Biometrical Journal **xx** (2020) xx, zzz–zzz / DOI: 10.1002/bimj.200100000

Estimating the prevalence of two or more diseases using outcomes from multiplex group testing

Md S. Warasi [1], **Joshua M. Tebbs*** [2], **Christopher S. McMahan** [3], and **Christopher R. Bilder** [4]

[1] Department of Mathematics and Statistics, Radford University, Radford, VA 24142, USA
[2] Department of Statistics, University of South Carolina, Columbia, SC 29208, USA
[3] School of Mathematical and Statistical Sciences, Clemson University, Clemson, SC 29634, USA
[4] Department of Statistics, University of Nebraska-Lincoln, Lincoln, NE 68583, USA

Received zzz, revised zzz, accepted zzz

When screening a population for infectious diseases, pooling individual specimens (e.g., blood, swabs, urine, etc.) can provide enormous cost savings when compared to testing specimens individually. In the biostatistics literature, testing pools of specimens is commonly known as group testing or pooled testing. Although estimating a population-level prevalence with group testing data has received a large amount of attention, most of this work has focused on applications involving a single disease, such as HIV. Modern methods of screening now involve testing pools and individuals for multiple diseases simultaneously through the use of multiplex assays. Hou et al. (2017, *Biometrics*) and Hou et al. (2020, *Biostatistics*) recently proposed group testing protocols for multiplex assays. However, only the aspects of case identification were studied in those articles (i.e., estimation was not considered). In this article, we describe Bayesian methods to estimate population-level disease probabilities from implementing these protocols or any other multiplex group testing protocol which might be carried out in practice. Our estimation methods can be used with multiplex assays for two or more diseases while incorporating the possibility of test misclassification for each disease. We use chlamydia and gonorrhea testing data collected at the State Hygienic Laboratory at the University of Iowa to illustrate our work. We also provide an online R resource practitioners can use to implement the methods in this article.

Key words: Bayesian estimation; Latent response; Multiplex assay; Pooled testing; Screening

1 Introduction

In group testing applications, individual specimens are combined into pools and tests are performed on the pools for a binary outcome (e.g., positive/negative, etc.). Individuals from negative pools are diagnosed to be negative, while individuals from positive pools are tested further to determine which ones are positive. Dorfman (1943) is credited with introducing this method to test American soldiers for syphilis during World War II. In this seminal paper, non-overlapping pools of individual specimens are formed in the first stage of testing, and positive pools are resolved by testing each individual one by one in the second stage. When the probability of disease is small, group testing protocols that implement a larger number of stages (Quinn et al., 2000; Pilcher et al., 2005) and/or overlapping pools (Martin et al., 2013) can further reduce the number of tests needed to identify positive individuals. The infectious disease literature documents numerous applications of group testing, including for HIV (Westreich et al., 2008), HBV and HCV (Hourfar et al., 2008; Stramer et al., 2013), chlamydia and gonorrhea (Lindan et al., 2005), and West Nile virus (Busch et al., 2005). Group testing has most recently played an important role in reducing laboratory testing loads when diagnosing individuals for SARS-CoV-2 (Abdalhamid et al., 2020; Bilder, Tebbs, and McMahan, 2021).

*Corresponding author: e-mail: tebbs@stat.sc.edu, Phone: +1 803-576-8765, Fax: +1 803-777-4048

Statistical research in group testing generally falls into one of two categories: case identification and estimation. In the case identification problem, the goal is to characterize the efficiency and accuracy of group testing protocols with the usual objectives of minimizing the expected number of tests and/or maximizing classification accuracy (Kim et al., 2007). On the other hand, the estimation problem involves estimating a population-level probability of disease (Liu et al., 2012; Huang et al., 2017) or covariate-adjusted probabilities by using regression methods (Delaigle and Meister, 2011; Wang et al., 2014; McMahan et al., 2017). In both the case identification and estimation problems, the performance of group testing and its ability to offer cost-effective screening and surveillance has been noted extensively. However, most of the existing research in group testing, including those articles referenced above, has focused on a single disease.

In this article, we consider the estimation problem in group testing when multiplex assays are used to test specimens for multiple diseases simultaneously. Our work is motivated by the screening practices for *Chlamydia trachomatis* (CT) and *Neisseria gonorrhoeae* (NG), the bacteria that lead to chlamydia and gonorrhea, respectively. Individuals are often tested for these bacteria using duplex assays. For example, the Aptima Combo 2 Assay (Hologic, Marlborough, USA) and the cobas CT/NG assay (Roche, Basel, Switzerland) are two assays that are commonly used by laboratories for this purpose. Recent advances in technology have seen the development of multiplex assays for a variety of infections. For example, the BD MAX CT/GC/TV Assay (Becton, Dickinson and Company, Franklin Lakes, USA) tests for CT, NG, and *Trichomonas vaginalis* (TV) simultaneously, and the Allplex STI Essential Assay (Seegene, Seoul, South Korea) detects CT, NG, TV, and *Mycoplasma genitalium* (de Salazar et al., 2019). Commonly used triplex assays for HIV, HBV, and HCV have been compared in Stramer et al. (2013). Multiplex assays have been recently authorized by the U.S. Food and Drug Administration for simultaneous detection of influenza and SARS-CoV-2 (Roche, 2020).

When compared to research for single diseases, estimating multiple population-level disease probabilities from group testing data has received far less attention. The original work on this problem is attributed to Hughes-Oliver and Rosenberger (2000), who developed D-optimal designs for estimation when assays are 100% accurate. Ding and Xiong (2015) and Li et al. (2017) proposed optimal designs to estimate probabilities for multiple independent diseases and two correlated diseases, respectively, while allowing for testing error. A practical limitation in these articles is that the methods are based only on outcomes from initially formed pools; that is, subsequent testing results from resolving positive pools are not incorporated. Another limitation is that the assay accuracy rates are assumed to be 100% for each disease or they are assumed to be known. In some applications, reasonable estimates may be available for disease-specific sensitivities and specificities. A more flexible approach is to regard these population-level parameters as unknown and then estimate them simultaneously with the disease probabilities. This is the approach we espouse in this article.

Estimation for group testing with multiplex assays is challenging. When incorporating test misclassification, (a) the true disease statuses of each specimen tested are latent and are likely correlated and (b) the available data from group testing protocols may include multiple (possibly misclassified) testing outcomes on the same individual. Tebbs, McMahan, and Bilder (2013) and Warasi et al. (2016) proposed estimation methods for two-stage Dorfman group testing protocols. In this article, we propose a Bayesian framework to estimate population-level disease probabilities and assay accuracy probabilities from *any* group testing protocol which uses multiplex assays. This includes higher-stage hierarchical and array-based protocols recently proposed in Hou et al. (2017) and Hou et al. (2020), respectively, and any other protocol that might be used in practice. In other words, the model framework we present herein is invariant to how the multiplex outcomes are recorded. Therefore, we can compare the accuracy and precision of population-level estimates for different group testing protocols which use multiplex assays. Until now, such a comparison has been missing in the literature.

Subsequent sections of this article are organized as follows. In Section 2, we describe notation, state assumptions, and derive the observed data likelihood which is applicable for any group testing protocol using multiplex assays. In Section 3, we present the specifics of our Bayesian estimation approach when assay accuracy probabilities (sensitivities and specificities) are known, including prior model specification

and data augmentation steps to construct an efficient posterior sampling algorithm. In Section 4, we then generalize our approach to allow for assay accuracy probabilities to be unknown. In Section 5, we provide simulation evidence to assess the performance of our estimation methods and provide a comparison of the estimates for different group testing protocols. In Section 6, we illustrate our work by using CT/NG data collected at the State Hygienic Laboratory (SHL) at the University of Iowa. In Section 7, we conclude with a summary discussion. Additional details and simulation evidence are provided in the Supporting Information.

2 Notation and preliminaries

Suppose N individuals are to be tested for $K \geq 2$ diseases using a group testing protocol. We assume all diagnostic test results are obtained from a multiplex assay which provides a positive/negative diagnosis for each disease each time it is used (on specimen pools or on individual specimens). For example, the SHL at the University of Iowa uses the Aptima Combo 2 Assay (AC2A) to detect CT and NG simultaneously; see Section 6. The current protocol at the SHL tests specimen pools (usually of size 4) with the AC2A. Pools testing positively for either disease are then resolved by testing each individual specimen with the same assay. Disease diagnoses are then determined from the individual tests.

To focus our ideas, the presentation in this article will assume there are $K = 2$ diseases (e.g., CT/NG, etc.). Generalizing our approach to $K \geq 2$ diseases is straightforward and is thus relegated to the Supporting Information. Let $\widetilde{\mathbf{Y}}_i = (\widetilde{Y}_{i1}, \widetilde{Y}_{i2})'$ denote a vector of binary random variables which encode the true disease statuses of the ith individual, with $\widetilde{Y}_{ik} = 1(0)$ denoting the individual is truly positive (negative) for the kth disease, for $i = 1, 2, ..., N$ and $k = 1, 2$. We assume the $\widetilde{\mathbf{Y}}_i$'s are mutually independent with probability mass function $\text{pr}(\widetilde{Y}_{i1} = \widetilde{y}_1, \widetilde{Y}_{i2} = \widetilde{y}_2|\mathbf{p}) = p_{00}^{(1-\widetilde{y}_1)(1-\widetilde{y}_2)} p_{10}^{\widetilde{y}_1(1-\widetilde{y}_2)} p_{01}^{(1-\widetilde{y}_1)\widetilde{y}_2} p_{11}^{\widetilde{y}_1\widetilde{y}_2}$, where $\widetilde{y}_1, \widetilde{y}_2 \in \{0, 1\}$, $\mathbf{p} = (p_{00}, p_{10}, p_{01}, p_{11})'$, and $p_{00} + p_{10} + p_{01} + p_{11} = 1$. Therefore, the joint distribution of the true disease status vectors for all N individuals; i.e., $\widetilde{\mathbf{Y}} = (\widetilde{\mathbf{Y}}_1', \widetilde{\mathbf{Y}}_2', \cdots, \widetilde{\mathbf{Y}}_N')'$, is given by

$$\pi(\widetilde{\mathbf{Y}}|\mathbf{p}) = \prod_{i=1}^{N} p_{00}^{(1-\widetilde{Y}_{i1})(1-\widetilde{Y}_{i2})} p_{10}^{\widetilde{Y}_{i1}(1-\widetilde{Y}_{i2})} p_{01}^{(1-\widetilde{Y}_{i1})\widetilde{Y}_{i2}} p_{11}^{\widetilde{Y}_{i1}\widetilde{Y}_{i2}}. \quad (1)$$

Note that estimating $\mathbf{p}$ using Equation (1) would be straightforward if individual testing were used and the multiplex assay were 100% accurate for each disease. Otherwise, the random vector $\widetilde{\mathbf{Y}}$ is best regarded as latent.

The observed data in group testing consist of diagnostic test results collected as part of a testing protocol. These protocols are typically completed over $S \geq 2$ stages, where, within each stage, pooled or individual specimens are tested in response to the results from the previous stage. For example, as noted earlier, the SHL uses an $S = 2$ stage protocol where pools of specimens are tested in the first stage and individual specimens from positive pools are tested in the second. Hou et al. (2017) evaluated the utility of hierarchical group testing protocols using a larger number of stages, showing that $S = 3$ stage protocols conferred the smallest number of tests when screening for CT/NG in four western states in the US (Alaska, Idaho, Oregon, and Washington). A three-stage hierarchical protocol uses an intermediate second stage with smaller-sized subpools; e.g., first-stage pools of size 9, three second-stage pools of size 3, individual testing in the third stage. Hou et al. (2020) later proposed $S = 2$ and $S = 3$ stage multiplex protocols which use array testing. In these (non-hierarchical) protocols, testing results arise from pooling rows and columns of overlapping specimens arranged in an array-like configuration.

In this article, we propose an estimation framework which is applicable for *any* group testing protocol using multiplex assays. To maintain this level of generality, we need notation that helps us track pool membership. Define the index set $\mathcal{P}_j \subseteq \{1, 2, ..., N\}$, $j = 1, 2, ..., J$, which identifies which individuals contribute to the jth pool; that is, $i \in \mathcal{P}_j$ when the ith individual is in the jth pool. Let $\widetilde{\mathbf{Z}}_j = (\widetilde{Z}_{j1}, \widetilde{Z}_{j2})'$ denote a vector of binary random variables encoding the true status of the jth pool,

where $\widetilde{Z}_{jk} = I(\sum_{i \in \mathcal{P}_j} \widetilde{Y}_{ik} > 0)$, for $j = 1, 2, ..., J$ and $k = 1, 2$. In other words, the jth pool is truly positive (truly negative) for the kth disease if the pool contains at least one (no) positive individual(s) for disease k. Again, due to the effects of imperfect testing, the $\widetilde{\mathbf{Z}}_j$'s are not observed. Instead, we observe $\mathbf{Z}_j = (Z_{j1}, Z_{j2})'$, a vector of binary random variables encoding the test results for the jth pool, where $Z_{jk} = 1(0)$ if the jth pool tests positively (negatively) for the kth disease.

To allow for imperfect testing, we need to relate the observed testing results in $\mathbf{Z}_j$ to the true disease statuses in $\widetilde{\mathbf{Z}}_j$. We assume $\text{pr}(Z_{jk} = 1|\widetilde{Z}_{jk} = 1) = S_{e:jk}$ and $\text{pr}(Z_{jk} = 0|\widetilde{Z}_{jk} = 0) = S_{p:jk}$, for $j = 1, 2, ..., J$ and $k = 1, 2$. That is, $S_{e:jk}$ ($S_{p:jk}$) is the sensitivity (specificity) of the multiplex assay when testing the jth pool for the kth disease. Our notation emphasizes $S_{e:jk}$ and $S_{p:jk}$ are "pool specific," affording us the flexibility to allow for different multiplex assays to be used and/or to have testing accuracy of one multiplex assay be a function of the size of the jth pool; see Section 4. The conditional distribution of the observed data $\mathbf{Z} = (\mathbf{Z}_1', \mathbf{Z}_2', \cdots, \mathbf{Z}_J')'$ given the individuals' true disease statuses $\widetilde{\mathbf{Y}}$ is given by

$$\pi(\mathbf{Z}|\widetilde{\mathbf{Y}}, \boldsymbol{\delta}) = \prod_{j=1}^{J} \prod_{k=1}^{2} S_{e:jk}^{Z_{jk}\widetilde{Z}_{jk}} (1 - S_{e:jk})^{(1-Z_{jk})\widetilde{Z}_{jk}} S_{p:jk}^{(1-Z_{jk})(1-\widetilde{Z}_{jk})} (1 - S_{p:jk})^{Z_{jk}(1-\widetilde{Z}_{jk})}, \quad (2)$$

where $\boldsymbol{\delta}$ is a vector that contains all assay accuracy probabilities; i.e., the $S_{e:jk}$'s and $S_{p:jk}$'s for $j = 1, 2, ..., J$ and $k = 1, 2$. Note that in writing Equation (2), we assume that testing results in $\mathbf{Z}$ are conditionally independent given the true statuses in $\widetilde{\mathbf{Y}}$ and the values of $S_{e:jk}$ and $S_{p:jk}$ for one disease do not depend on the true status of the other disease; see Hou et al. (2020). Combining Equations (1) and (2), we can express the distribution of the observed data from *any* group testing protocol as

$$\pi(\mathbf{Z}|\mathbf{p}, \boldsymbol{\delta}) = \sum_{\widetilde{\mathbf{Y}} \in \{0,1\}^{2N}} \pi(\widetilde{\mathbf{Y}}|\mathbf{p}) \pi(\mathbf{Z}|\widetilde{\mathbf{Y}}, \boldsymbol{\delta}), \quad (3)$$

where $\{0, 1\}^{2N}$ denotes the collection of all possible realizations of $\widetilde{\mathbf{Y}}$. This distribution is obtained by marginalizing the joint distribution of the observed testing responses and the individuals' latent statuses, that is, by summing $\pi(\mathbf{Z}, \widetilde{\mathbf{Y}}|\mathbf{p}, \boldsymbol{\delta}) = \pi(\widetilde{\mathbf{Y}}|\mathbf{p})\pi(\mathbf{Z}|\widetilde{\mathbf{Y}}, \boldsymbol{\delta})$ over $\widetilde{\mathbf{Y}}$. This marginalization process requires computing the sum over the 2^{2N} possible realizations of $\widetilde{\mathbf{Y}}$, which can be computationally prohibitive in practical settings. For example, the Iowa CT/NG data we consider in Section 6 involves $N = 14450$ individuals.

