for subsequent early pregnancy loss in women with pregnancy loss

Nan Ding[1], Peili Wang[1], Xiaoping Wang[1] and Fang Wang[1*]

Abstract

Objectives The aim of this study is to construct a nomogram for predicting subsequent early pregnancy loss in women with a history of pregnancy loss, which may increase well-being and the capacity for managing reproductive options.

Materials and methods We conducted a retrospective analysis of medical records from women with a history of pregnancy loss at the Reproductive Medicine Center of Lanzhou University Second Hospital between January 2019 and December 2022. A cohort of 718 patients was selected for the study. We structured our data into a training set of 575 cases (80% of the cohort) and a test set of 143 cases (20%). To identify significant predictors, we applied a stepwise forward algorithm guided by the Akaike Information Criterion (AIC) to the training set. Model validation was conducted using the test set. For the validation process, we employed various methods to assess the predictive power and accuracy of the model. Receiver Operating Characteristic (ROC) curves provided insights into the model's ability to distinguish between outcomes effectively. Calibration curves assessed the accuracy of the probability predictions against actual outcomes. The clinical utility of the model was further evaluated through Decision Curve Analysis, which quantified the net benefits at various threshold probabilities. In addition, a nomogram was developed to visually represent the risk factors.

Results Among the 36 candidate variables initially considered, 10 key predictors were identified through logistic regression analysis and incorporated into the nomogram. These selected variables include age, education, thrombin time (TT), antithrombin III (AT-III), D-dimer levels, 25-hydroxy Vitamin D, immunoglobulin G(IgG), complement components C4, anti-cardiolipin antibody (ACA) and lupus anticoagulant (LA). In addition, based on clinical experience, the number of previous pregnancy losses was also included as a predictive variable. The prediction model revealed an area under the curve (AUC) of approximately 0.717 for the training set and 0.725 for the validation set. Calibration analysis indicated satisfactory goodness-of-fit, with a Hosmer–Lemeshow test yielding a $\chi 2$ value of 7.78 ($p = 0.55$). Decision curve analysis confirmed the clinical utility of the nomogram. Internal validation confirmed the robust performance of the predictive model.

Conclusions The constructed nomogram provides a valuable tool for predicting subsequent early pregnancy loss in women with a history of pregnancy loss. This nomogram can assist clinicians and patients in making informed decisions regarding the management of pregnancy and improving clinical outcomes.

*Correspondence:
Fang Wang
ery_fwang@lzu.edu.cn

Trial Registration: This study was registered in the Chinese Clinical Trial Registry under the registration number ChiCTR2000039414 on October 27, 2020. The registration was done retrospectively.

Keywords Early pregnancy loss, Logistic regression, Prediction model

Introduction

Pregnancy loss, defined as the loss of a pregnancy before fetal viability at less than 24 weeks of gestation, has a multifaceted etiology, and it affects 12–15% of pregnancies [1, 2]. Statistics indicate approximately 23 million miscarriages annually, equating to 44 pregnancy losses per minute [3]. Pregnancy loss not only impacts a patient's desire to have children but also imposes significant psychological and social burdens [4]. In many cultures, particularly in low- and middle-income countries, societal stigma surrounding childlessness or unmet reproductive expectations exacerbates the emotional challenges faced by affected women [5]. This distress often extends to their families, further affecting quality of life.

Anatomical uterine defects, chromosomal abnormalities, endometrial dysfunction, thrombophilia, infectious agents, immune-related factors, and environmental influences all contribute to pregnancy loss [6, 7]. However, the causative factors for approximately 40–50% of cases remain elusive [8]. Its multifactorial nature, diverse clinical manifestations, and lack of standardized and efficacious treatments render pregnancy loss a significant clinical challenge [9]. According to the ESHRE guidelines for recurrent pregnancy loss. The standard workup for pregnancy loss typically includes genetic screening, hormonal assessments, ultrasound imaging. In addition, the evaluation includes testing for autoimmune conditions, particularly antiphospholipid syndrome, as well as thyroid function. The guidelines emphasize a targeted diagnostic approach, focusing on identifying modifiable risk factors while avoiding unnecessary and invasive testing [1]. At present, although a more comprehensive etiological screening is currently offered to couples who have experienced a previous pregnancy loss, predicting and preventing subsequent pregnancy losses remains a difficult task. When a pregnancy loss occurs, couples need accurate information about their prospects of becoming parents and suitable assistance should be provided to lessen the psychological stress brought on by miscarriage [10]. While numerous studies have devised predictive models for miscarriage risk based on individual factors, such as immune or endocrine factors, there remains a dearth of research focusing on establishing predictive models considering multiple etiological factors [11–13]. Developing a dependable predictive model assumes paramount importance in identifying high-risk patients for adverse pregnancy outcomes and furnishing them with preventive guidance.

Early prediction of subsequent pregnancy loss risk using preconception indicators may hold particular value, enabling clinicians to optimize management strategies for patients with a history of pregnancy loss. In addition, accurately assessing individual patient risk may alleviate some anxiety associated with the unpredictability of pregnancy outcomes. Nomograms, graphical representations delineating the likelihood of an event occurring for each patient, are increasingly favored among clinicians due to their simplicity, reproducibility, and capacity for personalized risk assessment [14]. This study aims to systematically provide clinically specific information regarding pregnancy loss management and assist clinicians in decision-making.

Materials and methods

We constructed the predicting model in a retrospective study in Lanzhou University Second Hospital Reproductive Medicine Center from January 2019 to December 2022. The study received approval from the ethical committee of Lanzhou University (2019A-231). In addition, this study was registered in the Chinese Clinical Trial Registry with the registration number ChiCTR2000039414 on October 27, 2020. In this study, Inclusion and Exclusion Criteria are as follows: inclusion criteria: (1) patients with one or more clinically confirmed intrauterine pregnancy losses before the 24th week of gestation. Pregnancy loss was defined as the spontaneous demise of a pregnancy visualized by ultrasonography, which could include: anembryonic pregnancy (empty gestational sac), Yolk sac-only pregnancy (gestational sac with a yolk sac but no fetal pole), Embryonic loss (before 10 weeks of gestation), Fetal loss (at or after 10 weeks of gestation) [4], (2) Patients aged 18 years or older. The exclusion criteria were as follows: (1) chromosomal abnormalities in the couple, (2) adverse obstetric outcomes, such as ectopic pregnancy, hydatidiform mole, and biochemical pregnancy, (3) congenital uterine anomalies, and (4) patients with unknown or missing pregnancy outcomes.

