Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Web Appendix for "Capturing the Pool Dilution Effect in Group Testing Regression: A Bayesian Approach"

Stella Self*[1] | Christopher McMahan[2] | Stefani Mokalled[2]

[1]Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, South Carolina, United States of America

[2]School of Mathematical and Statistical Sciences, Clemson University, South Carolina, United States of America

Correspondence
*Stella Self Email: scwatson@mailbox.sc.edu

Present Address
Discovery Building, 901 Greene St, Columbia SC, 29208

APPENDIX A

This web appendix provides a discussion of the various approaches to solving the label-switching problem which commonly arises when fitting Bayesian mixture models, as well as specific details for how the label-switching problem was addressed in the simulation study and data application described in the corresponding manuscript. As noted in Section 2 of the paper, the label-switching problem may be handled in a variety of ways. One approach is to impose a set of constraints on the parameter space which discourage or preclude label-switching. Richardson and Green (1997) use such an approach for a normal mixture model with an unknown number of components[1]. In our context, this approach amounts to assuming normal biomarker distributions and constraining the mean of the negative distribution to be less than that of the positive distribution. When the mean is a nonlinear function of multiple parameters, as is the case for the gamma and log-normal distributions, the set of necessary constraints becomes more complex. This problem is compounded if the distributions are heavily skewed, in which case the mean is not a reliable indicator of central tendency, and constraints may need to involve the median or mode. In general, deriving and implementing an appropriate set of constraints is non-trivial. Jasra et al. (2020) note that imposing constraints on the parameter space can sometime inadvertently induce highly informative prior distributions whose influence is poorly understood[2]. They recommend the use of priors informed by expert opinion instead.

In practice, there is often pre-existing information regarding the biomarker levels of positive and negative individuals. When such information exists, it may be used to set informative prior distribution on the biomarker distribution parameters to avoid the label switching problem. In fact, Jasra et al. recommend using this approach in their review of solutions to the label-switching problem[2], as do Branscum et al. (2008) who consider the label switching problem specifically in the context of Bayesian mixture models for continuous-outcome ELISA test data[3]. Setting informative priors need not be complicated; it is often sufficient to chose a prior distribution whose mean and variance are informed by published point estimates of the underlying biomaker levels of positive and negative individuals, with the strength of those distributions chosen to match the degree of confidence in the available data. However, a few more nuanced approaches do exist. Chung et. al (2004) show that using informative priors can out perform a constraint-based approach in certain circumstances, (including when the necessary constraint is a linear equation

involving only 2 parameters)[4]. Kunkel and Peruggia (2020) use informative priors to solve the label switching problem for finite mixtures of normal distributions[5]. Both of these approaches rely on data-driven priors and necessitate assigning a few observations to certain distributions. In general, in the group testing context in which only the pooled responses are observed, such an approach is not possible. However, if the data contains a mixture of pooled and individual-level responses, or if a few labeled positive and negative specimens are available, these approaches or similar ones may be implemented. It is of course possible that for a particular application, there may be no previous studies or expert opinion to inform the choice of prior distributions (e.g. in the assay development stage or the earliest stages of disease emergence). However, under such conditions, researchers performing assay development will almost surely have access to individual samples from confirmed positive and negative individuals. These samples can be used to guide the selection of informative prior distributions.

An alternate approach to resolving the label-switching problem is to apply a relabeling algorithm either online or during post processing so that similar groups of observations always receive the same label. A wide variety of such algorithms exist[6,7,8]. To our knowledge, no relabeling methods have been developed for cases where the observations to be relabeled are latent (and thus the latent data themselves differ across the posterior sample). The development of such an approach is an excellent area for future investigation.

To address the label switching problem in our simulation study, we placed informative Gamma prior distributions on the biomarker concentration distribution parameters $\boldsymbol{\theta}$ as recommended by Jasra et al.(2005)[2] and Bransum et al. (2008)[3]. To mimic the scenario most likely to occur in practice, we assumed only published summary statistics of the biomarker distributions are available and that the researchers do not have direct access to measurements from known positive and negative individuals. To assess the impact of error in the available summary statistics, summary statistics were simulated under the assumption of measurement error. Specifically, for each simulated dataset we generated 100 additional individual biomarker concentrations from the negative and positive distributions with measurement error. These observations were distinct from the N observations used to create the pools. Precisely, for each dataset, for $i = 1, 2, ...100$ we generate

$$V_{i0} = Z_{i0} + \epsilon_{i0}, \quad V_{i1} = Z_{i1} + \epsilon_{i1}$$

where $Z_{i0} \overset{ind}{\sim} f_{\zeta^-}(z|\boldsymbol{\theta}_0)$, $Z_{i1} \overset{ind}{\sim} f_{\zeta^+}(z|\boldsymbol{\theta}_1)$, and $\epsilon_{ik} \overset{ind}{\sim} \text{Normal}(0, \tau^2)$, $k = 0, 1$. For $k = 0, 1$, we assume $\alpha_k \sim \text{Gamma}(\alpha_{\alpha_k}, \gamma_{\alpha_k})$ and $\gamma_k \sim \text{Gamma}(\alpha_{\gamma_k}, \gamma_{\gamma_k})$. Allowing $\overline{V}_k$ and $\overline{s}_k^2$, $k = 0, 1$, to denote the sample mean and variance of the reference samples, we define $\hat{\alpha}_k = \overline{V}_k^2/\overline{s}_k^2$ and $\hat{\gamma}_k = \overline{V}_k/\overline{s}_k^2$. Thus if a random variable $\zeta \sim \text{Gamma}(\hat{\alpha}_k, \hat{\beta}_k)$, then $E(\zeta) = \overline{V}_k$ and $\text{var}(\zeta) = \overline{s}_k^2$, that is, the mean and variance of ζ are equal to the sample mean and variance of the reference biomarker concentration samples. We wish to assign Gamma prior distributions to α_k and γ_k having mean $\hat{\alpha}_k$ and $\hat{\gamma}_k$ with relatively small variance. Specifically, we assign $\alpha_{\alpha_k} = 10\hat{\alpha}_k$, $\alpha_{\gamma_k} = 10\hat{\gamma}_k$, and $\gamma_{\alpha_k} = \gamma_{\gamma_k} = 10$, resulting in prior means of $\hat{\alpha}_k$ and $\hat{\beta}_k$ and prior variances of $\hat{\alpha}_k/10$ and $\hat{\gamma}_k/10$, respectively.

