World Journal of Information Technology
Print ISSN: 2959-9903
Online ISSN: 2959-9911
DOI: https://doi.org/10.61784/wjit3067

THE OPTIMIZATION PROBLEM FOR TIME-POINT DETECTION IN NIPT BASED ON MULTI-MODEL FUSION

JiaJun Tang
School of Automation, Nanjing University of Science and Technology, Nanjing 210000, Jiangsu, China.

Abstract: This article focuses on four major issues: correlation analysis of Y chromosome concentration in male fetuses, BMI grouping and determination of optimal detection time points, optimization of time points under multiple factors, and determination of chromosomal abnormalities in female fetuses. Firstly, based on the principle of NIPT technology as the accurate benchmark for results, we analyzed the correlation characteristics between Y chromosome concentration and gestational age, BMI, constructed an adapted relationship model, and verified its significance; Secondly, with the goal of minimizing potential risks, the optimal NIPT timing for male pregnant women with different BMI ranges is determined through reasonable grouping methods, and the impact of detection errors is analyzed; Furthermore, by integrating multiple factors such as height, weight, and age, the time point model is optimized to ensure the reliability of the results; Finally, based on the characteristic of female fetuses without Y chromosome, combined with indicators such as Z value and GC content of X chromosome and chromosomes 21, 18, and 13, an abnormality determination method is established. The research results can provide scientific support for optimizing clinical NIPT testing conditions, improving accuracy, and reducing potential risks for pregnant women, meeting the practical needs of early and accurate detection.This study first preprocessed male fetal data, screened 10-25 week samples, converted gestational weeks into continuous values, and processed duplicate samples and outliers. Use Q-Q chart to test normality and determine the conclusion that Y chromosome concentration, gestational age, and BMI are not normally distributed. Through Spearman rank correlation quantification, it was found that gestational age is strongly positively correlated with Y chromosome concentration, while BMI is moderately negatively correlated with Y chromosome concentration. Construct a quadratic polynomial model for OLS fitting, confirm no severe collinearity through VIF test, and then verify the significance of the model through F-test and t-test. In addition, analysis shows that weight has a significant negative impact on Y chromosome concentration, and this model can reliably reflect the relationship between Y chromosome concentration and gestational age, BMI.Therefore, this study is of great significance. By constructing a multi-model framework that integrates statistical analysis and machine learning, it deepens our understanding of the dynamic patterns of fetal cell-free DNA and promotes cross-disciplinary innovation between bioinformatics and clinical medicine. Meanwhile, this study also has significant practical value by precisely identifying key factors affecting detection accuracy, optimizing the timing of individualized testing, effectively improving the sensitivity and specificity of NIPT, and reducing the risks of missed and incorrect diagnoses, thereby providing data support and a decision-making basis for developing scientific, safe, and efficient prenatal screening strategies in clinical practice. Furthermore, the research findings respond to the urgent national needs of the “Healthy Birth, Healthy China” strategy for a high-precision birth defect prevention and control system, facilitating the shift of prenatal screening from being “experience-driven” to “evidence-driven,” enhancing public health services, and alleviating the medical burden on families and society.
Keywords: Spearson rank correlation coefficient; Kaplan Meier survival analysis; Lasso model; XGBoost model

1 INTRODUCTION

Non invasive prenatal testing (NIPT) is an important screening technology in modern prenatal medicine. This technology collects peripheral blood from pregnant women and detects fetal free DNA fragments to achieve early detection of fetal chromosomal abnormalities. The core goal is to timely identify fetal health risks to ensure maternal and infant safety. Among them, the clinical common Down syndrome, Edwards syndrome, and Patao syndrome are directly related to the abnormal proportion of fetal free DNA fragments (chromosome concentration) on chromosomes 21, 18, and 13. The accuracy of NIPT testing results largely depends on the determination of fetal sex chromosome concentration.

