Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 8, No. 3, 2026

Prediction of The Earliest Detection Time of Fetal DNA Concentration in Noninvasive Prenatal Testing Based on Multivariate and Logistic Regression Model

Jia Zhou*, Huipeng Li

School of Software, Pingdingshan University, Pingdingshan 467000, China
* **Corresponding author**: Jia Zhou (Email: zhou45121@gmail.com)

Abstract: Fetal DNA concentration (FF) is a critical indicator affecting the accuracy and stability of non-invasive prenatal testing (NIPT). An excessively low FF may lead to test failure or false negatives. To address the lack of quantitative prediction for the earliest safe testing time at which clinical thresholds (FF ≥4%) are met, this study constructed multivariate linear regression and Logistic regression models based on publicly available sample data. The models analyzed the influence of gestational age and maternal body mass index (BMI) on FF levels and the probability of meeting the threshold, and introduced a gestational age quadratic term to establish a nonlinear model. The results demonstrated that BMI was a significant negative predictor of both FF and the probability of meeting the threshold. Further, the study derived the earliest gestational age at which the probability of meeting the threshold ≥95% for different BMI stratifications and validated model stability using the Bootstrap method. This research provides a quantitative basis for individualized NIPT testing timing.

Keywords: Non-Invasive Prenatal Testing, Fetal DNA Concentration, Logistic Regression, BMI Stratification Analysis.

1. Introduction

Non-Invasive Prenatal Testing (NIPT) is an important technique for screening chromosomal abnormalities based on cell-free fetal DNA (cffDNA) extracted from the peripheral blood of pregnant women. Fetal Fraction (FF), defined as the proportion of fetal-derived cell-free DNA in the total cell-free DNA in maternal circulation, serves as a core indicator that determines the accuracy and stability of NIPT results. It is generally acknowledged that NIPT results are highly reliable when the fetal DNA fraction is no less than 4%; conversely, an excessively low FF may increase the risk of assay failure or false-negative outcomes, which can severely compromise the clinical utility of non-invasive prenatal screening[1-3].Previous studies have consistently demonstrated that fetal DNA fraction shows a positive correlation with gestational age, meaning that FF levels gradually rise as pregnancy progresses. Meanwhile, multiple lines of evidence have confirmed a significant negative correlation between FF and maternal Body Mass Index (BMI), indicating that overweight and obese women tend to have lower fetal DNA concentrations in their peripheral blood[4-6]. Despite these well-established associations, most existing research remains limited to descriptive correlation analyses between FF and its influencing factors. In particular, current literature lacks a quantitative prediction model for the earliest safe testing time at which the effective detection threshold can be stably achieved. Furthermore, few studies have comprehensively accounted for individual differences across various maternal BMI categories, leading to a shortage of personalized guidance for clinical testing timing[7-8]. To fill these critical gaps, the present study was designed and conducted based on publicly available clinical sample datasets. In this research, we constructed multivariate regression models and Logistic regression models to systematically investigate the independent and combined effects of gestational age and maternal BMI on fetal DNA fraction levels and the probability of reaching the clinically acceptable threshold. By incorporating these predictive models, we aimed to accurately estimate the earliest gestational age at which the probability of achieving FF≥4% reaches at least 95% for different BMI stratifications. In addition, appropriate statistical methods were adopted to validate the reliability and robustness of the established models. The findings of this study are expected to provide a solid quantitative and statistical basis for formulating individualized NIPT testing strategies, optimizing the timing of non-invasive prenatal screening for pregnant women with different BMI levels, reducing the rate of sample rejection due to insufficient fetal DNA concentration, and further improving the accuracy and clinical application value of non-invasive prenatal testing.

2. Data sources and methods

2.1. Data sources

The research data was sourced from the China College Students Mathematical Modeling Competition Public Data Platform (https://www.mcm.edu.cn/html_cn/node/03c91a444e62eee81a3740fa97a461a6.html). The original data included over 1,000 male fetal samples and over 600 female fetal samples. This study only selected male fetal samples for modeling analysis.

2.2. Descriptive statistics

This study first conducted descriptive statistical analysis on gestational age, BMI, and FF. Continuous variables were expressed as mean ± standard deviation to preliminarily understand the distribution characteristics of the sample. The calculated results of these characteristics are presented in Table 1 below.

