Statistics
in Medicine

Research Article

Received 24 September 2014, Accepted 15 June 2015 Published online 14 July 2015 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6578

A general regression framework for group testing data, which incorporates pool dilution effects

Dewei Wang,[a] Christopher S. McMahan[b*†] and Colin M. Gallagher[b]

Group testing, through the use of pooling, has been widely implemented as a more efficient means to screen individuals for infectious diseases. Typically, in these settings, practitioners are tasked with the complimentary goals of both case identification and estimation. For these purposes, many group testing strategies have been proposed, which address issues such as preserving anonymity in estimation studies, quality control, and classification. In general, these strategies require that a significant number of the individuals be retested, either in pools or individually. In order to provide practitioners with a general methodology that can be used to accurately and precisely analyze data of this form, herein, we propose a binary regression framework that can incorporate data arising from any group testing strategy. Further, we relax previously made assumptions regarding testing error rates by relating the diagnostic testing results to the latent biological marker levels of the individuals being tested. We investigate the finite sample performance of our proposed methodology through simulation and by applying our techniques to hepatitis B data collected as part of a study involving Irish prisoners.

Keywords: biomarker; measurement error; pool testing; sensitivity; specificity

1. Introduction

The origin of group testing is typically attributed to Dorfman [1], who proposed the use of pooling as a means to reduce the time and cost associated with screening military inductees for syphilis during World War II. In general, group testing begins by collecting specimens (e.g., blood, urine, and plasma) from individuals, which are then physically combined to form a pooled specimen. The pooled specimen is then tested for the infection of interest, and the observed response provides classification information; that is, it provides evidence of whether or not the pool contains at least one positive member. Further, the data collected as part of a group testing classification process can subsequently be used for the purposes of estimation, for example, estimating the prevalence rate of an infection. Since its advent, group testing has been implemented for the purposes of screening for infectious diseases [2–4], identifying lead compounds in drug discovery [5], detecting rare mutations in genetics [6], and screening for viral agents in the case of bioterrorism [7]. These techniques have been used to screen millions of blood donations, both in the USA and abroad, for human immunodeficiency virus, hepatitis B virus (HBV), and hepatitis C virus [8]. Further, group testing is also routinely used to screen for other infectious diseases; for example, Lindan *et al.* [9] noted that 12% of the medical screening labs in the USA use pool testing for chlamydia screening. Consequently, given the ever increasing use of group testing techniques, the goal of this work is to develop reliable regression techniques that can be used to analyze data arising from any group testing process.

In many infectious disease screening applications, it is of primary interest to diagnose each individual as either being positive or negative for the infection of interest. To facilitate this goal, the classification

[a]*Department of Statistics, University of South Carolina, Columbia, SC 29028, U.S.A.*
[b]*Department of Mathematical Sciences, Clemson University, Clemson, SC 29634, U.S.A.*
[*]*Correspondence to: Christopher S. McMahan, O-110 Martin Hall, Department of Mathematical Sciences, Clemson University, Clemson, SC 29634, U.S.A.*
[†]*E-mail: mcmahan2@clemson.edu*

protocol presented in [1] suggested that if a pool tested negative, then each contributing individual should be diagnosed as uninfected. On the other hand, if a pool tested positive, then it should be "decoded" by retesting each contributing specimen separately. This testing protocol is commonly referred to as Dorfman decoding and is widely implemented in practice because of its simplicity. Since this seminal work, many variants of Dorfman's decoding algorithm have been proposed in an effort to reduce testing cost and/or increase classification accuracy; see [10] for a thorough review. In addition, group testing strategies have also been developed for the purposes of preserving anonymity in estimation studies [11] and for quality control purposes [12, 13]. In virtually all of the aforementioned situations, the associated group testing algorithm may require that a number of the individuals' specimens be assigned to multiple pools and/or be tested individually.

The use of group testing techniques, as a cost effective data collection mechanism in estimation studies, was first proposed by Thompson [14] and has since received a great deal of attention in the statistical literature. Many of the earlier works in this area use group testing data to estimate population-level characteristics, such as the proportion of infected individuals; for a review, see [15, 16] and the references therein. Extending these earlier works, authors have also proposed binary regression models that relate individual-level covariate information (e.g., age, race, and gender) to the testing responses observed from assaying pooled specimens [17–23]. All of the these regression methods proceed under the assumption that each of the individuals are assigned to exactly one pool, and make use of the testing responses observed from assaying these pools to perform estimation and inference. Therefore, these regression techniques cannot be used to analyze data arising from classification studies. Merging the goals of estimation and classification, Xie [24] and Zhang *et al.* [25] allow for the incorporation of additional retesting information gained from decoding positive pools. Further, Zhang *et al.* [25] illustrated that regression parameter estimates obtained from incorporating decoding information are more efficient than those based on individual-level testing data, when the assay being used is imperfect. That is, these authors were able to show that more precise estimation and inference can be realized through the analysis of group testing data than can be obtained through the analysis of individual-level testing data, and at a fraction of the data collection cost.

To account for assay measurement error, all of the aforementioned group testing regression techniques proceed under simplifying assumptions; that is, they assume that the testing error rates, sensitivity and specificity, are known and constant and are functionally independent of the pool size. In some applications, these assumptions may be reasonable [26], while in others, they may not [27, 28]. In general, a diagnostic test measures the concentration of a biological marker (biomarker) within a specimen, and its binary response indicates whether or not this measurement exceeds a predetermined threshold. Consequently, the composition of a pooled specimen, in terms of the number of positive and negative individuals, plays a key role in determining the testing error rates. For example, a specimen that might test positive, when tested individually, may be "diluted" past the assay's threshold of detection when pooled with multiple negative specimens. Several authors have developed methods that account for the 'dilution effect' when group testing is being used to estimate a population proportion (e.g., [29–32]). More recently, McMahan *et al.* [33] proposed a binary regression model, which identifies pool specific testing error rates based on the distributions of the latent biomarker concentration levels of the individuals. Further, these authors demonstrated that proceeding under the traditional assumptions, when they are invalid, may lead to severely biased estimation and inference. It is important to note that the methodology outlined in [33] was developed solely for the regression analysis of testing responses obtained from pools; that is, it does not allow for the incorporation of decoding/retesting information.

In this paper, we propose a general binary regression model that allows for the incorporation of information that may arise from all variants of group testing schemes, to include decoding algorithms. Generalizing the approach in [33] to data arising from individuals' specimens being assigned to multiple (statistically dependent) pools, we account for the 'dilution effect' by acknowledging the underlying mechanistic structure of the assay being employed. More pointedly, our modeling technique appropriately accounts for assay measurement error in order to avoid biased estimation and inference. Through simulation studies, we identify settings under which estimates obtained by our methodology are as efficient as those that arise from the analysis of individual-level testing data. We also illustrate that competing group testing regression methods, which proceed under the traditional assumptions regarding assay measurement error, may result in severely biased estimation and inference. Our proposed methodology could potentially benefit many agencies, such as those previously mentioned, that have adopted group testing strategies for the purposes of case identification and disease surveillance. That is, these organizations are privy to group testing data, which includes decoding information, and thus could benefit from the development of reliable binary regression methodology, which can be used to analyze the same.

The remainder of this article is organized as follows. In Section 2, we propose a general binary regression framework that can handle any type of data that arise from group testing strategies. To account for assay measurement error, we then derive explicit expressions that relate the observed testing outcomes to the underlying biomarker concentration levels that are being measured. In Section 3, we investigate the finite sample performance of our proposed methodology and compare our approach to existing modeling techniques. To further illustrate the performance of our proposed procedure, we also apply our regression methodology to hepatitis B data in Section 4. We conclude with a summary discussion in Section 5.

2. General notation and methodology

We consider the situation in which group testing is to be implemented for the purposes of screening N individuals for a binary characteristic of interest, such as infection status. In general, this process begins by collecting specimens (e.g., blood and urine) from individuals and assigning each of these specimens to exactly one of J groups of size n_j, for $j = 1, ..., J$. Within each group, these specimens are then tested according to a group testing strategy; for example, Dorfman decoding [1], halving [34], or array testing (AT) [35]. Depending on the goal of the study, this process may necessitate that a given individual be involved in several testing outcomes. For example, in the classification problem, the testing of pooled and/or individual specimens continues until each subject can be diagnosed as either positive or negative.

To formalize our notation in this context, we begin by defining $\mathcal{G}_j = \{1, ..., n_j\}$ to be the collection of indices corresponding to the specimens assigned to the j-th group, such that for each of the K_j observed testing responses associated with this group, we may identify the individuals involved in the l-th test by $\mathcal{P}_{jl} \subseteq \mathcal{G}_j$, for $l = 1, ..., K_j$, where we use l as a testing index. More explicitly, $\mathcal{P}_{jl}$ corresponds to the individuals in the j-th group whose specimens were pooled and assayed by the l-th test. We assume that under the selected group testing scheme, each individual in the j-th group should be tested at least once (i.e., $\cup_{l=1}^{K_j} \mathcal{P}_{jl} = \mathcal{G}_j$) and that pooling specimens across groups does not occur. On the other hand, we allow for the situation in which a specimen may belong to multiple pools within a given group (i.e., we do not require $\mathcal{P}_{jl} \cap \mathcal{P}_{jl'} = \emptyset$ for all l and l'), and we do not restrict attention to schemes that begin with master pool testing (MT) (i.e., we do not mandate that $\mathcal{G}_j \in \{\mathcal{P}_{j1}, ..., \mathcal{P}_{jK_j}\}$).

Let $Z_{\mathcal{P}_{jl}}$ denote the binary response observed from assaying the pool formed from amalgamating the $\mathcal{P}_{jl}$ individual specimens, such that $Z_{\mathcal{P}_{jl}} = 1$ indicates that the pool tested positive, $Z_{\mathcal{P}_{jl}} = 0$ otherwise. We collect all of the observed testing outcomes associated with the j-th group into the binary vector $\boldsymbol{Z}_j = (Z_{\mathcal{P}_{j1}}, \ldots, Z_{\mathcal{P}_{jK_j}})^{\mathrm{T}}$. Consequently, $\boldsymbol{Z}_j$ is a correlated binary vector that cannot be divided into two independent sub-vectors (otherwise, one could treat this group as two separate groups). Additionally, we assume that $\boldsymbol{Z}_j$ and $\boldsymbol{Z}_{j'}$ are independent, for all $j \neq j'$, which we believe to be reasonable because we do not allow for pooling specimens across groups. For the purpose of clarity, Table 1(a) provides three simple examples to illustrate the use and flexibility of our set notation.

