

RESEARCH ARTICLE OPEN ACCESS

Combining Biomarkers to Improve Diagnostic Accuracy in Detecting Diseases With Group-Tested Data

Jin Yang¹  | Wei Zhang²  | Paul S. Albert³  | Aiyi Liu¹  | Zhen Chen¹ 

¹Biostatistics and Bioinformatics Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, USA | ²Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China | ³Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

Correspondence: Zhen Chen (chenzhe@mail.nih.gov)

Received: 13 November 2023 | **Revised:** 22 August 2024 | **Accepted:** 11 September 2024

Funding: Research of J. Yang, A. Liu, and Z. Chen were supported by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). Research of W. Zhang was supported by the National Natural Science Foundation of China (NSFC: # 12001522 and # 72091212) and National Key R&D Program of China (# 2022YFA1004800). Research of P. Albert was supported by the Intramural Research Program of the National Cancer Institute (NCI). This research has been conducted using the UK Biobank Resource under Application Number 86711.

Keywords: AUC | differential misclassification | joint model | multiple biomarkers

ABSTRACT

We consider the problem of combining multiple biomarkers to improve the diagnostic accuracy of detecting a disease when only group-tested data on the disease status are available. There are several challenges in addressing this problem, including unavailable individual disease statuses, differential misclassification depending on group size and number of diseased individuals in the group, and extensive computation due to a large number of possible combinations of multiple biomarkers. To tackle these issues, we propose a pairwise model fitting approach to estimating the distribution of the optimal linear combination of biomarkers and its diagnostic accuracy under the assumption of a multivariate normal distribution. The approach is evaluated in simulation studies and applied to data on chlamydia detection and COVID-19 diagnosis.

1 | Introduction

In biomedical and epidemiological studies, diagnostic biomarkers are often used to distinguish diseased individuals from healthy ones in a potentially fast and economical way. For biomarkers with continuous results, the receiver operating characteristic (ROC) curve is a standard statistical tool to evaluate and compare their accuracy, consider Hanley and McNeil [1], Wieand et al. [2], Zou and Hall [3], Yang, Lu, and Zhao [4], and Yin and Tian [5], among others. The ROC curve of a biomarker is a plot of its sensitivity versus 1 minus specificity. The area or partial area under the ROC curve (AUC or pAUC) is a popular summary index, with large values indicating high diagnostic accuracy.

Although widely used, a single biomarker is usually not accurate enough and may incur considerable classification errors. It can be beneficial to use multiple biomarkers simultaneously so that one may obtain a combined biomarker with improved diagnostic accuracy. Under the assumption of multivariate normality, Su and Liu [6] derived the optimal linear combination of multiple biomarkers based on AUC. Following their work, Liu, Schisterman, Zhu [7] developed an optimal linear combination method by maximizing pAUC.

In many cases, it is possible that disease data are not available at individual subject level, due to cost and/or privacy consideration. In these situations, group testing has been advocated as

Jin Yang and Wei Zhang are the co-first authors.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Published 2024. This article is a U.S. Government work and is in the public domain in the USA. *Statistics in Medicine* published by John Wiley & Sons Ltd.

an efficient alternative to reduce cost and protect confidentiality. The method was first proposed by Dorfman [8] to screen for syphilitic antigen in the U.S. army recruits based on pooled blood samples. Moreover, group testing has been shown to potentially improve statistical efficiency. Some works focus on the estimating the prevalence of disease, see Soberl and Elashoff [9], Le [10], Gastwirth and Hammick [11], Farrington [12], Hepworth [13], Hughes-Oliver and Rosenberger [14], Turner et al. [15], Warasi et al. [16], Malinovsky, Haber, and Albert [17]; some focus on individual classification, see Chen and Swallow [18], Brookmeyer [19], Zhang et al. [20], Black et al. [21], Hepworth and Watson [22], Zhang et al. [23]; and others focus on multiplex assays, see Tebbs, McMahan, and Bilder [24], Li, Liu, Xiong [25], Warasi et al. [26], among others.

A practical scenario for which our method is particularly useful is when the conventional “gold standard” tests are prohibitively expensive and time-consuming, making it desirable to develop less expensive and more convenient procedures, with reasonable accuracy. To the best of our knowledge, little work has been done in the literature on diagnostic biomarkers in the context of group testing scheme, Zhang et al. [27] proposed a nonparametric estimation approach for a single biomarker when only group-based test results on the disease status are available. In the present paper, we proposed a cost-effective alternative by combining multiple biomarkers that are less expensive, more convenient to measure, and offer higher diagnostic accuracy of the disease in the context of group testing. This problem has several challenges, including (1) unavailable individual disease statuses, (2) differential misclassification which may depend on the group size and the number of diseased individuals in the group [28], and (3) extensive computational demands due to the large number of possible combinations of multiple biomarkers. To tackle these issues, we propose to estimate the distributions of biomarkers in the diseased and non-diseased populations based on the pairwise bivariate fitting procedure of Fieuws and Verbeke [29] and Kassahun–Yimer et al. [30]. Furthermore, we will construct an estimator of the ROC curve for the optimal linear combination of multiple biomarkers, providing a robust approach for quick and accurate disease diagnosis.

The article is organized as follows. We present the estimation strategies for the distributions of multiple biomarkers in Section 2 and the ROC curve of their optimal linear combination in Section 3. Extensive simulation studies are conducted in Section 4 to compare the statistical efficiency of estimators based on group and individual testings. In Section 5, we illustrate our methods with data on chlamydia detection from NHANES and COVID-19 diagnosis from UK Biobank. Conclusions and further research directions are given in Section 6.

2 | Methodology

2.1 | Likelihood Function of the Multivariate Model

We assume the values of biomarkers can be observed on each individual subject. However, the binary disease status (e.g., disease and non-disease) is only observed at group level, that is, the binary testing is only performed on pooled samples rather than

individual sample. Our aim is to find the linear combination of the individual biomarkers that maximize the AUC.

To be specific, let $\mathcal{X} = (X_1, \dots, X_m)^T$ denote the measurements of m biomarkers with the distribution H and the probability density function (pdf) h , and the symbol “ T ” standing for the transpose of a vector or matrix. Let D be the true binary status of a disease with the prevalence $\Pr(D = 1) = p$ and K be the observed disease status with specificity $\delta_0 = \Pr(K = 0|D = 0)$ and sensitivity $\delta_1 = \Pr(K = 1|D = 1)$, respectively. In this article, our main interest is to estimate the diagnostic accuracy of multiple biomarkers. To this end, we need to estimate the conditional distributions of \mathcal{X} in non-diseased and diseased population, which are denoted by F and G , respectively. Let f and g be the corresponding density functions of F and G . It is easy to get $H = (1 - p)F + pG$.

