ISSN 2616-5775 Vol. 9, Issue 6: 74-83, DOI: 10.25236/AJCIS.2026.090609

Research on NIPT Timing Selection and Fetal Abnormality Judgment Based on Clustering Optimization and Dynamic Risk Control

Rui Ge[1,*], Guanting Chen[2], Yilin Yang[1]

[1]*School of Statistics and Mathematics, Henan Finance University, Zhengzhou, China, 450046*
[2]*School of Computer Science and Artificial Intelligence, Henan Finance University, Zhengzhou, China, 450046*
**Corresponding author: 13253456096@163.com*

Abstract: *This paper focuses on solving the problems of timing selection and evaluation for fetal chromosomal abnormality detection in Non-Invasive Prenatal Testing (NIPT). Based on actual clinical data, it comprehensively applies correlation analysis, clustering algorithms, risk control models and machine learning methods. The study successfully constructs a fetal Y-chromosome concentration prediction model, a BMI grouping optimization model, a multi-factor clustering model and a female fetal abnormality judgment model. Furthermore, it obtains the optimal strategies for various problems in NIPT. For the correlation between fetal Y-chromosome concentration and indicators such as gestational age, BMI and age of pregnant women, this paper uses the Pearson correlation coefficient equation to analyze the linear correlation between Y-chromosome concentration and gestational age, BMI and age. Then, it constructs a multiple linear regression model of Y-chromosome concentration with gestational age, BMI and age. Finally, F-test and t-test are used to verify the overall significance of the model and the statistical significance of each variable. To reasonably group the BMI of pregnant women carrying male fetuses, this paper uses the silhouette coefficient to compare and analyze various clustering algorithms, and selects the optimal K-Means++ clustering algorithm to divide them into two groups. The ranges of statistical quantities of the two groups are calculated, and then a dynamic risk control model is constructed using the quantile estimation algorithm to determine the optimal detection time point. The results show that when BMI is in the range of [26.6, 32.4], the optimal detection time is 24.11 gestational weeks; when BMI is in the range of [32.5, 41.1], the optimal detection time is 21.33 gestational weeks. By calculating the actual delayed risk ratio, the risk is controlled below 5% and passes the test. With the help of the linear fitting relationship model for auxiliary analysis, the relationship between BMI and the time to meet the standard within the group is further clarified. Finally, the Bootstrap resampling algorithm is used to verify the stability of the model in predicting the optimal detection time point.*

Keywords: *K-Means++ Clustering; Quantile Optimization; Multivariate Clustering*

1. Introduction

NIPT, or Non-Invasive Prenatal Testing, is a prenatal testing technology for early detection of fetal health status. It collects maternal blood to obtain and detect cell-free fetal DNA fragments, and then analyzes whether fetal chromosomes are abnormal. Clinically, three types of diseases caused by chromosomal abnormalities are the most common: Down syndrome (chromosome 21), Edwards syndrome (chromosome 18) and Patau syndrome (chromosome 13)[1]. The accuracy of non-invasive prenatal testing is mainly based on the concentration of fetal sex chromosomes. That is, when the Y-chromosome concentration of male fetuses reaches or exceeds 4% during gestational age testing, and there is no abnormality in the X-chromosome of female fetuses, the NIPT results are basically accurate. Moreover, the risk of treatment for unhealthy fetuses increases with the increase of gestational age [2].

According to clinical experience, the Y-chromosome concentration of male fetuses is also significantly correlated with gestational age and the body mass index (BMI) of pregnant women. To ensure the accuracy of detection and the timeliness of subsequent treatment, pregnant women are usually grouped to determine their NIPT detection time points. However, each pregnant woman has other individual differences such as age and pregnancy status. Determining the NIPT detection time points for all pregnant women only based on experience will greatly affect the accuracy of detection

ISSN 2616-5775 Vol. 9, Issue 6: 74-83, DOI: 10.25236/AJCIS.2026.090609

results.