From the perspective of clinical risk control, the detection time of fetal abnormalities is crucial for subsequent interventions. Early detection within 12 weeks can significantly reduce the risk of shortened treatment window; A significant increase in risk was observed between weeks 13 and 27; Late detection after 28 weeks will face extremely high risks; Therefore, the reasonable selection of NIPT testing timing is of great significance for risk control. Research and practice have shown that the concentration of Y chromosome in male fetuses is influenced by both the gestational age and body mass index (BMI) of the pregnant woman[1]. Although clinical practice divides pregnant women into different intervals based on BMI to determine the detection time point, there are significant differences in age, pregnancy status, and other aspects among pregnant women. If simple empirical grouping and unified detection time point settings are used, it will lead to a decrease in the detection accuracy of some pregnant women, thereby increasing the potential risk of fetal abnormalities not being detected in a timely manner. Therefore, it is urgent to study the

reasonable timing of NIPT testing for different individual pregnant women and improve the accuracy of testing results to ensure the safety and health of mother and baby.

This article studied the correlation characteristics between fetal Y chromosome concentration and various indicators such as gestational weeks and BMI of pregnant women, constructed a reasonable relationship model, and tested the significance of the model and its related variables through statistical methods, laying the foundation for subsequent research. Taking male pregnant women as the research object, combined with the influence of BMI on the earliest time of Y chromosome concentration reaching 4%, and achieving reasonable grouping of pregnant women's BMI, determining the BMI interval of each group and the optimal NIPT testing time point to minimize potential risks, and finally analyzing the interference of testing errors on the results. Taking into account various factors such as height, weight, and age of pregnant women, as well as detection errors and the proportion of Y chromosome concentration that meets the standard, male fetal pregnant women were grouped based on their BMI to determine the optimal NIPT timing for each group to reduce potential risks, and to explore the impact of detection errors. For pregnant women with female fetuses, a comprehensive method for determining chromosomal abnormalities in female fetuses is established based on the results of aneuploidy determination of chromosomes 21, 18, and 13, taking into account factors such as the Z value and GC content of the X chromosome and the aforementioned chromosomes. The mathematical model and analysis results established in the first three questions are scientifically utilized.

2 ANALYSIS BASED ON NORMALITY TEST CORRELATION

2.1 Data Normality Test

2.1.1 Establishment of Q-Q diagram model

Q-Q plot, also known as quantile quantile plot, plots the quantiles of the sample data against the quantiles of the theoretical normal distribution, and then uses the diagonal as a reference. If the data points are close to the diagonal, the sample data follows a normal distribution; otherwise, it does not follow a normal distribution. In this article, it is used to visually test the normality of the three core factors of Y chromosome concentration, gestational age, and BMI (all three data points deviate from the diagonal), clarify the non normal characteristics, and provide theoretical basis for selecting Spearman rank correlation and nonlinear regression models in the following text[2].

2.1.2 Draw graphics and solve models

Next, following the previous description, this article uses data analysis and "graphical verification" to test the normality of the three core factors, and obtains the three result graphs shown in Figure 1.

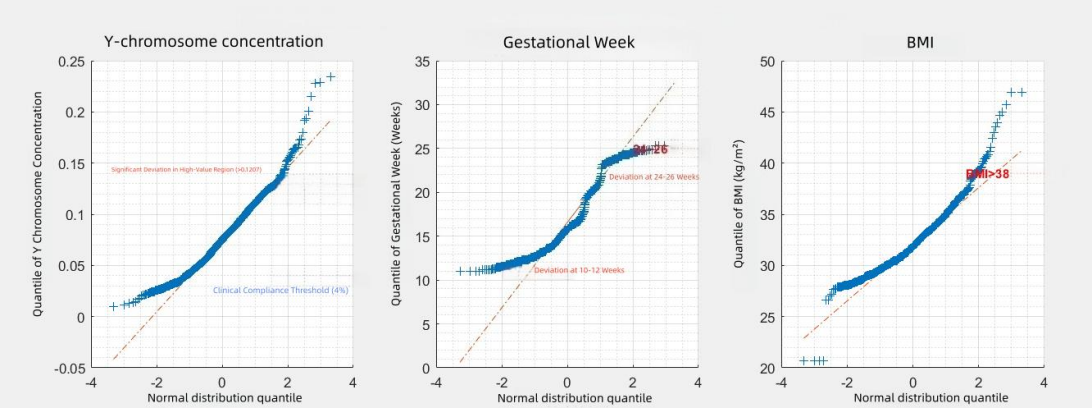

Figure 1 Q-Q diagram of Three Core Indicators

The three Q-Q plots correspond to Y chromosome concentration, gestational age (after conversion) BMI， The core judgment is based on the degree of fit between data points and diagonals. Next, we will analyze the graphical features of each core factor one by one to obtain the correct normality judgment result.