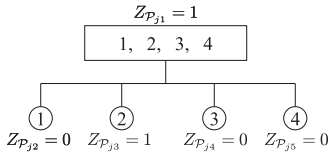
Let $T_{ij} = 1$ denote that the i-th individual in the j-th group is truly positive, $T_{ij} = 0$ otherwise, for $i = 1, ..., n_j$ and $j = 1, ..., J$. For notational convenience, we collect all of the statuses associated with the j-th group into the binary vector $\boldsymbol{T}_j = (T_{1j}, \ldots, T_{n_j j})^{\mathrm{T}}$. It is important to note that when the assay being used is imperfect, T_{ij} is unobservable, even under individual testing. For modeling purposes, we assume that the infection probability for the i-th individual in the j-th group is related to the linear predictor $\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}$ through a monotone and differentiable link function $\eta(\cdot)$, where $\boldsymbol{x}_{ij} = (1, x_{ij1}, \ldots, x_{ijp})^{\mathrm{T}}$ is a $(p+1)$-dimensional vector of covariates and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is the corresponding vector of regression parameters; that is, $\mathrm{pr}(T_{ij} = 1 | \boldsymbol{x}_{ij}) = \eta^{-1}(\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta})$. We assume throughout that the statuses T_{ij} are conditionally independent, given the covariates $\boldsymbol{x}_{ij}$. We also assume that conditional on the true statuses of the individuals, the observed testing responses are independent of all measured covariates. For a brief discussion of these and all subsequent modeling assumptions, please see Appendix A of the Supplementary Information.

It is well known that ignoring diagnostic testing errors when performing inference may lead to severely biased estimation (e.g., [33]). To account for imperfect testing, we first let $\mathcal{T}_j$ denote the collection of all possible outcomes of $\boldsymbol{T}_j$. Then for any $\boldsymbol{t}_j \in \mathcal{T}_j$ and for any possible realization $\boldsymbol{z}_j$ of $\boldsymbol{Z}_j$, we define

Table I. Illustration of notation under three common group testing algorithms; Dorfman testing, three-stage halving, and array testing.

(a) Illustration of notation

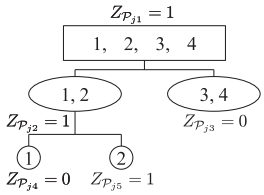
Example 1: Dorfman testing



Stage 1: Specimens collected from subjects $\mathcal{G}_j = \{1, 2, 3, 4\}$ are assigned to a master pool, which tests positive. The testing response is denoted by $Z_{\mathcal{P}_{j1}} = 1$, where $\mathcal{P}_{j1} = \{1, 2, 3, 4\}$.

Stage 2: Dorfman retesting reverts to individual testing, resulting in the four additional testing responses;
$Z_{\mathcal{P}_{j2}} = 0$, where $\mathcal{P}_{j2} = \{1\}$,
$Z_{\mathcal{P}_{j3}} = 1$, where $\mathcal{P}_{j3} = \{2\}$,
$Z_{\mathcal{P}_{j4}} = 0$, where $\mathcal{P}_{j4} = \{3\}$,
$Z_{\mathcal{P}_{j5}} = 0$, where $\mathcal{P}_{j5} = \{4\}$.

Example 2: Three-stage halving

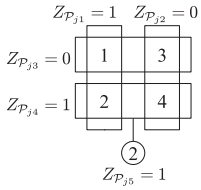


Stage 1: Exactly the same as **Stage 1** under Dorfman testing.

Stage 2: Halving then divides the positive master pool into two equally sized sub-pools, which are then tested, resulting in the two additional testing responses;
$Z_{\mathcal{P}_{j2}} = 1$, where $\mathcal{P}_{j2} = \{1, 2\}$,
$Z_{\mathcal{P}_{j3}} = 0$, where $\mathcal{P}_{j3} = \{3, 4\}$.

Stage 3: Of the sub-pools, one tests negative (requiring no further testing) and one tests positive. The positive sub-pool is decoded by retesting each specimen belonging to it individually, resulting in the two additional testing responses;
$Z_{\mathcal{P}_{j4}} = 0$, where $\mathcal{P}_{j4} = \{1\}$,
$Z_{\mathcal{P}_{j5}} = 1$, where $\mathcal{P}_{j5} = \{2\}$.

Example 3: Array testing

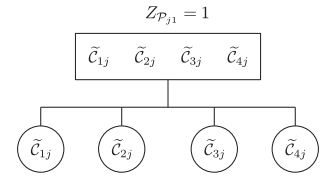


Stage 1: Specimens collected from subjects $\mathcal{G}_j = \{1, 2, 3, 4\}$ are assigned to a 2×2 array, row (column) pools are formed from combining specimens that share a common row (column). Row and column pools are tested, resulting in four testing responses;
$Z_{\mathcal{P}_{j1}} = 1$, where $\mathcal{P}_{j1} = \{1, 2\}$,
$Z_{\mathcal{P}_{j2}} = 0$, where $\mathcal{P}_{j2} = \{3, 4\}$,
$Z_{\mathcal{P}_{j3}} = 0$, where $\mathcal{P}_{j3} = \{1, 3\}$,
$Z_{\mathcal{P}_{j4}} = 1$, where $\mathcal{P}_{j4} = \{2, 4\}$.

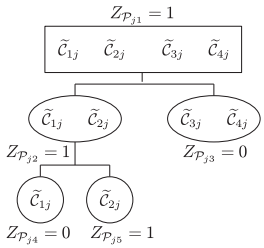
Stage 2: Specimens belonging to the intersection of positive rows and columns are retested individually, resulting in one additional testing response;
$Z_{\mathcal{P}_{j5}} = 1$, where $\mathcal{P}_{j5} = \{2\}$.

(b) Relating Testing Outcomes to Latent Biomarker Levels

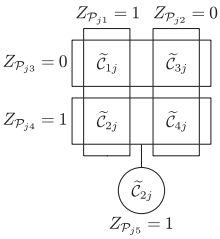
Let specimens 1, 2, 3, and 4 (above) have biomarker levels $\widetilde{\mathcal{C}}_{1j} = 1$, $\widetilde{\mathcal{C}}_{2j} = 7$, $\widetilde{\mathcal{C}}_{3j} = 1$, and $\widetilde{\mathcal{C}}_{4j} = 1$, respectively.
Assume that the test being employed measures biomarker levels without error and that the threshold is $t(c) = 2$, for all c.



Testing responses:
$Z_{\mathcal{P}_{j1}} = I\{4^{-1}\sum_{i\in\mathcal{P}_{j1}} \widetilde{\mathcal{C}}_{ij} > t(4)\} = 1$
$Z_{\mathcal{P}_{j2}} = I\{\widetilde{\mathcal{C}}_{1j} > t_0\} = 0$
$Z_{\mathcal{P}_{j3}} = I\{\widetilde{\mathcal{C}}_{2j} > t_0\} = 1$
$Z_{\mathcal{P}_{j4}} = I\{\widetilde{\mathcal{C}}_{3j} > t_0\} = 0$
$Z_{\mathcal{P}_{j5}} = I\{\widetilde{\mathcal{C}}_{4j} > t_0\} = 0$



Testing responses:
$Z_{\mathcal{P}_{j1}} = I\{4^{-1}\sum_{i\in\mathcal{P}_{j1}} \widetilde{\mathcal{C}}_{ij} > t(4)\} = 1$
$Z_{\mathcal{P}_{j2}} = I\{2^{-1}\sum_{i\in\mathcal{P}_{j2}} \widetilde{\mathcal{C}}_{ij} > t(2)\} = 1$
$Z_{\mathcal{P}_{j3}} = I\{2^{-1}\sum_{i\in\mathcal{P}_{j3}} \widetilde{\mathcal{C}}_{ij} > t(2)\} = 0$
$Z_{\mathcal{P}_{j4}} = I\{\widetilde{\mathcal{C}}_{1j} > t_0\} = 0$
$Z_{\mathcal{P}_{j5}} = I\{\widetilde{\mathcal{C}}_{2j} > t_0\} = 1$



Testing responses:
$Z_{\mathcal{P}_{j1}} = I\{2^{-1}\sum_{i\in\mathcal{P}_{j1}} \widetilde{\mathcal{C}}_{ij} > t(2)\} = 1$
$Z_{\mathcal{P}_{j2}} = I\{2^{-1}\sum_{i\in\mathcal{P}_{j2}} \widetilde{\mathcal{C}}_{ij} > t(2)\} = 0$
$Z_{\mathcal{P}_{j3}} = I\{2^{-1}\sum_{i\in\mathcal{P}_{j3}} \widetilde{\mathcal{C}}_{ij} > t(2)\} = 0$
$Z_{\mathcal{P}_{j4}} = I\{2^{-1}\sum_{i\in\mathcal{P}_{j4}} \widetilde{\mathcal{C}}_{ij} > t(2)\} = 1$
$Z_{\mathcal{P}_{j5}} = I\{\widetilde{\mathcal{C}}_{2j} > t_0\} = 1$

$M_j(\boldsymbol{z}_j, \boldsymbol{t}_j) = \text{pr}(\boldsymbol{Z}_j = \boldsymbol{z}_j \mid \boldsymbol{T}_j = \boldsymbol{t}_j)$. By an application of the law of total probability, one can relate the observed testing outcomes, given the observed pooling structure, to the individual-level covariates as follows:

$$\begin{aligned}
\text{pr}\left(\boldsymbol{Z}_j = \boldsymbol{z}_j | \boldsymbol{x}_{1j}, ..., \boldsymbol{x}_{n_j j}\right) &= \sum_{\boldsymbol{t}_j \in \mathcal{T}_j} \text{pr}\left(\boldsymbol{Z}_j = \boldsymbol{z}_j \mid \boldsymbol{T}_j = \boldsymbol{t}_j, \boldsymbol{x}_{1j}, ..., \boldsymbol{x}_{n_j j}\right) \text{pr}\left(\boldsymbol{T}_j = \boldsymbol{t}_j | \boldsymbol{x}_{1j}, ..., \boldsymbol{x}_{n_j j}\right) \\
&= \sum_{\boldsymbol{t}_j \in \mathcal{T}_j} \text{pr}\left(\boldsymbol{Z}_j = \boldsymbol{z}_j \mid \boldsymbol{T}_j = \boldsymbol{t}_j\right) \prod_{i=1}^{n_j} \text{pr}\left(T_{ij} = t_{ij} | \boldsymbol{x}_{ij}\right) \\
&= \sum_{\boldsymbol{t}_j \in \mathcal{T}_j} \left\{ M_j\left(\boldsymbol{z}_j, \boldsymbol{t}_j\right) \prod_{i=1}^{n_j} \left\{ t_{ij} \eta^{-1}\left(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta}\right) + \left(1 - t_{ij}\right)\left[1 - \eta^{-1}\left(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta}\right)\right] \right\} \right\}.
\end{aligned}$$

Recall that, under classification algorithms, such as Dorfman decoding, three-stage halving (TH), and AT, $\boldsymbol{Z}_j$ is a correlated binary vector. By calculating the joint probability of observing z_j, one implicitly accounts for the dependence that exists between these observed testing responses. Consequently, the observed data log-likelihood can be expressed as

$$l(\boldsymbol{\beta}) = \sum_{j=1}^{J} \log \mathcal{R}\left(z_j, \boldsymbol{x}_{1j}, ..., \boldsymbol{x}_{n_j j}, \boldsymbol{\beta}\right), \tag{1}$$

where $\mathcal{R}(z_j, \boldsymbol{x}_{1j}, ..., \boldsymbol{x}_{n_j j}, \boldsymbol{\beta}) = \text{pr}(\boldsymbol{Z}_j = z_j | \boldsymbol{x}_{1j}, ..., \boldsymbol{x}_{n_j j})$. If each $M_j(z_j, \boldsymbol{t}_j)$ were known, then one could easily estimate $\boldsymbol{\beta}$ by directly maximizing (1) with respect to $\boldsymbol{\beta}$, although, in practice, this is not the case. Consequently, in the next section, we provide details on how to evaluate $M_j(z_j, \boldsymbol{t}_j)$ based on the underlying characteristics of the assay being employed.