In the data application presented in Section 5 of the manuscript, we assume that the distribution of the OD readings of the individuals obey a gamma distribution; i.e., $\text{OD}_i|Y_i = y_i \sim \text{Gamma}(\alpha_y, \gamma_y)$, where OD_i denotes the optical density reading taken on the ith individual. To set priors, we assume that $\alpha_k \sim \text{Gamma}(\alpha_{\alpha_k}, \gamma_{\alpha_k})$ and $\gamma_k \sim \text{Gamma}(\alpha_{\gamma_k}, \gamma_{\gamma_k})$. By specifying the hyperparameters of these priors, we can inject information about the biomarker distributions into the analysis. For illustrational purposes, we consider four scenarios regarding the available information on the biomarker distributions; namely, a limited information, inaccurate information, high quality information, and perfect information setting. For the limited information scenario, the prior distributions on $\boldsymbol{\theta}$ are specified as in the simulation study, using observations from 25 randomly selected positive and negative individuals to serve as the reference samples. In the inaccurate information scenario, we specify $\alpha_{\alpha_0} \sim \text{Gamma}(10^{-7}, 10^{-3})$, $\gamma_{\gamma_0} \sim \text{Gamma}(10^{-3}, 10^{-1})$, $\alpha_{\alpha_1} \sim \text{Gamma}(50.625, 22.5)$, and $\gamma_{\gamma_1} \sim \text{Gamma}(0.225, 1.5)$. Noting that the mean (variances) of the OD readings in the observed data are 0.1 (0.007) and 5.3 (13.1) for the negative and positive individuals, respectively, it is easy to see that these specifications lead to egregiously misspecified priors for the biomarker distributions. For the high quality information scenario, the 900 selected observations were divided into negative and positive individuals and maximum likelihood estimation (MLE) was used to fit a gamma distribution to each subpopulation; denote the MLE estimates as $\breve{\alpha}_k$ and $\breve{\gamma}_k$. Subsequently, gamma priors having the estimated MLEs as means and a variance of 0.1 were then assigned to to the parameters. Procceeding in this fashion is equivalent to having a great deal of information about the biomarker distribution. For the perfect information scenario, no prior distributions were assigned; instead the MLEs obtained from fitting the aforementioned gamma distributions to the positive/negative subpopulations were specified as the “true" value of these parameters.

APPENDIX B

This web appendix provides additional details regarding the estimation procedure described in Section 2, including detailed specifications of the full conditional distributions and a step-by-step explanation of the Markov chain Monte Carlo (MCMC) posterior sampling algorithm. Recall that $\boldsymbol{\beta}|\boldsymbol{Y}, \boldsymbol{\psi} \sim N(\boldsymbol{\mu}_{\beta}^{*}, \boldsymbol{\Sigma}_{\beta}^{*})$. Allowing $\boldsymbol{\Sigma}_{\psi}$ to be the diagonal matrix with $\boldsymbol{\Psi}$ on the diagonal and $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, ..., \kappa_n)'$, we have that

$$\begin{aligned}\boldsymbol{\Sigma}_{\beta}^{*} &= (\boldsymbol{X}'\boldsymbol{\Sigma}_{\psi}\boldsymbol{X} + \boldsymbol{\Sigma}_{\beta}^{-1})^{-1} \\ \boldsymbol{\mu}_{\beta}^{*} &= \boldsymbol{\Sigma}_{\beta}^{*}\boldsymbol{X}'\boldsymbol{\kappa}.\end{aligned}$$

Recall also that $\tau^2|\mathscr{C}, \boldsymbol{\zeta} \sim IG\left(\alpha_{\tau}^{*}, \beta_{\tau}^{*}\right)$, where $\alpha_{\tau}^{*} = J/2 + \alpha_{\tau}$ and $\beta_{\tau}^{*} = (\boldsymbol{C} - \boldsymbol{\zeta}_{\boldsymbol{\mathcal{P}}})'(\boldsymbol{C} - \boldsymbol{\zeta}_{\boldsymbol{\mathcal{P}}})/2 + \beta_{\tau}$. Finally, recall that $Y_i|\zeta_i, \boldsymbol{\beta}, \boldsymbol{\theta}_0, \boldsymbol{\theta}_1 \sim$ Bernoulli(p_i^{*}), where

$$\begin{aligned}p_i^{*} &= \frac{p_{1i}}{p_{0i} + p_{1i}} \\ p_{0i} &= f_{\zeta^-}(\zeta_i|\boldsymbol{\theta}_0)\{1 - g^{-1}(\boldsymbol{x}_i\boldsymbol{\beta})\} \\ p_{1i} &= f_{\zeta^+}(\zeta_i|\boldsymbol{\theta}_1)g^{-1}(\boldsymbol{x}_i\boldsymbol{\beta}).\end{aligned}$$

Web Figure 1 contains a step-by-step description of the MCMC posterior sampling algorithm. Web Figure 2 provides a detailed description of the Metropolis step used to sample the ζ_is, while Web Figures 3 and 4 provide similar information for the Metropolis Hastings step used to sample $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$, respectively.

WEB APPENDIX C: ADDITIONAL SIMULATION STUDY RESULTS

Herein we provide details on the additional numerical studies that were conducted to validate the performance of the proposed methodology. These studies were conducted in the exact same fashion as those considered in Section 4 of the corresponding manuscript, with minor alterations. These alterations consider alternate group testing strategies (**Studies 1-2**) and alternate biomarker distribution configurations (**Studies 3-5**).

Study 1: This study was aimed at assessing the performance of the proposed approach under rectangular array testing; in contrast to the square arrays considered in the main document. This study considered generating individual level data under the settings of model M1 and biomarker model D2 as described in the corresponding manuscript. These individuals were then assigned to a 5×10 arrays, with the testing being simulated according to the same strategy as outlined in the main document. The proposed approach was then used to analyze the data arising from testing the rectangular arrays in the exact same fashion as was described in the manuscript. Web table 4 provides the results from this additional study. From these results, we see that the proposed approach continues to perform well, with estimation accuracy that is comparable to that attained from testing 5×5 and 10×10 arrays.

Study 2: This study was aimed at assessing the performance of our estimation methodology when tasked to analyze data arising from array testing with master pool testing. To this end, individual level data was generated under the settings of model M1 and biomarker model D1 as described in the manuscript. These individuals were then assigned master pools of size c^2, for $c \in \{3, 5\}$. Positive master pools were then formed into $c \times c$ arrays and subjected to array testing. The true optimal classification thresholds were assumed to be known, and a single round of model fitting was performed with the goal of regression estimation. Web table 5 provides the results from this additional study. From these results, we see that the regression performance of this approach is similar to that of array testing under regression model M1 and biomarker model D1 (see Web Table 1), though it requires substantially fewer tests.

Study 3: This study was aimed at assessing the performance of the proposed approach when the biomarker levels of the negative individuals are 0 or near 0, while being uncertain about the biomarker distribution of the positive individuals. In such a scenario, our method can be implemented treating $\boldsymbol{\theta}_0$ as a known constant while $\boldsymbol{\theta}_1$ is estimated. This study was performed to assess the performance of our method in such a scenario under model M1. In this study, the biomarker levels of negative and positive individuals were generated from Gamma$(0.25, 0.5)$ and Gamma$(20, 1)$ distributions, respectively, and testing

was conducted in the exact same fashion as was described in the corresponding manuscript. Model fitting followed the same strategy outlined in Section 4 with the marked difference that θ_0 was set to $(0.25, 0.5)$. Web Table 6 provides the results from this study. From these results, we see that the proposed approach continues to perform well. In fact, this setting is easier from an estimation perspective than those considered in the corresponding manuscript.

Study 4: This study was aimed at assessing the performance of the proposed approach under different biomarker models. In this study, we generate individuals true statuses according to model M1 and generate biomarker concentrations for positive and negative individuals from a $N(20, 2)$ and $N(12, 1)$, respectively. For testing, we consider only 5×5 array testing. The proposed approach was implemented in the exact same fashion as was described in the manuscript. Web Table 7 provides the results from this study. From these results, we see that the proposed approach continues to perform well.