Suppose that F and G are multivariate normal distributions, that is,

$$\begin{aligned} \mathcal{X}^T | D = 0 &\sim N(\boldsymbol{\mu}^{(0)}, \boldsymbol{\Sigma}^{(0)}) = F \\ \mathcal{X}^T | D = 1 &\sim N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}^{(1)}) = G \end{aligned} \quad (1)$$

with mean vectors $\boldsymbol{\mu}^{(0)} = (\mu_1^{(0)}, \dots, \mu_m^{(0)})^T$ and $\boldsymbol{\mu}^{(1)} = (\mu_1^{(1)}, \dots, \mu_m^{(1)})^T$, and covariance matrices

$$\boldsymbol{\Sigma}^{(0)} = \left\{ \sigma_{tt'}^{(0)} \right\}, \boldsymbol{\Sigma}^{(1)} = \left\{ \sigma_{tt'}^{(1)} \right\}, \text{ where } t, t' = 1, \dots, m$$

Consider N subjects that are randomly divided into n groups with sizes $J_i, i = 1, \dots, n$, where $J_1 + \dots + J_n = N$. The vector \mathcal{X} can be observed on each subject, yielding the observation $\{X_{ij,1}, \dots, X_{ij,m}, j = 1, \dots, J_i, i = 1, \dots, n\}$. Denote $\mathbb{X}_{i,t} = (X_{i1,t}, \dots, X_{iJ_i,t})^T, t = 1, \dots, m$. Thus $(\mathbb{X}_{i,1}, \dots, \mathbb{X}_{i,m}) = ((X_{i1,1}, \dots, X_{iJ_i,1})^T, \dots, (X_{i1,m}, \dots, X_{iJ_i,m})^T)$, for $i = 1, \dots, n$. Denote the group-tested results of the disease by $\tilde{K}_i, i = 1, \dots, n$, and let D_{ij} be the true disease status of the j th subject in the i th group and define $\tilde{D}_i = \max\{D_{i1}, \dots, D_{iJ_i}\}$ which is the true disease status of group i . For each group i , we assume that the specificity of the test remains unchanged, that is, $\Pr(\tilde{K}_i = 0 | \tilde{D}_i = 0) = \delta_0$, and the sensitivity is differential which depends on the group size J_i and the number of diseased subjects d_i in the group, see Haber, Malinovsky, and Albert [28], Hwang [31], and Hung and Swallow [32], denoted by $\Pr(\tilde{K}_i = 1 | \tilde{D}_i = 1) = \delta_1^*(J_i, d_i)$.

Given the true group disease status \tilde{D}_i , we assume the biomarkers’ level $(\mathbb{X}_{i,1}, \dots, \mathbb{X}_{i,m})$ and group-tested results \tilde{K}_i are independent. Thus, we can obtain

$$h\{(\mathbb{X}_{i,1}, \dots, \mathbb{X}_{i,m}) | \tilde{K}_i = k, \tilde{D}_i = d\} = h\{(\mathbb{X}_{i,1}, \dots, \mathbb{X}_{i,m}) | \tilde{D}_i = d\}$$

where $k, d \in \{0, 1\}$. Denote $h\{(\mathbb{X}_{i,1}, \dots, \mathbb{X}_{i,m}), \tilde{K}_i\}$ is the joint density function of $(\mathbb{X}_{i,1}, \dots, \mathbb{X}_{i,m})$ and \tilde{K}_i . Let $I(\cdot)$ be an indicator function. Then the likelihood function for the observed data is given by

$$\begin{aligned} L &= \prod_{i=1}^n h\{(\mathbb{X}_{i,1}, \dots, \mathbb{X}_{i,m}), \tilde{K}_i = 0\}^{I(\tilde{K}_i=0)} \\ &\quad \prod_{i=1}^n h\{(\mathbb{X}_{i,1}, \dots, \mathbb{X}_{i,m}), \tilde{K}_i = 1\}^{I(\tilde{K}_i=1)} \end{aligned} \quad (2)$$

where

$$\begin{aligned}
& h\{\mathbb{X}_{i,1}, \dots, \mathbb{X}_{i,m}, \tilde{K}_i = 0\} \\
&= \delta_0(1-p)^{J_i} \prod_{j=1}^{J_i} f(X_{ij,1}, \dots, X_{ij,m}) \\
&+ \sum_{\{\nu_1, \dots, \nu_{J_i}\} \in \mathcal{A}} \left\{ \prod_{j=1}^{J_i} f(X_{ij,1}, \dots, X_{ij,m})^{1-\nu_j} g(X_{ij,1}, \dots, X_{ij,m})^{\nu_j} \right\} \\
&\times \{1 - \delta_1^*(J_i, d_i)\} p^{d_i} (1-p)^{J_i-d_i} \\
& h\{\mathbb{X}_{i,1}, \dots, \mathbb{X}_{i,m}, \tilde{K}_i = 1\} \\
&= (1-\delta_0)(1-p)^{J_i} \prod_{j=1}^{J_i} f(X_{ij,1}, \dots, X_{ij,m}) \\
&+ \sum_{\{\nu_1, \dots, \nu_{J_i}\} \in \mathcal{A}} \left\{ \prod_{j=1}^{J_i} f(X_{ij,1}, \dots, X_{ij,m})^{1-\nu_j} g(X_{ij,1}, \dots, X_{ij,m})^{\nu_j} \right\} \\
&\delta_1^*(J_i, d_i) p^{d_i} (1-p)^{J_i-d_i}
\end{aligned} \tag{3}$$

where $\mathcal{A} = \{\{\nu_1, \dots, \nu_{J_i}\} \mid \nu_j \in \{0, 1\}, \nu_1 + \dots + \nu_{J_i} = d_i > 0\}$.

In the present article, we focus on equal group size, which is most commonly used choice in practice. The results can be extended unequal size. Assume $J_1 = \dots = J_n = N/n \triangleq J$. With the likelihood function L , we can use the maximum likelihood estimation (MLE) approach to estimate F and G . The log-likelihood function for $\{(\mathbb{X}_{i,1}^T, \dots, \mathbb{X}_{i,m}^T), \tilde{K}_i\}^T : i = 1, \dots, n\}$ is given by

$$\begin{aligned}
& l_n\{p; \boldsymbol{\mu}^{(0)}, \boldsymbol{\Sigma}^{(0)}, \boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}^{(1)}\} \\
&= \sum_{i=1}^n I(\tilde{K}_i = 1) \log \left[(1-\delta_0)(1-p)^J \prod_{j=1}^J f(X_{ij,1}, \dots, X_{ij,m}) \right. \\
&+ \sum_{\{\nu_1, \dots, \nu_{J_i}\} \in \mathcal{A}} \left\{ \prod_{j=1}^J f(X_{ij,1}, \dots, X_{ij,m})^{1-\nu_j} g(X_{ij,1}, \dots, X_{ij,m})^{\nu_j} \right\} \\
&\left. \delta_1^*(J, d) p^d (1-p)^{J-d} \right] \\
&+ \sum_{i=1}^n I(\tilde{K}_i = 0) \log \left[\delta_0(1-p)^J \prod_{j=1}^J f(X_{ij,1}, \dots, X_{ij,m}) \right. \\
&+ \sum_{\{\nu_1, \dots, \nu_{J_i}\} \in \mathcal{A}} \left\{ \prod_{j=1}^J f(X_{ij,1}, \dots, X_{ij,m})^{1-\nu_j} g(X_{ij,1}, \dots, X_{ij,m})^{\nu_j} \right\} \\
&\left. \{1 - \delta_1^*(J, d)\} p^d (1-p)^{J-d} \right]
\end{aligned} \tag{4}$$

where

$$\begin{aligned}
& f(X_{ij,1}, \dots, X_{ij,m}) \\
&= \frac{1}{(2\pi)^{m/2} \{\det(\boldsymbol{\Sigma}^{(0)})\}^{1/2}} \\
&\exp \left[-\frac{1}{2} \{ \boldsymbol{\mathcal{X}}^{(0)} - \boldsymbol{\mu}^{(0)} \}^T \{ \boldsymbol{\Sigma}^{(0)} \}^{-1} \{ \boldsymbol{\mathcal{X}}^{(0)} - \boldsymbol{\mu}^{(0)} \} \right] \\
& g(X_{ij,1}, \dots, X_{ij,m}) \\
&= \frac{1}{(2\pi)^{m/2} \{\det(\boldsymbol{\Sigma}^{(1)})\}^{1/2}} \\
&\exp \left[-\frac{1}{2} \{ \boldsymbol{\mathcal{X}}^{(1)} - \boldsymbol{\mu}^{(1)} \}^T \{ \boldsymbol{\Sigma}^{(1)} \}^{-1} \{ \boldsymbol{\mathcal{X}}^{(1)} - \boldsymbol{\mu}^{(1)} \} \right]
\end{aligned}$$

