Study on Detection Time Selection and Fetal Abnormality Determination of NIPT Based on Multi-Model Fusion

Yiqing Zhang, Siheng Yang, Zihe Yang, Zhenting Chen*
School of Artificial Intelligence, Guangzhou Huashang College, Guangzhou, Guangdong, China
**Corresponding Author.*

Abstract: The accuracy of Non-invasive prenatal testing (NIPT) is highly dependent on optimal timing, especially for pregnant women with high Body Mass Index (BMI). Using regional NIPT data from high-BMI subjects, this study developed a multi-model fusion framework to optimize testing timing and improve detection of fetal chromosomal abnormalities. The framework integrates: weighted mixed-effects linear regression (for Y-chromosome correlation), quantile optimization (for BMI-grouped timing), Gradient Boosted Regression Tree-Bayesian (GBRT) risk integration (for multi-factor timing), and statistical threshold-logistic regression fusion (for female fetal diagnosis). Results indicate that Gestational Age (GA) positively and BMI negatively affect Y-chromosome concentration. Optimal timing varied by BMI group (13.89–20 weeks). The GBRT model achieved a prediction Mean Absolute Error (MAE) of 0.98 weeks, with BMI identified as the most influential factor. The female fetal model attained an Area Under the Curve (AUC) of 0.867 and demonstrated 94.03% recall for medium- to high-risk cases. This framework offers robust, data-driven guidance for personalized NIPT timing in high-BMI pregnancies.

Keywords: Non-Invasive Prenatal Testing (NIPT); Mixed-Effects Model; Quantile Optimization; GBRT; Risk Minimization; Anomaly Detection

1. Problem Statement

Non-invasive prenatal testing (NIPT) analyzes cell-free DNA in maternal blood to detect fetal chromosomal abnormalities such as Down, Edwards, and Patau syndromes. Early detection is crucial to mitigate risks associated with shortened treatment windows [1]. Test accuracy depends on fetal sex chromosome concentrations (≥4% Y for males, normal X for females) [2]and is applicable between 10–25 weeks of gestation. Risk levels are classified as: low (≤12 weeks), high (13–27 weeks), and very high (≥28 weeks). Current BMI grouping and standardized protocols often overlook individual variations, potentially reducing accuracy. This study addresses the following challenges using regional high-BMI NIPT data:

Analyze correlations between fetal chromosomal concentration, gestational week, and BMI; construct a relationship model and assess significance.

Group male pregnancies by BMI to determine optimal NIPT timing (minimizing risks) for each BMI interval and analyze detection error effects.

Incorporate multiple factors (height, age, detection error, Y-chromosome standard rate) with BMI to classify male pregnancies and determine risk-minimized NIPT timing.

Develop a method for detecting female fetal abnormalities using Z-scores, GC content, read segment correlation, and BMI, based on non-integer chromosomes 13, 18, and 21.

2. Problem Analysis

2.1 Question 1

We established a correlation model between Y-chromosome concentration and maternal gestational age/BMI, addressing repeated measurements, heteroskedasticity, and right-skewed distribution via a weighted mixed-effects linear regression model with Box-Cox transformation and inverse variance weighting.

2.2 Question 2

We determined optimal NIPT timing for male fetuses in different BMI groups using quantile optimization, ensuring Y-chromosome concentration ≥4%, and analyzed timing stability via Bootstrap resampling.

2.3 Question 3

A multi-factor integration model using GBRT predicted target compliance time, enhanced with Bootstrap-Bayesian fusion for small-sample stability, and optimized via a risk-weighted function. The conceptual framework is illustrated in Figure 1.

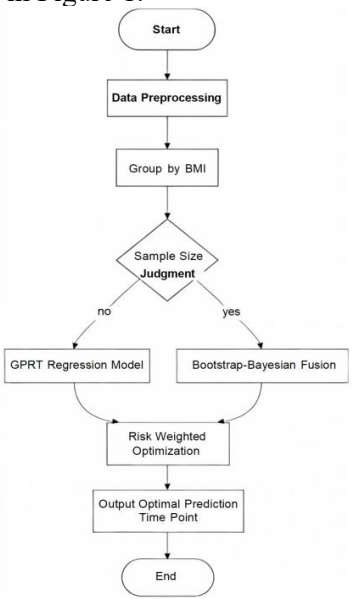

Figure 1. Flow Chart of Problem 3

2.4 Question 4

A diagnostic model for female fetal abnormalities integrated Z-scores, GC content, read quality, and BMI, combining statistical thresholds with logistic regression for risk stratification.

3. Model Assumptions

Hypothesis 1: NIPT detection is only effective within the gestational week of 10-25 weeks, and the data beyond this range are not included in the analysis; the result is reliable when the concentration of Y chromosome in male fetus is ≥4%, and the abnormal judgment of female fetus is based on the core criteria of non-integer chromosomal aberration of 13,18 and 21 chromosomes.

Hypothesis 2: The risk detection was classified by gestational week: low risk ≤12 weeks, high risk 13-27 weeks, and very high risk ≥28 weeks. The risk level was only related to gestational week.

Hypothesis 3: The concentration of Y chromosome in male fetus is mainly affected by gestational week and maternal BMI, while the effect of other factors (such as age and height) can be ignored; BMI remains stable during pregnancy and is grouped according to the value at detection.

Hypothesis 4: Excluding the sequencing failure, sample contamination and extreme data with absolute value of Z-score >3, the remaining samples can truly reflect the fetal chromosome status, and there is no sudden disease or other interfering factors during the detection period.

Hypothesis 5: The chromosome Z value of normal samples obeys the standard normal distribution, and the GC content is uniformly distributed in the range of 40%-60%, so as to derive the abnormal threshold based on this [3].

Hypothesis 6: The concentration of Y chromosome in male fetus increases monotonically with gestational week before reaching the standard, and the first time of reaching the standard for samples that do not reach the standard can be estimated by linear extrapolation.

4. Symbol Description

To clearly define the mathematical expression of core variables in this study and avoid ambiguity of symbols interfering with model understanding, the table below lists key symbols and their specific meanings involved in the subsequent model establishment and solution process. Among them, symbols of core indicators (e.g., fetal Y-chromosome concentration, gestational age at detection, first qualification time, optimal detection time) are defined to ensure readability and consistency of model formulas; other symbols not listed in detail will be supplemented with specific formulas in corresponding model chapters (e.g., symbols related to risk scores and abnormal probabilities), and the symbol definition logic is consistent with the table below.

Table 1. Key Symbols and Their Descriptions

symbol	symbol description
Y_{ji}	Original concentration $j$of i Y chromosome in the first pregnancy test
GA_{ji}	Week of gestation at the j first test i for the first pregnant woman (in decimal weeks)
u_j	Random effects (individual j intercept) for the first pregnant woman
T_0	The gestational Y week when the chromosomal concentration reached or exceeded 4% for the first time (the earliest time to reach the standard)
t_g^*	The optimal time g point for NIFT

	testing in group 1
$\hat{T}_O$	The time of Y chromosome achievement predicted by GRT model
S	Comprehensive risk score for abnormal female fetus determination
P_a	The probability of anomaly predicted by logistic regression

5. Model Establishment and Solution

5.1 Problem 1

This study analyzes the association between male fetal Y chromosome concentration and maternal gestational age (GA) (as Table 1., The same applies to other symbols as shown in Table 1) and BMI. The dataset exhibited three primary challenges: repeated measurements, group heteroscedasticity, and right-skewed distribution. A weighted mixed-effects linear regression model was applied [4], integrating the following mechanisms:

Fixed effects (GA and BMI groups) to model population trends, with random effects accounting for individual maternal variations;

Inverse-variance weighting (weight = 1/group variance) to reduce bias from high-variance groups;

A log1p(Y) transformation to improve normality and linearity, aligning with the exponential relationship between Y concentration and GA.