Study 5: This study was aimed at assessing the performance of the proposed approach under different biomarker models. In this study, we generate individuals true statuses according to model M1 and generate biomarker concentrations for positive and negative individuals from a lognormal$(4, 0.25)$ and lognormal$(2.5, 3)$, respectively. For testing, we consider only 5×5 array testing. The proposed approach was implemented in the exact same fashion as was described in the manuscript. Web Table 8 provides the results from this study. From these results, we see that the proposed approach continues to perform well.

WEB APPENDIX D

This web appendix provides instructions for implementing our method in practice using the code available at https://github.com/scwatson812/GT_Dilution. After performing an initial round of testing (under any protocol), the user should create two csv files. The first file should contain a row for each individual tested with a column containing a unique individual identifier and additional columns containing the covariates for the regression model. The second file should contain a row for each test performed, with a column containing a unique identifier for each test, a column containing the observed biomarker concentration from each test, a column containing the number of individuals who contributed to each pool, and additional columns storing the unique identifiers of the individuals who contributed to each pool. Examples of these files are provided on Github. After the creation of these files, the Rscript file Data_Application can then be run to apply our method. When running the file, selection windows will automatically open to allow the user to input the individual data (first) and the pool file (second). After the code finishes running, files containing parameters estimates, standard deviations, and 95% credible intervals will be written to the working directory, along with a file containing the unique test identifiers for pools which were classified as positive. The user can then perform retesting on these positive pools (using any retesting protocol), append the retesting results to the pool csv file, and run the Data_Applicaiton Rscript again. This process can be repeated as many times as desired. Note: The prior distributions for the biomarker distribution parameters are determined from an estimate of the mean and variance specified on lines 212-216. The user should specify values relevant to their particular application.

MCMC Algorithm for Posterior Sampling:
1 : Initialize all parameters;
2 : for $g := 1$ to G do
for $i := 1$ to n do
Sample ζ_i(Metropolis Step, see Web Figure 2);
od;
Sample τ^2(Gibbs Step);
Sample $\boldsymbol{Y}$(Gibbs Step);
Sample $\boldsymbol{\Psi}$(Gibbs Step);
Sample $\boldsymbol{\beta}$(Gibbs Step);
Sample θ_{01}(Metropolis Hastings Step, see Web Figure 3);
Sample θ_{02}(Metropolis Hastings Step, see Web Figure 3);
Sample θ_{11}(Metropolis Hastings Step, see Web Figure 4);
Sample θ_{12}(Metropolis Hastings Step, see Web Figure 4);
if $g \mod 100 = 0$ then adjust std. dev. of proposal distributions to target 30%-70% acceptance ratio;
fi;

Web Figure 1 The MCMC Algorithm used to sample from the posterior distribution described in Section 2.

Metropolis Hasting Step for Sampling the ζ_is:
1 : Sample ζ_i^* from $N(\zeta_i, \sigma^2_{\zeta i})$;
2 : Compute:
$\mathrm{a}_1 = \prod_{j \in \mathcal{A}_i} f(C_j;\ \zeta^*_{\mathcal{P}_j}, \tau^2) f_{\zeta^-}(\zeta_i^*;\ \boldsymbol{\theta}_0)^{1-Y_i} f_{\zeta^+}(\zeta_i^*;\ \boldsymbol{\theta}_1)^{Y_i}$
$\mathrm{a}_2 = \prod_{j \in \mathcal{A}_i} f(C_j;\ \zeta_{\mathcal{P}_j}, \tau^2) f_{\zeta^-}(\zeta_i;\ \boldsymbol{\theta}_0)^{1-Y_i} f_{\zeta^+}(\zeta_i;\ \boldsymbol{\theta}_1)^{Y_i}$
$\mathrm{a} = \mathrm{a}_1 / a_2$;
3 : Sample r from Binomial($\min\{a, 1\}$);
4 : Update $\zeta_i = \zeta_i^* r + \zeta_i(1 - r)$;

Web Figure 2 The Metropolis step used to sample from the full conditional distributions of the ζ_is.

Metropolis Hasting Step for Sampling the θ_{0k}s:
1 : Sample θ^*_{0k} from $TN(\theta_{0k}, \sigma^2_{\theta 0k}, \mathcal{S}(1))$;
2 : Compute:
$\mathrm{a}_1 = \prod_{i=1}^N f_{\zeta^-}(\zeta_i;\ \boldsymbol{\theta}_0^*)^{1-Y_i} \pi(\boldsymbol{\theta}_0^*) \varphi(\theta_{0k};\ \theta^*_{0k}, \sigma^2_{\theta 0k}, \mathcal{S}(1))$
$\mathrm{a}_2 = \prod_{i=1}^N f_{\zeta^-}(\zeta_i;\ \boldsymbol{\theta}_0)^{1-Y_i} \pi(\boldsymbol{\theta}_0) \varphi(\theta^*_{0k};\ \theta_{0k}, \sigma^2_{\theta 0k}, \mathcal{S}(1))$
$\mathrm{a} = \mathrm{a}_1 / a_2$
3 : Sample r from Binomial($\min\{a, 1\}$);
4 : Update $\theta_{0k} = \theta^*_{0k} r + \theta_{0k}(1 - r)$;
Note: $\varphi(a;\ b, c, S)$ denotes the probability density function of a truncated normal random variable with mean b variance c.

Web Figure 3 The Metropolis Hasting step used to sample from the full conditional distributions of the $\boldsymbol{\theta}_{0k}s$.

Metropolis Hasting Step for Sampling the θ_{1k}s:
1 : Sample θ^*_{1k} from $TN(\theta_{1k}, \sigma^2_{\theta 1k}, S(1))$;
2 : Compute:
$\mathrm{a}_1 = \prod_{i=1}^{N} f_{\zeta^+}(\zeta_i;\ \boldsymbol{\theta}^*_1)^{1-Y_i}\pi(\boldsymbol{\theta}^*_1)\varphi(\theta_{1k};\ \theta^*_{1k}, \sigma^2_{\theta 1k}, S(1))$
$\mathrm{a}_2 = \prod_{i=1}^{N} f_{\zeta^+}(\zeta_i;\ \boldsymbol{\theta}_1)^{1-Y_i}\pi(\boldsymbol{\theta}_1)\varphi(\theta^*_{1k};\ \theta_{1k}, \sigma^2_{\theta 1k}, S(1))$
$\mathrm{a} = \mathrm{a}_1/a_2$
3 : Sample r from Binomial(min$\{a, 1\}$);
4 : Update $\theta_{1k} = \theta^*_{1k} r + \theta_{1k}(1 - r)$;
Note: $\varphi(a;\ b, c, S)$ denotes the probability density function of a truncated normal random variable with mean b variance c.

Web Figure 4 The Metropolis Hasting step used to sample from the full conditional distributions of the $\boldsymbol{\theta}_{1k}s$.

Web Table 1 Simulation Results: The table provides the bias of the posterior mean estimate, average estimated posterior standard deviation (SD), empirical coverage probability of 95% credible intervals (CP), and sample standard deviation (SSD) of the estimated regression coefficients from regression model M1 with biomarker model D1 obtained from individual (A1), Dorfman (A2), and array (A3) testing. The testing accuracy using the estimated optimal threshold (ET) is reported in the form of true negative (TN), true positive (TP), false negative (FN), and false positive (FP) rates. The testing accuracy using the true optimal threshold (TT) is included for comparison. The average number of tests used (#) is also reported.

