Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 12
ESAC 2026

Quantitative Study on the Correlation Between Fetal Y Chromosome Concentration in NIPT and Maternal Indicators Based on Multiple Regression and Quantile Regression, and the Optimal Detection Timing

Haolin Zhou

School of Information Science and Technology (School of Cyber Security), Guangdong University of Foreign Studies, Guangzhou, China, 510006

20241003239@mail.gdufs.edu.cn

Abstract. This study focuses on the quantitative association analysis between fetal Y chromosome concentration and maternal indicators, and the determination of the optimal timing for non-invasive prenatal testing (NIPT) based on BMI grouping. The aim is to optimize NIPT strategies and minimize potential risks to pregnant women. Regarding the association characteristics between Y chromosome concentration and maternal indicators such as gestational age and BMI, Pearson correlation analysis first revealed a negative correlation between Y concentration and BMI (-0.1683) and a positive correlation with gestational age (0.09842), though both linear relationships were weak. Building upon this, a multivariate polynomial regression model incorporating quadratic and interaction terms was constructed, with parameters fitted using least squares. F-test validation confirmed the model's overall statistical significance ($p<0.001$). To determine the optimal testing time point, considering that K-means clustering might cause sample size imbalance, this study employed quartile grouping to reasonably categorize the BMI of pregnant women carrying male fetuses, ensuring an even distribution of sample sizes across groups. Subsequently, a quantile regression model was established with the objective of minimizing the potential risk to pregnant women. The results revealed optimal NIPT testing timepoints of 12.59 weeks, 12.87 weeks, 13.06 weeks, and 13.31 weeks for the four BMI groups, respectively. Detection error increased with rising BMI.

Keywords: Multivariate polynomial regression; Quantile regression; Optimal timing for NIPT.

1. Introduction

Non-invasive prenatal testing (NIPT) is a crucial prenatal screening technique that analyzes fetal chromosomal abnormalities—particularly conditions like Down syndrome caused by chromosomal abnormalities in chromosomes 21, 18, and 13—by collecting fetal cell-free DNA fragments from maternal blood [1-2]. NIPT accuracy is highly dependent on fetal sex chromosome concentration. For male fetuses, Y chromosome concentration must reach or exceed 4% to ensure reliable results. While testing typically occurs between 10 and 25 weeks of gestation, earlier detection reduces potential risks associated with a narrowing therapeutic window. Clinical practice indicates strong correlations between fetal Y chromosome concentration and both gestational age and maternal body mass index (BMI). Therefore, to safeguard human health and mitigate potential risks, this study addresses two core issues: quantifying the association between fetal Y chromosome concentration and maternal indicators, and determining the optimal timing for NIPT based on BMI-based gestational age grouping. The research methodology first employs Pearson correlation coefficients to analyze the characteristics of Y chromosome concentration in relation to BMI and gestational week. Subsequently, a multivariate polynomial regression model (incorporating quadratic and interaction terms) is constructed to quantify the functional relationship among these three variables and perform significance testing. Regarding optimal timing, this study grouped maternal BMI for male fetuses using quartiles [3]. A quantile regression model was then established to determine the optimal NIPT testing window that minimizes maternal risk, while also analyzing the impact of testing errors.

2. Quantitative Correlation Analysis of Fetal Y Chromosome Concentration with Maternal Indicators

By referring to the article "The Impact of Maternal Age, Gestational Week, and Body Mass Index on the Proportion of Fetal Cell-Free DNA in Maternal Peripheral Blood" [4-5], it is known that the fetal Y-chromosome concentration is generally negatively correlated with BMI and positively correlated with the gestational week of detection. Based on the above analysis premises and data preprocessing operations, the solution is carried out step by step below.

Since only males have Y-chromosomes, the female fetal test data is not required in the process of solving the relationship between the Y-chromosome concentration and indicators such as the pregnant woman's gestational week and BMI. It can be known from the additional information of the problem that the GC content of the fetus (the proportion of guanine G and cytosine C in the sequence, an important indicator for evaluating the quality of sequencing data) is usually distributed between 40% and 60%. If the GC content result of a certain sequencing is outside the range of 40% - 60%, it indicates that the quality of this sequencing is problematic, and further processing of abnormal data is required. In addition, it is observed that the "gestational week of detection" column in the attachment is in the form similar to "12w + 3" in the originally given data, which should be converted into a real value during data preprocessing (12w+3→(12×7+3)/7→14.1428). The above preprocessing steps are all completed in the "(V) Data Preprocessing" section. The sub-flowchart for the establishment and solution of the model is shown in the following cure 1:

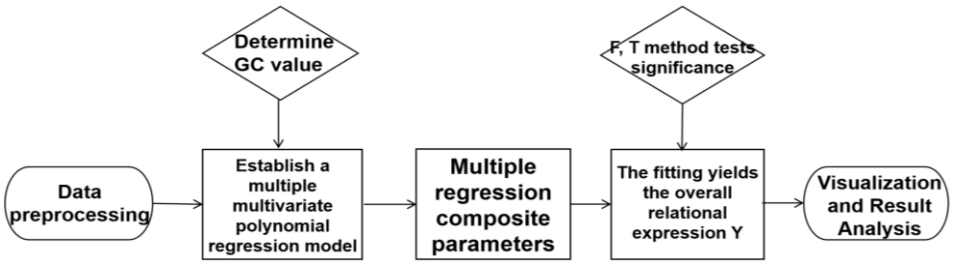

Figure 1. Flowchart

2.1 Analysis of Correlation Characteristics of Main Indicators

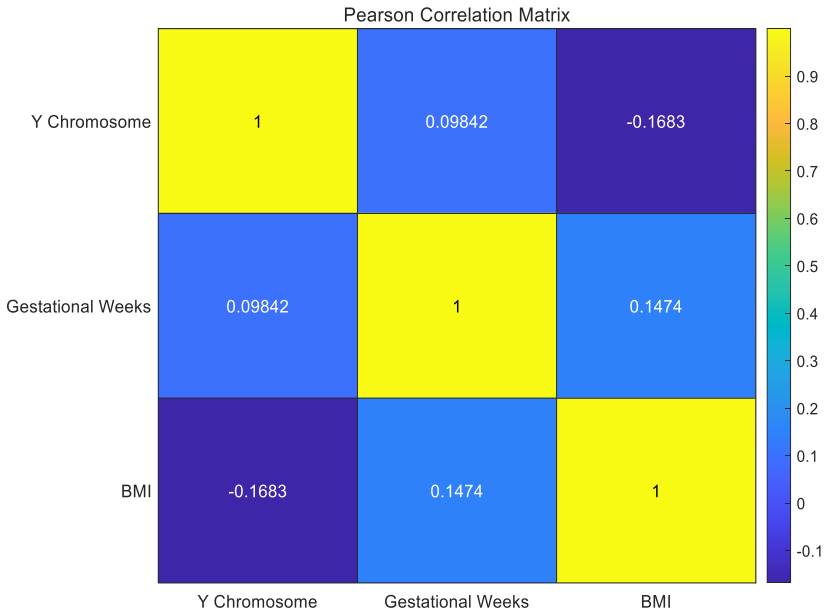

Figure 2. Pearson Correlation Coefficient Matrix

According to the requirements of the problem, MATLAB programming tools are used to solve the correlation characteristics. Obviously, to analyze the correlation between two variables, the Pearson coefficient should be used to determine the positive or negative correlation between the Y-chromosome

concentration and BMI as well as the gestational week [6-7]. By calling the internal function to solve the Pearson correlation matrix, the Pearson coefficients are obtained as -0.1683 and 0.09842 respectively (as shown in the following figure). It is found that although the Y-chromosome concentration has a negative correlation with BMI and a positive correlation with the gestational week, the linear relationship between them is weak. Therefore, in the following model, dimension expansion and interaction terms are carried out, and polynomial regression is established to test the existing nonlinear relationship. In the regression results, the p-value of BMI^2 is 0.048416 (less than 0.05), and the p-value of the square of the gestational week of detection is 0.0059535 (very small), which all indicate that the nonlinear relationship between the Y-chromosome concentration and BMI as well as the gestational week is significant. Pearson correlation coefficient matrix is shown in figure 2.

2.2 Establishment of a Multiple Polynomial Regression Relationship Model

The purpose of this stage is to establish an accurate mathematical model to quantitatively analyze the impact of the pregnant woman's BMI and gestational week of detection on the Y-chromosome concentration. According to the correlation characteristics between the Y-chromosome concentration and indicators such as the gestational week of detection and BMI in 2.1, it is known that the fetal Y-chromosome concentration should be negatively correlated with BMI with a Pearson coefficient of -0.1513, and positively correlated with the gestational week of detection with a Pearson coefficient of 0.1266. By referring to relevant materials [8-9], it is known that the Y-chromosome concentration is not only related to the BMI index and the gestational week of detection, but also closely related to the pregnant woman's age and X-chromosome concentration. Combined with the problem analysis and data exploration, a multiple polynomial regression model is selected as the core modeling tool. In the multiple polynomial regression function, the independent variables are BMI and the preprocessed gestational week of detection, and the dependent variable is the Y-chromosome concentration. In addition, random noise based on a normal distribution is added during the function fitting process to make the model more random.