To accurately analyze the suitable NIPT time points for each pregnant woman and evaluate their accuracy, NIPT data of pregnant women in a certain region are used to establish a mathematical model to solve the problem. In the actual research, it is necessary to consider the impact of failed sequencing caused by uncertain factors and the situation of multiple blood collections or multiple tests with one blood collection for some pregnant women to increase the certainty of detection results. (Data source: Official website of Mathematical Modeling Contest http://www.mcm.edu.cn/)

This paper analyzes the correlation characteristics between fetal Y-chromosome concentration and indicators such as gestational age and BMI of pregnant women, and establishes corresponding models. Firstly, it identifies the main factors affecting Y-chromosome concentration and their action directions. However, simple linear regression cannot fully reflect individual differences, so multivariate statistical analysis methods are needed. It is proposed to initially study the relationship between variables through univariate correlation analysis, then establish a multiple linear regression model to quantitatively analyze the influence degree of each factor, and finally verify the statistical significance of the model through significance tests. At the same time, pregnant women carrying male fetuses are grouped according to BMI, and the optimal NIPT detection time point for each group is determined to minimize potential risks. The core is to find a BMI grouping scheme that can not only reflect the homogeneity within the group but also ensure the heterogeneity between groups, and determine the risk-controllable detection time point according to the distribution of time to meet the standard. It is proposed to use clustering algorithms to achieve BMI grouping, determine the optimal time point by quantile estimation method, and finally use resampling method to verify the stability of the model and analyze the error impact[3].

2. Study on the Linear Relationship between Chromosome Concentration and Pregnant Women's Characteristics

2.1 Univariate Correlation Analysis

To initially study the linear relationship between Y-chromosome concentration and pregnant women's characteristics and provide directional basis for subsequent regression modeling, this paper selects three key variables of pregnant women: gestational age, BMI and age, and uses Pearson correlation coefficient to test the correlation between each pair of variables respectively. The formula of Pearson correlation coefficient is as follows:

For variables X and Y:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{1}$$

Among them, the value range of the correlation coefficient: r>0 indicates positive correlation, r<0 indicates negative correlation, and the closer r is to 1, the stronger the correlation.

This paper calculates the Pearson correlation coefficient and draws a correlation matrix heat map:

Table 1 Correlation Analysis between Y-Chromosome Concentration and Various Variables

Correlated Variables	Correlation Coefficient (r)	p-value
Gestational Age at Detection	0.1215	0.0023
Pregnant Women's BMI	-0.1584	0.0001
Pregnant Women's Age	-0.1054	0.0082

The following results are obtained through the above correlation analysis:

It can be seen from Table 1 that Y-chromosome concentration is significantly positively correlated with gestational age at detection, that is, Y concentration generally increases with the increase of gestational age; it is significantly negatively correlated with pregnant women's BMI, that is, pregnant women with higher BMI tend to have lower Y-chromosome concentration; it is significantly negatively correlated with pregnant women's age, that is, Y-chromosome concentration has a slight downward trend with the increase of age.

ISSN 2616-5775 Vol. 9, Issue 6: 74-83, DOI: 10.25236/AJCIS.2026.090609

The correlation matrix heat map is shown in Figure 1:

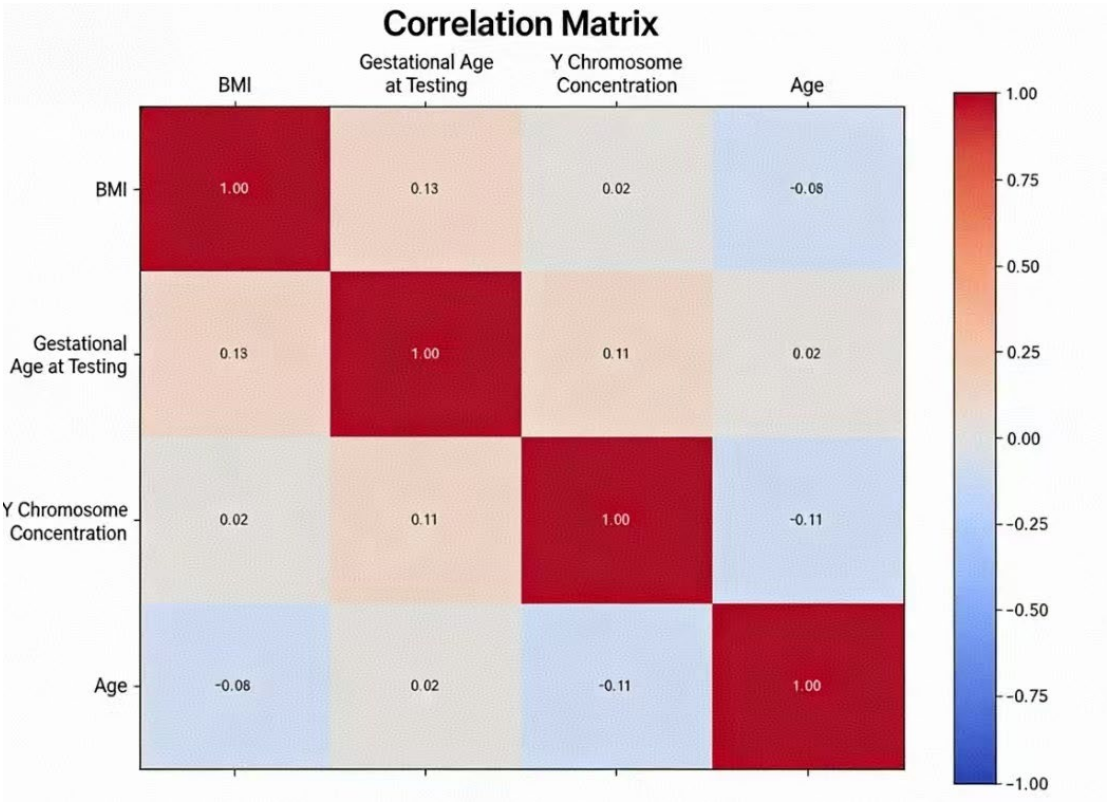

Figure 1 Correlation Matrix Heat Map

It can be seen from the row or column where Y-chromosome concentration is located that it is weakly negatively correlated with pregnant women's BMI (-0.16) and age (-0.11), and weakly positively correlated with gestational age at detection (0.11); the absolute values of correlation coefficients between each independent variable are relatively small, indicating that there is a weak multicollinearity between the independent variables selected by the model.

The results of correlation analysis indicate that there is value for further modeling and analysis between variables.

2.2 Multiple Regression Modeling

Univariate analysis can only reflect the relationship between two corresponding variables and cannot exclude the mutual interference between variables[4]. To further evaluate the independent influence of these three factors on Y-chromosome concentration, this paper establishes a multiple regression model under the condition of controlling the mutual influence between variables, so as to quantitatively describe the simultaneous influence of pregnant women's BMI, gestational age at detection and age on Y concentration. The form of multiple linear regression model is as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon \tag{2}$$

Among them: Y is the dependent variable, $X_1, \ldots, X_k$ are independent variables, β_0 is the intercept, $\beta_1, \ldots, \beta_k$ are regression coefficients, and $\varepsilon \sim N(0, \sigma^2)$ is the random error. The fitting results are shown in Table 2:

Table 2 Multiple Linear Regression Analysis of Influencing Factors of Y-Chromosome Concentration

ISSN 2616-5775 Vol. 9, Issue 6: 74-83, DOI: 10.25236/AJCIS.2026.090609

Variables	Regression Coefficient	Standard Error	t-value	p-value	95% Confidence Interval Lower Limit	95% Confidence Interval Upper Limit
Constant Term	0.4772	0.0244	19.539	<0.001	0.4292	0.5251
Pregnant Women's BMI	-0.2456	0.0542	-4.527	<0.001	-0.3521	-0.1391
Gestational Age at Detection	0.1263	0.0335	3.773	<0.001	0.0606	0.1920
Pregnant Women's Age	-0.1180	0.0433	-2.726	0.007	-0.2031	-0.0330

The following results are obtained through the above analysis:

The intercept term β_0 = 0.4772;the coefficient of pregnant women's BMI β_1 = -0.2456, that is, when other conditions remain unchanged, for each 1-unit increase in BMI,Y concentration decreases by about 0.246 on average; the coefficient of gestational age at detection β_2 =0.1263, that is, for each 1-week increase in gestational age, Y concentration increases by about 0.126 on average; the coefficient of pregnant women's age β_3 =-0.1180, that is, for each 1-year increase in pregnant women's age, Y concentration decreases by about 0.118 on average.