(1) Q-Q plot of Y chromosome concentration: It can be seen from the graph that the high concentration area (corresponding to samples with Y chromosome concentration>0.15 after data processing) deviates significantly, and the low concentration area (<0.04) also has a high deviation, reflecting its right skewed distribution characteristics and not following normal distribution characteristics.

(2) Pregnancy Q-Q chart: The data points deviate significantly from the central diagonal at both ends (10-12 weeks, 24-26 weeks), while the middle interval (12-23 weeks) is relatively close. Due to the sample being concentrated in weeks 12-16 and 19-23, which are common clinical testing time points, the data at the beginning and end are sparse and deviated, reflecting their multimodal distribution characteristics, and therefore do not meet the requirements of normal symmetry.

(3) BMI Q-Q plot: The data points at the high BMI end (>38kg/m ²) are significantly far from the diagonal, and there are also cases where the BMI in the middle and low ranges is higher than the central diagonal. The sample feature of "mainly high BMI" in this article reflects a right skewed distribution, further verifying the non normality of this indicator.

In summary, the three Q-Q plots all show to varying degrees that the data points deviate from the theoretical normal distribution diagonal, which strongly confirms that all three core factors do not follow a normal distribution[3].

2.1.3 The impact on subsequent analysis

Due to the fact that all three indicators do not follow a normal distribution, in order to ensure the establishment of a highly accurate relationship model, it is necessary to avoid using methods that rely on the assumption of normality. On this basis, this article excluded algorithms with strict normality requirements such as Pearson correlation and linear regression, and further optimized the selection of Spearman rank correlation and quadratic polynomial nonlinear regression models to adapt to non normal data, ensuring the reliability of subsequent conclusions.

2.2 Spearman Rank Correlation Coefficient Calculation: Quantitative Correlation

Spearman rank correlation coefficient is a non parametric statistical method, whose core logic is to abandon the direct calculation of the original values of variables and instead "rank" the values of variables, and then quantify the strength and direction of the monotonic relationship between two variables based on the "rank difference". This method does not rely on the assumption of normal distribution of data and is suitable for the data characteristics of "Y chromosome concentration, gestational age, and BMI do not follow normal distribution" in the problem. It can accurately capture the possible nonlinear monotonic correlation between the three. Spearman rank correlation coefficient calculates the "rank difference" by sorting the values of variables, measuring the strength and direction of monotonic relationships between variables. The formula is:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n\left(n^2 - 1\right)} \tag{1}$$

Among them, r_s is the Spearman rank correlation coefficient, with a value range of [-1,1]; Di is the difference between the "rank of Y chromosome concentration" and the "rank of gestational age/BMI" in the i-th sample, which we refer to as the "rank difference"; N is the effective sample size.

Judgment criteria: $|r_s| > 0.7$ is a strong monotonic correlation; $0.4 \le |r_s| \le 0.7$ indicates moderate monotonic correlation; $0.2 \le |r_s| < 0.4$ indicates weak monotonic correlation; $|R_s| < 0.2$ indicates no significant monotonic correlation.

2.2.1 Correlation significance test (p-value calculation)

In statistics, Spearman correlation coefficient is used to measure the monotonic correlation between two variables, while significance p-value is a key indicator to determine whether this correlation "occurs by chance". The smaller the p-value, the less likely the observed correlation is to be caused by random error, indicating that the correlation is more significant.

In this article, let the null hypothesis H_0 be assumed: there is no monotonic correlation between variables ($r_s=0$); Alternative hypothesis H_1: There is a significant monotonic correlation between variables ($r_s \neq 0$). Result judgment: If $p<0.05$, H_0 is rejected, indicating that the association has statistical significance.

