Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 10(2): 3523–3560.
DOI: 10.3934/math.2025163
Received: 06 November 2024
Revised: 05 February 2025
Accepted: 11 February 2025
Published: 24 February 2025

Research article

Single-index logistic model for high-dimensional group testing data

Changfu Yang[1,†], Wenxin Zhou[2,†], Wenjun Xiong[1,*], Junjian Zhang[1,*] and Juan Ding[2]

[1] School of Mathematics and Statistics, Guangxi Normal University, Guilin 541004, China
[2] School of Mathematics, Hohai University, Nanjing 210098, China
† Changfu Yang and Wenxin Zhou contributed equally to this work.

* **Correspondence:** Email: wjxiong@gxnu.edu.cn, jjzhang@gxnu.edu.cn.

Abstract: Group testing is an efficient screening method that reduces the number of tests by pooling multiple samples, making it especially effective in low-prevalence settings. This strategy gained significant attention during the COVID-19 pandemic, and has since been applied to detect various infectious diseases, including HIV, chlamydia, gonorrhea, influenza, and Zika virus. In this paper, we introduce a semi-parametric logistic single-index model for analyzing high-dimensional group testing data, which is particularly flexible in capturing complex nonlinear relationships. The proposed method achieves variable selection by parameter regularization, which proves especially beneficial for extracting relevant information from high-dimensional data. The performance of the model is evaluated through simulations across four group testing strategies: master pool testing, Dorfman testing, halving testing, and array testing. Further validation is provided using real-world data. The results demonstrate that our approach offers a flexible and robust tool for analyzing high-dimensional group testing data, with important applications in epidemiology and public health.

Keywords: group testing; latent variable; single-index model; high-dimensional data; variable selection; EM algorithm
Mathematics Subject Classification: 62G08, 62P10

1. Introduction

Group testing, or pooled testing, was first introduced by Dorfman [1] to identify syphilis infections among U.S. Army personnel during World War II. This approach involves combining specimens (e.g., blood, plasma, urine, swabs) from multiple individuals and conducting a single test to check for infection. According to Dorfman's procedure, if the combined sample tests negative, all individuals in this sample can be confirmed disease-free. Conversely, a positive result necessitates further testing to identify the affected individuals. This strategy gained prominence during the COVID-19 pandemic [2–4] and has been applied to detect various infectious diseases, including HIV [5,6], chlamydia and gonorrhea [7], influenza [8], and the Zika virus [9]. The primary motivation for pooled testing lies in its economic efficiency; for instance, the State Hygienic Laboratory at the University of Iowa saved approximately $3.1 million over five years by employing a modified Dorfman protocol for testing chlamydia and gonorrhea among residents of Iowa [10,11].

Despite its cost-effectiveness, group testing poses significant challenges for statistical analysis due to the absence of individual response data [12]. However, advancements in digital technology have provided access to rich covariate information, including demographic data, electronic health records, genomic data, lifestyle data, physiological monitoring data, imaging data, and environmental variables [13]. Integrating these covariates into various statistical models for group testing has been shown to enhance accuracy and robustness, as evidenced by studies from Mokalled et al. [14], Huang and Warasi [15], Haber et al. [16]. This integration leads to improved estimations of individual risk probabilities, thereby reducing the number of tests required and overall costs.

In managing covariates, single-index models offer advantages, such as less restrictive assumptions, good interpretability, and adaptability to high-dimensional data [17]. For high-dimensional single-index models, Radchenko [18] proposed a novel estimation method based on L_1 regularization, extending it to generalized linear models. Elmezouar et al. [19] developed a functional single index expectile model with a nonparametric estimator to address spatial dependency in financial data, showing strong consistency and practical applicability. Chen and Samworth [20] explored generalized additive models, deriving non-parametric estimators for each additive component by maximizing the likelihood function, and adapted this approach to generalized additive index models. Kereta et al. [21] employed a k-nearest neighbor estimator, enhanced by geodesic metrics, to extend local linear regression for single-index models. However, research on generalized semi-parametric single-index models in high-dimensional contexts remains limited, particularly in group testing applications, which are still underexplored.

Most current integrations of covariate information with group testing are developed based on parametric regression models. For example, Wang et al. [22] introduced a comprehensive binary regression framework, while McMahan et al. [11] developed a Bayesian regression framework. Gregory et al. [23] adopted an adaptive elastic net method, which remains effective as data dimensionality increases. Ko et al. [24] compared commonly used group testing procedures with group lasso regarding true positive selection in high-dimensional genomic data analysis. Furthermore, nonparametric regression methods have gained traction for applying covariates in group testing. Delaigle and Hall [25] proposed a nonparametric method for estimating conditional probabilities and testing specificity and sensitivity, addressing the unique dilution effects and complex data structures inherent in group testing. Self et al. [26] introduced a Bayesian generalized additive regression method to tackle dilution effects further, while Yuan et al. [12] developed a semiparametric monotone regression model using the expectation-maximization (EM) algorithm to navigate the complexities of group testing data. Zuo et al. [27] proposed a more flexible generalized nonparametric additive model, utilizing B-splines and group lasso methods for model estimation in high-dimensional data.

This article proposes a generalized single-index group testing model aimed at enhancing flexibility in addressing various nonlinear models and facilitating the selection of important variables. Given the absence of individual disease testing results in group testing data, the EM algorithm is employed to perform the necessary calculations for the model. B-spline functions are utilized to approximate the nonlinear unknown smooth functions, with model parameters estimated by maximizing the likelihood function. In modern group testing, a substantial amount of individual covariate information is typically collected during sample testing. Consequently, a penalty term is incorporated into the likelihood function, promoting the construction of a sparse model and enabling effective variable selection. We apply the method to four group testing strategies: master pool, Dorfman, halving, and array. The method is evaluated using both simulated and real data.

The remaining sections are organized as follows. Section 2 introduces our model with B-spline approximation, detailing the corresponding algorithm employing the EM algorithm. Section 3 elaborates on the E-step in the EM algorithm, facilitating the acceleration of our algorithm's convergence. Sections 4 and 5 present comprehensive simulations and real data application, demonstrating the method's robust performance. Finally, we conclude our findings and provide some discussion in Section 6.

2. Primary results and methodological advancements

2.1. Logistic single-index model for high-dimensional covariates

Consider a dataset comprising n individuals. For each $i \in \{1, 2, \ldots, n\}$, let the true disease status of the i-th individual be denoted by $\tilde{Y}_i \in \{0, 1\}$, where $\tilde{Y}_i = 1$ indicates disease presence, and $\tilde{Y}_i = 0$ indicates absence. Additionally, the dataset includes covariate information for each individual, represented as $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{iq_n})^T \in \mathbb{R}^{q_n}$, where $\mathbb{R}^{q_n}$ denotes a q_n-dimensional real vector space. We assume the number of covariates q_n is high-dimensional.

Let the risk probability for the i-th individual be defined as $p_i = \Pr(\tilde{Y}_i = 1 \mid \boldsymbol{X}_i)$, where $i \in \{1, 2, \ldots, n\}$. In many cases, the influence of covariates may be nonlinear; imposing linearity can result in inaccurate estimations. This study explores nonlinear scenarios, assuming p_i follows a flexible logistic single-index model, expressed as

$$\Pr(\tilde{Y}_i = 1 | \boldsymbol{X}_i) = \frac{\exp[g(\boldsymbol{X}_i^\top \boldsymbol{\beta})]}{1 + \exp[g(\boldsymbol{X}_i^\top \boldsymbol{\beta})]}, \tag{2.1}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_{q_n})^\top \in \mathbb{R}^{q_n}$ represents the unknown parameters, and $g(\cdot)$ is an unknown smooth function capturing the relationship between covariates and risk probabilities.

In semiparametric single-index models, the true parameters are generally considered non-identifiable without imposed constraints. To ensure the identifiability of $\boldsymbol{\beta}$, we impose a classical constraint: $\beta_1 = \sqrt{1 - \|\boldsymbol{\beta}_{-1}\|_2^2}$, where $\boldsymbol{\beta}_{-1} = (\beta_2, \beta_3, \ldots, \beta_{q_n})^\top \in \mathbb{R}^{q_n - 1}$, and $\| \cdot \|_2$ denotes the L_2-norm. Note that both the function $g(\cdot)$ and the coefficient $\boldsymbol{\beta}$ in the single-index model are unknown. The L_2-norm constraint $\|\boldsymbol{\beta}\|_2 = 1$ is crucial for the identifiability of $\boldsymbol{\beta}$ as shown by Carroll et al. [28], Zhu et al. [29], Lin et al. [30], Cui et al. [31], and Guo et al. [32]. We assume that the true parameter $\boldsymbol{\beta}^*$ is sparse, defining the true model as $\mathcal{M}^* = \{j \in \{1, 2, \ldots, q_n\} : \beta_j^* \neq 0\}$.

For $i \in \{1, 2, \ldots, n\}$, $\tilde{Y}_i$ follows a Binomial distribution with parameter p_i, denoted as $\tilde{Y}_i \sim \text{Binom}(1, p_i)$. In traditional single-index model studies, the true status $\tilde{\mathcal{Y}} = \{\tilde{Y}_i, i = 1, 2, \ldots, n\}$, is directly observable. However, in group testing, $\tilde{\mathcal{Y}}$ is unobservable [33]. This paper investigates parameter estimation and statistical inference of single-index models based on group testing data. Moreover, if a group test result is positive, further testing is required to identify infected individuals.

These results may depend on shared characteristics, leading to correlations within group test outcomes, complicating the modeling.

In group testing, we partition n individuals into J groups, denoted as $\mathcal{P}_{1,1}, \mathcal{P}_{2,1}, \ldots, \mathcal{P}_{J,1}$. Here, $\mathcal{P}_{j,1}$ represents the initial index set of individuals for the j-th group, ensuring $\cup_{j=1}^{J} \mathcal{P}_{j,1} = \{1, 2, \ldots, n\}$. For $j \in \{1, 2, \ldots, J\}$, if any testing result for $\mathcal{P}_{j,1}$ is positive, further testing may be warranted. Define $\mathcal{Z}_j = \{Z_{j,l},\ l = 1, 2, \ldots, L_j\}$ as the set of testing outcomes for the j-th group, where L_j denotes the total number of tests conducted within j-th group. Each $Z_{j,l} \in \{0, 1\}$, where $Z_{j,l} = 0$ indicates a negative result and $Z_{j,l} = 1$ indicates a positive result. If $Z_{j,1} = 0$, then $L_j = 1$; otherwise, $L_j \geq 1$. Let $\boldsymbol{\mathcal{P}}_j = \{\mathcal{P}_{j,l}, l = 1, 2, \ldots, L_j\}$, where $\mathcal{P}_{j,l}$ corresponds to the individuals associated with $Z_{j,l}$. Define $\tilde{\mathcal{Z}}_j = \{\tilde{Z}_{j,l}, l = 1, 2, \ldots, L_j\}$ as the true status corresponding to $\mathcal{Z}_j$. The true statuses of individuals determine the group's true status, defined as $\tilde{Z}_{j,l} = I\left(\sum_{i \in \mathcal{P}_{j,l}} \tilde{Y}_i\right)$, where $I(\cdot)$ denotes the indicator function.

In practical applications, measurement error of the test kits exists. We define $S_e = \Pr(Z_{j,l} = 1 \mid \tilde{Z}_{j,l} = 1)$ as sensitivity, representing the probability of correctly identifying positive samples, and $S_p = \Pr(Z_{j,l} = 0 \mid \tilde{Z}_{j,l} = 0)$ as specificity, denoting the probability of correctly identifying negative samples, where $l \in \{1, 2, \ldots, L_j\}$ and $j \in \{1, 2, \ldots, J\}$. According to the definitions of S_e and S_p, given the true status $\tilde{Z}_{j,l}$, the group's testing results satisfy $Z_{j,l} | \tilde{Z}_{j,l} \sim \mathrm{Binom}\left(1, \mathrm{S_e}^{\tilde{Z}_{j,l}} (1 - \mathrm{S_p})^{1-\tilde{Z}_{j,l}}\right)$.

Our approach is based on two widely accepted fundamental assumptions in group testing. The first assumption is that S_e and S_p are independent of group size, supported by various studies [34–37]. The second assumption posits that, given the true statuses of individuals in the j-th group $\{\tilde{Y}_i, i \in \mathcal{P}_{j,1}\}$, the group's true statuses $\tilde{\mathcal{Z}}_j$ are mutually independent, as supported by previous research [23, 34, 35].

We apply our method to four group testing methods: master pool testing, Dorfman testing, halving testing, and array testing. Figure 1 illustrates the process of four testing methods: (a) Master pool testing, where a group of individuals (e.g., $\mathcal{P}_{j,1}$ consisting of individuals 1, 2, 3, and 4) is tested as a whole to obtain the group testing result $Z_{j,1}$; (b) Dorfman testing, where initially the same group testing as in master pool testing is conducted, and if the result of the master pool testing is positive ($Z_{j,1} = 1$), each individual in the group is then tested separately to obtain individual testing results $Z_{j,2}$, $Z_{j,3}$, $Z_{j,4}$, and $Z_{j,5}$; (c) Halving testing, where the entire group (e.g., $\mathcal{P}_{j,1}$) is tested as a whole, and if the result is positive ($Z_{j,1} = 1$), the group is divided into two subgroups (e.g., $\mathcal{P}_{j,2}$ and $\mathcal{P}_{j,3}$) for subgroup testing, and if the result of subgroup testing is positive (e.g., $Z_{j,2} = 1$), individuals in the positive subgroup are then tested individually; and (d) Array testing, where multiple individuals (e.g., 16 individuals) are arranged in an array for group testing to obtain multiple group testing results such as $Z_{j,1}$, and if a specific group testing result is positive (e.g., $Z_{j,1} = 1$), further subgroup testing is performed (e.g., obtaining results $Z_{j,2}$, $Z_{j,3}$), and if the group testing results for both the row and column where an individual is located are positive (e.g., $Z_{j,3} = Z_{j,4} = Z_{j,7} = 1$), the individuals (e.g., 6-th individual and 10-th individual) are then tested.

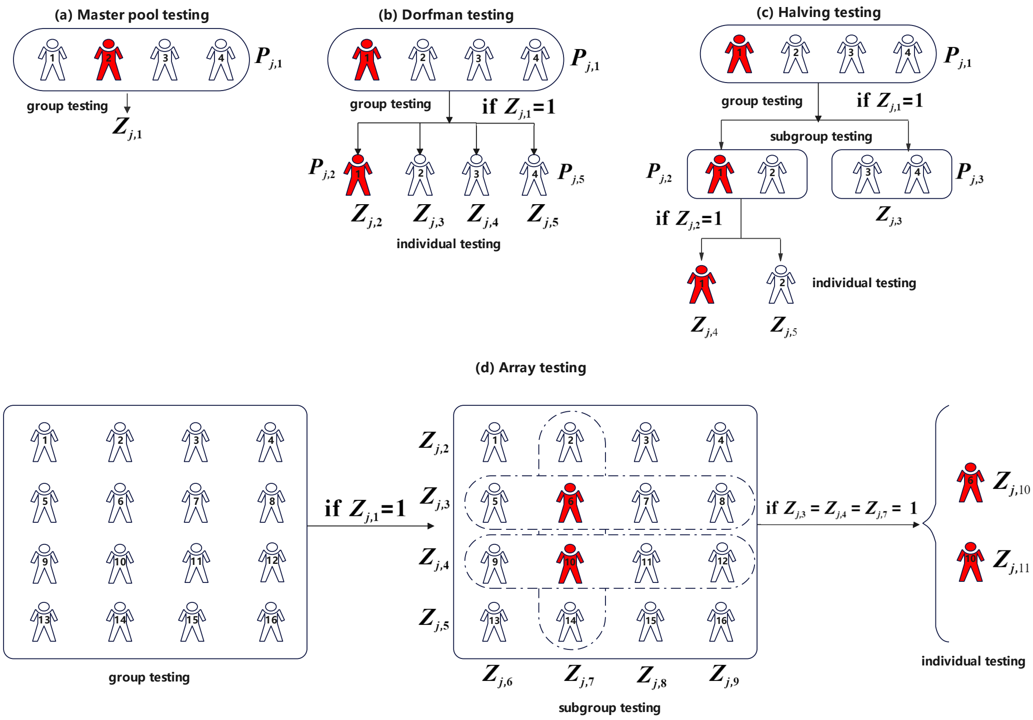

Figure 1. A flowchart of four group testing procedure.

Due to the nature of group testing, the true status of individuals, denoted as $\tilde{\mathcal{Y}}$, remains unknown. Our objective is to estimate $\boldsymbol{M}^*$, $\boldsymbol{\beta}^*$, and $g(\cdot)$ based on observed data $\mathcal{Z} = \{\mathcal{Z}_j,\ j = 1, 2, \ldots, J\}$ and covariate information $\mathbb{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n)^T \in \mathbb{R}^{n\times q_n}$ to ascertain individual risk probabilities. The likelihood function based on the observed data $\mathcal{Z}$ is defined as

$$P(\mathcal{Z}|\mathbb{X}) = \sum_{\tilde{\mathcal{Y}}\in\{0,1\}^n} P(\mathcal{Z}|\tilde{\mathcal{Y}})P(\tilde{\mathcal{Y}}|\mathbb{X}), \tag{2.2}$$

where

$$P(\mathcal{Z} \mid \tilde{\mathcal{Y}}) = \prod_{j=1}^{J}\prod_{l=1}^{L_j} P\left(Z_{j,l} \mid \tilde{\mathcal{Y}}_{\mathcal{P}_{j,l}}\right),$$

and $\tilde{\mathcal{Y}}_{\mathcal{P}_{j,l}} = \{\tilde{Y}_i,\ i \in \mathcal{P}_{j,l}\}$ represents the set of true statuses for individuals in $\mathcal{P}_{j,l}$. Furthermore, the conditional probability $P\left(Z_{j,l} \mid \tilde{\mathcal{Y}}_{\mathcal{P}_{j,l}}\right)$ is expressed as

$$P\left(Z_{j,l} \mid \tilde{\mathcal{Y}}_{\mathcal{P}_{j,l}}\right) = \left\{S_e^{\tilde{Z}_{j,l}}\left(1 - S_p\right)^{1-\tilde{Z}_{j,l}}\right\}^{Z_{j,l}}\left\{(1 - S_e)^{\tilde{Z}_{j,l}} S_p^{1-\tilde{Z}_{j,l}}\right\}^{1-Z_{j,l}}.$$

The likelihood function for the true disease status $\tilde{\mathcal{Y}}$ can be written as

$$P(\tilde{\mathcal{Y}}|\mathbb{X}) = \prod_{i=1}^{n} p_i^{\tilde{Y}_i}(1 - p_i)^{1-\tilde{Y}_i}.$$

Combining this with the logistic single-index model defined in (2.1), we obtain the log-likelihood function for $\tilde{\mathcal{Y}}$:

$$\ln P(\tilde{\mathcal{Y}}|\mathbb{X}) = \sum_{i=1}^{n}\left\{\tilde{Y}_i g(\boldsymbol{X}_i^\top\boldsymbol{\beta}) - \ln\left(1 + \exp\left[g(\boldsymbol{X}_i^\top\boldsymbol{\beta})\right]\right)\right\}. \tag{2.3}$$

