WEP Warwick Evans Publishing

International Journal of Public Health and Medical Research

ISSN: 3006-7707 (Print), ISSN: 3006-3752 (Online) | Volume 6, Number 2, Year 2026
DOI: https://doi.org/10.62051/ijphmr.v6n2.06
Journal homepage: https://wepub.org/index.php/IJPHMR/index

Research on Dynamic Timing Decision-Making for Non-Invasive Prenatal Testing Based on Risk-Error Co-Optimization

Yanting Liu *

School of Mathematics and Artificial Intelligence, Chongqing University of Arts and Sciences, Chongqing, China

*Corresponding Author: 17265343973@163.com

ABSTRACT

Addressing the risks faced by high-BMI pregnant women in non-invasive prenatal testing (NIPT)—namely delayed attainment of sufficient fetal cell-free DNA (cfDNA) concentration and missed optimal intervention windows—this study establishes a data-driven system for optimizing testing timepoints. The research first preprocesses raw samples by converting gestational age into continuous numerical variables and develops evaluation metrics encompassing clinical intervention risks and testing accuracy. To address the limitation of static grouping in distinguishing individual differences, the K-means clustering algorithm was employed to reclassify pregnant women into five mutually exclusive clusters based on BMI characteristics. Kaplan-Meier survival analysis was then applied to reveal the temporal evolution of Y chromosome concentration attainment rates across different BMI levels. Building upon this foundation, a nonlinear optimization model was constructed to balance sequencing failure risk and missed therapeutic window risk. By minimizing total risk and sequencing error, optimal detection timepoints were determined for each group. Validation results demonstrated that the optimal detection timepoint for the high-BMI group was significantly delayed compared to the low-BMI group, with a maximum deviation of up to 4 weeks. Finally, sensitivity analysis via Monte Carlo simulation confirmed the model's robustness to sequencing perturbations.

KEYWORDS

Non-invasive prenatal testing; K-means clustering; Risk-error optimization model

1. INTRODUCTION

With the widespread adoption of prenatal screening technologies, non-invasive prenatal testing (NIPT) has become the primary method for detecting fetal chromosomal aneuploidy. However, its accuracy heavily depends on the concentration of cell-free fetal DNA (cfDNA) in maternal blood. For high-BMI pregnant women, physiological factors such as blood dilution often result in fetal DNA concentrations below the 4% detection threshold, necessitating delayed testing. Previous studies predominantly employed fixed gestational age recommendations or coarse BMI stratification, failing to adequately balance the urgency of clinical intervention with the randomness of sequencing errors [1-2]. This approach leaves some pregnant women facing elevated decision-making risks. To address this, this section proposes a BMI-stratified approach for synergistic optimization of NIPT timing and risk assessment. Its innovation lies in abandoning traditional subjective grouping. Instead, it employs K-means clustering for high-density BMI samples and introduces a novel composite objective function incorporating Cox proportional hazards models and sequencing error weights, enabling multi-criteria trade-offs for optimal testing timing. The overall research framework follows a logical sequence: "data cleansing and format conversion → adaptive stratified clustering → survival analysis of target achievement probability → risk-error model construction → optimal timing decision and

sensitivity validation." This aims to provide precise, personalized clinical testing recommendations for pregnant women with varying constitutions [3-4].

Solving the Collaborative Optimization and Risk Assessment of NIPT Timing Based on BMI Stratification. Through refined data preprocessing, gestational age was converted to decimal format. Three tiers of clinical risk were defined based on testing periods, alongside establishing an accuracy evaluation system based on consistency between birth outcomes and test results. The grouping strategy employed K-means clustering to identify five core BMI clusters, addressing statistical inefficiency caused by uneven raw data distribution [5]. Kaplan-Meier analysis visually demonstrated the slow decline in the probability curve for the high-BMI group (Group 0), confirming that the gestational week required for Y-chromosome concentration to reach the 4% threshold was significantly later. Subsequently, a complex mathematical optimization model was developed. This model evaluates the cost of missing the therapeutic window through risk weighting ω and quantifies the probability of sequencing failure using the Cox proportional hazards model. Solving the objective function $\min\{W(B, T)-\alpha Err-\gamma Risk\}$ revealed that the optimal detection timing for the low BMI group (Group 4) was 18.5 weeks, while the extremely high BMI group (Group 0) required delaying detection to 22.5 weeks to ensure a 91.1% pass rate. Monte Carlo simulations further reveal that despite higher sequencing quality variability in the high-BMI group, the model's comprehensive error metrics remain controllable through dynamic threshold adjustments, validating the scientific rationale of the stratified decision strategy [6-7].