				Estimation				Classification				
N^*	$A^\dagger$	$c^\mp$		β_0	β_1	β_2		TN	TP	FN	FP	#
900	A1		Bias (SD)	-0.22(0.47)	0.08(0.25)	0.10(0.39)	ET	1.00	1.00	0.00	0.00	900.00
			CP95(SSD)	0.95(0.48)	0.94(0.25)	0.94(0.41)	TT	1.00	1.00	0.00	0.00	
	A2	3	Bias (SD)	-0.20(0.47)	0.08(0.25)	0.08(0.39)	ET	1.00	1.00	0.00	0.00	431.83
			CP95(SSD)	0.94(0.52)	0.95(0.27)	0.95(0.38)	TT	1.00	1.00	0.00	0.00	
		5	Bias (SD)	-0.16(0.47)	0.08(0.25)	0.03(0.39)	ET	1.00	0.99	0.01	0.00	392.93
			CP95(SSD)	0.95(0.49)	0.95(0.26)	0.95(0.39)	TT	1.00	1.00	0.00	0.00	
		10	Bias (SD)	-0.22(0.49)	0.10(0.26)	0.07(0.40)	ET	1.00	0.98	0.02	0.00	471.46
			CP95(SSD)	0.91(0.55)	0.93(0.29)	0.91(0.45)	TT	1.00	0.98	0.02	0.00	
	A3	3	Bias (SD)	-0.17(0.47)	0.07(0.25)	0.05(0.39)	ET	1.00	1.00	0.00	0.00	653.73
			CP95(SSD)	0.96(0.46)	0.95(0.25)	0.94(0.40)	TT	1.00	1.00	0.00	0.00	
		5	Bias (SD)	-0.16(0.47)	0.07(0.25)	0.04(0.39)	ET	1.00	0.99	0.01	0.00	438.23
			CP95(SSD)	0.95(0.49)	0.94(0.26)	0.94(0.41)	TT	1.00	0.99	0.01	0.00	
		10	Bias (SD)	-0.16(0.47)	0.06(0.25)	0.05(0.39)	ET	1.00	0.96	0.04	0.00	356.42
			CP95(SSD)	0.94(0.48)	0.96(0.27)	0.96(0.39)	TT	1.00	0.96	0.04	0.00	
1800	A1		Bias (SD)	-0.08(0.32)	0.03(0.17)	0.02(0.27)	ET	1.00	1.00	0.00	0.00	1800.00
			CP95(SSD)	0.94(0.33)	0.94(0.18)	0.93(0.28)	TT	1.00	1.00	0.00	0.00	
	A2	3	Bias (SD)	-0.10(0.32)	0.04(0.17)	0.04(0.27)	ET	1.00	1.00	0.00	0.00	862.73
			CP95(SSD)	0.93(0.35)	0.94(0.18)	0.92(0.28)	TT	1.00	1.00	0.00	0.00	
		5	Bias (SD)	-0.10(0.32)	0.04(0.18)	0.03(0.27)	ET	1.00	1.00	0.00	0.00	781.73
			CP95(SSD)	0.95(0.34)	0.94(0.18)	0.94(0.30)	TT	1.00	1.00	0.00	0.00	
		10	Bias (SD)	-0.08(0.33)	0.02(0.18)	0.02(0.28)	ET	1.00	0.98	0.02	0.00	937.82
			CP95(SSD)	0.92(0.37)	0.92(0.20)	0.93(0.30)	TT	1.00	0.98	0.02	0.00	
	A3	3	Bias (SD)	-0.07(0.32)	0.03(0.17)	0.00(0.27)	ET	1.00	1.00	0.00	0.00	1307.12
			CP95(SSD)	0.96(0.32)	0.94(0.18)	0.95(0.27)	TT	1.00	1.00	0.00	0.00	
		5	Bias (SD)	-0.08(0.32)	0.03(0.17)	0.03(0.27)	ET	1.00	0.99	0.01	0.00	874.26
			CP95(SSD)	0.94(0.34)	0.94(0.18)	0.95(0.28)	TT	1.00	0.99	0.01	0.00	
		10	Bias (SD)	-0.10(0.33)	0.04(0.18)	0.04(0.27)	ET	1.00	0.96	0.04	0.00	714.27
			CP95(SSD)	0.94(0.32)	0.95(0.17)	0.93(0.28)	TT	1.00	0.96	0.04	0.00	

∗N indicates the sample size.

† A indicates the retesting protocol.

∓ c indicates the pool size.

Web Table 2 Simulation Results: The table provides the bias of the posterior mean estimate, average estimated posterior standard deviation (SD), empirical coverage probability of 95% credible intervals (CP), and sample standard deviation (SSD) of the estimated regression coefficients from regression model M2 with biomarker model D1 obtained from individual (A1), Dorfman (A2), and array (A3) testing. The testing accuracy using the estimated optimal threshold (ET) is reported in the form of true negative (TN), true positive (TP), false negative (FN), and false positive (FP) rates. The testing accuracy using the true optimal threshold (TT) is included for comparison. The average number of tests used (#) is also reported.

				Estimation				Classification				
N^*	$A^\dagger$	$c^\mp$		β_0	β_1	β_2		TN	TP	FN	FP	#
900	A1		Bias (SD)	-0.06(0.22)	0.00(0.11)	0.04(0.25)	ET	1.00	1.00	0.00	0.00	900.00
			CP95(SSD)	0.94(0.23)	0.95(0.12)	0.94(0.27)	TT	1.00	1.00	0.00	0.00	
	A2	3	Bias (SD)	-0.04(0.22)	0.00(0.11)	0.02(0.25)	ET	1.00	1.00	0.00	0.00	594.34
			CP95(SSD)	0.94(0.23)	0.96(0.11)	0.95(0.25)	TT	1.00	1.00	0.00	0.00	
		5	Bias (SD)	-0.05(0.22)	0.00(0.11)	0.03(0.25)	ET	1.00	1.00	0.00	0.00	618.13
			CP95(SSD)	0.95(0.23)	0.95(0.12)	0.95(0.26)	TT	1.00	1.00	0.00	0.00	
		10	Bias (SD)	-0.03(0.23)	0.01(0.11)	0.00(0.27)	ET	1.00	0.99	0.01	0.00	747.62
			CP95(SSD)	0.92(0.27)	0.92(0.13)	0.90(0.32)	TT	1.00	0.99	0.01	0.00	
	A3	3	Bias (SD)	-0.05(0.22)	0.00(0.11)	0.04(0.25)	ET	1.00	1.00	0.00	0.00	753.95
			CP95(SSD)	0.93(0.23)	0.95(0.11)	0.94(0.26)	TT	1.00	1.00	0.00	0.00	
		5	Bias (SD)	-0.06(0.22)	0.01(0.11)	0.04(0.25)	ET	1.00	0.99	0.01	0.00	605.21
			CP95(SSD)	0.94(0.23)	0.96(0.11)	0.95(0.26)	TT	1.00	1.00	0.00	0.00	
		10	Bias (SD)	0.00(0.22)	0.00(0.11)	-0.02(0.25)	ET	1.00	0.98	0.02	0.00	676.37
			CP95(SSD)	0.94(0.24)	0.93(0.12)	0.94(0.26)	TT	1.00	0.98	0.02	0.00	
1800	A1		Bias (SD)	-0.03(0.16)	0.01(0.08)	0.01(0.18)	ET	1.00	1.00	0.00	0.00	1800.00
			CP95(SSD)	0.96(0.15)	0.96(0.08)	0.95(0.18)	TT	1.00	1.00	0.00	0.00	
	A2	3	Bias (SD)	-0.03(0.16)	0.00(0.08)	0.03(0.18)	ET	1.00	1.00	0.00	0.00	1190.83
			CP95(SSD)	0.92(0.17)	0.94(0.08)	0.93(0.19)	TT	1.00	1.00	0.00	0.00	
		5	Bias (SD)	-0.02(0.16)	0.00(0.08)	0.01(0.18)	ET	1.00	1.00	0.00	0.00	1235.73
			CP95(SSD)	0.94(0.17)	0.94(0.08)	0.94(0.19)	TT	1.00	1.00	0.00	0.00	
		10	Bias (SD)	-0.02(0.16)	0.00(0.08)	0.01(0.18)	ET	1.00	0.99	0.01	0.00	1504.40
			CP95(SSD)	0.92(0.18)	0.94(0.08)	0.92(0.21)	TT	1.00	0.99	0.01	0.00	
	A3	3	Bias (SD)	-0.03(0.16)	0.00(0.08)	0.03(0.18)	ET	1.00	1.00	0.00	0.00	1506.97
			CP95(SSD)	0.96(0.16)	0.95(0.08)	0.94(0.18)	TT	1.00	1.00	0.00	0.00	
		5	Bias (SD)	-0.01(0.16)	0.00(0.08)	0.01(0.17)	ET	1.00	0.99	0.01	0.00	1213.08
			CP95(SSD)	0.94(0.16)	0.94(0.08)	0.94(0.18)	TT	1.00	1.00	0.00	0.00	
		10	Bias (SD)	0.00(0.16)	-0.01(0.08)	-0.02(0.18)	ET	1.00	0.98	0.02	0.00	1345.42
			CP95(SSD)	0.93(0.18)	0.93(0.08)	0.92(0.20)	TT	1.00	0.98	0.02	0.00	