3 Estimation with known assay accuracy probabilities

To incorporate prior knowledge about the coinfection probabilities in $\mathbf{p}$ and the assay accuracy probabilities in $\boldsymbol{\delta}$, we take a Bayesian approach as in Warasi et al. (2016) who considered two-stage protocols only. In this section, we consider the simpler setting where the assay accuracy probabilities in $\boldsymbol{\delta}$ are known. This assumption is then relaxed in Section 4.

3.1 Posterior sampling

We assume *a priori* that $\mathbf{p} \sim \text{Dirichlet}(\boldsymbol{\alpha})$; i.e., the prior distribution for $\mathbf{p}$ is given by

$$\pi(\mathbf{p}) = B(\boldsymbol{\alpha}) p_{00}^{\alpha_{00}-1} p_{10}^{\alpha_{10}-1} p_{01}^{\alpha_{01}-1} p_{11}^{\alpha_{11}-1},$$

where $B(\boldsymbol{\alpha})$ is a normalizing constant and $\boldsymbol{\alpha} = (\alpha_{00}, \alpha_{10}, \alpha_{01}, \alpha_{11})'$ is a vector of known hyperparameters. Based on the observed data $\mathbf{Z}$, we then update our knowledge about $\mathbf{p}$ through its posterior distribution, given by $\pi(\mathbf{p}|\mathbf{Z}, \boldsymbol{\delta}) \propto \pi(\mathbf{Z}|\mathbf{p}, \boldsymbol{\delta})\pi(\mathbf{p})$. Unfortunately, this distribution involves $\pi(\mathbf{Z}|\mathbf{p}, \boldsymbol{\delta})$ whose calculation in Equation (3) is generally infeasible. Therefore, to facilitate posterior estimation, we develop a Markov chain Monte Carlo (MCMC) sampling algorithm that can draw realizations from $\pi(\mathbf{p}|\mathbf{Z}, \boldsymbol{\delta})$.

At the crux of this development is a data augmentation step which involves introducing individuals' true disease statuses as "missing data." Define the vector $\widetilde{\mathbf{V}}_i = (\widetilde{V}_{(00)i}, \widetilde{V}_{(10)i}, \widetilde{V}_{(01)i}, \widetilde{V}_{(11)i})'$ so that $\widetilde{V}_{(00)i} = 1$ when $\widetilde{\mathbf{Y}}_i' = (0,0)$, $\widetilde{V}_{(10)i} = 1$ when $\widetilde{\mathbf{Y}}_i' = (1,0)$, and so on. We introduce $\widetilde{\mathbf{V}}_i$ because it uniquely encodes the true disease status of the ith individual, and we can work out its full conditional distribution. Specifically, $\widetilde{\mathbf{V}}_i|\widetilde{\mathbf{Y}}_{-i}, \mathbf{p}, \mathbf{Z}, \boldsymbol{\delta} \sim$ multinomial$(p^*_{(00)i}, p^*_{(10)i}, p^*_{(01)i}, p^*_{(11)i})$, where $\widetilde{\mathbf{Y}}_{-i}$ aggregates all N true disease status vectors except $\widetilde{\mathbf{Y}}_i$. Closed-form expressions for $p^*_{(00)i}$, $p^*_{(10)i}$, $p^*_{(01)i}$, and $p^*_{(11)i}$ are given in Appendix A in the Supporting Information. In addition, from Equation (1) and the form of the prior $\pi(\mathbf{p})$, it is easy to verify the full conditional $\pi(\mathbf{p}|\widetilde{\mathbf{Y}})$ is also Dirichlet with parameter $\boldsymbol{\alpha}^* = (\alpha^*_{00}, \alpha^*_{10}, \alpha^*_{01}, \alpha^*_{11})'$, where $\alpha^*_{uv} = \alpha_{uv} + \sum_{i=1}^N \widetilde{V}_{(uv)i}$ and $\widetilde{V}_{(uv)i} = \widetilde{Y}^u_{i1}(1-\widetilde{Y}_{i1})^{1-u}\widetilde{Y}^v_{i2}(1-\widetilde{Y}_{i2})^{1-v}$, for $u, v \in \{0,1\}$. These two distributions, $\pi(\widetilde{\mathbf{V}}_i|\widetilde{\mathbf{Y}}_{-i}, \mathbf{p}, \mathbf{Z}, \boldsymbol{\delta})$ and $\pi(\mathbf{p}|\widetilde{\mathbf{Y}})$, can be used to construct an efficient algorithm to sample from $\pi(\mathbf{p}|\mathbf{Z}, \boldsymbol{\delta})$ as we now describe.

POSTERIOR SAMPLING ALGORITHM

1. Initialize $\mathbf{p}^{(0)} = (p^{(0)}_{00}, p^{(0)}_{10}, p^{(0)}_{01}, p^{(0)}_{11})'$, where $\sum_{u=0}^1 \sum_{v=0}^1 p^{(0)}_{uv} = 1$, and then simulate $\widetilde{\mathbf{Y}}^{(0)}_i = (\widetilde{Y}^{(0)}_{i1}, \widetilde{Y}^{(0)}_{i2})'$, for $i = 1, 2, ..., N$, from the population-level multinomial model at $\mathbf{p}^{(0)}$. Set $g = 1$.

2. For $i = 1, 2, ..., N$, sample $\widetilde{\mathbf{V}}^{(g)}_i = (\widetilde{V}^{(g)}_{(00)i}, \widetilde{V}^{(g)}_{(10)i}, \widetilde{V}^{(g)}_{(01)i}, \widetilde{V}^{(g)}_{(11)i})'$ from

   $$\widetilde{\mathbf{V}}_i|\widetilde{\mathbf{Y}}^{(g)}_{-i}, \mathbf{p}^{(g-1)}, \mathbf{Z}, \boldsymbol{\delta} \sim \text{multinomial}(p^*_{(00)i}, p^*_{(10)i}, p^*_{(01)i}, p^*_{(11)i}),$$

   where $\widetilde{\mathbf{Y}}^{(g)}_{-i} = (\widetilde{\mathbf{Y}}^{(g)'}_1, \cdots, \widetilde{\mathbf{Y}}^{(g)'}_{i-1}, \widetilde{\mathbf{Y}}^{(g-1)'}_{i+1}, \cdots, \widetilde{\mathbf{Y}}^{(g-1)'}_N)'$ and $\widetilde{\mathbf{Y}}^{(g)}_i = (\widetilde{V}^{(g)}_{(10)i} + \widetilde{V}^{(g)}_{(11)i}, \widetilde{V}^{(g)}_{(01)i} + \widetilde{V}^{(g)}_{(11)i})'$.

3. Sample $\mathbf{p}^{(g)}$ from $\mathbf{p}|\widetilde{\mathbf{Y}}^{(g)} \sim$ Dirichlet$(\boldsymbol{\alpha}^*)$, where $\widetilde{\mathbf{Y}}^{(g)} = (\widetilde{\mathbf{Y}}^{(g)'}_1, \widetilde{\mathbf{Y}}^{(g)'}_2, \cdots, \widetilde{\mathbf{Y}}^{(g)'}_N)'$.

4. Set $g = g + 1$ and repeat steps 2-4 while $g < G$, the number of Gibbs iterates.

Two remarks are in order. First, it is worth emphasizing the multinomial cell probabilities in Step 2 are functions of the observed data in $\mathbf{Z}$; see Appendix A in the Supporting Information. This is why the algorithm above can be implemented with *any* group testing protocol using multiplex assays, that is, different protocols will give rise to different types of observed data $\mathbf{Z}$ but the sampling procedure remains unchanged. Second, in practice, we recommend selecting the number of Gibbs iterates G to be large; e.g., $G = 10000$, after discarding the first thousand or so iterates for burn-in purposes. Note that this algorithm is extremely fast because all conditional distributions are in closed form. For inference, the sample mean of the G iterates can be used as an estimate of the posterior mean of $\mathbf{p}$; i.e., $E(\mathbf{p}|\mathbf{Z}, \boldsymbol{\delta})$, and credible intervals can be constructed by using the appropriate sample quantiles.

3.2 Maximum *a posteriori* estimation

It is well known that group testing is most beneficial when the probability of disease is low. Otherwise, most initial pools could test positively and the motivation for pooling specimens would diminish. In our multiplex setting, this means p_{00}, the probability an individual is disease free, may be close to unity, and the population-level parameters p_{10}, p_{01}, p_{11}, and marginal probabilities $p_{1+} = p_{10} + p_{11}$ and $p_{+1} = p_{01} + p_{11}$ may all be close to zero depending on the diseases under investigation. Because of these constraints on the parameter space, the marginal posterior distributions from $\pi(\mathbf{p}|\mathbf{Z}, \boldsymbol{\delta})$ may be heavily skewed and summarizing the posterior distribution with a mean (or median) estimate may be unwise.

In such instances, reporting a posterior mode may be more sensible. We therefore describe an approach to find the maximum *a posteriori* (MAP) estimate; i.e., the posterior mode of $\pi(\mathbf{p}|\mathbf{Z}, \boldsymbol{\delta})$. Using the same

missing data conceptualization in Section 3.1, we use the expectation-maximization (EM) algorithm to maximize $\pi(\mathbf{p}|\mathbf{Z},\boldsymbol{\delta})$. This algorithm involves evaluating $Q(\mathbf{p},\mathbf{p}^{(t)})$, the conditional expectation of the logarithm of the augmented posterior $\pi(\mathbf{Z},\widetilde{\mathbf{Y}}|\boldsymbol{\delta})\pi(\mathbf{p}) = \pi(\widetilde{\mathbf{Y}}|\mathbf{p})\pi(\mathbf{Z}|\widetilde{\mathbf{Y}},\boldsymbol{\delta})\pi(\mathbf{p})$ given the observed data and current parameter value $\mathbf{p}^{(t)}$, and then maximizing it as a function of $\mathbf{p}$. One then iterates between these two steps until convergence. This can be accomplished by using the steps described below.

MAP ESTIMATION VIA EM ALGORITHM

1. Initialize $\mathbf{p}^{(0)} = (p_{00}^{(0)}, p_{10}^{(0)}, p_{01}^{(0)}, p_{11}^{(0)})'$, where $\sum_{u=0}^{1}\sum_{v=0}^{1} p_{uv}^{(0)} = 1$, and then simulate $\widetilde{\mathbf{Y}}_i^{(0)} = (\widetilde{Y}_{i1}^{(0)}, \widetilde{Y}_{i2}^{(0)})'$, for $i = 1, 2, ..., N$, from the population-level multinomial model at $\mathbf{p}^{(0)}$. Set $t = 0$.
2. (E-Step): For $i = 1, 2, ..., N$,
   - sample $\widetilde{\mathbf{V}}_i^{(g)} = (\widetilde{V}_{(00)i}^{(g)}, \widetilde{V}_{(10)i}^{(g)}, \widetilde{V}_{(01)i}^{(g)}, \widetilde{V}_{(11)i}^{(g)})'$, for $g = 1, 2, ..., G$, from $\widetilde{\mathbf{V}}_i|\widetilde{\mathbf{Y}}_{-i}^{(g)}, \mathbf{p}^{(t)}, \mathbf{Z}, \boldsymbol{\delta} \sim$ multinomial$(p_{(00)i}^{*}, p_{(10)i}^{*}, p_{(01)i}^{*}, p_{(11)i}^{*})$, where G is the number of Gibbs iterates;
   - calculate the sample mean $G^{-1}\sum_{g=1}^{G}(\widetilde{V}_{(00)i}^{(g)}, \widetilde{V}_{(10)i}^{(g)}, \widetilde{V}_{(01)i}^{(g)}, \widetilde{V}_{(11)i}^{(g)})'$ as an estimate of the conditional expectation $E[(\widetilde{V}_{(00)i}, \widetilde{V}_{(10)i}, \widetilde{V}_{(01)i}, \widetilde{V}_{(11)i})'|\mathbf{Z}, \boldsymbol{\delta}; \mathbf{p}^{(t)}]$.
3. (M-Step): Calculate $\mathbf{p}^{(t+1)}$ using the solution in Appendix B in the Supporting Information; i.e., this maximizer depends on $E[(\widetilde{V}_{(00)i}, \widetilde{V}_{(10)i}, \widetilde{V}_{(01)i}, \widetilde{V}_{(11)i})'|\mathbf{Z}, \boldsymbol{\delta}; \mathbf{p}^{(t)}]$ and exists in closed form.
4. Set $t = t + 1$, and repeat steps 2-4 until the maximum absolute difference in $\mathbf{p}^{(t+1)} - \mathbf{p}^{(t)}$ is less than ϵ, where ϵ is small.

We again make brief remarks. First, because Step 2 uses a Gibbs sampler to estimate the conditional expectation $E[(\widetilde{V}_{(00)i}, \widetilde{V}_{(10)i}, \widetilde{V}_{(01)i}, \widetilde{V}_{(11)i})'|\mathbf{Z}, \boldsymbol{\delta}; \mathbf{p}^{(t)}]$, calculating the MAP estimate of $\mathbf{p}$ takes longer than simply summarizing $\pi(\mathbf{p}|\mathbf{Z},\boldsymbol{\delta})$ with the posterior mean (or median) from Section 3.1. However, because the M-Step solution exists in closed form, this additional time required is generally not prohibitive. Second, when a uniform prior distribution $\pi(\mathbf{p})$ is used; i.e., setting $\alpha_{00} = \alpha_{10} = \alpha_{01} = \alpha_{11} = 1$, the MAP estimate of $\mathbf{p}$ coincides with the maximum likelihood estimate (MLE) of $\mathbf{p}$, a potential preference for users wanting to report frequentist-based estimates. Finally, when the assay accuracy probabilities in $\boldsymbol{\delta}$ are unknown, our simulation results in Section 5 demonstrate that MAP estimates of $\boldsymbol{\delta}$ can be more accurate than other posterior estimates of $\boldsymbol{\delta}$. We now generalize our methodology to allow for this situation.

4 Estimation with unknown assay accuracy probabilities

Our goal now is to estimate the population-level coinfection probabilities in $\mathbf{p}$ and the assay accuracy probabilities in $\boldsymbol{\delta}$ jointly. As we demonstrate, this can be accomplished by taking our algorithms in Section 3 and adding appropriate steps for the conditional distribution and MAP solution of $\boldsymbol{\delta}$. Such an extension is practically useful in incorporating the uncertainty in $\boldsymbol{\delta}$. For example, although manufacturers will typically report values of sensitivity and specificity for multiplex assays (for each disease) in their product literature, these values are usually obtained from small pilot studies involving specimens whose true disease statuses are known in advance. The practice of ostensibly regarding these values as "correct" can lead to two potential problems. First, doing so ignores the sampling error incurred from having to estimate these values in small feasibility experiments. Second, the population under investigation (e.g., high-risk females in Iowa, etc.) may differ substantially from the one which was used to validate the multiplex assay initially.

Extending the approach in McMahan et al. (2017) for single diseases, let $S_{e:(l)k}$ and $S_{p:(l)k}$ denote the sensitivity and specificity of the lth assay for the kth disease, for $k = 1, 2$ and $l = 1, 2, ..., L$, and let $\mathcal{M}(l) = \{j : \text{the } l\text{th assay tests pool } j\}$ denote the index set of the specimens tested by the lth assay, for $j = 1, 2, ..., J$. Our use of the set $\mathcal{M}(l)$ simply allows us to reparameterize the exposition in Section 2. For

example, at the SHL in Iowa, the AC2A assay is used for all specimens tested in pools and individually. If this assay performs the same when testing pools and individuals, then $L = 1$ and the parameter vector $\boldsymbol{\delta} = (S_{e:(1)1}, S_{e:(1)2}, S_{p:(1)1}, S_{p:(1)2})'$. On the other hand, if the performance of the AC2A depends on whether pools or individuals are tested, one could envision one set of assay accuracy probabilities for pools ($l = 1$) and a separate set for individuals ($l = 2$). This situation would correspond to $L = 2$ and the parameter vector would become $\boldsymbol{\delta} = (S_{e:(1)1}, S_{e:(1)2}, S_{p:(1)1}, S_{p:(1)2}, S_{e:(2)1}, S_{e:(2)2}, S_{p:(2)1}, S_{p:(2)2})'$.