Sample average:

$$\bar{x} = \frac{1}{n}\sum_{i=0}^{n} x_i \quad (1)$$

Sample standard deviation:

$$s^2 = \frac{1}{n}\sum_{k=0}^{n}(x_i - \bar{x})^2 \quad (2)$$

Table 1. Basic characteristics of the sample

Indicator	Count	Mean	Std	Min	25%	50%	75%	Max
GA_weeks (Gestational Age)	896.00	15.46	2.83	11.00	13.14	15.14	17.00	21.86
Maternal BMI	896.00	31.99	2.58	20.70	30.11	31.64	33.59	39.90
FF (Y Chromosome Concentration, %)	896.00	7.50	3.12	1.00	5.00	7.33	9.66	22.85

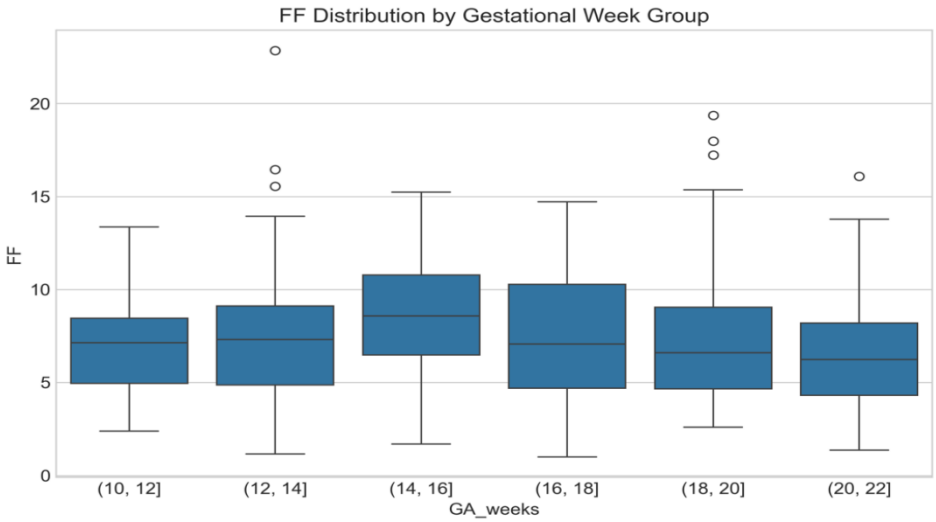

Figure 1. FF distribution by gestational week groups

Combined with the data in Figure 1, it can be seen that the FF fluctuates with the gestational age.

2.3. Correlation analysis

Pearson correlation coefficient was used to analyze the correlation between FF and gestational age, as well as the correlation between FF and BMI. Pearson

Correlation coefficent:

$$r=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}} \quad (3)$$

Test of significance:

$$t=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad (4)$$

Table 2. Results of Correlation Analysis

	GA_weeks	BMI	FF
GA_weeks	1.000	0.087	-0.026
BMI	0.087	1.000	-0.156
FF	-0.026	-0.156	1.000

The results in Table 2 demonstrate a significant negative correlation between BMI and FF（P<0.001）.

2.4. Multivariate regression analysis

In case:

$$W=\text{Gestational age, } B=\text{BMI} \quad (5)$$

Establish a linear regression model:

$$FF_i=\beta_0+\beta_1\times W_i+\beta_2\times B_i+\varepsilon_i \quad (6)$$

Where β_0 is the constant term, β_1 and β_2 are regression coefficients, and ε_i is the random error.

Suppose:

$$\varepsilon_i \sim N(0,\sigma^2) \quad (7)$$

Least squares estimation:

$$\hat{\beta}=(X^TX)^{-1}X^TY \quad (8)$$

Regression significance test:

$$t=\frac{\hat{\beta}j}{SE(\hat{\beta}j)} \quad (9)$$

Coefficient of determination, determination coefficient:

$$R^2=1-\frac{SS_{res}}{SS_{tot}} \quad (10)$$

Adjust R^2:

$$R^2_{adj}=1-\frac{(1-R^2)(n-1)}{n-k-1} \quad (11)$$

The linear regression results obtained from the above formulas are shown in Table 3.

Table 3. Linear Regression Results

Variable	Coefficient (coef)	Std. Error	t-value	P-value (P>|t|)	95% CI [Lower]	95% CI [Upper]
Constant (const)	13.7271	1.365	10.054	0.000	11.047	16.407
Gestational Age (GA)	-0.0135	0.037	-0.368	0.713	-0.086	0.058
Maternal BMI	-0.1883	0.040	-4.678	0.000	-0.267	-0.109

Quadratic nonlinear model

To improve the model fitting effect, a quadratic term of gestational age was further incorporated to construct a nonlinear model:

$$FF=\beta_0+\beta_1 W+\beta_2 B+\beta_3\times W^2+\varepsilon \quad (12)$$

Where β_0 is the constant term, β_1, β_2, β_3 are regression coefficients, and ε is the random error.