First, the general form of the model is defined. It can be easily seen from the results of the "Analysis of Correlation Characteristics of Main Indicators" above that the Y-chromosome concentration is correlated with both BMI and the gestational week of detection, but the correlation relationships are weak. Therefore, when determining the model form, the quadratic terms of the two independent variables BMI^2 and (gestational week of detection) ^{2}, the interaction term (BMI · gestational week of detection), the square of the male fetal X-chromosome concentration X^2, and the pregnant woman's age are introduced to improve the model fitting degree. To sum up, the final function Y is defined as follows:

$$\mathrm{Y} = \beta_0 + \beta_1 \mathrm{x}_1 + \beta_2 \mathrm{x}_2 + \beta_3 \mathrm{x}_1^2 + \beta_4 \mathrm{x}_2^2 + \beta_5 \mathrm{x}_1 \mathrm{x}_2 + \beta_6 \mathrm{x}_3 + \beta_7 \mathrm{x}_4 + \epsilon \quad (1)$$

In the formula, Y represents the chromosome concentration; x_1 is the independent variable BMI; x_2 is the gestational week of detection (g_week); x_3 is the X-chromosome concentration; x_4 is the pregnant woman's age. Their quadratic forms are x_1^2 and x_2^2 respectively, and the interaction form is $\mathrm{x}_1\mathrm{x}_2$. β_0, β_1, β_2, β_3, β_4, β_5, β_6, and β_7 are the coefficients of each term respectively; ϵ is the random noise value, and $\epsilon \sim \mathrm{N}(0, \sigma^2)$. The physical meaning of this model is that the Y-chromosome concentration is jointly determined by a constant term β_0, linear effect terms β_1 and β_2, curvature effect terms β_3 and β_4, and random noise ϵ.

After completing the definition of the above function, the model is now extended to a matrix form to facilitate the subsequent solution of the model. Assuming that finally n rows of preprocessed valid data are involved in the calculation, the matrix expression $\mathrm{Y}_{\mathrm{total}}$ is as follows: Among them, $\mathrm{Y}_{\mathrm{total}}$ is the matrix expression of the final Y-chromosome concentration result data, and β_{total} and $\mathrm{X}_{\mathrm{total}}$ are defined similarly [10].

$$\mathrm{Y_{total}} = \beta_{\mathrm{total}} \cdot \mathrm{X_{total}} + \epsilon, \epsilon \sim \mathrm{N}(0, \sigma^2) \quad (2)$$

The method of least squares is a mathematical optimization technique that finds the best function fit for the data by minimizing the sum of the squares of the errors. Specifically, for a given set of data points, assuming that there is a certain functional relationship (such as a linear relationship, a polynomial relationship, etc.) between them, the method of least squares aims to determine the unknown parameters in the function so that the sum of the squares of the deviations (i.e., errors) between the function and the actual data points is minimized. Then, the ordinary least squares (OLS) method is used for parameter estimation on Equation (2). The ultimate goal is to find a derived parameter matrix β'_{total} such that the residual sum of squares (RSS) is minimized. To sum up, the final objective function RSS is: Among them, $\mathrm{Y_{total}} - \mathrm{X_{total}} \cdot \beta_{\mathrm{total}}$ is the error value, and T is the matrix transposition symbol. By taking the derivative of RSS with respect to β_{total} and setting the derivative to zero, the closed-form solution β'_{total} of the parameter estimation value can be obtained.

$$min_{\beta} RSS = min_{\beta}[(Y_{total} - X_{total} \cdot \beta_{total}) \cdot (Y_{total} - X_{total} \cdot \beta_{total})^T] \quad (3)$$

$$\beta'_{total} = (X_{total}^T \cdot X_{total})^{-1} \cdot X_{total}^T \cdot Y_{total} \quad (4)$$

Among them, $\mathrm{X_{total}^T}$ is the transpose matrix of $\mathrm{X_{total}}$, $(\mathrm{X_{total}^T} \cdot \mathrm{X_{total}})^{-1}$ represents the inverse of the matrix, and β'_{total} is the optimal parameter solution obtained after calculation. This solution provides the optimal combination of parameters that minimizes the overall difference between the model's predicted values and the actual values.