Under our reparameterization, the distribution of the observed data $\mathbf{Z}$ from *any* group testing protocol in Equation (3) can be written as

$$\pi(\mathbf{Z}|\mathbf{p},\boldsymbol{\delta}) = \sum_{\widetilde{\mathbf{Y}}\in\{0,1\}^{2N}} \Bigg[\pi(\widetilde{\mathbf{Y}}|\mathbf{p}) \prod_{l=1}^{L} \prod_{j\in\mathcal{M}(l)} S_{e:(l)k}^{Z_{jk}\widetilde{Z}_{jk}} (1 - S_{e:(l)k})^{(1-Z_{jk})\widetilde{Z}_{jk}} \times S_{p:(l)k}^{(1-Z_{jk})(1-\widetilde{Z}_{jk})} (1 - S_{p:(l)k})^{Z_{jk}(1-\widetilde{Z}_{jk})} \Bigg],$$

where now both $\mathbf{p}$ and the assay accuracy probabilities in $\boldsymbol{\delta}$ are regarded as unknown. To incorporate the uncertainty in $\boldsymbol{\delta}$, we use beta prior distributions for each sensitivity and specificity parameter, that is, $S_{e:(l)k} \sim \text{beta}(a_{lk}, b_{lk})$ and $S_{p:(l)k} \sim \text{beta}(c_{lk}, d_{lk})$, for $k = 1, 2$ and $l = 1, 2, ..., L$. If all prior distributions are independently specified, then the posterior distribution of $\mathbf{p}$ and $\boldsymbol{\delta}$ satisfies $\pi(\mathbf{p},\boldsymbol{\delta}|\mathbf{Z}) \propto \pi(\mathbf{Z}|\mathbf{p},\boldsymbol{\delta})\pi(\mathbf{p}) \prod_{l=1}^{L} \pi(S_{e:(l)k})\pi(S_{p:(l)k})$, where $\pi(\mathbf{Z}|\mathbf{p},\boldsymbol{\delta})$ is given above and $\pi(S_{e:(l)k})$ and $\pi(S_{p:(l)k})$ denote the beta priors. As noted earlier, data from multiplex assay feasibility studies can be used to elicit informative prior distributions for $S_{e:(l)k}$ and $S_{p:(l)k}$. Of course, in the absence of any prior knowledge, uniform priors can also be used.

Both the posterior sampling and EM algorithms in Section 3 can be generalized to estimate $\mathbf{p}$ and $\boldsymbol{\delta}$ simultaneously. To sample from $\pi(\mathbf{p},\boldsymbol{\delta}|\mathbf{Z})$, we note that $S_{e:(l)k}|\mathbf{Z},\widetilde{\mathbf{Y}} \sim \text{beta}(a^*_{lk}, b^*_{lk})$ and $S_{p:(l)k}|\mathbf{Z},\widetilde{\mathbf{Y}} \sim \text{beta}(c^*_{lk}, d^*_{lk})$, where $a^*_{lk} = a_{lk} + \sum_{j\in\mathcal{M}(l)} Z_{jk}\widetilde{Z}_{jk}$, $b^*_{lk} = b_{lk} + \sum_{j\in\mathcal{M}(l)}(1 - Z_{jk})\widetilde{Z}_{jk}$, $c^*_{lk} = c_{lk} + \sum_{j\in\mathcal{M}(l)}(1 - Z_{jk})(1 - \widetilde{Z}_{jk})$, and $d^*_{lk} = d_{lk} + \sum_{j\in\mathcal{M}(l)} Z_{jk}(1 - \widetilde{Z}_{jk})$. Therefore, because all other conditionals remain unchanged, one can take the posterior sampling algorithm described in Section 3.1 and add one additional step. Similarly, to calculate the MAP estimate of $\mathbf{p}$ and $\boldsymbol{\delta}$, the EM algorithm in Section 3.2 can be easily amended. The conditional expectation of the logarithm of the augmented posterior given the observed data and current parameter value, now written $Q(\mathbf{p},\boldsymbol{\delta},\mathbf{p}^{(t)},\boldsymbol{\delta}^{(t)})$, also has a closed-form solution in the M-step. The complete algorithms are given in Appendix C in the Supporting Information.

5 Simulation evidence

We performed a comprehensive simulation study to evaluate the performance of our estimation methods. This study included examining three hierarchical group testing protocols (H2, H3, and H4) from Hou et al. (2017) and one array testing protocol (AT) from Hou et al. (2020). We now briefly describe these protocols.

5.1 Multiplex protocols and simulation description

A hierarchical group testing protocol is carried out by first testing a master pool of individual specimens. If this pool tests negatively, then each individual in the pool is declared to be negative. If this pool tests positively, then the master pool is divided into non-overlapping subpools of specimens. Two-stage protocols (H2) revert to individual testing in the second stage, while higher-stage protocols use smaller sized subpools during intermediate stages of testing before individual testing is used in the final stage. In AT, individual specimens are arranged in a square array configuration forming row and column master pools

which are tested in the first stage. Individuals in positive row/column intersections are tested in the second stage along with other individuals whose statuses are potentially unknown because of testing errors; see Hou et al. (2020). The overarching message from Hou et al. (2017) and Hou et al. (2020) is that higher-stage hierarchical protocols (H3, H4) and AT can substantially reduce the number of tests needed when compared to H2, especially when the probability of at least one disease $1 - p_{00}$ is small.

This prompts an obvious question. When compared to H2, how do H3, H4, and AT perform in terms of estimation? One might hypothesize that because H3, H4, and AT generally require fewer tests, fewer observations would be available and thus the estimation performance for these algorithms might be degraded. On the other hand, it could be that H3 and H4 implement more tests "where it counts," that is, on individuals who are more likely to be positive, and AT uses master pools (rows and columns) that consist of overlapping individuals. In the presence of testing errors, more replicate tests on potentially positive individuals may actually improve estimation−despite H3, H4, and AT requiring fewer tests overall.

[Table 1 here.]

We simulated the execution of each protocol (H2, H3, H4, and AT) using two configurations of the disease probabilities, $\mathbf{p} = (0.95, 0.02, 0.02, 0.01)'$ and $(0.990, 0.004, 0.004, 0.002)'$; we henceforth call these Configurations I and II, respectively. The first configuration was chosen to represent the overall prevalence of CT/NG, while the second configuration allows for two rarer diseases, each with a marginal probability of $0.004 + 0.002 = 0.006$. For each of H2, H3, H4, and AT, Table 1 lists the specific protocol which minimizes the expected number of tests when $S_{e:(1)k} = 0.95$ and $S_{p:(1)k} = 0.99$, for $k = 1, 2$. For example, the entry "$5 : 1$" for H2 under Configuration I means that master pools of size 5 (and individual testing in the second stage) reduces the number of tests as much as possible on average among all two-stage hierarchical protocols. Similarly, the entry "$9 : 3 : 1$" for H3 means that master pools of size 9 are used in the first stage, three subpools of size 3 are used in the second stage, and individual testing is used in the third. We determine these pool sizes using the optimization methods described in Hou et al. (2017) and Hou et al. (2020).

For each protocol and disease probability configuration, we simulated the true disease statuses of $N = 5000$ individuals and randomly assigned these individuals to appropriately sized master pools. We then simulated the testing outcomes on pools and individuals (allowing for potential testing errors) to produce data that would be available for estimation purposes. This entire process was repeated $B = 500$ times, providing us with 500 independent data sets for each protocol under Configurations I and II. Note that in some cases smaller-sized master pools were formed when there were remainder individuals. For example, 555 master pools of size 9 were formed for the "$9 : 3 : 1$" H3 protocol listed in Table 1; the remaining 5 individuals were tested in a master pool of size 5 and resolved using H2. This practice of using H2 for remainder pools was applied uniformly in all cases to ensure a fair comparison among the protocols.

5.2 Simulation results

Tables 2 and 3 show the estimation results for both disease probability configurations when the assay accuracy probabilities in $\boldsymbol{\delta} = (S_{e:(1)1}, S_{e:(1)2}, S_{p:(1)1}, S_{p:(1)2})'$ are unknown. Table 2 depicts the results for estimating $\mathbf{p}$ while Table 3 for estimating $\boldsymbol{\delta}$. When $\boldsymbol{\delta}$ is known (the scenario in Section 3), we provide a table of estimation results in Appendix D in the Supporting Information. The simulation which produced these results used flat priors for both $\mathbf{p}$ and $\boldsymbol{\delta}$, that is, $\mathbf{p} \sim$ Dirichlet$(\mathbf{1}_4)$, $S_{e:(1)k} \sim$ beta$(1, 1)$, and $S_{p:(1)k} \sim$ beta$(1, 1)$, for $k = 1, 2$. We selected these noninformative priors for two reasons. First, these distributions give us the most challenging case for estimation because no (useful) prior information is injected into the model. Second, our use of a flat prior for $\mathbf{p}$ produces MAP estimates which should largely coincide with the MLE of $\mathbf{p}$. When the Dirichlet$(\mathbf{1}_4)$ distribution is specified *a priori*, MAP and maximum likelihood estimates of $\mathbf{p}$ are identical for the scenario of known $\boldsymbol{\delta}$ and should be approximately equal with unknown $\boldsymbol{\delta}$. It is worth noting that eliciting informative priors on $\mathbf{p}$ and $\boldsymbol{\delta}$ is straightforward in our model

framework. The *power prior* approach in Ibrahim et al. (2015) may be a useful technique for this purpose. To view an application of power prior elicitation, refer to Warasi et al. (2016).

[Table 2 here.]

In Tables 2-3, we present the sample mean ("Est") of the posterior mean (Mean) and MAP estimates calculated from $B = 500$ independent data sets. Also, presented are the sample standard deviation (SD) of the 500 posterior means and the averaged estimated posterior standard deviation (SE) for each parameter. We also recorded the number of tests needed for each protocol (H2, H3, H4, and AT) to classify each of the $N = 5000$ individuals as being positive/negative for each disease. The initial value $\mathbf{p}^{(0)}$ may be specified from historical, pilot study, or any reasonable contextual estimate of $\mathbf{p}$. The mean estimate for each parameter is calculated from 2000 posterior realizations, which are obtained by sampling $G = 10000$ posterior draws and then keeping every 5th. Note that 2000 initial posterior draws have been discarded as a burn-in period before sampling the 10000. For MAP estimation, $G = 10000$ Gibbs samples drawn from the multinomial$(p^*_{(00)i}, p^*_{(10)i}, p^*_{(01)i}, p^*_{(11)i})$ have been used without thinning to approximate $E[(\widetilde{V}_{(00)i}, \widetilde{V}_{(10)i}, \widetilde{V}_{(01)i}, \widetilde{V}_{(11)i})'|\mathbf{Z}, \boldsymbol{\delta}; \mathbf{p}^{(t)}]$ in the E-Step. Again, 2000 initial draws have been discarded. For more information on model fitting, see the R code provided with this article.

[Table 3 here.]

Table 2 shows that both mean and MAP estimates of $\mathbf{p}$ are accurate for each configuration of disease probabilities. The posterior standard deviation SE is also estimated well (SE and SD match closely) and generally unaffected by protocols H2, H3, H4 and AT. However, when estimation for $S_{e:(1)k}$ and $S_{p:(1)k}$ is of interest, we observe many interesting results. As we move from H2 to H3, H4, or AT, the accuracy of the estimates of $S_{e:(1)k}$ and $S_{p:(1)k}$ in Table 3 improves noticeably and the posterior distributions become more compact (smaller SE). This is especially evident when the disease probabilities are smaller (Configuration II). That is, in low-prevalence disease scenarios, our estimation techniques with an advanced protocol (H3, H4, or AT) can provide much better accuracy and precision in the estimates of $S_{e:(1)k}$ and $S_{p:(1)k}$. The simulation also reveals that better accuracy can be achieved from MAP estimation. This occurs when the posterior distribution is skewed so the posterior mode rather than mean (or median) better represents the parameters of interest. To assess our estimation techniques in a more challenging situation, we replicate the simulation with $L = 2$ assays in the Supporting Information (Appendix D). We again find that the estimation accuracy is satisfactory and the estimation precision is better with H3, H4, and AT.

The benefits that can be realized from a more advanced protocol would be more obvious when the expended tests are also taken into consideration. For example, when switching from H2 to H3, the reduction in the number of tests for Configuration I is 14%, which for Configuration II is 36%. Despite such reduction in the observed data (test responses), the accuracy and precision of the estimates as discussed above are not compromised but rather improved.

6 Chlamydia and gonorrhea data application

CT and NG are the two most common sexually transmitted infections in the US (CDC, 2022). If left untreated, individuals can develop serious health related complications, including pelvic inflammatory disease, infertility, and ectopic pregnancies. Public health agencies across the US perform surveillance activities for these infections. For individuals in Iowa, the SHL at the University of Iowa performs thousands of tests using H2 protocol and the AC2A. Using those observed test results, our goal here is to apply the proposed estimation techniques to better understand the prevalence of infection.

Our data consist of the CT/NG test results of 14450 females tested over a one-year period of time. Urine and cervical swab specimens were collected from these individuals at different locations in Iowa and transported into the laboratory SHL for testing purposes. Of the 14450 females, 4402 contributed

urine specimens while 10048 contributed cervical swabs, which were then prepared for testing using the H2 protocol and the usual individual testing for a smaller portion of the samples. The AC2A was used to diagnose all specimens, whether pooled or individual. The H2 protocol was used for 9618 specimens yielding 2408 pooled test responses in stage 1 (2395 pools of size 4, 12 pools of size 3, and 1 pool of size 1) and 4832 individual retest responses in stage 2. Individual testing was used for the remaining 4832. Using the tests, all individuals were completely classified as either positive or negative for the diseases. Note that the individuals were cross-classified based on their specimen type (urine or swab) before performing the tests. Thus, the entire procedure results in two data sets of pooled/individual test responses as well as CT/NG classification results for each individual. We provide a summary of the classification results in Appendix E of the Supporting Information.

Our estimation techniques, unlike those available in the literature, can jointly model the test responses observed from the H2 protocol and individual testing. We, however, use the classification results and generate test responses so a fair comparison of H2 can be made with more advanced protocols. Particularly, we treat the classification results as individual true statuses, assign them into the initial pools, and simulate pooled/individual test responses using H2, H3, and AT as described in Section 5. We do not use H4 or other higher-stage protocols because the prevalence of CT/NG in the Iowa data is too large. The pool sizes that we use are shown in Table 4. For the prevalence level of CT/NG, these pool sizes overall minimize the expected number of tests; see Hou et al. (2017) and Hou et al. (2020). We use accuracy values, $S_{e:(1)k}$ and $S_{e:(1)k}$ for $k = 1, 2$, of the AC2A for simulating all test responses; the accuracy values are reported in the package inserts of the AC2A at `www.hologic.com`; see Table 4. To average out the Monte Carlo simulation error, $B = 500$ data sets are generated by repeating these steps for each specimen stratum.

The posterior sampling algorithm and the EM algorithm in Section 4 are used for estimation assuming that $\boldsymbol{\delta} = (S_{e:(1)1}, S_{e:(1)2}, S_{p:(1)1}, S_{p:(1)2})'$ is unknown. We find the mean and MAP estimates of the coinfection probabilities in $\mathbf{p} = (p_{00}, p_{10}, p_{01}, p_{11})'$, where

$$
\begin{aligned}
p_{00} &= \text{proportion of individuals negative for both CT and NG;} \\
p_{10} &= \text{proportion of individuals positive for CT but negative for NG;} \\
p_{01} &= \text{proportion of individuals negative for CT but positive for NG;} \\
p_{11} &= \text{proportion of individuals positive for both CT and NG.}
\end{aligned}
$$

Then, estimates of the marginal probabilities $p_{1+} = p_{10} + p_{11}$ and $p_{+1} = p_{01} + p_{11}$ or other parameters can be easily calculated. We also estimate the assay accuracy probabilities in $\boldsymbol{\delta}$. Again, we use the flat priors $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4)$, $S_{e:(1)k} \sim \text{beta}(1, 1)$, and $S_{p:(1)k} \sim \text{beta}(1, 1)$, for $k = 1, 2$. In both algorithms, we use $G = 10000$ Gibbs iterates after removing the initial 2000 iterates as described in Section 5. For one data set, the posterior sampling algorithm is completed in 46 (89) seconds for urine (swab) strata, while the EM algorithm is completed in 94 (439) seconds, on a computer that has an Intel 3.6 GHz processor and 32 GB of memory. The mean and MAP estimates along with the estimated posterior standard deviation (SE) calculated from the $B = 500$ data sets are presented in Tables 5-6.