2. SOLVING THE COLLABORATIVE OPTIMIZATION AND RISK ASSESSMENT OF NIPT TIMING BASED ON BMI STRATIFICATION

2.1. Data Preprocessing

2.1.1. Data Merging

To facilitate model establishment and optimize algorithm performance, we extract columns including serial number, pregnant woman code, GC content, gestational age at detection, pregnant woman's BMI, Y chromosome concentration, chromosomal aneuploidy, and fetal health status [8-9].

2.1.2. Conversion of Gestational Age to Decimal Form

In the actual calculation process, we need to convert the data format into an input acceptable to the algorithm, i.e., convert the original format "weeks + days" into a decimal form with "weeks" as the unit.

2.1.3. Establishment of Risk Assessment Column

According to the problem description, we classify the risk of shortened treatment window as Level 1 if the gestational age at detection is less than 10 weeks, Level 3 if greater than 28 weeks, and Level 2 if between 13-27 weeks [10].

2.1.4. Setting of Evaluation Criterion Column

To better demonstrate the accuracy of this NIPT, we represent the columns of fetal health status and chromosomal aneuploidy with 0-1. If the two columns are consistent, we record it as 1 (accurate detection result); if inconsistent, we record it as 0.

2.2. Initial Grouping of Raw Data

In the problem, BMI grouping is given as five groups: [20, 28), [28, 32), [32, 36), [36, 40), and above 40. By observing the data and analyzing the given condition that most pregnant women have high BMI, we question the initial grouping of the data provided in the problem. Next, we will draw boxplots and survival curves for these five BMI groups for preliminary analysis, and then judge

whether these five groups of data can be used as the grouping data for solving the optimal NIPT time point for each group through descriptive statistics. Initial BMI Grouping Statistics are shown in table 1.

Table 1. Initial BMI Grouping Statistics

BMI Group	Mean Value	Median Value
Above 40	19.55	20.14
[28, 32)	13.64	12.79
[32, 36)	14.69	13.29
[36, 40)	18.29	16.71

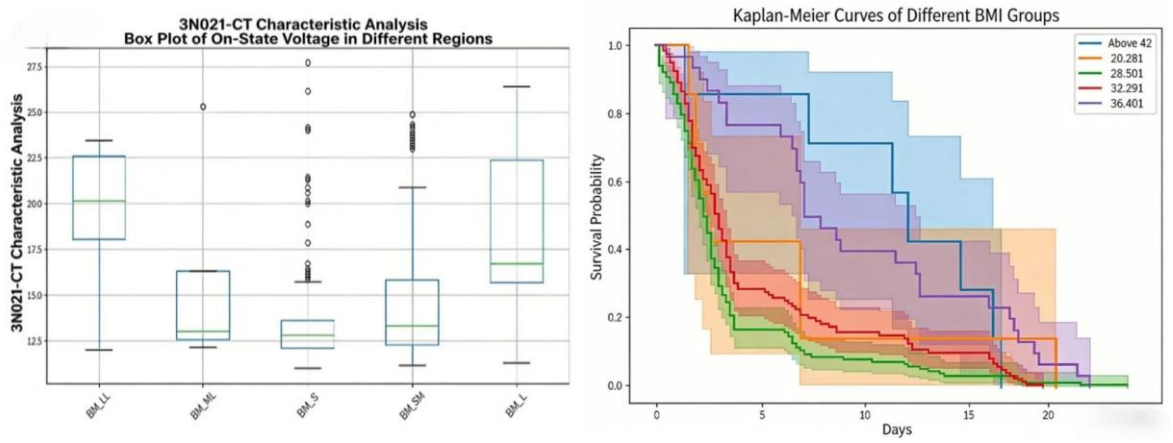

Figure 1. Boxplots and Kaplan-Meier curves of initial BMI groups

From the figure 1, the trend and separation degree of the survival curves show that the grouping of raw data does not well distinguish the differences under different BMI levels, and it reflects that the sample size between some groups is not large, and the survival spaces are relatively close; from the mean and median data on the right, there are differences in the qualified gestational age of different BMIs, but it is observed that the sample size of each group varies greatly, and the reference value obtained is small. Therefore, we abandon the grouping given in the original problem and select the optimal number of grouping clusters through K-means for grouping. However, the optimal number of grouping clusters is 2, which means that the 267 pregnant women have high similarity in BMI. Combined with relevant literature and Figure 2 (BMI changes of 267 pregnant women with gestational age at detection), the BMI of pregnant women will fluctuate and increase with the increase of gestational weeks. To solve the truly robust and reasonable NIPT time point, ensure the safety of pregnant women and improve the accuracy of decision-making in coping with complex clinical trade-offs, in the preliminary modeling exploration, we divide the higher BMI segment into Group 0 and the lower BMI segment into Group 4 based on mutation key point cutting, and then divide the middle dense area into 3 groups, so as to minimize the risk while obtaining the optimal NIPT time point. BMI changes of 267 pregnant women are shown in figure 2.