Since the smooth function $g(\cdot)$ is unknown, we approximate it using B-spline functions. Let the support interval of $g(\cdot)$ be $[a, b]$. We partition $[a, b]$ at points $a = d_0 < d_1 < \ldots < d_N < b = d_{N+1}$ into several segments, referred to as knots or internal nodes. This division generates subintervals $I_k = [d_k, d_{k+1})$ for $0 \leq k \leq N-1$ and $I_N = [d_N, d_{N+1}]$, ensuring that

$$\frac{\max_{0\leq k\leq N} |d_k - d_{k+1}|}{\min_{0\leq k\leq N} |d_k - d_{k+1}|} \leq M,$$

where $M \in (0, \infty)$. The B-spline basis functions of order q are denoted as $\boldsymbol{\Phi}(\cdot) = \left(\phi_1(\cdot), \phi_2(\cdot), \ldots, \phi_S(\cdot)\right)^{\top} \in \mathbb{R}^S$, with $S = N + q$. Thus, $g(\cdot)$ can be approximated as

$$g(\cdot) \approx \sum_{s=1}^{S} \phi_s(\cdot)\gamma_s,$$

where γ_s are the spline coefficients to be estimated [38]. Denote $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_S)^{\top} \in \mathbb{R}^S$. We approximate $g(\boldsymbol{X}_i^T\boldsymbol{\beta})$ as

$$g(\boldsymbol{X}_i^T\boldsymbol{\beta}) = \boldsymbol{\Phi}^{\top}(\boldsymbol{X}_i^{\top}\boldsymbol{\beta})\boldsymbol{\gamma},$$

where $\boldsymbol{\Phi}(\boldsymbol{X}_i^{\top}\boldsymbol{\beta}) = \left(\phi_1(\boldsymbol{X}_i^{\top}\boldsymbol{\beta}), \phi_2(\boldsymbol{X}_i^{\top}\boldsymbol{\beta}), \ldots, \phi_S(\boldsymbol{X}_i^{\top}\boldsymbol{\beta})\right)^{\top}$. Therefore, we approximate p_i by using a spline function, and denote the spline approximation of p_i as p_{iB}, which is defined as follows:

$$p_{iB} = \frac{\exp\left[\boldsymbol{\Phi}^{\top}(\boldsymbol{X}_i^{\top}\boldsymbol{\beta})\boldsymbol{\gamma}\right]}{1 + \exp\left[\boldsymbol{\Phi}^{\top}(\boldsymbol{X}_i^{\top}\boldsymbol{\beta})\boldsymbol{\gamma}\right]}. \tag{2.4}$$

In the following, we use the spline approximation p_{iB} of p_i to construct the log-likelihood function and the objective function in the subsequent EM algorithm. Thus, the log-likelihood function (2.3) for $\tilde{\mathcal{Y}}$ can be reformulated as

$$\ln P_B(\tilde{\mathcal{Y}}|\mathbb{X}) = \sum_{i=1}^{n} \left\{\tilde{Y}_i \boldsymbol{\Phi}^{\top}\left(\boldsymbol{X}_i^{\top}\boldsymbol{\beta}\right)\boldsymbol{\gamma} - \ln\left(1 + \exp\left[\boldsymbol{\Phi}^{\top}(\boldsymbol{X}_i^{\top}\boldsymbol{\beta})\boldsymbol{\gamma}\right]\right)\right\}.$$

Furthermore, the target likelihood function (2.2) can be represented as

$$P_B(\mathcal{Z}|\mathbb{X}) = \sum_{\tilde{\mathcal{Y}}\in\{0,1\}^n} P(\mathcal{Z}|\tilde{\mathcal{Y}})P_B(\tilde{\mathcal{Y}}|\mathbb{X}).$$

By employing spline approximation, we transform the estimation problem of $\boldsymbol{\beta}^*_{-1}$ and $g(\cdot)$ into estimating $\boldsymbol{\beta}^*_{-1}$ and $\boldsymbol{\gamma}$.

For high-dimensional group testing data, we aim to estimate $\boldsymbol{\beta}^*_{-1}$ using the penalized approach within a single-index model framework. The penalized log-likelihood function is defined as follows:

$$\ln P_B(\mathcal{Z}|\mathbb{X}) - \sum_{j=2}^{q_n} P_\lambda(\beta_j), \tag{2.5}$$

where $P_\lambda(\cdot)$ is the penalty function and λ is a tuning parameter. We consider three common penalty functions: LASSO [39], SCAD [40], and MCP [41]. Specifically, for LASSO, $P_\lambda(x) = \lambda|x|$. For SCAD, it is defined as

$$P_\lambda(x) = \begin{cases} \lambda|x| & \text{if } |x| \leq \lambda, \\ \frac{-x^2+2\delta\lambda|x|-\lambda^2}{2(\delta-1)} & \text{if } \lambda < |x| \leq \delta\lambda, \\ \frac{(\delta+1)\lambda^2}{2} & \text{if } |x| > \delta\lambda, \end{cases}$$

where $\delta > 2$. In MCP, the penalty function is given by

$$P_\lambda(x) = \begin{cases} \lambda|x| - \frac{x^2}{2\delta} & \text{if } |x| \le \delta\lambda, \\ \frac{1}{2}\delta\lambda^2 & \text{if } |x| > \delta\lambda, \end{cases}$$

with $\delta > 1$. The following section will detail the parameter estimation process.

2.2. EM algorithm for regularized single-index model in group testing

The penalized log-likelihood function (2.5) lacks the individual latent status $\tilde{\mathcal{Y}}$. The complete data penalized log-likelihood function can be expressed as

$$\ln P_B(\mathcal{Z}, \tilde{\mathcal{Y}}|\mathbb{X}) - \sum_{j=2}^{q_n} P_\lambda(\beta_j) = \ln P(\mathcal{Z}|\tilde{\mathcal{Y}}) + \ln P_B(\tilde{\mathcal{Y}}|\mathbb{X}) - \sum_{j=2}^{q_n} P_\lambda(\beta_j). \tag{2.6}$$

Notably, $\ln P(\mathcal{Z}|\tilde{\mathcal{Y}})$ depends solely on known parameters S_e and S_p, allowing us to disregard it in computations. The presence of the latent variable $\tilde{\mathcal{Y}}$ complicates direct maximization of the complete data penalized log-likelihood function (2.6). Therefore, we employ the EM algorithm, comprising two steps: the Expectation (E) step, and the Maximization (M) step.

In the E step, given the observed data $\mathcal{Z}$ and the parameters from the t-th iteration $(\boldsymbol{\beta}_{-1}^{(t)}, \boldsymbol{\gamma}^{(t)})$, calculate the following function:

$$\begin{aligned} S^{(t)}(\boldsymbol{\beta}_{-1}, \boldsymbol{\gamma}) =& \mathbb{E}\left\{ \sum_{i=1}^{n} \left\{ \tilde{Y}_i \boldsymbol{\Phi}^\top (\boldsymbol{X}_i^\top \boldsymbol{\beta}) \boldsymbol{\gamma} - \ln\left(1 + \exp\left[\boldsymbol{\Phi}^\top(\boldsymbol{X}_i^\top \boldsymbol{\beta})\boldsymbol{\gamma}\right]\right)\right\} \Big| \mathcal{Z}, \boldsymbol{\beta}_{-1}^{(t)}, \boldsymbol{\gamma}^{(t)} \right\} - \sum_{j=2}^{q_n} P_\lambda(\beta_j) \\ =& \sum_{i=1}^{n} \left\{ w_i^{(t)} \boldsymbol{\Phi}^\top (\boldsymbol{X}_i^\top \boldsymbol{\beta}) \boldsymbol{\gamma} - \ln\left(1 + \exp\left[\boldsymbol{\Phi}^\top(\boldsymbol{X}_i^\top \boldsymbol{\beta})\boldsymbol{\gamma}\right]\right)\right\} - \sum_{j=2}^{q_n} P_\lambda(\beta_j), \end{aligned} \tag{2.7}$$

where $w_i^{(t)} = \mathbb{E}[\tilde{Y}_i|\mathcal{Z}, \boldsymbol{\gamma}^{(t)}, \boldsymbol{\beta}_{-1}^{(t)}]$, $i = 1, 2, \ldots, n$. The calculation of the $w_i^{(t)}$ varies among the four grouping testing methods, which will be discussed in Section 3.

In the M step, we update $\boldsymbol{\beta}_{-1}^{(t+1)}$ and $\boldsymbol{\gamma}^{(t+1)}$, respectively. Initially, we update $\boldsymbol{\gamma}^{(t+1)}$ by maximizing:

$$S^{(t)}(\boldsymbol{\beta}_{-1}^{(t)}, \boldsymbol{\gamma}) = \sum_{i=1}^{n} \left\{ w_i^{(t)} \boldsymbol{\Phi}^\top \left(\boldsymbol{X}_i^\top \boldsymbol{\beta}^{(t)}\right) \boldsymbol{\gamma} - \ln\left(1 + \exp\left[\boldsymbol{\Phi}^\top(\boldsymbol{X}_i^\top \boldsymbol{\beta}^{(t)})\boldsymbol{\gamma}\right]\right)\right\} - \sum_{j=2}^{q_n} P_\lambda(\beta_j^{(t)}). \tag{2.8}$$

Subsequently, we maximize $S^{(t)}(\boldsymbol{\beta}_{-1}, \boldsymbol{\gamma}^{(t+1)})$ to update the parameters $\boldsymbol{\beta}_{-1}^{(t+1)}$:

$$S^{(t)}(\boldsymbol{\beta}_{-1}, \boldsymbol{\gamma}^{(t+1)}) = \sum_{i=1}^{n} \left\{ w_i^{(t)} \boldsymbol{\Phi}^\top (\boldsymbol{X}_i^\top \boldsymbol{\beta}) \boldsymbol{\gamma}^{(t+1)} - \ln\left(1 + \exp\left[\boldsymbol{\Phi}^\top(\boldsymbol{X}_i^\top \boldsymbol{\beta})\boldsymbol{\gamma}^{(t+1)}\right]\right)\right\} - \sum_{j=2}^{q_n} P_\lambda(\beta_j). \tag{2.9}$$

Given that $\boldsymbol{\beta}_{-1}$ appears in each B-spline basis function $\phi(\boldsymbol{X}_i^\top \boldsymbol{\beta})$, direct iteration presents challenges. Let $\tilde{g}^{(t)}(\boldsymbol{X}_i^T \boldsymbol{\beta}) = \boldsymbol{\Phi}^\top(\boldsymbol{X}_i^T \boldsymbol{\beta})\boldsymbol{\gamma}^{(t+1)}$. We apply the approach by Guo et al. [42], approximating $\tilde{g}^{(t)}(\boldsymbol{X}_i^T \boldsymbol{\beta})$ via a first-order Taylor expansion

$$\tilde{g}^{(t)}(\boldsymbol{X}_i^T \boldsymbol{\beta}) \approx \tilde{g}^{(t)}(\boldsymbol{X}_i^T \boldsymbol{\beta}^{(t)}) + \tilde{g}^{(t)\prime}(\boldsymbol{X}_i^T \boldsymbol{\beta}^{(t)}) \boldsymbol{X}_i^T \boldsymbol{J}(\boldsymbol{\beta}^{(t)})(\boldsymbol{\beta}_{-1} - \boldsymbol{\beta}_{-1}^{(t)}),$$

where $\boldsymbol{J}(\boldsymbol{\beta}) = \partial\boldsymbol{\beta}/\partial\boldsymbol{\beta}_{-1} = \left(-\boldsymbol{\beta}_{-1}/\sqrt{1-\|\boldsymbol{\beta}_{-1}\|_2^2}, \boldsymbol{I}_{q_n-1}\right)^{\top}$ represents the Jacobian matrix of size $q_n\times(q_n-1)$ and $\boldsymbol{I}_{q_n-1}$ denotes the (q_n-1)-dimensional identity matrix. This approximation is incorporated into $S^{(t)}(\boldsymbol{\beta}_{-1}, \boldsymbol{\gamma}^{(t+1)})$ to maximize the expression and update $\boldsymbol{\beta}_{-1}^{(t+1)}$. Therefore, we approximate $S^{(t)}(\boldsymbol{\beta}_{-1}, \boldsymbol{\gamma}^{(t+1)})$ by $\tilde{S}^{(t)}(\boldsymbol{\beta}_{-1}, \boldsymbol{\gamma}^{(t+1)})$ as follows:

$$\begin{aligned}\tilde{S}^{(t)}(\boldsymbol{\beta}_{-1}, \boldsymbol{\gamma}^{(t+1)}) &= \sum_{i=1}^{n}\left\{w_i^{(t)}\tilde{g}^{(t)}(\boldsymbol{X}_i^{\top}\boldsymbol{\beta}) - \ln\left(1+\exp\left[\tilde{g}^{(t)}(\boldsymbol{X}_i^{\top}\boldsymbol{\beta})\right]\right)\right\} - \sum_{j=2}^{q_n}P_\lambda(\beta_j)\\ &= \sum_{i=1}^{n}\Bigg\{w_i^{(t)}\left[\tilde{g}^{(t)}(\boldsymbol{X}_i^T\boldsymbol{\beta}^{(t)}) + \tilde{g}^{(t)\prime}(\boldsymbol{X}_i^T\boldsymbol{\beta}^{(t)})\boldsymbol{X}_i^T\boldsymbol{J}(\boldsymbol{\beta}^{(t)})(\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^{(t)})\right]\\ &\quad - \ln\left(1+\exp\left[\tilde{g}^{(t)}(\boldsymbol{X}_i^T\boldsymbol{\beta}^{(t)}) + \tilde{g}^{(t)\prime}(\boldsymbol{X}_i^T\boldsymbol{\beta}^{(t)})\boldsymbol{X}_i^T\boldsymbol{J}(\boldsymbol{\beta}^{(t)})(\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^{(t)})\right]\right)\Bigg\}\\ &\quad - \sum_{j=2}^{q_n}P_\lambda(\beta_j).\end{aligned} \tag{2.10}$$

We employ stochastic gradient descent [43] and coordinate descent [44] to update $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$, respectively. Let $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\beta}}_{-1}$ denote the estimated parameters, and $\hat{\mathcal{M}} = \{j \in \{1, 2, \ldots, q_n\} : \hat{\beta}_j \neq 0\}$ represent the estimated model. Furthermore, $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\beta}}_{-1}$ can be used to calculate individual risk probabilities and guide subsequent testing strategies. In summary, the EM algorithm offers a structured approach to handle the latent variable $\tilde{\mathcal{Y}}$ and estimate model parameters. The detailed steps of this method are summarized in Algorithm 1.

Algorithm 1 Regularized single-index model for group testing.

Input: $\mathcal{Z}$, $\mathbb{X}$, t_{max} and initialization $(\boldsymbol{\beta}_{-1}^{(0)}, \boldsymbol{\gamma}^{(0)})$.
For: $t = 0, 1, 2, \ldots, t_{max}$

- **Step 1 (E-step):** In the E step, given the parameters $(\boldsymbol{\beta}_{-1}^{(t)}, \boldsymbol{\gamma}^{(t)})$ and $\mathcal{Z}$, calculate the conditional expectation $S^{(t)}(\boldsymbol{\beta}_{-1}, \boldsymbol{\gamma})$ in (2.7).
- **Step 2 (M-step):** Update the iterative parameters $\boldsymbol{\beta}_{-1}^{(t+1)}$ and $\boldsymbol{\gamma}^{(t+1)}$ in two substeps:
  1. Update $\boldsymbol{\gamma}^{(t+1)}$ by maximizing $S^{(t)}(\boldsymbol{\beta}_{-1}^{(t)}, \boldsymbol{\gamma})$ in (2.8).
  2. Update $\boldsymbol{\beta}_{-1}^{(t+1)}$ by maximizing $\tilde{S}^{(t)}(\boldsymbol{\beta}_{-1}, \boldsymbol{\gamma}^{(t+1)})$ in (2.10).

End for: Repeat steps 1 and 2 until parameters converge or reach the maximum number of iterations t_{max}.
Output: The estimates $\hat{\boldsymbol{\beta}}_{-1}$ and $\hat{\boldsymbol{\gamma}}$.

3. Calculation of conditional expectations

Implementing Algorithm 1 requires deriving formulas to calculate the conditional expectations of individuals' true statuses. These expressions are essential for the effective application of the EM algorithm in various testing scenarios. Common group testing methods include master pool testing, Dorfman testing, halving testing, and array testing. We have derived the conditional expectation formula of these methods under our methodological framework, which will facilitate our other calculations.

For master pool testing, samples are divided into J distinct groups, with each sample assigned to only one group, and each group undergoes a single test without subsequent testing. When the i-th individual is assigned to the j-th group, consider two cases for $w_i^{(t)}$:

While $Z_j = 0$,

$$w_i^{(t)} = \frac{P(\tilde{Y}_i = 1, Z_j = 0)}{P(Z_j = 0)} = \frac{P(Z_j = 0|\tilde{Y}_i = 1)P(\tilde{Y}_i = 1)}{P(Z_j = 0)}.$$

Due to

$$\begin{aligned} P(Z_j = 1) =& P(Z_j = 1|\tilde{Z}_j = 1)P(\tilde{Z}_j = 1) + P(Z_j = 1|\tilde{Z}_j = 0)P(\tilde{Z}_j = 0) \\ =& S_e[1 - \prod_{i\in\mathcal{P}_j}(1 - p_{iB}^{(t)})] + (1 - S_p)\prod_{i\in\mathcal{P}_j}(1 - p_{iB}^{(t)}) \\ =& S_e + (1 - S_e - S_p)\prod_{i\in\mathcal{P}_j}(1 - p_{iB}^{(t)}), \end{aligned}$$

let $\Delta_j = S_e + (1 - S_e - S_p)\prod_{i\in\mathcal{P}_j}(1 - p_{iB}^{(t)})$, where p_{iB} is an approximate result of p_i in (2.4). Therefore,

$$P(Z_j = 0) = 1 - [S_e + (1 - S_e - S_p)\prod_{i\in\mathcal{P}_j}(1 - p_{iB}^{(t)})] = 1 - \Delta_j.$$

Then,

$$w_i^{(t)} = \frac{(1 - S_e)\cdot p_{iB}^{(t)}}{1 - [S_e + (1 - S_e - S_p)\prod_{i\in\mathcal{P}_j}(1 - p_{iB}^{(t)})]} = \frac{(1 - S_e)\cdot p_{iB}^{(t)}}{(1 - \Delta_j)}.$$

While $Z_j = 1$,

$$\begin{aligned} w_i^{(t)} &= P(\tilde{Y}_i = 1|Z_j = 1) \\ &= \frac{p(Z_j = 1|\tilde{Y}_i = 1)P(\tilde{Y}_i = 1)}{P(Z_j = 1)} \\ &= \frac{S_e\cdot p_{iB}^{(t)}}{S_e + (1 - S_e - S_p)\prod_{i\in\mathcal{P}_j}(1 - p_{iB}^{(t)})} \\ &= \frac{S_e\cdot p_{iB}^{(t)}}{\Delta_j}. \end{aligned}$$

In conclusion,

$$w_i^{(t)} = \begin{cases} P(\tilde{Y}_i = 1|Z_j = 0) = (1 - S_e)\cdot p_{iB}^{(t)}/(1 - \Delta_j), & \text{if} \quad Z_j = 0, \\ P(\tilde{Y}_i = 1|Z_j = 1) = S_e\cdot p_{iB}^{(t)}/\Delta_j, & \text{if} \quad Z_j = 1. \end{cases}$$

We apply our method to four group testing algorithms: master pool testing, Dorfman testing, halving testing, and array testing. For other algorithms, detailed expressions can be found in Appendix C. Using these expressions, we apply the EM algorithm to estimate the model parameters.