2.2 | Estimation Procedure

It is not easy to estimate the parameters by directly maximizing (4), especially when m is large. To address this issue, we employ the pairwise fitting approach which uses the pairwise models to represent the original multivariate model, see Fieuws and Verbeke [29], Kassahun-Yimer et al. [30], and Fieuws, Verbeke, and Molenberghs [33]. The general idea is that the parameters in the full multivariate model can be identified from all pairwise models (the bivariate model for each pair of biomarkers). Related to pseudo-likelihood estimation (Besag [34]), the estimation procedure can be summarized as the following:

- Instead of maximizing the log-likelihood of the multivariate model, log-likelihoods of the following form will be maximized separately

$$\sum_{i=1}^n l_{rs,i}(\Omega_{r,s}) \tag{5}$$

where $r = 1, \dots, m-1, s = r+1, \dots, m$, and $\Omega_{r,s}$ is the vector containing all parameters in the bivariate model corresponding to the specific pair (r, s) , and the log-likelihood function for the r th and s th biomarkers can be written as

$$l_{rs} \left\{ p; \boldsymbol{\mu}_r^{(0)}, \boldsymbol{\mu}_s^{(0)}, \boldsymbol{\mu}_r^{(1)}, \boldsymbol{\mu}_s^{(1)}, \boldsymbol{\Sigma}_r^{(0)}, \boldsymbol{\Sigma}_s^{(0)}, \boldsymbol{\Sigma}_r^{(1)}, \boldsymbol{\Sigma}_s^{(1)} \right\} \tag{6}$$

which is the same as (4) with $f(X_{ij,1}, \dots, X_{ij,m})$ and $g(X_{ij,1}, \dots, X_{ij,m})$ replaced by $f(X_{ij,r}, X_{ij,s})$ and $g(X_{ij,r}, X_{ij,s})$, respectively.

- For simplicity, (5) can be rewritten as $\sum_{i=1}^n l_{p,i}(\Omega_p)$ with $p = 1, \dots, W$. So we note that there are a total of $W = m(m-1)/2$ possible pairs.
- Let $\hat{\Omega}_{rs}$ be the maximizer of $l_n(\Omega_{rs})$ with respect to Ω_{rs} and $\hat{\Omega}$ be the stacked vector containing all pair-specific parameter vectors $\hat{\Omega}_{rs}$. Then, the final estimator for each parameter in Ω can be obtained by averaging all pair-specific estimators in $\hat{\Omega}$. The estimators of F and G can be subsequently derived and are denoted by \hat{F} and \hat{G} .

The above procedure has the advantage when it is impossible or too time consuming to fit, as summarized by Fieuws and Verbeke [29]: “the pairwise approach is able to yield unbiased estimators with robust variances which indicates that it can capture the true sampling variability.” It is possible that the approach may result in an estimator of covariance matrix, that is not positive definite. That is because we are averaging pairwise covariance, particularly in large dimension. When the correlation are high, a non-positive definite covariance can occur. In such a case, we can shrink the off diagonal element toward zero resulting in a positive definite covariance matrix, see Devlin, Gnanadesikan, and Kettenring [35].

3 | ROC Curves and AUC

Without loss of generality, we assume that the values of the biomarkers in the diseased population tend to be larger than that in the non-diseased population. Consider a biomarker that follows the normal distributions $N(\mu_1^{(1)}, \{\sigma_1^{(1)}\}^2)$ and $N(\mu_1^{(0)}, \{\sigma_1^{(0)}\}^2)$

in the diseased and non-diseased population, respectively. The ROC curve of the biomarker is given by

$$\text{ROC}(q) = \Phi\{u_1 - v_1\Phi^{-1}(1 - q)\} \quad (7)$$

for $0 \leq q \leq 1$, where $u_1 = (\mu_1^{(1)} - \mu_1^{(0)})/\sigma_1^{(1)}$ and $v_1 = \sigma_1^{(1)}/\sigma_1^{(0)}$, Φ is the cumulative distribution function of the standard normal distribution.

For the case of multiple biomarkers, Su and Liu [6] gave an explicit form for the best linear combination β , which has the form $c \times \mu^T(\Sigma^{(1)} + \Sigma^{(0)})^{-1}$, where c is the reciprocal of the first element of $\mu^T(\Sigma^{(1)} + \Sigma^{(0)})^{-1}$ and $\mu = \mu^{(1)} - \mu^{(0)}$, see chapter 4 in Zou et al. [36]. Based on the results of Su and Liu [6], the ROC curve and AUC for the best linear combination of multiple biomarkers are given by

$$\text{ROC}(q) = \Phi\{u - v\Phi^{-1}(1 - q)\} \quad (8)$$

and

$$\text{AUC} = \Phi\left\{\sqrt{\mu^T(\Sigma^{(0)} + \Sigma^{(1)})^{-1}\mu}\right\} \quad (9)$$

where $u = \frac{\beta^T(\mu^{(1)} - \mu^{(0)})}{\sqrt{\beta^T\Sigma^{(0)}\beta}}$ and $v = \frac{\sqrt{\beta^T\Sigma^{(1)}\beta}}{\sqrt{\beta^T\Sigma^{(0)}\beta}}$.

4 | Simulations

In evaluating the performance of our proposed approach through simulations, we have two goals in mind: (1) to compare multiple biomarkers versus single biomarker and (2) to compare group-tested results versus individual-tested results. Due to study budget and other constraints, it may not be always possible to include all N subjects in testing. As a result, we considered two scenarios when we approached the individual tested data, with the first using individual tested results of all $N = nJ$ subjects (which are called *full individual testing*) and the second using a random sample of group n from the nJ subjects (which are called *random individual testing*).

In this section, we consider three biomarkers. The group size J is chosen from $\{1, 2, 5\}$, where $J = 1$ corresponds to individual testing. The sensitivity δ_1 and specificity δ_0 were selected from $\{0.90, 0.95, 1.00\}$. We specified δ_1^* using the model of Hung and Swallow [32] as $\delta_1^* = \delta_1 d / \{d + \lambda(J - d)\}$, where d represents the number of diseased individuals in a group, with $\lambda = 0.02$. The prevalence p is set to be 0.01 or 0.02. For a given sample size $N = nJ$, we generated the random sample of the biomarkers $(\mathbb{X}_{1,1}^T, \dots, \mathbb{X}_{1,m}^T)$ and disease status \tilde{K} using the following procedure. We first individually simulated the true disease status $\{D_{ij} : 1 \leq i \leq n; 1 \leq j \leq J\}$ for all subjects from a Bernoulli distribution with probability p . For the subjects with $D = 0$, values of biomarkers were generated from the multivariate normal distribution with mean $(0, 0, 0)$ and covariance matrix with diagonal elements 1, 1, and 1 and pairwise correlation coefficients $1/3, 1/3$, and $1/3$; otherwise, they were generated from the multivariate normal distribution with mean $(1.6, 1.5, 1.2)$ and covariance matrix with diagonal elements $1.4^2, 1.5^2$, and 1.3^2 and pairwise correlation coefficients 0.5, 0.6, and 0.3, respectively. Then we randomly divided the N subjects into n groups of

size J . We generated the group-tested result (\tilde{K}) from a Bernoulli distribution with the probability $1 - \delta_0$ for the groups with all $D = 0$ and the probability δ_1^* for the groups with at least one $D = 1$. The true value of AUC for the three biomarkers and the best linear combination of the biomarkers are 0.8238, 0.7973, 0.7678, and 0.8579, respectively.