Based on published literature and clinical expertise, we selected candidate variables for further analysis. Medical records were meticulously reviewed, and the following data were assessed: maternal demographic

characteristics, such as age, education, race, and pregestational body mass index (BMI). Also considered were the number of prior pregnancy losses a patient had experienced. Clinical characteristics encompassed thyroid-related indicators, such as free triiodothyronine (FT3), free thyroxine (FT4), and thyroid stimulating hormone (TSH). Lipid metabolism variables included total cholesterol (TC), triglycerides (TG), high-density lipoprotein (HDL), and low-density lipoprotein (LDL). Glucose metabolism was assessed through fasting blood glucose (FBG), fasting insulin (FINS), and homeostatic model assessment of insulin resistance (HOMA-IR). We also evaluated coagulation-related indicators including prothrombin time (PT) and its derivatives, activated partial thromboplastin time (APTT), fibrinogen (FIB), D-dimer, and fibrin degradation products (FDP). Immune-related variables assessed were immunoglobulins (IgA, IgG, and IgM), complement components (C3, C4), and specific antibodies such as anti-nuclear antibody (ANA), rheumatoid factor (RF), and lupus anticoagulant (LA). Additional variables included homocysteine (HCY) and 25-hydroxy Vitamin D levels. This comprehensive set of variables were chosen to explore potential predictors of early pregnancy outcomes (Supplementary Table 1).

In this study, we analyzed a data set initially comprising 1065 cases. Data cleaning involved the exclusion of cases based on the following: 28 cases were excluded due to missing pregnancy outcomes, 8 cases due to chromosomal abnormalities, 3 cases of hydatidiform mole, 2 cases of ectopic pregnancy, and 7 cases of biochemical pregnancy. This resulted in a screened data set of 1,017 cases. Further exclusion involved removing cases with more than 25% missing indicators, reducing the data set to 718 cases. In our data set of 718 cases, most variables not only exhibited some degree of missing data but also contained outliers. To address these outliers, we employed the Tukey method, a widely used technique for detecting anomalies in various fields. The basic principle of the Tukey method involves calculating the interquartile range (IQR) of the data and then multiplying it by a factor (typically 1.5 or 3) to establish a threshold[15, 16]. Data points that exceed this threshold are considered outliers. In this study, we used a factor of 3 to define and manage outliers, ensuring the robustness of our analysis. Finally, variables such as age, education, and number of pregnancy losses had no missing data. BMI and race had minimal missing values, at 1.53% and 0.14%, respectively. Thyroid-related indicators showed higher missingness, with FT3 and FT4 each missing in approximately 10.86% and 10.58% of cases, respectively. TSH was missing in 8.22% of the cases. Coagulation markers also showed significant missing data: PT, PT%, and PT-R were missing in 14.21%, 14.48%, and 14.62% of cases, respectively, with similar rates for INR, APTT, and TT. Immune-related markers and other variables such as D-dimer, HCY, 25-OH Vitamin D, and various immunoglobulins (IgG, IgA, IgM) were less frequently missing, ranging from approximately 3.76–9.75%. Lipid and glucose metabolism markers such as total cholesterol, triglycerides, HDL, LDL, and HOMA-IR were moderately missing, with HOMA-IR notably missing in 16.02% of cases. The above missing data were subsequently imputed using multiple imputation techniques to maximize the utility of the remaining data [17]. Finally, the refined data set was then divided into a training set comprising 80% of the cases (575 cases) and a test set comprising 20% (143 cases). The detailed flowchart is shown in Fig. 1.

After data cleaning, we employed logistic regression to model the risk of early pregnancy loss, identifying significant predictors and estimating their effect using train data set. Model validation was conducted using the test data set, with performance evaluated via Receiver Operating Characteristic (ROC) curves and Calibration curves. Additional analyses included a Nomogram for visualizing the risk factors and their weights, and

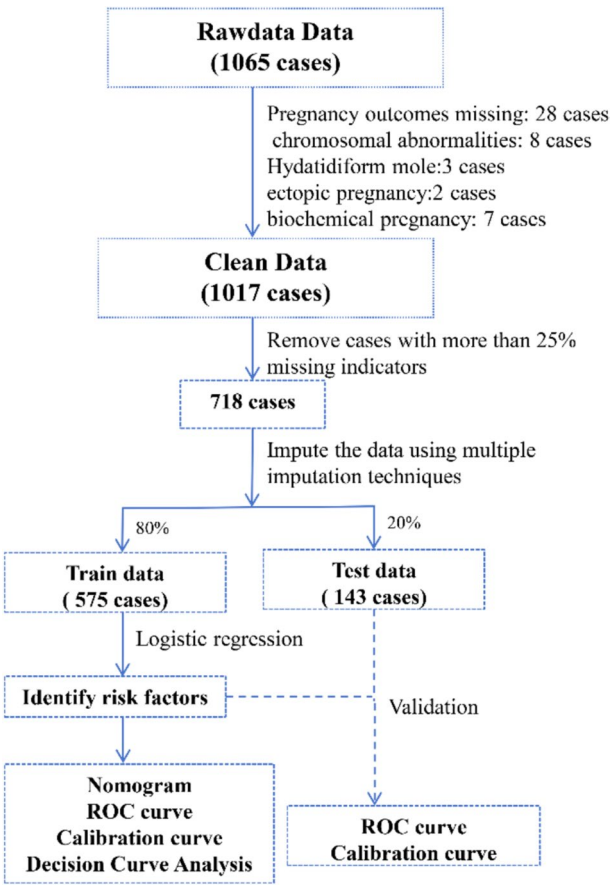

Fig. 1 Flowchart of study design and participant selection

Decision Curve Analysis to evaluate the clinical usefulness of the model. DCA evaluates the clinical utility of prediction models by calculating the net benefit at various threshold probabilities for intervention. It compares the model's performance with two extreme strategies: intervention for all and intervention for none. The net benefit is derived from true positives minus false positives, adjusted by the threshold probability. DCA helps determine the threshold, where the model provides the greatest benefit, offering insights into its real-world applicability and clinical value. This robust approach ensures comprehensive model validation and aims to improve the understanding and management of early pregnancy loss.