[Table 6 here.]

The overall findings in the simulation study (Section 5) are revealed again with the CT/NG data. The mean/MAP estimates of the parameters in $\mathbf{p}$ and $\boldsymbol{\delta}$ remain nearly the same across the protocols. When moving from H2 to H3 or AT, the estimator efficiency improves (i.e., SE decreases) slightly for estimating $\mathbf{p}$ but improves considerably for estimating $\boldsymbol{\delta}$, which is realized with a large amount of cost savings. For example, the H3 protocol requires 6.3% fewer tests in the urine stratum and 6.9% fewer tests in the swab stratum; i.e., H3 uses approximately 560 fewer tests in total but still produces equivalent or better estimates. Therefore, the SHL may benefit if its current two-stage hierarchical protocol is replaced by three-stage hierarchical or array testing. Because CT/NG screening is conducted in other states as well, the practical benefits by adopting such advanced group testing protocols would be substantial.

7 Discussion

We have developed general estimation techniques for group testing data with two or more diseases. This work generalizes the two-stage Bayesian approach presented in Warasi et al. (2016) and the two-stage maximum likelihood approach in Tebbs, McMahan, and Bilder (2013). Overall, the estimates produced by our work are accurate. We have demonstrated in Sections 5-6 that a higher-stage hierarchical or array protocol can provide equivalent or better precision in prevalence estimation but much better precision in estimating the assay accuracy probabilities, all of which can be achieved with a substantially smaller testing budget. Together our work, which focuses on estimation, and the optimality work in Hou et al. (2017) and Hou et al. (2020), which focuses on case identification, can serve as a useful approach for cost-effective surveillance of infectious diseases. We present an R function in Appendix F of the Supporting Information, which can be used to implement the posterior sampling algorithm and the EM algorithm. The code with documentation and examples is provided with the article and also made available at `https://github.com/mswarasi/General-MultiplexBayes`.

The best feature of our model structure is its flexibility and generality. The model can accommodate data observed using any multiplex group testing protocols, including those currently available in the literature. Another flexibility is that the assay accuracy probabilities, $S_{e:(l)k}$ and $S_{p:(l)k}$, are defined in the model as test specific. In other words, L different multiplex assays can be used for testing J pools (or individuals), where $L \leq J$. This flexibility would be especially useful in scenarios where separate screening and confirmatory assays are used. However, caution needs to be exercised in this case. When L is large (e.g., 3 or more assays), the model can be weakly identifiable or unidentifiable so informative priors would be necessary for reliable estimation. For ease of exposition, we have presented the estimation techniques with two diseases. Extending the techniques to the general case of two or more diseases, $K \geq 2$, is straightforward; refer to Appendix G in the Supporting Information.

Acknowledgements This research was funded by Grant R01 AI121351 from the National Institutes of Health.

Conflict of Interest

The authors have declared no conflict of interest.

Data Availability Statement

Simulated data that support the findings of this study can be generated with the code provided as Supporting Information. The real data used in this manuscript (Section 6) are summarized in Appendix E of the Supporting Information.

References

Abdalhamid, B., Bilder, C., McCutchen, E., Hinrichs, S., Koepsell, S., and Iwen, P. (2020). Assessment of specimen pooling to conserve SARS CoV-2 testing resources. *American Journal of Clinical Pathology* **153**, 715–718.

Bilder, C., Tebbs, J., and McMahan, C. (2021). Discussion on "Is group testing ready for prime-time in disease identification". *Statistics in Medicine* **40**, 3881–3886.

Busch, M., Caglioti, S., Robertson, E., McAuley, J., Tobler, L., Kamel, H., Linnen, J., Venkatakrishna, S., Tomasulo, P., and Kleinman, S. (2005). Screening the blood supply for West Nile virus RNA by nucleic acid amplification testing. *New England Journal of Medicine* **353**, 460–467.

Centers for Disease Control and Prevention (CDC). Sexually transmitted disease surveillance 2020. Atlanta, GA: Department of Health and Human Services. https://www.cdc.gov/std/statistics/2020/default.htm. Retrieved September 29, 2022.

Delaigle, A. and Meister, A. (2011). Nonparametric regression analysis for group testing data. *Journal of the American Statistical Association* **106**, 640–650.

de Salazar, A., Espadafor, B., Fuentes-Lopez, A., Barrientos-Duran, A., Salvador, L., Alvarez, M., and Garcia, F. (2019). Comparison between Aptima Assays (Hologic) and the Allplex STI Essential Assay (Seegene) for the diagnosis of sexually transmitted infections. *PLoS One* **14**, e022439.

Dorfman, R. (1943). The detection of defective members of large populations. *Annals of Mathematical Statistics* **14**, 436–440.

Ding, J. and Xiong, W. (2015). Robust group testing for multiple traits with misclassification. *Journal of Applied Statistics* **42**, 2115–2125.

Hou, P., Tebbs, J., Bilder, C., and McMahan, C. (2017). Hierarchical group testing for multiple infections. *Biometrics* **73**, 656–665.

Hou, P., Tebbs, J., Wang, D., McMahan, C., and Bilder, C. (2020). Array testing for multiplex assays. *Biostatistics* **21**, 417–431.

Hourfar, M., Jork, C., Schottstedt, V., Weber-Schehl, M., Brixner, V., Busch, M., Geusendam, G., Gubbe, K., Mahnhardt, C., Mayr-Wohlfar, W., Pichl, L., Roth, W., Schmidt, M., Seifried, E., and Wright, D. (2008). Experience of German Red Cross blood donor services with nucleic acid testing: Results of screening more than 30 million blood donations for human immunodeficiency virus, hepatitis C virus, and hepatitis B virus. *Transfusion* **48**, 1558–1566.

Huang, S., Huang, M., Shedden, K., and Wong, W. (2017). Optimal group testing designs for estimating prevalence with uncertain testing errors. *Journal of the Royal Statistical Society: Series B* **79**, 1547–1563.

Hughes-Oliver, J. and Rosenberger, W. (2000). Efficient estimation of the prevalence of multiple rare traits. *Biometrika* **87**, 315–327.

Ibrahim, J., Chen, M., Gwon, Y., Chen, F. (2015). The power prior: theory and applications. *Statistics in Medicine* **34**, 3724–3749.

Kim, H., Hudgens, M., Dreyfuss, J., Westreich, D., and Pilcher, C. (2007). Comparison of group testing algorithms for case identification in the presence of testing error. *Biometrics* **63**, 1152–1163.

Lindan, C., Mathur, M., Kumta, S., Jerajani, H., Gogate, A., Schachter, J., and Moncada, J. (2005). 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 Virology* **43**, 1674–1677.

Li, Q., Liu, A., and Xiong, W. (2017) D-optimality of group testing for joint estimation of correlated rare diseases with misclassification. *Statistica Sinica* **27**, 823–838.

Liu, A., Liu, C., Zhang, Z., and Albert, P. (2012). Optimality of group testing in the presence of misclassification. *Biometrika* **99**, 245–251.

Martin, E., Salaru, G., Mohammed, D., Coombs, R., Paul, S., and Cadoff, E. (2013). Finding those at risk: Acute HIV infection in Newark, NJ. *Journal of Clinical Virology* **58**, e24–e28.

McMahan, C., Tebbs, J., Hanson, T., and Bilder, C. (2017). Bayesian regression for group testing data. *Biometrics* **73**, 1443–1452.

Pilcher, C., Fiscus, S., Nguyen, T., Foust, E., Wolf, L., Williams, D., Ashby, R., O'Dowd, J., McPherson, J., Stalzer, B., Hightow, L., Miller, W., Eron, J., Cohen, M., and Leone, P. (2005). Detection of acute infections during HIV testing in North Carolina. *New England Journal of Medicine* **352**, 1873–1883.

Quinn, T., Brookmeyer, R., Kline, R., Shepherd, M., Paranjape, R., Mehendale, S., Gadkari, D., and Bollinger, R. (2000). Feasibility of pooling sera for HIV-1 viral RNA to diagnose acute primary HIV-1 infection and estimate HIV incidence. *AIDS* **14**, 2751–2757.

Roche (2020). Roche receives FDA Emergency Use Authorization for the cobas SARS-CoV-2 & Influenza A/B Test for use on the cobas 6800/8800 Systems. https://www.roche.com/media/releases/med-cor-2020-09-04. Retrieved October 9, 2020.

Stramer, S., Krysztof, D., Brodsky, J., Fickett, T., Reynolds, B., Dodd, R., and Kleinman, S. (2013). Comparative analysis of triplex nucleic acid test assays in United States blood donors. *Transfusion* **53**, 2525–2537.

Stramer, S., Notari, E., Krysztof, D., and Dodd, R. (2013). Hepatitis B virus testing by minipool nucleic acid testing: Does it improve blood safety? *Transfusion* **53**, 2449–2458.

Tebbs, J., McMahan, C., and Bilder C. (2013). Two-stage hierarchical group testing for multiple infections with application to the Infertility Prevention Project. *Biometrics* **69**, 1064–1073.

Wang, D., McMahan, C., Gallagher, C., and Kulasekera, K. (2014). Semiparametric group testing regression models. *Biometrika* **101**, 587–598.

Warasi, M., Tebbs, J., McMahan C., and Bilder, C. (2016). Estimating the prevalence of multiple diseases from two-stage hierarchical pooling. *Statistics in Medicine* **35**, 3851–3864.

Westreich, D., Hudgens, M., Fiscus, S., and Pilcher, C. (2008). Optimizing screening for acute human immunodeficiency virus infection with pooled nucleic acid amplification tests. *Journal of Clinical Microbiology* **46**, 1785–1792.

Table 1 Multiplex group testing protocols. Hierarchical protocols H2, H3, and H4 use two, three, and four stages, respectively (Hou et al., 2017). The array testing (AT) protocol uses square arrays (Hou et al., 2020). For each protocol, the configurations listed below minimize the expected number of tests per individual specimen. "Configuration I" uses $\mathbf{p} = (0.95, 0.02, 0.02, 0.01)'$ and "Configuration II" uses $\mathbf{p} = (0.990, 0.004, 0.004, 0.002)'$.

Configuration I		Configuration II	
Protocol	Pool sizes	Protocol	Pool sizes
H2	$5:1$	H2	$11:1$
H3	$9:3:1$	H3	$25:5:1$
H4	$18:6:3:1$	H4	$48:12:4:1$
AT	11×11	AT	29×29

Table 2 Simulation evidence for the characteristics of posterior mean (Mean) and MAP estimates of $\mathbf{p} = (p_{00}, p_{10}, p_{01}, p_{11})'$ when assay accuracy probabilities are **unknown**. Estimates ("Est") are averages over $B = 500$ Monte Carlo data sets, "SD" is the sample standard deviation of the 500 estimates, and "SE" is the estimated posterior standard deviation. Flat priors have been used for all parameters; i.e., $\mathbf{p} \sim$ Dirichlet$(\mathbf{1}_4)$, $S_{e:(1)k} \sim$ beta$(1,1)$, and $S_{p:(1)k} \sim$ beta$(1,1)$, for $k = 1, 2$. The average number of tests (over $B = 500$ data sets) is also shown; the percent reduction (in parentheses) is compared to the two-stage hierarchical protocol H2 in Tebbs, McMahan, and Bilder (2013) and Warasi et al. (2016). The specific protocols used (H2, H3, H4, and AT) are identified in Table 1.

			H2			H3			H4			AT		
			Est	SD	SE	Est	SD	SE	Est	SD	SE	Est	SD	SE
Configuration I	$p_{00} = 0.95$ $p_{10} = 0.02$ $p_{01} = 0.02$ $p_{11} = 0.01$	Mean	0.949	0.0035	0.0039	0.949	0.0033	0.0034	0.949	0.0034	0.0033	0.950	0.0034	0.0034
			0.021	0.0024	0.0026	0.020	0.0022	0.0022	0.020	0.0021	0.0022	0.020	0.0023	0.0022
			0.021	0.0025	0.0026	0.020	0.0021	0.0022	0.020	0.0023	0.0022	0.020	0.0022	0.0022
			0.010	0.0014	0.0015	0.010	0.0015	0.0015	0.010	0.0014	0.0015	0.010	0.0015	0.0015
		MAP	0.950	0.0035	0.0039	0.950	0.0032	0.0034	0.950	0.0034	0.0033	0.950	0.0033	0.0034
			0.020	0.0024	0.0026	0.020	0.0022	0.0022	0.020	0.0021	0.0022	0.020	0.0022	0.0022
			0.020	0.0024	0.0026	0.020	0.0021	0.0022	0.020	0.0022	0.0022	0.020	0.0022	0.0022
			0.010	0.0014	0.0015	0.010	0.0015	0.0015	0.010	0.0014	0.0015	0.010	0.0015	0.0015
	Avg. number of tests			2164.3			1859.0 (14.1%)			1858.3 (14.1%)			1728.0 (20.2%)	
Configuration II	$p_{00} = 0.990$ $p_{10} = 0.004$ $p_{01} = 0.004$ $p_{11} = 0.002$	Mean	0.988	0.0018	0.0021	0.989	0.0016	0.0016	0.989	0.0015	0.0016	0.989	0.0014	0.0016
			0.005	0.0012	0.0013	0.004	0.0010	0.0011	0.004	0.0010	0.0010	0.004	0.0009	0.0010
			0.005	0.0012	0.0013	0.004	0.0010	0.0011	0.004	0.0010	0.0010	0.004	0.0009	0.0010
			0.003	0.0007	0.0008	0.002	0.0006	0.0007	0.002	0.0006	0.0007	0.002	0.0006	0.0007
		MAP	0.990	0.0018	0.0021	0.990	0.0016	0.0016	0.990	0.0015	0.0016	0.990	0.0015	0.0016
			0.004	0.0011	0.0013	0.004	0.0010	0.0011	0.004	0.0010	0.0010	0.004	0.0010	0.0010
			0.004	0.0012	0.0013	0.004	0.0010	0.0011	0.004	0.0010	0.0010	0.004	0.0009	0.0010
			0.002	0.0007	0.0008	0.002	0.0006	0.0007	0.002	0.0006	0.0007	0.002	0.0006	0.0007
	Avg. number of tests			1054.3			671.0 (36.4%)			581.4 (44.9%)			751.4 (28.7%)	

Table 3 Simulation evidence for the characteristics of posterior mean (Mean) and MAP estimates of $\boldsymbol{\delta} = (S_{e:(1)1}, S_{e:(1)2}, S_{p:(1)1}, S_{p:(1)2})'$ for two configurations of $\mathbf{p}$. Estimates ("Est") are averages over $B = 500$ Monte Carlo data sets, "SD" is the sample standard deviation of the 500 estimates, and "SE" is the estimated posterior standard deviation. Flat priors have been used for all parameters; i.e., $\mathbf{p} \sim$ Dirichlet$(\mathbf{1}_4)$, $S_{e:(1)k} \sim$ beta$(1,1)$, and $S_{p:(1)k} \sim$ beta$(1,1)$, for $k = 1, 2$. "Configuration I" uses $\mathbf{p} = (0.95, 0.02, 0.02, 0.01)'$ and "Configuration II" uses $\mathbf{p} = (0.990, 0.004, 0.004, 0.002)'$. The specific protocols used (H2, H3, H4, and AT) are identified in Table 1.