Table 4. Results of Nonlinear Models

Variable	Coefficient (coef)	Std. Error	t-value	P-value (P>|t|)	95% CI [Lower]	95% CI [Upper]
Constant (const)	-1.8029	3.583	-0.503	0.615	-8.835	5.229
Gestational Age (GA)	1.9830	0.428	4.631	0.000	1.143	2.823
GA Squared (GA_sq)	-0.0617	0.428	-4.679	0.000	-0.088	-0.036
Maternal BMI	-0.1913	0.040	-4.806	0.000	-0.269	-0.113

According to the calculation results in Table 4:

(1) The pregnancy week was significantly positive for a single item($P<0.001$).

(2) The second term of gestational age was significantly negative($P<0.001$).

(3) BMI remains a significant negative factor($P<0.001$).

The FF shows an inverted U-shaped trend with gestational weeks.

Combining with Figure 2, it can be analyzed that the adjusted R^2 of the nonlinear model increased from 0.022 to 0.045, indicating superior fitting performance compared to the linear model.

Comparison of model goodness using AIC:

$$AIC=2k-2\ln(L) \quad (13)$$

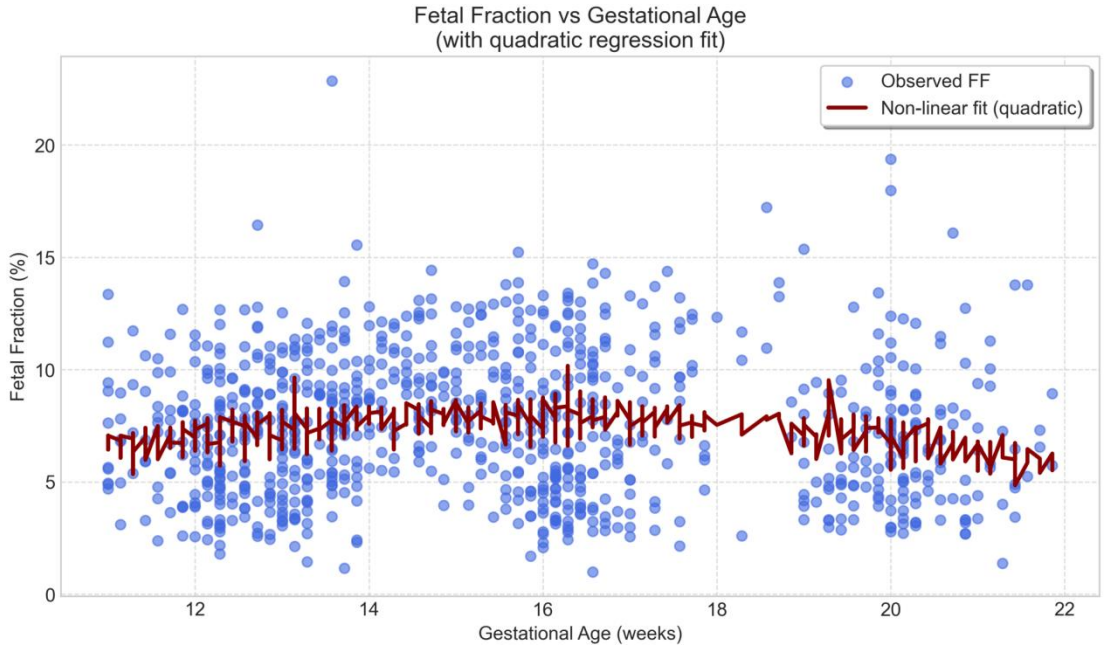

Figure 2. Scatter plot of FF versus gestational weeks and nonlinear fitting curve

2.5. Probability model of success rate

Construct a Logistic regression model to analyze the probability of FF≥4%. Let P be the probability of FF≥4%, then:

$$P=\frac{1}{1+\exp[-(\theta_0+\theta_1 W+\theta_2 B)]} \quad (14)$$

When $P\geq 0.95$, determine the minimum gestational age as the earliest safe testing time.