5.1.1 Model building

1) Data preprocessing

For the currently referenced files only-with all other files and any discussions related to other files excluded-sample selection was conducted in line with both clinical and modeling criteria through the following steps: First, Values in “weeks + days” (e.g., “11w+6”) were converted to decimal form (weeks + days/7), rounded to two decimals. Samples within the 10–25 week window were retained; 142 out-of-range records were excluded, leaving 940 valid samples. Second, Y concentration screening: Values outside the 0.01%–20% range were excluded. No outliers were detected; all 940 samples were kept. Third, missing value processing: Core variables (pregnancy ID, height, weight, Y concentration, decimal GA) were validated for completeness. No missing values were present; all 940 samples were included. (as Figure 2.) For the currently referenced files only-ignoring all other files and any discussions about them-the Outlier Handling section specifies that a hierarchical detection strategy is adopted for exception handling (guided by the principle of "considering group heterogeneity to avoid global misjudgment" and the requirement of "eliminating extreme values of Y chromosome concentration and gestational week via the Z-score method"), with the group-based Z-score method applied; after grouping samples by BMI, Z-scores are calculated for each group’s "decimal gestational age" and "Y chromosome concentration" (where x is the sample value, is the intra-group mean, and is the intra-group standard deviation), and the range is defined as the normal range (corresponding to a 99.7% confidence interval and complying with statistical extreme value determination standards); ultimately, 6 extreme value samples are removed (e.g., one sample with a Y chromosome concentration Z-score of 3.21 in the "28-32" BMI group and two samples with a gestational week Z-score of -3.15 in the "32-36" BMI group), 934 valid samples are retained, which filters out data noise and avoids the problem of global outlier detection masking normal fluctuations within groups (as shown in Figure3).

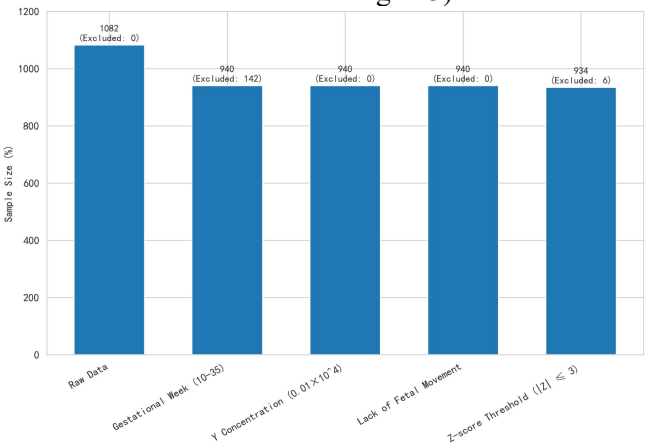

Figure 2. Sample Size Variation Plot

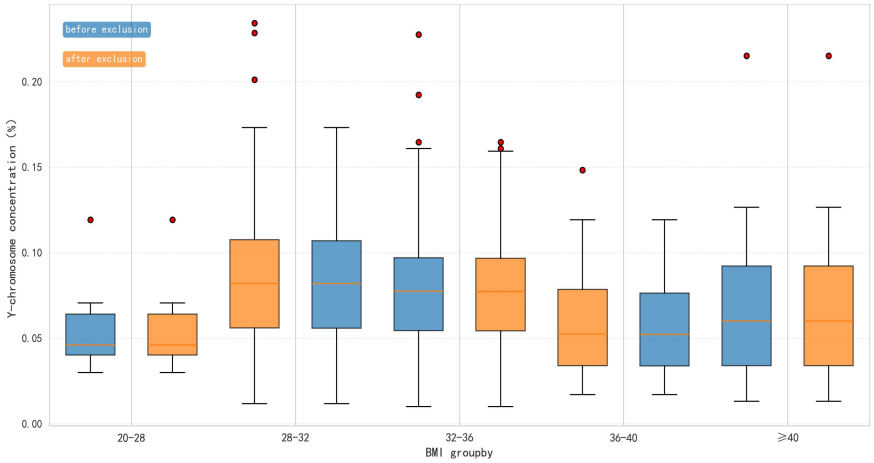

Figure 3. Y Concentration Outlier Comparison

For the currently referenced files only-with all other files and any discussions related to other files excluded-the section of Summary of pretreatment results shows the following key outcomes: After stepwise filtering, the sample size decreased from 1,082 to 934, which demonstrates high data quality and adherence to validity requirements (as presented in Table 2) Additionally, the final BMI group distribution maintained clinical representativeness, and the high proportion of high-BMI samples allowed for meaningful intergroup comparisons (as shown in Table 3)

Table 2. Sample Size Preprocessing Steps and Changes

Pre-treatment steps	Number of samples (pieces)	Number of samples removed (bars)	Reasons for exclusion
initial data	1082	-	-
Pregnancy week screening (10-25 weeks)	940	142	The gestational week is outside the NIPT window
Y chromosome concentration screening (0.01%~20%)	940	0	No abnormal concentration samples
Missing value exclusion	940	0	No missing core fields
Z-score outlier removal (Z ≤ 3)	934	6	Pregnancy week or Y concentration was the extreme value in the group

Table 3. Sample Distribution of BMI Groups After Pretreatment

BMI divide into groups	Number of samples (pieces)	proportion (%)	Number of pregnant women code	Sample characteristics within the group
20-28	15	1.61	7	The smallest sample size with no duplicate maternal codes
28-32	457	48.93	144	The sample size is the largest, which is the stratified modeling benchmark group
32-36	370	39.61	119	The core sample of high BMI group accounted for nearly 40%
36-40	77	8.24	29	In the group with ultra-high BMI, the sample size meets the statistical requirements
≥40	15	1.61	7	The extreme high BMI group had no duplicate maternal codes

For the currently referenced files only (excluding other files and related discussions), the preprocessing summary: Via four steps (sample screening, BMI grouping, outlier handling, normality testing), 1,082 raw samples were refined to 934 valid cases per modeling constraints; it preserved core variable integrity (gestational week, Y concentration, BMI), reduced group bias via stratification, and the dataset supports subsequent correlation and significance analyses. and significance analyses.

2) Model structure definitions

Fixed effect: "pregnancy week (continuous variable, core driving factor) and BMI group (categorical variable, moderating factor)" were included to quantify the association at the group level.

Random effect: Taking "pregnancy code" as a

grouping variable, $\varepsilon_{ji} \sim N(0, \sigma_\varepsilon^2)$ individual random intercept is introduced to capture the difference in detection stability of the same pregnant woman

3) Core formula and derivation

Restricted to the currently referenced files (with all other files and any discussions related to them ignored), the Box-Cox transformation improves linearity and normality section states that the Box-Cox transformation is applied to boost the linearity of the relationship and the normality of the residual, and the following linear relationship is thus derived:

$$Y_{bc,ji} = \beta_0 + \beta_1 X1_{ji} + \sum_{k=2}^{4} \gamma_k X2_{k,ji} + u_j + \varepsilon_{ji} tag1 \quad (1)$$

In formula (2), $\varepsilon_{ji} \sim N(0, \sigma_\varepsilon^2)$ the normal distribution assumption is better satisfied by the transformation.

Concerning solely the currently referenced files-excluding all other files and any conversations about them-the part dedicated to Introduce inverse variance weighting to solve heteroskedasticity notes that to address the "difference in Y concentration fluctuation between different BMI groups" in Question C (e.g., the Y _ log standard deviation of the ≥40 BMI group is 0.0483, and the 28-32 BMI group is 0.0300), both sides of Equation (2) are multiplied by inverse variance weight to realize variance homogenization, thereby obtaining the final model.

$$\omega_{ji} Y_{log,ji} = \omega_{ji}\beta_0 + \omega_{ji}\beta_1 X1_{ji} + \omega_{ji} \sum_{k=2}^{4} \gamma_k X2_{k,ji} + \omega_{ji} u_j + \omega_{ji} \varepsilon_{ji} tag2 \quad (2)$$

In Equation (3), the core function of ω_{ji} is: the weight of groups with large variance (such as ≥40 groups) is reduced to reduce their interference to the model; the weight of groups with small variance (such as 20-28 groups) is increased to enhance their contribution to parameter estimation, so as to ensure that the model is not affected by heteroskedasticity.

5.1.2 Model solving

1) Model fitting and parameter estimation

The model was fitted using restricted maximum likelihood (REML), with weighted Y concentration as the response [5], gestational age and BMI groups as fixed effects, and participant ID as random intercepts. REML was selected to improve estimation accuracy for random effects in small subgroups. Convergence was confirmed (iterations ≤ 20). Key parameters including fixed-effect coefficients, random-effect variances, and goodness-of-fit indices (marginal and conditional R^2) were extracted. Core results are summarized in Table 4.