∗N indicates the sample size.

† A indicates the retesting protocol.

∓ c indicates the pool size.

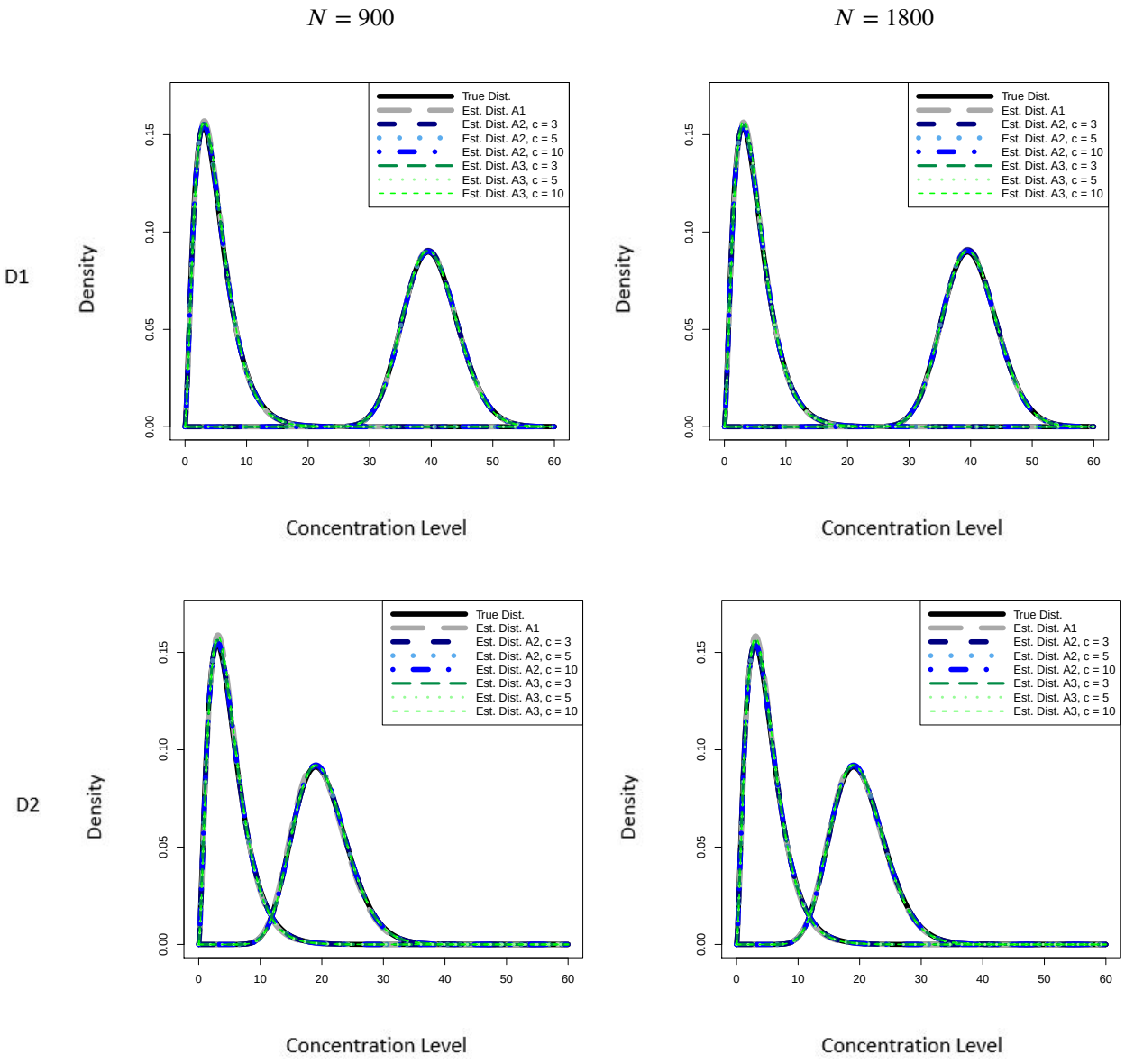
Web Table 3 Simulation Results: The table provides the bias of the posterior mean estimate, average estimated posterior standard deviation (SD), empirical coverage probability of 95% credible intervals (CP), and sample standard deviation (SSD) of the estimated regression coefficients from regression model M2 with biomarker model D2 obtained from individual (A1), Dorfman (A2), and array (A3) testing. The testing accuracy using the estimated optimal threshold (ET) is reported in the form of true negative (TN), true positive (TP), false negative (FN), and false positive (FP) rates. The testing accuracy using the true optimal threshold (TT) is included for comparison. The average number of tests used (#) is also reported.