Definition of pregnancy outcome

The primary outcome was early pregnancy loss, defined as pregnancy loss occurring before 10 weeks of gestation, which includes cases of an empty gestational sac, gradual cessation of embryonic development, embryonic or fetal death, and expulsion of the embryo and its appendages. Clinical ongoing pregnancy was defined as the presence of an intrauterine gestational sac with a detectable fetal heartbeat on ultrasound after 10 weeks of gestation.

Statistical analyses

Continuous variables were reported as medians with interquartile ranges (Q1–Q3), and categorical variables were presented as percentages. The Chi-squared test was utilized to analyze categorical variables, while the Wilcoxon–Mann–Whitney test was applied for continuous variables, as appropriate. The primary outcome investigated was early pregnancy loss. We conducted univariate logistic regression analyses to explore the relationship between each variable and the outcome. Variables with a p value less than 0.2 were initially selected for further analysis. For the multivariate logistic regression, we adopted a backward stepwise approach guided by the Akaike Information Criterion (AIC). The AIC balances model fit and complexity, favoring models that achieve better predictive performance without overfitting. This method was used to refine the selection of predictors that contributed significantly to the model's overall explanatory power, without solely relying on p values. This comprehensive approach ensures that the final model incorporates all relevant predictors to accurately identify risk factors associated with early pregnancy loss.

For model performance evaluation, we assessed both discrimination and calibration. Discrimination was quantified by calculating the area under the receiver operating characteristic curve (AUC) and the calibration was evaluated by calibration plots accompanied by the Hosmer–Lemeshow test. Calibration plots were used to visualize the model's performance, where the closer the observed line is to the 45-degree diagonal line, the better the calibration. The Hosmer–Lemeshow test assesses the goodness-of-fit by comparing observed and expected event rates. A p value greater than 0.05 from the HL test indicates little evidence of a significant departure from the expected values, suggesting that the model fits well. A nomogram was formulated based on the results of multivariate logistic regression analysis. Decision curve analysis (DCA) was employed to ascertain the clinical utility of the nomogram by quantifying net benefits at different threshold probabilities in the validation data set. Statistical analyses were performed using R statistical software (https://www.r-project.org/,version 4.1.3). Two-sided p values < 0.05 were considered statistically significant.

Results

Patient characteristics

In this study, 718 pregnant women who had at least one prior pregnancy loss were finally included. Prior to modeling, all included variables were comprehensively analyzed for their association with early pregnancy loss. In our study of 718 patients, Table 1 meticulously analyzes the baseline clinical features, distinguishing between individuals with clinical ongoing pregnancies ($n=550$) and those experiencing early pregnancy loss ($n=168$). The analysis identified several potential key variables significantly associated with early pregnancy loss. Older age was a prominent factor, with older patients more likely to experience loss ($p=0.002$). Educational attainment also played a role, where lower education levels correlated with higher risk ($p=0.05$). A history of multiple pregnancy losses was notably linked to increased risk ($p=0.009$). Biochemical and physiological markers such as thyroid function (FT3, $p=0.026$), thrombin time (TT, $p=0.008$), antithrombin III (AT-III, $p<0.001$), D-dimer levels ($p=0.005$), homocysteine (HCY, $p=0.04$), and 25-hydroxy Vitamin D ($p<0.001$) also showed significant associations. In addition, immunological factors such as Immunoglobulin G (IgG, $p=0.018$), anticardiolipin antibodies (ACA, $p<0.001$), and lupus anticoagulant (LA, $p<0.001$) were identified as significant predictors of pregnancy loss. Furthermore, we divided the data set into a training set and a validation set using an 8:2 split, with the purpose of developing and subsequently validating our predictive model. This division allows us to establish the model using the training set and then rigorously test its accuracy and generalizability with the validation set. The baseline characteristics of these two subsets are comprehensively detailed in Supplementary Tables 2, 3, ensuring that the features remain consistent across both data sets to facilitate reliable and robust model training and validation.