We assessed the proposed estimation approach under varying number of groups (*full individual* vs. *random individual testing*), group size ($J = 1, 2, 5$), prevalence ($p = 0.01, 0.02$), and misclassification rate ($1 - \delta_0 = 1 - \delta_1 = 0, 0.05, 0.1$). For a more comprehensive comparison, we also consider the AUCs of individual biomarkers based on group-tested data. The evaluations were based on commonly used criteria including bias, variance, 95% coverage probability (CP) and the average confidence interval length (ACIL). We used the bootstrap procedure which resamples the group-level observations $\{(\mathbb{X}_{i,1}, \dots, \mathbb{X}_{i,m}), \tilde{K}_i\} : i = 1, \dots, n\}$ with replacement to estimate CP and ACIL. Throughout, 200 simulation data were obtained and 300 bootstrap replicates were generated within each simulation.

In the scenario of *full individual testing*, we fixed the total number of subjects at $N = 15\,000$. Table 1 presents the performance of the prevalence estimator in a finite sample setting. Regardless of the true prevalence level (across the two row blocks), number of biomarkers (across the columns), or misclassification rate (across the rows within each row block), the estimates are all close to the true values, with coverage probabilities close to the nominal level. The estimator exhibits increased statistical efficiency (indicated by smaller variance or shorter ACIL) as the misclassification error ($1 - \delta_0$ and $1 - \delta_1$) decreases, resulting in consistent findings with Zhang et al. [27] who focused on a single biomarker. For example, when prevalence is $p = 0.02$, misclassification rate are both equal to 0.1 ($\delta_1 = \delta_0 = 0.9$), and the group size is $J = 5$, the relative efficiency of the estimator under our proposed approach is about 0.77 (0.0317/0.0410) in comparison to $J = 1$. Moreover, the estimators based on the combined biomarker can improve the efficiency over that based on a single biomarker [27]. For example, when $p = 0.02$, $\delta_1 = \delta_0 = 0.9$, and $J = 5$, the relative efficiency of the estimator based on the combined biomarker is 0.76, which is less than 1. In essence, utilizing multiple biomarkers with group-tested data achieves double efficiency gains, one from group-tested data in the presence of misclassification and the other from multiple biomarkers.

Table 2 summarizes the performance of the AUC estimator in a finite sample setting. As expected, combining biomarker (“Combined”) yields larger AUC estimates, indicating improved diagnostic accuracy compared to individual biomarkers (“Individual”). For example, when $p = 0.01$, $\delta_0 = \delta_1 = 0.9$, and $J = 2$, the AUCs of the single biomarkers are 0.8304, 0.8096 and 0.7730, respectively, and the AUC of combined biomarker is 0.8721. Similar to the findings for the prevalence estimator, as the misclassification error decreases, the variance of the estimators decreases. Furthermore, the AUC estimator of the combined biomarkers is always more efficient than their single biomarker counterparts, as its variance is smaller than all those in the individual setting. Focusing on the estimators based on the combined biomarker, we can see that, when there is

TABLE 1 | Simulation results for the prevalence estimator based on the group and full individual testing approaches: estimate (Est), bias (Bias), variance (Var), coverage probability (CP) and average confidence interval length (ACIL) of the estimators for the individual (B-1, B-2, B-3) and combined biomarkers.

		$p = 0.01$															
		Individual						Combined									
		B-1		B-2		B-3		B-1		B-2		B-3		Combined			
$\delta_0 = \delta_1$	J	Est	Bias	Var	CP (ACIL)	Est	Bias	Var	CP (ACIL)	Est	Bias	Var	CP (ACIL)	Est	Bias	Var	CP (ACIL)
0.90	1	0.9979	-0.0021	0.0672	94.00% (0.0107)	0.9907	-0.0093	0.0674	92.00% (0.0108)	1.0002	0.0002	0.0880	91.50% (0.0118)	1.0074	0.0074	0.0258	95.00% (0.0069)
	2	0.9915	-0.0085	0.0470	95.50% (0.0090)	0.9835	-0.0165	0.0502	93.00% (0.0089)	0.9941	-0.0059	0.0449	95.50% (0.0093)	1.0066	0.0066	0.0245	95.50% (0.0065)
	5	1.0059	0.0059	0.0310	94.00% (0.0070)	1.0003	0.0003	0.0309	95.00% (0.0070)	1.0067	0.0067	0.0316	96.50% (0.0072)	1.0107	0.0107	0.0238	93.00% (0.0059)
0.95	1	1.0063	0.0063	0.0414	90.45% (0.0076)	0.9863	-0.0137	0.0422	91.96% (0.0078)	1.0008	0.0008	0.0423	94.47% (0.0084)	1.0041	0.0041	0.0201	95.98% (0.0058)
	2	1.0001	0.0001	0.0254	96.50% (0.0065)	0.9968	-0.0032	0.0267	95.00% (0.0065)	1.0000	0.0000	0.0252	95.50% (0.0066)	1.0086	0.0086	0.0180	94.00% (0.0053)
	5	1.0106	0.0106	0.0160	94.95% (0.0052)	1.0075	0.0075	0.0153	96.46% (0.0052)	1.0089	0.0089	0.0154	95.96% (0.0052)	1.0129	0.0129	0.0137	94.44% (0.0047)
1.00	1	1.0063	0.0063	0.0074	95.50% (0.0031)	1.0063	0.0063	0.0073	95.50% (0.0031)	1.0060	0.0060	0.0073	95.50% (0.0031)	1.0144	0.0144	0.0067	95.00% (0.0030)
	2	1.0065	0.0065	0.0075	96.00% (0.0032)	1.0064	0.0064	0.0075	95.50% (0.0032)	1.0064	0.0064	0.0075	95.50% (0.0032)	1.0137	0.0137	0.0070	95.50% (0.0031)
	5	1.0078	0.0078	0.0080	94.97% (0.0033)	1.0073	0.0073	0.0080	94.47% (0.0034)	1.0077	0.0077	0.0081	93.97% (0.0034)	1.0136	0.0136	0.0072	93.97% (0.0032)

		$p = 0.02$															
		Individual						Combined									
		B-1		B-2		B-3		B-1		B-2		B-3		Combined			
$\delta_0 = \delta_1$	J	Est	Bias	Var	CP (ACIL)	Est	Bias	Var	CP (ACIL)	Est	Bias	Var	CP (ACIL)	Est	Bias	Var	CP (ACIL)
0.90	1	1.9949	-0.0051	0.0787	97.50% (0.0122)	1.9876	-0.0124	0.0826	94.50% (0.0121)	1.9965	-0.0035	0.0902	95.00% (0.0129)	1.9981	-0.0019	0.0410	95.00% (0.0086)
	2	2.0027	0.0027	0.0536	96.00% (0.0097)	1.9904	-0.0096	0.0572	95.00% (0.0099)	1.9992	-0.0008	0.0588	96.50% (0.0100)	2.0007	0.0007	0.0341	96.00% (0.0079)
	5	1.9995	-0.0005	0.0420	94.50% (0.0079)	1.9963	-0.0037	0.0419	93.50% (0.0079)	2.0045	0.0045	0.0419	94.00% (0.0079)	2.0077	0.0077	0.0317	94.00% (0.0070)
0.95	1	2.0024	0.0024	0.0559	91.50% (0.0087)	1.9897	-0.0103	0.0506	93.00% (0.0088)	2.0011	0.0011	0.0579	92.00% (0.0091)	1.9979	-0.0021	0.0341	94.50% (0.0069)
	2	1.9948	-0.0052	0.0335	93.50% (0.0073)	1.9918	-0.0082	0.0340	95.00% (0.0073)	1.9963	-0.0037	0.0339	93.50% (0.0073)	2.0021	0.0021	0.0277	92.50% (0.0064)
	5	1.9985	-0.0015	0.0236	96.00% (0.0062)	2.0010	0.0010	0.0243	95.50% (0.0062)	2.0014	0.0014	0.0246	95.50% (0.0063)	2.0066	0.0066	0.0220	94.50% (0.0058)
1.00	1	2.0005	0.0005	0.0139	94.50% (0.0044)	2.0002	0.0002	0.0139	94.00% (0.0044)	1.9998	-0.0002	0.0139	95.00% (0.0044)	2.0108	0.0108	0.0132	94.00% (0.0043)
	2	2.0006	0.0006	0.0141	96.00% (0.0045)	2.0008	0.0008	0.0142	95.50% (0.0045)	2.0011	0.0011	0.0142	95.00% (0.0045)	2.0097	0.0097	0.0133	95.00% (0.0044)
	5	1.9995	-0.0005	0.0165	92.50% (0.0047)	2.0002	0.0002	0.0163	93.50% (0.0047)	2.0007	0.0007	0.0166	93.00% (0.0047)	2.0084	0.0084	0.0155	93.00% (0.0046)