In addition, to simulate a real modeling environment, noise is introduced here. Noise can be understood as random and irregular physical or acoustic phenomena. A common noise model is a random process, which describes the behavior of a set of random variables that change over time. Mathematically, probability and statistical theories can be used to describe and quantify this uncertainty. In a noisy environment, these random variables may represent various different sounds or vibrations. In this paper, noise can be understood as the influence of other secondary factors (such as the pregnant woman's age, living habits, and the external environment) on the Y-chromosome concentration, in addition to the main factors (BMI and gestational week of detection) that cause changes in the Y-chromosome concentration. In the model of this section, the noise variance σ^2 is estimated unbiasedly by the residuals: Among them, ϵ'^2 is the estimated noise variance, RSS is the residual sum of squares, k refers to the number of parameters in Equation (1), excluding the constant term parameter β_0, i.e., $\mathrm{k} = 8 - 1 = 7$. $\mathrm{y_i}$ and $\mathrm{y_i'}$ are the actual value and predicted value of the i-th data in the $\mathrm{Y_{total}}$ matrix respectively, and n refers to the number of valid data after preprocessing. The meaning of σ is the magnitude of the random fluctuations that the model fails to explain. To sum up, combining the above Equations (1) - (5), the final total model equation $\mathrm{Y'_{total}}$ established is: Among them, $\mathrm{Y'_{total}}$ is the predicted value of the Y-chromosome concentration matrix; $\beta'_{\mathrm{i}}(\mathrm{i} = 0, \ldots, 7)$ represents the final fitting value of the i-th parameter; $\mathrm{x_1}$ and $\mathrm{x_2}$ are the BMI index and the gestational week of detection respectively; $\mathrm{x_1^2}$ and $\mathrm{x_2^2}$ are their quadratic terms respectively; and $\mathrm{x_1 x_2}$ is the interaction term.

$$\epsilon'^2 = \frac{RSS}{n-k-1} \cdot \frac{\sum_{i=1}^{n} (y_i - y_i')^2}{n-5} \quad (5)$$

$$Y'_{total} = \beta'_0 + \beta'_1 x_1 + \beta'_2 x_2 + \beta'_3 x_1^2 + \beta'_4 x_2^2 + \beta'_5 x_1 x_2 + \beta'_6 x_3 + \beta'_7 x_4 + \epsilon \quad (6)$$

The coefficients β'_1 and β'_3 of the total model (6) quantify the (linear and nonlinear) impact of BMI on the Y-chromosome concentration; β'_2 and β'_4 quantify the (linear and nonlinear) impact of the gestational week index on the Y-chromosome concentration; β'_5 quantifies the impact of the

interaction term between BMI and the gestational week of detection on the Y-chromosome concentration; and finally, β_6' and β_7' quantify the impact of the X-chromosome concentration and the pregnant woman's age on the Y-chromosome concentration respectively. By inputting the new pregnant woman's BMI, gestational week of detection, X-chromosome concentration, and age into the model, the Y-chromosome concentration in her blood can be predicted. In addition, based on the results of the model fitting, it is possible to determine which factors are significant influencing factors and provide the confidence interval of the predicted values. This modeling stage lays a solid foundation for the subsequent model solution, significance test, and result analysis.

2.3 Solution of the Relationship Model

Relationships between BMI, Gestational Week of Detection, Pregnant Woman's Age and Y-Chromosome Concentration is shown in figure 3. The program is written using MATLAB programming software, and the solution results are as follows: $R^2 = 0.2364$, adjusted $R^2 = 0.2314$, $RMSE = 0.027322$, $MAE = 0.022227$. The result of the total model equation Y'_{total} is:

$$Y'_{total} = -0.089809 + 0.013912x_1 - 0.002874x_2 - 0.000259x_1^2 + 0.000028x_2^2 + 0.000072x_5 + 0.348371x_6 - 0.001035x_7 + \epsilon \quad (7)$$