2.1. *Evaluation of misclassification*

To account for imperfect testing and pool dilution effects, we generalize the methodology described in [29, 30, 33]. Following the work of these authors, we proceed under the standard convention that a diagnostic test classifies a specimen as positive (negative) if its measured biomarker concentration is above (below) a predetermined threshold. For generality, we allow the assay threshold, which we denote by $t(c)$, for a pool to vary with the number of specimens, say c, of which it is composed. Typically, the specification of $t(c)$ has proceeded in one of two fashions. In particular, Tu *et al.* [36] specified that $t(c) = t_0$ for any pool size c, where t_0 is the assay threshold under individual testing. This approach has also been implemented in the infectious disease screening literature (e.g., [37]). Alternatively, to account for the effect of pooling, Vansteelandt *et al.* [38] specified that $t(c) = t_0/c$. Note that if the diagnostic test is subject to limits of detection, the latter thresholding strategy may not be reasonable for larger pool sizes. Both of these thresholding strategies are investigated in Sections 3 and 4. In either case, we first derive a closed-form expression for $M_j(z_j, \boldsymbol{t}_j)$ in terms of the relevant biomarker distributions under an arbitrary thresholding strategy.

To this end, we define $\tilde{C}_{ij}$ to be the true biomarker concentration level for the i-th individual in the j-th group, and we assume that conditional on the individual's true status, $\tilde{C}_{ij} | T_{ij} \sim f_{\tilde{C}|T_{ij}} = T_{ij} f_{\tilde{C}^+} + (1 - T_{ij}) f_{\tilde{C}^-}$, where $f_{\tilde{C}^+}$ and $f_{\tilde{C}^-}$ are the probability density functions for the biomarker concentration levels of the infected and uninfected individuals, respectively. We initially assume that these biomarker distributions are known; this assumption is later relaxed in Section 4. Define $\widetilde{\boldsymbol{C}}_j = (\widetilde{C}_{1j}, \ldots, \widetilde{C}_{n_j j})^{\text{T}}$, for each j, to be the collection of the true biomarker levels for the n_j individuals assigned to the j-th group. To account for the underlying structure of the assay being employed, we are left to relate $\widetilde{\boldsymbol{C}}_j$ to the testing outcomes $\boldsymbol{Z}_j$.

When pooled assessments are being made, we assume that the true biomarker concentration of the pool is the arithmetic average of the biomarker concentrations of the individual specimens contributing to the pool; that is, letting $\widetilde{C}_{\mathcal{P}_{jl}}$ denote the biomarker concentration for the pool consisting of the $\mathcal{P}_{jl}$ individuals, we assume that $\widetilde{C}_{\mathcal{P}_{jl}} = c_{jl}^{-1} \sum_{i \in \mathcal{P}_{jl}} \widetilde{C}_{ij}$, where c_{jl} denotes the cardinality of the set $\mathcal{P}_{jl}$. We view this assumption to be reasonable, as long as the individual specimens being pooled are of equal volume. Additionally, this assumption is common among the biomarker pooling literature [39–43] and has previously been assumed in the group testing estimation literature [29, 30, 33]. To simplify this relationship, we define the design vector associated with the test of the pool consisting of the $\mathcal{P}_{jl}$ individuals to be $\boldsymbol{D}_{\mathcal{P}_{jl}} = c_{jl}^{-1} \mathbf{1}_{\mathcal{P}_{jl}}$, where $\mathbf{1}_{\mathcal{P}_{jl}}$ is a n_j-dimensional binary vector whose $\mathcal{P}_{jl}$-th components are 1 and all others are 0. Using this notation, we can express the pool biomarker concentration levels as $\widetilde{C}_{\mathcal{P}_{jl}} = \boldsymbol{D}_{\mathcal{P}_{jl}}^{\text{T}} \widetilde{\boldsymbol{C}}_j$. It is important to point out that in the presence of measurement error, each $\widetilde{C}_{\mathcal{P}_{jl}}$ is unobservable.

We now derive our expression for $M_j(\boldsymbol{z}_j, \boldsymbol{t}_j)$ in terms of the aforementioned biomarker distributions. To account for assay measurement error, we let $C_{\mathcal{P}_{jl}}$ denote the error laden measurement of $\widetilde{C}_{\mathcal{P}_{jl}}$, and we assume that conditional on the true biomarker concentration levels, $C_{\mathcal{P}_{jl}} \overset{ind}{\sim} f_{C|\widetilde{C}_{\mathcal{P}_{jl}}}$, for $l = 1, ..., K_j$ and $j = 1, ..., J$. Thus, the observed testing responses, under our classification rule, are given by $Z_{\mathcal{P}_{jl}} = I\{C_{\mathcal{P}_{jl}} > t(c_{jl})\}$. For purposes of clarity, we provide a simple illustration of how testing responses are derived in this context in Table 1(b). Noting this relationship, we are able to write the probabilities associated with the observed testing outcomes in terms of the measured biomarker concentrations; for example, $\text{pr}(Z_{\mathcal{P}_{jl}} = z_{\mathcal{P}_{jl}}) = \text{pr}\{C_{\mathcal{P}_{jl}} \in A(z_{\mathcal{P}_{jl}}, c_{jl})\}$, where $A(0, c) = \{u : u \leq t(c)\}$ and $A(1, c) = \{u : u > t(c)\}$.

Using this relationship, we can express $M_j(\boldsymbol{z}_j, \boldsymbol{t}_j)$ as follows:

$$\begin{aligned} M_j\left(\boldsymbol{z}_j, \boldsymbol{t}_j\right) &= \text{pr}\left(Z_{\mathcal{P}_{j1}} = z_{\mathcal{P}_{j1}}, ..., Z_{\mathcal{P}_{jK_j}} = z_{\mathcal{P}_{jK_j}} \mid \boldsymbol{T}_j = \boldsymbol{t}_j\right) \\ &= \text{pr}\left\{C_{\mathcal{P}_{j1}} \in A\left(z_{\mathcal{P}_{j1}}, c_{j1}\right), ..., C_{\mathcal{P}_{jK_j}} \in A\left(z_{\mathcal{P}_{jK_j}}, c_{jK_j}\right) \mid \boldsymbol{T}_j = \boldsymbol{t}_j\right\} \\ &= \text{pr}\left\{\boldsymbol{C}_j \in \boldsymbol{A}\left(\boldsymbol{z}_j, \mathbf{c}_j\right) \mid \boldsymbol{T}_j = \boldsymbol{t}_j\right\}, \end{aligned}$$

where $\boldsymbol{C}_j = \left(C_{\mathcal{P}_{j1}}, ..., C_{\mathcal{P}_{jK_j}}\right)^{\mathrm{T}}$, $\boldsymbol{A}\left(\boldsymbol{z}_j, \mathbf{c}_j\right) = A\left(z_{\mathcal{P}_{j1}}, c_{j1}\right) \times \cdots \times A\left(z_{\mathcal{P}_{jK_j}}, c_{jK_j}\right)$, and $\mathbf{c}_j = \left(c_{j1}, ..., c_{jK_j}\right)^{\mathrm{T}}$. Based on the probability density functions $f_{\widetilde{C}|T_{ij}}$ and $f_{C|\widetilde{C}_{\mathcal{P}_{jl}}}$, the conditional probability density function of $\boldsymbol{C}_j$ given $\boldsymbol{T}_j = \boldsymbol{t}_j$ is

$$f_{\boldsymbol{C}_j|\boldsymbol{T}_j=\boldsymbol{t}_j}(\boldsymbol{u}) = \int \prod_{l=1}^{K_j} f_{C|\widetilde{C}_{\mathcal{P}_{jl}}=\boldsymbol{D}^{\mathrm{T}}_{\mathcal{P}_{jl}}\boldsymbol{y}}(u_l) \prod_{i=1}^{n_j} f_{\widetilde{C}|T_{ij}=t_{ij}}\left(y_{ij}\right) d\boldsymbol{y}, \tag{2}$$

where $\boldsymbol{u} = \left(u_1, \ldots, u_{K_j}\right)^{\mathrm{T}}$ and $\boldsymbol{y} = \left(y_{1j}, \ldots, y_{n_j j}\right)^{\mathrm{T}}$. Finally,

$$M_j\left(\boldsymbol{z}_j, \boldsymbol{t}_j\right) = \int_{\boldsymbol{A}(\boldsymbol{z}_j, \mathbf{c}_j)} f_{\boldsymbol{C}_j|\boldsymbol{T}_j=\boldsymbol{t}_j}(\boldsymbol{u}) d\boldsymbol{u}. \tag{3}$$

By relating the observed testing outcomes to the latent biomarker concentration levels of the individuals in this fashion, we have accomplished two primary goals: First, we are able to appropriately account for the effect of imperfect testing, and second, we are able to account for the dependence that exists between the observed testing responses. It is worthwhile to point out that under MT, expression (3) is equivalent to the results presented in [33].

One should note that the integrals in (2) and (3) are multidimensional if $n_j > 1$ and if the individuals in the j-th group are involved in more than one test, respectively. In general, these integrals may be difficult to evaluate analytically, but this challenge is easily overcome using Monte Carlo techniques, as will be illustrated in Sections 3 and 4. It is possible to obtain a closed-form expression of $f_{\boldsymbol{C}_j|\boldsymbol{T}_j=\boldsymbol{t}_j}$, if one assumes that $\widetilde{C}_{ij}|T_{ij} = 1 \sim N(\mu_+, \sigma^2_+)$, $\widetilde{C}_{ij}|T_{ij} = 0 \sim N(\mu_-, \sigma^2_-)$ and $C|\widetilde{C} \sim N(\widetilde{C}, \rho^2)$. Although a special case, these distributional assumptions are common among the pooled biomarker literature (e.g., [39–41, 44]). For further details on how to evaluate (3), see Appendix B of the Supplementary Information.

2.2. *Maximum likelihood estimation*

Using the observed data $\{(\boldsymbol{z}_j, \boldsymbol{x}_{1j}, ..., \boldsymbol{x}_{n_j j}), j = 1, ..., J\}$, one can estimate $\boldsymbol{\beta}$ by maximizing (1) directly after using (3) to evaluate $M_j(\boldsymbol{z}_j, \boldsymbol{t}_j)$, for all $\boldsymbol{t}_j \in \mathcal{T}_j$. We denote the resulting maximum likelihood estimator (MLE) as $\widehat{\boldsymbol{\beta}}$. The standard theoretical properties for MLEs hold for $\widehat{\boldsymbol{\beta}}$ under the assumption that the group sizes remain finite so that $J \to \infty$ as $N \to \infty$. The variance–covariance matrix of $\widehat{\boldsymbol{\beta}}$ can be estimated by the negative inverse Hessian of (1) evaluated at $\widehat{\boldsymbol{\beta}}$ and can be used to conduct typical Wald-type inference.