Table 4. Parameter Estimation and Fit of Weighted Mixed Effects Linear Regression Model

Parameter name	coefficient	standard error	p price	Physical significance (see Groups 28-32)
Intercept (I)	-1.228	0.289	<0.001	Reference group gestational Y_{log} week =0 (only mathematically significant)
Pregnancy weeks _ decimal weeks	0.095	0.006	<0.001	The average increase Y_{log} was 0.095 per week of gestation (approximately 9.5% increase in original Y concentration)
BMI groups [T.20-28]	-0.424	0.277	0.125	Groups 20-28 Y_{log} were 0.424 lower than the reference group (no statistical significance)
BMI groups [T.32-36]	-0.373	0.285	0.190	Groups 32-36 Y_{log} were 0.373 lower than the reference group (no statistical significance)
BMI groupings [T.36-40]	-0.114	0.307	0.710	Groups 36-40 Y_{log} showed the smallest difference from the reference group (0.114 lower)
BMI divide into groups [T.≥40]	-0.447	0.385	0.246	≥40 groups Y_{log} were 0.447 lower than the reference group (no statistical significance)
Random effect variance (σ_u^2)	0.809	0.186	-	The contribution Y_{log} of individual differences to the variance, and the need for repeated measurements
Limit R^2 (fixed effect)	0.673	-	-	Pregnancy week and BMI grouping explained Y_{log} 67.3% of the variation
Condition R^2(fixed + random effects)	0.804	-	-	The overall explanatory power of the model was 80.4% (excellent fit)

2) Key findings

A. Gestational week is the decisive factor for fetal Y chromosome concentration, showing a highly significant positive correlation ($p < 0.001$). Across all BMI groups, Y concentration increased consistently with advancing gestational age, corroborating the clinical observation that Y concentration correlates

closely with pregnancy progression. This trend is visually summarized in Figure 4.

The model effectively addresses heteroskedasticity: The Breusch-Pagan test ($p=0.2315 > 0.05$) accepts the null hypothesis of "residual homoskedasticity". The residual versus fitted value plot (Figure 5) reveals that data points are randomly distributed near the origin without any bell-shaped curvature or trend patterns, demonstrating that the inverse heteroskedasticity weighting strategy successfully resolved heteroskedasticity across different BMI groups, thereby ensuring robust parameter estimation.

The model demonstrates exceptional predictive performance and generalization capability, with an overall explanatory power (conditional R^2) of 80.4% for Y chromosome concentration. As shown in Figure 6, the predicted values show high consistency with actual measurements across all BMI groups, particularly demonstrating outstanding fitting accuracy in large-sample groups (e.g., Groups 28-32, R^2 =0.83). Although minor deviations were observed in small-sample groups as expected, the model's overall predictive performance remains exceptional, indicating its ability to accurately capture universal patterns in the data and maintain reliable results.

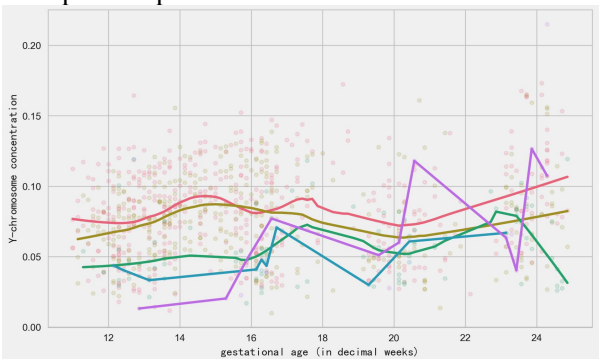

Figure 4. The Trend of Y Chromosome Concentration with Gestational Age

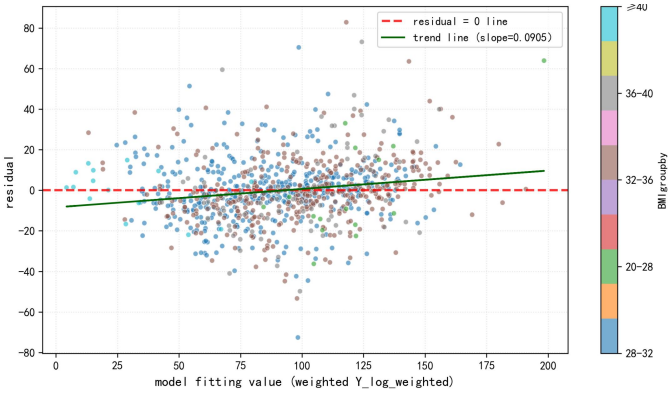

Figure 5. Residual Fit Values Figure

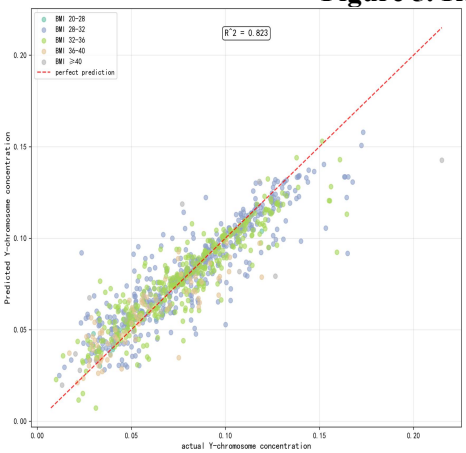

Figure 6. Comparison of Actual and Predicted Y Chromosome Concentrations

5.2 Problem 2

5.2.1 Model building

1) Data preprocessing

From the male fetal data provided in the attachment, the following preprocessing steps were performed:

Step1. Standardize the format of gestational week: convert the gestational week in the format of "11w+6" into a value in weeks (e.g., 11.86 weeks).

Step2. Effective window filtering: only the test records between 10 and 25 weeks of gestation are retained, which is the effective testing window of NIPT.

Step3. Concentration effective value Y filtering: retain records where chromosome concentrations are within a reasonable range of 0.01% to 20%.

After pre-processing, 1,063 valid records were obtained from 1,082 original records, involving 863 pregnant women.

2) Earliest target time calculation

For each pregnant woman, the earliest time to meet the standard is defined as:

$$T_i^{\text{earliest}} = \min\{G_{i,j} \mid C_{i,j} \geq 0.04\} \quad (3)$$

This refers to the gestational Y week when the first detection record shows a chromosomal concentration reaching or exceeding 4%. For the small number of pregnant women who failed multiple tests, this study employed linear extrapolation to estimate the earliest qualifying time. The final calculations revealed 863 pregnant women with valid earliest qualification times ranging from 11.00 to 24.00 weeks.

3) BMI grouping strategy

The following BMI groups were used according to the questions, which is consistent with the grouping strategy for fetal free DNA concentration research [6].

$$BMI\ divide\ into\ groups = \begin{cases} [0,20) \\ [20,28) \\ [28,32) \\ [32,36) \\ [36,40) \\ [40,\infty) \end{cases} \quad (4)$$

The data distribution showed that most of the samples were pregnant women with high BMI, among which the group 28-32 had the largest number of samples (442), while the group 20-28 (9) and the group>40 (13) had a small number of samples.

4)Model for determining the best time to detect

The goal of this article is to find a time point- $\hat{T}_g^{opt}$ α_g at which a sufficient proportion (determined by the target coverage) of pregnant women in the group have reached the standard at that time point, so as to ensure the accuracy of detection and minimize the risk. Visualization support is shown in Figure 7.