			H2			H3			H4			AT		
			Est	SD	SE	Est	SD	SE	Est	SD	S3	Est	SD	SE
Configuration I	$S_{e:(1)1} = 0.95$ $S_{e:(1)2} = 0.95$ $S_{p:(1)1} = 0.99$ $S_{p:(1)2} = 0.99$	Mean	0.941	0.023	0.027	0.946	0.015	0.015	0.947	0.012	0.012	0.947	0.014	0.015
			0.942	0.025	0.027	0.946	0.014	0.015	0.948	0.011	0.012	0.948	0.015	0.015
			0.990	0.004	0.004	0.989	0.003	0.004	0.989	0.003	0.004	0.989	0.004	0.005
			0.990	0.004	0.004	0.989	0.003	0.004	0.990	0.004	0.004	0.988	0.005	0.005
		MAP	0.950	0.023	0.027	0.950	0.014	0.015	0.950	0.012	0.012	0.950	0.014	0.015
			0.951	0.025	0.027	0.950	0.014	0.015	0.950	0.011	0.012	0.951	0.014	0.015
			0.990	0.004	0.004	0.990	0.003	0.004	0.990	0.003	0.004	0.990	0.004	0.005
			0.990	0.004	0.004	0.990	0.003	0.004	0.990	0.003	0.004	0.990	0.004	0.005
Configuration II	$S_{e:(1)1} = 0.95$ $S_{e:(1)2} = 0.95$ $S_{p:(1)1} = 0.99$ $S_{p:(1)2} = 0.99$	Mean	0.890	0.057	0.067	0.927	0.034	0.037	0.937	0.026	0.028	0.931	0.031	0.037
			0.889	0.055	0.067	0.928	0.033	0.037	0.938	0.026	0.028	0.933	0.031	0.036
			0.990	0.004	0.005	0.989	0.005	0.005	0.988	0.005	0.006	0.988	0.005	0.005
			0.989	0.004	0.005	0.988	0.004	0.005	0.988	0.005	0.006	0.988	0.005	0.006
		MAP	0.941	0.058	0.067	0.947	0.037	0.037	0.950	0.026	0.028	0.949	0.034	0.037
			0.940	0.058	0.067	0.948	0.036	0.037	0.950	0.027	0.028	0.951	0.034	0.036
			0.990	0.004	0.005	0.990	0.005	0.005	0.990	0.005	0.006	0.990	0.006	0.005
			0.990	0.004	0.005	0.990	0.005	0.005	0.990	0.005	0.006	0.990	0.006	0.006

Table 4 CT/NG testing protocols. Hierarchical protocols H2 and H3 use two and three stages, respectively (Hou et al., 2017). The array testing (AT) protocol uses square arrays (Hou et al., 2020). For each protocol, the configurations listed below minimize the expected number of tests per individual specimen overall. The Aptima Combo 2 Assay accuracy values, reported in the Package Inserts at `www.hologic.com`, are also shown.

	Protocol	Pool sizes			Sensitivity	Specificity
Urine	H2	$4:1$	Urine	CT	$S_{e:(1)1} = 0.947$	$S_{p:(1)1} = 0.989$
	H3	$9:3:1$		NG	$S_{e:(1)2} = 0.913$	$S_{p:(1)2} = 0.993$
	AT	8×8				
Swab	H2	$4:1$	Swab	CT	$S_{e:(1)1} = 0.942$	$S_{p:(1)1} = 0.976$
	H3	$9:3:1$		NG	$S_{e:(1)2} = 0.992$	$S_{p:(1)2} = 0.987$
	AT	8×8				

Table 5 CT/NG data application results for the characteristics of posterior mean (Mean) and MAP estimates of $\mathbf{p} = (p_{00}, p_{10}, p_{01}, p_{11})'$ when assay accuracy probabilities are **unknown**. Estimates ("Est") are averages over $B = 500$ Monte Carlo data sets and "SE" is the estimated posterior standard deviation. Flat priors have been used for all parameters; i.e., $\mathbf{p} \sim$ Dirichlet($\mathbf{1}_4$), $S_{e:(1)k} \sim$ beta$(1,1)$, and $S_{p:(1)k} \sim$ beta$(1,1)$, for $k = 1, 2$. The average number of tests (over $B = 500$ data sets) is also shown; the percent reduction (in parentheses) is compared to the two-stage hierarchical protocol H2 in Tebbs, McMahan, and Bilder (2013) and Warasi et al. (2016). The sample size, N, in each stratum is reported. The specific protocols used (H2, H3, and AT) are identified in Table 4.

				H2		H3		AT	
Stratum		CT	NG	Est	SE	Est	SE	Est	SE
Urine $N = 4402$	Mean	$-$	$-$	$\widehat{p}_{00} = 0.907$	0.0065	$\widehat{p}_{00} = 0.908$	0.0055	$\widehat{p}_{00} = 0.909$	0.0054
		$+$	$-$	$\widehat{p}_{10} = 0.081$	0.0062	$\widehat{p}_{10} = 0.081$	0.0053	$\widehat{p}_{10} = 0.080$	0.0051
		$-$	$+$	$\widehat{p}_{01} = 0.007$	0.0018	$\widehat{p}_{01} = 0.006$	0.0013	$\widehat{p}_{01} = 0.006$	0.0013
		$+$	$+$	$\widehat{p}_{11} = 0.006$	0.0013	$\widehat{p}_{11} = 0.005$	0.0012	$\widehat{p}_{11} = 0.005$	0.0011
	MAP	$-$	$-$	$\widehat{p}_{00} = 0.908$	0.0065	$\widehat{p}_{00} = 0.908$	0.0055	$\widehat{p}_{00} = 0.908$	0.0054
		$+$	$-$	$\widehat{p}_{10} = 0.081$	0.0062	$\widehat{p}_{10} = 0.081$	0.0053	$\widehat{p}_{10} = 0.081$	0.0051
		$-$	$+$	$\widehat{p}_{01} = 0.006$	0.0018	$\widehat{p}_{01} = 0.006$	0.0013	$\widehat{p}_{01} = 0.006$	0.0013
		$+$	$+$	$\widehat{p}_{11} = 0.005$	0.0013	$\widehat{p}_{11} = 0.005$	0.0012	$\widehat{p}_{11} = 0.005$	0.0011
	Avg. number of tests			2489.6		2332.9 (6.3%)		2333.3 (6.3%)	
Swab $N = 10048$	Mean	$-$	$-$	$\widehat{p}_{00} = 0.907$	0.0052	$\widehat{p}_{00} = 0.908$	0.0039	$\widehat{p}_{00} = 0.908$	0.0038
		$+$	$-$	$\widehat{p}_{10} = 0.081$	0.0051	$\widehat{p}_{10} = 0.082$	0.0038	$\widehat{p}_{10} = 0.081$	0.0037
		$-$	$+$	$\widehat{p}_{01} = 0.006$	0.0011	$\widehat{p}_{01} = 0.006$	0.0008	$\widehat{p}_{01} = 0.006$	0.0008
		$+$	$+$	$\widehat{p}_{11} = 0.005$	0.0009	$\widehat{p}_{11} = 0.005$	0.0007	$\widehat{p}_{11} = 0.005$	0.0007
	MAP	$-$	$-$	$\widehat{p}_{00} = 0.908$	0.0052	$\widehat{p}_{00} = 0.909$	0.0039	$\widehat{p}_{00} = 0.909$	0.0038
		$+$	$-$	$\widehat{p}_{10} = 0.081$	0.0051	$\widehat{p}_{10} = 0.081$	0.0038	$\widehat{p}_{10} = 0.081$	0.0037
		$-$	$+$	$\widehat{p}_{01} = 0.006$	0.0011	$\widehat{p}_{01} = 0.005$	0.0008	$\widehat{p}_{01} = 0.005$	0.0008
		$+$	$+$	$\widehat{p}_{11} = 0.005$	0.0009	$\widehat{p}_{11} = 0.005$	0.0007	$\widehat{p}_{11} = 0.005$	0.0007
	Avg. number of tests			$\overline{T} = 5802.8$		$\overline{T} = 5400.0(6.9)$		$\overline{T} = 5354.7(7.7)$	

Table 6 CT/NG data application results for the characteristics of posterior mean (Mean) and MAP estimates of $\boldsymbol{\delta} = (S_{e:(1)1}, S_{e:(1)2}, S_{p:(1)1}, S_{p:(1)2})'$. Estimates ("Est") are averages over $B = 500$ Monte Carlo data sets and "SE" is the estimated posterior standard deviation. Flat priors have been used for all parameters; i.e., $\mathbf{p} \sim$ Dirichlet$(\mathbf{1}_4)$, $S_{e:(1)k} \sim$ beta$(1,1)$, and $S_{p:(1)k} \sim$ beta$(1,1)$, for $k = 1, 2$. The sample size, N, in each stratum is reported. The specific protocols used (H2, H3, and AT) are identified in Table 4.

			H2		H3		AT	
Stratum	Accuracy		Est	SE	Est	SE	Est	SE
Urine $N = 4402$	$S_{e:(1)1} = 0.947$ $S_{e:(1)2} = 0.913$ $S_{p:(1)1} = 0.989$ $S_{p:(1)2} = 0.993$	Mean	$\widehat{S}_{e:(1)1} = 0.948$	0.022	$\widehat{S}_{e:(1)1} = 0.946$	0.012	$\widehat{S}_{e:(1)1} = 0.948$	0.011
			$\widehat{S}_{e:(1)2} = 0.875$	0.066	$\widehat{S}_{e:(1)2} = 0.899$	0.034	$\widehat{S}_{e:(1)2} = 0.904$	0.032
			$\widehat{S}_{p:(1)1} = 0.987$	0.007	$\widehat{S}_{p:(1)1} = 0.987$	0.006	$\widehat{S}_{p:(1)1} = 0.985$	0.008
			$\widehat{S}_{p:(1)2} = 0.993$	0.003	$\widehat{S}_{p:(1)2} = 0.993$	0.002	$\widehat{S}_{p:(1)2} = 0.993$	0.002
		MAP	$\widehat{S}_{e:(1)1} = 0.948$	0.022	$\widehat{S}_{e:(1)1} = 0.947$	0.012	$\widehat{S}_{e:(1)1} = 0.948$	0.011
			$\widehat{S}_{e:(1)2} = 0.911$	0.066	$\widehat{S}_{e:(1)2} = 0.911$	0.034	$\widehat{S}_{e:(1)2} = 0.914$	0.032
			$\widehat{S}_{p:(1)1} = 0.989$	0.007	$\widehat{S}_{p:(1)1} = 0.989$	0.006	$\widehat{S}_{p:(1)1} = 0.989$	0.008
			$\widehat{S}_{p:(1)2} = 0.993$	0.003	$\widehat{S}_{p:(1)2} = 0.993$	0.002	$\widehat{S}_{p:(1)2} = 0.993$	0.002
Swab $N = 10048$	$S_{e:(1)1} = 0.942$ $S_{e:(1)2} = 0.992$ $S_{p:(1)1} = 0.976$ $S_{p:(1)2} = 0.987$	Mean	$\widehat{S}_{e:(1)1} = 0.940$	0.019	$\widehat{S}_{e:(1)1} = 0.941$	0.009	$\widehat{S}_{e:(1)1} = 0.941$	0.009
			$\widehat{S}_{e:(1)2} = 0.936$	0.040	$\widehat{S}_{e:(1)2} = 0.985$	0.009	$\widehat{S}_{e:(1)2} = 0.986$	0.009
			$\widehat{S}_{p:(1)1} = 0.976$	0.006	$\widehat{S}_{p:(1)1} = 0.976$	0.005	$\widehat{S}_{p:(1)1} = 0.975$	0.007
			$\widehat{S}_{p:(1)2} = 0.988$	0.002	$\widehat{S}_{p:(1)2} = 0.987$	0.002	$\widehat{S}_{p:(1)2} = 0.987$	0.002
		MAP	$\widehat{S}_{e:(1)1} = 0.942$	0.019	$\widehat{S}_{e:(1)1} = 0.942$	0.009	$\widehat{S}_{e:(1)1} = 0.942$	0.009
			$\widehat{S}_{e:(1)2} = 0.984$	0.040	$\widehat{S}_{e:(1)2} = 0.991$	0.009	$\widehat{S}_{e:(1)2} = 0.991$	0.009
			$\widehat{S}_{p:(1)1} = 0.976$	0.006	$\widehat{S}_{p:(1)1} = 0.976$	0.005	$\widehat{S}_{p:(1)1} = 0.976$	0.007
			$\widehat{S}_{p:(1)2} = 0.987$	0.002	$\widehat{S}_{p:(1)2} = 0.987$	0.002	$\widehat{S}_{p:(1)2} = 0.987$	0.002

Web-based Supporting Materials for "Estimating the prevalence of two or more diseases using outcomes from multiplex group testing"

Md S. Warasi[1], Joshua M. Tebbs*[2], Christopher S. McMahan[3], and Christopher R. Bilder[4]

[1]Department of Mathematics and Statistics, Radford University, Radford, VA 24142, USA
[2]Department of Statistics, University of South Carolina, Columbia, SC 29208, USA
[3]School of Mathematical and Statistical Sciences, Clemson University, Clemson, SC 29634, USA
[4]Department of Statistics, University of Nebraska-Lincoln, Lincoln, NE 68583, USA
*email: tebbs@stat.sc.edu

Web Appendix A. *Conditional multinomial distribution.*

The full conditional distribution $\widetilde{\mathbf{V}}_i|\widetilde{\mathbf{Y}}_{-i}, \mathbf{p}, \mathbf{Z}, \boldsymbol{\delta} \sim$ multinomial$(p^*_{(00)i}, p^*_{(10)i}, p^*_{(01)i}, p^*_{(11)i})$ with cell probabilities $p^*_{(uv)i} = \zeta^i_{uv}/\zeta^i$, for $u, v \in \{0, 1\}$ and $i = 1, 2, ..., N$, is used in the posterior sampling algorithm and the EM algorithm in Section 3, where

$$\begin{aligned}
\zeta^i_{00} &= p_{00} \prod_{j \in \mathcal{A}_i} \prod_{k=1}^{2} \left(S^{Z_{jk}}_{e:jk} \overline{S}^{1-Z_{jk}}_{e:jk}\right)^{\gamma_{ijk}} \left(S^{1-Z_{jk}}_{p:jk} \overline{S}^{Z_{jk}}_{p:jk}\right)^{1-\gamma_{ijk}} \\
\zeta^i_{10} &= p_{10} \prod_{j \in \mathcal{A}_i} S^{Z_{j1}}_{e:j1} \overline{S}^{1-Z_{j1}}_{e:j1} \left(S^{Z_{j2}}_{e:j2} \overline{S}^{1-Z_{j2}}_{e:j2}\right)^{\gamma_{ij2}} \left(S^{1-Z_{j2}}_{p:j2} \overline{S}^{Z_{j2}}_{p:j2}\right)^{1-\gamma_{ij2}} \\
\zeta^i_{01} &= p_{01} \prod_{j \in \mathcal{A}_i} \left(S^{Z_{j1}}_{e:j1} \overline{S}^{1-Z_{j1}}_{e:j1}\right)^{\gamma_{ij1}} \left(S^{1-Z_{j1}}_{p:j1} \overline{S}^{Z_{j1}}_{p:j1}\right)^{1-\gamma_{ij1}} S^{Z_{j2}}_{e:j2} \overline{S}^{1-Z_{j2}}_{e:j2} \\
\zeta^i_{11} &= p_{11} \prod_{j \in \mathcal{A}_i} \prod_{k=1}^{2} S^{Z_{jk}}_{e:jk} \overline{S}^{1-Z_{jk}}_{e:jk},
\end{aligned}$$

$\zeta^i = \sum_{u=0}^{1} \sum_{v=0}^{1} \zeta^i_{uv}$, $\overline{S}_{e:jk} = 1 - S_{e:jk}$, $\overline{S}_{p:jk} = 1 - S_{p:jk}$, $\mathcal{A}_i = \{j : i \in \mathcal{P}_j\}$, $\mathcal{P}_{j(i)} = \mathcal{P}_j \backslash \{i\}$, and $\gamma_{ijk} = I(\sum_{i' \in \mathcal{P}_{j(i)}} \widetilde{Y}_{i'k} > 0)$.