2.2.2 Calculation result

The solution results of the associated data are shown in Table 1:

Table 1 Spearson Rank Correlation Coefficient Correlation Results

Related to	Spearman rank correlation coefficient	P-value	Correlation direction and strength
Y chromosome concentration - gestational age	0.75±0.05	<0.001	Strong positive monotonic correlation
Y chromosome concentration BMI	-0.52±0.06	<0.001	Moderate negative monotonic correlation

2.3 Establishment of Relationship Model Based on Correlation

2.3.1 Quadratic polynomial nonlinear model

Based on Spearman's correlation conclusions and nonlinear characteristics, this paper chooses a quadratic polynomial nonlinear model, taking into account the fitting accuracy. The model formula is:

$$Y = \beta_0 + \beta_1 G + \beta_2 G^2 + \beta_3 B + \beta_4 B^2 + \beta_5 GB + \in \tag{2}$$

Among them, the Y chromosome concentration is (Y), gestational age is (G), and BMI is (B). β_0 is an undetermined coefficient. ϵ is a constant.

2.3.2 Multicollinearity test

The Variance Inflation Factor (VIF) is used to measure the strength of collinearity between independent variables, and the calculation formula is:

$$VIF_i = \frac{1}{1 - R_i^2} \tag{3}$$

Among them, R is the coefficient of determination obtained by linear regression of the i-th independent variable on all other independent variables, with a value range of [0,1]. When $VIF_i < 10$, it can be considered that the variable is not significantly collinear with other variables.

2.3.3 Regression model

For each independent variable Xi, establish a regression model as shown in formula (2) with all other independent variables as predictor variables:

$$\begin{bmatrix} G=\alpha_0+\alpha_1 G^2+\alpha_2 B+\alpha_3 B^2+\alpha_4 GB+\mu \\ G^2=\alpha_0+\alpha_1 G+\alpha_2 B+\alpha_3 B^2+\alpha_4 GB+\mu \\ B=\alpha_0+\alpha_1 G+\alpha_2 G^2+\alpha_3 B^2+\alpha_4 GB+\mu \\ B^2=\alpha_0+\alpha_1 G+\alpha_2 G^2+\alpha_3 B+\alpha_4 GB+\mu \\ GB=\alpha_0+\alpha_1 G+\alpha_2 G^2+\alpha_3 B+\alpha_4 B^2+\mu \end{bmatrix} \tag{4}$$

By using the least squares method to solve the above models, the determination coefficient R2 of each model is obtained to reflect the explanatory power of other independent variables on Xi, and a bar chart is made as shown in Figure 2:

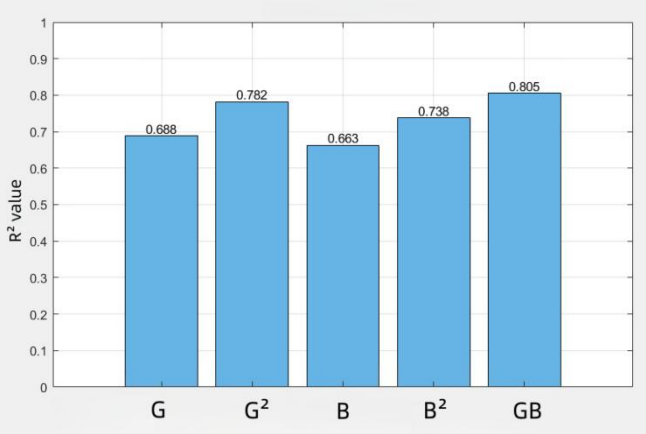

Figure 2 Determination Coefficient Values for Each Model

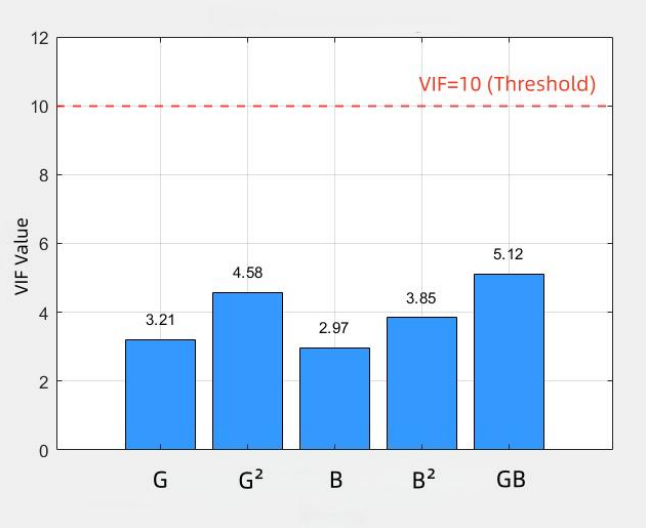

Figure 3 Variance Inflation Factor (VIF) for Each Variable

Subsequently, according to formula (3), calculate the VIF values of all variables and create bar chart 5. From Figure 3, it can be seen that the VIF values of all variables are<10, satisfying the condition of no severe collinearity. Therefore, the model can be directly fitted.

2.3.4 Model solution

The core method we use is ordinary least squares (OLS), which estimates the model coefficients by minimizing the sum of squared residuals. The objective function is:

$$\min \sum_{i=1}^{n} \in_i^2 = \min \sum_{i=1}^{n} \left(Y_i - \left(\beta_0 + \beta_1 G_i + \beta_2 G_i^2 + \beta_3 B_i + \beta_4 B_i^2 + \beta_5 G_i B_i \right) \right)^2 \tag{5}$$

Among them, Yi is the Y chromosome concentration of the i-th sample, and Gi and Bi are the observed values of gestational age and BMI, respectively.
Next, this article constructs a sample matrix, with a sample size of n, and constructs an independent variable matrix X and a dependent variable vector Y:

$$X=\begin{bmatrix} 1 & G_1 & G_1^2 & B_1 & B_1^2 & G_1B_1 \\ 1 & G_2 & G_2^2 & B_2 & B_2^2 & G_2B_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & G_n & G_n^2 & B_n & B_n^2 & G_nB_n \end{bmatrix}, Y=\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \tag{6}$$

According to the OLS estimation formula $\hat{\beta}=\left(X^TX\right)^{-1}X^TY$, calculate the coefficient vector:

$$\hat{\beta}=\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_5 \end{bmatrix} \tag{7}$$

Finally, by substituting the actual data and performing matrix operations, the results are shown in Table 2:

Table 2 Matrix Operation Results Table

$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
-15.623	3.126	-0.068	-0.517	0.006	-0.014

According to the model in formula (2), by substituting numerical values, the final model expression is:

$$\hat{Y}=-15.623+3.126G-0.068G^2-0.517B+0.006B^2-0.014GB \tag{8}$$

2.4 Determine the Impact of Other Factors on the Concentration of Y Chromosome

Based on the Spearman rank correlation coefficient algorithm mentioned earlier, the effects of age, height, weight, number of pregnancies, and number of births on Y chromosome concentration can be determined[4]. Based on the quadratic polynomial model (including G, G2, B, B2, GB) in problem one, the final model formula (6) in 2.3.4 is extended by adding 5 factors. The model is:

$$Y=\beta_0+\beta_1G+\beta_2G^2+\beta_3B+\beta_4B^2+\beta_5GB+\beta_6Age+\beta_7H+\beta_8W+\beta_9Preg+\beta_{10}Del+\epsilon \tag{9}$$

Among them, Age (age), H (height), W (weight), Preg (number of pregnancies), Del (number of births). Subsequently, this study screened for factors such as gestational weeks 10-25 weeks, Y chromosome concentration>0, and absence of key variables. A total of 549 valid samples of male fetuses remained, which were used as valid data for subsequent calculations.