Table 5. Multivariate logistic regression analysis of BMI and gestational age in pregnant women

Variable	Coefficient (coef)	Std. Error	z-value	P-value (P>|z|)	95% CI [Lower]	95% CI [Upper]
Constant (Intercept)	5.6773	1.262	4.499	0.000	3.204	8.150
Gestational Age (GA)	-0.0375	0.034	-1.116	0.265	-0.103	0.028
Maternal BMI	-0.1021	0.037	-2.782	0.005	-0.174	-0.030

The Logistic regression results in Table 5 indicate that BMI is a significantly negative predictor ($P=0.005$).

2.6. BMI fractimal analysis

Grouping according to the WHO recommended criteria:

(1) Low body weight: BMI < 18.5

(2) Normal type: 18.5 ≤ BMI < 24

(3) Overweight: 24 ≤ BMI < 28

(4) Corpulence: BMI ≥ 28

The results and trends of the grouping are shown in Table

6 and Figure 3 below.

Table. 6. Earliest gestational week with FF≥4% probability≥95%

BMI Group	Representative BMI	Earliest GA (weeks)
Normal (18.5–23.9)	21.0	15.7–16.7
Overweight (24–27.9)	26.0	10.0–11.0
Obese I (28–34.9)	32.0	10.0–11.0
Obese II (≥35)	37.0	10.0–11.0

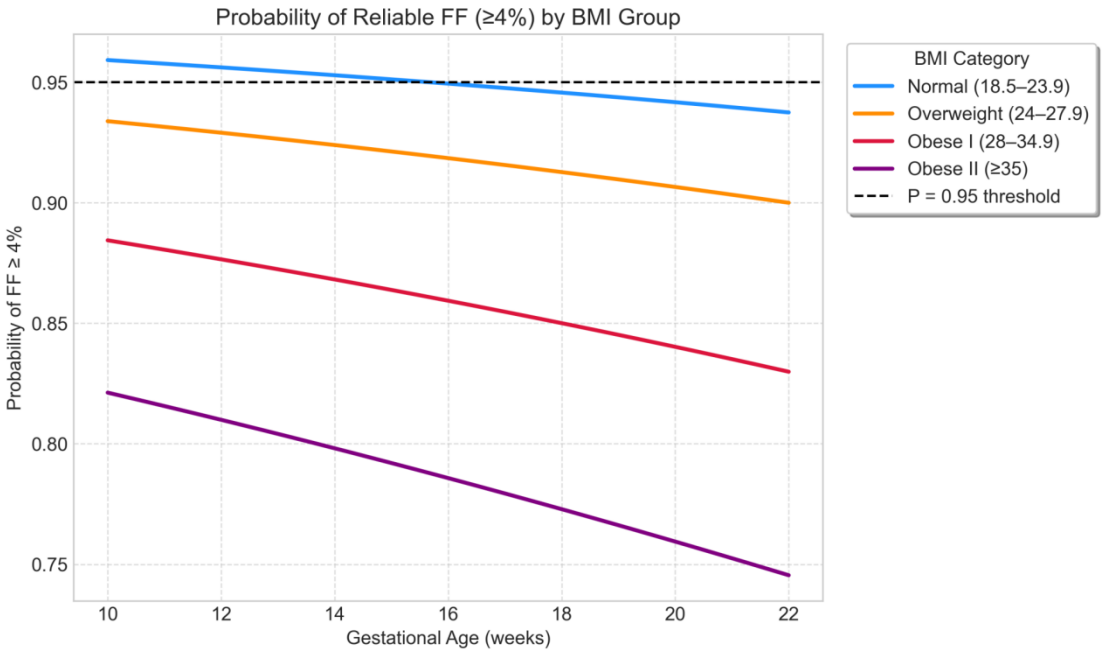

Figure 3. Achievement probability curves for different BMI groups

2.7. Stability analysis

Bootstrap method was used for 1000 repeated sampling to observe the fluctuation of regression coefficients. Meanwhile, the measurement error of ±0.3% was simulated, and the compliance time was recalculated.

Bootstrap Standard error:

$$SE_{boot}=\sqrt{\frac{1}{B-1}\sum(\beta_{(b)}-\beta)^2} \quad (15)$$

Table 7. Stability Sampling Table

	Intercept	GA_weeks	BMI
0.025	2.4015	-0.0977	-0.1883
0.975	8.6171	0.0346	-0.0090

The results in Table 7 indicate that the BMI coefficient remains stable, while the gestational age coefficient exhibits significant fluctuations.

3. Result

3.1. Describe statistical results

In the male fetus samples, the mean gestational age was 15.46 weeks, the mean BMI was 31.99 kg/m², and the mean FF was 7.50%.

3.2. Results of correlation analysis

(1) There was no significant correlation between FF and gestational age. (r=-0.026, P=0.443).