Note: Entries of Est and Bias are multiplied by 100, and entries of Var are multiplied by 10000 for better presentation. p is the prevalence, δ_0 and δ_1 are the specificity and sensitivity, J is the group size, and B-1, B-2, B-3 stand for individual biomarker 1, 2, 3, respectively.

TABLE 2 | Simulation results for the AUC estimator based on the group and *full individual testing* approaches: estimate (Est), bias (Bias) and variance (Var), coverage probability (CP) and average confidence interval length (ACIL) of the estimators for the individual (B-1, B-2, B-3) and combined biomarkers.

p = 0.01															
Individual															
B-1				B-2				B-3				Combined			
$\delta_0 = \delta_1$	<i>J</i>	Est	Bias	Var	CP (ACIL)	Est	Bias	Var	CP (ACIL)	Est	Bias	Var	CP (ACIL)		
0.90	1	0.8317	0.0079	0.0558	87.50% (0.2676)	0.8086	0.0113	0.0545	91.00% (0.2818)	0.7811	0.0133	0.0708	91.00% (0.3291)		
	2	0.8304	0.0066	0.0426	91.00% (0.2546)	0.8096	0.0123	0.0615	86.00% (0.2557)	0.7730	0.0052	0.0666	84.00% (0.2860)		
0.95	5	0.8219	-0.0019	0.0406	89.50% (0.2309)	0.8087	0.0114	0.0418	88.50% (0.2347)	0.7742	0.0064	0.0580	89.00% (0.2834)		
	1	0.8302	0.0064	0.0371	86.43% (0.2135)	0.8057	0.0084	0.0413	89.95% (0.2314)	0.7792	0.0114	0.0455	90.45% (0.2540)		
1.00	2	0.8249	0.0011	0.0262	91.00% (0.2035)	0.8015	0.0042	0.0316	92.00% (0.2102)	0.7687	0.0009	0.0360	94.00% (0.2359)		
	5	0.8231	-0.0007	0.0234	93.43% (0.1924)	0.7993	0.0020	0.0265	93.94% (0.1963)	0.7709	0.0031	0.0320	94.95% (0.2246)		
1.00	1	0.8231	-0.0007	0.0042	91.00% (0.0760)	0.7963	-0.0010	0.0040	97.00% (0.0839)	0.7690	0.0012	0.0049	93.00% (0.0827)		
	2	0.8264	0.0026	0.0057	94.00% (0.0898)	0.7969	-0.0004	0.0055	95.50% (0.0974)	0.7671	-0.0007	0.0076	94.00% (0.1043)		
	5	0.8211	-0.0027	0.0093	96.98% (0.1234)	0.8029	0.0056	0.0098	91.96% (0.1276)	0.7682	0.0004	0.0154	91.96% (0.1491)		

p = 0.02															
Individual															
B-1				B-2				B-3				Combined			
$\delta_0 = \delta_1$	<i>J</i>	Est	Bias	Var	CP (ACIL)	Est	Bias	Var	CP (ACIL)	Est	Bias	Var	CP (ACIL)		
0.90	1	0.8285	0.0047	0.0190	93.50% (0.1791)	0.8021	0.0048	0.0207	94.50% (0.1867)	0.7744	0.0066	0.0255	93.00% (0.2018)		
	2	0.8212	-0.0026	0.0182	91.50% (0.1548)	0.8029	0.0056	0.0162	92.50% (0.1642)	0.7687	0.0009	0.0207	94.00% (0.1817)		
0.95	5	0.8187	-0.0051	0.0132	95.00% (0.1487)	0.8016	0.0043	0.0159	93.00% (0.1483)	0.7615	-0.0063	0.0213	91.50% (0.1809)		
	1	0.8248	0.0010	0.0128	92.00% (0.1381)	0.8002	0.0029	0.0123	95.00% (0.1420)	0.7729	0.0051	0.0134	96.00% (0.1508)		
1.00	2	0.8234	-0.0004	0.0101	94.00% (0.1253)	0.8011	0.0038	0.0117	92.00% (0.1287)	0.7718	0.0040	0.0105	97.00% (0.1359)		
	5	0.8251	0.0013	0.0091	92.00% (0.1178)	0.8005	0.0032	0.0112	93.50% (0.1255)	0.7667	-0.0011	0.0135	94.50% (0.1493)		
1.00	1	0.8234	-0.0004	0.0021	92.50% (0.0541)	0.7959	-0.0014	0.0020	93.50% (0.0591)	0.7679	0.0001	0.0022	95.00% (0.0587)		
	2	0.8227	-0.0011	0.0032	92.50% (0.0652)	0.7958	-0.0015	0.0032	93.50% (0.0703)	0.7660	-0.0018	0.0034	96.00% (0.0746)		
	5	0.8239	0.0001	0.0054	89.50% (0.0871)	0.7971	-0.0002	0.0063	93.00% (0.0921)	0.7674	-0.0004	0.0078	93.00% (0.1079)		

Note: Entries of Var are multiplied 10 for better presentation. *p* is the prevalence, δ_0 and δ_1 are specificity and sensitivity, *J* is the group size, and B-1, B-2, B-3 stand for the individual biomarker 1, 2, 3, respectively.

no misclassification ($\delta_0 = \delta_1 = 1.00$), the superiority of group testing disappears, with yielding larger variance than that based on individual testing. Moreover, the AUC estimates based on group testing are close to that based on individual testing. However, when misclassification exists ($\delta_0, \delta_1 < 1.00$), there is always an estimator based on group testing that yields smaller variance than the individual testing counterpart. For example, when $p = 0.01$, $\delta_0 = \delta_1 = 0.95$, the AUC estimator achieves the smallest variance at $J = 5$. These results suggest that the AUC estimators based on group testing can be more efficient than those based on individual testing when the test is subject to misclassification.

Moreover, we tested the smaller sample size $N = 10\,000$ with larger prevalence $p = 0.05, 0.1$, and keep the remaining settings, see the result in Appendix A of the Supporting Information.

In summary, the simulations demonstrate these findings: (1) the proposed approach of using multiple biomarkers can improve both accuracy and statistical efficiency in the AUC estimates, compared to those of using a single biomarker, (2) as misclassification rates decrease, the proposed estimator becomes more efficient; and (3) when the test is subject to misclassification, the AUC estimates based on group testing can be superior (in both accuracy and efficiency) to those based on individual testing.