One may note that the evaluation of $M_j(\boldsymbol{z}_j, \boldsymbol{t}_j)$ over all $\boldsymbol{t}_j \in \mathcal{T}_j$ could pose a significant computational burden, especially if n_j is large. To obviate this difficulty, we point out that $M_j(\boldsymbol{z}_j, \boldsymbol{t}_j)$ is free of $\boldsymbol{\beta}$ and can therefore be calculated before numerical optimization routines are implemented. To further alleviate this computational burden, we have developed efficient algorithms for computing these terms under two of the most popular group testing schemes: Dorfman decoding and TH; these algorithms are provided in Appendix C of the Supplementary Information. In conjunction with the aforementioned algorithms, we have had little difficulty implementing a quasi-Newton optimization routine in R for the purposes of identifying the MLE. Depending on the complexity of the group testing strategy (e.g., AT), it may not be feasible to directly maximize the observed data log-likelihood using numerical techniques. In these situations, our methodology can still be implemented through the use of an expectation maximization algorithm, which is provide in Appendix D of the Supplementary Information.

3. Simulation study

In this section, we illustrate the performance of our proposed methodology through simulation and compare our results to those obtained from more traditional group testing regression techniques. These traditional methods generally proceed under the assumption that the testing error rates, sensitivity (S_e) and

specificity (S_p), are known and constant and do not depend on the pool size. More explicitly, the testing error rates are the same for all pool sizes, to include individual-level testing. The sensitivity (specificity) of an assay is typically defined to be the probability that the assay will classify a specimen as positive (negative) given it is truly positive (negative). To incorporate retesting information, previously proposed methods have been developed under the assumption that the testing outcomes for pools (individuals) are independent given their true statuses. Under these assumptions, the conditional probability of observing z_j, given the individuals' true latent statuses $\boldsymbol{t}_j$, can be expressed as

$$M_j\left(z_j, \boldsymbol{t}_j\right) = \prod_{l=1}^{K_j} \left\{ S_e^{z_{p_{jl}} \widetilde{z}_{p_{jl}}} (1 - S_e)^{\left(1 - z_{p_{jl}}\right) \widetilde{z}_{p_{jl}}} (1 - S_p)^{z_{p_{jl}} \left(1 - \widetilde{z}_{p_{jl}}\right)} S_p^{\left(1 - z_{p_{jl}}\right)\left(1 - \widetilde{z}_{p_{jl}}\right)} \right\}, \tag{4}$$

where $\widetilde{z}_{p_{jl}} = I\{\sum_{i \in p_{jl}} t_{ij} > 0\}$ is the true status of the pool being tested. Substituting the aforementioned expression into (1) and maximizing directly result in obtaining an estimate of $\boldsymbol{\beta}$, under these more traditional assumptions.

3.1. Data generation and model fitting

In this study, we consider the following models:

(1) $\text{logit}\{\text{pr}(T_{ij} = 1 \mid x_{ij1})\} = \beta_0 + \beta_1 x_{ij1}$; $\boldsymbol{\beta} = (\beta_0, \beta_1)^{\text{T}} = (-3, 2)^{\text{T}}$,
(2) $\text{logit}\{\text{pr}(T_{ij} = 1 \mid x_{ij1})\} = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij1}^2$; $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^{\text{T}} = (-3, 1, 0.5)^{\text{T}}$,
(3) $\text{logit}\{\text{pr}(T_{ij} = 1 \mid x_{ij1}, x_{ij2})\} = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2}$; $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^{\text{T}} = (-3, 2, 1)^{\text{T}}$,

where $x_{ij1} \sim \text{N}(0, 0.75^2)$ and $x_{ij2} \sim \text{Bernoulli}(0.1)$. These model choices emulate situations in which group testing could be employed and provide for mean prevalences ranging from 8–10%. These models were also studied in [33]. Motivated by the data application in Section 4, gamma distributions were chosen for the individual biomarker concentrations; specifically, $\widetilde{C}|T = 1 \sim \text{gamma}(2, \tau)$, and $\widetilde{C}|T = 0 \sim \text{gamma}(0.2, 1.4)$, where the rate parameter $\tau \in \{0.3, 0.4, 0.5\}$. To account for assay measurement error, we specified the conditional distribution of the measured concentration levels to be $C|\widetilde{C} \sim N\{\widetilde{C}, (0.1\widetilde{C})^2\}$. The assay threshold was chosen to be $t_0 = 1$, so that the specificity under individual testing would be $S_p = 0.971$, while the sensitivities would be $S_e = 0.962, 0.937, 0.908$ corresponding to $\tau = 0.3, 0.4, 0.5$, respectively.

Individual-level covariates, $\boldsymbol{x}_i$, and infection probabilities, p_i, were randomly generated according to the aforementioned models, for $i = 1, ..., N$, where $N = 3600$. Each individual's true status, T_i, was then determined according to a Bernoulli distribution having success probability p_i. The corresponding biomarker concentration level, $\widetilde{C}_i$, was then generated according to either $\widetilde{C}_i|T_i = 1 \sim \text{gamma}(2, \tau)$ or $\widetilde{C}_i|T_i = 0 \sim \text{gamma}(0.2, 1.4)$, depending on the value of T_i. For each model and value of τ, this process was used to generate 500 independent data sets of the form $\{(\widetilde{C}_i, \boldsymbol{x}_i), i = 1, ..., N\}$. Note that the slight change to the subscript notation in the individual-level data is meant to highlight the fact that the individuals have not been assigned to an initial group; once assignment occurs, the notation developed in Section 2 is readopted.

In what follows, we describe how group testing data were generated based on the individual-level biomarker data. As one might expect, the data structure is highly dependent on the particular group testing strategy being employed. For the purposes of our simulation, we have opted to investigate four of the most common strategies: MT, Dorfman testing (DT), TH, and AT. In order to levy pool diagnoses, we considered two methods of specifying the assay threshold for pooled specimens; that is, we considered letting $t(c) = t_0$ and $t(c) = t_0/c$ as was suggested in [37, 38], respectively.

Under MT, all individuals within a given group are pooled together, and the pool is tested, with no further testing being implemented. Thus, to create MT data, we first randomly assigned each of the individuals to one of J groups of size n; that is, $n_j = n$, for $j = 1, ..., J$, where $n \in \{2, 4, 6\}$ and $J = N/n$. The testing response vector under MT for the j-th group is given by $\boldsymbol{Z}_j^{MT} = Z_{\mathcal{P}_{j1}}$, where $\mathcal{P}_{j1} = \mathcal{G}_j = \{1, ..., n\}$, and is determined by $Z_{\mathcal{P}_{j1}} = I\{C_{\mathcal{P}_{j1}} > t(n)\}$, where $C_{\mathcal{P}_{j1}} \sim N\{\widetilde{C}_{\mathcal{P}_{j1}}, (0.1\widetilde{C}_{\mathcal{P}_{j1}})^2\}$ and $\widetilde{C}_{\mathcal{P}_{j1}} = n^{-1} \sum_{i=1}^{n} \widetilde{C}_{ij}$. MT, unlike DT, TH, and AT, is not a decoding algorithm; that is, it does not levy a diagnosis for each individual.

The first decoding algorithm that we consider is DT, which specifies that a group whose master pool test is negative requires no further screening, but in those cases where the master pool test is positive, the group is resolved by retesting each subject individually. To create DT data, we again consider the

grouping strategy described for MT and note that the testing response vector under DT for the j-th group is identical to that under MT, if the master pool test is negative. Alternatively, if the master pool test is positive, the testing response vector is given by $\boldsymbol{Z}_j^{DT} = (Z_{\mathcal{P}_{j1}}, Z_{\mathcal{P}_{j2}}, ..., Z_{\mathcal{P}_{jK_j}})^{\mathrm{T}}$, where $Z_{\mathcal{P}_{j1}}$ is determined as discussed previously, $K_j = n + 1$, and $\mathcal{P}_{jl} = \{l - 1\}$, for $l = 2, ..., K_j$. The response $Z_{\mathcal{P}_{jl}}$, for $l = 2, ..., K_j$, corresponds to individually testing the $(l-1)$-th subject and is determined according to $Z_{\mathcal{P}_{jl}} = I\{C_{\mathcal{P}_{jl}} > t_0\}$, where $C_{\mathcal{P}_{jl}} \sim N\{\widetilde{C}_{ij}, (0.1\widetilde{C}_{ij})^2\}$ and $i = l - 1$.

The other two decoding algorithms considered in this study are TH and AT. Under TH, when a positive master pool response is observed, the positive group is randomly divided into two equally sized subgroups, and these subgroups are then tested. If a subgroup tests negative, then testing is complete; alternatively, if a subgroup tests positive, then all contributing subjects are retested individually [e.g., Table 1(a)]. For the purpose of this study, we again randomly assign individuals to pools of size n, where $n \in \{4, 6\}$, and use the threshold $t(n/2)$ to diagnose the subgroups in the second stage. Note that the use of TH and DT is equivalent with groups of size $n = 2$ and consequently, this group size was not considered for TH. Similarly, in implementing AT, we adopted the testing protocol outlined in [10] for square AT (without MT) for $n \times n$ arrays, where $n \in \{4, 6\}$. It is easy to show that AT is more costly (in terms of the number of tests required to complete classification) than individual-level testing when $n = 2$, and we therefore do not consider $n = 2$ in this study. For brevity, we have chosen not to explicitly describe the construction of the response vectors for TH and AT, but the process was similar to the approach described previously.

The regression methodology discussed in Section 2 was applied to each of the group testing data sets. In particular, $\boldsymbol{\beta}$ was estimated by directly maximizing (1) using a quasi-Newton optimization routine in R, for MT, DT, and TH. To ease computational burdens, the efficient algorithms for DT and TH, provided in Appendix C of the Supplementary Information, were utilized when maximizing (1). In contrast, analyzing data arising from AT is more complex, and consequently, the EM algorithm provided in Appendix D of the Supplementary Information was implemented for the purpose of obtaining the MLE of $\boldsymbol{\beta}$. The evaluation of $M_j(\boldsymbol{z}_j, \boldsymbol{t}_j)$ under all group testing strategies was accomplished using the Monte Carlo techniques described in Appendix B of the Supplementary Information. It is worthwhile to point out that under MT, our proposed methodology reduces to the technique proposed in [33]. Thus, comparisons between the model fits based on MT, DT, TH, and AT data allow one to assess the performance of our proposed methodology when compared with that of this existing technique. Further, these comparisons also allow one to assess the benefits, in terms of parameter estimation and inference, of including decoding information when it is available, for example, in studies in which practitioners are tasked to perform both classification and estimation. For the purposes of comparison, we also fit the regression models that proceed under the more traditional assumptions using the reformulation presented in (4) and the appropriate individual S_e and S_p levels. Additionally, for each of the biomarker data sets, we also generated subject-level testing responses (i.e., $n = 1$) and fit the individual data model.