Table 1 Patients baseline clinical features analysis

Variables	Total ($n=718$)	Clinical ongoing pregnancy ($n=550$)	Early pregnancy loss ($n=168$)	*p* value
Age(year)	30 (27, 32)	29 (27, 32)	30 (28, 33)	0.002
BMI(kg/m^2)	22 (20, 23.9)	21.9 (19.9, 23.8)	22 (20.3, 23.9)	0.229
Education, *n* (%)				0.05
High school or below	249 (34.7)	178 (32.4)	71 (42.3)	
University	443 (61.7)	350 (63.6)	93 (55.4)	
Graduate school or above	26 (3.6)	22 (4)	4 (2.4)	
Race, *n* (%)				0.823
Han Chinese	659 (91.8)	506 (92)	153 (91.1)	
Other ethnicities	59 (8.2)	44 (8)	15 (8.9)	
Number of PL, *n* (%)				0.009
1(%)	197 (27.4)	166 (30.2)	31 (18.5)	
2(%)	294 (40.9)	220 (40)	74 (44)	
≥ 3(%)	227 (31.6)	164 (29.8)	63 (37.5)	
FT3(pmol/L)	5.2 ± 0.5	5.3 ± 0.5	5.2 ± 0.5	0.026
FT4(pmol/L)	15.9 ± 2.1	15.8 ± 2	16 ± 2.2	0.377
TSH(pmol/L))	2.5 (1.7, 3.3)	2.5 (1.7, 3.3)	2.6 (1.8, 3.3)	0.219
PT(s)	11.1 (10.7, 11.6)	11.1 (10.7, 11.6)	11.1 (10.7, 11.6)	0.867
PT(%)	98.9 (91, 106)	99 (91, 106)	98 (91.4, 105)	0.532
PT-R	1 (1, 1)	1 (1, 1)	1 (1, 1)	0.241
INR	1 (1, 1)	1 (1, 1)	1 (1, 1)	0.211
APTT(s)	31.3 (27.8, 33.9)	31.3 (27.1, 33.9)	31.3 (29.1, 33.9)	0.36
FIB(g/L)	2.8 (2.4, 3.2)	2.8 (2.4, 3.2)	2.8 (2.4, 3.2)	0.449
TT (s)	14.6 (13.6, 16.5)	14.6 (13.6, 17.1)	14.4 (13.4, 15.3)	0.008
AT-III(%)	105.8 ± 12.3	106.6 ± 12.3	103 ± 12	< 0.001
D-dimer(μg/mL)	0.2 (0.1, 0.3)	0.2 (0.1, 0.3)	0.2 (0.1, 0.3)	0.005
HCY (μmol/L)	11 (9.2, 12.9)	10.8 (9, 12.8)	11.4 (9.9, 13)	0.04
25-hydroxy vitamin D(ng/mL)	11.2 (8.7, 14.3)	11.6 (9, 14.7)	10.1 (7.5, 13.3)	< 0.001
IgG (g/L)	12.8 (11.4, 14.3)	12.7 (11.3, 14.3)	13.1 (11.9, 14.6)	0.018
IgA (g/L)	2.1 (1.7, 2.8)	2.1 (1.7, 2.8)	2.2 (1.7, 2.8)	0.968
IgM (g/L)	1.5 (1.2, 2)	1.5 (1.2, 1.9)	1.5 (1.2, 2)	0.976
C3 (g/L)	1.1 (1, 1.3)	1.1 (1, 1.3)	1.2 (1, 1.3)	0.099
C4 (g/L)	0.3 (0.2, 0.3)	0.3 (0.2, 0.3)	0.3 (0.2, 0.3)	0.105
FBG (mmol/L)	5 (4.7, 5.2)	4.9 (4.7, 5.2)	5 (4.7, 5.2)	0.387
FINS (mU/L)	9.9 (7, 13.3)	10.1 (6.8, 13.4)	9.5 (7.3, 13.1)	0.833
HOMA-IR	2.2 (1.5, 3.1)	2.2 (1.5, 3.1)	2.2 (1.6, 3.1)	0.858
TC (mmol/L)	3.9 (3.5, 4.3)	3.9 (3.4, 4.3)	4 (3.5, 4.5)	0.064
TG (mmol/L)	0.9 (0.7, 1.3)	0.9 (0.7, 1.3)	0.9 (0.7, 1.3)	0.982
HDL (mmol/L)	1.3 (1.2, 1.6)	1.3 (1.2, 1.6)	1.3 (1.1, 1.5)	0.815
LDL(mmol/L)	2.4 (2.1, 2.9)	2.4 (2.1, 2.8)	2.4 (2.1, 3)	0.052
ANA, *n* (%)				0.359
Negative	616 (85.8)	476 (86.5)	140 (83.3)	
Positive	102 (14.2)	74 (13.5)	28 (16.7)	
RF, *n* (%)				0.831
Negative	562 (78.3)	432 (78.5)	130 (77.4)	
Positive	156 (21.7)	118 (21.5)	38 (22.6)	
ACA, *n* (%)				< 0.001
Negative	605 (84.3)	478 (86.9)	127 (75.6)	
Positive	113 (15.7)	72 (13.1)	41 (24.4)	
β2GP1, *n* (%)				0.811

Table 1 (continued)

Variables	Total ($n=718$)	Clinical ongoing pregnancy ($n=550$)	Early pregnancy loss ($n=168$)	*p* value
Negative	680 (94.7)	522 (94.9)	158 (94)	
Positive	38 (5.3)	28 (5.1)	10 (6)	
LA, *n* (%)				< 0.001
Negative	686 (95.5)	535 (97.3)	151 (89.9)	
Positive	32 (4.5)	15 (2.7)	17 (10.1)	

Table 2 Univariate and multivariable logistics regression analyses identified the risk factors associated with early pregnancy loss

Variables	Univariate analysis			Multivariate analysis (Based on the backward stepwise logistic regression (LR) method)		
	β	OR (95% CI)	*p* value	*β*	OR (95% CI)	*p* value
Age	0.067	1.07 (1.02–1.12)	0.007	0.056	1.06 (1–1.12)	0.034
High school or below		1			1	
University	− 1.34	0.26 (0.06–1.15)	0.076	− 0.438	0.65 (0.42–1)	0.051
Graduate school or above	− 0.364	0.7 (0.46–1.04)	0.078	− 0.438	0.65 (0.42–1)	0.051
History of one PL		1				
History of two PL	0.497	1.64 (0.97–2.77)	0.063			
History of three or more PL	0.602	1.83 (1.06–3.13)	0.029			
FT3	− 0.439	0.64 (0.43–0.97)	0.034			
APTT	0.029	1.03 (0.99–1.07)	0.152			
TT	− 0.095	0.91 (0.83–1)	0.053	− 0.097	0.91 (0.81–1.02)	0.092
AT-III	− 0.03	0.97 (0.95–0.99)	0	− 0.028	0.97 (0.96–0.99)	0.002
D-dimer	1.724	5.6 (1.26–24.91)	0.024	1.522	4.58 (0.83–25.16)	0.08
HCY	0.049	1.05 (0.99–1.11)	0.099			
25-Hydroxy Vitamin D	− 0.088	0.92 (0.87–0.96)	0	− 0.089	0.91 (0.87–0.96)	0.001
IgG	0.062	1.06 (0.98–1.16)	0.136	0.078	1.08 (0.99–1.18)	0.093
C4	1.696	5.45 (0.67–44.12)	0.112	1.823	6.19 (0.58–65.69)	0.13
TC	0.292	1.34 (1.02–1.75)	0.033			
LDL	0.276	1.32 (0.97–1.78)	0.075			
ANA(Negative)		1				
ANA(Positive)	0.404	1.5 (0.87–2.57)	0.142			
ACA(Negative)		1				
ACA(Positive)	0.857	2.36 (1.44–3.86)	0.001	0.891	2.44 (1.41–4.21)	0.001
LA(Negative)		1				
LA(Positive)	1.237	3.45 (1.48–8)	0.004	1.222	3.39 (1.34–8.57)	0.01

In Table 2, using a backward stepwise logistic regression method guided by the Akaike Information Criterion (AIC), we identified several significant predictors of early pregnancy loss. Age was confirmed as a risk factor, with each additional year slightly increasing the likelihood of pregnancy loss (OR = 1.06, 95%CI 1.0–1.12, $p = 0.034$). Education level also influenced outcomes, revealing that lower educational attainment, specifically high school or below, was associated with a higher risk compared to higher educational levels. Thrombin time (TT), antithrombin III (AT-III), D-dimer, 25-hydroxy Vitamin D, and Immunoglobulin G (IgG), each played a role, with significant findings particularly for D-dimer and Vitamin D, indicating their critical involvement in the pathology of early pregnancy loss. Complement component 4 (C4), although not reaching traditional statistical significance, was included due to its predictive relevance as suggested by AIC. Anticardiolipin antibodies (ACA) and lupus anticoagulant (LA) were among the strongest predictors, significantly elevating the risk of adverse pregnancy

outcomes (ACA: OR = 2.44, 95%CI 1.41–4.21, p = 0.001; LA: OR = 3.39, 95%CI 1.34–8.57, p = 0.01). This analysis underscores the complex interplay of demographic, clinical, and biochemical factors in predicting early pregnancy loss.