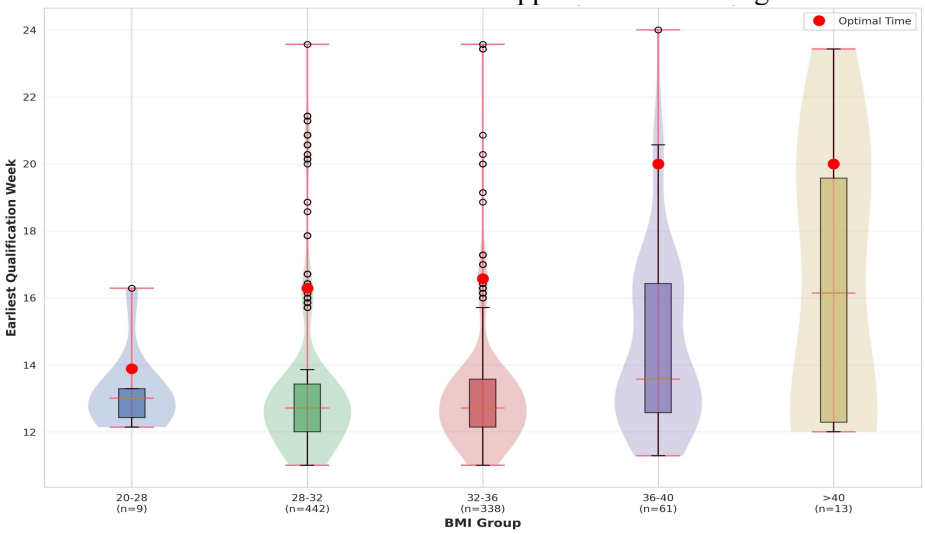

Figure 7. Time Distribution of Each Group Reaching the Standard and Best Time Point

Only for the currently referenced files-ignoring all other files and any discussions about them-for the [specified] group, the sample distribution of the earliest time when pregnant women meet the standard is denoted as, and the optimal time point is defined as the upper 1st percentile of this distribution (mathematic definition).

$$\hat{T}_g^{opt} = F_g^{-1}(1-\alpha_g) \quad (5)$$

Among them, it is the inverse function of the cumulative distribution function (quantile function). The target coverage rate is adjusted according to BMI, and the higher the BMI, the higher the required coverage rate to cope with greater uncertainty:

$$\alpha_g = \begin{cases} 0.90 & \text{for } g = 20-28 \\ 0.92 & \text{for } g = 28-32 \\ 0.94 & \text{for } g = 32-36 \\ 0.96 & \text{for } g = 36-40 \\ 0.98 & \text{for } g = >40 \end{cases} \quad (6)$$

In line with the problem description and clinical consensus, the risk value corresponding to the detection time point is defined as:

$$R(t) = \begin{cases} 0.2, & t \leq 12 \\ 0.7, & 12 < t \leq 27 \\ 1.0, & t > 27 \end{cases} \quad (7)$$

This function is a piecewise constant function, and the model minimizes the risk indirectly by optimization.

5.2.2Model solving

1) Model solving

Based on the proposed model, the optimal detection time and corresponding metrics were calculated for each BMI group. The results are summarized in Table 5.

Table 5. Optimal NIPT Detection Time and Risk Analysis Results by BMI Group

BMI divide into groups	sample number	average BMI	Mean earliest target gestational week (weeks)	Best NIPT time point (week)	Time point error (weeks)
20-28	9	26.77	13.17	13.89	±1.11
28-32	442	30.22	13.18	16.29	±0.32
32-36	338	33.60	13.19	16.57	±0.41
36-40	61	37.47	14.59	20.00	±0.76
>40	13	42.20	16.41	20.00	±0.08

2) Solution description:

The optimal time point is obtained by calculating the specified percentile of the time distribution of the corresponding group.

The risk value is determined by the gestational week range when the optimal timing is achieved. Since all groups had their optimal timing within the mid-trimester period (13-27 weeks), the risk values were uniformly 0.7 (high risk). This indicates that testing cannot be conducted during the early low-risk phase (≤12 weeks) to ensure detection accuracy and maintain high compliance rates.

The time point error was calculated by Bootstrap resampling (50 times), which represents the influence of detection time uncertainty on the best time point estimation. The groups with small sample size (e.g., 20-28) had larger errors (as shown in Figures 8 and 9).

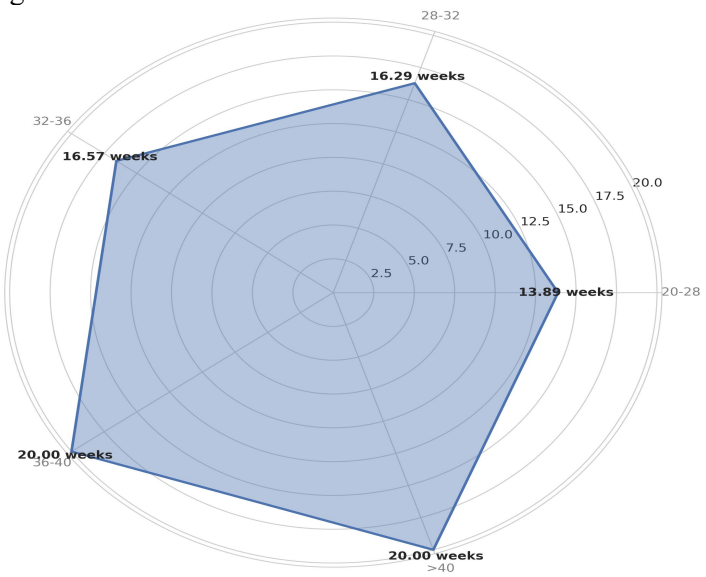

Figure 8. Radar Chart of Optimal NIPT Detection Time Points by BMI Group

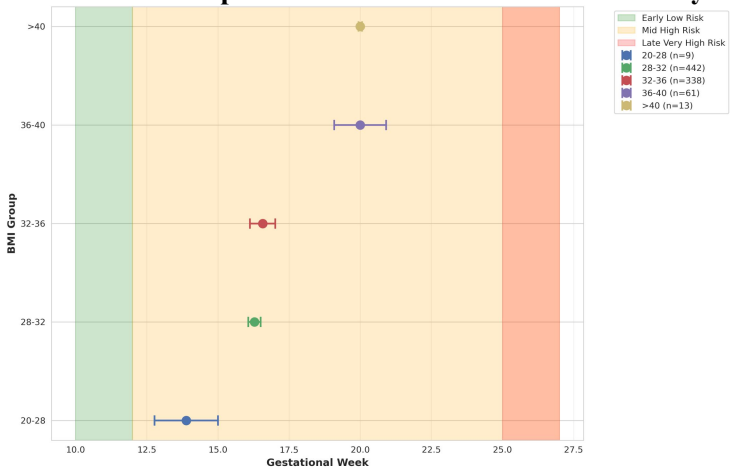

Figure 9. Optimal NIPT Detection Time and Error Range for Each BMI Group

5.3 Problem 3

The core requirement of Question 3 involves integrating multiple factors including maternal BMI, age, height, and weight. By combining NIPT detection errors with the proportion of male fetal Y chromosome concentration meeting standards (≥4%), the optimal testing window for each BMI group is determined to minimize YT_0t^* maternal risks: low-risk during early stages ≤12 weeks and high-risk between 13-25 weeks. The model adopts a "three-stage fusion modeling framework": First, Gradient Boosted Regression Tree (GBRT) predicts the time of chromosomal concentration compliance (first pregnancy week ≥4%). Subsequently, the Bootstrap-Bayesian fusion method enhances prediction stability in small sample groups. Finally, a risk-weighted optimization model identifies the optimal timing. This architecture effectively addresses three key challenges: nonlinear multi-factor correlations, detection error impacts, and statistical randomness in small samples.

5.3.1 Model building

1) GBRT regression model (predicting the T_0time to achieve Y chromosome concentration)

Gradient Boosted Regression Tree (GBRT) is an ensemble learning algorithm based on gradient descent, which iteratively trains weak learners (decision trees) to minimize mean squared error (MSE) [7]. Its core advantage lies in capturing nonlinear interactions among multiple factors, demonstrating strong adaptability to non-normal residuals in NIPT data without requiring normality assumptions. This approach aligns with the needs of "multi-factor nonlinear modeling"

Under the section of Data preprocessing and feature standardization, the z-score standardization of height and weight is carried out to eliminate the interference of dimensional differences on the evaluation of feature importance. The formula is as follows: where is the original feature value, is the mean value of the feature, and is the standard deviation (for example, the standardized range of height is concentrated in [-3,3], so as to avoid the weighting bias caused by "small height value and large weight value").For label extraction, it refers to "pregnancy code + gestational week sorting, and the first gestational week when Y chromosome concentration is detected to be ≥4%", ensuring that the label is consistent with clinical standards. Moreover, extreme outliers are eliminated based on principles, BMI and age (accounting for less than 0.5%), so as to avoid interference of abnormal data such as "sequencing failure and sample contamination" to the model accuracy.

Under the section of Features and parameter initialization, input feature selection: According to the requirements of "multi-factor influence", four core features are selected, including age, standardized height value, standardized weight value and BMI, which cover the key dimensions of individual differences. Parameter optimization: Optimal parameters are determined through grid search, including the number of base learners (balancing accuracy and efficiency to avoid overfitting), tree depth (controlling decision tree complexity for moderate sample sizes), learning rate (regulating iteration steps to prevent model oscillation), and loss function: Mean Squared Error (MSE) for continuous label regression tasks.