4. Simulation study

In this section, we assess the performance of the proposed method using simulated datasets. The generation of covariates follows the approach described by Guo et al. [42]. Specifically, covariates $\mathbb{X} \in \mathbb{R}^{n \times q_n}$ are drawn from a truncated multivariate normal distribution. We first generate covariates from $N(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} \in \mathbb{R}^{q_n \times q_n}$ and $\Sigma_{ij} = 0.5^{|i-j|}$ for $1 \leq i, j \leq q_n$. These covariates are then truncated to the range $(-2, 2)$ to obtain $\mathbb{X}$. We consider logistic single-index models to describe $p_i = \Pr(\tilde{Y}_i = 1 \mid \boldsymbol{X}_i)$, with the function $g(\boldsymbol{X}_i^\top \boldsymbol{\beta})$ in the model (2.1) defined as follows,

Example 4.1. *We set* $n = 500$ *and* $\boldsymbol{\beta}^* = \left(\frac{3}{\sqrt{15.25}}, \frac{2.5}{\sqrt{15.25}}, 0, \ldots, 0\right)^\top$. *We consider two scenarios:* $q_n = 50$ *and* $q_n = 100$. *The model is described as follows:*

$$g(\boldsymbol{X}_i^\top \boldsymbol{\beta}^*) = \exp(\boldsymbol{X}_i^\top \boldsymbol{\beta}^*) - 7.$$

Under this setting, the disease prevalence is approximately 8.93%.

Example 4.2. *We set* $n = 1000$ *and* $\boldsymbol{\beta}^* = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, 0, \ldots, 0\right)^\top$. *We consider two scenarios:* $q_n = 100$ *and* $q_n = 500$. *The model is described as follows:*

$$g(\boldsymbol{X}_i^\top \boldsymbol{\beta}^*) = \boldsymbol{X}_i^\top \boldsymbol{\beta}^* (1 - \boldsymbol{X}_i^\top \boldsymbol{\beta}^*) + \exp(\boldsymbol{X}_i^\top \boldsymbol{\beta}^*) - 6.$$

In this example, the disease prevalence is approximately 11.41%.

Example 4.3. *We set* $q_n = 50$ *and* $\boldsymbol{\beta}^* = \left(\frac{9}{\sqrt{181}}, \frac{8}{\sqrt{181}}, \frac{6}{\sqrt{181}}, 0, \ldots, 0\right)^\top$. *We consider two scenarios:* $n = 500$ *and* $n = 1000$. *The model is described as follows:*

$$g(\boldsymbol{X}_i^\top \boldsymbol{\beta}^*) = \boldsymbol{X}_i^\top \boldsymbol{\beta}^* (1 - \boldsymbol{X}_i^\top \boldsymbol{\beta}^*) + 0.5 \cdot \sin\left(\frac{\pi \boldsymbol{X}_i^\top \boldsymbol{\beta}^*}{2}\right) - 6.$$

In this example, the disease prevalence is approximately 9.42%.

Example 4.4. *We set* $q_n = 100$ *and* $\boldsymbol{\beta}^* = (0.5, 0.5, 0.5, 0.5, 0, \ldots, 0)^\top$. *Two scenarios are considered:* $n = 750$ *and* $n = 1000$. *The model is described as follows:*

$$g(\boldsymbol{X}_i^\top \boldsymbol{\beta}^*) = \boldsymbol{X}_i^\top \boldsymbol{\beta}^* (1 - \boldsymbol{X}_i^\top \boldsymbol{\beta}^*) + \exp(\boldsymbol{X}_i^\top \boldsymbol{\beta}^*) + 0.1 \cdot \sin\left(\frac{\pi \boldsymbol{X}_i^\top \boldsymbol{\beta}^*}{2}\right) - 6.$$

In this scenario, the disease prevalence is approximately 10.32%.

In our simulation study, we employed four group testing algorithms: master pool testing (MPT), Dorfman testing (DT), halving testing (HT), and array testing (AT) to evaluate the model. For MPT, DT, and HT, the group size was set to 4, while in AT, individuals were arranged in a 4×4 array. Both sensitivity and specificity were fixed at $S_e = S_p = 0.98$. Based on the methodologies of Fan and Li [40] and Zhang [41], we set δ values of 3.7 and 2 for SCAD and MCP, respectively. Each scenario was simulated $B = 100$ times, where $\hat{\boldsymbol{\beta}}^{[b]}$ denotes the estimated $\boldsymbol{\beta}^*$ in the b-th simulation, with $b \in \{1, 2, \ldots, B\}$.

Following the approach of Guan et al. [45], we measured the estimation accuracy of $\hat{\beta}_j$ ($j = 1,2,3,4$) using the mean squared error (MSE), defined as

$$\text{MSE} = \frac{1}{B}\sum_{b=1}^{B}(\beta_j^* - \hat{\beta}_j^{[b]})^2,\ \ j = 1,2,3,4.$$

We utilized average mean squared error (AMSE) to assess the accuracy of $\hat{\boldsymbol{\beta}}$, consistent with methods employed by Wang and Yang [46]:

$$\text{AMSE} = \frac{1}{Bq_n}\sum_{b=1}^{B}\|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}^{[b]}\|_2^2.$$

Average mean absolute error (AMAE) was used to evaluate the estimation performance of $g(\cdot)$ and individual risk probabilities p_i [42]. The AMAE for $g(\cdot)$ is defined as

$$\text{AMAE}_g = \frac{1}{Bn}\sum_{b=1}^{B}\sum_{i=1}^{n}\left|g(\boldsymbol{X}_i^\top\boldsymbol{\beta}^*) - g(\boldsymbol{X}_i^\top\hat{\boldsymbol{\beta}}^{[b]})\right|,$$

while the AMAE for $\hat{p}_i^{[b]} = \frac{e^{g(X_i^\top\hat{\beta}^{[b]})}}{1+e^{g(X_i^\top\hat{\beta}^{[b]})}}$ is defined as

$$\text{AMAE}_p = \frac{1}{Bn}\sum_{b=1}^{B}\sum_{i=1}^{n}\left|p_i^* - \hat{p}_i^{[b]}\right|,$$

where $p_i^* = \frac{g(\boldsymbol{X}_i^\top\boldsymbol{\beta}^*)}{1+e^{g(X_i^\top\beta^*)}}$.

To evaluate variable selection performance, we employed true positive rate (TPR) and false positive rate (FPR). The FPR represents the proportion of false positives among identified predictors, while the TPR indicates the proportion of true positives among relevant predictors. Table 1 shows the results of variable selection.

Table 1. Four outcomes of variable selection.

Metric	Implication
True positive (TP)	Actual positive and predicted positive
False positive (FP)	Actual negative and predicted positive
False negative (FN)	Actual positive and predicted negative
True negative (TN)	Actual negative and predicted negative

TPR and FPR are defined as follows:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}},\quad \text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}.$$

The simulation results are summarized in Tables 2 to 5. As shown in the tables, the TPR was approximately 97%, with a very low FPR. The result shows that the probability that $\mathcal{M}^*$ is contained in $\hat{\mathcal{M}}$ is very close to 1. This demonstrates the notable performance of our model in variable selection.

The AMAE for $g(\cdot)$ and p_i was approximately 0.5 and 0.01, respectively. This shows that we have accurately captured the form of the unknown smooth function $g(\cdot)$ and are able to precisely predict the individual risk probability. The AMSE for the model parameters $\boldsymbol{\beta}$ was around 10^{-4}, while the AMSE for significant variables β_j was approximately 10^{-3}. This demonstrates the accuracy of our model in parameter estimation.

We set up two different sample sizes (n) or covariate scenarios (q_n) for each example. Results of Examples 4.1 and 4.2 suggest that our method maintains robust estimation performance as dimensionality increases in small sample scenarios. Furthermore, results of Examples 4.3 and 4.4 demonstrate that estimation accuracy improves with increased sample size. Figure 2 illustrates the estimation performance of $g(\cdot)$ and individual risk probabilities p_i, confirming our method's efficacy in estimating unknown functions and risk probabilities.

Table 2. Simulation results for Example 4.1.

						AMAE		AMSE	MSE	
Model	Setting	Test	Penalty	TPR	FPR	$g(\cdot)$	Prob	$\boldsymbol{\beta}$	β_1	β_2
Example 4.1 (n=500)	q_n=50	MPT	MCP	0.980	0.061	0.325	0.011	0.0003	0.0022	0.0015
		HT		0.985	0.003	0.413	0.007	0.0002	0.0068	0.0025
		DT		0.968	0.062	0.295	0.011	0.0003	0.0001	0.0004
		AT		0.987	0.035	0.388	0.009	0.0004	0.0019	0.0009
		MPT	SCAD	0.967	0.060	0.508	0.014	0.0003	0.0012	0.0021
		HT		0.988	0.001	0.479	0.008	0.0001	0.0021	0.0023
		DT		0.980	0.051	0.511	0.012	0.0003	0.0005	0.0004
		AT		0.974	0.063	0.432	0.009	0.0003	0.0035	0.0024
		MPT	LASSO	0.964	0.060	0.337	0.011	0.0003	0.0038	0.0051
		HT		0.986	0.003	0.436	0.006	0.0001	0.0003	0.0002
		DT		1.000	0.029	0.522	0.013	0.0001	0.0004	0.0003
		AT		0.981	0.034	0.320	0.007	0.0001	0.0006	0.0008
	q_n=100	MPT	MCP	0.985	0.010	0.511	0.009	0.0001	0.0004	0.0004
		HT		0.973	0.038	0.374	0.009	0.0002	0.0022	0.0033
		DT		0.986	0.023	0.338	0.010	0.0001	0.0004	0.0001
		AT		0.982	0.023	0.470	0.005	0.0001	0.0004	0.0005
		MPT	SCAD	0.987	0.031	0.265	0.013	0.0002	0.0002	0.0003
		HT		0.988	0.038	0.458	0.015	0.0005	0.0017	0.0001
		DT		0.978	0.051	0.451	0.011	0.0001	0.0008	0.0004
		AT		0.985	0.010	0.422	0.009	0.0001	0.0058	0.0047
		MPT	LASSO	0.987	0.026	0.478	0.010	0.0001	0.0008	0.0012
		HT		0.966	0.044	0.364	0.011	0.0003	0.0029	0.0052
		DT		0.984	0.031	0.503	0.012	0.0001	0.0001	0.0003
		AT		0.987	0.031	0.401	0.008	0.0001	0.0016	0.0014

Table 3. Simulation results for Example 4.2.

Model	Setting	Test	Penalty	TPR	FPR	AMAE $g(\cdot)$	AMAE Prob	AMSE $\boldsymbol{\beta}$	MSE β_1	MSE β_2	MSE β_3
Example 4.2 (n=1000)	q_n=100	MPT	MCP	0.980	0.001	0.569	0.011	0.0001	0.0006	0.0025	0.0073
		HT		0.974	0.001	0.626	0.012	0.0003	0.0059	0.0167	0.0054
		DT		0.971	0.027	0.601	0.012	0.0002	0.0035	0.0022	0.0019
		AT		0.986	0.010	0.582	0.011	0.0001	0.0059	0.0011	0.0032
		MPT	SCAD	0.970	0.019	0.551	0.010	$*$	0.0014	0.0006	0.0011
		HT		0.964	0.029	0.588	0.011	0.0001	0.0021	0.0041	0.0001
		DT		0.972	0.021	0.572	0.011	0.0001	0.0037	0.0002	0.0034
		AT		0.971	0.021	0.575	0.011	0.0001	0.0057	0.0005	0.0042
		MPT	LASSO	0.974	0.048	0.553	0.010	0.0001	0.0000	0.0002	0.0003
		HT		0.972	0.056	0.601	0.010	0.0001	0.0003	0.0001	0.0006
		DT		0.982	0.021	0.574	0.010	0.0001	0.0035	0.0001	0.0042
		AT		0.986	0.010	0.584	0.011	0.0001	0.0041	0.0002	0.0056
	q_n=500	MPT	MCP	0.964	0.011	0.562	0.013	0.0001	0.0005	0.0015	0.0042
		HT		0.972	0.010	0.670	0.018	0.0001	0.0056	0.0001	0.0115
		DT		0.987	0.011	0.567	0.012	$*$	0.0044	0.0003	0.0058
		AT		0.986	0.020	0.669	0.015	0.0001	0.0022	0.0108	0.0012
		MPT	SCAD	0.965	0.014	0.515	0.010	0.0001	0.0003	0.0055	0.0045
		HT		0.968	0.018	0.547	0.015	0.0001	0.0023	0.0112	0.0069
		DT		0.989	0.007	0.534	0.011	0.0001	0.0048	0.0001	0.0047
		AT		0.985	0.005	0.608	0.010	$*$	0.0042	0.0021	0.0007
		MPT	LASSO	0.978	0.006	0.536	0.012	0.0001	0.0013	0.0132	0.0104
		HT		0.970	0.002	0.644	0.015	0.0001	0.0000	0.0092	0.0126
		DT		0.987	0.005	0. 545	0.012	$*$	0.0015	0.0007	0.0019
		AT		0.981	0.002	0.526	0.012	$*$	0.0011	0.0093	0.0045

Symbol $*$ indicates value smaller than 0.0001.

Table 4. Simulation results for Example 4.3.

Model	Setting	Test	Penalty	TPR	FPR	AMAE $g(\cdot)$	AMAE Prob	AMSE $\boldsymbol{\beta}$	MSE β_1	MSE β_2	MSE β_3
Example 4.3 (q_n=50)	n=500	MPT	MCP	0.951	0.103	0.466	0.019	0.0003	0.0003	0.0009	0.0011
		HT		0.966	0.091	0.571	0.021	0.0005	0.0007	0.0036	0.0045
		DT		0.982	0.043	0.360	0.006	0.0001	0.0002	0.0001	0.0001
		AT		0.981	0.021	0.464	0.012	0.0001	0.0005	0.0009	0.0006
		MPT	SCAD	0.957	0.139	0.527	0.023	0.0005	0.0001	0.0031	0.0098
		HT		0.968	0.082	0.433	0.020	0.0004	0.0006	0.0001	0.0003
		DT		0.954	0.140	0.411	0.013	0.0002	0.0011	0.0018	0.0012
		AT		0.972	0.064	0.793	0.018	0.0002	0.0038	0.0021	0.0004
		MPT	LASSO	0.981	0.024	0.604	0.021	0.0003	0.0042	0.0014	0.0019
		HT		0.983	0.021	0.432	0.026	0.0001	0.0017	0.0005	0.0016
		DT		0.971	0.094	0.470	0.013	0.0002	0.0004	0.0014	0.0023
		AT		0.980	0.061	0.447	0.013	0.0002	0.0002	0.0004	0.0015
	n=1000	MPT	MCP	0.988	0.040	0.358	0.015	0.0002	0.0011	0.0024	0.0042
		HT		0.984	0.021	0.399	0.017	0.0006	0.0008	0.0009	0.0013
		DT		0.989	0.000	0.583	0.014	0.0001	0.0001	0.0019	0.0024
		AT		0.985	0.009	0.405	0.013	0.0001	0.0017	0.0041	0.0012
		MPT	SCAD	0.989	0.043	0.537	0.016	0.0002	0.0025	0.0004	0.0038
		HT		0.987	0.003	0.512	0.012	0.0001	0.0012	0.0032	0.0031
		DT		0.986	0.003	0.515	0.012	0.0001	0.0001	0.0002	0.0004
		AT		1.000	0.000	0.410	0.013	0.0001	0.0013	0.0022	0.0013
		MPT	LASSO	0.988	0.004	0.441	0.011	0.0002	0.0029	0.0012	0.0021
		HT		0.982	0.007	0.326	0.007	0.0001	0.0002	0.0004	0.0002
		DT		0.987	0.008	0.489	0.013	0.0001	0.0008	0.0001	0.0032
		AT		0.977	0.043	0.283	0.007	0.0001	0.0012	0.0024	0.0034

Table 5. Simulation results for Example 4.4.

Model	Setting	Test	Penalty	TPR	FPR	AMAE		AMSE	MSE			
						$g(\cdot)$	Prob	$\boldsymbol{\beta}$	β_1	β_2	β_3	β_4
Example 4.4 (q_n=100)	n=750	MPT	MCP	0.979	0.053	0.744	0.019	0.0004	0.0028	0.0015	0.0036	0.0076
		HT		0.959	0.100	0.970	0.027	0.0011	0.0024	0.0005	0.0022	0.0018
		DT		0.986	0.035	0.611	0.011	0.0001	0.0001	0.0025	0.0034	0.0016
		AT		0.984	0.043	0.789	0.013	0.0002	0.0012	0.0051	0.0026	0.0012
		MPT	SCAD	0.966	0.059	0.723	0.014	0.0003	0.0030	0.0078	0.0081	0.0001
		HT		0.978	0.069	0.576	0.014	0.0002	0.0004	0.0083	0.0052	0.0002
		DT		0.989	0.063	0.698	0.013	0.0003	0.0034	0.0155	0.0011	0.0051
		AT		0.981	0.052	0.671	0.022	0.0005	0.0078	0.0023	0.0085	0.0073
		MPT	LASSO	0.977	0.072	0.620	0.014	0.0003	0.0047	0.0041	0.0141	0.0002
		HT		0.964	0.069	0.680	0.015	0.0003	0.0018	0.0071	0.0073	0.0007
		DT		0.986	0.041	0.581	0.016	0.0003	0.0034	0.0090	0.0014	0.0005
		AT		0.984	0.065	0.679	0.016	0.0003	0.0001	0.0095	0.0065	0.0003
	n=1000	MPT	MCP	0.967	0.029	0.706	0.015	0.0002	0.0068	0.0097	0.0022	0.0015
		HT		0.986	0.001	0.818	0.012	0.0001	0.0035	0.0061	0.0007	0.0001
		DT		0.987	0.032	0.872	0.012	0.0002	0.0007	0.0074	0.0017	0.0016
		AT		0.988	0.037	0.800	0.027	0.0002	0.0013	0.0061	0.0002	0.0025
		MPT	SCAD	0.961	0.059	0.724	0.015	0.0002	0.0081	0.0087	0.0030	0.0006
		HT		0.974	0.010	0.779	0.013	0.0001	0.0036	0.0066	0.0012	0.0001
		DT		0.983	0.071	0.405	0 .010	0.0001	0.0013	0.0059	0.0008	0.0001
		AT		0.981	0.041	0.422	0.010	0.0001	0.0003	0.0009	0.0020	0.0011
		MPT	LASSO	0.977	0.029	0.819	0.017	0.0004	0.0057	0.0012	0.0083	0.0079
		HT		0.951	0.004	0.545	0.043	0.0001	0.0093	0.0004	0.0004	0.0025
		DT		0.985	0.021	0.408	0.009	0.0001	0.0002	0.0011	0.0026	0.0007
		AT		0.989	0.008	0.581	0.010	0.0001	0.0042	0.0003	0.0004	0.0008

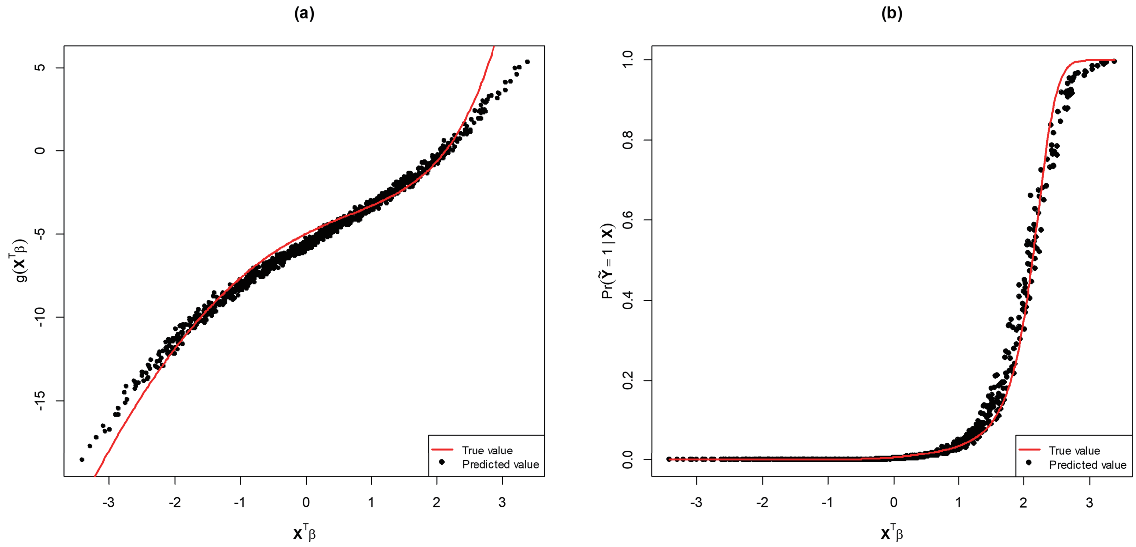

Figure 2. Estimation of unknown function (a) and risk probability (b) in Example 4.2, with $n = 1000$ and $q_n = 500$, using MPT and the SCAD penalty function.