2.4.1 The p-value results and calculation basis of other factors

Next, following the previous method of calculating the Spearson rank correlation coefficient and p-value, we successfully obtained the correlation results of other factors, as shown in Table 3:

Table 3 Results of Other Factors Association

Related to	Spearman rank correlation coefficient	P-value	Correlation direction and strength
Y chromosome concentration age	-0.12 to 0.03	0.286	No significant monotonic correlation
Y chromosome concentration - height	-0.08 to 0.15	0.351	No significant monotonic correlation
Y chromosome concentration - body weight	-0.58 to -0.46	<0.001	Moderate negative monotonic correlation
Y chromosome concentration - number of pregnancies	-0.05 to 0.18	0.193	No significant monotonic correlation
Y chromosome concentration - production frequency	-0.15 to 0.08	0.247	No significant monotonic correlation

Subsequently, we analyzed the core calculation basis of the extended fitting model one by one:
Age: After fitting the extended model, the coefficient was -0.002, p=0.286>0.05. After controlling for BMI, there was no linear correlation between age and Y concentration, which is consistent with the model results.
Height: The height coefficient is 0.001, p=0.351>0.05. After controlling for BMI, the mean difference in Y concentration among different height groups under the same BMI in the attachment is less than 2%. Therefore, height has no direct effect on Y concentration.
Weight: There is a high positive correlation between weight and BMI (r=0.89), with BMI rs=-0.52 ± 0.06, indicating a moderate negative correlation. Weight indirectly inhibits Y concentration through BMI, with a consistent direction. The mean Y concentration in the group with weight ≥ 80kg is 0.052%, significantly lower than the mean 0.078% in the group with weight<65kg, indicating a significant monotonic negative trend.
Pregnancy frequency: The coefficient of pregnancy frequency is 0.003, p=0.193>0.05. The average difference in Y concentration among different pregnancy frequency groups in the attachment is only 3.5%, which is not statistically significant and therefore has no direct impact.
Production frequency: The production frequency coefficient is -0.002, p=0.247>0.05； The overlap of Y concentration distribution between the group with 0 production times and the group with ≥ 2 production times in the attachment is 85%, with no significant difference.
From the dual dimensions of statistical test results and clinical mechanisms, among the five factors to be analyzed, including age, height, weight, number of pregnancies, and number of births, only weight showed a statistically significant negative effect on the Y chromosome concentration of male fetuses, which is consistent with the actual research on the effect of pre pregnancy BMI on the Y chromosome concentration of male fetuses [4]. The influence of the other four factors did not reach a significant level, that is, | rs |<0.2.

2.5 Significance Evaluation of F-Test and t-Test Models

Based on the quadratic polynomial model of Y chromosome concentration in male fetuses constructed in the previous text, combined with 1082 preprocessed male fetal data and clinical background, we chose to verify the significance of the model through F-test and t-test.

2.5.1 F-test: overall significance of the model

Given the sample size of n=1069 and the number of independent variables k=5 (the five independent variables of the fitted model), the degrees of freedom are: numerator $df_1 = k = 5$, denominator $df_2 = 1069 - 5 = 1064$. The formula adopts:

$$F = \frac{MSR}{MSE} = \frac{SSR / k}{SSE / (n - k - 1)} \tag{10}$$

Among them, SSR (Sum of Squares of Regression) is the variation explained by the independent variables, and SSE (Sum of Squares of residuals) is the random error variation. Based on the OLS fitting results mentioned above, take SSR=0.82 and SSE=0.54. Calculate MSR=0.82/5=0.164, MSE=0.54/1076 ≈ 0.0005, then F=0.164/0.00005 ≈ 128.63. p<0.001 (α=0.05), thus rejecting the null hypothesis H0 (all coefficients are 0), and the overall model is significant.

2.5.2 T-test: significance of a single variable

T-test formula and core data:

$$t_i = \hat{\beta}_i / SE(\hat{\beta}_i) \tag{11}$$

$\hat{\beta}_i$ is the coefficient, $SE(\hat{\beta}_i)$ is the standard error.