2.3 Model and Variable Significance Test

To verify the validity of the established multiple linear regression model and to test whether the influence of each independent variable on the dependent variable is truly effective rather than caused by random error, this paper conducts F-tests and T-tests on the regression model (corresponding respectively to the significance of the overall model and each regression coefficient statistically). The algorithms for the F-test and T-test are as follows:

F-test (Overall Significance): Tests whether all independent variables together have a significant effect on the dependent variable, null hypothesis $H_0 : \beta_1 = \ldots = \beta_k = 0$

$$F = \frac{SSR / k}{SSE / (n-k-1)} \sim F(k, n-k-1) \tag{3}$$

Among them, SSR is the sum of squares due to regression (variation explained by the independent variable), and SSE is the sum of squared errors (unexplained variation). T-test (variable significance): Tests whether a particular independent variable has a significant effect on the dependent variable, with the null hypothesis $H_0 : \beta_i = 0$

$$t = \frac{\hat{\beta}_i}{SE\left(\hat{\beta}_i\right)} \sim t(n-k-1) \tag{4}$$

Among them, $\hat{\beta}_i$ is the coefficient estimate, $SE(\hat{\beta}_i)$ is the standard error, n is the sample size, and k is the number of independent variables.

The test results are shown in Table 3:

Table 3 Multiple Linear Regression Analysis and F-test of Factors Affecting Y-chromosome Concentration

Variables	Regression Coefficient	Standard Error	t-value	p-value
Consant Term	0.4772	0.0244	19.539	<0.001
Pregnant Women's BMI	-0.2456	0.0542	-4.527	<0.001

ISSN 2616-5775 Vol. 9, Issue 6: 74-83, DOI: 10.25236/AJCIS.2026.090609

Gestational Age at Detection	0.1263	0.0335	3.773	<0.001
Pregnant Women's Age	-0.1180	0.0433	-2.726	0.007
F-test				0.00000005

It can be known from the test results:

(1)The p-value of F-test is <0.001, indicating that the regression model is highly significant overall. This result proves that at least one independent variable is meaningful for explaining the change of Y-chromosome concentration, and the null hypothesis that "all coefficients are zero" is not valid.

(2)The p-values of all independent variables are <0.05, indicating that under the condition of controlling other variables, pregnant women's BMI, gestational age at detection and age are all significant influencing factors of Y-chromosome concentration.

On this basis, the direction and intensity of the influence are further analyzed:

(1)The coefficient of pregnant women's BMI is negative, that is, BMI has a significant negative impact on Y-chromosome concentration, and the higher the BMI, the lower the Y-chromosome concentration tends to be. Moreover, the absolute value of BMI coefficient is the largest among the three variables, indicating that it is the strongest negative influencing factor.

(2)The coefficient of gestational age at detection is positive, that is, gestational age has a significant positive impact on Y-chromosome concentration, and Y-chromosome concentration will increase slowly with the increase of gestational age.

(3)The coefficient of pregnant women's age is negative, that is, age has a significant negative impact on Y-chromosome concentration, and the older the pregnant woman, the lower the Y-chromosome concentration tends to be. According to clinical experience, the Y-chromosome concentration of male fetuses is also significantly correlated with gestational age and the body mass index (BMI) of pregnant women. To ensure the accuracy of detection and the timeliness of subsequent treatment, pregnant women are usually grouped to determine their NIPT detection time points. However, each pregnant woman has other individual differences such as age and pregnancy status. Determining the NIPT detection time points for all pregnant women only based on experience will greatly affect the accuracy of detection results.

To accurately analyze the suitable NIPT time points for each pregnant woman and evaluate their accuracy, NIPT data of pregnant women in a certain region are used to establish a mathematical model to solve the problem. In the actual research, it is necessary to consider the impact of failed sequencing caused by uncertain factors and the situation of multiple blood collections or multiple tests with one blood collection for some pregnant women to increase the certainty of detection results. The coefficient of pregnant women's age is -0.1180, indicating that age has a significant negative impact on Y-chromosome concentration, that is, the older the pregnant woman, the lower the Y-chromosome concentration tends to be[5].

Finally, in the multiple linear regression model, the standard error is an important indicator to evaluate the estimation accuracy of regression coefficients, which shows the possible fluctuation range of coefficient estimates under different sample conditions. The smaller the standard error, the more stable the estimation of the coefficient. The standard errors of each variable coefficient based on the model output are shown in Table 4:

Table 4 Standard Errors

Const	Pregnant Women's BMI	Gestational Age at Detection	Age
0.024423	0.054249	0.033476	0.043302

It can be seen from the standard error analysis data that the coefficients of the independent variables Const, gestational age at detection and age are all significant (p<0.05), which proves that the model established in this study successfully identifies the significant influence of gestational age at detection and age on Y-chromosome concentration. Although the coefficient of BMI is relatively large, its p-value is extremely low (p=0.054249), and its significance strongly supports the core conclusion that "BMI is an important factor affecting Y-chromosome concentration".