Moreover, we aim to evaluate our method's performance under different group sizes. Using Example 4.4, we investigated group sizes of 2, 4, 6, and 8 with the Dorfman algorithm and LASSO penalty function. Results are presented in Table 6, reporting the means of $\hat{\beta}_j$ for $j = 1, 2, 3, 4$. The simulation results indicate that our method consistently delivers strong estimation performance across various group sizes. At the same time, we set up comparative experiments with different S_e and S_p, and the simulation results are shown in Tables 8 to 11 in Appendix A. As shown in these tables, our model maintains a certain level of stability, ensuring that $\mathcal{M}^*$ is still contained within $\hat{\mathcal{M}}$.

Table 6. Simulation results for different group size.

Model	Setting	Group Size	TPR	FPR	AMAE		MEAN			
					$g(\cdot)$	Prob	β_1	β_2	β_3	β_4
Example 4 (q_n=100)	n=750	2	0.970	0.015	0.611	0.011	0.452	0.465	0.478	0.460
		4	0.965	0.020	0.581	0.016	0.445	0.405	0.464	0.477
		6	0.986	0.041	0.627	0.009	0.519	0.497	0.487	0.495
		8	0.973	0.020	0.594	0.012	0.471	0.467	0.484	0.477
	n=1000	2	0.974	0.014	0.447	0.009	0.468	0.484	0.473	0.485
		4	0.964	0.018	0.408	0.009	0.489	0.468	0.450	0.474
		6	0.985	0.021	0.440	0.011	0.486	0.478	0.443	0.471
		8	0.974	0.010	0.466	0.009	0.494	0.494	0.447	0.437

5. Application to real data

In this section, we validate the effectiveness of our method using the diabetes dataset from the National Health and Nutrition Examination Survey (NHANES) conducted between 1999 and 2004.

NHANES is a probability-based cross-sectional survey representing the U.S. population, collecting demographic, health history, and behavioral information through household interviews. Participants were also invited to equip mobile examination centers for detailed physical, psychological, and laboratory assessments. The dataset is accessible at `https://wwwn.cdc.gov/Nchs/Nhanes/`.

The dataset comprises $n = 5515$ records and 17 variables, categorizing individuals as diabetic or non-diabetic. Covariates include age ($\boldsymbol{X}_1$), waist circumference ($\boldsymbol{X}_2$), BMI ($\boldsymbol{X}_3$), height ($\boldsymbol{X}_4$), weight ($\boldsymbol{X}_5$), smoking age ($\boldsymbol{X}_6$), alcohol use ($\boldsymbol{X}_7$), leg length ($\boldsymbol{X}_8$), total cholesterol ($\boldsymbol{X}_9$), hypertension ($\boldsymbol{X}_{10}$), education level ($\boldsymbol{X}_{11}$), household income ($\boldsymbol{X}_{12}$), family history ($\boldsymbol{X}_{13}$), physical activity ($\boldsymbol{X}_{14}$), gender ($\boldsymbol{X}_{15}$), and race ($\boldsymbol{X}_{16}$). Notably, nominal variables from $\boldsymbol{X}_{10}$ to $\boldsymbol{X}_{16}$ are transformed using one-hot encoding, resulting in $q_n = 47$ covariates per individual. The first nine variables are continuous, while the remainder are binary. A detailed explanation of the variables as well as the content of the questionnaire can be found at `https://wwwn.cdc.gov/Nchs/Nhanes/search/default.aspx`. For convenience, the nominal variables are explained in Table 12. in Appendix B.

For $i \in \{1, 2, \dots, n\}$, we define $\tilde{Y}_i = 1$ for diabetes and $\tilde{Y}_i = 0$ for non-diabetes. Individual covariate information is represented as $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \dots, X_{iq_n})^\top$. We construct the following single-index model for the probability of diabetes risk for the i-th individual:

$$\Pr(\tilde{Y}_i = 1|\boldsymbol{X}_i) = \frac{\exp[g(\boldsymbol{X}_i^\top \boldsymbol{\beta})]}{1 + \exp[g(\boldsymbol{X}_i^\top \boldsymbol{\beta})]},$$

where the smooth function $g(\cdot)$ is unknown, and our objective is to estimate the coefficients $\boldsymbol{\beta}$.

To verify the accuracy of our method, we compare the results with those obtained from two other methods. The first method is penalized logistic regression (PLR), which uses the true individual status, $\tilde{Y}_i$. This method is implemented using the R package "glmnet". The second method is the adaptive elastic net for group testing (aenetgt) data, as introduced by Gregory et al. [23]. This approach utilizes group testing data and employs a penalized Expectation-Maximization (EM) algorithm to fit an adaptive elastic net logistic regression model. The R package "aenetgt" is used for implementation. We generate Dorfman group testing data with a group size of 6, setting both sensitivity and specificity at $S_e = S_p = 0.98$.

To ensure comparability, we adhere to the standardization techniques referenced in Cui et al. [31]. First, we center the covariates to facilitate the comparison of relative effects across different explanatory variables. Second, we normalize the PLR and aenetgt coefficients by dividing them by their L_2-norm, as follows:

$$\hat{\boldsymbol{\beta}}_{PLR}^{norm} = \frac{\hat{\boldsymbol{\beta}}_{PLR}}{||\hat{\boldsymbol{\beta}}_{PLR}||_2}, \quad \hat{\boldsymbol{\beta}}_{aenet}^{norm} = \frac{\hat{\boldsymbol{\beta}}_{aenet}}{||\hat{\boldsymbol{\beta}}_{aenet}||_2},$$

thereby obtaining coefficients with unit norm. This enables a comparison of regression coefficients from PLR, aenetgt, and the single-index group testing model.

The estimated coefficients from the three models are summarized in Table 7, and the parameter estimation of our method is denoted as $\hat{\beta}_{our}$. In this study, the estimated coefficients for age, $\hat{\beta}_{PLR}^{norm}$ and $\hat{\beta}_{our}$, are 0.280 and 0.307, respectively, indicating that the risk of diabetes increases with age, consistent with the findings of Turi et al. [47]. However, the coefficient $\hat{\beta}_{aenet}^{norm}$ is close to zero. For waist circumference, the coefficients $\hat{\beta}_{PLR}^{norm}$, $\hat{\beta}_{our}$, and $\hat{\beta}_{aenet}^{norm}$ are 0.178, 0.194, and 0.271, respectively, suggesting a positive association between waist circumference and diabetes risk, which is supported by Bai et al. [48] and Snijder et al. [49]. In addition, all three methods also identified leg length [50], hypertension [51], race [52], family history [53], and sex [54] as variables associated with diabetes. These covariates are widely recognized as being related to diabetes in the biomedical field [55].

We found that the covariable physical activity is associated with diabetes, but the aenetgt method failed to identify this association. The results of a study by Yu et al. [55], which used the same dataset as ours, are consistent with this finding. In addition, we found that education level was also a covariable associated with diabetes ($\hat{\beta}_{PLR}^{norm}$ and $\hat{\beta}_{our}$ are -0.523 and -0.335). Evidence for this association can also be found in the study by Aldossari et al. [56], and in this dataset, the probability that these participants will not develop diabetes is 100%. We also identified that household income is associated with diabetes, which is consistent with the study by Yen et al. [57]. In this dataset, the probability of developing diabetes for those who refused to answer about their household income is 60%. Furthermore, our model yields results similar to those obtained by the PLR method, which uses individual observations ($\tilde{\mathcal{Y}}$), suggesting that our method is able to extract information from group observations.

Table 7. Estimated coefficients for the real data model.

Variable	$\hat{\beta}_{PLR}^{norm}$	$\hat{\beta}_{our}$	$\hat{\beta}_{aenet}^{norm}$	**Variable**	$\hat{\beta}_{PLR}^{norm}$	$\hat{\beta}_{our}$	$\hat{\beta}_{aenet}^{norm}$	**Variable**	$\hat{\beta}_{PLR}^{norm}$	$\hat{\beta}_{our}$	$\hat{\beta}_{aenet}^{norm}$
age	0.280	0.307	-0.085	**Family history**				**Household income**			
waist circumference	0.178	0.194	0.271	family history1	0.000	0.000	0.000	household income1	0.000	0.000	0.000
BMI	0.000	0.000	0.000	family history2	-0.492	-0.567	-0.466	household income2	0.024	0.000	0.000
height	0.000	0.000	0.000	family history9	0.000	0.000	0.000	household income3	0.000	0.000	0.000
weight	0.000	0.000	0.000	**Physical activity**				household income4	0.000	-0.069	0.000
smoking age	0.000	0.007	0.000	physical activity1	0.000	0.056	0.000	household income5	0.000	0.000	0.000
alcohol use	0.009	0.013	0.000	physical activity2	-0.086	-0.018	0.000	household income6	0.000	0.000	0.000
leg length	-0.048	-0.100	-0.043	physical activity3	-0.134	-0.039	0.000	household income7	0.000	0.000	0.000
total cholesterol	0.000	0.000	0.000	physical activity4	-0.088	0.000	0.000	household income8	0.001	0.065	0.000
Hypertension				physical activity9	0.000	0.000	0.000	household income9	0.000	0.000	0.000
hypertension1	0.000	0.000	0.000	**Sex**				household income10	0.000	0.000	0.000
hypertension2	-0.350	-0.372	-0.641	sex1	-0.010	0.000	0.000	household income11	0.000	0.000	0.000
Education				sex2	-0.237	-0.225	-0.424	household income12	0.000	0.000	0.000
education1	0.000	0.000	0.000	**race**				household income13	0.000	0.000	0.000
education2	0.000	0.000	0.000	race1	0.000	0.000	0.000	household income77	0.000	0.231	0.000
education3	0.000	0.000	0.000	race2	-0.019	-0.073	0.000	household income99	0.000	0.000	0.000
education4	0.000	0.000	0.000	race3	-0.399	-0.380	-0.330				
education5	-0.014	-0.052	0.000	race4	0.000	0.000	0.000				
education7	-0.523	-0.335	0.000	race5	0.000	0.124	0.000				

6. Conclusions and discussion

This study presents a group testing framework based on a logistic regression single-index model for disease screening in low-prevalence environments. By employing B-splines to estimate unknown functions and incorporating penalty functions, our approach achieves high flexibility in capturing the relationships between covariates and individual risk probabilities while accurately identifying important variables. To address potential computational challenges in individual disease status estimation, we implemented an iterative EM algorithm for model estimation. Our simulation experiments demonstrate the proposed method's performance in high-dimensional covariate contexts with limited sample sizes, while application to real data confirms its efficacy. Our framework offers a unified approach for various group testing methods, showcasing its practical application value.

Despite these promising outcomes, our study acknowledges several limitations. First, our model assumes that sensitivity and specificity of testing are independent of group size, which may not always hold in practical applications. Second, data quality and variations in the testing population can impact the model's applicability. Therefore, exploring how to integrate prior information to enhance model accuracy and practical value remains a critical research direction. Furthermore, the potential high dimensionality of individual covariates poses significant challenges, necessitating the development of models capable of handling ultra-high-dimensional data.

Future research could explore the following directions. Firstly, examining model performance under varying group testing configurations, such as changes in testing errors and group sizes, could yield valuable insights. Secondly, investigating methods to incorporate additional prior knowledge to improve estimation accuracy is a worthwhile endeavor. Additionally, considering computational efficiency, developing faster algorithms for processing large-scale datasets will be a key focus for future work.

Author contributions

Changfu Yang: Methodolog, formal analysis, writing-original draft; Wenxin Zhou: Methodology, formal analysis; Wenjun Xiong: Conceptualization, methodology, writing-original draft, funding acquisition; Junjian Zhang: Conceptualization, methodology, writing-review and editing, funding acquisition; Juan Ding: Conceptualization, formal analysis, writing-review and editin, funding acquisition. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant Nos. 12361055, 11801102), Guangxi Natural Science Foundation (2021GXNSFAA220054), and the Fundamental Research Funds for the Central Universities (B240201095).

Conflict of interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

1. R. Dorfman, The detection of defective members of large populations, *Ann. Math. Statist.*, **14** (1943), 436–440. http://dx.doi.org/10.1214/aoms/1177731363
2. S. Mallapaty, The mathematical strategy that could transform coronavirus testing, *Nature,* **583** (2020), 504–505. http://dx.doi.org/10.1038/d41586-020-02053-6
3. L. Mutesa, P. Ndishimye, Y. Butera, J. Souopgui, A. Uwineza, R. Rutayisire, et al., A pooled testing strategy for identifying SARS-CoV-2 at low prevalence, *Nature*, **589** (2021), 276–280. http://dx.doi.org/10.1038/s41586-020-2885-5
4. W. Chen, C. Tatsuoka, X. Lu, HiBGT: High-performance Bayesian group testing for COVID-19, In: *2022 IEEE 29th international conference on high performance computing, data, and analytics (HiPC)*, 2022, 176–185. https://doi.org/10.1109/HiPC56025.2022.00033
5. D. J. Westreich, M. G. Hudgens, S. A. Fiscus, C. D. Pilcher, Optimizing screening for acute human immunodeficiency virus infection with pooled nucleic acid amplification tests, *J. Clin. Microbiol.*, **46** (2008), 1785–1792. http://dx.doi.org/10.1128/jcm.00787-07
6. M. Krajden, D. Cook, A. Mak, K. Chu, N. Chahil, M. Steinberg, et al., Pooled nucleic acid testing increases the diagnostic yield of acute HIV infections in a high-risk population compared to 3rd and 4th generation HIV enzyme immunoassays, *J. Clin. Virol.*, **61** (2014), 132–137. http://dx.doi.org/10.1016/j.jcv.2014.06.024
7. J. L. Lewis, V. M. Lockary, S. Kobic, Cost savings and increased efficiency using a stratified specimen pooling strategy for Chlamydia trachomatis and Neisseria gonorrhoeae, *Sex. Transm. Dis.*, **39** (2012), 46–48. http://dx.doi.org/10.1097/OLQ.0b013e318231cd4a
8. T. T. Van, J. Miller, D. M. Warshauer, E. Reisdorf, D. Jernigan, R. Humes, et al., Pooling Nasopharyngeal/Throat swab specimens to increase testing capacity for influenza viruses by PCR, *J. Clin. Microbiol.*, **50** (2012), 891–896. http://dx.doi.org/10.1128/jcm.05631-11
9. P. Saá, M. Proctor, G. Foster, D. Krysztof, C. Winton, J. M. Linnen, et al., Investigational testing for Zika virus among U.S. blood donors, *N. Engl. J. Med.*, **378** (2018), 1778–1788. http://dx.doi.org/10.1056/NEJMoa1714977
10. J. M. Tebbs, C. S. McMahan, C. R. Bilder, Two-stage hierarchical group testing for multiple infections with application to the infertility prevention project, *Biometrics*, **69** (2013), 1064–1073. http://dx.doi.org/10.1111/biom.12080
11. C. S. McMahan, J. M. Tebbs, T. E. Hanson, C. R. Bilder, Bayesian regression for group testing data, *Biometrics*, **73** (2017), 1443–1452. http://dx.doi.org/10.1111/biom.12704
12. A. Yuan, J. Piao, J. Ning, J. Qin, Semiparametric isotonic regression modelling and estimation for group testing data, *Can. J. Stat.*, **49** (2021), 659–677. http://dx.doi.org/10.1002/cjs.11581
13. J. Ding, W. J. Xiong, Robust group testing for multiple traits with misclassification, *J. Appl. Stat.*, **42** (2015), 2115–2125. https://doi.org/10.1080/02664763.2015.1019841
14. S. C. Mokalled, C. S. McMahan, J. M. Tebbs, D. Andrew Brown, C. R. Bilder, Incorporating the dilution effect in group testing regression, *Stat. Med.*, **40** (2021), 2540–2555. http://dx.doi.org/10.1002/sim.8916

15. X. Z. Huang, M. S. S. Warasi, Maximum likelihood estimators in regression models for error-prone group testing data, *Scand. J. Stat.*, **44** (2017), 918–931. http://dx.doi.org/10.1111/sjos.12282

16. G. Haber, Y. Malinovsky, P. S. Albert, Sequential estimation in the group testing problem, *Sequential Anal.*, **37** (2018), 1–17. http://dx.doi.org/10.1080/07474946.2017.1394716

17. J. L. Horowitz, *Semiparametric and nonparametric methods in econometrics*, 1 Eds., New York: Springer, 2009. https://doi.org/10.1007/978-0-387-92870-8

18. P. Radchenko, High dimensional single index models, *J. Multivariate Anal.*, **139** (2015), 266–282. http://dx.doi.org/10.1016/j.jmva.2015.02.007

19. Z. C. Elmezouar, F. Alshahrani, I. M. Almanjahie, S. Bouzebda, Z. Kaid, A. Laksaci, Strong consistency rate in functional single index expectile model for spatial data, *AIMS Mathematics*, **9** (2024), 5550–5581. http://dx.doi.org/10.3934/math.2024269

20. Y. N. Chen, R. J. Samworth, Generalized additive and index models with shape constraints, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **78** (2016), 729–754. http://dx.doi.org/10.1111/rssb.12137

21. Z. Kereta, T. Klock, V. Naumova, Nonlinear generalization of the monotone single index model, *Inf. Inference*, **10** (2021), 987–1029. http://dx.doi.org/10.1093/imaiai/iaaa013

22. D. Wang, C. S. McMahan, C. M. Gallagher, A general regression framework for group testing data, which incorporates pool dilution effects, *Stat. Med.*, **34** (2015), 3606–3621. http://dx.doi.org/10.1002/sim.6578

23. K. B. Gregory, D. Wang, C. S. McMahan, Adaptive elastic net for group testing, *Biometrics*, **75** (2019), 13–23. http://dx.doi.org/10.1111/biom.12973

24. H. Ko, K. Kim, H. Sun, Multiple group testing procedures for analysis of high-dimensional genomic data, *Genomics Inform.*, **14** (2016), 187–195. http://dx.doi.org/10.5808/gi.2016.14.4.187

25. A. Delaigle, P. Hall, Nonparametric methods for group testing data, taking dilution into account, *Biometrika*, **102** (2015), 871–887. http://dx.doi.org/10.1093/biomet/asv049

26. S. Self, C. McMahan, S. Mokalled, Capturing the pool dilution effect in group testing regression: A Bayesian approach, *Stat. Med.*, **41** (2022), 4682–4696. http://dx.doi.org/10.1002/sim.9532

27. X. L. Zuo, J. Ding, J. J. Zhang, W. J. Xiong, Nonparametric additive regression for high-dimensional group testing data, *Mathematics*, **12** (2024), 686. http://dx.doi.org/10.3390/math12050686