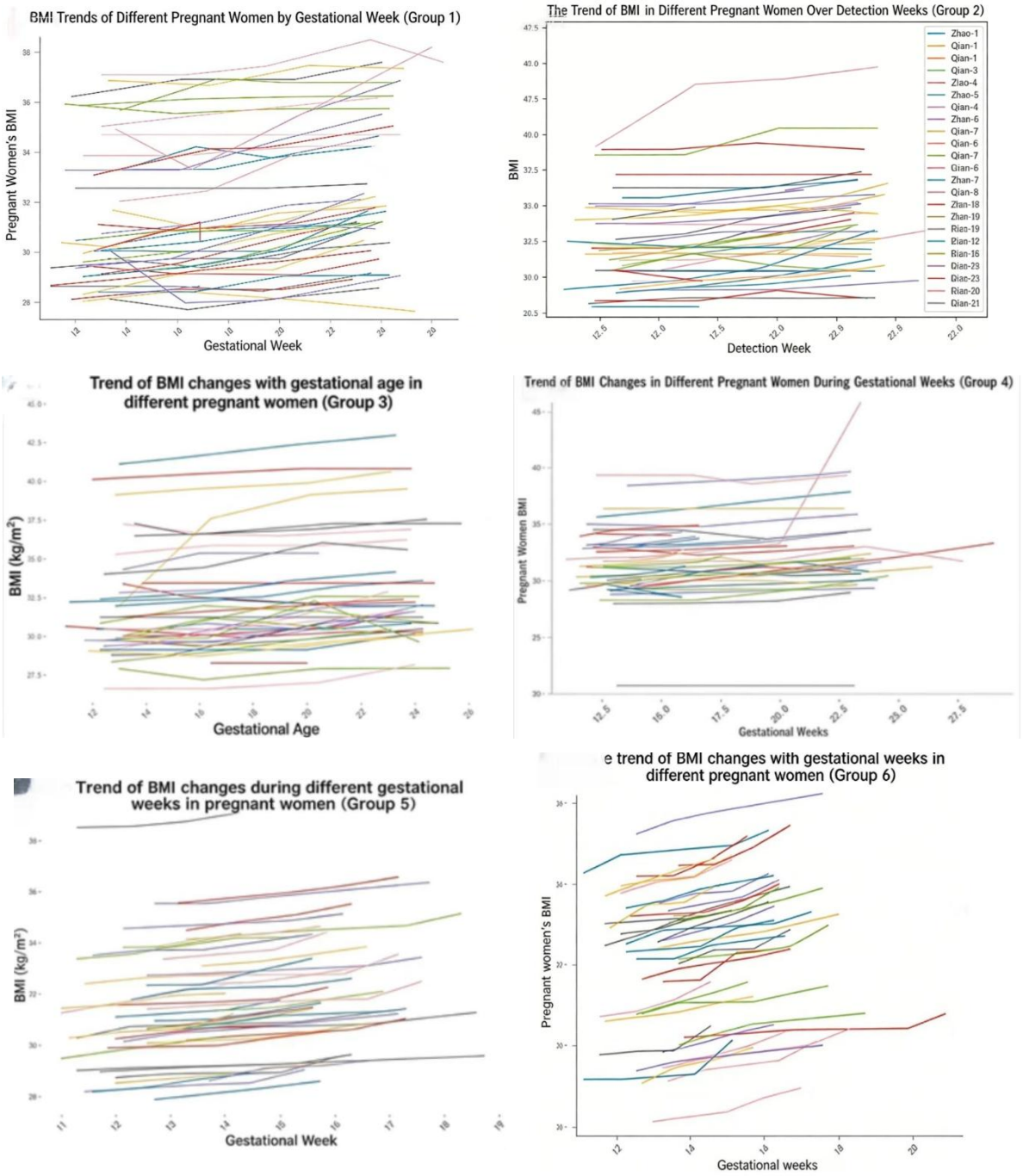

Figure 2. BMI changes of 267 pregnant women with gestational age at detection (Trends of BMI changes of different pregnant women with gestational age at detection (Group 1) to (Group 6))