ISSN 2616-5775 Vol. 9, Issue 6: 74-83, DOI: 10.25236/AJCIS.2026.090609

3. Study on Pregnant Women's BMI and Time to Meet the Standard

The main variables used in this study are pregnant women's BMI and time to meet the standard (i.e., the time when Y-chromosome concentration reaches 4% for the first time). All male fetus samples are extracted from the basic preprocessed data set, and all samples that fail to meet the standard in all tests are excluded. The BMI value of each pregnant woman is taken as the value at the first detection to avoid deviations caused by multiple measurements. Through the above processing, 239 valid data points are obtained, including the main variables used in the study: pregnant women's BMI and time to meet the standard.

3.1 Determination of Clustering Method and BMI Grouping Scheme

To find the optimal BMI grouping scheme that ensures high similarity of BMI within the group and obvious separation between groups, this paper systematically compares the performance of four clustering algorithms under different numbers of groups (K), and uses the silhouette coefficient as the evaluation index of clustering quality. The closer the silhouette coefficient is to 1, the better the clustering effect[6][7].

The definition of silhouette coefficient is as follows:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{5}$$

Among them, $a(i)$is the average distance of sample i within its own cluster, and $b(i)$is the average distance of sample i from the nearest neighboring cluster. The overall silhouette coefficient is the average value of $s(i)$ of all samples.

Calculations are performed for each clustering method when the number of groups K=2, 3, 4, 5 respectively, and the results are shown in Table 5:

Table 5 Comparison of Silhouette Coefficients of Different Clustering Methods and K Values

Clustering Method	K=2	K=3	K=4	K=5	Optimal K	Optimal Score
K-Means	0.6007	0.5897	0.5385	0.5417	2	0.6007
Equal-width Ordered Clustering	0.5392	0.5560	0.4626	0.5166	3	0.5560
Equal-frequency Ordered Clustering	0.5180	0.4384	0.4517	0.4451	2	0.5180
Hierarchical Clustering	0.5691	0.5117	0.5318	0.5094	2	0.5691

It can be known from the above analysis that the K-Means++ method has the highest silhouette coefficient (0.6007) when K=2, and its clustering effect is significantly better than other methods. Therefore, K-Means (K=2) is selected as the final grouping method.

The K-Means++ algorithm is run to output the sample sets of the two clusters, and their clustering distribution is shown in Figures 2 to 4:

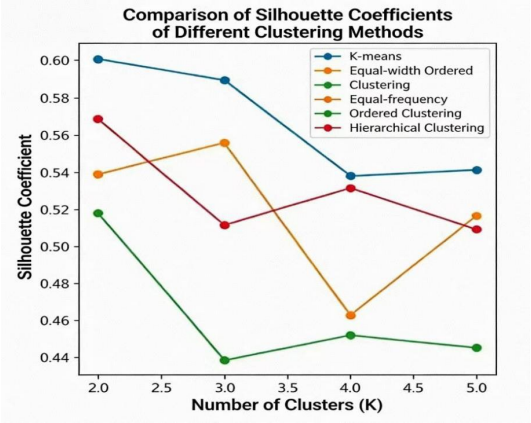

Figure 2 Comparison Chart of Various Clustering Algorithms

ISSN 2616-5775 Vol. 9, Issue 6: 74-83, DOI: 10.25236/AJCIS.2026.090609

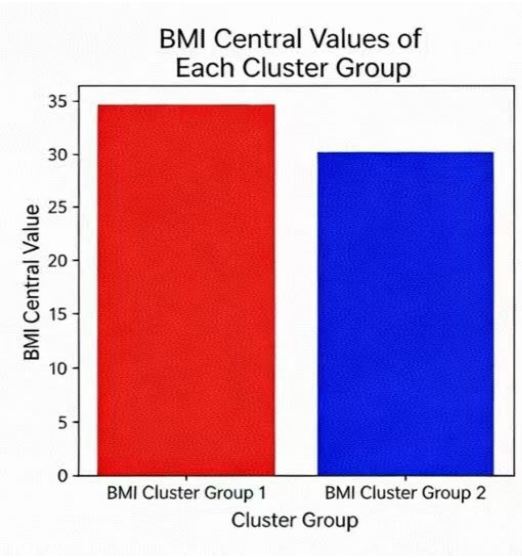

Figure 3 BMI Center Values of Each Cluster Group

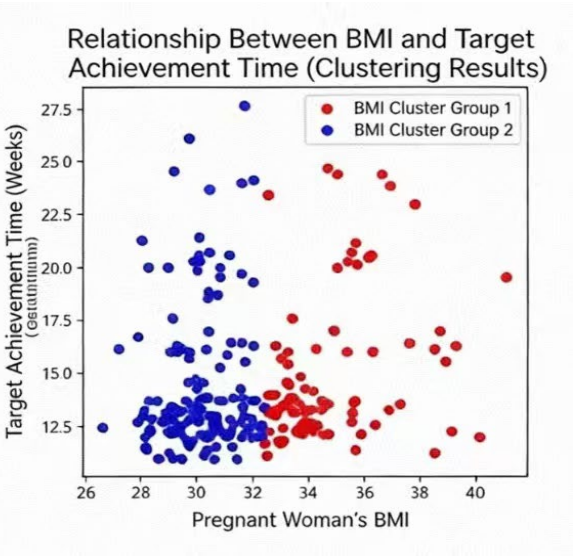

Figure 4 Relationship between BMI and Time to Meet the Standard

The specific grouping intervals are: BMI Cluster Group 1 (high BMI group): BMI ∈ [32.5, 41.1], BMI Cluster Group 2 (low BMI group): BMI ∈ [26.6, 32.4]. The above analysis shows that the two groups are clearly separated in BMI, and the time to meet the standard tends to be slightly delayed with the increase of BMI.

3.2 Determination of Optimal NIPT Time Point

After obtaining the BMI grouping based on K-Means++ (K=2) and the distribution characteristic information of time to meet the standard, this paper calculates the descriptive statistics of time to meet the standard for each BMI group, including mean, standard deviation and value range. The descriptive statistics of time to meet the standard of the two groups are shown in Table 6:

Table 6 Descriptive Statistics of Time to Meet the Standard of the Two Groups

BMI Cluster Group	Mean of Time to Meet the Standard	Standard Deviation of Time to Meet the Standard	Range of Time to Meet the Standard
Group 1	15.04	3.62	11.14-26.43
Group 2	14.42	3.38	11.00-27.71

The above results suggest that BMI has a certain impact on the time to meet the standard.

To determine the 'optimal testing time point' for the time distribution of compliance in each group, this paper establishes a dynamic risk control model: At this time point, the probability that the target achievement time of pregnant women within the group is later than a specified time $T_{optimal}$ (i.e., the risk of delay) should not exceed an acceptable upper limit α.According to clinical practice, this study sets the risk threshold α=5%.Its solution is equivalent to solving the (1-α) quantile of the distribution of time to meet the standard, that is, $T_{optimal} = F^{-1}(0.95)$, where F is the empirical distribution function[8].

Finally, it is determined that:

BMI Cluster Group 1 (high BMI group): The optimal detection time $T_1^* = 24.11$ gestational weeks.

BMI Cluster Group 2 (low BMI group): The optimal detection time $T_2^* = 21.33$ gestational weeks.

To verify the risk control effect, this paper calculates the actual delayed risk ratio, that is, the proportion of samples in the group whose time to meet the standard is later than the optimal time point. The results show that the actual risks of the two groups are successfully controlled below 5%, which proves the effectiveness of the model. The data of BMI grouping characteristics and optimal detection

ISSN 2616-5775 Vol. 9, Issue 6: 74-83, DOI: 10.25236/AJCIS.2026.090609

time points are shown in Table 7:

Table 7 BMI Grouping Characteristics and Optimal Detection Time Points

Group	BMI Range	Sample Size	Mean of Time to Meet the Standard (Weeks)	Standard Deviation of Time to Meet the Standard	Optimal Detection Time Point (Weeks)	Actual Risk Ratio
BMI Group1	32.5-41.1	90	15.04	3.62	24.11	4.44%
BMI Group2	26.6-32.4	149	14.42	3.38	21.33	4.70%