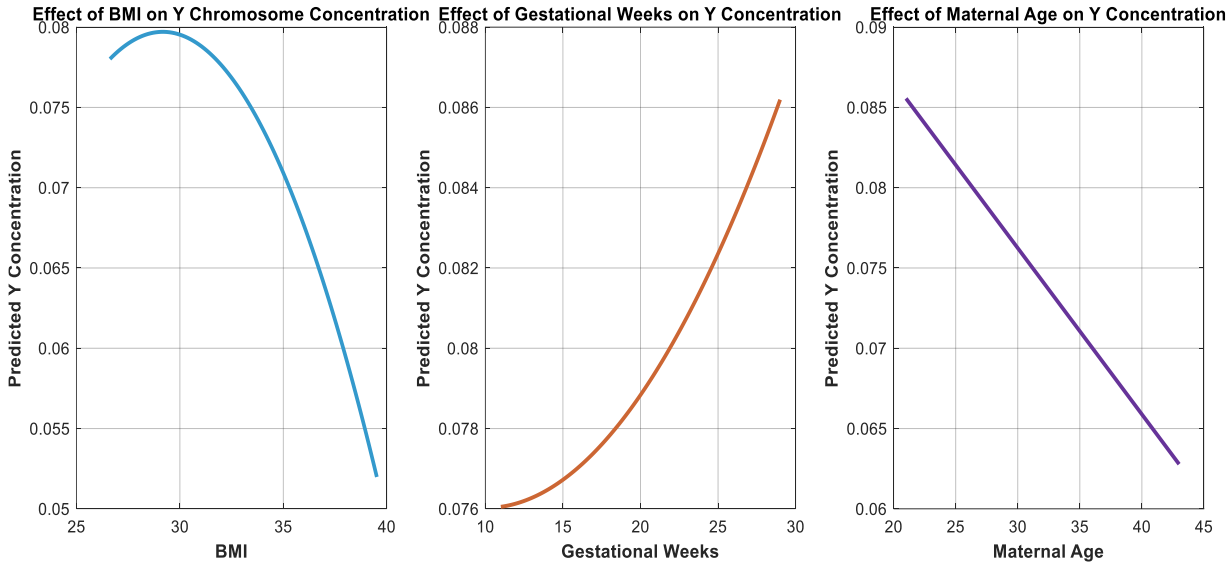

Figure 3. Relationships between BMI, Gestational Week of Detection, Pregnant Woman's Age and Y-Chromosome Concentration

2.4 Significance Test

For the multiple polynomial regression model established earlier, the F-test and T-test are used here. By calculating the overall F-statistic and the significance test of each coefficient, the model significance is verified. The F-statistic (for the overall significance of the model) and the T-coefficient significance matrix are still solved by writing a program in MATLAB. It can be known from the calculation results output by the code that the final result of the F-statistic is 47.5044, and the F-test is passed; the T-test coefficient matrix is shown in Figure 3. The smaller the p-value in the T-test coefficient significance matrix, the stronger the correlation between the parameter and the dependent variable (Y-chromosome concentration), and vice versa. It can be observed that the first-order terms of BMI and the gestational week of detection, and the second-order term of the gestational week of detection have an insignificant relationship with the Y-chromosome concentration, while the second-order term of BMI and the interaction term have a relatively significant relationship with the Y-chromosome concentration. The model significance test is relatively successful. Significance matrix of each coefficient calculated by T-Test is shown in table 1.

Table 1. Significance Matrix of Each Coefficient Calculated by T-Test

Coefficient	Estimated Value	Standard Error	t-Statistic	p-Value
β_0	-0.089809	0.112989	-0.7948	4.2688e-01
β_1	0.013912	0.006761	2.0576	3.9870e-02
β_2	-0.002874	0.003007	-0.9557	3.3946e-01
β_3	-0.000259	0.000105	-2.4798	1.3299e-02
β_4	0.000028	0.000056	0.5058	6.1313e-01
β_5	0.000072	0.000078	0.9203	3.5763e-01
β_6	0.348371	0.022643	15.3854	0.0000e+00
β_7	-0.001035	0.000229	-4.5313	6.5205e-06

2.5 Analysis of the Solution Results of the Relationship Model

From the calculation results in 2.3, the p-value is 6.9311e-59, which is used to test the overall significance of the model. If $p < 0.001$, it indicates that the null hypothesis that "all coefficients are zero" is rejected, and it is believed that there is a significant regression relationship between at least one independent variable and the dependent variable, and the model is statistically valid. Obviously, the p-value is less than 0.001, so the model is valid. The Root Mean Square Error (RMSE) is 0.027322, which means that the average difference between the Y-concentration values predicted by the model and the actual values is approximately 2.73%. The error is not large, and this is a core indicator for measuring the prediction accuracy of the model. The Mean Absolute Error (MAE) is 0.022227, indicating that the mean absolute error is 2.22%, which is close to the RMSE. This indicates that the error distribution is relatively symmetric, and there are no extreme outliers. To sum up, the multiple polynomial regression function for the Y-chromosome concentration performs well in some test results, and the model has certain reference value.