				Estimation				Classification				
N^*	$A^\dagger$	$c^\mp$		β_0	β_1	β_2		TN	TP	FN	FP	#
900	A1		Bias (SD)	-0.02(0.30)	0.00(0.13)	-0.03(0.32)	ET	0.96	0.98	0.02	0.04	900.00
			CP95(SSD)	0.93(0.66)	0.94(0.13)	0.95(0.51)	TT	0.96	0.98	0.02	0.04	
	A2	3	Bias (SD)	-0.13(0.33)	0.01(0.13)	0.11(0.34)	ET	0.97	0.93	0.07	0.03	651.05
			CP95(SSD)	0.94(0.37)	0.94(0.13)	0.95(0.37)	TT	0.97	0.94	0.06	0.03	
		5	Bias (SD)	-0.13(0.35)	0.02(0.14)	0.08(0.36)	ET	0.97	0.91	0.09	0.03	657.60
			CP95(SSD)	0.93(0.45)	0.95(0.15)	0.94(0.44)	TT	0.97	0.91	0.09	0.03	
		10	Bias (SD)	-0.14(0.40)	0.01(0.15)	0.08(0.41)	ET	0.97	0.92	0.08	0.03	744.48
			CP95(SSD)	0.96(0.66)	0.95(0.15)	0.95(0.48)	TT	0.97	0.92	0.08	0.03	
	A3	3	Bias (SD)	-0.08(0.31)	0.01(0.12)	0.06(0.32)	ET	0.98	0.90	0.10	0.02	787.74
			CP95(SSD)	0.95(0.39)	0.94(0.13)	0.96(0.39)	TT	0.98	0.89	0.11	0.02	
		5	Bias (SD)	-0.12(0.34)	0.02(0.13)	0.09(0.35)	ET	0.98	0.86	0.14	0.02	647.45
			CP95(SSD)	0.92(0.44)	0.95(0.13)	0.93(0.44)	TT	0.98	0.88	0.12	0.02	
		10	Bias (SD)	-0.09(0.24)	0.01(0.10)	0.03(0.24)	ET	0.98	0.85	0.15	0.02	656.62
			CP95(SSD)	0.95(0.57)	0.94(0.10)	0.95(0.32)	TT	0.97	0.87	0.13	0.03	
1800	A1		Bias (SD)	0.02(0.19)	-0.01(0.08)	-0.01(0.20)	ET	0.96	0.98	0.02	0.04	1800.00
			CP95(SSD)	0.91(0.20)	0.92(0.09)	0.95(0.20)	TT	0.96	0.98	0.02	0.04	
	A2	3	Bias (SD)	-0.05(0.20)	0.01(0.09)	0.04(0.21)	ET	0.97	0.94	0.06	0.03	1307.71
			CP95(SSD)	0.95(0.21)	0.95(0.09)	0.95(0.20)	TT	0.97	0.94	0.06	0.03	
		5	Bias (SD)	-0.05(0.21)	0.00(0.09)	0.04(0.22)	ET	0.97	0.91	0.09	0.03	1314.64
			CP95(SSD)	0.96(0.21)	0.96(0.09)	0.94(0.22)	TT	0.97	0.91	0.09	0.03	
		10	Bias (SD)	-0.10(0.26)	0.01(0.11)	0.03(0.26)	ET	0.97	0.92	0.08	0.03	1477.56
			CP95(SSD)	0.95(0.77)	0.95(0.10)	0.96(0.42)	TT	0.97	0.92	0.08	0.03	
	A3	3	Bias (SD)	-0.02(0.20)	0.00(0.09)	0.01(0.21)	ET	0.98	0.90	0.10	0.02	1575.45
			CP95(SSD)	0.93(0.22)	0.95(0.09)	0.92(0.22)	TT	0.98	0.89	0.11	0.02	
		5	Bias (SD)	-0.01(0.20)	0.00(0.09)	0.01(0.21)	ET	0.97	0.86	0.14	0.03	1298.77
			CP95(SSD)	0.93(0.22)	0.96(0.09)	0.95(0.21)	TT	0.97	0.87	0.13	0.03	
		10	Bias (SD)	-0.08(0.23)	0.01(0.10)	0.06(0.24)	ET	0.97	0.86	0.14	0.03	1318.60
			CP95(SSD)	0.94(0.26)	0.95(0.10)	0.94(0.24)	TT	0.97	0.87	0.13	0.03	

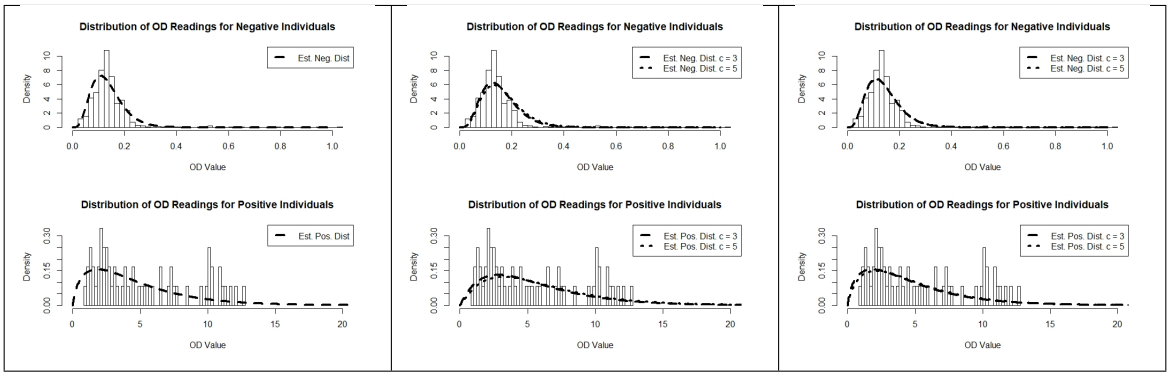
∗N indicates the sample size.

† A indicates the retesting protocol.

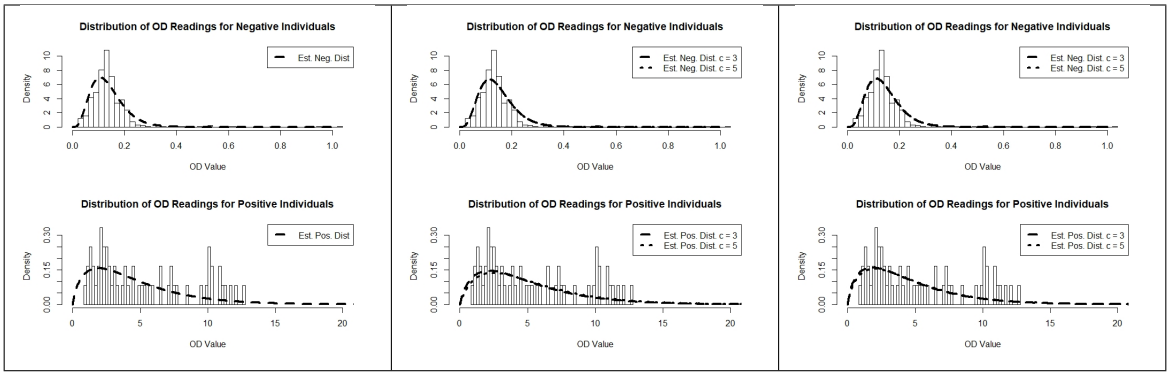
∓ c indicates the pool size.