On the basis of determining the optimal detection time point, to further study the relationship between BMI and time to meet the standard within each group, this paper establishes a linear regression model for each group to assist the analysis. The fitting results are shown in Figure 5:

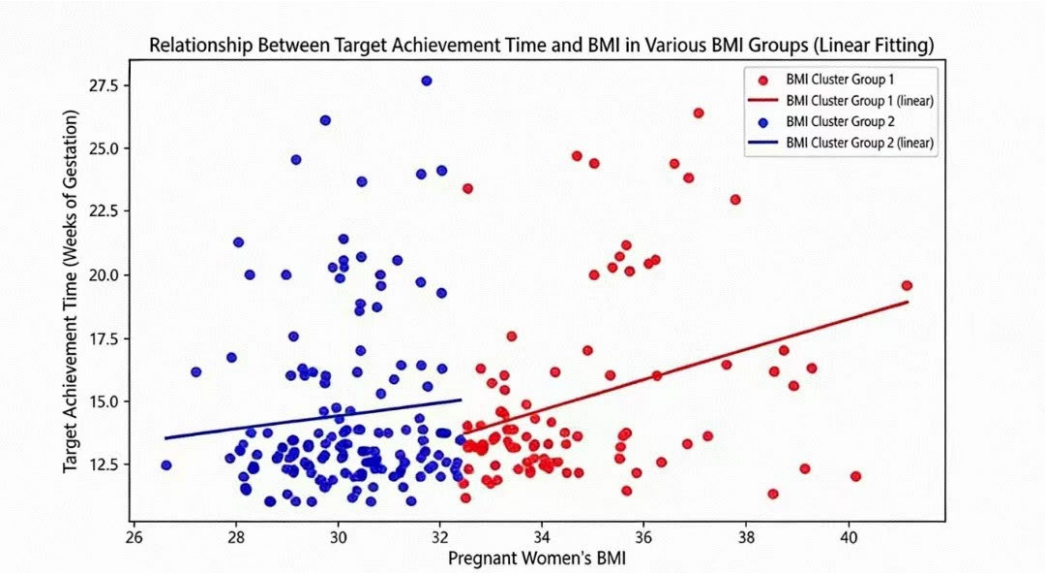

Figure 5 Relationship between Time to Meet the Standard and BMI of Each BMI Group (Linear Fitting)

It can be seen from Figure 5 that within each group, the time to meet the standard tends to be slightly delayed with the increase of BMI, but the slope of the regression line is small, that is, this tendency is not obvious, which proves that BMI is not the main influencing factor of time to meet the standard within each group. Therefore, this paper adopts the risk control strategy based on the overall distribution (quantile method) to determine the optimal time point, rather than the prediction based on the regression model.

3.3 Stability Analysis of the Model

Finally, to test the stability of the optimal time point estimation results, this paper conducts Bootstrap sampling with replacement: 100 Bootstrap data sets are constructed by repeated sampling (B=100 times) from the original data, and then the above clustering and optimization process is re-executed on each new data set. Thus, 100 estimated values of the optimal time point are obtained, and the mean and 95% confidence interval of the optimal time point estimated values are calculated. The calculation results are as follows: BMI Cluster Group 1: the Bootstrap mean of the optimal time point is 23.56 gestational weeks, 95% CI [20.54, 24.63]; BMI Cluster Group 2: the Bootstrap mean of the optimal time point is 21.74 gestational weeks, 95% CI [20.00, 24.26].

The above estimated values of the optimal time point are all reasonable and fall within the 95% confidence interval. It is proved that the model is not sensitive to random fluctuations, and its optimal time point estimation results have good stability[9].

3.4 Analysis of the Impact of Errors on Results

The possible errors in this study are mainly the measurement error of time to meet the standard and the measurement error of BMI.

ISSN 2616-5775 Vol. 9, Issue 6: 74-83, DOI: 10.25236/AJCIS.2026.090609

Table 8 Analysis Results

Group	BMI Range	Optimal Detection Time Point (Gestational Weeks)	Number of Data Points	Risk Ratio	Bootstrap Mean	Bootstrap Confidence Interval	Model Type
BMI Group 1	32.5-41.1	24.10857	90	0.044444	23.55546	[20.54,24.63]	linear
BMI Group 2	26.6-32.4	21.33429	149	0.04698	21.73931	[20.00,24.26]	linear

It can be concluded from Table 8:

(1)If there is non-systematic random error in the measurement of time to meet the standard, the Bootstrap test has proved the stability of our model to such errors; if there is systematic deviation (such as all measured values are too small), it will directly lead to the synchronous shift of the estimation of the optimal time point T_{optimal} .The impact degree can be quantified through sensitivity analysis[10].