(2) There is a significant negative correlation between FF and BMI. (r=-0.156, P<0.001).

3.3. Regression analysis results

Multivariate regression analysis revealed:

(1) Gestational age as a significant positive influencing factor (P<0.001);

(2) BMI was a significantly negative influencing factor (P<0.001).

The inclusion of the second-term of gestational age improved the model's adjusted R^2 from 0.022 to 0.045, indicating a non-linear trend (inverted U-shape) in FF with gestational age, where the nonlinear model outperforms the linear model. Bootstrap replication demonstrated a narrow fluctuation range for BMI regression coefficients (95% CI [-0.1883, -0.0090]), confirming stability. The gestational age coefficient's confidence interval containing zero suggests its influence is unstable in this sample. With a simulated measurement error of ±0.3%, the recommended testing interval variation should not exceed 0.5 weeks, demonstrating the model's robustness against measurement noise.

4. Conclusion

This study systematically analyzed the effects of gestational age and BMI on fetal free DNA (FF) concentration in male fetuses using multivariate regression and Logistic models based on publicly available sample data, and established a prediction model for FF attainment time. The results demonstrated that BMI was a significant negative predictor of FF (r=-0.156, P<0.001; regression coefficient approximately-0.19, P<0.001), indicating that higher maternal body mass index (BMI) was associated with lower

fetal FF concentration, consistent with clinically established patterns. The nonlinear model revealed a trend of gestational age variation on FF, with better fitting than the linear model (adjusted R^2 increased from 0.022 to 0.045), suggesting that FF may exhibit non-uniform changes with gestational age. Logistic regression further confirmed that BMI was an independent negative predictor of FF ≥4% attainment (P=0.005), and the model stability was well validated by Bootstrap. Based on the above analysis, this study provides reference recommendations for NIPT sampling timing in high-BMI populations: normal-weight pregnant women may be sampled around 15–16 weeks, while overweight and obese pregnant women are advised to delay sampling until after 15 weeks to effectively reduce the risk of test failure due to low FF and improve detection reliability. This study clearly demonstrated the dominant negative effect of BMI on FF in high-BMI samples, offering clinical guidance. Future research could expand the proportion of low-BMI samples and incorporate additional influencing factors for validation to refine individualized testing strategies. The individualized detection recommendations proposed in this study based on BMI stratification can help reduce the detection failure rate caused by low fetal concentration and improve the clinical application efficiency of NIPT. Future research could incorporate more clinical variables for model optimization and conduct multicenter validation studies to further enhance the generalizability and practicality of the model.

References

[1] Deng Chenxi, Xu Jing, Li Xiaomei, et al. Analysis of influencing factors of fetal free DNA concentration in non-invasive prenatal testing for pregnant women in China [J]. China Maternal and Child Health, 2022, 37(15): 2785-2789.

[2] Jiang Y, Zhu X, Wang L, et al. Factors influencing fetal fraction in noninvasive prenatal testing: a systematic review and meta-analysis[J]. Frontiers in Pediatrics, 2022, 10: 845714.

[3] Qiao L, Wang J, Zhang L, et al. The influence of maternal characteristics on fetal fraction in noninvasive prenatal testing[J]. American Journal of Translational Research, 2019, 11(12): 7665-7673.

[4] Wang E, Hou Y, Li J, et al. Maternal obesity and low fetal fraction in noninvasive prenatal testing: a retrospective cohort study[J]. BMC Pregnancy and Childbirth, 2021, 21(1): 456.

[5] Liu F, Zhang W, Chen L, et al. Effect of maternal BMI on fetal free DNA concentration in non-invasive prenatal testing [J]. Chinese Journal of Obstetrics and Gynecology, 2023, 58(4): 289-294.

[6] Zhang H, Gao Y, Li Q, et al. Gestational age and maternal BMI as predictors of fetal fraction in cell-free DNA-based prenatal testing: a large-scale Chinese cohort study[J]. Prenatal Diagnosis, 2020, 40(11): 1425-1433.

[7] Zhao Jing, Wang Xiaoyan, Li Na, et al. Discussion on the influencing factors and clinical significance of low fetal DNA concentration in non-invasive prenatal testing [J]. China Journal of Eugenics and Genetics, 2024, 32(2): 145-150.

[8] Hou Y, Wang E, Chen L, et al. Individualized timing of noninvasive prenatal testing based on fetal fraction prediction models: a retrospective analysis[J]. Journal of Maternal-Fetal & Neonatal Medicine, 2023, 36(1): 2187075.