2.3. Re-grouping of Sample Data

When analyzing the data of pregnant women, we find that to increase the reliability of detection results, some pregnant women have multiple blood collections and multiple detections or one blood collection and multiple detections. It is difficult for us to compare the qualified rate of Y chromosome data of male fetuses among the entire pregnant woman group with different BMIs. The K-means algorithm can help us discover grouping patterns in the data. For example, clustering based on the pregnant woman's BMI can divide pregnant women into five groups according to special points, which helps to consider each pregnant woman as much as possible, thereby reducing the potential risk of each pregnant woman in non-invasive prenatal testing and determining the optimal NIPT time point.

2.3.1. K-means Sample Grouping and BMI Clusters

Based on the core background of the pregnant woman's BMI, we split the samples into $K = 5$ mutually exclusive clusters through the K-means method according to the base of the pregnant woman's BMI. The range of each group is as table 2:

Table 2. K-means Clustering Results

Group	BMI Range
Group 0	38.2-46.9
Group 1	35.1-38.2
Group 2	32.0-35.1
Group 3	29.7-32.0
Group 4	20.7-29.7

From the table, it can be seen that we first divide the higher or lower BMI into two groups, which are Group 0 (38.2-46.9) and Group 4 (20.7-29.7) respectively. The other three clusters cover the middle BMI (29.7-37.9), which is the range with the most dense sample size. Through the K-means method, pregnant women are effectively grouped by BMI, providing a basis for solving the NIPT time point later.

2.3.2. Correlation Analysis of Sample Grouping

(1) Kaplan-Meier Method

Also known as the product-limit estimation method, it is a non-parametric statistical method that measures the time elapsed from a certain starting event (such as disease diagnosis, product start-up, etc.) to an endpoint event (such as death, product failure, etc.).

(2) Exploring Differences in Qualified Gestational Age among Different BMI Groups

As known from the problem, the BMI of pregnant women with male fetuses is the main factor affecting the earliest qualified time of fetal Y chromosome concentration (i.e., the earliest time when the concentration reaches or exceeds 4%). After classification by BMI, this method can be used to explore the situation of pregnant women with different BMI categories in these time-related events. The survival curve drawn by the Kaplan-Meier method can intuitively show the qualified situation of different groups at each time point, and then analyze the optimal NIPT time point of pregnant women with different BMI categories.

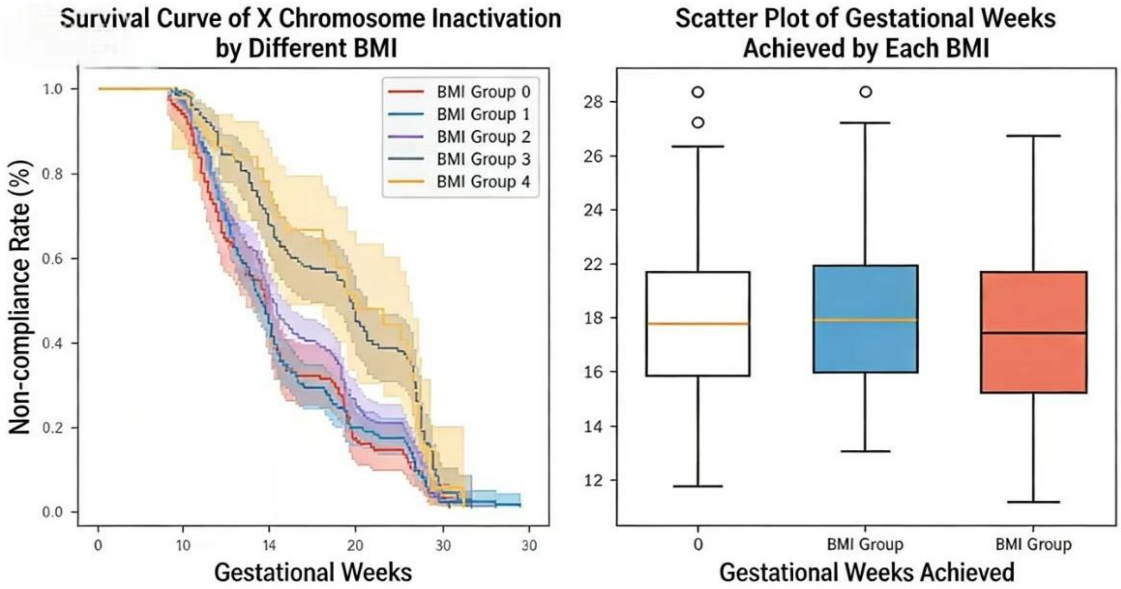

Figure 3. Survival curves of Y chromosome qualification for different BMI groups and Boxplots of qualified gestational age for each BMI group