Construction of a predictive nomogram model in women after cervical cerclage

The discriminatory power and generalizability of our models were evaluated using receiver operating characteristic (ROC) curves. For our logistic regression model, the analysis revealed an area under the curve (AUC) of approximately 0.717 for the training set and 0.725 for the validation set (Fig. 2A, C). According to standard thresholds, these AUC values indicate good predictive capabilities (0.7–0.8). Calibration using the Hosmer–Lemeshow test yielded satisfactory goodness-of-fit across both sets, affirming the model's accuracy in estimating the probability of early pregnancy loss. The calibration curves (Fig. 2B, D) further demonstrated the high accuracy of the predictions, showing close alignment between expected outcomes and actual occurrences.

A nomogram was constructed incorporating the significant predictive factors identified through multivariable logistic regression. This visual tool allows for the calculation of a cumulative score by summing the points for each parameter, providing a direct measure of the clinical probability of early pregnancy loss (Fig. 3B). In addition, we have provided a supplementary example file (Supplementary example) with a detailed example to facilitate better interpretation and utilization of the nomogram. Decision curve analysis (DCA), depicted in Fig. 3A, confirmed that the model performs well across a range of decision thresholds, supporting its utility in clinical decision-making. This analysis showed that the model is beneficial for clinical use, providing a robust tool for assessing risk and aiding in the management of early pregnancy loss.

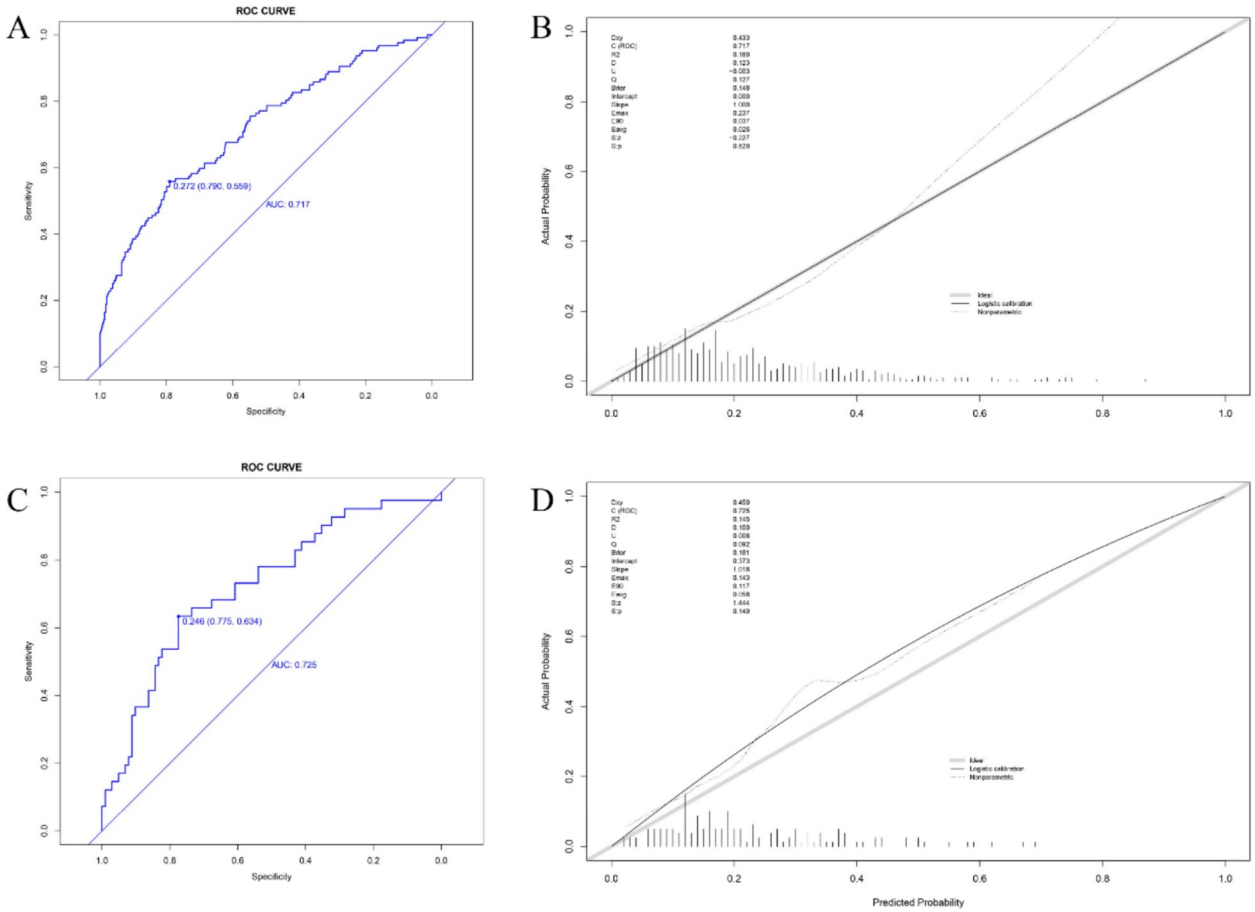

Fig. 2 A, B ROC and Calibration curves for the training set, while C and D depict those for the validation set, highlighting the performance and validation of the logistic regression model