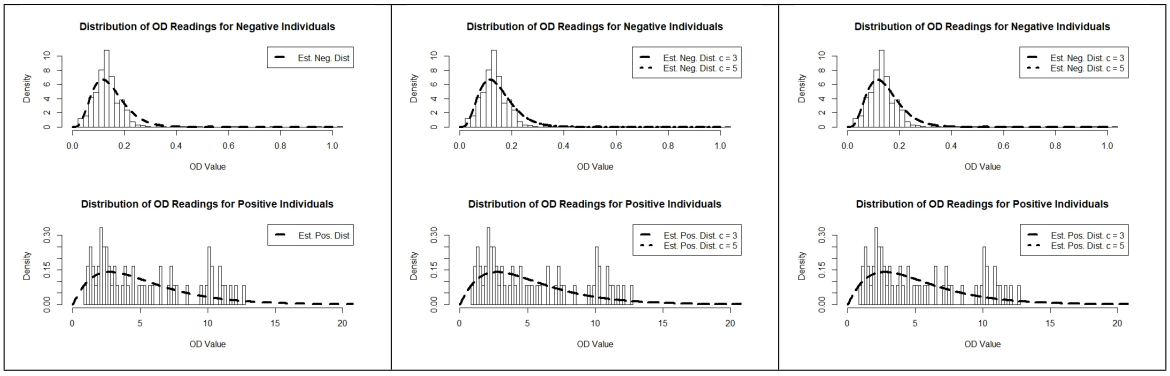
Web Figure 5 Simulation Results: Summary of the posterior mean estimates of $\hat{\theta}$ from model M2, under biomarker distributions D1 (top row) and D2 (bottom row) obtained from individual (A1), Dorfman (A2), and array (A3) testing. The curves represent the average estimate of the biomarker distributions, plotted against the true densities. The left column corresponds to $N = 900$ and the right to $N = 1800$.



Web Figure 6 Data Application Results: The figure displays both the empirical and the average estimated biomarker distributions for negative and positive individuals obtained from individual (left), Dorfman (center) and array (right) testing using limited information for the HBV data application. Note that the axis scales differ between the histograms of the negative and positive individuals.



Web Figure 7 Data Application Results: The figure displays both the empirical and the average estimated biomarker distributions for negative and positive individuals obtained from individual (left), Dorfman (center) and array (right) testing using high quality information for the HBV data application. Note that the axis scales differ between the histograms of the negative and positive individuals.



Web Figure 8 Data Application Results: The figure displays both the empirical and the average estimated biomarker distributions for negative and positive individuals obtained from individual (left), Dorfman (center) and array (right) testing using perfect information for the HBV data application. Note that the axis scales differ between the histograms of the negative and positive individuals.

Web Table 4 Simulation Results: The table provides the bias of the posterior mean estimate, average estimated posterior standard deviation (SD), empirical coverage probability of 95% credible intervals (CP), and sample standard deviation (SSD) of the estimated regression coefficients from regression model M1 with biomarker model D2 obtained from array testing with 105 arrays. The testing accuracy using the estimated optimal threshold (ET) is reported in the form of true negative (TN), true positive (TP), false negative (FN), and false positive (FP) rates. The testing accuracy using the true optimal threshold (TT) is included for comparison. The average number of tests used (#) is also reported.

				Estimation				Classification				
N^*	$A^\dagger$	$c^\mp$		β_0	β_1	β_2		TN	TP	FN	FP	#
900	A3	10 ×5	Bias (SD)	-0.32(0.74)	0.12(0.37)	0.10(0.51)	ET	0.98	0.80	0.20	0.02	447.66
			CP95(SSD)	0.92(0.86)	0.93(0.40)	0.93(0.55)	TT	0.97	0.84	0.16	0.03	
1800			Bias (SD)	-0.13(0.47)	0.05(0.24)	0.01(0.33)	ET	0.98	0.80	0.20	0.02	889.70
			CP95(SSD)	0.94(0.50)	0.93(0.26)	0.95(0.34)	TT	0.97	0.84	0.16	0.03	

∗N indicates the sample size.

† A indicates the retesting protocol.

∓ c indicates the pool size.

Web Table 5 Simulation Results: The table provides the bias of the posterior mean estimate, average estimated posterior standard deviation (SD), empirical coverage probability of 95% credible intervals (CP), and sample standard deviation (SSD) of the estimated regression coefficients from regression model M1 with biomarker model D1 obtained from array testing with master pool testing. The testing accuracy using the true optimal threshold (TT) is reported in the form of true negative (TN), true positive (TP), false negative (FN), and false positive (FP) rates. The average number of tests used (#) is also reported.

			Estimation				Classification				
N^*	$c^\mp$		β_0	β_1	β_2		TN	TP	FN	FP	#
900	3	Bias (SD)	-0.13(0.47)	0.07(0.25)	0.01(0.39)	TT	1.00	0.98	0.02	0.00	384.63
		CP95(SSD)	0.95(0.52)	0.92(0.28)	0.93(0.41)						
	5	Bias (SD)	-0.16(0.47)	0.05(0.26)	0.05(0.40)	TT	1.00	0.96	0.04	0.00	366.98
		CP95(SSD)	0.94(0.49)	0.95(0.25)	0.94(0.42)						
1800	3	Bias (SD)	-0.08(0.32)	0.03(0.18)	0.01(0.27)	TT	1.00	0.98	0.02	0.00	772.18
		CP95(SSD)	0.93(0.34)	0.92(0.19)	0.94(0.27)						
	5	Bias (SD)	-0.07(0.33)	0.02(0.18)	0.03(0.27)	TT	1.00	0.96	0.04	0.00	740.82
		CP95(SSD)	0.95(0.33)	0.95(0.18)	0.94(0.29)						

∗N indicates the sample size.

∓ c indicates the pool size.

Web Table 6 Simulation Results: The table provides the bias of the posterior mean estimate, average estimated posterior standard deviation (SD), empirical coverage probability of 95% credible intervals (CP), and sample standard deviation (SSD) of the estimated regression coefficients from regression model M1 with the biomarker distribution of negative individuals concentrated at 0, obtained from individual (A1), Dorfman (A2), and array (A3) testing. The testing accuracy using the estimated optimal threshold (ET) is reported in the form of true negative (TN), true positive (TP), false negative (FN), and false positive (FP) rates. The testing accuracy using the true optimal threshold (TT) is included for comparison. The average number of tests used (#) is also reported.