3.2. *Simulation results*

Table II provides summary statistics of the 500 estimates of $\boldsymbol{\beta}$ obtained from model (2) for all considered group testing algorithms under the two thresholding strategies, when $n \in \{2, 4, 6\}$ and $\tau = 0.5$. From these results, we see that the maximum likelihood estimates of $\boldsymbol{\beta}$ obtained by our proposed approach exhibit little, if any, evidence of bias, across all considered configurations. The parameter estimates based on MT data do exhibit bias, and this bias tends to increase with the group size n. For MT, this effect is likely explained by the fact that the number of observed testing responses, used to estimate $\boldsymbol{\beta}$, decreases as n increases. One will also notice that the standard deviations of the estimated regression coefficients, from our proposed method, are predominantly in agreement with the corresponding average standard errors, suggesting that the variance–covariance matrix of $\widehat{\boldsymbol{\beta}}$ is being estimated correctly, across all considered configurations and group testing strategies. Further, as n increases, so does the variability in $\boldsymbol{\beta}$; this is an expected phenomenon because the number of testing responses, used to estimate $\boldsymbol{\beta}$, decreases as n increases. However, this effect is attenuated for the data collected by the decoding algorithms (DT, TH, and AT), which is explained by the addition of the retesting information associated with decoding positive pools. Consequently, these results exhibit the strengths of our proposed methodology when compared with those of the technique presented in [33]; that is, by allowing for the inclusion of retesting information, when available, more reliable estimation and inference can be obtained.

Table II also provides summary statistics of the 500 estimates of $\boldsymbol{\beta}$, which were obtained from the traditional group testing regression methodology, that is, the regression model that proceeds under the

Table II. Simulation results for model (2) having regression parameters $\boldsymbol{\beta} = (-3, 1, 0.5)^{\mathrm{T}}$.

When $t(c) = t_0$:			Proposed model				Traditional model			
	n	Measure	MT	DT	TH	AT	MT	DT	TH	AT
$\widehat{\beta}_0$	2	Bias(Cov)	−0.02(0.95)	−0.02(0.96)	−−(−−)	−−(−−)	−0.45(**0.12**)	−0.04(0.95)	−−(−−)	−−(−−)
		SD(SE)	0.12(0.12)	0.11(0.11)	−−(−−)	−−(−−)	0.17(0.16)	0.10(0.10)	−−(−−)	−−(−−)
	4	Bias(Cov)	−0.04(0.95)	−0.03(0.95)	−0.00(0.95)	−0.03(0.95)	−1.07(**0.00**)	−0.55(**0.00**)	−0.46(**0.02**)	−0.60(**0.00**)
		SD(SE)	0.17(0.17)	0.13(0.13)	0.13(0.13)	0.13(0.13)	0.32(0.29)	0.13(0.13)	0.12(0.12)	0.12(0.12)
	6	Bias(Cov)	−0.08(0.96)	−0.03(0.95)	−0.00(0.95)	−0.05(0.93)	−1.88(**0.00**)	−1.02(**0.00**)	−0.87(**0.00**)	−1.07(**0.00**)
		SD(SE)	0.26(0.24)	0.15(0.15)	0.15(0.15)	0.15(0.15)	0.70(0.57)	0.20(0.18)	0.15(0.14)	0.15(0.15)
$\widehat{\beta}_1$	2	Bias(Cov)	0.05(0.96)	0.03(0.95)	−−(−−)	−−(−−)	0.31(0.92)	−0.08(0.85)	−−(−−)	−−(−−)
		SD(SE)	0.19(0.18)	0.13(0.13)	−−(−−)	−−(−−)	0.38(0.31)	0.11(0.10)	−−(−−)	−−(−−)
	4	Bias(Cov)	0.11(0.95)	0.03(0.95)	0.03(0.96)	0.03(0.94)	0.77(**0.86**)	0.05(0.93)	−0.04(**0.89**)	−0.02(**0.87**)
		SD(SE)	0.38(0.32)	0.15(0.15)	0.15(0.14)	0.15(0.14)	0.81(0.65)	0.20(0.17)	0.16(0.14)	0.17(0.15)
	6	Bias(Cov)	0.21(0.94)	0.04(0.96)	0.04(0.96)	0.08(0.96)	1.45(**0.89**)	0.29(**0.90**)	0.05(0.93)	0.04(**0.91**)
		SD(SE)	0.64(0.50)	0.18(0.18)	0.18(0.17)	0.20(0.18)	1.41(1.14)	0.40(0.30)	0.23(0.20)	0.24(0.21)
$\widehat{\beta}_2$	2	Bias(Cov)	−0.02(0.96)	−0.01(0.94)	−−(−−)	−−(−−)	−0.16(**0.89**)	−0.06(0.94)	−−(−−)	−−(−−)
		SD(SE)	0.14(0.14)	0.11(0.10)	−−(−−)	−−(−−)	0.23(0.19)	0.09(0.09)	−−(−−)	−−(−−)
	4	Bias(Cov)	−0.06(0.94)	−0.01(0.95)	−0.02(0.94)	−0.01(0.94)	−0.47(**0.74**)	−0.15(**0.80**)	−0.14(**0.73**)	−0.15(**0.72**)
		SD(SE)	0.27(0.23)	0.12(0.12)	0.13(0.12)	0.13(0.12)	0.40(0.32)	0.14(0.12)	0.12(0.10)	0.12(0.10)
	6	Bias(Cov)	−0.11(0.93)	−0.02(0.95)	−0.03(0.96)	−0.01(0.94)	−0.77(**0.74**)	−0.28(**0.69**)	−0.22(**0.66**)	−0.22(**0.66**)
		SD(SE)	0.40(0.34)	0.15(0.15)	0.15(0.15)	0.17(0.15)	0.60(0.50)	0.22(0.18)	0.15(0.13)	0.16(0.13)
When $t(c) = t_0/c$:			Proposed model				Traditional model			
	n	Measure	MT	DT	TH	AT	MT	DT	TH	AT
$\widehat{\beta}_0$	2	Bias(Cov)	−0.02(0.96)	−0.01(0.96)	−−(−−)	−−(−−)	0.39(**0.04**)	0.49(**0.00**)	−−(−−)	−−(−−)
		SD(SE)	0.13(0.13)	0.12(0.12)	−−(−−)	−−(−−)	0.10(0.10)	0.08(0.08)	−−(−−)	−−(−−)
	4	Bias(Cov)	−0.03(0.96)	−0.02(0.95)	−0.01(0.95)	−0.03(0.95)	0.68(**0.00**)	0.54(**0.00**)	0.64(**0.00**)	0.62(**0.00**)
		SD(SE)	0.17(0.18)	0.12(0.12)	0.12(0.12)	0.12(0.12)	0.10(0.10)	0.08(0.08)	0.08(0.08)	0.08(0.08)
	6	Bias(Cov)	−0.09(0.97)	−0.01(0.97)	−0.01(0.94)	−0.02(0.97)	0.86(**0.00**)	0.56(**0.00**)	0.70(**0.00**)	0.75(**0.00**)
		SD(SE)	0.25(0.25)	0.12(0.12)	0.12(0.12)	0.12(0.12)	0.12(0.13)	0.08(0.08)	0.08(0.08)	0.07(0.08)
$\widehat{\beta}_1$	2	Bias(Cov)	0.04(0.93)	0.02(0.95)	−−(−−)	−−(−−)	−0.18(**0.65**)	−0.22(**0.24**)	−−(−−)	−−(−−)
		SD(SE)	0.22(0.21)	0.14(0.14)	−−(−−)	−−(−−)	0.12(0.12)	0.08(0.08)	−−(−−)	−−(−−)
	4	Bias(Cov)	0.11(0.94)	0.02(0.96)	0.02(0.97)	0.02(0.95)	−0.24(**0.60**)	−0.20(**0.35**)	−0.23(**0.20**)	−0.25(**0.15**)
		SD(SE)	0.35(0.34)	0.14(0.14)	0.14(0.13)	0.14(0.13)	0.15(0.16)	0.08(0.08)	0.08(0.08)	0.08(0.08)
	6	Bias(Cov)	0.21(0.94)	0.02(0.96)	0.02(0.96)	0.02(0.98)	−0.26(**0.66**)	−0.18(**0.44**)	−0.21(**0.26**)	−0.22(**0.25**)
		SD(SE)	0.62(0.50)	0.14(0.14)	0.14(0.14)	0.13(0.13)	0.20(0.20)	0.09(0.09)	0.08(0.08)	0.08(0.08)
$\widehat{\beta}_2$	2	Bias(Cov)	−0.02(0.97)	−0.01(0.96)	−−(−−)	−−(−−)	−0.02(0.98)	−0.04(0.95)	−−(−−)	−−(−−)
		SD(SE)	0.15(0.14)	0.10(0.10)	−−(−−)	−−(−−)	0.09(0.10)	0.07(0.07)	−−(−−)	−−(−−)
	4	Bias(Cov)	−0.06(0.94)	−0.01(0.96)	−0.01(0.96)	−0.00(0.95)	−0.04(0.97)	−0.02(0.96)	−0.04(0.94)	−0.05(**0.91**)
		SD(SE)	0.23(0.22)	0.10(0.10)	0.10(0.10)	0.11(0.10)	0.12(0.13)	0.07(0.07)	0.07(0.07)	0.07(0.07)
	6	Bias(Cov)	−0.09(0.92)	−0.01(0.96)	−0.01(0.96)	−0.00(0.95)	−0.02(0.97)	−0.01(0.96)	−0.03(0.95)	−0.03(0.94)
		SD(SE)	0.35(0.29)	0.10(0.10)	0.10(0.10)	0.10(0.10)	0.16(0.17)	0.07(0.08)	0.07(0.07)	0.07(0.07)

MT, master pool testing; DT, Dorfman testing; TH, three-stage halving; AT, array testing; Bias, empirical bias; Cov, coverage probabilities; SD, standard deviation; SE, standard error.

Presented results include the Bias and SD of the 500 estimates of $\boldsymbol{\beta}$, when $n \in \{2, 4, 6\}$ and $\tau = 0.5$. The average SE and estimated 95% Wald Cov are also provided. Assuming a 99% confidence level for the coverage probabilities, the margin of error is 0.03. Estimates outside this margin of error are shown in bold.

assumption that $M_j(\mathbf{z}_j, \boldsymbol{t}_j)$ can be expressed as (4). From these results, one will note that the regression parameter estimates obtained by the traditional modeling approach exhibit a larger bias when compared with the estimates resulting from our proposed method, across all considered configurations and group testing strategies. This finding likely explains why the estimated coverage probabilities for the traditional modeling approach tends to be incongruously small given the specified confidence level, while those associated with our technique remain at their nominal level, regardless of the group size or group testing procedure.