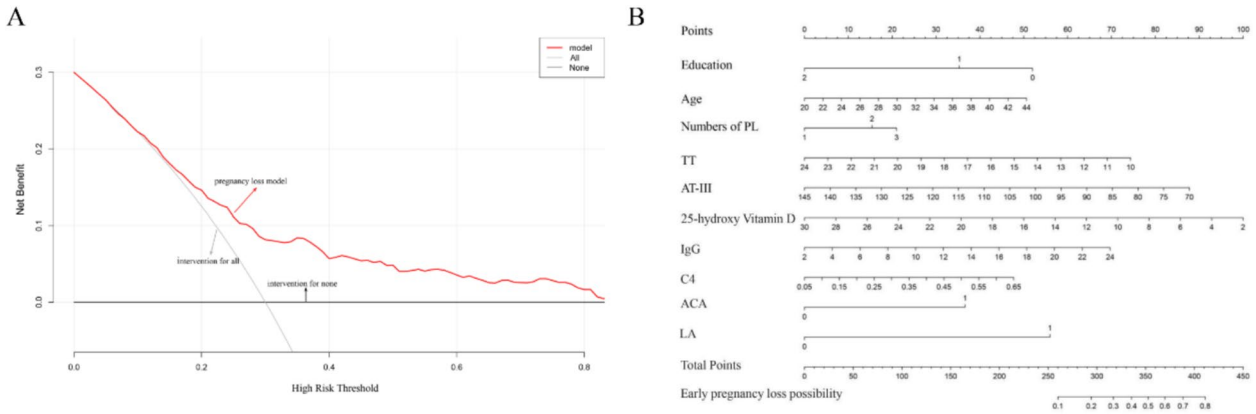

Fig. 3 A Decision curve analysis (DCA), highlighting the clinical benefits of the model across different decision thresholds. (Intervention for all: this line represents the scenario, where the intervention is applied to all patients. Intervention for none: this line represents the scenario, where no intervention is applied to any patients. Pregnancy loss model: this line represents the net benefit of using the model for decision-making). B Nomogram, a graphical representation used to predict individual probabilities of early pregnancy loss

Discussion

The causes of early pregnancy loss remain incompletely understood. While many cases are attributed to chromosomal abnormalities, immunological, anatomical, hormonal, and thrombotic factors are also known to play significant roles. In this study, we developed a comprehensive predictive model to better understand the multifactorial nature of early pregnancy loss. Using a multivariable logistic regression model refined through a backward stepwise approach based on the Akaike Information Criterion (AIC), we identified several key predictors: maternal age, educational levels, thrombin time (TT), antithrombin III (AT-III), D-dimer, vitamin D levels, immunoglobulin G (IgG), complement component C4, anticardiolipin antibodies (ACA), and lupus anticoagulant (LA). In addition, based on clinical experience and widely recognized clinical indicators, the number of previous pregnancy losses was also incorporated as a predictive variable.

The relationship between increasing maternal age and early pregnancy loss is well-documented. The risk associated with maternal age showing a strong biological gradient, where the risk increases with advancing maternal age [18, 19]. Women aged between 20 and 29 year experience the lowest risk of miscarriage, estimated at 12%. This risk steeply increases to 65% in women aged 45 years and older [20]. The increasing risk with age is primarily attributed to the heightened frequency of embryonic chromosomal anomalies, such as trisomies, particularly trisomy 16, which is the most common cause of miscarriage and increases linearly with age from 20 to 40 years [21, 22]. Similarly, the number of previous miscarriages plays a critical role in predicting the risk of subsequent miscarriages. Women with no history of miscarriage have the lowest risk, approximately 11%. However, this risk escalates with each subsequent miscarriage, increasing by about 10% with each event, reaching up to 42% in women who have experienced three or more previous miscarriages [23]. This pattern demonstrates a clear biological gradient and suggests a cumulative effect of previous pregnancy losses on the risk of future losses. In this study, we also found advanced maternal age is associated with a higher risk of pregnancy loss.

In this study, lower educational levels were found to correlate with increased risks of early pregnancy loss in our model. Educational attainment serves as a surrogate marker for socioeconomic status, which influences access to healthcare and health-related behaviors. It is essential to acknowledge the broader implications of educational attainment on women's reproductive health. Education plays a pivotal role in equipping women with the knowledge necessary to navigate the complexities of fertility and the associated risks of miscarriage [24]. Educated women are more likely to seek healthcare advice early and adhere to prenatal care regimens, which can mitigate some risk factors associated with pregnancy loss. Moreover, education can indirectly affect miscarriage rates through its impact on socioeconomic status. Women with higher educational levels often have better access to healthcare resources, including fertility treatments and preventive care, which can decrease the likelihood of pregnancy loss [25, 26]. They are also more likely to delay childbearing as they pursue higher education and careers, which, while increasing the risk of age-related fertility decline, also means that they are better informed about the potential risks and are more likely to access medical advice and interventions promptly. In conclusion, while education itself does not directly cause

or prevent miscarriage, its significant influence on health literacy and access to care positions it as a crucial determinant in managing the risk associated with miscarriage.

Thrombophilia, a disruption in blood clotting processes, is often linked with pregnancy loss [27]. This study identified significant hypercoagulability among women with a history of missed abortions, evidenced by elevated levels of D-dimer and reduced activity of thrombin time (TT) and antithrombin III (AT-III). These findings suggest that abnormal coagulation processes play a critical role in the etiology of pregnancy loss [28, 29]. Elevated D-dimer levels indicate increased fibrin turnover and clot formation, which are atypical in the general population. This marker, a byproduct of fibrin degradation, is crucial for identifying thrombotic conditions that may underpin missed abortions [30]. Furthermore, the observed reduction in AT-III levels—critical for regulating the coagulation cascade—suggests a compromised ability to control clot formation, thereby increasing the risk of complications during pregnancy.AT-III is a crucial regulator of the coagulation cascade, inhibiting several enzymes of the system, including thrombin [31]. Lower levels of AT-III suggest a diminished capacity to regulate clot formation effectively, potentially leading to an increased risk of clot-related complications during pregnancy [32]. The implications of these findings are profound, suggesting that screening for coagulation abnormalities could be crucial in early pregnancy, especially among women with a history of pregnancy loss.

Immunological factors are pivotal in pregnancy maintenance, where abnormal immune responses can lead to fetal rejection. Elevated levels of IgG and C4 in this study suggest an autoimmune component to early pregnancy loss. IgG, which includes various subclasses, has been implicated in autoimmune responses that may compromise pregnancy. To some extent, elevated levels of specific IgG antibodies can induce a state of heightened immune response against the trophoblast, leading to increased risk of miscarriage [33]. Complement C4, part of the classical pathway of the complement system, is another crucial immunological factor associated with PL. Elevated levels of C4 may reflect an overactive immune response that can lead to fetal–placental unit damage [34]. The study indicated that high levels of C4 are correlated with increased pregnancy loss, suggesting that excessive activation of the complement system might directly affect the integrity and function of the placental bed, leading to miscarriage [34]. Clinically, these findings underscore the potential for targeted interventions that modulate these immune responses. Treatments that regulate the activity of specific IgG antibodies or control the activation of the complement system might help reduce the incidence of PL in high-risk patients.