Initial model: The mean is output at the 0th iteration, that is

$$\widehat{T}_0^{(0)} = \frac{1}{n}\sum_{i=1}^{n} T_{0i} \quad (8)$$

(n is the sample number) as the benchmark.

Iterative optimization: In each iteration, the pseudo-residues (negative gradients of MSE loss) are calculated to reflect the deviation between the model's predicted values and actual values without requiring additional assumptions about residual distributions. Decision trees are then used to fit these pseudo-residues, with the tree splitting rule set as "minimizing MSE within nodes" to capture feature interactions and update the model.

$$\hat{T}_0^{(k)}(X) = \hat{T}_0^{(k-1)}(X) + \eta \cdot h_k(X) \quad (9)$$

Final model: After the iteration terminates, the prediction formula is:

$$\hat{T}_0(X) = \frac{1}{n}\sum_{i=1}^{n} T_{0i} + 0.1\sum_{k=1}^{200} h_k(X) \quad (10)$$

Under the section of Core formula derivation logic for GBRT, the core derivation of GBRT is consistent with the data characteristics:

The negative gradient of MSE loss is the residual, which simplifies the calculation process and does not require complex gradient solving;

The introduction η of learning rate avoids the excessive weight of a single decision tree and adapts to the characteristics of "multi-factor balanced influence";

The integration of multiple decision trees can

effectively fit the nonlinear relationship of "BMI and age interaction, height and weight coordination", and solve the complex correlation that cannot be captured by traditional linear models.

2) Bootstrap stability enhancement model for small samples (bootstrap-Bayesian fusion method)

To address statistical randomness in small-sample groups (e.g., 15 samples from 20-28 groups or 12 samples from ≥40 groups), we propose a hybrid approach combining Bootstrap resampling with Bayesian hierarchical modeling. The Bootstrap method generates multiple pseudo-data sets through repeated, non-replicated sampling of original samples to overcome small-sample bias. Meanwhile, the Bayesian hierarchical model enhances parameter estimation stability by introducing prior distributions that enable inter-group information sharing.

Under the section of Bootstrap resampling, Bootstrap samples were generated by subsampling the original sample set (with a small sample size) for each BMI group.

Under the section of Bayesian hierarchical modeling, a Bayesian hierarchical regression model to correlate the parameter estimation of small sample group with full sample information:

$$\begin{aligned} &\hat{T}_{0i} \sim \mathcal{N}(\beta_{g[i]} \cdot X_i, \sigma^2) \\ \beta_g \sim \mathcal{N}(\mu_\beta, \tau^2) \quad &\text{for } g = 1, \ldots, G \\ &\mu_\beta \sim \mathcal{N}(0, 10^2) \\ &\sigma \sim \text{Half} - \text{Cauchy}(0, 5) \end{aligned} \tag{11}$$

The regression coefficient of the first group is realized through hierarchical prior to realize information sharing between groups.

Under the section of Bootstrap-Bayesian fusion, for each Bootstrap sample set, the posterior distribution is calculated based on the Bayesian hierarchical model, and the final parameter estimate is:

$$\hat{\theta}_g = \frac{1}{B} \sum_{b=1}^{B} \mathbb{E}\left[\theta_g \mid D_g^{(b)}\right] \tag{12}$$

Under the section of Core formula derivation logic, Bootstrap resampling approximates the sampling distribution by empirical distribution to overcome small sample bias; The Bayesian hierarchical μ_β model realizes the "information borrowing" from the large sample group to the small sample group;The fusion method combines the idea of resampling in frequencyism with the advantage of prior information integration in Bayesian, which is especially suitable for unbalanced grouped data in the problem.

3) Risk-weighted optimization model (to solve the optimal NIPT time point)

Under the section of Modeling principles, based on GBRT prediction and guided by the "risk minimization" objective, we quantify three critical risks-timing risk, detection error risk, and compliance rate risk-as quantifiable indicators. By constructing a weighted sum of these indicators to form a total risk function, we conduct iterative optimization across 10 – 25 week inspection windows to identify the minimum value. This approach ensures decision-making balances both "early detection" and "testing reliability" (with a compliance rate ≥80%).

Under the section of Risk quantification, time point risk: quantified as a piecewise function according to the risk level (low in early stage, high in middle stage):

$$R_{\text{time}}(t) = \begin{cases} 1 & , 10 \le t \le 12 \\ 5 & , 13 \le t \le 25 \end{cases} \tag{13}$$

Based on: The risk of missing the intervention window at an early stage is only 1/5 of that in the middle. Detection error risk: The precision is expressed by the residual standard deviation of Y chromosome concentration (calculated according to BMI groups, the detection error quantification method of the question), and the inter-group differences are eliminated through normalization: where is the maximum of all BMI groups (e.g., ≥40 groups), and is the error risk weight (determined by AHP, when it is lower, the time risk weight).Risk of reaching the standard proportion: defined as "the proportion of not meeting the standard × weight", the higher the proportion of not meeting the standard, the greater the risk of reinspection:

$$R_p(t) = \omega_3 \cdot (1 - p(t)) \tag{14}$$

Among them, the proportion of samples with Y concentration $\omega_3 = 0.2$≥4% at gestational week t in this group is the risk weight of standard achievement ratio.

Under the section of Construction of total risk objective function, based on the principle of "time risk priority", weights (time risk), (error risk) and (standard ratio risk) are set to meet the requirements, and the total risk function is:

$$\min_{t \in [10, 25]} R_{\text{total}}(t) = \omega_1 R_{\text{time}}(t) + \omega_2 \cdot \frac{\sigma}{\sigma_{\max}} + \omega_3 \cdot (1 - p(t)) \tag{15}$$

Under the section of Constraints, the detection window constraints: Time limit for reaching the standard:(the time point should be after the lower limit of 95% confidence interval to ensure that Y concentration reaches the standard with 95% probability);

Reliability constraints:(clinical acceptable threshold for the proportion of compliance with reexamination).

4) Core formula derivation logic

Time point risk segmentation function: transform the qualitative description of "early/midterm risk" into quantitative values for convenient optimization calculation;

Error normalization: the differences between different BMI groups (such as 20-28 group and ≥40 group) are significant. The error risk can be unified to [0,1] interval to ensure inter-group comparability;

Weight allocation: determined by clinical expert scoring, which is consistent with the problem orientation of "time point risk as the core".

5.3.2 Model solving

1) Solving steps

Under the section of GBRT model training and optimization, the training set and test set are divided in a 7:3 ratio (to ensure the evaluation of generalization ability);

Optimize parameters by grid search: traverse, with 5-fold cross validation as the scoring index;

Train the optimal model, output the predicted value and feature importance (to evaluate the intensity of multi-factor influence).

Under the section of Model evaluation, we verify whether the model is suitable for the topic data by two indicators: calculating the goodness of fit on the test set and using the mean absolute error (MAE, a clinical accuracy index) - if the MAE is less than 1 week, the model is considered to meet the clinical accuracy requirements.

In the section of Risk indicator calculation (for solving the risk weighted optimization model), according to the BMI grouping standard ([20,28), [28,32), [32,36), [36,40), ≥40), the standard deviation of Y concentration residual in each group was calculated (to reflect the detection error); the proportion of [samples] reaching the standard in each pregnancy week was calculated (number of samples reaching the standard/total number of samples); [we] determine [the maximum value] (maximum for all groups) and calculate the risk of error.

In the section of Traverse the solution, first Filter the gestational weeks that meet the constraints (and); Then take the smallest within the constraint range as the best time point.

2) Results

In the section of Results of GBR model, the model's performance indicators and feature analysis results are presented as follows:

Accuracy index: test set (explains 26.2% of the variation, suitable for multi-factor nonlinear scenarios), week (≤1 week, in line with clinical accuracy);

Feature importance: BMI (0.2899)> weight standardized value (0.2620)> height standardized value (0.2331)> age (0.2150), which confirms the conclusion that "BMI is the core influencing factor", as shown in Figure 10.