				Estimation				Classification				
N	A	c		β_0	β_1	β_2		TN	TP	FN	FP	#
900	A1		Bias (SD)	-0.15(0.47)	0.05(0.25)	0.04(0.39)	ET	1.00	1.00	0.00	0.00	900.00
			CP95(SSD)	0.95(0.49)	0.94(0.26)	0.94(0.41)	TT	1.00	1.00	0.00	0.00	
	A2	3	Bias (SD)	-0.17(0.47)	0.06(0.25)	0.08(0.39)	ET	1.00	1.00	0.00	0.00	431.84
			CP95(SSD)	0.94(0.52)	0.96(0.25)	0.92(0.44)	TT	1.00	1.00	0.00	0.00	
		5	Bias (SD)	-0.14(0.47)	0.05(0.25)	0.06(0.39)	ET	1.00	1.00	0.00	0.00	390.33
			CP95(SSD)	0.91(0.55)	0.93(0.28)	0.93(0.42)	TT	1.00	1.00	0.00	0.00	
		10	Bias (SD)	-0.20(0.48)	0.08(0.26)	0.06(0.39)	ET	1.00	0.99	0.01	0.00	456.76
			CP95(SSD)	0.92(0.52)	0.94(0.27)	0.95(0.43)	TT	1.00	1.00	0.00	0.00	
	A3	3	Bias (SD)	-0.19(0.48)	0.08(0.26)	0.05(0.39)	ET	1.00	1.00	0.00	0.00	653.64
			CP95(SSD)	0.94(0.50)	0.95(0.26)	0.95(0.40)	TT	1.00	1.00	0.00	0.00	
		5	Bias (SD)	-0.15(0.47)	0.05(0.25)	0.04(0.39)	ET	1.00	1.00	0.00	0.00	436.26
			CP95(SSD)	0.95(0.48)	0.93(0.27)	0.94(0.40)	TT	1.00	1.00	0.00	0.00	
		10	Bias (SD)	-0.19(0.48)	0.07(0.26)	0.05(0.40)	ET	1.00	0.99	0.01	0.00	345.15
			CP95(SSD)	0.93(0.53)	0.91(0.30)	0.93(0.41)	TT	1.00	0.99	0.01	0.00	
1800	A1		Bias (SD)	-0.09(0.33)	0.03(0.18)	0.02(0.27)	ET	1.00	1.00	0.00	0.00	1800.00
			CP95(SSD)	0.94(0.33)	0.95(0.18)	0.94(0.28)	TT	1.00	1.00	0.00	0.00	
	A2	3	Bias (SD)	-0.10(0.33)	0.04(0.18)	0.03(0.27)	ET	1.00	1.00	0.00	0.00	861.82
			CP95(SSD)	0.95(0.34)	0.94(0.19)	0.92(0.28)	TT	1.00	1.00	0.00	0.00	
		5	Bias (SD)	0.00(0.32)	-0.03(0.17)	0.01(0.27)	ET	1.00	1.00	0.00	0.00	777.01
			CP95(SSD)	0.91(0.36)	0.89(0.21)	0.95(0.29)	TT	1.00	1.00	0.00	0.00	
		10	Bias (SD)	-0.08(0.33)	0.03(0.18)	0.04(0.27)	ET	1.00	0.99	0.01	0.00	926.22
			CP95(SSD)	0.95(0.34)	0.94(0.19)	0.97(0.26)	TT	1.00	1.00	0.00	0.00	
	A3	3	Bias (SD)	-0.09(0.33)	0.04(0.18)	0.02(0.27)	ET	1.000	0.99	0.01	0.00	1307.49
			CP95(SSD)	0.96(0.33)	0.96(0.18)	0.95(0.28)	TT	1.00	1.00	0.00	0.00	
		5	Bias (SD)	-0.13(0.33)	0.05(0.18)	0.04(0.27)	ET	1.00	1.00	0.00	0.00	868.98
			CP95(SSD)	0.92(0.35)	0.94(0.19)	0.96(0.27)	TT	1.00	1.00	0.00	0.00	
		10	Bias (SD)	-0.04(0.33)	0.00(0.18)	-0.01(0.28)	ET	1.00	0.99	0.01	0.00	688.36
			CP95(SSD)	0.93(0.37)	0.91(0.21)	0.93(0.34)	TT	1.00	0.99	0.01	0.00	

∗N indicates the sample size.

† A indicates the retesting protocol.

∓ c indicates the pool size.

Web Table 7 Simulation Results: The table provides the bias of the posterior mean estimate, average estimated posterior standard deviation (SD), empirical coverage probability of 95% credible intervals (CP), and sample standard deviation (SSD) of the estimated regression coefficients from regression model M1 with normal biomarker distributions obtained from array testing with 5×5 arrays. The testing accuracy using the estimated optimal threshold (ET) is reported in the form of true negative (TN), true positive (TP), false negative (FN), and false positive (FP) rates. The testing accuracy using the true optimal threshold (TT) is included for comparison. The average number of tests used (#) is also reported.

				Estimation				Classification				
N^*	$A^\dagger$	$c^\mp$		β_0	β_1	β_2		TN	TP	FN	FP	#
900	A3	5	Bias (SD)	-0.18(0.49)	0.07(0.26)	0.03(0.40)	ET	1.00	0.91	0.09	0.00	442.42
			CP95(SSD)	0.91(0.55)	0.95(0.28)	0.95(0.41)	TT	1.00	0.91	0.09	0.00	
1800			Bias (SD)	-0.09(0.33)	0.04(0.18)	0.02(0.28)	ET	1.00	0.91	0.09	0.00	885.21
			CP95(SSD)	0.94(0.35)	0.95(0.19)	0.94(0.29)	TT	1.00	0.91	0.09	0.00	

∗N indicates the sample size.

† A indicates the retesting protocol.

∓ c indicates the pool size.

Web Table 8 Simulation Results: The table provides the bias of the posterior mean estimate, average estimated posterior standard deviation (SD), empirical coverage probability of 95% credible intervals (CP), and sample standard deviation (SSD) of the estimated regression coefficients from regression model M1 with lognormal biomarker distributions obtained from array testing with 5×5 arrays. The testing accuracy using the estimated optimal threshold (ET) is reported in the form of true negative (TN), true positive (TP), false negative (FN), and false positive (FP) rates. The testing accuracy using the true optimal threshold (TT) is included for comparison. The average number of tests used (#) is also reported.

				Estimation				Classification				
N^*	$A^\dagger$	$c^\mp$		β_0	β_1	β_2		TN	TP	FN	FP	#
900	A3	5	Bias (SD)	-0.16(0.48)	0.06(0.26)	0.05(0.40)	ET	1.00	0.96	0.04	0.00	447.13
			CP95(SSD)	0.95(0.51)	0.94(0.27)	0.95(0.40)	TT	1.00	0.96	0.04	0.00	
1800			Bias (SD)	-0.12(0.33)	0.05(0.18)	0.05(0.28)	ET	1.00	0.96	0.04	0.00	890.54
			CP95(SSD)	0.96(0.33)	0.94(0.18)	0.96(0.26)	TT	1.00	0.96	0.04	0.00	

∗N indicates the sample size.

† A indicates the retesting protocol.

∓ c indicates the pool size.

References

1. Richardson S, Green PJ. On Bayesian Analysis of Mixtures with an Unknown Number of Components (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 1997; 59(4): 731-792. doi: https://doi.org/10.1111/1467-9868.00095

2. Jasra A, Holmes CC, Stephens DA. Markov Chain Monte Carlo Methods and the Label Switching Problem in Bayesian Mixture Modeling. *Statistical Science* 2005; 20(1): 50–67.

3. Branscum AJ, Johnson WO, Hanson TE, Gardner IA. Bayesian semiparametric ROC curve estimation and disease diagnosis. *Statistics in Medicine* 2008; 27(13): 2474-2496. doi: https://doi.org/10.1002/sim.3250

4. Chung H, Loken E, Schafer JL. Difficulties in Drawing Inferences With Finite-Mixture Models. *The American Statistician* 2004; 58(2): 152-158. doi: 10.1198/0003130043286

5. Kunkel D, Peruggia M. Anchored Bayesian Gaussian mixture models. *Electronic Journal of Statistics* 2020; 14(2): 3869 – 3913. doi: 10.1214/20-EJS1756

6. Rodríguez CE, Walker SG. Label Switching in Bayesian Mixture Models: Deterministic Relabeling Strategies. *Journal of Computational and Graphical Statistics* 2014; 23(1): 25-45. doi: 10.1080/10618600.2012.735624

7. Stephens M. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2000; 62(4): 795-809. doi: https://doi.org/10.1111/1467-9868.00265

8. Papastamoulis P, Iliopoulos G. An Artificial Allocations Based Solution to the Label Switching Problem in Bayesian Analysis of Mixtures of Distributions. *Journal of Computational and Graphical Statistics* 2010; 19(2): 313-331. doi: 10.1198/jcgs.2010.09008

How to cite this article: Self, S., C. McMahan, and S. Mokalled (2021), A Bayesian Group Testing Regression Procedure, *Statistics in Medicine*, *XX*.