In Figure 1, we provide a comparison between our proposed method and the analogous approach that would be utilized to analyze individual-level testing data. This comparison is meant to examine the two key aspects of our proposed approach: first, the variability of the regression parameter estimates, and second, the reduction in testing expenditure. In this situation, observing a testing response on each

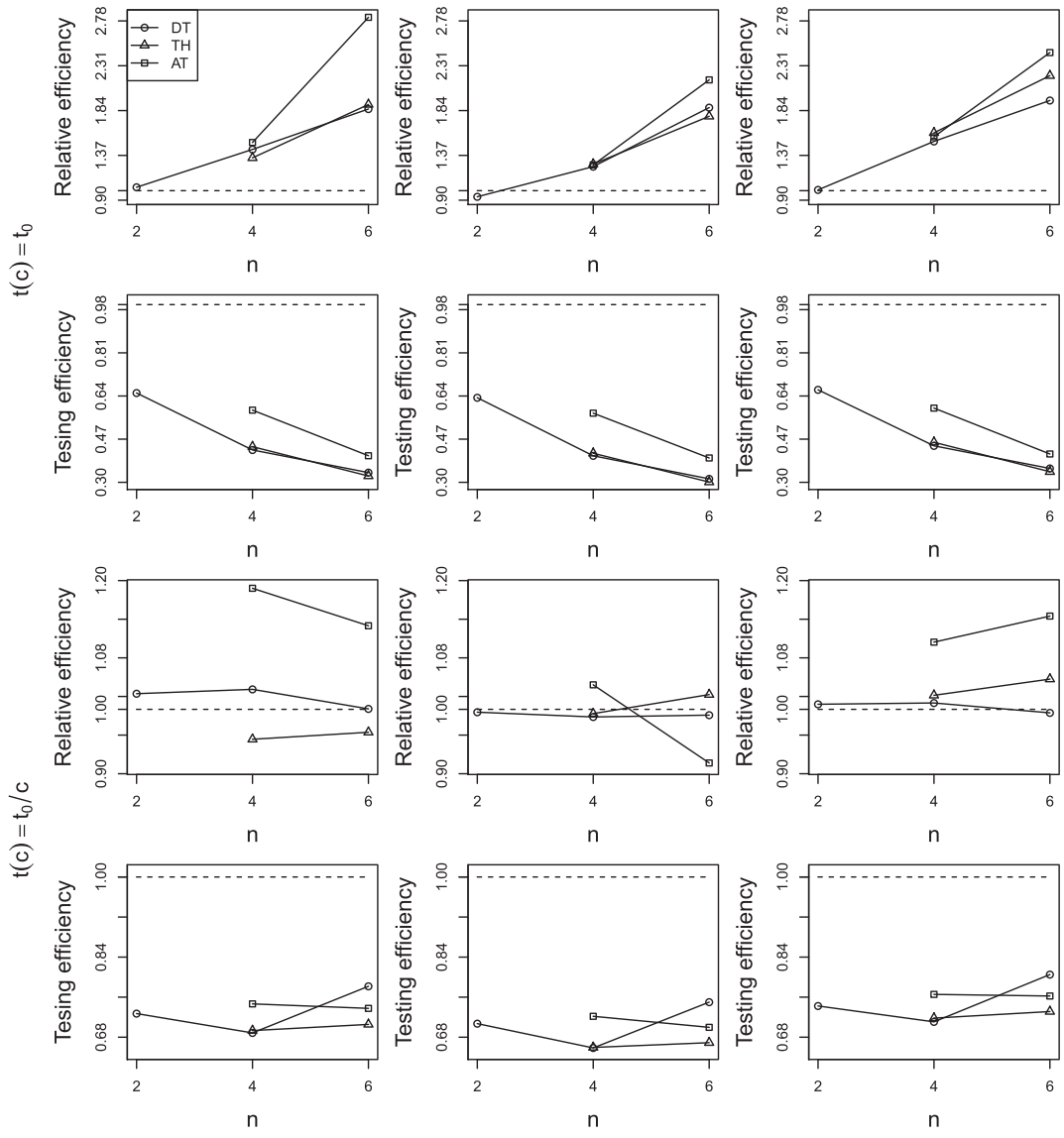


Figure 1. Simulation results concerning the efficiency of the parameter estimates obtained from modeling group testing data, resulting from different algorithms. Presented are results for model (1) (left), model (2) (middle), and model (3) (right) across all considered group sizes (n), when τ_0.5. We define the testing efficiency to be the ratio between the average number of tests performed by a group testing algorithm and the number of tests required to conduct individual-level testing; i.e., N. The relative efficiency is defined to be the ratio between $\text{MSE}(\widehat{\boldsymbol{\beta}})$ obtained from modeling group testing data and the $\text{MSE}(\widehat{\boldsymbol{\beta}})$ from modeling individual-level testing data, where $\text{MSE}(\widehat{\boldsymbol{\beta}}) = tr\{E[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\mathrm{T}}]\}$.

individual can be thought of as having the maximal amount of information. Consequently, a comparison between the variability of the estimates of $\boldsymbol{\beta}$ obtained by our approach to the variability of the estimates resulting from analyzing individual-level testing data allows one to see what impact pooling has on the efficiency of parameter estimation. In particular, in Figure 1, we provide a plot of the relative efficiency, which is defined to be the ratio between the mean squared error of the estimates obtained from analyzing group testing data and the mean squared error of the estimates resulting from the individual data model, as well as the percentage reduction in testing cost obtained through the use of DT, TH, and AT, across all considered configurations when $\tau = 0.5$. These results suggest that the estimates obtained from data collected by a group testing decoding algorithm can be roughly as efficient as those obtained from the analysis of individual-level data, but at a significantly reduced cost of testing.

In Appendix E of the Supplementary Information, we provide a complete summary of our simulation results across all considered group testing strategies and values of n, when $\tau = 0.5$. The results

under other considered settings of τ were practically identical and are therefore omitted. In addition to the simulation study discussed earlier, we have also performed simulations that allow for different biomarker distributional assumptions (e.g., normal, Weibull, and log-normal) and the use of different group testing algorithms. We have also investigated the characteristics of our regression methodology under the situation in which the biomarker distributions are unknown and have to be estimated. The findings from these alternate studies are congruous with the results presented herein, and we have therefore opted not to include them. Further, we have also performed a model selection study; the results of which indicate that the proposed approach preforms as well, if not better, than the traditional modeling technique, with respect to identifying the correct model; for further details, see Appendix E of the Supplementary Information.

4. Irish hepatitis B virus data

To further illustrate our methodology, we apply our techniques to the hepatitis B data previously analyzed by [33]. This data set was originally compiled in an effort to assess the prevalence of antibodies to the HBV in Irish prisoners; for further details, see [45]. In the original study, oral fluid specimens were collected from each individual and were subsequently tested through the use of the Murex ICE enzyme immunoassay; the observed optical density (OD) readings from this process were then recorded. All positive results were then confirmed using an in-house radioimmunoassay. The data set also provides a diagnosed status for each of the individuals, which we will treat as their true infection status. At the time of specimen collection, covariate information (e.g., age, gender, and drug use) was collected via voluntary questionnaire from participating subjects. After the individuals with missing predictor variables and/or testing information were removed, we are left with a sample of size $N = 1098$, that is, 60 HBV-positive and 1038 HBV-negative individuals. Note that in this and similar settings, practitioners are typically tasked with the complimentary goals of both classification and estimation, thus providing a scenario in which our proposed methodology could be adopted. The main purpose of this study is to compare the performance of our group testing regression methods to those that proceed under the more traditional assumptions.

One will note that the aforementioned information was collected on the individual level; that is, for all N individuals, we have access to their OD reading and covariate information. Using this information, we are able to artificially construct group testing data. Proceeding in this fashion allows us to assess the performance of our methodology across a wide variety of settings (e.g., various group sizes, grouping schemes, and thresholding strategies), which would not be possible otherwise. Additionally, this approach, which allows for comparisons between estimates obtained from group testing and individual-level data, has become a common practice in the group testing regression literature (e.g., [22]). To create group testing data, we first note that the OD readings, which were available to us, are simply a measurement of the underlying antibody concentration levels. We assume that the observed OD readings are linearly related to the true antibody concentration levels and were measured without error. This assumption was also made in [33] and from other related studies appears to be reasonable (e.g., [46]). Subsequently, we may determine the OD reading for a pool formed from combining the $\mathcal{P}_{jl}$ individuals by $\widetilde{C}_{\mathcal{P}_{jl}} = c_{jl}^{-1} \sum_{i \in \mathcal{P}_{jl}} \widetilde{C}_{ij}$. Notice that we use the $\widetilde{C}$ notation defined in Section 2.1 to represent the OD readings. In practice, this assumption may be questionable for various reasons (e.g., pools are constructed of specimens of unequal volume), and as such, we have conducted a sensitivity analysis designed to assess the impact of violations of this assumption on our proposed approach; see Appendix F of the Supplementary Information. Testing outcomes for pools were then determined by $Z_{\mathcal{P}_{jl}} = I\{\widetilde{C}_{\mathcal{P}_{jl}} > t(c_{jl})\}$, where we considered the thresholding strategies $t(c) = t_0$ and $t(c) = t_0/c$. In practice, the assay threshold for individual-level testing (t_0) is known, but regretfully, this value was not provided in the data. Thus, to choose a reasonable value of t_0, we first partition all 1098 OD readings into two sets $\text{OD}^+ = \{\widetilde{C}_i : T_i = 1\}$ and $\text{OD}^- = \{\widetilde{C}_i : T_i = 0\}$. We then select t_0 to minimize the discrepancies between the individuals' true statuses and their diagnosed statuses based on the OD readings; that is,

$$t_0 = \arg\max_t \left\{ \sum_{\widetilde{C}_i \in \text{OD}^+} I\left(\widetilde{C}_i > t\right) + \sum_{\widetilde{C}_i \in \text{OD}^-} I\left(\widetilde{C}_i < t\right) \right\}.$$

In this study, we did not have access to the underlying distribution of the OD readings for the positive and negative individuals, which we denote by $f_{\widetilde{C}^+}$ and $f_{\widetilde{C}^-}$, respectively. Consequently, we estimated these distributions through the use of training data. Specifically, density estimation proceeded under the assumption that the OD readings followed a parametric model, and we considered three such models: gamma, Weibull, and log-normal. Two training data sets were created by randomly sampling 10 observations from OD^+ and 44 observations from OD^-. Using the training data, we obtained the estimates $\widehat{f}_{\widetilde{C}^+}$ and $\widehat{f}_{\widetilde{C}^-}$ of $f_{\widetilde{C}^+}$ and $f_{\widetilde{C}^-}$, respectively, through the use of maximum likelihood techniques. In order to fit the traditional regression models, we calculated the sensitivity and specificity of individual-level testing to be $S_e = \int_{t_0}^{\infty} \widehat{f}_{\widetilde{C}^+}(x)dx$ and $S_p = \int_{\infty}^{t_0} \widehat{f}_{\widetilde{C}^-}(x)dx$, respectively. To implement our regression methodology, we calculated $M_j(\boldsymbol{z}_j, \boldsymbol{t}_j)$ as follows:

Table III. Irish hepatitis B virus data.