Besides that, we added a new simulation where the sensitivity and specificity are mis-specified and have reported the corresponding results in Appendix B of Supporting Information. In addition, the findings are similar in the scenario of *random individual testing* where the number of groups n is fixed at 8000 (for both $J = 1$ and $J = 2, 5$). We presented the result in Appendix C of Supporting Information.

5 | Applications

5.1 | COVID-19 Detection

We first applied the proposed approach to detecting COVID-19 with data from the UK Biobank, an international health resource enabling research into the genetic and lifestyle determinants of common diseases. Over 500 000 participants (aged 50–81 on March 16, 2020) from the UK general population were recruited between 2006 and 2010 (aged 40–69), see Satter et al. [37]. Data were available for the period February 25, 2021 to July 12, 2021.

Blood samples and symptoms information of the participants were collected and analyzed in this study. In our analysis, we used the self-antibody test from the UK Biobank. This test recruited UK Biobank participants to perform a SARS-CoV-2 antibody self-test using a lateral flow device (Fortress Fast COVID-19 device) at home and to report their result to UK Biobank. We focused on participants whose disease statuses are obtained using IgM (https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/c19_antibody_p1_overview.pdf).

Total bilirubin (TBIL), aspartate aminotransferase (AST), and alanine aminotransferase (ALT) levels are three common clinical measurements which are often used in the studies of COVID-19, see Liu et al. [38], Ali [39], and Kasapoglu et al. [40]. For illustration, we used them as the biomarkers for COVID-19 infection. While the UK Biobank did not employ group testing to detect COVID-19, we presented a hypothetical scenario using group-tested data. This approach is justifiable because the self-antibody test and the three biomarkers relied on different specimens. We simulated group-tested outcomes to determine the presence of COVID-19 independently. To achieve this, we considered the testing results in the dataset as the true disease statuses of the subjects and randomly assigned the self-antibody test specimens to groups of size K . The values of the three biomarkers for each subject remained unchanged.

To apply our proposed method, we first transformed the data to normality using the Box-Cox transformation. After removing the missing values of IgM, TBIL, ALT, and AST, $N = 11,837$ independent observations of $(\mathcal{X}, K)^T$ were included in the final analysis, out of which 548 subjects tested positive for IgM in the self-antibody test. To investigate the diagnostic ability of the three biomarkers (TBIL, ALT, and AST), we estimated the ROC curve of their best linear combination, which required the distribution of the combination in the non-infected and infected population. For comparison, we also calculated the AUC estimator of the best linear combination of the three biomarkers based on individual-tested results. Based on the UK Biobank data, we set the sensitivity and specificity of the IgM test to be $\delta_1 = 0.952$ and $\delta_0 = 0.960$ (see <https://biobank.ndph.ox.ac.uk/showcase/refer.cgi?id=4513>), and assumed that $\delta_1^*(J, d) = \delta_1 d / \{d + \lambda(J - d)\}$ with $\lambda = 0.02$.

Table 3 presents the AUC estimators of the three individual biomarkers and their best linear combination based on the group-tested results and individual-tested results. The variance of the AUC estimator is calculated using 1000 bootstrap

TABLE 3 | Analysis of the UK Biobank COVID-19 data: estimates (Est) and variances (Var) of AUC for the three individual biomarker (TBIL, ALT, and AST) and their best linear combination based on individual ($J = 1$) and group testing ($J = 2, 5$) approaches.

J	Individual						Combined		
	TBIL		ALT		AST		Est	Var	95% CI
	Est	Var	Est	Var	Est	Var			
1	0.5318	0.0029	0.5213	0.0021	0.5218	0.0022	0.5730	0.0057	(0.5264, 0.6196)
2	0.5189	0.0021	0.5178	0.0017	0.5209	0.0021	0.5616	0.0041	(0.5217, 0.6015)
5	0.5278	0.0043	0.5316	0.0044	0.5260	0.0034	0.5914	0.0079	(0.5362, 0.6467)

Note: Entries of Var are multiplied 10 for better presentation. J is the size of each group.

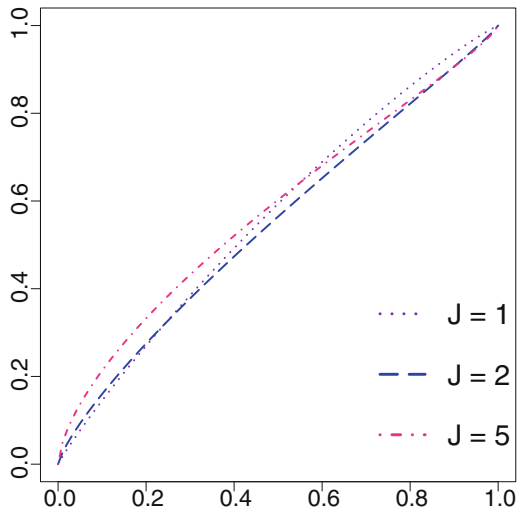


FIGURE 1 | ROC curve estimates of the best linear combination of the three biomarkers (TBIL, ALT, and AST) for the group sizes of $J = 1, 2, 5$, using data from UK Biobank.

replicates. From the table, it can be found that the AUC estimator of the best combination is always larger than that of individual biomarkers. Among the group sizes considered, the best efficiency is attained by the AUC estimator under $J = 2$. The ROC curve of the best linear combination of the three biomarkers is displayed in Figure 1.

5.2 | Chlamydia Detection

Another real data analysis concerns genital chlamydia infections and is conducted data from the National Health and Nutrition Examination Survey (NHANES; <https://www.cdc.gov/nchs/nhanes/index.htm>), a large-scale population study aimed at evaluating the health and nutritional well-being of individuals across the United States.

The NHANES study collected urine samples from participants aged 18–39 and tests for genital chlamydia infections using the DNA strand displacement amplification method. The publicly released data includes assay results of eligible participants. Chlamydia, caused by *Chlamydia trachomatis*, is a common sexually transmitted disease that can affect the levels of monocyte, neutrophils, and albumin, as reported in Datta et al. [41], Naglak, Morrison, and Morrisons [42], and Stoke and Isada [43].

In this analysis, we considered using monocyte, neutrophils, and albumin as biomarkers for chlamydia infections.

We collected data on chlamydia and three biomarkers including monocyte, neutrophils, and albumin from six consecutive and independent surveys of NHANES (1999–2000, 2001–2002, 2003–2004, 2005–2006, 2007–2008, 2009–2010). To address the potential impact of oversampling and the complex survey design, we performed resampling on the data from each two-year survey dataset. This resampling was done with replacement, using sampling weights proportional to the probabilities, while keeping the sample size consistent with the original dataset. Subsequently, we combined these resampled datasets to create a large sample.

After excluding subjects with missing values for chlamydia, monocyte, neutrophils, and albumin, our final working dataset consisted of $N = 12\,424$ independent observations of $(\mathcal{X}, K)^T$. Among these observations, 221 subjects tested positive for chlamydia. Similar to the UK Biobank data, the NHANES study did not utilize group testing to detect chlamydia. Since the detection of chlamydia infections and the measurement of monocyte, neutrophils, and albumin used different specimens, we generated group-tested outcomes for disease presence independently. We achieved this by considering the testing results available in the current dataset. Similarly, we use the Box-Cox transformation to transform the original data of monocyte, neutrophils, and albumin. We estimated the ROC curve for the optimal linear combination of the three biomarkers using a process similar to the COVID-19 example, except that we assumed a specificity of $\delta_0 = 0.99$ and a sensitivity of $\delta_1 = 0.9$, and assumed that $\delta_1^*(J, d) = \delta_1 d / \{d + \lambda(J - d)\}$ with $\lambda = 0.02$.