3. Determination of Optimal Timing for Non-Invasive Prenatal Testing Based on BMI Grouping

Clinically, it has been shown that the BMI of pregnant women carrying male fetuses is a key factor affecting the time it takes for the fetal Y-chromosome concentration to reach the standard. The pregnant women carrying male fetuses are reasonably grouped according to their BMI indicators to determine the BMI range and the optimal NIPT detection time for each group, so as to minimize the potential risks and analyze the impact of detection errors. The solution is carried out step by step below figure 4:

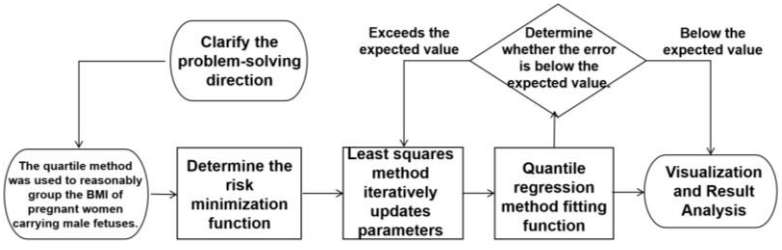

Figure 4. Flowchart

3.1 Establishment of the Model

Step 1: Reasonable Grouping of BMI of Pregnant Women Carrying Male Fetuses

First, it is necessary to determine the method for grouping pregnant women according to the BMI indicator. If the K-means clustering method is used for grouping in this problem, it will have a fatal flaw: the number of people in each group varies greatly. The consequence of this phenomenon is that

some groups may not be able to optimize to a good parameter due to too little training data during the parameter fitting process, thereby affecting the final minimum risk value. To ensure a reasonable number of groups and a relatively similar sample size in each group, after discussion, it is decided to break the conventional method of K-means clustering grouping, innovate on the basis of the original commonly used models in combination with the actual data situation, and use the quartile method to group BMI. First, all the BMIs of pregnant women carrying male fetuses are sorted from low to high. Then, the quartiles Q1 (25%) (i.e., the first quartile), Q2 (50%) (i.e., the median), and Q3 (75%) (i.e., the third quartile) are found in the sorting. The subsequent grouping logic is as follows: the first group is $\mathrm{BMI} \leq \mathrm{Q1}$, the second group is $\mathrm{Q1} < \mathrm{BMI} \leq \mathrm{Q2}$, the third group is $\mathrm{Q2} < \mathrm{BMI} \leq \mathrm{Q3}$, and the fourth group is $\mathrm{BMI} > \mathrm{Q3}$.

Step 2: Establish an Optimization Function to Minimize the Potential Risks of Pregnant Women and Solve for the Optimal NIPT Time

Parameter fitting. Since the detection time cannot be too early (resulting in sequencing failure) or too late (resulting in health risks), a quantile regression model (with BMI as the independent variable and gestational week of detection as the dependent variable) is used for this problem to perform 20% quantile regression prediction. That is, it is appropriate to select the quantile parameter $\tau = 0.2$. The model is established as the following formula: Among them, b_0 is the bias term, and b_1 is the weight of the BMI index. Then, according to the quantile regression [3], iterative optimization is performed to determine b_0 and b_1. The method of least squares is a mathematical optimization technique that finds the best function fit for the data by minimizing the sum of the squares of the errors. Specifically, for a given set of data points, assuming that there is a certain functional relationship between them, the method of least squares aims to determine the unknown parameters in the function, which are b_0 and b_1 here, so that the sum of the squares of the deviations (i.e., errors) between the function and the actual data points is minimized. A pair of b_0 and b_1 is calculated using the ordinary least squares (OLS) method as the starting point of the iteration to solve the parameter matrix b: Among them, X represents the matrix composed of BMI indicators, and y represents the matrix composed of gestational week of detection indicators. Next, according to the currently obtained parameters b_0 and b_1, the predicted value $\overline{\mathrm{y}}_\mathrm{i}$ of each point is calculated.