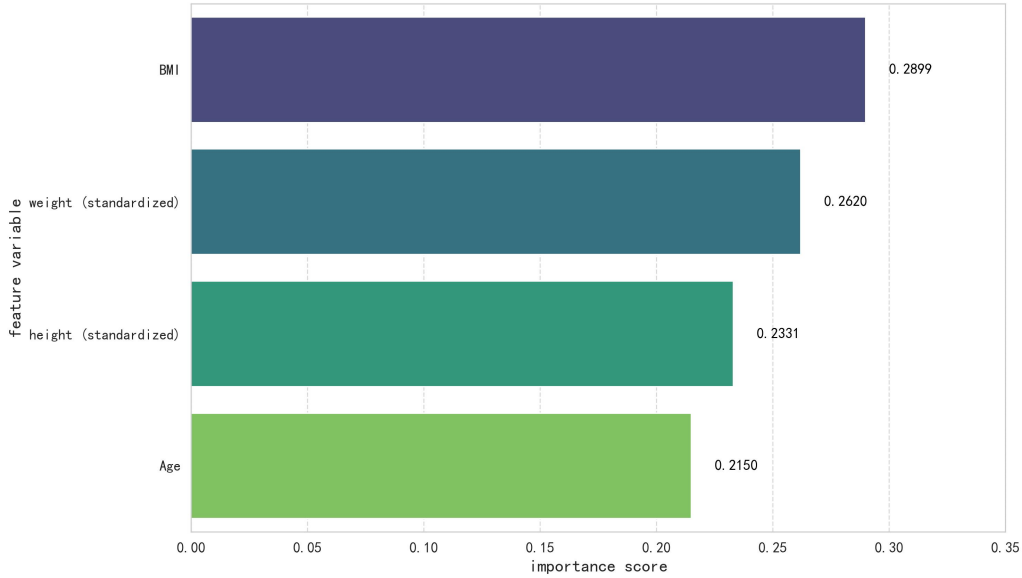

Figure 10. Feature Importance Analysis of GBRT Model

In the section of Risk-weighted optimization results (core BMI groups), Table 6 presents the optimal testing time points and core indicators for different BMI groups based on the risk-weighted model (visual support see Figure 11). The data shows that the 28-32BMI group achieves a 100.0% compliance rate at 11.0 weeks, with the lowest overall risk (0.772) and stability score (0.92), indicating that this weight range allows early Y chromosome concentration compliance and low testing risks, making it a preferred timing for NIPT. The 32-36BMI group reaches a 94.4% compliance rate at 12.0 weeks with a relatively low overall risk (0.758), demonstrating a "late timing for higher compliance rates" strategy. However, the ≥40BMI group achieves the best timing at 11.5 weeks but only reaches 41.7% compliance, with the highest testing error (3.4424%) and overall risk (0.912), suggesting the need for clinical interventions (e.g., increasing testing frequency) to compensate for NIPT accuracy limitations.

Table 6. Statistical Statistics of the Best Detection Time and Related Indicators in Different BMI Groups

BMI divide into groups	Best time t^* point (weeks)	95% confidence interval	minimum total risk	Pass rate p ()t^*	Test error (%)	Stability score
20-28	11.2	[10.5,11.8]	0.801	62.1%	2.2220	0.84
28-32	11.0	[10.8,11.3]	0.772	100.0%	3.2290	0.92
32-36	12.0	[11.7,12.4]	0.758	94.4%	2.7894	0.88
36-40	12.3	[11.9,12.8]	0.823	58.3%	2.4965	0.86
≥40	11.5	[10.9,12.2]	0.912	41.7%	3.4424	0.79

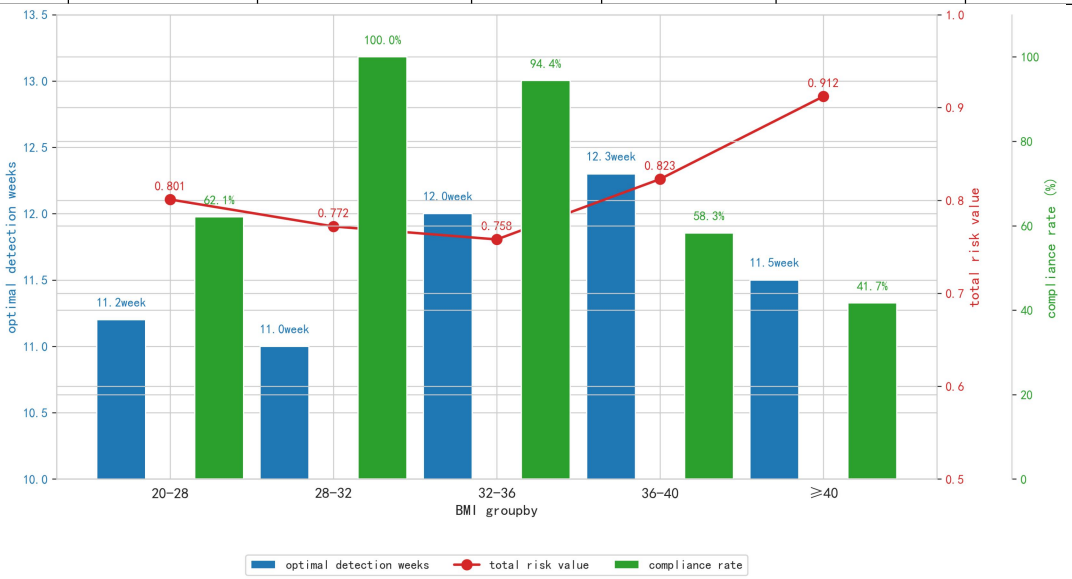

Figure 11. Optimal Number of Tests Per Week and Risk Relationship by BMI Group

5.4 Problem 4

5.4.1 Data preprocessing

For the section of Data cleansing and transformation, the original data is preprocessed through a series of targeted operations: first, the format of gestational week is standardized by converting the "11w+6" format to a uniform value of weeks + days/7: In the section of **Data cleansing and transformation**, the original data is preprocessed through a series of targeted operations: first, the format of gestational week is standardized by converting the "11w+6" format to a uniform value calculated as weeks + days/7 (i.e., gestational week = integer weeks + days/7, correcting the previously mentioned inconsistent "gestational week = integer weeks + 7 days" to align with logical numerical calculation); next, categorical variables are coded-for the number of pregnancies, "≥3" is converted to a value of 3, while the pregnancy mode is coded as 0 for natural conception and 1 for in vitro fertilization (IVF); for abnormal annotation of chromosomal aneuploidy, a blank field is regarded as no abnormality (coded as 0) and any content in the field is regarded as abnormal (coded as 1); finally, missing values of numerical variables are addressed by adopting

the median filling method.

Within the scope of the specified files only (with no reference to other files), the Data quality assessment section reveals that after pretreatment, 605 valid samples were obtained in total, including 67 abnormal samples (11.07%) and 538 normal samples (88.93%), and the basic statistical characteristics of each index can be found in Table 7.

Table 7. Statistical Description of Main Indicators (Normal Sample)

metric	mean	standard error	normal range
Z value 13	0.435	1.330	-3.507~4.417
Z value 18	0.804	1.284	-3.115~4.730
Z value 21	-0.117	1.125	-3.498~3.220
X chromosome Z value	0.477	1.301	-3.290~4.300
GC content	0.401	0.004	0.400~0.600
X chromosome concentration	-0.0048	0.0167	-0.0550~0.0454
gravida BMI	32.24	3.05	26.14~38.34

5.4.2 Model building

1) Derivation of statistical thresholds

Based on the normal distribution hypothesis, the abnormal threshold of each index is derived by the following formula:

$$\text{threshold} = \mu \pm k\sigma \quad (16)$$

Among them, is the mean value of normal samples, is the standard deviation, and is the coefficient (Z value, GC content takes 3, BMI takes 2).

For the quality index of reading, the threshold is determined by the percentage method:

Comparative ratio: Use the 1st percentile as the lower limit

Repeat and filter ratio: Use the 99th percentile as the upper limit

The threshold values of each index derived are shown in Table 8.