It can be seen from Figure 3 that the probability curve of unqualified rate in the high BMI group (Group 0) decreases slowly, indicating that the time for the Y chromosome concentration of male fetuses to reach 4% is relatively late, while that in the low BMI group (Group 4) decreases rapidly and can reach 4% quickly. It can be inferred that the high BMI group needs to wait for a longer gestational age at detection, and BMI is negatively correlated with Y chromosome concentration. Intuitively from the boxplot, the higher the BMI of pregnant women, the later the gestational age required for their Y chromosome concentration to reach the 4% standard threshold.

(3) Qualified Rate of Different BMI Groups

Since high BMI has a great impact on the gestational age at detection, to reduce the potential risk of pregnant women as much as possible, we have made a statistics on the Y chromosome concentration that pregnant women in each group with different BMIs can reach. The statistics prove that the qualified rate of the high BMI group is the lowest, laying the foundation for the risk assessment below.

2.4. Model for Reducing Potential Risks and Finding the Optimal NIPT Time Point

To determine the optimal NIPT time point for each group, we first need to define the qualified time, analyze the earliest qualified time of each divided BMI group, clarify the risk objective function and detection accuracy function in the detection data, give the expression of the objective function, and solve the model combined with multiple constraints.

2.4.1. Parameter Determination

T_i: The gestational age at detection when the Y chromosome concentration first reaches ≥ 0.04 in all detection records of the 267 pregnant women. If the concentration of a pregnant woman does not meet the standard in all detection records, it is recorded as an 'unobserved event', the so-called right censoring. Parameter Description is shown in table 3.

Table 3. Parameter Description

Parameter	Description	Parameter	Description
T_i	Earliest qualified time of the i-th pregnant woman	R_{total}	Total risk objective function
w_r	Risk assessment weight	W_i	Gestational age at detection
P_c	Qualified rate of each BMI group	d_i	Error value
$P_m(t)$	Qualified proportion at gestational age t	γ	Risk function weight
α	Error expression weight	t^*	Optimal time point for 90% confidence interval

2.4.2. Establishment of Risk Function

As known from the problem, we need to consider two mutually restrictive core risks: one is the 'sequencing failure risk' that the fetal Y chromosome concentration does not meet the clinically testable standard due to too early detection time point; the other is the 'risk of missing the optimal time point' that the treatment window period (the time required from infection with pathogens to the production of antibodies) is shortened when non-invasive prenatal testing is performed after ensuring that the Y chromosome concentration meets the standard.

For the risk of missing the optimal time point, according to the prompt given in the problem, we use R_i to evaluate the risk weight of the sample, and the gestational age at detection is set as W_i, then R_i is expressed as:

$$R_i = \begin{cases} 1, & (W_i \leq 12) \\ 2, & (12 < R_i < 28) \\ 3, & (W_i \geq 28) \end{cases} \tag{1}$$

For the risk caused by sequencing failure, we define the Cox proportional hazards model:

$$h(t|BMI) = h_0(t)\exp(\beta \cdot BMI) \tag{2}$$

Where $h(t)$ represents the risk that the Y chromosome concentration reaches 0.04 at the gestational age at detection t, and the magnitude and sign of β determine the impact of BMI on the selection of gestational age at detection. As known from Problem 1 and relevant references, there is a negative correlation between BMI and DNA concentration.

The risk function can be expressed as the sum of the quotient of the risk weight corresponding to a certain time point and the qualified probability of the BMI group and $h(t)$ in the established Cox proportional hazards model:

$$R_t = \frac{R_i}{P_c} + h(t) \tag{3}$$

Where R_t represents the total risk of the model, and P_c refers to the Y chromosome concentration qualified probability under different BMI groupings ($c \in [0, 1, 2, 3, 4]$). The smaller the qualified probability, the greater the risk of NIPT for pregnant women in the BMI group.

2.4.3. Establishment of Error Expression

In addition, the problem states that the normal GC content should be between 0.04-0.06. Excessively high or low GC content has potential risks that may cause sequencing quality problems and errors.