28. R. J. Carroll, J. Fan, I. Gijbels, M. P. Wand, Generalized partially linear single-index models, *J. Amer. Statist. Assoc.*, **92** (1997), 477–489. https://doi.org/10.1080/01621459.1997.10474001

29. L. Zhu, L. Xue, Empirical likelihood confidence regions in a partially linear single-index model, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **68** (2006), 549–570. https://doi.org/10.1111/j.1467-9868.2006.00556.x

30. W. Lin, K. B. Kulasekera, Identifiability of single-index models and additive-index models, *Biometrika*, **94** (2007), 496–501. https://doi.org/10.1093/biomet/asm029

31. X. Cui, W. K. Härdle, L. X. Zhu, The EFM approach for single-index models, *Ann. Statist.*, **39** (2011), 1658–1688. http://dx.doi.org/10.1214/10-aos871

32. X. Guo, C. Z. Niu, Y. P. Yang, W. L. Xu, Empirical likelihood for single index model with missing covariates at random, *Statistics*, **49** (2015), 588–601. http://dx.doi.org/10.1080/02331888.2014.881826

33. W. Xiong, J. Ding, W. Zhang, A. Liu, Q. Li, Nested group testing procedure, *Commun. Math. Stat.*, **11** (2023), 663–693. http://dx.doi.org/10.1007/s40304-021-00269-0

34. J. X. Lin, D. W. Wang, Q. Zheng, Regression analysis and variable selection for two-stage multiple-infection group testing data, *Stat. Med.*, **38** (2019), 4519–4533. http://dx.doi.org/10.1002/sim.8311

35. D. Wang, C. S. McMahan, C. M. Gallagher, K. B. Kulasekera, Semiparametric group testing regression models, *Biometrika*, **101** (2014), 587–598. https://doi.org/10.1093/biomet/asu007

36. B. A. Zhang, C. R. Bilder, J. M. Tebbs, Group testing regression model estimation when case identification is a goal, *Biometrical J.*, **55** (2013), 173–189. http://dx.doi.org/10.1002/bimj.201200168

37. S. Vansteelandt, E. Goetghebeur, T. Verstraeten, Regression models for disease prevalence with diagnostic tests on pools of serum samples, *Biometrics*, **56** (2000), 1126–1133. http://dx.doi.org/10.1111/j.0006-341x.2000.01126.x

38. C. Boor, *A practical guide to splines*, New York: Springer, 1978.

39. R. Tibshirani, Regression shrinkage and selection via the Lasso, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **58** (1996), 267–288. http://dx.doi.org/10.1111/j.2517-6161.1996.tb02080.x

40. J. Fan, R. Z. Li, Variable selection via nonconcave penalized likelihood and its Oracle properties, *J. Amer. Statist. Assoc.*, **96** (2001), 1348–1360. http://dx.doi.org/10.1198/016214501753382273

41. C. H. Zhang, Nearly unbiased variable selection under minimax concave penalty, *Ann. Statist.*, **38** (2010), 894–942. http://dx.doi.org/10.1214/09-aos729

42. W. C. Guo, X. H. Zhou, S. J. Ma, Estimation of optimal individualized treatment rules using a covariate-specific treatment effect curve with high-dimensional covariates, *J. Amer. Statist. Assoc.*, **116** (2021), 309–321. http://dx.doi.org/10.1080/01621459.2020.1865167

43. J. Duchi, E. Hazan, Y. Singer, Adaptive subgradient methods for online learning and stochastic optimization, *J. Mach. Learn. Res.*, **12** (2011), 2121–2159.

44. P. Breheny, J. Huang, Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection, *Ann. Appl. Stat.*, **5** (2011), 232–253. http://dx.doi.org/10.1214/10-aoas388

45. S. J. Guan, M. T. Zhao, Y. H. Cui, Variable selection for single-index varying-coefficients models with applications to synergistic G x E interactions, *Electron. J. Statist.*, **17** (2023), 823–857. http://dx.doi.org/10.1214/23-ejs2117

46. L. Wang, L. J. Yang, Spline estimation of single-index models, *Statist. Sinica*, **19** (2009), 765–783.

47. K. N. Turi, D. M. Buchner, D. S. Grigsby-Toussaint, Predicting risk of type 2 diabetes by using data on easy-to-measure risk factors, *Prev. Chronic. Dis.*, **14** (2017), 160244. http://dx.doi.org/10.5888/pcd14.160244

48. K. Bai, X. Chen, R. Song, W. Shi, S. Shi, Association of body mass index and waist circumference with type 2 diabetes mellitus in older adults: a cross-sectional study, *BMC Geriatr.*, **22** (2022), 489. http://dx.doi.org/10.1186/s12877-022-03145-w

49. M. B. Snijder, P. Z. Zimmet, M. Visser, J. M. Dekker, J. C. Seidell, J. E. Shaw, Independent and opposite associations of waist and hip circumferences with diabetes, hypertension and dyslipidemia: the AusDiab Study, *Int. J. Obes.*, **28** (2004), 402–409. http://dx.doi.org/10.1038/sj.ijo.0802567

50. C. Wittenbecher, O. Kuxhaus, H. Boeing, N. Stefan, M. B. Schulze, Associations of short stature and components of height with incidence of type 2 diabetes: Mediating effects of cardiometabolic risk factors, *Diabetologia*, **62** (2019), 2211–2221. https://doi.org/10.1007/s00125-019-04978-8

51. G. Colussi, A. Da Porto, A. Cavarape, Hypertension and type 2 diabetes: Lights and shadows about causality, *J. Hum. Hypertens.*, **34** (2020), 91–93. http://dx.doi.org/10.1038/s41371-019-0268-x

52. S. E. Richards, C. Wijeweera, A. Wijeweera, Lifestyle and socioeconomic determinants of diabetes: Evidence from country-level data, *PLoS ONE*, **17** (2022), e0270476. https://doi.org/10.1371/journal.pone.0270476

53. J. Su, J. Y. Zhou, R. Tao, Y. N. Wan, Y. Qin, Y. Lu, et al., Association between family history of diabetes and incident diabetes of adults: A prospective study, *Chin. J. Prev. Med.*, **54** (2020), 828–833. https://doi.org/10.3760/cma.j.cn112150-20200212-00091

54. S. Clotet-Freixas, O. Zaslaver, M. Kotlyar, C. Pastrello, A. T Quaile, C. M. McEvoy, et al., Sex differences in kidney metabolism may reflect sex-dependent outcomes in human diabetic kidney disease, *Sci. Transl. Med.*, **16** (2024), eabm2090. https://doi.org/10.1126/scitranslmed.abm2090

55. M. Yu, T. Liu, R. Valdez, M. Gwinn, M. J. Khoury, Application of support vector machine modeling for prediction of common diseases: The case of diabetes and pre-diabetes, *BMC Med. Inform. Decis. Mak.*, **10** (2010), 16. https://doi.org/10.1186/1472-6947-10-16

56. K. K. Aldossari, A. Aldiab, J. M. Al-Zahrani, S. H. Al-Ghamdi, M. Abdelrazik, M. A. Batais, et al., Prevalence of prediabetes, diabetes, and its associated risk factors among males in Saudi Arabia: A population-based survey, *J. Diabetes Res.*, **2018** (2018), 2194604. https://doi.org/10.1155/2018/2194604

57. F. S. Yen, J. C. C. Wei, J. S. Liu, C. M. Hwu, Parental income level and risk of developing type 2 diabetes in youth, *JAMA Netw. Open.*, **6** (2023), e2345812. https://doi.org/10.1001/jamanetworkopen.2023.45812

A. Appendix-A: Simulation results with different sensitivity and specificity settings

In this part, we tested the performance of four examples at different sensitivity and specificity, using the Dofman algorithm and the LASSO penalty function. The simulation results are shown in Tables 8 to 11.

Table 8. Example 4.1: Simulation results with different sensitivity and specificity settings.

					AMAE		AMSE	MSE	
Model	Setting	(S_e, S_p)	TPR	FPR	$g(\cdot)$	Prob	$\boldsymbol{\beta}$	β_1	β_2
Example 4.1 ($q_n = 50$)	n=500	(0.98,0.98)	1.000	0.029	0.522	0.013	0.0001	0.0004	0.0003
		(0.95,0.95)	0.987	0.020	0.474	0.011	0.0001	0.0003	0.0003
		(0.90,0.90)	0.982	0.036	0.532	0.011	0.0001	0.0006	0.0007
		(0.85,0.85)	0.984	0.040	0.578	0.016	0.0003	0.0001	0.0002

Table 9. Example 4.2: Simulation results with different sensitivity and specificity settings.

					AMAE		AMSE		MSE	
Model	Setting	(S_e, S_p)	TPR	FPR	$g(\cdot)$	Prob	$\boldsymbol{\beta}$	β_1	β_2	β_3
Example 4.2 ($q_n = 100$)	n=1000	(0.98,0.98)	0.982	0.021	0.574	0.010	0.0001	0.0035	0.0001	0.0042
		(0.95,0.95)	0.975	0.030	0.612	0.011	0.0001	0.0047	0.0001	0.0069
		(0.90,0.90)	0.978	0.020	0.556	0.012	0.0001	0.0023	0.0002	0.0049
		(0.85,0.85)	0.965	0.020	0.717	0.016	0.0004	0.0158	0.0002	0.0212

Table 10. Example 4.3: Simulation results with different sensitivity and specificity settings.

					AMAE		AMSE		MSE	
Model	Setting	(S_e, S_p)	TPR	FPR	$g(\cdot)$	Prob	$\boldsymbol{\beta}$	β_1	β_2	β_3
Example 4.3 ($q_n = 50$)	n=1000	(0.98,0.98)	0.987	0.008	0.489	0.013	0.0001	0.0008	0.0001	0.0032
		(0.95,0.95)	0.971	0.064	0.404	0.011	0.0003	0.0005	0.0033	0.0085
		(0.90,0.90)	0.963	0.048	0.465	0.011	0.0001	0.0002	0.0012	0.0055
		(0.85,0.85)	0.966	0.018	0.377	0.015	0.0004	0.0016	0.0023	0.0007

Table 11. Example 4.4: Simulation results with different sensitivity and specificity settings.

					AMAE		AMSE		MSE		
Model	Setting	(S_e, S_p)	TPR	FPR	$g(\cdot)$	Prob	$\boldsymbol{\beta}$	β_1	β_2	β_3	β_4
Example 4.4 ($q_n = 100$)	n=750	(0.98,0.98)	0.986	0.041	0.581	0.016	0.0003	0.0034	0.0090	0.0014	0.0005
		(0.95,0.95)	0.981	0.026	0.534	0.018	0.0001	0.0016	0.0045	0.0018	0.0005
		(0.90,0.90)	0.974	0.018	0.546	0.016	0.0002	0.0004	0.0024	0.0014	0.0028
		(0.85,0.85)	0.976	0.024	0.539	0.011	0.0002	0.0047	0.0085	0.0004	0.0039

B. Appendix-B: Meaning of the nominal variable

Table 12. Meaning of the nominal variable.

Variable	Implication	Variable	Implication
Hypertension circumstance		**Family history of diabetes**	
hypertension1	Have a history of hypertension	family history1	Blood relatives with diabetes
hypertension2	No history of hypertension	family history2	Blood relatives do not have diabetes
Education level		family history9	Not known if any blood relatives have diabetes
education1	Less Than 9th Grade	**Physical activity**	
education2	9 - 11th Grade (Includes 12th grade with no diploma)	physical activity1	Sit during the day and do not walk about very much
education3	High School Grad/GED or Equivalent	physical activity2	Stand or walk about a lot during the day, but do not have to carry or lift things very often
education4	Some College or AA degree	physical activity3	Lift light load or has to climb stairs or hills often
education5	College Graduate or above	physical activity4	Do heavy work or carry heavy loads
education7	Refuse to answer about the level of education	physical activity9	Don't know physical activity level
Household income		**Sex**	
household income1	0 to 4,999 $	sex1	Male
household income2	5,000 to 9,999 $	sex2	Female
household income3	10,000 to 14,999 $	**Race/Ethnicity**	
household income4	15,000 to 19,999 $	race1	Mexican American
household income5	20,000 to 24,999 $	race2	Other Hispanic
household income6	25,000 to 34,999 $	race3	Non - Hispanic White
household income7	35,000 to 44,999 $	race4	Non - Hispanic Black
household income8	45,000 to 54,999 $	race5	Other Race - Including Multi - Racial
household income9	55,000 to 64,999 $		
household income10	65,000 to 74,999 $		
household income11	75,000 and Over $		
household income12	Over 20,000 $		
household income13	Under 20,000 $		
household income77	Refusal to answer about household income		
household income99	Don't know household income		

C. Appendix-C: Calculation of conditional expectations

In this part, we derive the conditional expectation formulas for Dorfman testing, halving testing, and array testing within the framework of our method. Before proceeding, it is necessary to clarify some notations. Let $\mathcal{P}_j \setminus \{i\}$ represent the set of individuals in $\mathcal{P}_j$ excluding the i-th individual, and $|\mathcal{P}_j|$ denotes the number of individuals in $\mathcal{P}_j$. Let Y_i represent the test result of the i-th individual and $\mathcal{Y}_{\mathcal{P}_{j,l}} = \{Y_i, i \in \mathcal{P}_{j,l}\}$ represent the set of testing results of individuals in $\mathcal{P}_{j,l}$.

C.1. Dorfman testing

If the initial group testing result is negative, no re-testing is performed. However, if $Z_{j,1} = 1$, each individual in the group needs to undergo a separate re-testing.

1) When $Z_{j,1} = 0$, the result is the same as the master poor testing:

$$w_{i,0}^{(t)} = \frac{P(\tilde{Y}_i = 1, Z_{j,1} = 0)}{P(Z_{j,1} = 0)} = \frac{(1 - S_e) \cdot p_{iB}^{(t)}}{1 - [S_e + (1 - S_e - S_p) \prod_{i \in \mathcal{P}_j} (1 - p_{iB}^{(t)})]}.$$

2) When $Z_{j,1} = 1$, each individual in the group must undergo a separate re-test. In total, the group has undergone $|\mathcal{P}_j| + 1$ tests.

$$w_{i,1}^{(t)} = \frac{P(\tilde{Y}_i = 1, Z_{j,1}, \mathcal{Y}_{\mathcal{P}_j})}{P(Z_{j,1}, \mathcal{Y}_{\mathcal{P}_j})} = \frac{P(\tilde{Y}_i = 1) P(Z_{j,1}, \mathcal{Y}_{\mathcal{P}_j} | \tilde{Y}_i = 1)}{P(Z_{j,1}, \mathcal{Y}_{\mathcal{P}_j})}.$$

The denominator is

$$\begin{aligned}
P(Z_{j,1}, \mathcal{Y}_{\mathcal{P}_j}) &= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j}} P(Z_{j,1}, \mathcal{Y}_{\mathcal{P}_j} | \tilde{\mathcal{Y}}_{\mathcal{P}_j}) P(\tilde{\mathcal{Y}}_{\mathcal{P}_j}) \\
&= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j}} P(Z_{j,1} | \tilde{Z}_{j,1}) \prod_{i \in \mathcal{P}_j} P(Y_i | \tilde{Y}_i) P(\tilde{Y}_i) \\
&= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j}} [S_e^{\tilde{Z}_{j,1}} (1 - S_p)^{1 - \tilde{Z}_{j,1}}] \prod_{i \in \mathcal{P}_j} [S_e^{Y_i} (1 - S_e)^{(1 - Y_i)}]^{\tilde{Y}_i} \\
&\quad \times [(1 - S_p)^{Y_i} S_p^{(1 - Y_i)}]^{(1 - \tilde{Y}_i)} [p_{iB}^{(t)}]^{\tilde{Y}_i} [1 - p_{iB}^{(t)}]^{1 - \tilde{Y}_i} \\
&= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j}} [S_e^{\tilde{Z}_{j,1}} (1 - S_p)^{1 - \tilde{Z}_{j,1}}] \prod_{i \in \mathcal{P}_j} [S_e^{Y_i} (1 - S_e)^{(1 - Y_i)} p_{iB}^{(t)}]^{\tilde{Y}_i} \\
&\quad \times [(1 - S_p)^{Y_i} S_p^{(1 - Y_i)} (1 - p_{iB}^{(t)})]^{(1 - \tilde{Y}_i)}.
\end{aligned}$$

Thus, the numerator is

$$\begin{aligned}
P(\tilde{Y}_i = 1, Z_{j,1}, \mathcal{Y}_{\mathcal{P}_j}) &= P(Z_{j,1}, \mathcal{Y}_{\mathcal{P}_j} | \tilde{Y}_i = 1) P(\tilde{Y}_i = 1) \\
&= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j \setminus i}} P(Z_{j,1}, \mathcal{Y}_{\mathcal{P}_j} | \tilde{Y}_i = 1, \tilde{\mathcal{Y}}_{\mathcal{P}_j \setminus i}) P(\tilde{\mathcal{Y}}_{\mathcal{P}_j \setminus i}) P(\tilde{Y}_i = 1) \\
&= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j \setminus i}} P(Z_{j,1} | \tilde{Z}_{j,1} = 1) P(Y_i | \tilde{Y}_i = 1) \prod_{i \in \mathcal{P}_j \setminus \{i\}} P(Y_i | \tilde{Y}_i) P(\tilde{Y}_i) P(\tilde{Y}_i = 1)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j \setminus i}} S_e S_e^{Y_i}(1-S_e)^{(1-Y_i)} \prod_{i \in \mathcal{P}_j \setminus \{i\}} [S_e^{Y_i}(1-S_e)^{(1-Y_i)}]^{\tilde{Y}_i} \\
&\quad \times [(1-S_p)^{Y_i} S_p^{(1-Y_i)}]^{(1-\tilde{Y}_i)} [p_{iB}^{(t)}]^{\tilde{Y}_i} [1-p_{iB}^{(t)}]^{1-\tilde{Y}_i} p_{iB}^{(t)} \\
&= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j \setminus i}} S_e^{1+Y_i}(1-S_e)^{(1-Y_i)} p_{iB}^{(t)} \times \prod_{i \in \mathcal{P}_j \setminus \{i\}} [S_e^{Y_i}(1-S_e)^{(1-Y_i)} p_{iB}^{(t)}]^{\tilde{Y}_i} \\
&\quad \times [(1-S_p)^{Y_i} S_p^{(1-Y_i)}(1-p_{iB}^{(t)})]^{(1-\tilde{Y}_i)}.
\end{aligned}
$$

Therefore, the final expression is

$$
w_i^{(t)} = Z_{j,1} w_{i,1}^{(t)} + (1 - Z_{j,1}) w_{i,0}^{(t)}.
$$

C.2. Halving testing

Assume that the maximum number of partitions required during testing is two. Let the test result of the first testing be $Z_{j,1}$. At this time, the set of all unpartitioned individuals is $\mathcal{P}_{j,1}$, which contains $|\mathcal{P}_j|$ individuals. After the first partition, the partitioning method is to divide into two equal parts, with the two subsets of individuals being $\mathcal{P}_{j,2}$ and $\mathcal{P}_{j,3}$, respectively. The responses of the second testing are $Z_{j,2}$ and $Z_{j,3}$. There are five types of testing results in halving testing.