The presence of ACA and LA as markers of APS has been robustly linked to the increased risk of pregnancy loss in our model. These antibodies not only enhance the risk of forming blood clots but also affect the placental environment by disrupting the normal immune tolerance required during pregnancy [35, 36]. This immune-mediated attack is presumed to cause placental insufficiency, which is a significant pathophysiological mechanism leading to miscarriage. The placenta, being the critical interface for nutrient and gas exchange between the mother and fetus, when compromised, directly impacts fetal survival. The presence of these antibodies suggests an aberrant immune response, where instead of supporting pregnancy, the immune system acts against it, leading to pregnancy complications or losses [37]. Furthermore, the data from the literature and findings from this study suggest that LA and ACA can be predictive of adverse pregnancy outcomes [38, 39]. Monitoring these antibodies in women with a history of pregnancy losses could guide the implementation of targeted interventions, such as anticoagulation therapies, which have been shown to improve pregnancy outcomes in women with APS. In conclusion, the role of ACA and LA in mediating adverse pregnancy outcomes through mechanisms involving immune dysfunction and placental insufficiency underlines the importance of screening for APS in women with recurrent pregnancy losses.

Vitamin D is essential for the immune system and placental function. Vitamin D's influence on pregnancy outcomes can be linked to its crucial role in the immune system and the modulation of placental biology [40]. The enzyme responsible for activating vitamin D, CYP27B1, is expressed in the maternal decidua and fetal trophoblasts early in pregnancy, indicating that vitamin D could be integral to successful implantation and early pregnancy maintenance [41]. Deficiencies in vitamin D might lead to insufficient immunological adaptation during pregnancy, potentially increasing the risk of rejection of the fetal allograft. Furthermore, vitamin D is vital for the proper functioning of the placenta [42].It influences trophoblast invasion and placental spiral artery remodeling—critical processes for establishing adequate blood flow to the growing fetus [43]. Abnormalities in these processes are often observed in miscarriages, suggesting that optimal vitamin D levels could support proper placental development and function, thus reducing miscarriage risks [44]. This aligns with findings from various studies that correlate higher pre-pregnancy and early pregnancy vitamin D levels with lower rates of miscarriage [45]. Our results support the notion that maintaining adequate vitamin D levels during pregnancy is critical for reducing the risk of miscarriage.

The predictors identified in this study provide crucial insights into the etiology of early pregnancy loss and suggest targeted interventions that could potentially mitigate this risk. Future research should focus on prospective studies to validate these findings and explore the effectiveness of interventions based on individual risk profiles determined by our predictive model.

Strengths

This study presents several strengths that contribute to the development of a predictive model for early pregnancy loss. First, the use of a logistic regression-based model ensures a robust and interpretable approach to predicting subsequent pregnancy loss in women with a history of pregnancy loss. This model allows for the integration of multiple clinical and laboratory variables, providing a comprehensive risk assessment. Second, the model's development was rigorously validated using internal validation techniques, which enhances its reliability. Furthermore, the study provides an in-depth analysis of various predictive factors, some of which have not been fully explored in previous research, thus contributing valuable insights to the field. Finally, the model's potential clinical utility is demonstrated through the nomogram, offering a user-friendly tool for clinicians to assess individual patient risks and make more informed decisions.

Limitation

This study has several limitations that need to be acknowledged. First, its retrospective design and the involvement of a single center limit the generalizability of the findings. The study population was exclusively Chinese, and the inclusion of both consecutive and non-consecutive pregnancy losses may further restrict the applicability of the model to other populations or settings. Another key limitation is the lack of external validation. To ensure the model's real-world applicability, it will need to be validated in diverse populations and across multiple centers. Finally, missing data affected the study, with 25% of the original cohort excluded due to incomplete information, potentially introducing selection bias.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40001-025-02361-5.

Supplementary material 1

Supplementary material 2

Author contributions
The study conception and design were performed by N.D and F.W. Material preparation, data collection, and analysis were performed by XP. W, N.D and PL. W. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding
This work was funded by the Science Foundation of Lanzhou University (Grant No. 071100132 and 071100186), the Medical Innovation and Development Project of Lanzhou University (Grant No. lzuyxcx-2022-137), and the Science Foundation of Lanzhou University Second Hospital (Grant No. YJS-BD-19).

Availability of data and materials
The data that support the findings of this study are available on request from the corresponding author. Researchers who are interested in working together on our study are more than welcome to collaborate. Contact the paper's corresponding author, Fang Wang, ery_fwang@lzu.edu.cn.

Declarations

Ethics approval and consent to participate
The study received approval from the ethical committee of Lanzhou University (2019A-231). All subjects gave informed consent before participation. All methods were performed in accordance with the Declaration of Helsinki and the relevant regulations.

Competing interests
The authors declare no competing interests.

Author details
[1]Reproductive Medicine Center, The Second Hospital of Lanzhou University, No.82, Cuiying Road, Chengguan District, Lanzhou, Gansu, China.

Received: 25 September 2024 Accepted: 5 February 2025
Published online: 12 February 2025

References

1. Bender Atik R, et al. ESHRE guideline: recurrent pregnancy loss. Hum Reprod Open. 2018;2018(2):004.
2. Zhao QY, et al. Decidual macrophages in recurrent spontaneous abortion. Front Immunol. 2022;13: 994888.
3. Quenby S, et al. Miscarriage matters: the epidemiological, physical, psychological, and economic costs of early pregnancy loss. Lancet. 2021;397(10285):1658–67.
4. Dimitriadis E, et al. Recurrent pregnancy loss. Nat Rev Dis Primers. 2020;6(1):98.
5. Coomarasamy A, et al. Recurrent miscarriage: evidence to accelerate action. Lancet. 2021;397(10285):1675–82.
6. Robinson GE. Pregnancy loss. Best Pract Res Clin Obstet Gynaecol. 2014;28(1):169–78.
7. Pinar MH, et al. Early pregnancy losses: review of nomenclature, histopathology, and possible etiologies. Fetal Pediatr Pathol. 2018;37(3):191–209.
8. Yu N, Kwak-Kim J, Bao S. Unexplained recurrent pregnancy loss: Novel causes and advanced treatment. J Reprod Immunol. 2023;155: 103785.
9. Tsonis O, et al. Management of recurrent miscarriages: an overview of current evidence. Curr Opin Obstet Gynecol. 2021;33(5):370–7.
10. Bardos J, et al. A national survey on public perceptions of miscarriage. Obstet Gynecol. 2015;125(6):1313–20.
11. Mu F, et al. Predicting risk of subsequent pregnancy loss among women with recurrent pregnancy loss: an immunological factor-based multivariable model. Am J Reprod Immunol. 2024;91(3): e13837.
12. Sakthiswary R, et al. Antinuclear antibodies predict a higher number of pregnancy loss in unexplained recurrent pregnancy loss. Clin Ter. 2015;166(2):e98-101.
13. Casikar I, et al. Prediction of successful expectant management of first trimester miscarriage: development and validation of a new mathematical model. Aust N Z J Obstet Gynaecol. 2013;53(1):58–63.