			Proposed model			Traditional model		
	n	IT	DT	TH	AT	DT	TH	AT
Under the assumption that the OD readings follow a gamma distribution								
$\widehat{\beta}_0$	2	−2.90(0.06)	−2.87(0.06)	– –(– –)	– –(– –)	−2.86(0.06)	– –(– –)	– –(– –)
	4		−2.78(0.10)	−2.78(0.09)	−2.77(0.09)	−2.68(0.07)	−2.66(0.07)	−2.65(0.07)
	6		−2.85(0.11)	−2.82(0.09)	−2.84(0.10)	−2.27(0.07)	−2.22(0.08)	−2.29(0.09)
$\widehat{\beta}_1$	2	1.09(0.10)	1.14(0.10)	– –(– –)	– –(– –)	1.15(0.11)	– –(– –)	– –(– –)
	4		1.09(0.10)	1.10(0.10)	1.10(0.10)	1.08(0.10)	1.09(0.10)	1.10(0.10)
	6		1.10(0.11)	1.12(0.10)	1.10(0.11)	0.80(0.07)	0.83(0.07)	0.82(0.08)
$\widehat{\beta}_2$	2	−0.31(0.06)	−0.33(0.06)	– –(– –)	– –(– –)	−0.33(0.06)	– –(– –)	– –(– –)
	4		−0.32(0.06)	−0.32(0.05)	−0.32(0.06)	−0.33(0.06)	−0.33(0.05)	−0.31(0.06)
	6		−0.32(0.06)	−0.33(0.06)	−0.32(0.06)	−0.29(0.05)	−0.30(0.05)	−0.31(0.05)
Under the assumption that the OD readings follow a log-normal distribution								
$\widehat{\beta}_0$	2	−2.92(0.06)	−2.90(0.05)	– –(– –)	– –(– –)	−2.88(0.06)	– –(– –)	– –(– –)
	4		−2.84(0.09)	−2.84(0.09)	−2.82(0.09)	−2.70(0.07)	−2.69(0.07)	−2.69(0.06)
	6		−2.89(0.08)	−2.87(0.07)	−2.85(0.07)	−2.30(0.07)	−2.26(0.09)	−2.25(0.08)
$\widehat{\beta}_1$	2	1.09(0.11)	1.14(0.10)	– –(– –)	– –(– –)	1.14(0.10)	– –(– –)	– –(– –)
	4		1.10(0.10)	1.12(0.10)	1.10(0.10)	1.07(0.09)	1.09(0.10)	1.07(0.11)
	6		1.10(0.11)	1.12(0.11)	1.10(0.11)	0.80(0.07)	0.83(0.08)	0.85(0.06)
$\widehat{\beta}_2$	2	−0.31(0.06)	−0.32(0.06)	– –(– –)	– –(– –)	−0.32(0.06)	– –(– –)	– –(– –)
	4		−0.32(0.05)	−0.33(0.06)	−0.32(0.06)	−0.32(0.05)	−0.33(0.06)	−0.32(0.06)
	6		−0.31(0.06)	−0.33(0.06)	−0.32(0.06)	−0.28(0.05)	−0.30(0.05)	−0.31(0.05)
Under the assumption that the OD readings follow a Weibull distribution								
$\widehat{\beta}_0$	2	−2.88(0.06)	−2.86(0.06)	– –(– –)	– –(– –)	−2.84(0.07)	– –(– –)	– –(– –)
	4		−2.78(0.11)	−2.78(0.10)	−2.78(0.11)	−2.66(0.07)	−2.65(0.08)	−2.63(0.09)
	6		−2.81(0.15)	−2.80(0.12)	−2.82(0.14)	−2.25(0.08)	−2.20(0.09)	−2.22(0.10)
$\widehat{\beta}_1$	2	1.09(0.10)	1.13(0.10)	– –(– –)	– –(– –)	1.14(0.10)	– –(– –)	– –(– –)
	4		1.08(0.10)	1.11(0.10)	1.08(0.10)	1.07(0.10)	1.10(0.10)	1.09(0.10)
	6		1.08(0.11)	1.12(0.11)	1.08(0.10)	0.80(0.07)	0.84(0.08)	0.85(0.09)
$\widehat{\beta}_2$	2	−0.30(0.05)	−0.32(0.06)	– –(– –)	– –(– –)	−0.32(0.06)	– –(– –)	– –(– –)
	4		−0.32(0.06)	−0.33(0.06)	−0.33(0.06)	−0.33(0.06)	−0.33(0.06)	−0.32(0.06)
	6		−0.32(0.06)	−0.33(0.06)	−0.31(0.06)	−0.29(0.05)	−0.30(0.05)	−0.31(0.05)

OD, optical density; IT, individual-level testing data; DT, Dorfman testing; TH, three-stage halving; AT, array testing; Presented results include the sample mean (standard deviation) of the 1000 maximum likelihood estimates of $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^{\mathrm{T}}$, across all considered configurations under the thresholding strategy $t(c) = t_0/c$ and homogeneous grouping.

$$M_j\left(\boldsymbol{z}_j, \boldsymbol{t}_j\right) = \int \prod_{l=1}^{K_j} I\left\{\boldsymbol{D}_{\mathcal{P}_{jl}}^{\mathrm{T}} \boldsymbol{y} \in A\left(z_{\mathcal{P}_{jl}}, c_{jl}\right)\right\} \prod_{i=1}^{n_j} \widehat{f}_{\widetilde{C}}\left(y_{ij} \mid T_{ij} = t_{ij}\right) d\boldsymbol{y},$$

where $\widehat{f}_{\widetilde{C}}(\cdot \mid T) = T\widehat{f}_{\widetilde{C}^+}(\cdot) + (1 - T)\widehat{f}_{\widetilde{C}^-}(\cdot)$. Note that the aforementioned integral is difficult to analytically compute; consequentlys the Monte Carlo technique described in Appendix B of the Supplementary Information was used to approximate $M_j(\boldsymbol{z}_j, \boldsymbol{t}_j)$.

To make our comparisons, we consider the following simple second-order logistic model:

$$\operatorname{logit}\left\{\operatorname{pr}\left(T_{ij} = 1 \mid x_{ij}\right)\right\} = \beta_0 + \beta_1 x_{ij} + \beta_2 x_{ij}^2,$$

where x_{ij} denotes the age of the i-th individual in the j-th group. After the removal of the training data, there were $N = 1044$ observations remaining. We again specify a common group size n, where $n \in \{1, 2, 4, 6\}$, and we assigned each of the N individuals to one of the J groups. In this study, we considered both homogeneous and random grouping schemes; that is, individuals of a common age were grouped together (homogeneous grouping), or individuals were randomly assigned to groups (random grouping). The former grouping strategy has been shown to result in more efficient parameter estimation when compared with the latter (e.g., [18, 22]). However, homogeneous grouping is not always feasible in practical applications. The group testing strategies chosen for this study were DT, TH, and AT, and group testing data

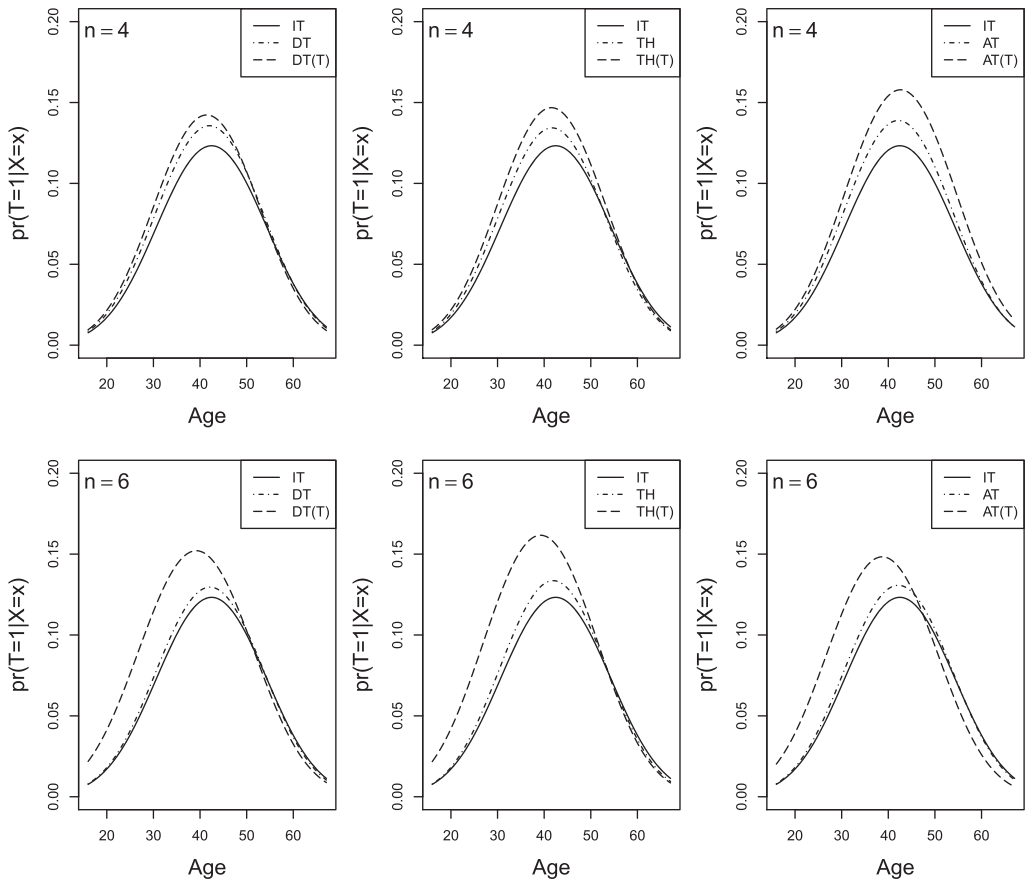


Figure 2. Irish hepatitis B virus data. Plots of the estimated regression functions averaged over the 1000 data sets for all considered group testing algorithms under the thresholding strategy $t(c) = t_0/c$, homogeneous grouping, $n \in \{4, 6\}$, and under the assumption that the optical density readings follow a gamma distribution. Within each figure, the average estimated regression function based on individual-level testing data (IT) is also provided as a standard by which comparisons can be made. From left to right, the figures present the regression estimates corresponding to optical density (DT), three-stage halving (TH), and array testing (AT). We use DT(T), TH(T), and AT(T) to denote the results obtained under the traditional modeling assumptions for the group testing algorithms DT, TH, and AT, respectively.

were subsequently generated in a similar fashion to the methods described in Section 3.1. For each of the group testing data sets, our proposed methodology was used to estimate the regression parameters. In order to compare our approach with existing techniques, we also estimated the regression parameters for each data set using the group testing regression models, which proceed under the traditional modeling assumptions. Further, to provide a standard by which comparisons can be made, we also fit the individual data model (i.e., $n = 1$). This process was repeated 1000 times for each pool size, with a new training data set being selected each time.