Web Appendix B. *Additional information for the EM algorithm.*

The EM algorithm in Section 3.2 uses the solution $\mathbf{p}^{(t+1)} = (p_{00}^{(t+1)}, p_{10}^{(t+1)}, p_{01}^{(t+1)}, p_{11}^{(t+1)})'$ in the M-step, where

$$p_{00}^{(t+1)} = \frac{1}{\breve{N}}\left[\alpha_{00} - 1 + \sum_{i=1}^{N} E(\widetilde{V}_{(00)i}|\mathbf{Z}, \boldsymbol{\delta}; \mathbf{p}^{(t)})\right]$$
$$p_{10}^{(t+1)} = \frac{1}{\breve{N}}\left[\alpha_{10} - 1 + \sum_{i=1}^{N} E(\widetilde{V}_{(10)i}|\mathbf{Z}, \boldsymbol{\delta}; \mathbf{p}^{(t)})\right]$$
$$p_{01}^{(t+1)} = \frac{1}{\breve{N}}\left[\alpha_{01} - 1 + \sum_{i=1}^{N} E(\widetilde{V}_{(01)i}|\mathbf{Z}, \boldsymbol{\delta}; \mathbf{p}^{(t)})\right]$$
$$p_{11}^{(t+1)} = \frac{1}{\breve{N}}\left[\alpha_{11} - 1 + \sum_{i=1}^{N} E(\widetilde{V}_{(11)i}|\mathbf{Z}, \boldsymbol{\delta}; \mathbf{p}^{(t)})\right],$$

$\breve{N} = N - 4 + \sum_{u=0}^{1}\sum_{v=0}^{1}\alpha_{uv}$, and $\widetilde{V}_{(uv)i} = \widetilde{Y}_{i1}^{u}(1-\widetilde{Y}_{i1})^{1-u}\widetilde{Y}_{i2}^{v}(1-\widetilde{Y}_{i2})^{1-v}$, for $u, v \in \{0, 1\}$. One can easiliy see that $\mathbf{p}^{(t+1)}$ with flat Dirichlet prior (i.e., when $\alpha_{00} = \alpha_{10} = \alpha_{01} = \alpha_{11} = 1$) reduces to the maximum likelihood solution provided in Tebbs et al. (2013).

Web Appendix C. *Algorithms for the estimation techniques with unknown $\boldsymbol{\delta}$.*

In Section 4, we have briefly described the posterior sampling algorithm and the EM algorithm for the scenario that the assay accuracy probabilities in $\boldsymbol{\delta}$ are unknown. We now present the algorithms completely.

<u>POSTERIOR SAMPLING ALGORITHM</u>

1. Initialize $\mathbf{p}^{(0)} = (p_{00}^{(0)}, p_{10}^{(0)}, p_{01}^{(0)}, p_{11}^{(0)})'$, where $\sum_{u=0}^{1}\sum_{v=0}^{1} p_{uv}^{(0)} = 1$, and simulate $\widetilde{\mathbf{Y}}_i^{(0)} = (\widetilde{Y}_{i1}^{(0)}, \widetilde{Y}_{i2}^{(0)})'$, for $i = 1, 2, ..., N$, from the population-level multinomial at $\mathbf{p}^{(0)}$. Initialize $S_{e:(l)k}^{(0)}$ and $S_{p:(l)k}^{(0)}$, for $k = 1, 2; l = 1, 2, ..., L$, and accumulate them into $\boldsymbol{\delta}^{(0)}$. Set $g = 1$.

2. For $i = 1, 2, ..., N$, sample $\widetilde{\mathbf{V}}_i^{(g)} = (\widetilde{V}_{(00)i}^{(g)}, \widetilde{V}_{(10)i}^{(g)}, \widetilde{V}_{(01)i}^{(g)}, \widetilde{V}_{(11)i}^{(g)})'$ from

$$\widetilde{\mathbf{V}}_i|\widetilde{\mathbf{Y}}_{-i}^{(g)}, \mathbf{p}^{(g-1)}, \boldsymbol{\delta}^{(g-1)}, \mathbf{Z} \sim \text{multinomial}(p_{(00)i}^{*}, p_{(10)i}^{*}, p_{(01)i}^{*}, p_{(11)i}^{*}),$$

   where $\widetilde{\mathbf{Y}}_{-i}^{(g)} = (\widetilde{\mathbf{Y}}_1^{(g)'}, \cdots, \widetilde{\mathbf{Y}}_{i-1}^{(g)'}, \widetilde{\mathbf{Y}}_{i+1}^{(g-1)'}, \cdots, \widetilde{\mathbf{Y}}_N^{(g-1)'})'$ and $\widetilde{\mathbf{Y}}_i^{(g)} = (\widetilde{V}_{(10)i}^{(g)} + \widetilde{V}_{(11)i}^{(g)}, \widetilde{V}_{(01)i}^{(g)} + \widetilde{V}_{(11)i}^{(g)})'$.

3. Sample $\mathbf{p}^{(g)}$ from $\mathbf{p}|\widetilde{\mathbf{Y}}^{(g)} \sim \text{Dirichlet}(\boldsymbol{\alpha}^{*})$, where $\widetilde{\mathbf{Y}}^{(g)} = (\widetilde{\mathbf{Y}}_1^{(g)'}, \widetilde{\mathbf{Y}}_2^{(g)'}, \cdots, \widetilde{\mathbf{Y}}_N^{(g)'})'$.

4. Sample $S_{e:(l)k}^{(g)}$ from $S_{e:(l)k}|\mathbf{Z}, \widetilde{\mathbf{Y}}^{(g)} \sim \text{beta}(a_{lk}^{*}, b_{lk}^{*})$ and $S_{p:(l)k}^{(g)}$ from $S_{p:(l)k}|\mathbf{Z}, \widetilde{\mathbf{Y}}^{(g)} \sim \text{beta}(c_{lk}^{*}, d_{lk}^{*})$, for $k = 1, 2; l = 1, 2, ..., L$, and accumulate them into $\boldsymbol{\delta}^{(g)}$.

5. Set $g = g + 1$ and repeat steps 2-5 while $g < G$, the number of Gibbs iterates.

For the EM algorithm, we have the closed-form solutions $\mathbf{p}^{(t+1)}$ and $\boldsymbol{\delta}^{(t+1)}$, where $\mathbf{p}^{(t+1)}$ is the same as shown in Appendix B and the assay accuracy probabilities in $\boldsymbol{\delta}^{(t+1)}$ are

$$S_{e:(l)k}^{(t+1)} = \frac{1}{\breve{J}_{lk}} \left[a_{lk} - 1 + \sum_{j \in \mathcal{M}(l)} Z_{jk} E(\widetilde{Z}_{jk} | \mathbf{Z}; \mathbf{p}^{(t)}, \boldsymbol{\delta}^{(t)}) \right]$$

$$S_{p:(l)k}^{(t+1)} = \frac{1}{J_{lk}^{*}} \left[c_{lk} - 1 + \sum_{j \in \mathcal{M}(l)} (1 - Z_{jk}) \{1 - E(\widetilde{Z}_{jk} | \mathbf{Z}; \mathbf{p}^{(t)}, \boldsymbol{\delta}^{(t)})\} \right],$$

for $k = 1, 2$ and $l = 1, 2, ..., L$, where $\breve{J}_{lk} = a_{lk} + b_{lk} - 2 + \sum_{j \in \mathcal{M}(l)} E(\widetilde{Z}_{jk} | \mathbf{Z}; \mathbf{p}^{(t)}, \boldsymbol{\delta}^{(t)})$, $J_{lk}^{*} = c_{lk} + d_{lk} - 2 + \sum_{j \in \mathcal{M}(l)} \{1 - E(\widetilde{Z}_{jk} | \mathbf{Z}; \mathbf{p}^{(t)}, \boldsymbol{\delta}^{(t)})\}$, and $E(\widetilde{Z}_{jk} | \mathbf{Z}; \mathbf{p}^{(t)}, \boldsymbol{\delta}^{(t)})$ is approximated using Gibbs samples in the E-Step as shown below.

MAP ESTIMATION VIA EM ALGORITHM

1. Initialize $\mathbf{p}^{(0)} = (p_{00}^{(0)}, p_{10}^{(0)}, p_{01}^{(0)}, p_{11}^{(0)})'$, where $\sum_{u=0}^{1} \sum_{v=0}^{1} p_{uv}^{(0)} = 1$, and simulate $\widetilde{\mathbf{Y}}_i^{(0)} = (\widetilde{Y}_{i1}^{(0)}, \widetilde{Y}_{i2}^{(0)})'$, for $i = 1, 2, ..., N$, from the population-level multinomial at $\mathbf{p}^{(0)}$. Initialize $S_{e:(l)k}^{(0)}$ and $S_{p:(l)k}^{(0)}$, for $k = 1, 2; l = 1, 2, ..., L$, and accumulate into $\boldsymbol{\delta}^{(0)}$. Set $t = 0$.

2. (E-Step): For $i = 1, 2, ..., N$,
   - sample $\widetilde{\mathbf{V}}_i^{(g)} = (\widetilde{V}_{(00)i}^{(g)}, \widetilde{V}_{(10)i}^{(g)}, \widetilde{V}_{(01)i}^{(g)}, \widetilde{V}_{(11)i}^{(g)})'$, for $g = 1, 2, ..., G$, from the conditional $\widetilde{\mathbf{V}}_i | \widetilde{\mathbf{Y}}_{-i}^{(g)}, \mathbf{p}^{(t)}, \boldsymbol{\delta}^{(t)}, \mathbf{Z} \sim$ multinomial$(p_{(00)i}^{*}, p_{(10)i}^{*}, p_{(01)i}^{*}, p_{(11)i}^{*})$, where G is the number of Gibbs iterates;
   - calculate the sample mean $G^{-1} \sum_{g=1}^{G} (\widetilde{V}_{(00)i}^{(g)}, \widetilde{V}_{(10)i}^{(g)}, \widetilde{V}_{(01)i}^{(g)}, \widetilde{V}_{(11)i}^{(g)})'$ as an estimate of $E[(\widetilde{V}_{(00)i}, \widetilde{V}_{(10)i}, \widetilde{V}_{(01)i}, \widetilde{V}_{(11)i})' | \mathbf{Z}; \mathbf{p}^{(t)}, \boldsymbol{\delta}^{(t)}]$;
   - calculate the sample mean $G^{-1} \sum_{g=1}^{G} \widetilde{Z}_{jk}^{(g)}$ as an estimate of $E(\widetilde{Z}_{jk} | \mathbf{Z}; \mathbf{p}^{(t)}, \boldsymbol{\delta}^{(t)})$, where $\widetilde{Z}_{jk}^{(g)} = I(\sum_{i \in \mathcal{P}_j} \widetilde{Y}_{ik}^{(g)} > 0)$, $\widetilde{Y}_{i1}^{(g)} = \widetilde{V}_{(10)i}^{(g)} + \widetilde{V}_{(11)i}^{(g)}$, and $\widetilde{Y}_{i2}^{(g)} = \widetilde{V}_{(01)i}^{(g)} + \widetilde{V}_{(11)i}^{(g)}$.

3. (M-Step): Calculate $\mathbf{p}^{(t+1)}$ and $\boldsymbol{\delta}^{(t+1)}$ using the solutions above.

4. Set $t = t + 1$, and repeat steps 2-4 until the maximum absolute difference in $(\mathbf{p}'^{(t+1)}, \boldsymbol{\delta}'^{(t+1)}) - (\mathbf{p}'^{(t)}, \boldsymbol{\delta}'^{(t)})$ is less than ϵ, where ϵ is small.

Web Appendix D. *Additional simulation results.*

In Section 5 of the article, we have described data simulation with $L = 1$ assay. Tables 2-3 of the article show the simulation evidence with the assumption that $\boldsymbol{\delta}$ is unknown, but Table D.1 in this appendix shows the simulation evidence assuming $\boldsymbol{\delta}$ is known. We herein perform an additional simulation study with $L = 2$ assays assuming that the assay accuracy probabilities in $\boldsymbol{\delta} = (S_{e:(1)1}, S_{e:(1)2}, S_{p:(1)1}, S_{p:(1)2}, S_{e:(2)1}, S_{e:(2)2}, S_{p:(2)1}, S_{p:(2)2})'$ are unknown. The description is provided below.

For data simulation, the pooling protocols and disease probability configurations used here are the same as those in Section 5. The assay accuracy probabilities that we now use are $S_{e:(1)k} = 0.95$ and $S_{p:(1)k} = 0.98$ for assay 1 and $S_{e:(2)k} = 0.98$ and $S_{p:(2)k} = 0.99$ for assay 2, for $k = 1, 2$. For all protocols, assay 1 is used for pooled sample testing and assay 2 is used for individual testing in the final stage. For example, when the testing protocol is H4 (four-stage hierarchical), $S_{e:(1)k} = 0.95$ and $S_{p:(1)k} = 0.98$, for $k = 1, 2$, are used for simulating pooled responses in stages 1-3 but $S_{e:(2)k} = 0.98$ and $S_{p:(2)k} = 0.99$, for $k = 1, 2$, are used for simulating individual test responses in stage 4. As in Section 5, we use pool sizes that minimize the expected number of tests. Assuming $\boldsymbol{\delta}$ is unknown, we now estimate 12 parameters (4 parameters in $\mathbf{p}$ and 8 parameters in $\boldsymbol{\delta}$). We consider two scenarios of prior distribution.

- $\mathbf{p} \sim$ Dirichlet$(\mathbf{1}_4)$, $S_{e:(l)k} \sim$ beta$(1, 1)$, and $S_{p:(l)k} \sim$ beta$(1, 1)$, for $k = 1, 2$ and $l = 1, 2$.
- $\mathbf{p} \sim$ Dirichlet$(\mathbf{1}_4)$, $S_{e:(1)k} \sim$ beta$(109, 6.7)$, $S_{e:(2)k} \sim$ beta$(100, 3)$, $S_{p:(1)k} \sim$ beta$(100, 3)$, and $S_{p:(2)k} \sim$ beta$(55.2, 1.6)$, for $k = 1, 2$.

In the first scenario, priors for all parameters are noninformative, as in the article; the results are shown in Tables D.3-D.5. In the second, the assay accuracy parameters use informative beta priors; the results are in Tables D.6-D.8. In each of the beta informative priors, the mode is located at the true value. For example, the mode of beta$(100, 3)$ is 0.98. Overall, the estimation results are accurate and the finding discovered in Section 5 with $L = 1$ assay is retained. As one would expect, estimates with informative priors become more precise.

For convenience, tables presented in this section are listed below.

- Table D.1: $L = 1$ assay is used where $\boldsymbol{\delta}$ is **known** (results from Section 5).
  - Results for estimating $\mathbf{p}$ for Configurations I-II.
- Table D.2: Pool sizes with $L = 2$ assays.
- Tables D.3-D.5: $L = 2$ assays are used where $\boldsymbol{\delta}$ is **unknown** and **noninformative priors** have been used for all parameters. Results from the same posterior distribution are split into three tables.
  - Table D.3: Results for estimating $\mathbf{p}$ for Configurations I-II.
  - Table D.4: Results for estimating $\boldsymbol{\delta}$ for Configuration I.

  - Table D.5: Results for estimating $\boldsymbol{\delta}$ for Configuration II.

- Tables D.6-D.8: $L = 2$ assays are used where $\boldsymbol{\delta}$ is **unknown** and **informative priors** have been used for the assay accuracy parameters. Results from the same posterior distribution are split into three tables.
  - Table D.6: Results for estimating $\mathbf{p}$ for Configurations I-II.
  - Table D.7: Results for estimating $\boldsymbol{\delta}$ for Configuration I.
  - Table D.8: Results for estimating $\boldsymbol{\delta}$ for Configuration II.

Table D.1: Simulation evidence for the characteristics of posterior mean (Mean) and MAP estimates of $\mathbf{p} = (p_{00}, p_{10}, p_{01}, p_{11})'$ when assay accuracy probabilities are **known**. Estimates ("Est") are averages over $B = 500$ Monte Carlo data sets, "SD" is the sample standard deviation of the 500 estimates, and "SE" is the estimated posterior standard deviation. The Dirichlet flat prior has been used for $\mathbf{p}$; i.e., $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4)$. The average number of tests (over $B = 500$ data sets) is also shown; the percent reduction (in parentheses) is compared to the two-stage hierarchical protocol H2 in Tebbs, McMahan, and Bilder (2013) and Warasi et al. (2016). The specific protocols used (H2, H3, H4, and AT) are identified in Table 1 of the article.