14. Ohori Tatsuo G, et al. Nomogram as predictive model in clinical practice. Gan To Kagaku Ryoho. 2009;36(6):901–6.
15. Bai YY, et al. Reference intervals of systemic immune-inflammation index, neutrophil-to-lymphocyte ratio, lymphocyte-to-monocyte ratio, and platelet-to-lymphocyte ratio during normal pregnancy in China. Eur Rev Med Pharmacol Sci. 2023;27(3):1033–44.
16. Johansen MB, Christensen PA. A simple transformation independent method for outlier definition. Clin Chem Lab Med. 2018;56(9):1524–32.
17. D'Aloisio AA, et al. Prenatal and infant exposures and age at menarche. Epidemiology. 2013;24(2):277–84.
18. Female age-related fertility decline. Committee Opinion No 589. Fertil Steril, 2014. **101**(3): 633–4.
19. Sauer MV. Reproduction at an advanced maternal age and maternal health. Fertil Steril. 2015;103(5):1136–43.
20. Kling C, Hedderich J, Kabelitz D. Fertility after recurrent miscarriages: results of an observational cohort study. Arch Gynecol Obstet. 2018;297(1):205–19.
21. Hardy K, et al. Temporal changes in chromosome abnormalities in human spontaneous abortions: results of 40 years of analysis. Am J Med Genet A. 2016;170(10):2671–80.
22. Stephenson MD, Awartani KA, Robinson WP. Cytogenetic analysis of miscarriages from couples with recurrent miscarriage: a case-control study. Hum Reprod. 2002;17(2):446–51.
23. Coomarasamy A, et al. Micronized vaginal progesterone to prevent miscarriage: a critical evaluation of randomized evidence. Am J Obstet Gynecol. 2020;223(2):167–76.
24. Tong F, et al. The epidemiology of pregnancy loss: global burden, variable risk factors, and predictions. Hum Reprod. 2024;39(4):834–48.
25. Kramer MS, et al. Socio-economic disparities in pregnancy outcome: why do the poor fare so poorly? Paediatr Perinat Epidemiol. 2000;14(3):194–210.
26. Stephansson O, et al. The influence of socioeconomic status on stillbirth risk in Sweden. Int J Epidemiol. 2001;30(6):1296–301.
27. Lu X, et al. Prothrombotic state of patients with unexplained recurrent spontaneous abortion. Int J Gynaecol Obstet. 2015;131(2):161–5.
28. Chen D, et al. Coagulation status in women with a history of missed abortion. Reprod Sci. 2024;31(2):480–7.
29. Wang T, et al. Prediction of thrombophilia in patients with unexplained recurrent pregnancy loss using a statistical model. Int J Gynaecol Obstet. 2017;138(3):283–7.
30. Yao H, Ji Y, Zhou Y. Analysis of blood coagulation indexes, thromboelastogram and autoantibodies in patients with recurrent pregnancy loss. Pak J Med Sci. 2022;38(7):2005–10.
31. Sharma A, et al. Polymorphisms in factor V and antithrombin III gene in recurrent pregnancy loss: a case-control study in Indian population. J Thromb Thrombolysis. 2015;39(4):481–8.
32. Wang P, et al. Predictive value of thromboelastography parameters combined with antithrombin III and D-Dimer in patients with recurrent spontaneous abortion. Am J Reprod Immunol. 2019;82(4): e13165.
33. Opatrny L, et al. Association between antiphospholipid antibodies and recurrent fetal loss in women without autoimmune disease: a metaanalysis. J Rheumatol. 2006;33(11):2214–21.
34. Sugiura-Ogasawara M, et al. Complement as a predictor of further miscarriage in couples with recurrent miscarriages. Hum Reprod. 2006;21(10):2711–4.
35. Gaigne L, et al. Anti-cytokine autoantibodies: review of the literature. Rev Med Interne. 2022;43(9):528–36.
36. Abramson J, Husebye ES. Autoimmune regulator and self-tolerance—molecular and clinical aspects. Immunol Rev. 2016;271(1):127–40.
37. Schreiber K, Hunt BJ. Managing antiphospholipid syndrome in pregnancy. Thromb Res. 2019;181(Suppl 1):S41-s46.
38. Schreiber K, Hunt BJ. Pregnancy and antiphospholipid syndrome. Semin Thromb Hemost. 2016;42(7):780–8.
39. de Assis V, Giugni CS, Ros ST. Evaluation of recurrent pregnancy loss. Obstet Gynecol. 2024;143(5):645–59.
40. Tamblyn JA, et al. Vitamin D and miscarriage: a systematic review and meta-analysis. Fertil Steril. 2022;118(1):111–22.
41. Gonçalves DR, et al. Recurrent pregnancy loss and vitamin D: a review of the literature. Am J Reprod Immunol. 2018;80(5): e13022.
42. Schröder-Heurich B, Springer CJP, von Versen-Höynck F. Vitamin D effects on the immune system from periconception through pregnancy. Nutrients. 2020. https://doi.org/10.3390/nu12051432.
43. Xu QH, et al. Altered vitamin D metabolism is involved in the dysregulation of γδT cell function and their crosstalk with trophoblasts in recurrent pregnancy loss. Am J Reprod Immunol. 2023;89(6): e13581.
44. Hou W, et al. Decreased serum vitamin D levels in early spontaneous pregnancy loss. Eur J Clin Nutr. 2016;70(9):1004–8.
45. Yan X, et al. Decreased expression of the vitamin D receptor in women with recurrent pregnancy loss. Arch Biochem Biophys. 2016;606:128–33.

Publisher's Note