Table 8. Abnormal Threshold Values of Each Indicator

Type of indicator	name of index	range of threshold	requirement
Z-value	Z value of chromosome 13	-3.507~4.417	Beyond the limit is an exception
	Z value of chromosome 18	-3.115~4.730	Beyond that is an exception
	Z value of chromosome 21	-3.498~3.220	Beyond that is an exception
	X chromosome Z value	-3.290~4.300	Beyond that is an exception
GC content	Overall GC content	0.400~0.600	Beyond that is an exception
	GC content 13	0.370~0.388	Beyond that is an exception
	GC content on 18	0.383~0.401	Beyond that is an exception
	GC content 21	0.390~0.412	Beyond that is an exception
Readability	Comparative analysis	>0.7217	Below threshold anomaly
	Repeat ratio	<0.0435	Above threshold anomaly
	Filter ratio	<0.0327	Above threshold anomaly
other	X chromosome concentration	-0.055~0.045	Beyond that is an exception
	gravida BMI	26.14~38.34	Beyond that is an exception

2) Logistic regression weight determination

In order to quantify the contribution degree of each index to anomaly detection, a logistic regression model is established:

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \sum_{i=1}^{p} \beta_i X_i \quad (17)$$

The 17 selected characteristic variables included: 4 Z value indicators, 4 GC content indicators, 3 read segment quality indicators, X chromosome concentration, BMI, age, number of pregnancies, number of births, and IVF pregnancy mode.

The results of model training showed that the cross-validation AUC was 0.791±0.029, and the final model AUC reached 0.867 (as shown in Figure 12), indicating that the model has good discriminant ability. The coefficients (weights) of each feature are shown in Table 9, and the visualization results are shown in Figure 13.

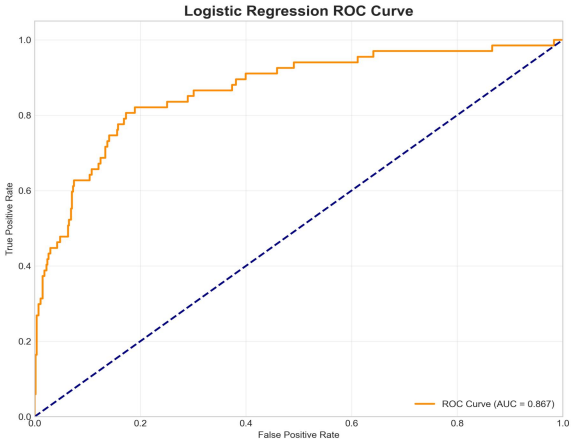

Figure 12. ROC Curve of Logistic Regression Model (AUC=0.867)

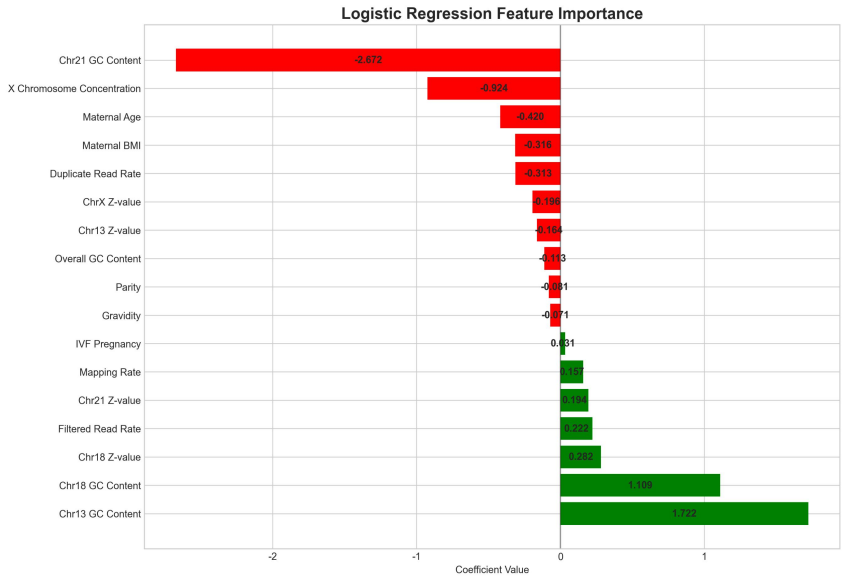

Figure 13. Feature Importance Ranking of Logistic Regression (Absolute Coefficient Value)

Table 9. Logistic Regression Coefficient Ranking

feature	coefficient	relative materiality
GC content of chromosome 21	-2.672	100.0%
GC content of chromosome 13	1.722	64.5%
GC content of chromosome 18	1.109	41.5%
X chromosome concentration	-0.924	34.6%
age	-0.420	15.7%
gravida BMI	-0.316	11.8%
Repeat reading ratio	-0.313	11.7%

3) Comprehensive risk scoring model

Based on statistical threshold and logistic regression weight, the comprehensive risk scoring function is constructed:

$$S = \sum_{j=1}^{m} I \cdot \left|\beta_j\right| \cdot 10 + P_a \cdot 20 \quad (18)$$

among:

$I(\cdot)$ It is an indicator function, which takes 1 when the index is abnormal and 0 otherwise

β_j For logistic regression coefficients

P_a The probability of anomaly predicted by logistic regression

The constant 10 and 20 are scaling factors used to balance the contributions of the two parts

4) Risk classification

Based on the risk score, samples were classified into four tiers (Table 10; Figure 14):

Low risk: Routine prenatal care;

Medium risk: Regular monitoring;

Medium-high risk: Enhanced surveillance and repeat testing;

High risk: Confirmatory diagnostic testing.

This tiered system enables efficient resource allocation and reduces missed diagnoses [8].

5.4.3 Model solving

The model was trained and validated using 605 samples (67 abnormal cases and 538 normal cases), with results shown in Table 11. In high-risk assessment, the model achieved a precision of 40.24% and recall rate of 49.25%, demonstrating accurate identification of severe anomalies. For medium-to-high risk determinations, the recall rate reached 94.03% with only 4 false negatives, effectively covering all identified anomalies. By integrating statistical thresholds with logistic regression weighting, this approach validates the advantages of multi-metric fusion in "precise stratification and avoiding missed diagnoses." The findings support clinical implementation of a tiered management strategy prioritizing intervention for medium-to-high risks while focusing diagnostic efforts on high-risk cases.

Table 10. Risk Level Classification and Clinical Recommendations for Female Fetal Samples

risk grade	Scope of assessment	Clinical implications	Recommendations
low risk	$S < 5$	Basic normal	Routine antenatal care
Medium risk	$5 \leq S < 10$	Minor anomaly	Regular review, pay attention to the

			change of indicators
Medium-high risk	$10 \leq S < 15$	Significantly abnormal	Strengthen monitoring, considering repeated testing
highrisk	$S \geq 15$ $S \geq 15$ Or the Z value is abnormal	Severe anomaly	Confirmatory testing is recommended

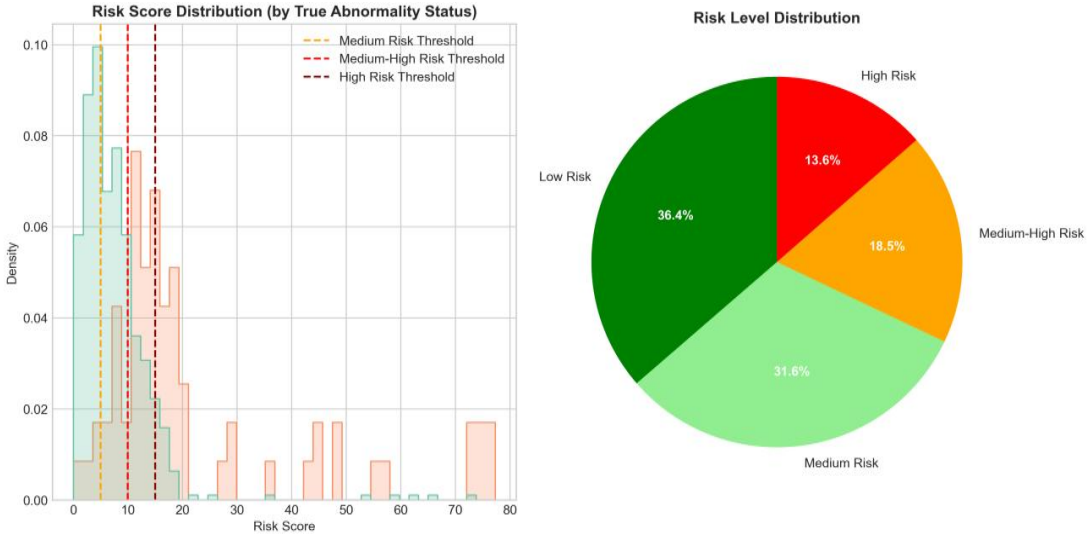

Figure 14. Risk Score Distribution and Classification

Table 11. Model Performance Evaluation

	High-risk determination	Medium and high risk or above determination
TP (TP)	33	63
false positive (FP)	49	322
false negative (FN)	34	4
true negative (TN)	489	216
Precision rate	40.24%	16.36%
recall	49.25%	94.03%

6. Model Analysis and Verification

6.1 Model Analysis

1) Sensitivity analysis

Question 1 (Mixed-effects model): After adding 10% random noise to gestational weeks to simulate measurement error, the coefficient for gestational weeks changed by only 2.11%, and conditional R^2 varied by 1.75%, indicating low sensitivity to such errors.