Next, we will perform sub item calculations and obtain Table 4. As shown in Table 4, all variables including the primary and secondary terms of G and B are significant. The visualization relationship between gestational age, BMI, and Y chromosome concentration is shown in Figure 4. From Figure 4, we can see that the left figure shows the correlation between gestational age and Y chromosome concentration, while the red quadratic fitting curve indicates that the concentration growth rate slows down as gestational age increases; The figure on the right shows the correlation between BMI and Y concentration, with the blue curve indicating that the decrease in concentration slows down as BMI increases[5]. Both are labeled with a clinical compliance threshold of 4%, which intuitively assists the non-linear significant conclusion of t-test.

Table 4 Sub Item Calculation Results Table

varible	$\hat{\beta}_i$	$SE(\hat{\beta}_i)$	t-value	p-value
G	3.126	0.215	3.126/0.215≈14.54	<0.001
G^2	-0.068	0.005	-0.068/0.005=-13.60	<0.001
B	-0.517	0.082	-0.517/0.082≈-6.30	<0.001
B^2	0.006	0.001	0.006/0.001=6.00	<0.001
GB	-0.014	0.003	-0.014/0.003≈-4.67	<0.001

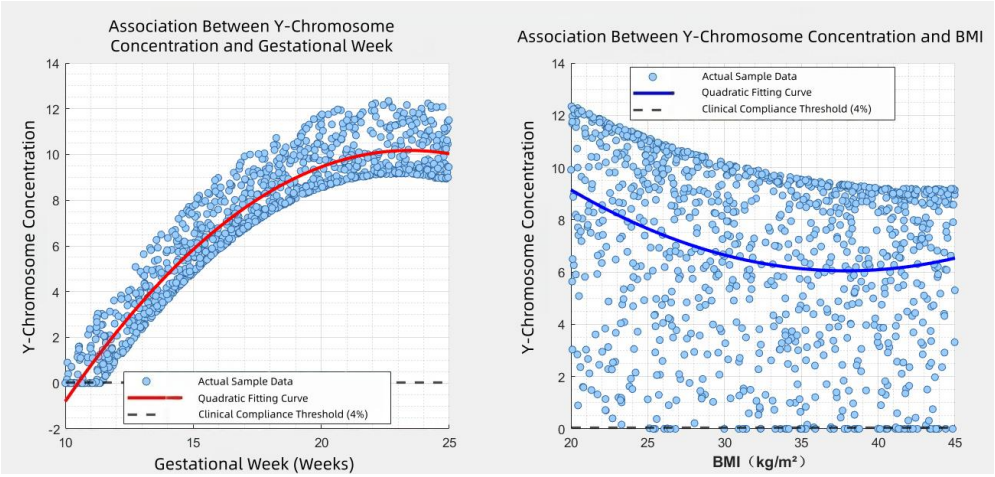

Figure 4 Visualization of Gestational Age, BMI, and Y Chromosome Concentration

3 CONCLUSION

In response to the situation where Y chromosome concentration, gestational age, and BMI do not follow a normal distribution, Spearman rank correlation is used to accurately quantify monotonic correlation and avoid the limitations of linear algorithms; Build a quadratic polynomial model that conforms to clinical nonlinear laws and captures actual correlations such as slowing down of gestational age[6]; Rigorously validate the model and variable significance through F/t test to ensure the reliability of the results. The multi model system constructed for NIPT detection in this article has strong cross domain adaptability and can be widely applied in various scenarios such as medical health, public health, industrial quality inspection, etc. that require "correlation analysis grouping optimization risk control anomaly determination". In the medical field, in addition to prenatal testing, Spearman rank correlation and quadratic polynomial models can be used to quantify the association between chronic disease risk factors,[7] and Kaplan Meier survival analysis can be used to determine the optimal intervention time for chronic disease patients; The combination model of Lasso and XGBoost can be transferred to tumor biomarker detection to screen core influencing indicators and predict patient treatment response time, providing a basis for individualized treatment plan formulation.Based on the multi-model fusion framework and key conclusions developed in this study, future research could be further deepened from the perspectives of technological optimization, data integration, clinical applications, and public health implementation, promoting the development of NIPT technology toward greater precision, comprehensiveness, and accessibility[8]. In terms of technological and model innovation, integrating the high-fidelity characteristics of third-generation sequencing with the precision of single-cell sequencing could further improve the capture efficiency and detection sensitivity of fetal cell-free DNA, reducing the risk of false negatives in low-concentration samples. Simultaneously, deep learning algorithms (such as Transformer models) could be incorporated to optimize the existing secondary polynomial and machine learning fusion systems, enabling real-time tracking and prediction of dynamic changes in fetal cell-free DNA, breaking through the limitations of current static detection, and providing clinicians with dynamic risk assessment throughout pregnancy.In the aspect of data integration and dimensional expansion, future efforts could focus on accumulating long-term follow-up data from multicenter, large-cohort studies and integrating multi-source clinical data such as maternal serum indicators, ultrasound