1) When $Z_{j,1} = 0$:

Only one testing is performed, and the process is the same as master pool testing. Since the result of one testing is negative, no further partitioning and testing are performed. At this time,

$$
\begin{aligned}
w_i^{(t)} = P(\tilde{Y}_i = 1 | Z_{j,1} = 0) &= \frac{P(\tilde{Y}_i = 1, Z_{j,1} = 0)}{P(Z_{j,1} = 0)} \\
&= \frac{P(\tilde{Y}_i = 1) P(Z_{j,1} = 0 | \tilde{Y}_i = 1)}{P(Z_{j,1} = 0)} \\
&= \frac{p_{iB}^{(t)}(1 - S_e)}{1 - [S_e + (1 - S_e - S_p) \prod_{i \in \mathcal{P}_j} (1 - p_{iB}^{(t)})]}.
\end{aligned}
$$

2) When $Z_{j,1} = 1, Z_{j,2} = 0, Z_{j,3} = 0$:

That is, the result of the first testing is $Z_{j,1} = 1$. Subsequently, the first partition is performed, dividing into two equal parts $\mathcal{P}_{j,2}$ and $\mathcal{P}_{j,3}$. Then, testings are performed on the two sets respectively, with the testing results being $Z_{j,2} = Z_{j,3} = 0$. At this time,

$$
\begin{aligned}
w_i^{(t)} &= P(\tilde{Y}_i = 1 | Z_{j,1} = 1, Z_{j,2} = 0, Z_{j,3} = 0) \\
&= \frac{P(Z_{j,1} = 1, Z_{j,2} = 0, Z_{j,3} = 0 | \tilde{Y}_i = 1) P(\tilde{Y}_i = 1)}{P(Z_{j,1} = 1, Z_{j,2} = 0, Z_{j,3} = 0)}.
\end{aligned}
$$

The denominator is

$$
\begin{aligned}
&P(Z_{j,1} = 1, Z_{j,2} = 0, Z_{j,3} = 0) \\
&= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_{j,1}}} P(Z_{j,1} = 1, Z_{j,2} = 0, Z_{j,3} = 0 | \tilde{\mathcal{Y}}_{\mathcal{P}_{j,1}}) P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}}) P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}})
\end{aligned}
$$

$$= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_{j,1}}} P(Z_{j,1} = 1|\tilde{\mathcal{Y}}_{\mathcal{P}_{j,1}}) P(Z_{j,2} = 0|\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}}) P(Z_{j,3} = 0|\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}}) P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}}) P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}})$$

$$= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_{j,1}}} \left[S_e^{\tilde{Z}_{j,1}} (1 - S_p)^{1-\tilde{Z}_{j,1}} \right] \left[(1 - S_e)^{\tilde{Z}_{j,2}} S_p^{1-\tilde{Z}_{j,2}} \right] \prod_{i \in \mathcal{P}_{j,2}} \left[p_{iB}^{(t)} \right]^{\tilde{Y}_i} \left[1 - p_{iB}^{(t)} \right]^{1-\tilde{Y}_i}$$

$$\times \left[(1 - S_e)^{\tilde{Z}_{j,3}} S_p^{1-\tilde{Z}_{j,3}} \right] \prod_{i \in \mathcal{P}_{j,3}} \left[p_{iB}^{(t)} \right]^{\tilde{Y}_i} \left[1 - p_{iB}^{(t)} \right]^{1-\tilde{Y}_i}$$

$$= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_{j,1}}} \left[S_e^{\tilde{Z}_{j,1}} (1 - S_p)^{1-\tilde{Z}_{j,1}} \right] \prod_{u=2}^{3} (1 - S_e)^{\tilde{Z}_{j,u}} S_p^{1-\tilde{Z}_{j,u}} \prod_{i \in \mathcal{P}_j} \left[p_{iB}^{(t)} \right]^{\tilde{Y}_i} \left[1 - p_{iB}^{(t)} \right]^{1-\tilde{Y}_i}.$$

Since the placement of the i-th individual in the sets $\mathcal{P}_{j,2}$ and $\mathcal{P}_{j,3}$ is symmetric, assume that i-th individual is placed in the set $\mathcal{P}_{j,2}$. Then, the numerator is

$$\begin{aligned}
& P(Z_{j,1} = 1, Z_{j,2} = 0, Z_{j,3} = 0, \tilde{Y}_i = 1) \\
=& P(Z_{j,1} = 1, Z_{j,2} = 0, Z_{j,3} = 0|\tilde{Y}_i = 1) P(\tilde{Y}_i = 1) \\
=& \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j \setminus i}} P(Z_{j,1} = 1, Z_{j,2} = 0, Z_{j,3} = 0|\tilde{Y}_i = 1, \tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}}) \times P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2} \setminus i}) P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}}) P(\tilde{Y}_i = 1) \\
=& \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j \setminus i}} P(Z_{j,1} = 1|\tilde{Z}_{j,1} = 1) P(Z_{j,2} = 0|\tilde{Z}_{j,2} = 1) P(Z_{j,3} = 0|\tilde{Z}_{j,3}) \times P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2} \setminus i}) P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}}) P(\tilde{Y}_i = 1) \\
=& \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j \setminus i}} S_e (1 - S_e) \prod_{i \in \mathcal{P}_{j,2} \setminus \{i\}} \left[p_{iB}^{(t)} \right]^{\tilde{Y}_i} \left[1 - p_{iB}^{(t)} \right]^{1-\tilde{Y}_i} (1 - S_e)^{\tilde{Z}_{j,3}} S_p^{1-\tilde{Z}_{j,3}} \prod_{i \in \mathcal{P}_{j,3}} \left[p_{iB}^{(t)} \right]^{\tilde{Y}_i} \left[1 - p_{iB}^{(t)} \right]^{1-\tilde{Y}_i} p_{iB}^{(t)} \\
=& \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j \setminus i}} S_e (1 - S_e)^{1+\tilde{Z}_{j,3}} S_p^{1-\tilde{Z}_{j,3}} p_{iB}^{(t)} \prod_{i \in \mathcal{P}_j \setminus \{i\}} \left[p_{iB}^{(t)} \right]^{\tilde{Y}_i} \left[1 - p_{iB}^{(t)} \right]^{1-\tilde{Y}_i}.
\end{aligned}$$

3) When $Z_{j,1} = 1, Z_{j,2} = 0, Z_{j,3} = 1$:

At this time, the second partitions are performed. The first partition divides all individuals into two sets, $\mathcal{P}_{j,2}$ and $\mathcal{P}_{j,3}$, with testing results $Z_{j,2} = 0$ and $Z_{j,3} = 1$, respectively. Individual testings are performed separately on the individuals in $\mathcal{P}_{j,3}$, and the set of testing results is $\mathcal{Y}_{\mathcal{P}_{j,3}}$. At this time,

$$\begin{aligned}
w_i^{(t)} &= P(\tilde{Y}_i = 1|Z_{j,1} = 1, Z_{j,2} = 0, Z_{j,3} = 1, \mathcal{Y}_{\mathcal{P}_{j,3}}) \\
&= \frac{P(Z_{j,1} = 1, Z_{j,2} = 0, Z_{j,3} = 1, \mathcal{Y}_{\mathcal{P}_{j,3}}|\tilde{Y}_i = 1) P(\tilde{Y}_i = 1)}{P(Z_{j,1} = 1, Z_{j,2} = 0, Z_{j,3} = 1, \mathcal{Y}_{\mathcal{P}_{j,3}})}.
\end{aligned}$$

The denominator is

$$\begin{aligned}
& P(Z_{j,1} = 1, Z_{j,2} = 0, Z_{j,3} = 1, \mathcal{Y}_{\mathcal{P}_{j,3}}) \\
=& \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j}} P(Z_{j,1} = 1, Z_{j,2} = 0, Z_{j,3} = 1, \mathcal{Y}_{\mathcal{P}_{j,3}}|\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}}, \tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}}) P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}}) P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}}) \\
=& \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j}} P(Z_{j,1} = 1|\tilde{\mathcal{Y}}_{\mathcal{P}_j}) P(Z_{j,2} = 0|\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}}) P(Z_{j,3} = 1|\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}})
\end{aligned}$$

$$
\begin{aligned}
&\times P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}})P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}})P(\mathcal{Y}_{\mathcal{P}_{j,3}}|\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}})\\
&=\sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j}} P(Z_{j,1}=1|\tilde{Z}_{j,1})P(Z_{j,2}=0|\tilde{Z}_{j,2})P(Z_{j,3}=1|\tilde{Z}_{j,3})\\
&\times\prod_{i\in\mathcal{P}_{j,3}} P(Y_i|\tilde{Y}_i)P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}})P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}})\\
&=\sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j}}\Big[S_e^{\tilde{Z}_{j,1}}(1-S_p)^{1-\tilde{Z}_{j,1}}\Big]\Big[(1-S_e)^{\tilde{Z}_{j,2}}S_p^{1-\tilde{Z}_{j,2}}\Big]\Big[S_e^{\tilde{Z}_{j,3}}(1-S_p)^{1-\tilde{Z}_{j,3}}\Big]\\
&\times\prod_{i\in\mathcal{P}_{j,2}}\Big[p_{iB}^{(t)}\Big]^{\tilde{Y}_i}\Big[1-p_{iB}^{(t)}\Big]^{1-\tilde{Y}_i}\prod_{i\in\mathcal{P}_{j,3}}\Big[p_{iB}^{(t)}\Big]^{\tilde{Y}_i}\Big[1-p_{iB}^{(t)}\Big]^{1-\tilde{Y}_i}\\
&\times\prod_{i\in\mathcal{P}_{j,3}}\Big[S_e^{Y_i}(1-S_e)^{1-Y_i}\Big]^{\tilde{Y}_i}\Big[(1-S_p)^{Y_i}S_p^{1-Y_i}\Big]^{1-\tilde{Y}_i}\\
&=\sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j}}\Big[S_e^{\tilde{Z}_{j,1}+\tilde{Z}_{j,3}}(1-S_p)^{2-\tilde{Z}_{j,1}-\tilde{Z}_{j,3}}\Big]\Big[(1-S_e)^{\tilde{Z}_{j,2}}S_p^{1-\tilde{Z}_{j,2}}\Big]\\
&\times\prod_{i\in\mathcal{P}_j}\Big[p_{iB}^{(t)}\Big]^{\tilde{Y}_i}\Big[1-p_{iB}^{(t)}\Big]^{1-\tilde{Y}_i}\prod_{i\in\mathcal{P}_{j,3}}\Big[S_e^{Y_i}(1-S_e)^{1-Y_i}\Big]^{\tilde{Y}_i}\Big[(1-S_p)^{Y_i}S_p^{1-Y_i}\Big]^{1-\tilde{Y}_i}.
\end{aligned}
$$

Since an i-th individual may belong to either set $\mathcal{P}_{j,2}$ or $\mathcal{P}_{j,3}$, the numerator is discussed accordingly.
(a) Assume that i-th individual belongs to set $\mathcal{P}_{j,2}$. Then, the numerator is

$$
\begin{aligned}
&P(\tilde{Y}_i=1,Z_{j,1}=1,Z_{j,2}=0,Z_{j,3}=1,\mathcal{Y}_{\mathcal{P}_{j,3}})\\
&=\sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j\setminus i}} P(Z_{j,1}=1,Z_{j,2}=0,Z_{j,3}=1,\mathcal{Y}_{\mathcal{P}_{j,3}}|\tilde{Y}_i=1,\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}\setminus i},\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}})\\
&\times P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}\setminus i},\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}})P(\tilde{Y}_i=1)\\
&=\sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j\setminus i}} P(Z_{j,1}=1|\tilde{Z}_{j,1}=1)P(Z_{j,2}=0|\tilde{Z}_{j,2}=1)P(Z_{j,3}=1|\tilde{Z}_{j,3})\\
&\times P(\mathcal{Y}_{\mathcal{P}_{j,3}}|\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}})P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}\setminus i})P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}})P(\tilde{Y}_i=1)\\
&=\sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j\setminus i}} S_e(1-S_e)S_e^{\tilde{Z}_{j,3}}(1-S_p)^{1-\tilde{Z}_{j,3}}\prod_{i\in\mathcal{P}_{j,2}\setminus\{i\}}\Big[p_{iB}^{(t)}\Big]^{\tilde{Y}_i}\Big[1-p_{iB}^{(t)}\Big]^{1-\tilde{Y}_i}\\
&\times\prod_{i\in\mathcal{P}_{j,3}}\Big[p_{iB}^{(t)}\Big]^{\tilde{Y}_i}\Big[1-p_{iB}^{(t)}\Big]^{1-\tilde{Y}_i}p_{iB}^{(t)}\\
&\times\prod_{i\in\mathcal{P}_{j,3}}\Big[S_e^{Y_i}(1-S_e)^{1-Y_i}\Big]^{\tilde{Y}_i}\Big[(1-S_p)^{Y_i}S_p^{1-Y_i}\Big]^{1-\tilde{Y}_i}.
\end{aligned}
$$

(b) Assume that i-th individual belongs to set $\mathcal{P}_{j,3}$. Then, the numerator is

$$
\begin{aligned}
&P(\tilde{Y}_i=1,Z_{j,1}=1,Z_{j,2}=0,Z_{j,3}=1,\mathcal{Y}_{\mathcal{P}_{j,3}})\\
&=\sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j\setminus i}} P(Z_{j,1}=1,Z_{j,2}=0,Z_{j,3}=1,\mathcal{Y}_{\mathcal{P}_{j,3}}|\tilde{Y}_i=1,\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}},\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}\setminus i})
\end{aligned}
$$

$$
\begin{aligned}
&\times P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}}, \tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}\setminus i})P(\tilde{Y}_i = 1)\\
&= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j \setminus i}} P(Z_{j,1} = 1|\tilde{Z}_{j,1} = 1)P(Z_{j,2} = 0|\tilde{Z}_{j,2})P(Z_{j,3} = 1|\tilde{Z}_{j,3} = 1)\\
&\times P(\mathcal{Y}_{\mathcal{P}_{j,3}\setminus i}|\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}\setminus i})P(Y_i|\tilde{Y}_i = 1)P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}})P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}\ \setminus i})P(\tilde{Y}_i = 1)\\
&= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j \setminus i}} S_e^2(1 - S_e)^{\tilde{Z}_{j,2}} S_p^{1-\tilde{Z}_{j,2}} p_{iB}^{(t)} S_e^{Y_i}(1 - S_e)^{1-Y_i}\\
&\times \prod_{i \in \mathcal{P}_{j,3}\setminus\{i\}} \left[S_e^{Y_i}(1 - S_e)^{1-Y_i}\right]^{\tilde{Y}_i} \left[(1 - S_p)^{Y_i} S_p^{1-Y_i}\right]^{1-\tilde{Y}_i}\\
&\times \prod_{i \in \mathcal{P}_j\setminus\{i\}} \left[p_{iB}^{(t)}\right]^{\tilde{Y}_i} \left[1 - p_{iB}^{(t)}\right]^{1-\tilde{Y}_i}.
\end{aligned}
$$

4) When $Z_{j,1} = 1, Z_{j,2} = 1$, and $Z_{j,3} = 0$, the process is the same as when $Z_{j,1} = 1, Z_{j,2} = 0$, and $Z_{j,3} = 1$, and the numerator needs to be discussed accordingly. At this time,

$$
\begin{aligned}
w_i^{(t)} &= P(\tilde{Y}_i = 1|Z_{j,1} = 1, Z_{j,2} = 1, Z_{j,3} = 0, \mathcal{Y}_{\mathcal{P}_{j,2}})\\
&= \frac{P(Z_{j,1} = 1, Z_{j,2} = 1, Z_{j,3} = 0, \mathcal{Y}_{\mathcal{P}_{j,2}}|\tilde{Y}_i = 1)P(\tilde{Y}_i = 1)}{P(Z_{j,1} = 1, Z_{j,2} = 1, Z_{j,3} = 0, \mathcal{Y}_{\mathcal{P}_{j,2}})}.
\end{aligned}
$$

First, the denominator is

$$
\begin{aligned}
&P(Z_{j,1} = 1, Z_{j,2} = 1, Z_{j,3} = 0, \mathcal{Y}_{\mathcal{P}_{j,2}})\\
&= \sum_{\tilde{\mathcal{Y}}} P(Z_{j,1} = 1, Z_{j,2} = 1, Z_{j,3} = 0, \mathcal{Y}_{\mathcal{P}_{j,2}}|\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}}, \tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}})P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}})P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}})\\
&= \sum_{\tilde{\mathcal{Y}}} P(Z_{j,1} = 1|\tilde{Z}_{j,1})P(Z_{j,2} = 1|\tilde{Z}_{j,2})P(Z_{j,3} = 0|\tilde{Z}_{j,3})\\
&\times \prod_{i \in \mathcal{P}_{j,2}} P(Y_i|\tilde{Y}_i)P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}})P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}})\\
&= \sum_{\tilde{\mathcal{Y}}} S_e^{\tilde{Z}_{j,1}+\tilde{Z}_{j,2}}(1 - S_p)^{2-\tilde{Z}_{j,1}-\tilde{Z}_{j,2}}(1 - S_e)^{\tilde{Z}_{j,3}} S_p^{1-\tilde{Z}_{j,3}}\\
&\times \prod_{i \in \mathcal{P}_j} [p_{iB}^{(t)}]^{\tilde{Y}_i}[1 - p_{iB}^{(t)}]^{1-\tilde{Y}_i} \prod_{i \in \mathcal{P}_{j,2}} [S_e^{Y_i}(1 - S_e)^{1-Y_i}]^{\tilde{Y}_i}[(1 - S_p)^{Y_i} S_p^{1-Y_i}]^{1-\tilde{Y}_i}.
\end{aligned}
$$

Next, the numerator is discussed.
(a) Assume that i-th individual belongs to set $\mathcal{P}_{j,2}$. Then, the numerator is

$$
\begin{aligned}
&P(\tilde{Y}_i = 1, Z_{j,1} = 1, Z_{j,2} = 1, Z_{j,3} = 0, \mathcal{Y}_{\mathcal{P}_{j,2}})\\
&= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j \setminus i}} P(Z_{j,1} = 1, Z_{j,2} = 1, Z_{j,3} = 0, \mathcal{Y}_{\mathcal{P}_{j,2}}|\tilde{Y}_i = 1, \tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}\setminus i}, \tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}})P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}\setminus i}, \tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}})P(\tilde{Y}_i = 1)\\
&= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j \setminus i}} P(Z_{j,1} = 1|\tilde{Z}_{j,1} = 1)P(Z_{j,2} = 1|\tilde{Z}_{j,2} = 1)P(Z_{j,3} = 0|\tilde{Z}_{j,3})
\end{aligned}
$$

$$\times P(\mathcal{Y}_{\mathcal{P}_{j,2}\backslash i}|\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}\backslash i})P(Y_i|\tilde{Y}_i = 1)P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}\backslash i})P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}})P(\tilde{Y}_i = 1)$$
$$= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j\backslash i}} S_e^2(1 - S_e)^{\tilde{Z}_{j,3}} S_p^{1-\tilde{Z}_{j,3}} S_e^{Y_i}(1 - S_e)^{1-Y_i} p_{iB}^{(t)}$$
$$\times \prod_{i\in\mathcal{P}_{j,2}\backslash\{i\}} [S_e^{Y_i}(1 - S_e)^{1-Y_i}]^{\tilde{Y}_i}[(1 - S_p)^{Y_i} S_p^{1-Y_i}]^{1-\tilde{Y}_i} \times \prod_{i\in\mathcal{P}_j\backslash\{i\}} [p_{iB}^{(t)}]^{\tilde{Y}_i}[1 - p_{iB}^{(t)}]^{1-\tilde{Y}_i}.$$