In order to assess misclassification error rates of the different thresholding strategies, we compared the individuals' true statuses with the diagnosed statuses obtained from the three group testing decoding algorithms. A summary of these results across all of the considered testing configurations is provided in Appendix F of the Supplementary Information. From these comparisons, we found that the thresholding strategy $t(c) = t_0$ resulted in an extremely high false negative rate; that is, under this strategy, many of the truly positive individuals were classified as negative. Consequently, we have chosen to focus our attention on the data arising from the thresholding strategy $t(c) = t_0/c$, which resulted in misclassification rates similar to that of individual-level testing. In practice, the thresholding strategy used in conjunction with a specified group testing process should be chosen with these characteristics in mind, especially if the practitioner is faced with the complimentary goals of classification and estimation; for further discussion, see Section 5. Table III provides a summary of the 1000 estimates of $\boldsymbol{\beta}$ across all considered configurations under our selected thresholding strategy and homogeneous pooling. Figure 2 provides plots of the estimated regression functions averaged over the 1000 replications for all considered group testing algorithms under the same thresholding and pooling strategies, when the OD readings are modeled using a gamma distribution. A complete summary of the results of this study is provided in Appendix F of the Supplementary Information. From these results, one will first note that the estimates obtained by our regression methodology appear to be more reliable when compared with the estimates resulting from the more traditional regression techniques, across all considered configurations. These results reinforce the main findings discussed in Section 3.2. Specifically, Figure 2 illustrates that the traditional regression methodology tends to drastically overestimate the age-specific probabilities of HBV infection for larger group sizes (e.g., $n = 6$), while the estimates from our method remain in agreement with the results from the individual-level data. These trends can also be observed in the summary of the estimates of $\boldsymbol{\beta}$ provided in Table III. The discrepancies between the estimates obtained by our method and those resulting from the individual data model are likely explained by the error introduced by having to estimate $f_{\widetilde{C}^+}$ and $f_{\widetilde{C}^-}$.

5. Discussion

In this paper, we have generalized the group testing regression methodology proposed in [33], to allow for the incorporation of testing information obtained from all group testing strategies. Our techniques could therefore be adopted by practitioners who have been tasked to perform studies in which classification and estimation are complimentary goals. We have illustrated that by incorporating retesting information, when available, a more precise analysis can be obtained, when compared with analyses that ignore this additional information. Further, we have shown that the regression methods, which operate under the traditional modeling assumptions, with regard to assay measurement error, can result in biased estimation and inference, when the assumptions are in fact invalid. Through simulation studies, we have identified settings in which our proposed techniques can result in parameter estimates that are as efficient as those based on individual-level data, and at a fraction of the cost of data collection. To further disseminate our work, we have developed software, programmed in R, that implements all of the proposed regression techniques, which can be downloaded at `http://people.stat.sc.edu/wang528/Research.html`.

In addition to the findings presented herein, we have also discovered situations in which the proposed method, when utilized to analyze data arising from a group testing decoding process (e.g., DT, TH, or AT), can result in obtaining estimates that are less variable than estimates that are obtained from the analysis of individual-level testing data. This feature has also been reported in [23, 25]. Consequently, a topic of future research could be to design group testing algorithms such that the resulting observed testing data can be used to obtain more efficient and accurate estimators, when compared with estimators resulting from individual-level testing data. Further, as was demonstrated in Section 4, the choice of the thresholding strategy in certain applications is paramount with respect to controlling false negative/positive rates. Thus, a future line of research could involve deriving thresholding strategies, which minimize these error rates.

References

1. Dorfman R. The detection of defective members of large populations. *Annals of Mathematical Statistics* 1943; **14**:436–440.
2. Cardoso M, Koerner K, Kubanek B. Mini-pool screening by nucleic acid testing for hepatitis B virus, hepatitis C virus, and HIV: preliminary results. *Transfusion* 1988; **38**:905–907.
3. Lewis J, Lockary V, Kobic S. Cost savings and increased efficiency using a stratified specimen pooling strategy for chlamydia trachomatis and neisseria gonorrhea. *Sexually Transmitted Diseases* 2012; **39**:46–48.
4. Van T, Miller J, Warshauer D, Reisdorf E, Jerrigan D, Humes R, Shult P. Pooling nasopharyngeal/throat swab specimens to increase testing capacity for influenza viruses by PCR. *Journal of Clinical Microbiology* 2012; **50**:891–896.
5. Remlinger K, Hughes Oliver J, Young S, Lam R. Statistical design of pools using optimal coverage and minimal collision. *Technometrics* 2006; **48**:133–143.
6. Gastwirth J. The efficiency of pooling in the detection of rare mutations. *The American Journal of Human Genetics* 2000; **67**:1036–4039.
7. Schmidt M, Roth W, Meyer H, Seifried E, Hourfar M. Nucleic acid test screening of blood donors for orthopoxviruses can potentially prevent dispersion of viral agents in case of bioterrorism. *Transfusion* 2005; **45**:399–403.
8. Stramer S, Notari E, Krysztof D, Dodd R. Hepatitis B virus testing by minipool nucleic acid testing: does it improve blood safety. *Transfusion* 2013; **53**:2525–2537.
9. Lindan C, Mathur M, Kumta S, Jerajani H, Gogate A, Schachter J, Moncada J. Utility of pooled urine specimens for detection of chlamydia trachomatis and neisseria gonorrhoeae in men attending public sexually transmitted infection clinics in Mumbai, India, by PCR. *Journal of Clinical Microbiology*; **43**:1674–1677.
10. Kim H, Hudgens M, Dreyfuss J, Westreich D, Pilcher C. Comparison of group testing algorithms for case identification in the presence of test error. *Biometrics* 2007; **63**:1152–1163.
11. Hammick P, Gastwirth J. Group testing for sensitive characteristics: extension to higher prevalence levels. *International Statistical Review/ Revue Internationale de Statistique* 1994; **62**:319–331.
12. Gastwirth J, Johnson W. A two-stage adaptive group-testing procedure for estimating small proportions. *Journal of the American Statistical Association* 1994; **89**:972–981.
13. Johnson W, Pearson L. Dual screening. *Biometrics* 1999; **55**:867–873.
14. Thompson K. Estimation of the proportion of vectors in a natural population of insects. *Biometrics* 1962; **18**:568–578.
15. Bilder CR, Tebbs JM. Empirical Bayesian estimation of the disease transmission probability in multiple-vector-transfer designs. *Biometrical Journal* 2005; **47**:502–516.
16. Liu A, Chunling L, Zhang Z, Albert P. Optimality of group testing in the presence of misclassification. *Biometrika* 2012; **99**:245–251.
17. Vansteelandt E, Goetghebeur E, Verstraeten T. Regression models for disease prevalence with diagnostic tests on pools of serum samples. *Biometrics* 2000; **56**:1126–1133.
18. Bilder CR, Tebbs JM. Bias, efficiency and agreement for group-testing regression models. *Journal of Statistical Computation and Simulation* 2009; **79**:67–80.
19. Chen P, Tebbs JM, Bilder CR. Group testing regression models with fixed and random effects. *Biometrics* 2009; **65**: 1270–1278.
20. Huang X, Tebbs JM. On latent-variable model misspecification in structural measurement error models for binary response. *Biometrics* 2009; **65**:710–718.
21. Delaigle A, Meister A. Nonparametric regression analysis for group testing data. *Journal of American Statistical Association* 2011; **106**:640–650.
22. Delaigle A, Hall P. Nonparametric regression with homogeneous group testing data. *Annals of Statistics* 2012; **40**:131–158.
23. Wang D, Zhou H, Kulasekera KB. A semi-local likelihood regression estimator of the proportion based on group testing data. *Journal of Nonparametric Statistics* 2013; **25**:209–221.
24. Xie M. Regression analysis of group testing samples. *Statistics in Medicine* 2001; **20**:1957–1969.
25. Zhang B, Bilder C, Tebbs J. Group testing regression model estimation when case identification is a goal. *Biometrical Journal* 2013; **55**:173–189.
26. Cahoon Young B, Chandler A, Livermore T, Gaudino J, Benjamin R. Sensitivity and specificity of pooled versus individual sera in a human immunodeficiency virus antibody prevalence Study. *Journal of Clinical Microbiology* 1989; **101**: 775–778.
27. Rovira A, Cano J, Munoz ZC. Feasibility of pooled-sample testing for the detection of porcine reproductive and respiratory syndrome virus antibodies on serum samples by ELISA. *Veterinary Microbiology* 2008; **130**:60–68.
28. Chang C, Wu Y, Pan Y, Chen Z, Wang C. Feasibility of human T-lymphotropic virus type I screening using pooled sera. *Journal of the Formosan Medical Association* 2002; **27**:1893–1895.
29. Wein L, Zenios S. Pooled testing for HIV screening: capturing the dilution effect. *Operations Research* 1996; **44**:543–569.
30. Zenios S, Wein L. Pooled testing for HIV prevalence estimation: exploiting the dilution effect. *Statistics in Medicine* 1988; **17**:1447–1467.
31. Yamamura K, Hino A. Estimation of the proportion of defective units by using group testing under the existence of a threshold of detection. *Communications in Statistics* 2007; **36**:949–957.
32. Zenios S, Wein L. Sample size for detecting and estimating the proportion of transgenic plants with narrow confidence intervals. *Seed Research Science* 2010; **20**:1–14.
33. McMahan CS, Tebbs JM, Bilder CR. Regression models for group testing data with pool dilution effects. *Biostatistics* 2013; **14**:284–298.
34. Litvak E, Tu X, Pagano M. Screening for presence of a disease by pooling sera samples. *Journal of the American Statistical Association* 1994; **89**:424–434.
35. Phatarfod RM, Sudbury A. The use of a square array scheme in blood testing. *Statistics in Medicine* 1994; **13**:2337–2343.
36. Tu X, Litvak E, Pagano M. Screening tests: can we get more by doing less?. *Statistics in Medicine* 1994; **13**:1905–1919.

37. Currie M, McNiven M, Yee T, Schiemer U, Bowden F. Pooling of clinical specimens prior to testing for Chlamydia trachomatis by PCR is accurate and cost saving. *Journal of Clinical Microbiology* 2004; **42**:4866–4867.
38. Vansteelandt S, Goetghebeur E, Thomas I, Mathys E, Van LF. On the viral safety of plasma pools and plasma derivatives. *Journal of the Royal Statistical Society: Series A* 2005; **168**:345–363.
39. Liu A, Schisterman E. Comparison of diagnostic accuracy of biomarkers with pooled assessments. *Biometrical Journal* 2003; **45**:631–644.
40. Liu A, Schisterman E, Teoh E. Sample size and power calculation in comparing diagnostic accuracy of biomarkers with pooled assessments. *Journal of Applied Statistics* 2004; **31**:49–59.
41. Mumford S, Schisterman E, Vexler A, Liu A. Pooling biospecimens and limits of detection: effects on ROC curve analysis. *Biostatistics* 2006; **4**:585–598.
42. Bondell H, Liu A, Schisterman E. Statistical inference based on pooled data: a moment-based estimating equation approach. *Journal of Applied Statistics* 2007; **34**:129–140.
43. Vexler A, Schisterman E, Liu A. Estimation of ROC curves based on stably distributed biomarkers subject to measurement error and pooling mixtures. *Statistics in Medicine* 2008; **27**:280–296.
44. Faraggi D, Reiser B, Schisterman E. ROC curve analysis for biomarkers based on pooled assessments. *Statistics in Medicine* 2003; **22**:2515–2527.
45. Allwright S, Bradley F, Long J, Barry J, Thornton L, Parry J. Prevalence of antibodies to hepatitis B, hepatitis C, and HIV and risk factors in Irish prisoners: results of a national cross sectional survey. *British Medical Journal* 2000; **321**:78–82.
46. Noborg U, Gusdal A, Pisa E, Hedrum A, Lindh M. Automated quantitative analysis of hepatitis B virus DNA by using the COBAS Amplicor HBV monitor test. *Journal of Clinical Microbiology* 1999; **37**:2793–2797. PMCID: PMC85382.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.