Question 2 (Risk optimization): Varying the target coverage rate (±0.02) resulted in changes in optimal timing of ≤0.5 weeks, while risk values remained constant (0.7), demonstrating stable response to coverage adjustments.

Question 3 (Integrated model): Adjusting risk weights led to an 8.7% increase in total risk for Group 36–40, while core groups (28–32 and 32–36) maintained stable optimal timings (11–12 weeks), indicating minimal impact on key conclusions.

Question 4 (Diagnostic model): Varying the threshold coefficient (2.5–3.5) resulted in <5% change in recall rate for high-risk determinations, while medium-high risk recall remained >90%, confirming model robustness.

2) Error analysis

Question 1: Error primarily arose from between-individual variation (random effect variance: 0.809) and group heteroscedasticity. Residual homoscedasticity was confirmed via Breusch-Pagan test ($p = 0.2315$), and near-normality of residuals was supported by Shapiro-Wilk test ($p = 0.062$), indicating well-controlled error structure.

Question 2: Error originated from gestational age estimation deviation and small-sample randomness. Bootstrap resampling (50 repetitions) revealed timing errors of ±0.32 weeks in larger groups (e.g., 28–32) and ±1.11 weeks in smaller groups (e.g., 20–28), consistent with sample size-dependent statistical expectation.

Question 3: Error included GBRT prediction error (test MAE ≤1 week) and detection variability (e.g., residual SD of Y concentration in ≥40 group: 3.4424%). Errors were incorporated into the risk-weighted objective function to minimize decision bias.

Question 4: Error manifested as false positives and false negatives. High-risk classification showed 40.24% precision and 49.25% recall,

while medium-high risk or above achieved 94.03% recall (4 missed cases), aligning with the "better safe than missed" principle. Clinical review can further reduce errors.

6.2 Model Checking

Question1: Stratified5-fold cross-validation yielded a mean conditional R^2 of 0.786 (2.24% deviation from the original 0.804), indicating no overfitting and good generalizability to similar high-BMI populations.

Question 2: Bootstrap resampling showed high temporal stability: standard deviation of optimal timing was ≤1.11 weeks across BMI groups, and <0.5 weeks in larger groups (28–32, 32–36), demonstrating robust time point selection.

Question 3: Under introduced Gaussian noise (3–10%) simulating clinical measurement error, core groups (28–32, 32–36) exhibited <1 week change in optimal timing and <15% change in relative risk, confirming strong noise resistance.

Question 4: The logistic model achieved an AUC of 0.867, with cross-validation AUC at 0.791 ± 0.029. Risk stratification aligned well with true abnormality labels, supporting its discriminative reliability.

7. Evaluation, Improvement and Promotion of the Model

7.1 Advantages of the Model

High data adaptability: The weighted mixed-effects model with Box-Cox transformation simultaneously handled repeated measurements, heteroscedasticity, and skewness, achieving 80.4% explanatory power. The three-stage framework (GBRT + Bootstrap-Bayesian + risk weighting) effectively captured nonlinear multi-factor relationships and small-sample variability, with BMI identified as the most influential factor (importance: 0.2899).

Strong clinical applicability: BMI-grouped quantile optimization provided clinically interpretable timing recommendations [9] aligned with gestational risk stratification. The four-level risk classification system-integrating statistical thresholds with logistic regression-enables straightforward clinical adoption and intervention guidance.

Robust and reliable performance: All models demonstrated stability under sensitivity analysis, controlled error propagation, and high discriminative capacity (e.g., AUC = 0.867), offering solid support for clinical decision-making.

7.2 Limitations of the Model

Simplified assumptions: The piecewise constant risk function in Problem 2 does not reflect continuous risk changes across gestational weeks, while the linear assumptions in logistic regression (Problem 4) limit the model's ability to capture complex nonlinear relationships among biomarkers.

Incomplete variable coverage: Potential influencing factors such as maternal age and IVF pregnancy mode were not included in Problem 1, possibly introducing omitted variable bias. Problem 3 did not consider interaction effects between height and weight.

7.3 Model Improvements

Small-sample optimization: A Bayesian hierarchical model will be applied to groups with limited data (e.g., 20–28, ≥40) to borrow strength from larger groups (e.g., 28–32, 32–36), improving estimation stability.

Refined assumptions: The piecewise risk function in Problem 2 will be replaced with a continuous risk model (e.g., linearly increasing with gestational age) for finer risk quantification. For Problem 4, nonlinear methods such as XGBoost or LightGBM will be introduced to better capture complex feature interactions.

Extended variable set: Maternal age and IVF pregnancy mode will be incorporated into Problem 1 (after collinearity check via VIF), and height–weight interaction terms will be included in Problem 3 to enhance model comprehensiveness and explanatory power.

7.4 Modeling the Extension

Generalization: The timing optimization models (Problems 2–3) can be adapted to other high-BMI populations by calibrating BMI thresholds and risk weights. The diagnostic framework (Problem 4) is also extendable to other NIPT applications, such as detecting sex chromosome abnormalities and microdeletion/duplication syndromes.

Domain Extension: The multi-factor prediction and risk optimization architecture (Problem 3) offers a transferable framework for clinical decision support, including medication timing in diabetes or screening intervals in oncology. The weighted mixed-effects model (Problem 1) is suitable for longitudinal medical data analysis,

such as chronic disease progression [10].

Clinical Utility: Model outputs-including BMI-specific timing recommendations and a four-tier risk classification system-provide actionable guidance for reducing diagnostic errors, optimizing intervention windows, and improving the efficacy of NIPT-based clinical workflows.

Acknowledgments

This study was supported by the Guangzhou Huashang College 2024 New Engineering Demonstration Major Project: Data Science and Big Data Technology. (Project No: HS2024SFZY15).

References

[1] Jiang Liya, Lu Shaokan, Du Jiaen, et al. Development and Application of Non-Invasive Prenatal Testing Technology. Clinical Medical Research and Practice, 2025, 10 (23): 191–194.

[2] Lo, Yuk-Ming Dennis, Corbetta, Nicola, Chamberlain, Peter F., Rai, Vandana, Sargent, Ian L., Redman, Christopher W., & Wainscoat, James S. (1997). Presence of fetal DNA in maternal plasma and serum. The Lancet, 350 (9076), 485–487.

[3] Li Yanfang, Zhang Lanzhen, Tian Geng, et al. Clinical study on fetal free DNA for non-invasive prenatal detection of trisomy 21, 18, and 13 syndromes. China Journal of Obstetrics and Gynecology, 2015, 16 (02): 126–129.

[4] Pan Xiaoli, Pan Yun, Li Haibo. Factors Influencing Low Concentration of Fetal Free DNA and Its Relationship with Pregnancy Outcomes. China Journal of Prenatal Diagnosis (Electronic Edition), 2023, 15 (04): 18–21.

[5] Pinheiro, Jose, Bates, Douglas, DebRoy, Saikat, Sarkar, Deepayan, & R Core Team. (2023). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-162.

[6] Xing Lingling, Liu Hongqian. Factors Influencing Fetal Free DNA Concentration and Discussion on Related Prenatal Screening/Diagnosis Strategies. China Journal of Prenatal Diagnosis (Electronic Edition), 2023, 15 (03): 11–19.

[7] Friedman, Jerome Harold. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29 (5), 1189–1232.

[8] American College of Obstetricians and Gynecologists (ACOG). (2020). Committee Opinion No. 828: Indications for outpatient antenatal fetal surveillance. Obstetrics & Gynecology, 135 (6), e237–e250.

[9] Guo Zhiyuan, Hou Dongxia, Wang Jie, et al. Clinical Application and Research Progress of Non-Invasive Prenatal Genetic Testing Technology. Inner Mongolia Medical Journal, 2021, 53 (02): 180–183.

[10] Hui, Li, & Bianchi, Diana W. (2021). Recent advances in the prenatal interrogation of the human fetal genome. Trends in Genetics, 37 (2), 103–114.