(2)The two groups obtained by this modeling have obvious separation (the distance band is about 0.1), so the slight measurement error of BMI that may affect the clustering accuracy will hardly lead to samples being incorrectly divided into another group. Therefore, such grouping is robust to BMI measurement errors[11].

4. Conclusion

This study investigates the optimal timing for detecting fetal chromosomal abnormalities in non-invasive prenatal testing (NIPT), aiming to provide scientific and precise guidance on testing time for clinical practice. The findings reveal that the concentration of Y-chromosome is positively correlated with gestational age, and negatively correlated with maternal body mass index (BMI) and age.

For data processing, the K-Means++ clustering algorithm was adopted in this study to stratify pregnant women into a high-BMI group ([32.5,41.1]) and a low-BMI group ([26.6,32.4]). Based on this grouping, the quantile estimation algorithm was used to determine the optimal testing time: the 24.11th week of gestation for the high-BMI group and the 21.33rd week for the low-BMI group. With the risk of testing delay controlled below 5%, this timing can reduce the risks of missed detection and false positive results.

Verification via Bootstrap resampling analysis demonstrated that the model has a strong ability to resist random fluctuations, with stable and reliable estimation results. The grouping protocol exhibits good robustness to BMI measurement errors, which do not affect the group classification or the determination of the optimal testing time.

From the perspective of application feasibility, the research results of this paper have high clinical application value. The determined optimal detection time points for different BMI groups provide clear and specific guidance for medical staff in practical operations, which helps to improve the accuracy and effectiveness of NIPT detection, thereby improving the level of prenatal diagnosis and providing strong support for ensuring fetal health. This study can further expand the research scope, include more factors affecting fetal chromosome concentration, such as pregnant women's living habits and genetic background, to build a more comprehensive and accurate prediction model.

References

[1] Li J, Liu Y, Zhang N, et al. Optimal planning of energy storage in active distribution network for congestion management and voltage support [J]. IEEE Transactions on Power Systems, 2020, 35 (5): 4120-4133.

[2] Li Z, Wang C, Li Y, et al. A dynamic reconfiguration method for active distribution networks considering distributed generation uncertainty [J]. International Journal of Electrical Power & Energy Systems, 2021, 124: 106337.

[3] Wang Y, Chen J, Chen X, et al. Short-term load forecasting for industrial customers based on TCN-LightGBM [J]. IEEE Transactions on Power Systems, 2021, 36 (3): 1984-1997.

[4] Li F, Han Y. A data-driven approach for customer baseline load estimation based on density peak

ISSN 2616-5775 Vol. 9, Issue 6: 74-83, DOI: 10.25236/AJCIS.2026.090609

clustering [J]. IEEE Transactions on Smart Grid, 2022, 13 (1): 653-664.

[5] LifeCycle Project-Maternal Obesity and Childhood Outcomes Study Group. Association of gestational weight gain with adverse maternal and infant outcomes [J]. JAMA, 2019, 321 (17): 1702-1715.

[6] Norton M E, Jacobsson B, Swamy G K, et al. Cell-free DNA analysis for noninvasive examination of trisomy [J]. New England Journal of Medicine, 2015, 372 (17): 1589-1597.

[7] Anifowose F, Labadin J, Abdulraheem A. Improving the prediction of petroleum reservoir characterization with a stacked generalization ensemble model of support vector machines [J]. Applied Soft Computing, 2017, 54: 376-388.

[8] Jain A K, Murty M N, Flynn P J. Data Clustering: A Review [J]. ACM Computing Surveys (CSUR), 2020, 53 (3): 1-38.

[9] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction [M]. 4th ed. Springer Science & Business Media, 2021: 124-156.

[10] James G, Witten D, Hastie T, et al. An Introduction to Statistical Learning: with Applications in R [M]. 2nd ed. Springer, 2021: 89-112.

[11] Efron B, Tibshirani R J. An Introduction to the Bootstrap [M]. Revised ed. Chapman & Hall/CRC, 2020: 67-92.