$$Q_{0.20}(t|BMI) = b_0 + b_1 BMI \tag{8}$$

$$b = (X^T X)^{-1} X^T y \tag{9}$$

$$\overline{y}_i = b_0 + b_1 x_i \tag{10}$$

Among them, x_i is the i-th indicator in the BMI indicator matrix X. Then, the residual r_i o1f each point is calculated: Among them, y_i represents the actual value of the gestational week of detection, and $\overline{\mathrm{y}}_\mathrm{i}$ represents the predicted value of the gestational week of detection. If $\mathrm{r} > 0$, the point is above the regression line; if $\mathrm{r} < 0$, the point is below the regression line. Finally, the residuals r_i are weighted, and then the weighted least squares method is used to update the coefficients. The following is the core part of quantile regression. First, a weight matrix $\mathrm{W} = \mathrm{diag}(\mathrm{w}_1, \mathrm{w}_2, \ldots, \mathrm{w}_\mathrm{n})$ is constructed. For each residual r_i:

$$r_i = y_i - \overline{y}_i \tag{11}$$

$$w_i = \begin{cases} \tau \cdot \frac{1}{|r_i| + \epsilon}, & r_i \geq 0 \\ (1-\tau) \cdot \frac{1}{|r_i| + \epsilon}, & r_i < 0 \end{cases} \tag{12}$$

Among them, τ is the quantile parameter, and ϵ is the random noise based on a normal distribution. Subsequently, the weighted least squares solution is solved:

$$b_{new} = (X^T W X)^{-1} X^T W y \tag{13}$$

Finally, the difference between the new parameter b_{new} and the old parameter b_{old} is compared. A expected value of 10^{-6} is specified. If the difference between the new and old parameters does not meet the convergence requirement, the iteration is continued until the absolute value of the difference is less than the following formula, and then it is considered that sufficient convergence has been achieved, and the iteration is stopped: Finally, the parameters b_0 and b_1 required by the model solution are obtained. At this point, the quantile regression model can predict the earliest time when the detection threshold (Y-chromosome concentration > 4%) can be reached in the gestational weeks (10 - 25 weeks) under different BMIs.

$$|b_{new} - b_{old}| \leq 10^{-6} \tag{14}$$

3.2 Solution of the Model

MATLAB is used for programming calculation, and the grouping results obtained are shown in the following table. It can be observed from the figure that the grouping based on BMI indicators is divided into four intervals, which are [26.62, 30.19], [30.20, 31.67], [31.68, 33.71], and [33.72, 39.52] respectively. The optimal NIPT times are 12.59 (with an error range of ±0.09), 12.87 (with an error range of ±0.05), 13.06 (with an error range of ±0.07), and 13.31 (with an error range of ±0.19) respectively. The results are visualized as shown in the following table 2 and figure 5:

Table 2. BMI Interval Distribution and Optimal NIPT Time Prediction for Each Group

Group	BMI Interval	Optimal NIPT Time
1	[26.62, 30.19]	12.59 weeks
2	[30.20, 31.67]	12.87 weeks
3	[31.68, 33.71]	13.06 weeks
4	[33.72, 39.52]	13.31 weeks

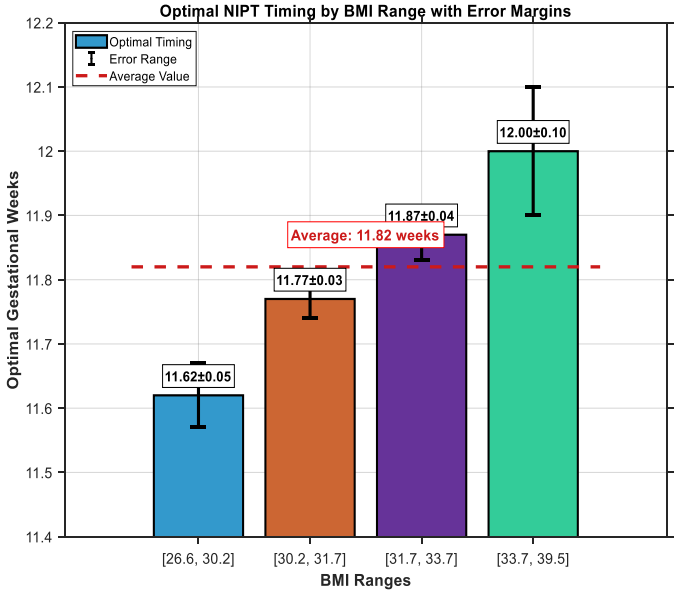

Figure 5. Optimal NIPT Detection Time and Error Range for Different BMI Intervals

3.3 Error Test and Result Analysis

It can be seen from the results that the higher the BMI, the later the gestational week is needed to ensure the success of the detection. This is because the circulating cell-free DNA of pregnant women with high BMI is diluted, resulting in a slow increase in Y-concentration. Women with low BMI can undergo NIPT as early as around 12.6 weeks, and there is almost no need to worry about detection failure; women with moderate BMI are recommended to be tested between 12.9 and 13.1 weeks for greater safety; the optimal time for women with high BMI is delayed to around 13.3 weeks, and there is a large individual difference. Doctors may need to delay the test to 13.5 weeks or even later according to specific circumstances to ensure the success rate. That is to say, as BMI increases, the optimal NIPT detection time is gradually delayed and the individual difference (detection error) increases. Pregnant women with high BMI need to be tested in a later gestational week to ensure the success rate. To sum up, in most cases, the detection error of pregnant women with medium or low BMI indicators will not have a negligible impact on the estimation of the optimal detection time, because their error ranges are relatively small.