Define sequencing error:

$$d_i = f(x) = \begin{cases} 0.04 - d_i, & d_i < 0.04 \\ d_i - 0.06, & d_i \geq 0.06 \\ 0, & 0.04 \leq d_i < 0.06 \end{cases} \quad (4)$$

By comparing the abnormalities caused by non-chromosomal aneuploidy in the problem with whether the fetus is normal after birth, we define the evaluation criterion column as A_i, which is accurately recorded as 1 (i.e., the result after the fetus is born is consistent with the result of detecting chromosomal aneuploidy), and inaccurately recorded as 0. From this, the accuracy rate expression is established:

$$L_i = \frac{1}{N}\sum_{i=1}^{N} A_i \quad (5)$$

Finally, we obtain the error expression:

$$E_t = d_i - L_i \quad (6)$$

Where N is the total number of retained sample entries, and d_i indicates the possible error of the test quality. The larger the error value, the lower the accuracy of the NIPT and the more likely there are sequencing quality problems.

2.4.4. Establishment and Solution of Objective Function

(1) Define Decision Variables

Assume that the overall pregnant women are divided into K mutually independent groups according to BMI, and this division is defined by K-1 BMI cut-off points d_k to form a decision vector D. At the same time, define an optimal detection time point $T_k(k \leq K)$ for each group to form a decision vector T. That is, our final solution is to find the optimal combination of cut-off points D^* and the optimal combination of time points T^* to minimize the risk function R_t.

To determine the optimal NIPT time point, we need to ensure that the Y chromosome concentration meets the standard on the premise of detecting as early as possible. To better balance the benefits between the two, suppose that in a certain BMI group, the proportion of qualified cases at gestational age t is $P_m(T \leq t)$, we can select the minimum gestational age at which the probability reaches a 90% confidence level as the optimal time point for the group.

$$t^* = \min\{t: P_m(T \leq t) \geq 0.9\} \quad (7)$$

Where T is the individual's earliest qualified time.

(2) Clarify Constraints

In this problem, we analyze the possible mutually exclusive risk quantities, define the total risk function, and quantify the qualified time of different BMIs. Secondly, considering the quality of NIPT, we observe that there are cases in the sample where the postnatal condition of the fetus is inconsistent with the chromosomal aneuploidy in the attachment. Combined with the range of GC content divided by the problem that can ensure the sequencing quality, we further establish the error expression, considering the potential risks and the test itself, to minimize the risk and error and achieve the purpose of solving the optimal NIPT time point.

(3) Establish Objective Function

$$\max\{P_m(t) - \alpha E_t - \gamma R_t\} \quad (8)$$

Where α represents the error weight, and γ is the risk weight (balancing the two).

Table 4. Optimal Detection Time Point and Related Indicators

BMI Cluster	Optimal NIPT Time Point (Weeks)	Qualified Rate at This Time Point	Detection Error Value	Potential Risk	Objective Function Value
0	22.5	0.911	0.032	2	0.505
1	19.0	0.914	0.021	2	0.510
2	20.5	0.903	0.025	2	0.498
3	20.0	0.908	0.023	2	0.503
4	18.5	0.933	0.018	2	0.529

From the table 4, it can be seen that with the increase of the pregnant woman's BMI, the optimal NIPT detection time point of pregnant women with male fetuses shows an obvious delay trend. From the comparison between the low BMI group 4 and the high BMI group 0, the optimal detection time points of the two differ by as much as 4 weeks. This indicates that pregnant women with higher BMI need to be detected later to meet the Y chromosome concentration standard.