			H2			H3			H4			AT		
True			Est	SE	SD	Est	SE	SD	Est	SE	SD	Est	SE	SD
Configuration I	$p_{00} = 0.95$ $p_{10} = 0.02$ $p_{01} = 0.02$ $p_{11} = 0.01$	Mean	0.949	0.0033	0.0032	0.949	0.0032	0.0031	0.949	0.0033	0.0033	0.949	0.0032	0.0032
			0.020	0.0021	0.0021	0.020	0.0021	0.0021	0.020	0.0021	0.0021	0.020	0.0021	0.0022
			0.020	0.0021	0.0020	0.020	0.0021	0.0021	0.020	0.0021	0.0022	0.020	0.0021	0.0022
			0.010	0.0015	0.0014	0.010	0.0014	0.0014	0.010	0.0014	0.0014	0.010	0.0015	0.0015
		MAP	0.950	0.0033	0.0031	0.950	0.0032	0.0032	0.950	0.0033	0.0033	0.950	0.0032	0.0031
			0.020	0.0021	0.0021	0.020	0.0021	0.0021	0.020	0.0021	0.0020	0.020	0.0021	0.0021
			0.020	0.0021	0.0021	0.020	0.0021	0.0020	0.020	0.0021	0.0022	0.020	0.0021	0.0021
			0.010	0.0015	0.0014	0.010	0.0014	0.0014	0.010	0.0014	0.0014	0.010	0.0015	0.0015
	Avg. number of tests			2164.3			1859.0 (14.1%)			1858.3 (14.1%)			1728.0 (20.2%)	
Configuration II	$p_{00} = 0.990$ $p_{10} = 0.004$ $p_{01} = 0.004$ $p_{11} = 0.002$	Mean	0.989	0.0015	0.0016	0.989	0.0015	0.0015	0.990	0.0015	0.0015	0.989	0.0015	0.0014
			0.004	0.0010	0.0010	0.004	0.0010	0.0010	0.004	0.0010	0.0009	0.004	0.0010	0.0009
			0.004	0.0010	0.0010	0.004	0.0010	0.0010	0.004	0.0010	0.0010	0.004	0.0010	0.0009
			0.002	0.0007	0.0007	0.002	0.0007	0.0006	0.002	0.0007	0.0006	0.002	0.0007	0.0006
		MAP	0.990	0.0015	0.0015	0.990	0.0015	0.0015	0.990	0.0015	0.0015	0.990	0.0015	0.0014
			0.004	0.0010	0.0009	0.004	0.0010	0.0009	0.004	0.0010	0.0009	0.004	0.0010	0.0009
			0.004	0.0010	0.0009	0.004	0.0010	0.0009	0.004	0.0010	0.0010	0.004	0.0010	0.0008
			0.002	0.0007	0.0006	0.002	0.0007	0.0006	0.002	0.0007	0.0006	0.002	0.0007	0.0006
	Avg. number of tests			1054.3			671.0 (36.4%)			581.4 (44.9%)			751.4 (28.7%)	

Table D.2: Multiplex group testing protocols with $L = 2$ assays. Hierarchical protocols H2, H3, and H4 use two, three, and four stages, respectively (Hou et al., 2017). The array testing (AT) protocol uses square arrays (Hou et al., 2020). For each protocol (except AT), the configurations listed below minimize the expected number of tests per individual specimen. "Configuration I" uses $\mathbf{p} = (0.95, 0.02, 0.02, 0.01)'$ and "Configuration II" uses $\mathbf{p} = (0.990, 0.004, 0.004, 0.002)'$.

Configuration I		Configuration II	
Protocol	Pool sizes	Protocol	Pool sizes
H2	$5:1$	H2	$11:1$
H3	$9:3:1$	H3	$24:6:1$
H4	$18:6:3:1$	H4	$36:12:4:1$
AT	11×11	AT	29×29

Note: For AT with $L = 2$ assays, the software that we use for optimal pool size calculation is not available yet. Thus, we use 29×29, the pool sizes reported in Table 1.

Table D.3: Simulation evidence with $L = 2$ assays when the assay accuracy probabilities are **unknown** and the priors used for these parameters are **noninformative**. The characteristics of posterior mean (Mean) and MAP estimates of $\mathbf{p} = (p_{00}, p_{10}, p_{01}, p_{11})'$ are presented. Estimates ("Est") are averages over $B = 500$ Monte Carlo data sets, "SD" is the sample standard deviation of the 500 estimates, and "SE" is the estimated posterior standard deviation. The priors used are $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4)$, $S_{e:(l)k} \sim \text{beta}(1,1)$, and $S_{p:(l)k} \sim \text{beta}(1,1)$, for $k = 1, 2$ and $l = 1, 2$. The average number of tests (over $B = 500$ data sets) is also shown; the percent reduction (in parentheses) is compared to the two-stage hierarchical protocol H2 in Tebbs, McMahan, and Bilder (2013) and Warasi et al. (2016). The specific protocols used (H2, H3, H4, and AT) are identified in Table D.2.

			H2			H3			H4			AT		
			Est	SD	SE	Est	SD	SE	Est	SD	SE	Est	SD	SE
Configuration I	$p_{00} = 0.95$ $p_{10} = 0.02$ $p_{01} = 0.02$ $p_{11} = 0.01$	Mean	0.946	0.0035	0.0042	0.949	0.0034	0.0035	0.949	0.0033	0.0034	0.948	0.0033	0.0036
			0.021	0.0025	0.0027	0.021	0.0024	0.0023	0.020	0.0022	0.0022	0.021	0.0023	0.0024
			0.022	0.0025	0.0028	0.021	0.0022	0.0023	0.020	0.0022	0.0022	0.021	0.0021	0.0024
			0.011	0.0015	0.0016	0.010	0.0015	0.0015	0.010	0.0015	0.0014	0.010	0.0015	0.0015
		MAP	0.949	0.0035	0.0042	0.950	0.0034	0.0035	0.950	0.0033	0.0034	0.950	0.0033	0.0036
			0.020	0.0024	0.0027	0.020	0.0023	0.0023	0.020	0.0022	0.0022	0.020	0.0023	0.0024
			0.020	0.0023	0.0028	0.020	0.0022	0.0023	0.020	0.0021	0.0022	0.020	0.0022	0.0024
			0.010	0.0015	0.0016	0.010	0.0015	0.0015	0.010	0.0015	0.0014	0.010	0.0015	0.0015
	Avg. number of tests			2248.5			1909.1 (15.1%)			1895.9 (15.7%)			1770.3 (21.3%)	
Configuration II	$p_{00} = 0.990$ $p_{10} = 0.004$ $p_{01} = 0.004$ $p_{11} = 0.002$	Mean	0.986	0.0025	0.0033	0.988	0.0016	0.0018	0.989	0.0015	0.0016	0.989	0.0015	0.0018
			0.006	0.0016	0.0021	0.005	0.0011	0.0012	0.004	0.0009	0.0010	0.004	0.0010	0.0011
			0.006	0.0018	0.0021	0.005	0.0012	0.0012	0.004	0.0011	0.0011	0.005	0.0010	0.0011
			0.003	0.0006	0.0009	0.002	0.0006	0.0007	0.002	0.0006	0.0007	0.002	0.0006	0.0007
		MAP	0.990	0.0017	0.0033	0.990	0.0017	0.0018	0.990	0.0015	0.0016	0.990	0.0016	0.0018
			0.004	0.0011	0.0021	0.004	0.0011	0.0012	0.004	0.0009	0.0010	0.004	0.0010	0.0011
			0.004	0.0012	0.0021	0.004	0.0012	0.0012	0.004	0.0011	0.0011	0.004	0.0011	0.0011
			0.002	0.0006	0.0009	0.002	0.0006	0.0007	0.002	0.0006	0.0007	0.002	0.0006	0.0007
	Avg. number of tests			1130.8			713.2 (36.9%)			610.8 (45.9%)			780.1 (31.0%)	

Table D.4: Simulation evidence with $L = 2$ assays for the characteristics of posterior mean (Mean) and MAP estimates of the parameters in $\boldsymbol{\delta} = (S_{e:(1)1}, S_{e:(1)2}, S_{p:(1)1}, S_{p:(1)2}, S_{e:(2)1}, S_{e:(2)2}, S_{p:(2)1}, S_{p:(2)2})'$. Estimates ("Est") are averages over $B = 500$ Monte Carlo data sets, "SD" is the sample standard deviation of the 500 estimates, and "SE" is the estimated posterior standard deviation. The priors used are $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4)$, $S_{e:(l)k} \sim \text{beta}(1,1)$, and $S_{p:(l)k} \sim \text{beta}(1,1)$, for $k = 1, 2$ and $l = 1, 2$. Results are shown for **Configuration I** which uses $\mathbf{p} = (0.95, 0.02, 0.02, 0.01)'$. The specific protocols used (H2, H3, H4, and AT) are identified in Table D.2.

		H2			H3			H4			AT		
		Est	SD	SE	Est	SD	SE	Est	SD	S3	Est	SD	SE
		0.924	0.033	0.042	0.943	0.020	0.021	0.946	0.014	0.015	0.941	0.021	0.023
		0.924	0.034	0.042	0.943	0.021	0.021	0.946	0.014	0.015	0.943	0.021	0.023
		0.983	0.005	0.007	0.979	0.006	0.006	0.979	0.006	0.006	0.981	0.007	0.009
		0.983	0.005	0.007	0.980	0.006	0.006	0.979	0.006	0.006	0.981	0.007	0.009
	Mean	0.944	0.021	0.037	0.971	0.014	0.017	0.971	0.014	0.016	0.965	0.017	0.022
$S_{e:(1)1} = 0.95$		0.944	0.022	0.036	0.971	0.014	0.017	0.971	0.013	0.016	0.965	0.016	0.022
$S_{e:(1)2} = 0.95$		0.989	0.004	0.005	0.988	0.005	0.006	0.987	0.005	0.006	0.989	0.005	0.005
$S_{p:(1)1} = 0.98$		0.990	0.004	0.004	0.988	0.005	0.005	0.988	0.005	0.006	0.988	0.005	0.005
$S_{p:(1)2} = 0.98$													
$S_{e:(2)1} = 0.98$		0.950	0.033	0.042	0.950	0.020	0.021	0.950	0.013	0.015	0.949	0.022	0.023
$S_{e:(2)2} = 0.98$		0.948	0.035	0.042	0.950	0.021	0.021	0.950	0.014	0.015	0.951	0.022	0.023
$S_{p:(2)1} = 0.99$		0.981	0.006	0.007	0.980	0.006	0.006	0.980	0.006	0.006	0.980	0.008	0.009
$S_{p:(2)2} = 0.99$		0.981	0.006	0.007	0.980	0.006	0.006	0.980	0.006	0.006	0.980	0.009	0.009
	MAP	0.973	0.026	0.037	0.980	0.015	0.017	0.980	0.014	0.016	0.979	0.019	0.022
		0.974	0.027	0.036	0.980	0.015	0.017	0.980	0.013	0.016	0.978	0.019	0.022
		0.990	0.004	0.005	0.990	0.005	0.006	0.989	0.006	0.006	0.991	0.005	0.005
		0.990	0.004	0.004	0.990	0.006	0.005	0.990	0.005	0.006	0.990	0.005	0.005

Table D.5: Simulation evidence with $L = 2$ assays for the characteristics of posterior mean (Mean) and MAP estimates of the parameters in $\boldsymbol{\delta} = (S_{e:(1)1}, S_{e:(1)2}, S_{p:(1)1}, S_{p:(1)2}, S_{e:(2)1}, S_{e:(2)2}, S_{p:(2)1}, S_{p:(2)2})'$. Estimates ("Est") are averages over $B = 500$ Monte Carlo data sets, "SD" is the sample standard deviation of the 500 estimates, and "SE" is the estimated posterior standard deviation. The priors used are $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4)$, $S_{e:(l)k} \sim \text{beta}(1,1)$, and $S_{p:(l)k} \sim \text{beta}(1,1)$, for $k = 1, 2$ and $l = 1, 2$. Results are shown for **Configuration II** which uses $\mathbf{p} = (0.990, 0.004, 0.004, 0.002)'$. The specific protocols used (H2, H3, H4, and AT) are identified in Table D.2.

		H2			H3			H4			AT		
		Est	SD	SE	Est	SD	SE	Est	SD	S3	Est	SD	SE
		0.805	0.087	0.126	0.904	0.048	0.055	0.933	0.031	0.035	0.911	0.038	0.056
		0.806	0.089	0.125	0.907	0.044	0.055	0.930	0.030	0.036	0.908	0.041	0.056
		0.984	0.006	0.009	0.981	0.007	0.009	0.979	0.008	0.009	0.982	0.007	0.011
		0.983	0.006	0.009	0.980	0.008	0.009	0.978	0.008	0.009	0.982	0.007	0.012
	Mean	0.865	0.048	0.092	0.933	0.030	0.051	0.940	0.026	0.046	0.923	0.029	0.059
$S_{e:(1)1} = 0.95$		0.864	0.047	0.093	0.934	0.031	0.050	0.940	0.029	0.045	0.923	0.032	0.059
$S_{e:(1)2} = 0.95$		0.989	0.005	0.005	0.986	0.007	0.008	0.983	0.008	0.010	0.987	0.006	0.006
$S_{p:(1)1} = 0.98$		0.989	0.005	0.005	0.986	0.007	0.008	0.984	0.008	0.010	0.987	0.006	0.006
$S_{p:(1)2} = 0.98$													
$S_{e:(2)1} = 0.98$		0.940	0.074	0.126	0.944	0.055	0.055	0.952	0.033	0.035	0.952	0.052	0.056
$S_{e:(2)2} = 0.98$		0.940	0.079	0.125	0.948	0.051	0.055	0.949	0.032	0.036	0.947	0.053	0.056
$S_{p:(2)1} = 0.99$		0.982	0.008	0.009	0.982	0.009	0.009	0.981	0.009	0.009	0.981	0.012	0.011
$S_{p:(2)2} = 0.99$		0.982	0.008	0.009	0.981	0.009	0.009	0.981	0.009	0.009	0.981	0.012	0.012
	MAP	0.961	0.061	0.092	0.977	0.032	0.051	0.980	0.028	0.046	0.978	0.039	0.059
		0.961	0.060	0.093	0.977	0.033	0.050	0.979	0.031	0.045	0.979	0.041	0.059
		0.990	0.005	0.005	0.990	0.007	0.008	0.990	0.008	0.010	0.990	0.006	0.006
		0.990	0.005	0.005	0.990	0.007	0.008	0.990	0.008	0.010	0.990	0.006	0.006

Table D.6: Simulation evidence with $L = 2$ assays when the assay accuracy probabilities are **unknown** and the priors used for these parameters are **informative**. The characteristics of posterior mean (Mean) and MAP estimates of $\mathbf{p} = (p_{00}, p_{10}, p_{01}, p_{11})'$ are presented. Estimates ("Est") are averages over $B = 500$ Monte Carlo data sets, "SD" is the sample standard deviation of the 500 estimates, and "SE" is the estimated posterior standard deviation. The priors used are $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4)$, $S_{e:(1)k} \sim \text{beta}(109, 6.7)$, $S_{e:(2)k} \sim \text{beta}(100, 3)$, $S_{p:(1)k} \sim \text{beta}(100, 3)$, and $S_{p:(2)k} \sim \text{beta}(55.2, 1.6)$, for $k = 1, 2$. The specific protocols used (H2, H3, H4, and AT) are identified in Table D.2.