The AUC estimators for individual biomarkers and their best linear combination, based on the group-tested and individual-tested results, are presented in Table 4. The variance of the AUC estimator is computed using 1000 bootstrap replicates. The table reveals that the AUC estimates for the best combination consistently outperform those for individual biomarkers, with an average increase of 10.31%. The AUC estimator achieves the best efficiency when $J = 2$, with a relative efficiency of 0.92 compared to the individual-tested results ($J = 1$). The ROC curve of the best linear combination of the three biomarkers is displayed in Figure 2.

TABLE 4 | Analysis of the chlamydia data: estimates (Est) and variances (Var) of AUC for the three individual biomarker (monocyte, neutrophils and albumin) and their best linear combination based on individual ($J = 1$) and group testing ($J = 2, 5$) approaches.

J	Individual						Combined		
	Monocyte		Neutrophils		Albumin		Est	Var	95% CI
	Est	Var	Est	Var	Est	Var			
1	0.5333	0.0052	0.5734	0.0191	0.5697	0.0076	0.6345	0.0076	(0.5804, 0.6885)
2	0.5267	0.0039	0.5594	0.0135	0.5293	0.0054	0.5859	0.0070	(0.5339, 0.6379)
5	0.5340	0.0070	0.5957	0.0133	0.5490	0.0102	0.6051	0.0089	(0.5466, 0.6636)

Note: Entries of Var are multiplied 10 for better presentation. J is the size of each group.

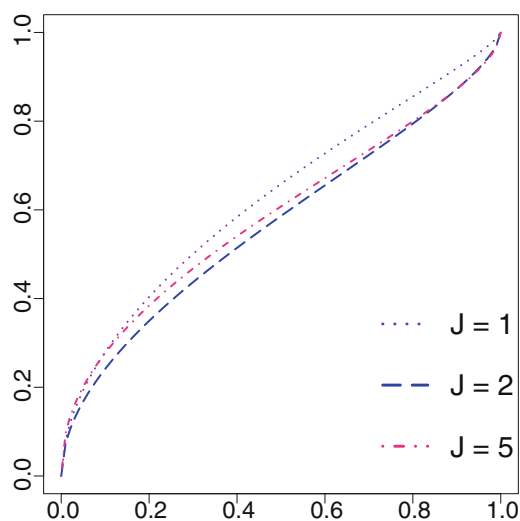


FIGURE 2 | ROC curve estimates of the best linear combination of the three biomarkers (Monocyte, Neutrophils, and Albumin) for the group sizes of $J = 1, 2, 5$, using data from the NHANES study.

6 | Discussion

In this article, we consider the problem of combining multiple biomarkers to improve the diagnostic accuracy using AUC and ROC curve as a criterion, when only group-based test results on the disease status are available while biomarker value at individual level can be observed.

To ease computation, we proposed a pairwise model fitting approach to estimate the conditional distributions of the optimal linear combination of the multiple biomarkers given the disease status and its diagnostic accuracy. This approach is advantageous whenever fitting the full multivariate model is not possible or too time-consuming. Extensive simulations show that the proposed estimators based on group testing can improve statistical efficiency over the individual testing in the presence of misclassification.

Here, we focus on normal distributions since obtaining the optimal linear combination and its corresponding AUC is challenging for a general multivariate distribution even the most common ones. This is because even for the most common multivariate distributions, it is very difficult to derive the distribution of a linear combination of the individual variate. The multivariate normal distributions are the only exception since it is well known that a linear combination of their variates is also normal, thus warrants an explicit form of the optimal linear combination and its AUC. It is worth pointing out that treating non-normal biomarkers as normal biomarkers may lead to substantial bias, see the results in Appendix D of Supporting Information.

In addition to the measures discussed in this article, alternative summary measures of a biomarker's diagnostic capability, such as the Youden index, offer a different perspective. The Youden index, represented as $J = \max_c \text{Sensitivity}(c) + \text{Specificity}(c) - 1$, quantifies the maximum potential diagnostic effectiveness of a biomarker by assigning equal importance to sensitivity and specificity. Exploring the extension of the findings presented

in this article to other diagnostic measures warrants further investigation and research.

We assumed that the misclassification of the single and pooled assay for the disease outcome is known. It is worth noting that mis-specifying the sensitivity and specificity may yield poor estimation. Haber et al. [28] showed the importance of estimating these misclassification processes in the population under study. See also sensitivity analysis results in Appendix B of the Supporting Information. Future research will focus on the efficient design of a validation study to estimate these processes.

Cautions need to be taken that when the prevalence is low and the sample size is small, resulting in a relatively small number of groups tested positive, the confidence intervals may have coverage below the expected level and may not be reliable.

Our work focuses on the Master pool testing. Other group testing protocols have been developed in the literature, such as array testing and master pool testing with retesting. With more data available from further testing, we expect that more efficient estimation for the ROC and its AUC can be obtained. Extending our method to other types of group testing data appears to be technically more challenging, and warrants further investigation.

Finally, if measuring the biomarkers' levels is costly or time consuming, one could also consider pooling biospecimen to measure the biomarkers to save cost and time. In this case, individual level is not observed on the biomarkers, but rather one observes an average of a biomarker's levels from individuals in the pool. This additional feature in the data structure creates more technical difficulty in deriving the likelihood, the ROC and its AUC. Future research is warranted in this direction.

Acknowledgments

Research of J. Yang, A. Liu, and Z. Chen were supported by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). Research of W. Zhang was supported by the National Natural Science Foundation of China (NSFC: # 12001522 and # 72091212) and National Key R&D Program of China (# 2022YFA1004800). Research of P. Albert was supported by the Intramural Research Program of the National Cancer Institute (NCI). This research has been conducted using the UK Biobank Resource under Application Number 86711.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The COVID-19 data that support the findings of this study are available from UK Biobank. Restrictions apply to the availability of these data, which were used under the license for this study. Data are available at <https://www.ukbiobank.ac.uk/> with the permission of UK Biobank. The chlamydia data that support the findings of this study are available in National Health and Nutrition Examination Survey (NHANES) at <https://www.cdc.gov/nchs/nhanes/index.htm>.

References

1. J. A. Hanley and B. J. McNeil, "The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve," *Radiology* 143 (1982): 29–36.