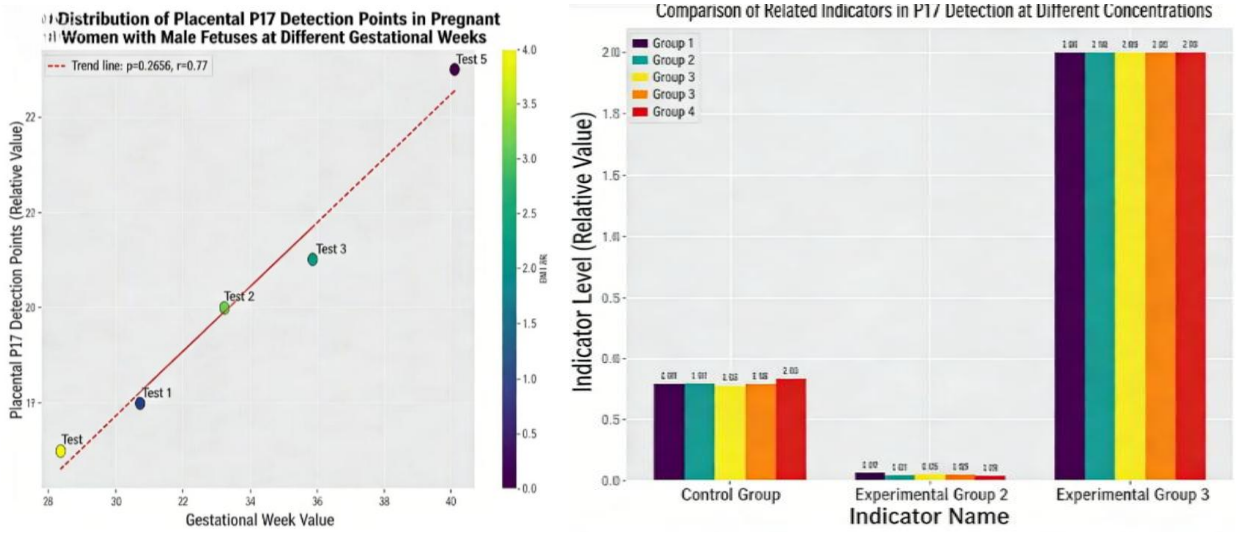

Figure 4. Distribution of optimal detection time points for different BMI clusters and Comprehensive error indicators of different BMI clusters

It can be seen from the figure 4 that the optimal detection time point shows an increasing trend with the increase of BMI, which means that the high BMI group often needs to take measures to delay the NIPT time point. At the same time, the higher BMI group often bears higher potential risks and sequencing quality errors, which indicates that the higher the BMI of pregnant women, the greater the potential risks. Combined with the literature, we speculate that the excessively high BMI of pregnant women may affect the concentration of fetal cell-free DNA, resulting in the slowest qualification speed and large detection errors. At the same time, we find that there is a certain proportional synchronous relationship between potential risks and detection errors, which indicates that on the contrary, pregnant women with lower BMI have higher stability and lower risks.

2.4.5. Sensitivity Analysis Using Monte Carlo Simulation Method

The Monte Carlo simulation method is adopted, which means that small probability events will eventually become large probability events through multiple attempts. To analyze the impact of errors on the optimal detection time point in different BMI groupings, we design a Monte Carlo experiment to analyze the impact of differences in BMI cluster groupings and changes in gestational age thresholds within BMI groupings on the optimal NIPT time point (The sensitivity code has been optimized to a certain extent using AI).

Use random numbers to randomly obtain feasible solutions of BMI groupings, record the current solution and the current objective function value, then continue to take a random feasible solution, and continuously select the optimal value to update to obtain the following images:

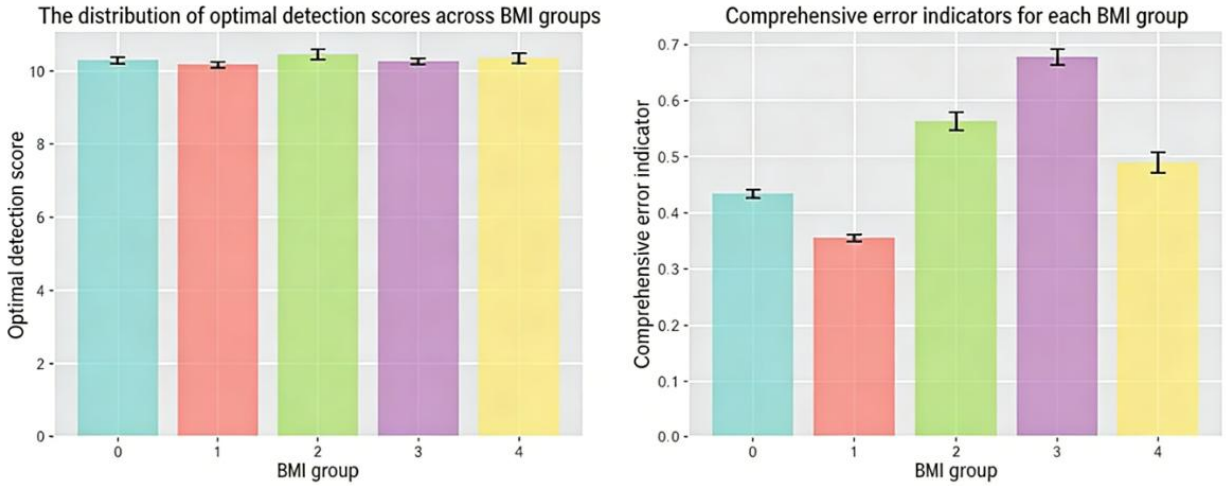

Figure 5. Monte Carlo analysis results (Distribution of optimal detection gestational age for each BMI cluster; Comprehensive error indicators of each BMI cluster)