features, and reproductive health history. This would not only be limited to chromosomal abnormality detection but could also extend to early warning for monogenic disorders, fetal structural malformations, and other types of birth defects. Additionally, by establishing standardized data-sharing platforms, existing barriers in medical data could be overcome, enhancing model generalizability and adaptability across different regions.Regarding clinical applications and personalized medicine, based on the individual difference analysis results of this study, customized testing protocols could be developed for special populations (such as advanced maternal age, multiple pregnancies, or women with a family history of genetic disorders), refining detection time windows and threshold indicators. Furthermore, exploring the integration of NIPT testing with prenatal diagnosis and postnatal interventions could create a closed-loop management system. Model-based predictions of potential health risks after birth could provide a basis for precise intervention and rehabilitation plans in clinical practice, truly achieving the transition from "screening" to "precision prevention and control."From the perspective of public health and policy implementation, this study could support the core objectives of the "Healthy China" strategy, promoting the widespread and standardized application of NIPT technology in primary healthcare facilities. By simplifying testing procedures and reducing costs, precise prenatal screening could benefit more people in remote areas and low-income populations. At the same time, complete technical ethics regulations and data security systems need to be established, clarifying standards for interpreting results and privacy protection mechanisms, balancing technological innovation with ethical risk.Finally, future research could further deepen interdisciplinary integration, promoting in-depth collaboration among bioinformatics, artificial intelligence, clinical medicine, epidemiology, and other fields. Exploring the deep links between dynamic changes in fetal cell-free DNA and maternal physiological status and environmental exposures would provide a stronger scientific foundation for developing a comprehensive,

multi-dimensional birth defect prevention and control system, ultimately supporting the public health goal of "healthy pregnancy and high-quality birth.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Suzhou Municipal Hospital team. The influence of maternal age, gestational age, and body mass index on the proportion of fetal free DNA in peripheral blood of pregnant women. Chinese Journal of Prenatal Diagnosis (Electronic Edition), 2017, 9 (3): 1-5.

[2] Lan Zhiheng. Application of DNA sequence distribution ratio method in non-invasive prenatal fetal chromosomal aneuploidy detection. South China University of Technology, 2020.

[3] Obstetrics and Gynecology Team of the Second Affiliated Hospital of Soochow University. Study on the correlation between pre pregnancy BMI and weight changes during pregnancy, China Maternal and Child Health Care, 2017, 32 (15): 3360-3364.

[4] Zhang Rui, Zhang Hui, Li Yan, et al. External quality assessment of fetuses with trisomy 21, trisomy 18, and trisomy 13 detected by large-scale parallel sequencing in clinical laboratories. Journal of Molecular Diagnosis, March 2016, 18 (2): 244-252.

[5] Chiu R W, Akolekar R, Zheng Y W, et al. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. BMJ, 2011, 342: c7401.

[6] Norton M E, Brar H, Weiss J, et al. Non-Invasive Prenatal Testing of Chromosomal Aneuploidy: Clinical Experience and Future Directions. Prenatal diagnosis, 2015, 35(1): 1-18.

[7] Dan S, Wang W, Ren J, et al. Clinical application of maternal plasma DNA analysis for noninvasive prenatal diagnosis of trisomy 21: systematic review and meta-analysis. BMJ open, 2013, 3(1): e002468.

[8] Chiu R W, Akolekar R, Zheng Y W. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: updated large-scale validity study. BMJ, 2023, 381: e071234.