(b) Assume that i-th individual belongs to set $\mathcal{P}_{j,3}$. Then, the numerator is

$$P(\tilde{Y}_i = 1, Z_{j,1} = 1, Z_{j,2} = 1, Z_{j,3} = 0, \mathcal{Y}_{\mathcal{P}_{j,2}})$$
$$=P(Z_{j,1} = 1, Z_{j,2} = 1, Z_{j,3} = 0, \mathcal{Y}_{\mathcal{P}_{j,2}}|\tilde{Y}_i = 1)P(\tilde{Y}_i = 1)$$
$$= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j\backslash i}} P(Z_{j,1} = 1, Z_{j,2} = 1, Z_{j,3} = 0, \mathcal{Y}_{\mathcal{P}_{j,2}}|\tilde{Y}_i = 1, \tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}}, \tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}\backslash i})P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}}, \tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}\backslash i})P(\tilde{Y}_i = 1)$$
$$= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j\backslash i}} P(Z_{j,1} = 1|\tilde{Z}_{j,1} = 1)P(Z_{j,2} = 1|\tilde{Z}_{j,2})P(Z_{j,3} = 0|\tilde{Z}_{j,3} = 1)$$
$$\times P(\mathcal{Y}_{\mathcal{P}_{j,2}}|\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}})P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}})P(\mathcal{Y}_{\mathcal{P}_{j,3}\backslash i}|\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}\backslash i})P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}\backslash i})P(\tilde{Y}_i = 1)$$
$$= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j\backslash i}} S_e(1 - S_e)S_e^{\tilde{Z}_{j,2}}(1 - S_p)^{1-\tilde{Z}_{j,2}} p_{iB}^{(t)} \times \prod_{i\in\mathcal{P}_{j,2}} [S_e^{Y_i}(1 - S_e)^{1-Y_i}]^{\tilde{Y}_i}$$
$$\times [(1 - S_p)^{Y_i} S_p^{1-Y_i}]^{1-\tilde{Y}_i} \prod_{i\in\mathcal{P}_j\backslash\{i\}} [p_{iB}^{(t)}]^{\tilde{Y}_i}[1 - p_{iB}^{(t)}]^{1-\tilde{Y}_i}.$$

5) When $Z_{j,1} = 1, Z_{j,2} = 1$, and $Z_{j,3} = 1$, two similar partitions are performed as above, and individual retests are conducted separately for all individuals in $\mathcal{P}_j$. At this time, $\mathcal{Y}_{\mathcal{P}_j} = \mathcal{Y}_{\mathcal{P}_{j,2}} \cup \mathcal{Y}_{\mathcal{P}_{j,3}}$, and we have

$$w_i^{(t)} =P(\tilde{Y}_i = 1|Z_{j,1} = 1, Z_{j,2} = 1, Z_{j,3} = 1, \mathcal{Y}_{\mathcal{P}_{j,2}}, \mathcal{Y}_{\mathcal{P}_{j,3}})$$
$$=\frac{P(Z_{j,1} = 1, Z_{j,2} = 1, Z_{j,3} = 1, \mathcal{Y}_{\mathcal{P}_j}|\tilde{Y}_i = 1)P(\tilde{Y}_i = 1)}{P(Z_{j,1} = 1, Z_{j,2} = 1, Z_{j,3} = 1, \mathcal{Y}_{\mathcal{P}_j})}.$$

The denominator is

$$P(Z_{j,1} = 1, Z_{j,2} = 1, Z_{j,3} = 1, \mathcal{Y}_{\mathcal{P}_j})$$
$$= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j}} P(Z_{j,1} = 1, Z_{j,2} = 1, Z_{j,3} = 1, \mathcal{Y}_{\mathcal{P}_j}|\tilde{\mathcal{Y}}_{\mathcal{P}_j})P(\tilde{\mathcal{Y}}_{\mathcal{P}_j})$$
$$= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j}} P(Z_{j,1} = 1|\tilde{Z}_{j,1})P(Z_{j,2} = 1|\tilde{Z}_{j,2})P(Z_{j,3} = 1|\tilde{Z}_{j,3}) \prod_{i\in\mathcal{P}_j} P(Y_i|\tilde{Y}_i)P(\tilde{Y}_i)$$
$$= \sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j}} S_e^{\tilde{Z}_{j,1}}(1 - S_p)^{1-\tilde{Z}_{j,1}} \prod_{u=2}^{3} S_e^{\tilde{Z}_{j,u}}(1 - S_p)^{1-\tilde{Z}_{j,u}}$$
$$\times \prod_{i\in\mathcal{P}_j} [S_e^{Y_i}(1 - S_e)^{1-Y_i} p_{iB}^{(t)}]^{\tilde{Y}_i}[(1 - S_p)^{Y_i} S_p^{1-Y_i}(1 - p_{iB}^{(t)})]^{1-\tilde{Y}_i}.$$

The results of i-th individual belonging to either set $\mathcal{P}_{j,2}$ or $\mathcal{P}_{j,3}$ are symmetric. Assume that i-th individual belongs to set $\mathcal{P}_{j,2}$. Then, the numerator is

$$
\begin{aligned}
&P(Z_{j,1}=1, Z_{j,2}=1, Z_{j,3}=1, \mathcal{Y}_{\mathcal{P}_j}, \tilde{Y}_i=1)\\
=&\sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j\setminus i}} P(Z_{j,1}=1, Z_{j,2}=1, Z_{j,3}=1, \mathcal{Y}_{\mathcal{P}_j}|\tilde{Y}_i=1, \tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}\setminus i}, \tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}})P(\tilde{\mathcal{Y}}_{\mathcal{P}_{j,2}\setminus i}, \tilde{\mathcal{Y}}_{\mathcal{P}_{j,3}})P(\tilde{Y}_i=1)\\
=&\sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j\setminus i}} P(Z_{j,1}=1|\tilde{Z}_{j,1}=1)P(Z_{j,2}=1|\tilde{Z}_{j,2}=1)P(Z_{j,3}=1|\tilde{Z}_{j,3})\\
&\times P(Y_i|\tilde{Y}_i=1)P(\mathcal{Y}_{\mathcal{P}_j\setminus i}|\tilde{\mathcal{Y}}_{\mathcal{P}_j\setminus i})P(\tilde{\mathcal{Y}}_{\mathcal{P}_j\setminus i})P(\tilde{Y}_i=1)\\
=&\sum_{\tilde{\mathcal{Y}}_{\mathcal{P}_j\setminus i}} S_e^2 S_e^{\tilde{Z}_{j,3}}(1-S_p)^{1-\tilde{Z}_{j,3}} p_{iB}^{(t)} S_e^{Y_i}(1-S_e)^{1-Y_i}\\
&\times \prod_{i\in\mathcal{P}_j\setminus\{i\}} \left[S_e^{Y_i}(1-S_e)^{1-Y_i}p_{iB}^{(t)}\right]^{\tilde{Y}_i}\left[(1-S_p)^{Y_i}S_p^{1-Y_i}(1-p_{iB}^{(t)})\right]^{1-\tilde{Y}_i}.
\end{aligned}
$$

C.3. Array testing

For convenience, assume that the set of all individuals is G, and all individuals can be arranged into an $R\times C$ array, that is, $G=\{(r,c), r\in R, c\in C\}$. Define $\mathcal{R}=(R_1, R_2, \cdots, R_R)$ and $C=(C_1, C_2, \cdots, C_C)$ as the collections of row and column testing results, respectively. Let $R=\max R_r$ and $C=\max C_c$. Furthermore, define $\tilde{R}_r=\max_c \tilde{Y}_{rc}$ and $\tilde{C}_c=\max_r \tilde{Y}_{rc}$ as the true result sets for rows and columns, respectively. Let Y_{rc} denote the testing result of the individual in the r-th row and c-th column of the array, and $\tilde{Y}_{rc}$ represents the true disease status of the individual in the r-th row and c-th column of the array. Let

$$
\begin{aligned}
Q=\{(s,t)\mid & R_s=1, C_t=1, 1\le s\le R, 1\le t\le C,\\
&\text{or } R_s=1, C_1=\cdots=C_C=0, 1\le s\le R,\\
&\text{or } R_1=\cdots=R_R=0, C_t=1, 1\le t\le C\}.
\end{aligned}
$$

$\mathcal{Y}_Q$ represents the collection of responses from all potentially positive individuals, and $\tilde{\mathcal{Y}}_Q$ denotes the true disease statuses of all potentially positive individuals. Let $\mathcal{Z}_G=(R,C)$ denote the group testing responses. Since $(r,c)\in G$, define

$$
\tilde{\mathcal{Y}}_{G\setminus(r,c)}=\left\{\tilde{Y}_{r'c'}, r'\in R\setminus\{r\}, c'\in C\setminus\{c\}\right\}.
$$

Then,

$$
w_{rc}^{(t)}=P(\tilde{Y}_{rc}=1\mid \mathcal{Z}_G, \mathcal{Y}_Q)=\frac{P(\tilde{Y}_{rc}=1, \mathcal{Z}_G, \mathcal{Y}_Q)}{P(\mathcal{Z}_G, \mathcal{Y}_Q)}.
$$

1) When $\mathcal{Z}_G=(0,0)$, there is no need to retest individuals within the group. At this time,

$$
w_{rc}^{(t)}=P(\tilde{Y}_{rc}=1\mid \mathcal{Z}_G=(0,0))=\frac{P(\tilde{Y}_{rc}=1, \mathcal{Z}_G=(0,0))}{P(\mathcal{Z}_G=(0,0))}.
$$

The denominator is

$$
\begin{aligned}
P(\mathcal{Z}_G = (0,0)) &= \sum_{\tilde{\mathcal{Y}}_G} P(\mathcal{Z}_G = (0,0)|\tilde{\mathcal{Y}}_G)P(\tilde{\mathcal{Y}}_G) \\
&= \sum_{\tilde{\mathcal{Y}}_G} P(\mathcal{R} = 0|\tilde{\mathcal{Y}}_G)P(C = 0|\tilde{\mathcal{Y}}_G)P(\tilde{\mathcal{Y}}_G) \\
&= \sum_{\tilde{\mathcal{Y}}_G} \Big[\prod_{r'=1}^{R} P\Big(R_{r'} = 0|\tilde{Y}_{r'1}, \tilde{Y}_{r'2}, \cdots, \tilde{Y}_{r'C}\Big)\Big]\Big[\prod_{c'=1}^{C} P\Big(C_{c'} = 0|\tilde{Y}_{1c'}, \tilde{Y}_{2c'}, \cdots, \tilde{Y}_{Rc'}\Big)\Big] \\
&\quad \times \prod_{r'\in R}\prod_{c'\in C} {p^{(t)}_{r'c'B}}^{\tilde{Y}_{r'c'}}\Big(1 - p^{(t)}_{r'c'B}\Big)^{1-\tilde{Y}_{r'c'}} \\
&= \sum_{\tilde{\mathcal{Y}}_G} \prod_{r'=1}^{R}\Big[(1 - S_e)^{\tilde{R}_{r'}} S_p^{1-\tilde{R}_{r'}}\Big]\prod_{c'=1}^{C}\Big[(1 - S_e)^{\tilde{C}_{c'}} S_p^{1-\tilde{C}_{c'}}\Big] \prod_{(r',c')\in G}\Big\{{p^{(t)}_{r'c'B}}^{\tilde{Y}_{r'c'}}\Big(1 - p^{(t)}_{r'c'B}\Big)^{1-\tilde{Y}_{r'c'}}\Big\} \\
&= \sum_{\tilde{\mathcal{Y}}_G} \prod_{r'=1}^{R}\prod_{c'=1}^{C}\Big[(1 - S_e)^{\tilde{R}_{r'}+\tilde{C}_{c'}} S_p^{2-\tilde{R}_{r'}-\tilde{C}_{c'}}\Big] \prod_{(r',c')\in G}\Big\{{p^{(t)}_{r'c'B}}^{\tilde{Y}_{r'c'}}\Big(1 - p^{(t)}_{r'c'B}\Big)^{1-\tilde{Y}_{r'c'}}\Big\}.
\end{aligned}
$$

The numerator is

$$
\begin{aligned}
P(\mathcal{Z}_G = (0,0), \tilde{Y}_{rc} = 1) &= P(\mathcal{Z}_G = (0,0)|\tilde{Y}_{rc} = 1)P(\tilde{Y}_{rc} = 1) \\
&= \sum_{\tilde{\mathcal{Y}}_{G\setminus(r,c)}} P(\mathcal{R} = 0, C = 0|\tilde{\mathcal{Y}}_{G\setminus(r,c)}, \tilde{Y}_{rc} = 1)P(\tilde{\mathcal{Y}}_{G\setminus(r,c)}) \\
&= \sum_{\tilde{\mathcal{Y}}_{G\setminus(r,c)}} P(R_r = 0|\tilde{R}_r = 1)\Big[\prod_{r'\in R\setminus\{r\}} P(R_{r'} = 0|\tilde{Y}_{r'1}, \tilde{Y}_{r'2}, \cdots, \tilde{Y}_{r'C})\Big] \\
&\quad \times P(C_c = 0|\tilde{C}_c = 1)\Big[\prod_{c'\in C\setminus\{c\}} P(C_{c'} = 0|\tilde{Y}_{1c'}, \cdots, \tilde{Y}_{Rc'})\Big] \\
&\quad \times \prod_{r'\in R\setminus\{r\}}\ \prod_{c'\in C\setminus\{c\}} {p^{(t)}_{r'c'B}}^{\tilde{Y}_{r'c'}}(1 - p^{(t)}_{r'c'B})^{1-\tilde{Y}_{r'c'}} p^{(t)}_{rcB} \\
&= \sum_{\tilde{\mathcal{Y}}_{G\setminus(r,c)}} (1 - S_e)^2 \prod_{r'\in R\setminus\{r\}} (1 - S_e)^{\tilde{R}_{r'}} S_p^{1-\tilde{R}_{r'}} \prod_{c'\in C\setminus\{c\}} (1 - S_e)^{\tilde{C}_{c'}} S_p^{1-\tilde{C}_{c'}} \\
&\quad \times \prod_{(r',c')\in G} {p^{(t)}_{r'c'B}}^{\tilde{Y}_{r'c'}}(1 - p^{(t)}_{r'c'B})^{1-\tilde{Y}_{r'c'}} p^{(t)}_{rcB} \\
&= \sum_{\tilde{\mathcal{Y}}_{G\setminus(r,c)}} \prod_{r'\in R\setminus\{r\}}\ \prod_{c'\in C\setminus\{c\}} (1 - S_e)^{2+\tilde{R}_i+\tilde{C}_c} S_p^{2-\tilde{R}_i-\tilde{C}_c} \\
&\quad \times \prod_{(r',c')\in G\setminus(r,c)} {p^{(t)}_{r'c'B}}^{\tilde{Y}_{r'c'}}(1 - p^{(t)}_{r'c'B})^{1-\tilde{Y}_{r'c'}} p^{(t)}_{rcB}.
\end{aligned}
$$

2) When $\mathcal{Z}_G \neq (0,0)$, $\mathcal{Z}_G = (R, C)$ has multiple scenarios, specifically $\mathcal{Z}_G = (R, C) = (1,0)$, $(R, C) = (0,1)$, and $(R, C) = (1,1)$. Therefore, when $\mathcal{Z}_G \neq (0,0)$, the following classifications can be discussed:

(a) When $(R, C) = (1, 0)$,

$$w_{rc}^{(t)} = P(\tilde{Y}_{rc} = 1 \mid \mathcal{Z}_G = (1,0)) = \frac{P(\tilde{Y}_{rc} = 1, \mathcal{Z}_G = (1,0))}{P(\mathcal{Z}_G = (1,0))}.$$

The denominator is

$$\begin{aligned}
P(\mathcal{Z}_G = (1,0)) &= \sum_{\tilde{\mathcal{Y}}_G} P(\mathcal{R} \neq 0, C = 0, \mathcal{Y}_Q | \tilde{\mathcal{Y}}_G) P(\tilde{\mathcal{Y}}_G) \\
&= \sum_{\tilde{\mathcal{Y}}_G} \Big[\prod_{r'=1}^{R} P\Big(R_{r'} | \tilde{Y}_{r'1}, \tilde{Y}_{r'2}, \ldots, \tilde{Y}_{r'C}\Big)\Big]\Big[\prod_{c'=1}^{C} P\Big(C_{c'} = 0 | \tilde{Y}_{1c'}, \tilde{Y}_{2c'}, \ldots, \tilde{Y}_{Rc'}\Big)\Big] \\
&\quad \times \Big[\prod_{(s,t)\in Q} P\Big(Y_{st} | \tilde{Y}_{st}\Big)\Big]\Big[\prod_{r'\in R}\prod_{c'\in C} {p_{r'c'B}^{(t)}}^{\tilde{Y}_{r'c'}} \Big(1 - p_{r'c'B}^{(t)}\Big)^{1-\tilde{Y}_{r'c'}}\Big] \\
&= \sum_{\tilde{\mathcal{Y}}_G} \prod_{r'=1}^{R} \Big[S_e^{R_{r'}} (1 - S_e)^{1-R_{r'}}\Big]^{\tilde{R}_{r'c}} \Big[(1 - S_p)^{R_{r'}} S_p^{1-R_{r'}}\Big]^{1-\tilde{R}_{r'}} \\
&\quad \times \prod_{c'=1}^{C} \Big[(1 - S_e)^{1-C_{c'}}\Big]^{\tilde{C}_{c'}} \Big[S_p^{1-C_{c'}}\Big]^{1-\tilde{C}_{c'}} \times \prod_{(s,t)\in Q} \Big[S_e^{Y_{st}} (1 - S_e)^{1-Y_{st}}\Big]^{\tilde{Y}_{st}} \Big[(1 - S_p)^{Y_{st}} S_p^{1-Y_{st}}\Big]^{1-\tilde{Y}_{st}} \\
&\quad \times \prod_{(r',c')\in G} {p_{r'c'B}^{(t)}}^{\tilde{Y}_{r'c'}} \Big(1 - p_{r'c'B}^{(t)}\Big)^{1-\tilde{Y}_{r'c'}}.
\end{aligned}$$

At this point, the numerator requires further discussion:

(i) If $(r, c) \in Q$ and $R_r = 1$ and $C_r = 0$, then

$$\begin{aligned}
&P(\tilde{Y}_{rc} = 1, \mathcal{Z}_G = (1,0)) \\
&= \sum_{\tilde{\mathcal{Y}}_{G\setminus(r,c)}} P(\mathcal{Z}_G = (1,0), \mathcal{Y}_Q \big| \tilde{Y}_{rc} = 1, \tilde{\mathcal{Y}}_{G\setminus(r,c)}) P(\tilde{Y}_{rc} = 1, \tilde{\mathcal{Y}}_{G\setminus(r,c)}) \\
&= \sum_{\tilde{\mathcal{Y}}_{G\setminus(r,c)}} P(\mathcal{R} \neq 0 \big| \tilde{Y}_{rc} = 1, \tilde{\mathcal{Y}}_{G\setminus(r,c)}) P(C = 0 \big| \tilde{Y}_{rc} = 1, \tilde{\mathcal{Y}}_{G\setminus(r,c)}) \\
&\quad \times P(\mathcal{Y}_Q \big| \tilde{Y}_{rc} = 1, \tilde{\mathcal{Y}}_{Q\setminus(r,c)}) P(\tilde{Y}_{rc} = 1) P(\tilde{\mathcal{Y}}_{G\setminus(r,c)}) \\
&= \sum_{\tilde{\mathcal{Y}}_{G\setminus(r,c)}} P\Big(R_r = 1 | \tilde{R}_r = 1\Big)\Big[\prod_{r'\in R\setminus\{r\}} P\Big(R_{r'} | \tilde{Y}_{r'1}, \tilde{Y}_{r'2}, \cdots, \tilde{Y}_{r'C}\Big)\Big] \\
&\quad \times P\Big(C_c = 0 | \tilde{C}_c = 1\Big)\Big[\prod_{c'\in C\setminus\{c\}} P\Big(C_{c'} = 0 | \tilde{Y}_{1c'}, \cdots, \tilde{Y}_{Rc'}\Big)\Big] P\Big(Y_{rc} | \tilde{Y}_{rc} = 1\Big) \\
&\quad \times \Big[\prod_{(s,t)\in Q\setminus(r,c)} P\Big(Y_{st} | \tilde{Y}_{st}\Big)\Big] p_{rcB}^{(t)} \times \Big[\prod_{r'\in R\setminus\{r\}}\prod_{c'\in C\setminus\{c\}} {p_{r'c'B}^{(t)}}^{\tilde{Y}_{r'c'}} \Big(1 - p_{r'c'B}^{(t)}\Big)^{1-\tilde{Y}_{r'c'}}\Big] \\
&= \sum_{\tilde{\mathcal{Y}}_{G\setminus(r,c)}} S_e^{1+Y_{rc}} \Big(1 - S_e\Big)^{2-Y_{rc}} p_{rcB}^{(t)} \\
&\quad \times \prod_{r'\in R\setminus\{r\}} \Big[S_e^{R_{r'}} \Big(1 - S_e\Big)^{1-R_{r'}}\Big]^{\tilde{R}_{r'c}} \Big[(1 - S_p)^{R_{r'}} S_p^{1-R_{r'}}\Big]^{1-\tilde{R}_{r'}}
\end{aligned}$$

$$\times \prod_{c'\in C\setminus\{c\}} \left[(1-S_e)^{1-C_{c'}}\right]^{\tilde{C}_{c'}} \left[S_p^{1-C_{c'}}\right]^{1-\tilde{C}_{c'}}$$
$$\times \prod_{(s,t)\in Q\setminus\{(r,c)\}} \left[S_e^{Y_{st}}\left(1-S_e\right)^{1-Y_{st}}\right]^{\tilde{Y}_{st}} \left[(1-S_p)^{Y_{st}} S_p^{1-Y_{st}}\right]^{1-\tilde{Y}_{st}}$$
$$\times \prod_{(r',c')\in G\setminus\{(r,c)\}} p_{r'c'B}^{(t)}{}^{\tilde{Y}_{r'c'}} \left(1-p_{r'c'B}^{(t)}\right)^{1-\tilde{Y}_{r'c'}}.$$

(ii) If $(r,c) \notin Q$, then $\mathcal{R} = 0, C \neq 0$, but $R_r = 0$ and $C_r = 0$:

$$\begin{aligned}
&P(\tilde{Y}_{rc} = 1, \mathcal{Z}_G = (1,0)) \\
&= \sum_{\tilde{\mathcal{Y}}_{G\setminus(r,c)}} P\Big(\mathcal{R}, C, \mathcal{Y}_Q | \tilde{Y}_{rc} = 1, \tilde{\mathcal{Y}}_{G\setminus(r,c)}\Big) P(\tilde{Y}_{rc} = 1, \tilde{\mathcal{Y}}_{G\setminus(r,c)}) \\
&= \sum_{\tilde{\mathcal{Y}}_{G\setminus(r,c)}} P\Big(\mathcal{R} | \tilde{Y}_{rc} = 1, \tilde{\mathcal{Y}}_{G\setminus(r,c)}\Big) P\Big(C | \tilde{Y}_{rc} = 1, \tilde{\mathcal{Y}}_{G\setminus(r,c)}\Big) P(\mathcal{Y}_Q \big| \tilde{\mathcal{Y}}_Q) P(\tilde{Y}_{rc} = 1) P(\tilde{\mathcal{Y}}_{G\setminus(r,c)}) \\
&= \sum_{\tilde{\mathcal{Y}}_{G\setminus(r,c)}} P\Big(R_r = 0 | \tilde{R}_r = 1\Big) \Big[\prod_{r'\in R\setminus\{r\}} P\Big(R_{r'} | \tilde{Y}_{r'1}, \tilde{Y}_{r'2}, \cdots, \tilde{Y}_{r'C}\Big)\Big] \\
&\quad \times P\Big(C_c = 0 | \tilde{C}_c = 1\Big) \Big[\prod_{c'\in C\setminus\{c\}} P\Big(C_{c'} = 0 | \tilde{Y}_{1c'}, \cdots, \tilde{Y}_{Rc'}\Big)\Big] \\
&\quad \times \Big[\prod_{(s,t)\in Q} P\Big(Y_{st} | \tilde{Y}_{st}\Big)\Big] p_{rcB}^{(t)} \prod_{(r',c')\in G\setminus\{(r,c)\}} p_{r'c'B}^{(t)}{}^{\tilde{Y}_{r'c'}} \left(1-p_{r'c'B}^{(t)}\right)^{1-\tilde{Y}_{r'c'}} \\
&= \sum_{\tilde{\mathcal{Y}}_{G\setminus(r,c)}} S_e^{Y_{rc}} \left(1-S_e\right)^{3-Y_{rc}} p_{rcB}^{(t)} \\
&\quad \times \prod_{r'\in R\setminus\{r\}} \left[S_e^{R_{r'}} (1-S_e)^{1-R_{r'}}\right]^{\tilde{R}_{r'c}} \left[(1-S_p)^{R_{r'}} S_p^{1-R_{r'}}\right]^{1-\tilde{R}_{r'c}} \\
&\quad \times \prod_{c'\in C\setminus\{c\}} \left[(1-S_e)^{1-C_{c'}}\right]^{\tilde{C}_{c'}} \left[S_p^{1-C_{c'}}\right]^{1-\tilde{C}_{c'}} \\
&\quad \times \prod_{(s,t)\in Q} \left[S_e^{Y_{st}} (1-S_e)^{1-Y_{st}}\right]^{\tilde{Y}_{st}} \left[(1-S_p)^{Y_{st}} S_p^{1-Y_{st}}\right]^{1-\tilde{Y}_{st}} \\
&\quad \times \prod_{(r',c')\in G\setminus\{(r,c)\}} \left[p_{r'c'B}^{(t)}{}^{\tilde{Y}_{r'c'}} \left(1-p_{r'c'B}^{(t)}\right)^{1-\tilde{Y}_{r'c'}}\right].
\end{aligned}$$

(b) When $(R,C) = (0,1)$, the denominator is

$$\begin{aligned}
P(\mathcal{Z}_G = (0,1)) &= \sum_{\tilde{\mathcal{Y}}_G} P(\mathcal{R} = 0, C \neq 0, \mathcal{Y}_Q | \tilde{\mathcal{Y}}_G) P(\tilde{\mathcal{Y}}_G) \\
&= \sum_{\tilde{\mathcal{Y}}_G} \Big[\prod_{r'=1}^{R} P(R'_r = 0 | \tilde{Y}_{r'1}, \tilde{Y}_{r'2}, \cdots, \tilde{Y}_{r'C})\Big]
\end{aligned}$$

$$
\begin{aligned}
&\times\Big[\prod_{c'=1}^{C} P(C_{c'}|\tilde{Y}_{1c'},\tilde{Y}_{2c'},\cdots,\tilde{Y}_{Rc'})\Big]\\
&\times\prod_{(s,t)\in Q} P(Y_{st}\big|\tilde{Y}_{st})\prod_{r'\in R}\prod_{c'\in C} {p^{(t)}_{r'c'B}}^{\tilde{Y}_{r'c'}}(1-p^{(t)}_{r'c'B})^{1-\tilde{Y}_{r'c'}}\\
=&\sum_{\tilde{\mathcal{Y}}_G}\prod_{r'=1}^{R}\Big[(1-S_e)^{1-R_{r'}}\Big]^{\tilde{R}_{r'}}\Big[S_p^{1-R_{r'}}\Big]^{1-\tilde{R}_{r'}}\\
&\times\prod_{c'=1}^{C}\Big[S_e^{C_{c'}}(1-S_e)^{1-C_{c'}}\Big]^{\tilde{C}_{c'}}\Big[(1-S_p)^{C_{c'}}S_p^{1-C_{c'}}\Big]^{1-\tilde{C}_{c'}}\\
&\times\prod_{(s,t)\in Q}\Big[S_e^{Y_{st}}(1-S_e)^{1-Y_{st}}\Big]^{\tilde{Y}_{st}}\Big[(1-S_p)^{Y_{st}}S_p^{1-Y_{st}}\Big]^{1-\tilde{Y}_{st}}\\
&\times\prod_{(r',c')\in G}\Big[{p^{(t)}_{r'c'B}}^{\tilde{Y}_{r'c'}}(1-p^{(t)}_{r'c'B})^{1-\tilde{Y}_{r'c'}}\Big].
\end{aligned}
$$

The numerator requires further discussion:

(i) If $(r,c)\in Q$ and $R_r=0$ and $C_c=1$, then

$$
\begin{aligned}
&P(\tilde{Y}_{rc}=1,\mathcal{Z}_G=(0,1))=P(\tilde{Y}_{rc}=1,\mathcal{R},\mathcal{C},\tilde{\mathcal{Y}}_Q)\\
=&\sum_{\tilde{\mathcal{Y}}_{G\backslash(r,c)}} P(\mathcal{R},\mathcal{C},\mathcal{Y}_Q|\tilde{Y}_{rc}=1,\tilde{\mathcal{Y}}_{G\backslash(r,c)})P(\tilde{Y}_{rc}=1,\tilde{\mathcal{Y}}_{G\backslash(r,c)})\\
=&\sum_{\tilde{\mathcal{Y}}_{G\backslash(r,c)}} S_e^{Y_{rc}}(3-S_e)^{1-Y_{rc}}p^{(t)}_{rcB}\prod_{r'\in R\backslash\{r\}}\Big[(1-S_e)^{1-R_{r'}}\Big]^{\tilde{R}_{r'c}}\Big[S_p^{1-R_{r'}}\Big]^{1-\tilde{R}_{r'c}}\\
&\times\prod_{c'\in C\backslash\{c\}}\Big[S_e^{C_{c'}}(1-S_e)^{1-C_{c'}}\Big]^{\tilde{C}_{c'}}\Big[(1-S_p)^{C_{c'}}S_p^{1-C_{c'}}\Big]^{1-\tilde{C}_{c'}}\\
&\times\prod_{(s,t)\in Q}\Big[S_e^{Y_{st}}(1-S_e)^{1-Y_{st}}\Big]^{\tilde{Y}_{st}}\Big[(1-S_p)^{Y_{st}}S_p^{1-Y_{st}}\Big]^{1-\tilde{Y}_{st}}\\
&\times\prod_{(r',c')\in G\backslash\{(r,c)\}}\Big[{p^{(t)}_{r'c'B}}^{\tilde{Y}_{r'c'}}(1-p^{(t)}_{r'c'B})^{1-\tilde{Y}_{r'c'}}\Big].
\end{aligned}
$$

(ii) If $(r,c)\notin Q$, then $\mathcal{R}=0$, $\mathcal{C}\neq 0$, but $R_r=0$ and $C_r=0$:

$$
\begin{aligned}
P(\tilde{Y}_{rc}=1,\mathcal{Z}_G=(0,1))=&P(\tilde{Y}_{rc}=1,\mathcal{R},\mathcal{C},\mathcal{Y}_Q)\\
=&\sum_{\tilde{\mathcal{Y}}_{G\backslash(r,c)}} P(\mathcal{R},\mathcal{C},\mathcal{Y}_Q|\tilde{Y}_{rc}=1,\tilde{\mathcal{Y}}_{G\backslash(r,c)})P(\tilde{Y}_{rc}=1,\tilde{\mathcal{Y}}_{G\backslash(r,c)})\\
=&\sum_{\tilde{\mathcal{Y}}_{G\backslash(r,c)}} S_e^{Y_{rc}}(3-S_e)^{1-Y_{rc}}p^{(t)}_{rcB}\prod_{r'\in R\backslash\{r\}}\Big[(1-S_e)^{1-R_{r'}}\Big]^{\tilde{R}_{r'c}}\Big[S_p^{1-R_{r'}}\Big]^{1-\tilde{R}_{r'c}}\\
&\times\prod_{c'\in C\backslash\{c\}}\Big[S_e^{C_{c'}}(1-S_e)^{1-C_{c'}}\Big]^{\tilde{C}_{c'}}\Big[(1-S_p)^{C_{c'}}S_p^{1-C_{c'}}\Big]^{1-\tilde{C}_{c'}}
\end{aligned}
$$

$$\times \prod_{(s,t)\in Q} \left[S_e^{Y_{st}} (1-S_e)^{1-Y_{st}} \right]^{\tilde{Y}_{st}} \left[(1-S_p)^{Y_{st}} S_p^{1-Y_{st}} \right]^{1-\tilde{Y}_{st}}$$

$$\times \prod_{(r',c')\in G\setminus\{(r,c)\}} ({p_{r'c'B}^{(t)}}^{\tilde{Y}_{r'c'}} (1-p_{r'c'B}^{(t)})^{1-\tilde{Y}_{r'c'}}).$$

(c) When $(R, C) = (1, 1)$, the denominator is

$$\begin{aligned}
P(\mathcal{Z}_G = (1,1)) &= \sum_{\tilde{\mathcal{Y}}} P(\mathcal{R}, C, \mathcal{Y}_Q | \tilde{\mathcal{Y}}) P(\tilde{\mathcal{Y}}) \\
&= \sum_{\tilde{\mathcal{Y}}} \left[\prod_{r'=1}^{R} P(R'_r | \tilde{Y}_{r'1}, \tilde{Y}_{r'2}, \cdots, \tilde{Y}_{r'C}) \right] \left[\prod_{c'=1}^{C} P(C_{c'} | \tilde{Y}_{1c'}, \tilde{Y}_{2c'}, \cdots, \tilde{Y}_{Rc'}) \right] \\
&\quad \times \prod_{(s,t)\in Q} P(Y_{st} | \tilde{Y}_{st}) \prod_{r'\in R} \prod_{c'\in C} {p_{r'c'B}^{(t)}}^{\tilde{Y}_{r'c'}} (1-p_{r'c'B}^{(t)})^{1-\tilde{Y}_{r'c'}} \\
&= \sum_{\tilde{\mathcal{Y}}} \prod_{r'=1}^{R} \left[S_e^{R_{r'}} (1-S_e)^{1-R_{r'}} \right]^{\tilde{R}_{r'c}} \left[(1-S_p)^{R_{r'}} S_p^{1-R_{r'}} \right]^{1-\tilde{R}_{r'}} \\
&\quad \times \prod_{c'=1}^{C} \left[S_e^{C_{c'}} (1-S_e)^{1-C_{c'}} \right]^{\tilde{C}_{c'}} \left[(1-S_p)^{C_{c'}} S_p^{1-C_{c'}} \right]^{1-\tilde{C}_{c'}} \\
&\quad \times \prod_{(s,t)\in Q} \left[S_e^{Y_{st}} (1-S_e)^{1-Y_{st}} \right]^{\tilde{Y}_{st}} \left[(1-S_p)^{Y_{st}} S_p^{1-Y_{st}} \right]^{1-\tilde{Y}_{st}} \\
&\quad \times \prod_{(r',c')\in G} \left\{ {p_{r'c'B}^{(t)}}^{\tilde{Y}_{r'c'}} (1-p_{r'c'B}^{(t)})^{1-\tilde{Y}_{r'c'}} \right\}.
\end{aligned}$$

For the numerator, we provide the following derivations:

(i) If $(r, c) \in Q$ and $R_r = 1$ and $C_c = 1$, then

$$\begin{aligned}
&P(\tilde{Y}_{rc} = 1, \mathcal{Z}_G = (1,1)) = P(\tilde{Y}_{rc} = 1, \mathcal{R}, C, \mathcal{Y}_Q) \\
&= \sum_{\tilde{\mathcal{Y}}_{G\setminus(r,c)}} P(\mathcal{R}, C, \mathcal{Y}_Q | \tilde{Y}_{rc} = 1, \tilde{\mathcal{Y}}_{G\setminus(r,c)}) P(\tilde{Y}_{rc} = 1, \tilde{\mathcal{Y}}_{G\setminus(r,c)}) \\
&= \sum_{\tilde{\mathcal{Y}}_{G\setminus(r,c)}} S_e^{2+Y_{rc}} (1-S_e)^{1-Y_{rc}} \cdot p_{rcB}^{(t)} \\
&\quad \times \prod_{r'\in R\setminus\{r\}} \left[S_e^{R_{r'}} (1-S_e)^{1-R_{r'}} \right]^{\tilde{R}_{r'c}} \left[(1-S_p)^{R_{r'}} S_p^{1-R_{r'}} \right]^{1-\tilde{R}_{r'c}} \\
&\quad \times \prod_{c'\in C\setminus\{c\}} \left[S_e^{C_{c'}} (1-S_e)^{1-C_{c'}} \right]^{\tilde{C}_{c'}} \left[(1-S_p)^{C_{c'}} S_p^{1-C_{c'}} \right]^{1-\tilde{C}_{c'}} \\
&\quad \times \prod_{(s,t)\in Q\setminus\{(r,c)\}} \left[S_e^{Y_{st}} (1-S_e)^{1-Y_{st}} \right]^{\tilde{Y}_{st}} \left[(1-S_p)^{Y_{st}} S_p^{1-Y_{st}} \right]^{1-\tilde{Y}_{st}} \\
&\quad \times \prod_{(r',c')\in G\setminus\{(r,c)\}} \left\{ {p_{r'c'B}^{(t)}}^{\tilde{Y}_{r'c'}} (1-p_{r'c'B}^{(t)})^{1-\tilde{Y}_{r'c'}} \right\}.
\end{aligned}$$

(ii) If $(r, c) \notin Q$, then $\mathcal{R} \neq 0$, $C \neq 0$, but $R_r = 0$ and $C_r = 0$:

$$
\begin{aligned}
&P(\tilde{Y}_{rc} = 1, \mathcal{Z}_G = (1,1)) = P(\tilde{Y}_{rc} = 1, \mathcal{R}, C, \mathcal{Y}_Q) \\
&= \sum_{\tilde{\mathcal{Y}}_{G\setminus(r,c)}} P(\mathcal{R}, C, \mathcal{Y}_Q | \tilde{Y}_{rc} = 1, \tilde{\mathcal{Y}}_{G\setminus(r,c)}) P(\tilde{Y}_{rc} = 1, \tilde{\mathcal{Y}}_{G\setminus(r,c)}) \\
&= \sum_{\tilde{\mathcal{Y}}_{G\setminus(r,c)}} (1 - S_e)^2 S_e^{Y_{rc}} (1 - S_e)^{1-Y_{rc}} p_{rcB}^{(t)} \\
&\quad \times \prod_{r' \in R\setminus\{r\}} \left[S_e^{R_{r'}} (1 - S_e)^{1-R_{r'}} \right]^{\tilde{R}_{r'c}} \left[(1 - S_p)^{R_{r'}} S_p^{1-R_{r'}} \right]^{1-\tilde{R}_{r'c}} \\
&\quad \times \prod_{c' \in C\setminus\{c\}} \left[S_e^{C_{c'}} (1 - S_e)^{1-C_{c'}} \right]^{\tilde{C}_{c'}} \left[(1 - S_p)^{C_{c'}} S_p^{1-C_{c'}} \right]^{1-\tilde{C}_{c'}} \\
&\quad \times \prod_{(s,t)\in Q} \left[S_e^{Y_{st}} (1 - S_e)^{1-Y_{st}} \right]^{\tilde{Y}_{st}} \left[(1 - S_p)^{Y_{st}} S_p^{1-Y_{st}} \right]^{1-\tilde{Y}_{st}} \\
&\quad \times \prod_{(r',c')\in G\setminus\{(r,c)\}} \left\{ p_{r'c'B}^{(t)\ \tilde{Y}_{r'c'}} (1 - p_{r'c'B}^{(t)})^{1-\tilde{Y}_{r'c'}} \right\}.
\end{aligned}
$$