It can be seen from the figure 5 that there are significant differences in error indicators for different BMI groupings. However, this phenomenon is in line with general expectations. The goal of this model is to ensure the detection success accuracy and detection safety of most pregnant women. From the perspective of improving accuracy, there are cases where pregnant women have multiple detections with one blood collection and multiple detections. In such a nested data model, our goal is not an unoptimized decision, but to provide a standardized solution based on all data.

3. CONCLUSIONS

This study systematically addresses the challenge of determining optimal NIPT testing timepoints for pregnant women of varying body types by integrating clustering algorithms with nonlinear optimization theory. Findings indicate a significant negative correlation between BMI and the rate of fetal cell-free DNA concentration attainment. Establishing a risk-error co-optimization model effectively mitigates sequencing failures caused by premature testing and clinical risks arising from delayed testing. The model successfully quantifies corresponding “safety windows” for five BMI groups, achieving an optimal balance between testing accuracy and clinical timeliness. Limitations include the model's risk classification relying on certain anthropological assumptions, and the error parameter settings in Monte Carlo simulations still depending on specific historical datasets, potentially limiting generalization across sequencing platforms in different medical institutions. Future research will explore integrating real-time fetal DNA concentration monitoring data with more sophisticated deep learning sequence prediction models. This approach aims to develop a technical pathway for dynamically tracking individual DNA accumulation curves in each pregnant woman, thereby advancing from “population stratification” to “real-time, precise individual prediction.”.

REFERENCES

[1] Shu X. A Fine-Tuned Study on Optimal Timing for Non-Invasive Prenatal Testing Based on Generalized Additive Mixed Models and Simulated Annealing Optimization [J]. World Journal of Information Technology, 2025, 3(6).

[2] Lin K, Xu J, Su H, et al. A multiplex digital PCR-based all-in-one non-invasive prenatal test for simultaneous detection of Fetal chromosomal aneuploidies and Fetal fraction quantification [J]. Microchemical Journal, 2026, 220116569-116569.

[3] Lillie M H, Scherr L C, Ratcliff L C, et al. Expanding the Theory of Persuasive Hope: Identifying Mechanisms of Hope's Effect on Non-Invasive Prenatal Testing Intentions [J]. Health communication, 2025, 1-11.

[4] Evans N W, Acherman J R, Kip T K, et al. Embedding Fetal Cardiologists in Maternal-Fetal-Medicine Clinics and Near Universal State-Wide Detection of Critical Cardiovascular Malformations: The Nevada Experience [J]. Journal of ultrasound in medicine: official journal of the American Institute of Ultrasound in Medicine, 2025.

[5] Zemanick T E, Putra M, Elfman H, et al. Response to Wynn et al. re: "False reassurance following single gene non-invasive prenatal testing for cystic fibrosis" [J]. Journal of cystic fibrosis: official journal of the European Cystic Fibrosis Society, 2025.

[6] Wynn J, Haddad A, Hoskovec J, et al. Response to Zemanick et al., "false reassurance following single gene non-invasive prenatal testing for cystic fibrosis" [J]. Journal of cystic fibrosis: official journal of the European Cystic Fibrosis Society, 2025.

[7] Freund Ś M, Tworkiewicz M, Kosiński A, et al. Sources and Level of Patient Knowledge Regarding Available Prenatal Diagnostic Methods and the Frequency of Their Use in the Polish Population [J]. Healthcare, 2025, 13(23):3168-3168.

[8] Hu J, Zhang C, Zeng L, et al. Clinical utility and limitations of non-invasive prenatal testing for chromosome 20 abnormalities: A retrospective study of 83 high-risk pregnancies [J]. International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics, 2025.

[9] Li T. Regression Analysis-Based Investigation of Factors Influencing Male Fetal Y Chromosome Concentration and Stratified Optimization of Optimal Timing for Non-Invasive Prenatal Testing [J]. Journal of Pharmaceutical and Medical Research, 2025, 7(3).

[10] Chen R. Non-Invasive Prenatal Detection Model for Female Fetal Chromosomal Aneuploidy Based on Xgboost [J]. Journal of Pharmaceutical and Medical Research, 2025, 7(3).