4. Conclusion

Quantitative analysis of the association between fetal Y chromosome concentration and maternal indicators revealed a negative correlation between Y chromosome concentration and BMI, and a positive correlation with gestational age. The constructed multivariate polynomial regression model, incorporating quadratic terms, interaction terms, and random noise, successfully fitted the nonlinear relationship between Y chromosome concentration and factors such as BMI and gestational age. The F-test and p-value ($p<0.001$) both indicated the overall statistical significance of the model. With an RMSE of 0.027322 and MAE of 0.022227, the model demonstrated acceptable predictive accuracy. To determine the optimal NIPT timing based on BMI-stratified groups, BMI was grouped using quartiles and optimized via quantile regression. The study identified optimal NIPT timings of 12.59 weeks, 12.87 weeks, 13.06 weeks, and 13.31 weeks for the four pregnant groups. Results indicate that the optimal NIPT timing progressively delays with increasing BMI. This occurs because circulating cell-free DNA is diluted in high-BMI pregnant women, leading to slower Y chromosome concentration increases. Concurrently, inter-individual variability also increases in the high-BMI group.

References

[1] Yan Xiaoje, Chen Hong, Li Yufeng, et al. Prenatal Diagnosis and Clinical Phenotype Analysis of a Fetus with Chromosome 3 Fragment Recombination and Chromosome 5 Fragment Deletion: A Case Report [J]. Chinese Journal of Eugenics and Genetics, 2025, 33 (06): 1408-1413.

[2] Deng Zehua, Sun Youpeng, Wen Lijuan, et al. Association between Maternal Peripheral Blood Cell-Free Fetal DNA Concentration and Gestational Hypertension [J]. Chinese Journal of Prenatal Diagnosis (Electronic Edition), 2025, 17 (02):1-6.

[3] Xu Xinnni, Zhang Zhuangping, Hou Yaping. Advances in the Application of Non-Invasive Prenatal Screening in Twin Pregnancies with Fetal Loss [J]. Chinese Journal of Prenatal Diagnosis (Electronic Edition), 2025, 17 (02): 39-44.

[4] Pan Haotian. Correlation Study Between Maternal Plasma Fetal Cell-Free DNA Concentration and Early-Onset Fetal Growth Restriction [D]. Hubei Medical College, 2025.

[5] Yang Jinhua, Ren Daoju, Li Xiaowei, et al. Accuracy and Feasibility of Non-invasive Fetal Free DNA RHE Blood Type Genotyping Technology [J]. Chinese Journal of Transfusion Medicine, 2025, 38 (03): 368-374.

[6] Jin Yuxia, Wang Lu, Mao Qiaoling, et al. Non-invasive screening for dominant single-gene hereditary diseases in fetuses based on UMI molecular tags [J]. Chinese Journal of Eugenics and Genetics, 2025, 33 (03): 680-685.

[7] Zhao Jia, Song Shuo. Research progress on maternal urine cell-free DNA in fetal chromosomal abnormality detection [J]. Women and Children's Health Journal, 2025, 4 (05): 27-31+36.

[8] Du Limin, Guo Mengyu, Liu Shuhan. Application of Maternal Peripheral Blood Cell-Free Fetal DNA Prenatal Testing in Diagnosing Trisomy 13, Trisomy 18, and Down Syndrome [J]. Journal of Practical Medical Technology, 2025, 32 (03): 204-206.

[9] Xie Zhehui. Application Research of Non-Invasive Prenatal Testing in Prenatal Screening for Fetal Chromosomal Disorders in Assisted Reproduction and Natural Pregnancies [D]. Lanzhou University, 2025.

[10] Li Miyun, Xiao Wenguang, Niu Guangxia, et al. Clinical Value Analysis of PlGF Combined with Fetal Cell-Free DNA in Predicting Preeclampsia Risk [J]. Chinese Journal of Eugenics and Genetics, 2025, 33 (02): 352-357.