2. S. Wieand, M. H. Gail, B. R. James, and K. L. James, "A Family of Nonparametric Statistics for Comparing Diagnostic Markers With Paired or Unpaired Data," *Biometrika* 76 (1989): 585–592.
3. K. H. Zou and W. J. Hall, "Semiparametric and Parametric Transformation Models for Comparing Diagnostic Markers With Paired Design," *Journal of Applied Statistics* 29 (2002): 803–816.
4. H. Yang, K. Lu, and Y. Zhao, "A Nonparametric Approach for Partial Areas Under ROC Curves and Ordinal Dominance Curves," *Statistica Sinica* 27, no. 1 (2017): 357–371.
5. J. Yin and L. Tian, "Joint Confidence Region Estimation for Area Under ROC Curve and Youden Index," *Statistics in Medicine* 33, no. 6 (2014): 985–1000.
6. J. Q. Su and J. S. Liu, "Linear Combinations of Multiple Diagnostic Markers," *Journal of the American Statistical Association* 88 (1993): 1350–1355.
7. A. Liu, E. Schisterman, and Y. Zhu, "On Linear Combinations of Biomarkers to Improve Diagnostic Accuracy," *Statistics in Medicine* 24, no. 1 (2005): 37–47.
8. R. Dorfman, "The Detection of Defective Members of Large Populations," *Annals of Mathematical Statistics* 14 (1943): 436–440.
9. M. Sobel and R. M. Elashoff, "Group Testing With a New Goal, Estimation," *Biometrika* 62, no. 1 (1975): 181–193.
10. C. T. Le, "A New Estimator for Infection Rates Using Pools of Variable Size," *American Journal of Epidemiology* 114, no. 1 (1981): 132–136.
11. J. L. Gastwirth and P. A. Hammick, "Estimation of Prevalence of a Rare Disease, Preserving Anonymity of Subjects by Group Testing: Application to Estimating the Prevalence of Aids Antibodies in Blood Donors," *Journal of Statistical Planning and Inference* 22 (1989): 15–27.
12. C. P. Farrington, "Estimating Prevalence by Group Testing Using Generalized Linear Models," *Statistics in Medicine* 11, no. 12 (1992): 1591–1597.
13. G. Hepworth, "Exact Confidence Intervals for Proportions Estimated by Group Testing," *Biometrics* 52 (1996): 1134–1146.
14. J. M. Hughes-Oliver and W. F. Rosenberger, "Efficient Estimation of the Prevalence of Multiple Rare Traits," *Biometrika* 87 (2000): 315–327.
15. D. W. Turner, J. D. Stamey, and D. M. Young, "Classic Group Testing With Cost for Grouping and Testing," *Computers & Mathematics with Applications* 58, no. 10 (2009): 1930–1935.
16. M. Warasi, J. Tebbs, C. McMahan, and C. Bilder, "Estimating the Prevalence of Multiple Diseases From Two-Stage Hierarchical Pooling," *Statistics in Medicine* 35 (2016): 3851–3864.
17. Y. Malinovsky, G. Haber, and P. S. Albert, "An Optimal Design for Hierarchical Generalized Group Testing," *Journal of the Royal Statistical Society, Series C* 69, no. 3 (2020): 607–621.
18. C. L. Chen and W. H. Swallow, "Using Group Testing to Estimate a Proportion, and to Test the Binomial Model," *Biometrics* 46 (1990): 1035–1046.
19. R. Brookmeyer, "Analysis of Multistage Pooling Studies of Biological Specimens for Estimating Disease Incidence and Prevalence," *Biometrics* 55, no. 2 (1999): 608–612.
20. B. Zhang, C. R. Bilder, and J. M. Tebbs, "Group Testing Regression Model Estimation When Case Identification Is a Goal," *Biometrical Journal* 55, no. 2 (2013): 173–189.
21. M. S. Black, C. R. Bilder, and J. M. Tebbs, "Optimal Retesting Configurations for Hierarchical Group Testing," *Journal of the Royal Statistical Society, Series C* 64, no. 4 (2015): 693–710.
22. G. Hepworth and R. Watson, "Revisiting Retesting in the Estimation of Proportions by Group Testing," *Communications in Statistics - Simulation and Computation* 46, no. 1 (2017): 261–274.
23. W. Zhang, A. Liu, Q. Li, and P. S. Albert, "Incorporating Retesting Outcomes for Estimation of Disease Prevalence," *Statistics in Medicine* 39, no. 6 (2020): 687–697.
24. J. M. Tebbs, C. S. McMahan, and C. R. Bilder, "Two-Stage Hierarchical Group Testing for Multiple Infections With Application to the Infertility Prevention Project," *Biometrics* 69, no. 4 (2013): 1064–1073.
25. Q. Li, A. Liu, and W. Xiong, "D-Optimality of Group Testing for Joint Prevalence Estimation of Correlated Rare Diseases in the Presence of Misclassification," *Statistics in Medicine* 27 (2017): 823–838.
26. M. S. Warasi, J. M. Tebbs, C. S. McMahan, and C. R. Bilder, "Estimating the Prevalence of Two or More Diseases Using Outcomes From Multiplex Group Testing," *Biometrical Journal* 65, no. 7 (2023): 2200270.
27. W. Zhang, A. Liu, Q. Li, and P. S. Albert, "Nonparametric Estimation of Distributions and Diagnostic Accuracy Based on Group-Tested Results With Differential Misclassification," *Biometrics* 76, no. 4 (2020): 1147–1156.
28. G. Haber, Y. Malinovsky, and P. S. Albert, "Is Group Testing Ready for Prime-Time in Disease Identification?," *Statistics in Medicine* 40, no. 17 (2021): 3865–3880.
29. S. Fieuws and G. Verbeke, "Pairwise Fitting of Mixed Models for the Joint Modeling of Multivariate Longitudinal Profiles," *Biometrics* 62, no. 2 (2006): 424–431.
30. W. Kassahun-Yimer, P. S. Albert, L. M. Lipsky, T. R. Nansel, and A. Liu, "A Joint Model for Multivariate Hierarchical Semicontinuous Data With Replications," *Statistical Methods in Medical Research* 28, no. 3 (2019): 858–870.
31. F. K. Hwang, "Group Testing With a Dilution Effect," *Biometrika* 63 (1976): 671–680.
32. M. Hung and W. Swallow, "Robustness of Group Testing in the Estimation of Proportions," *Biometrics* 55 (1999): 231–237.
33. S. Fieuws, G. Verbeke, and G. Molenberghs, "Random-Effects Models for Multivariate Repeated Measures," *Statistical Methods in Medical Research* 16 (2007): 387–397.
34. J. Besag, "Statistical Analysis of Non-lattice Data," *Statistician* 24, no. 3 (1975): 179–195.
35. S. J. Devlin, R. Gnanadesikan, and J. R. Kettenring, "Robust Estimation and Outlier Detection With Correlation Coefficients," *Biometrika* 62, no. 3 (1975): 531–545.
36. K. H. Zou, A. Liu, A. I. Bandos, L. Ohno-Machado, and H. E. Rockette, *Statistical Evaluation of Diagnostic Performance: Topics in ROC Analysis* (Boca Raton, FL: Chapman and Hall/CRC, 2012).
37. N. Sattar, F. K. Ho, J. M. Gill, et al., "BMI and Future Risk for COVID-19 Infection and Death Across Sex, Age and Ethnicity: Preliminary Findings From UK Biobank," *Diabetes and Metabolic Syndrome: Clinical Research and Reviews* 14, no. 5 (2020): 1149–1151.
38. Z. Liu, J. Li, W. Long, et al., "Bilirubin Levels as Potential Indicators of Disease Severity in Coronavirus Disease Patients: A Retrospective Cohort Study," *Frontiers in Medicine* 7 (2020), <https://doi.org/10.3389/fmed.2020.598870>.
39. N. Ali, "Relationship Between COVID-19 Infection and Liver Injury: A Review of Recent Data," *Frontiers in Medicine* 7, no. 458 (2020), <https://doi.org/10.3389/fmed.2020.00458>.
40. B. Kasapoglu, A. Yozgat, A. Tanoglu, G. Can, Y. S. Sakin, and M. Kekilli, "Gamma-Glutamyl-Transferase May Predict COVID-19 Outcomes in Hospitalised Patients," *International Journal of Clinical Practice* 75 (2021): e14933, <https://doi.org/10.1111/ijcp.14933>.
41. B. Datta, F. Njau, J. Thalmann, H. Haller, and A. D. Wagner, "Differential Infection Outcome of *Chlamydia Trachomatis* in Human Blood Monocytes and Monocyte-Derived Dendritic Cells," *BMC Microbiology* 14, no. 209 (2014), <https://doi.org/10.1186/s12866-014-0209-3>.

42. E. K. Naglak, S. G. Morrison, and R. P. Morrison, "Neutrophils Are Central to Antibody-Mediated Protection Against Genital Chlamydia," *Infection and Immunity* 85, no. 10 (2017): e00409–e00417.

43. G. V. Stokes and N. B. Isada, "Albumin Enhances Chlamydial Infectivity on Human Placental Cells," *Microbios* 65 (1991): 179–186.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.