			H2			H3			H4			AT		
			Est	SD	SE	Est	SD	SE	Est	SD	SE	Est	SD	SE
Configuration I	$p_{00} = 0.95$ $p_{10} = 0.02$ $p_{01} = 0.02$ $p_{11} = 0.01$	Mean	0.949	0.0033	0.0034	0.949	0.0032	0.0034	0.949	0.0033	0.0034	0.949	0.0033	0.0035
			0.021	0.0022	0.0023	0.020	0.0021	0.0022	0.020	0.0021	0.0022	0.021	0.0022	0.0023
			0.020	0.0022	0.0023	0.020	0.0021	0.0022	0.020	0.0021	0.0022	0.020	0.0022	0.0023
			0.010	0.0015	0.0015	0.010	0.0015	0.0014	0.010	0.0014	0.0014	0.010	0.0014	0.0015
		MAP	0.950	0.0032	0.0034	0.950	0.0032	0.0034	0.950	0.0033	0.0034	0.950	0.0032	0.0035
			0.020	0.0021	0.0023	0.020	0.0021	0.0022	0.020	0.0020	0.0022	0.020	0.0021	0.0023
			0.020	0.0021	0.0023	0.020	0.0021	0.0022	0.020	0.0021	0.0022	0.020	0.0021	0.0023
			0.010	0.0015	0.0015	0.010	0.0015	0.0014	0.010	0.0014	0.0014	0.010	0.0014	0.0015
Configuration II	$p_{00} = 0.990$ $p_{10} = 0.004$ $p_{01} = 0.004$ $p_{11} = 0.002$	Mean	0.989	0.0016	0.0016	0.989	0.0015	0.0016	0.989	0.0015	0.0015	0.989	0.0015	0.0016
			0.004	0.0010	0.0010	0.004	0.0009	0.0010	0.004	0.0010	0.0010	0.004	0.0010	0.0010
			0.004	0.0010	0.0011	0.004	0.0009	0.0010	0.004	0.0010	0.0010	0.004	0.0010	0.0010
			0.002	0.0006	0.0007	0.002	0.0007	0.0007	0.002	0.0006	0.0007	0.002	0.0006	0.0007
		MAP	0.990	0.0015	0.0016	0.990	0.0015	0.0016	0.990	0.0015	0.0015	0.990	0.0015	0.0016
			0.004	0.0009	0.0010	0.004	0.0009	0.0010	0.004	0.0010	0.0010	0.004	0.0010	0.0010
			0.004	0.0009	0.0011	0.004	0.0009	0.0010	0.004	0.0010	0.0010	0.004	0.0010	0.0010
			0.002	0.0006	0.0007	0.002	0.0007	0.0007	0.002	0.0006	0.0007	0.002	0.0006	0.0007

Table D.7: Simulation evidence with $L = 2$ assays for the characteristics of posterior mean (Mean) and MAP estimates of the parameters in $\boldsymbol{\delta} = (S_{e:(1)1}, S_{e:(1)2}, S_{p:(1)1}, S_{p:(1)2}, S_{e:(2)1}, S_{e:(2)2}, S_{p:(2)1}, S_{p:(2)2})'$. Estimates ("Est") are averages over $B = 500$ Monte Carlo data sets, "SD" is the sample standard deviation of the 500 estimates, and "SE" is the estimated posterior standard deviation. The priors used are $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4)$, $S_{e:(1)k} \sim \text{beta}(109, 6.7)$, $S_{e:(2)k} \sim \text{beta}(100, 3)$, $S_{p:(1)k} \sim \text{beta}(100, 3)$, and $S_{p:(2)k} \sim \text{beta}(55.2, 1.6)$, for $k = 1, 2$. Results are shown for **Configuration I** which uses $\mathbf{p} = (0.95, 0.02, 0.02, 0.01)'$. The specific protocols used (H2, H3, H4, and AT) are identified in Table D.2.

		H2			H3			H4			AT		
		Est	SD	SE	Est	SD	SE	Est	SD	S3	Est	SD	SE
		0.944	0.008	0.019	0.946	0.010	0.015	0.947	0.009	0.012	0.946	0.010	0.016
		0.943	0.008	0.019	0.946	0.011	0.015	0.947	0.009	0.012	0.946	0.010	0.016
		0.980	0.005	0.006	0.979	0.006	0.006	0.979	0.006	0.006	0.980	0.007	0.008
		0.980	0.005	0.006	0.980	0.006	0.006	0.979	0.006	0.006	0.980	0.007	0.008
	Mean	0.971	0.005	0.015	0.975	0.007	0.011	0.976	0.007	0.011	0.974	0.006	0.013
$S_{e:(1)1} = 0.95$		0.972	0.005	0.015	0.976	0.006	0.011	0.975	0.007	0.011	0.974	0.007	0.013
$S_{e:(1)2} = 0.95$		0.989	0.004	0.004	0.988	0.005	0.006	0.988	0.005	0.006	0.988	0.005	0.005
$S_{p:(1)1} = 0.98$		0.989	0.004	0.004	0.988	0.005	0.006	0.988	0.005	0.006	0.989	0.005	0.005
$S_{p:(1)2} = 0.98$													
$S_{e:(2)1} = 0.98$		0.950	0.008	0.019	0.950	0.010	0.015	0.950	0.009	0.012	0.950	0.010	0.016
$S_{e:(2)2} = 0.98$		0.950	0.008	0.019	0.950	0.010	0.015	0.950	0.009	0.012	0.950	0.009	0.016
$S_{p:(2)1} = 0.99$		0.980	0.005	0.006	0.980	0.005	0.006	0.980	0.005	0.006	0.980	0.005	0.008
$S_{p:(2)2} = 0.99$		0.980	0.004	0.006	0.980	0.005	0.006	0.980	0.005	0.006	0.980	0.005	0.008
	MAP	0.980	0.004	0.015	0.979	0.007	0.011	0.980	0.007	0.011	0.980	0.006	0.013
		0.980	0.004	0.015	0.980	0.006	0.011	0.980	0.007	0.011	0.980	0.006	0.013
		0.990	0.004	0.004	0.990	0.005	0.006	0.990	0.004	0.006	0.990	0.004	0.005
		0.990	0.004	0.004	0.990	0.004	0.006	0.990	0.004	0.006	0.990	0.004	0.005

Table D.8: Simulation evidence with $L = 2$ assays for the characteristics of posterior mean (Mean) and MAP estimates of the parameters in $\boldsymbol{\delta} = (S_{e:(1)1}, S_{e:(1)2}, S_{p:(1)1}, S_{p:(1)2}, S_{e:(2)1}, S_{e:(2)2}, S_{p:(2)1}, S_{p:(2)2})'$. Estimates ("Est") are averages over $B = 500$ Monte Carlo data sets, "SD" is the sample standard deviation of the 500 estimates, and "SE" is the estimated posterior standard deviation. The priors used are $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4)$, $S_{e:(1)k} \sim \text{beta}(109, 6.7)$, $S_{e:(2)k} \sim \text{beta}(100, 3)$, $S_{p:(1)k} \sim \text{beta}(100, 3)$, and $S_{p:(2)k} \sim \text{beta}(55.2, 1.6)$, for $k = 1, 2$. Results are shown for **Configuration II** which uses $\mathbf{p} = (0.95, 0.02, 0.02, 0.01)'$. The specific protocols used (H2, H3, H4, and AT) are identified in Table D.2.

		H2			H3			H4			AT		
		Est	SD	SE	Est	SD	SE	Est	SD	S3	Est	SD	SE
$S_{e:(1)1} = 0.95$	Mean	0.942	0.004	0.021	0.943	0.007	0.020	0.945	0.009	0.018	0.943	0.006	0.020
$S_{e:(1)2} = 0.95$		0.942	0.004	0.021	0.943	0.008	0.020	0.945	0.009	0.018	0.943	0.006	0.020
$S_{p:(1)1} = 0.98$		0.978	0.008	0.008	0.978	0.008	0.009	0.976	0.008	0.009	0.977	0.010	0.011
$S_{p:(1)2} = 0.98$		0.977	0.008	0.008	0.978	0.008	0.009	0.977	0.008	0.009	0.976	0.010	0.012
$S_{e:(2)1} = 0.98$		0.971	0.003	0.016	0.972	0.004	0.015	0.973	0.005	0.015	0.972	0.004	0.016
$S_{e:(2)2} = 0.98$		0.971	0.003	0.016	0.972	0.005	0.015	0.972	0.006	0.015	0.972	0.003	0.016
$S_{p:(2)1} = 0.99$		0.988	0.005	0.005	0.986	0.006	0.008	0.984	0.008	0.010	0.987	0.006	0.006
$S_{p:(2)2} = 0.99$		0.988	0.005	0.005	0.986	0.006	0.007	0.983	0.008	0.010	0.987	0.006	0.006
	MAP	0.950	0.004	0.021	0.950	0.007	0.020	0.951	0.009	0.018	0.950	0.007	0.020
		0.950	0.004	0.021	0.950	0.008	0.020	0.951	0.009	0.018	0.950	0.006	0.020
		0.980	0.006	0.008	0.980	0.006	0.009	0.980	0.006	0.009	0.980	0.006	0.011
		0.979	0.006	0.008	0.980	0.006	0.009	0.980	0.006	0.009	0.980	0.007	0.012
		0.980	0.002	0.016	0.980	0.004	0.015	0.980	0.005	0.015	0.980	0.004	0.016
		0.980	0.002	0.016	0.980	0.005	0.015	0.980	0.006	0.015	0.980	0.003	0.016
		0.990	0.004	0.005	0.990	0.005	0.008	0.990	0.006	0.010	0.990	0.005	0.006
		0.990	0.004	0.005	0.990	0.005	0.007	0.989	0.006	0.010	0.990	0.005	0.006

Web Appendix E. *Chlamydia and gonorrhea data.*

CT/NG classification results for the individuals tested at the SHL are shown in the table below. The disease classification has been determined by two-stage hierarchical testing and individual testing as described in Section 6 of the article.

Urine			Swab		
CT	NG	Count	CT	NG	Count
−	−	3998	−	−	9130
+	−	357	+	−	816
−	+	25	−	+	54
+	+	22	+	+	48
Total		4402	Total		10048

Web Appendix F. *R functions for Bayesian estimation.*

We provide two R functions to perform the simulations in Section 5 (with $L = 1$ assay) and Appendix D (with $L = 2$ assays). The function that uses one assay is described below, while the other function is presented at `https://github.com/mswarasi/General-MultiplexBayes`. The R code along with all simulation examples presented in this article is provided at the GitHub repository.

Usage

```
mult.gt.bayes(p0=c(.90,.06,.03,.01),delta0=c(.95,.95,.98,.98),Z,
              N,S,p.pr=rep(1,4),Se1.pr=c(1,1),
              Se2.pr=c(1,1),Sp1.pr=c(1,1),Sp2.pr=c(1,1),postGit=12000,
              emGit=12000,emburn=2000,emmaxit=200,emtol=1e-03,
              method=c("MAP","Bayes"),accuracy=c("unknown","known"))
```

Arguments

`p0`	An initial value of $\mathbf{p} = (p_{00}, p_{10}, p_{01}, p_{11})'$.
`delta0`	An initial value of $\boldsymbol{\delta} = (S_{e:(1)1}, S_{e:(1)2}, S_{p:(1)1}, S_{p:(1)2})'$. Used only when the assay accuracy probabilities are unknown.
`Z`	A matrix of the observed group testing data, Z.
`N`	Number of individuals tested (i.e., the sample size).
`S`	Maximum number of times an individual is tested in different pools.
`p.pr`	A 4×1 vector of Dirichlet hyperparameters for $\mathbf{p}$.
`Se1.pr`	A 2×1 vector of beta hyperparameters for $S_{e:(1)1}$.
`Se2.pr`	A 2×1 vector of beta hyperparameters for $S_{e:(1)2}$.
`Sp1.pr`	A 2×1 vector of beta hyperparameters for $S_{p:(1)1}$.
`Sp2.pr`	A 2×1 vector of beta hyperparameters for $S_{p:(1)2}$.
`postGit`	Number of Gibbs samples, G, to be drawn from the posterior dist.
`emGit`	Number of Gibbs samples, G, to be used in the EM algorithm.
`emburn`	Number of initial Gibbs samples to be discarded in the EM algorithm.
`emmaxit`	Maximum number of iterations the EM algorithm can run.
`emtol`	Convergence tolerance used in the EM algorithm.
`method`	Estimation method to be used: "MAP" or "Bayes". Defaults to "MAP".
`accuracy`	Whether assay accuracies are known or unknown. Defaults to "unknown".

Value

A list of components:

`prevalence`	An estimate of $\mathbf{p}$ for "MAP" but a matrix of posterior samples for "Bayes".
`accuracy`	An estimate of $\boldsymbol{\delta}$ for "MAP" but a matrix of posterior samples for "Bayes".
`convergence`	A binary indicator. For "MAP", 0 indicates successful convergence and 1 indicates that the EM algorithm reaches the max iteration. For "Bayes", convergence is always 0.

Web Appendix G. *Extensions to include $K \geqslant 2$ diseases.*

We show how the estimation techniques in the article can be extended to accommodate the general scenario that the number of diseases is 2 or more, i.e., $K \geqslant 2$. In doing so, we assume that the assay used in each stage of a group testing protocol can diagnose a pooled or individual specimen simultaneously for K diseases. An example of such assay is the Procleix Ultrio Assay (Grifols Diagnostics Solutions Inc.) which can detect HIV, HBV, and HCV from blood samples. The extensions are briefly described below.

Let $\widetilde{\mathbf{Y}}_i = (\widetilde{Y}_{i1}, \widetilde{Y}_{i2}, ..., \widetilde{Y}_{iK})'$ denote the vector of individual true statuses, where $\widetilde{Y}_{ik} = 1$ if individual i is truly positive for the kth disease and $\widetilde{Y}_{ik} = 0$ if otherwise, for $i = 1, 2, ..., N$ and $k = 1, 2, ..., K$. Let $\mathbf{p}$ denote a $2^K \times 1$ vector of the multinomial cell probabilities $\{p_\omega : \omega \in \Omega\}$ with $\sum_{\omega \in \mathbf{\Omega}} p_\omega = 1$, where $\omega = (\omega_1 \omega_2 \cdots \omega_K)$ is a K-tuple, $\omega_k \in \{0, 1\}$, and $\mathbf{\Omega}$ is a $2^K \times 1$ vector of ω_k's. When $K = 2$, we find $\mathbf{\Omega} = \{00, 10, 01, 11\}$ and $\mathbf{p} = (p_{00}, p_{10}, p_{01}, p_{11})'$ as in the article.

Let $\widetilde{\mathbf{Z}}_j = (\widetilde{Z}_{j1}, \widetilde{Z}_{j2}, ..., \widetilde{Z}_{jK})'$ denote the vector of true statuses of the jth pool, where $\widetilde{Z}_{jk} = I(\sum_{i \in \mathcal{P}_j} \widetilde{Y}_{ik} > 0)$ and $j = 1, 2, ..., J$. Denote by $\mathbf{Z}_j = (Z_{j1}, Z_{j2}, ..., Z_{jK})'$ the vector of test responses from the jth pool. The assay sensitivities and specificities are defined again as $\text{pr}(Z_{jk} = 1 | \widetilde{Z}_{jk} = 1) = S_{e:jk}$ and $\text{pr}(Z_{jk} = 0 | \widetilde{Z}_{jk} = 0) = S_{p:jk}$, for $k = 1, 2, ..., K$. As in the article, $\mathbf{Z} = (\mathbf{Z}_1{}', \mathbf{Z}_2{}', \cdots, \mathbf{Z}_J{}')'$ is the observed data and $\boldsymbol{\delta}$ is the collection of $S_{e:jk}$'s and $S_{p:jk}$'s. Using the same assumptions and prior distributions, the posterior sampling algorithm and the EM algorithm in Sections 3-4 can be easily extended for the general case $K \geqslant 2$.

Known assay accuracy probabilities: The full conditional of $\mathbf{p}$ is again Dirichlet; i.e., $\mathbf{p} | \widetilde{\mathbf{Y}} \sim \text{Dirichlet}(\boldsymbol{\alpha}^*)$, where $\boldsymbol{\alpha}^*$ is a $2^K \times 1$ vector of α^*_ω's, $\alpha^*_\omega = \alpha_\omega + \sum_{i=1}^N \widetilde{V}_{(\omega)i}$, and α_ω, for $\omega \in \mathbf{\Omega}$, are the Dirichlet hyperparameters. Let $\widetilde{\mathbf{V}}_i$ be a $2^K \times 1$ multinomial vector of $\{\widetilde{V}_{(\omega)i} : \omega \in \Omega\}$. Then we find the full conditional distribution $\widetilde{\mathbf{V}}_i | \widetilde{\mathbf{Y}}_{-i}, \mathbf{p}, \boldsymbol{\delta}, \mathbf{Z} \sim \text{multinomial}(\mathbf{p}^*_i)$, where $\mathbf{p}^*_i$ is a $2^K \times 1$ vector of the cell probabilities $p^*_{(\omega)i} = \zeta^i_\omega / \zeta^i$,

$$\zeta^i_\omega = p_\omega \prod_{j \in \mathcal{A}_i} \prod_{k=1}^{K} \left[S^{Z_{jk}}_{e:jk} \overline{S}^{1-Z_{jk}}_{e:jk} \right]^{\omega_k} \left[\left(S^{Z_{jk}}_{e:jk} \overline{S}^{1-Z_{jk}}_{e:jk} \right)^{\gamma_{ijk}} \left(S^{1-Z_{jk}}_{p:jk} \overline{S}^{Z_{jk}}_{p:jk} \right)^{1-\gamma_{ijk}} \right]^{1-\omega_k}$$

$\zeta^i = \sum_{\omega \in \Omega} \zeta^i_\omega$, and $\mathcal{A}_i$, $\widetilde{\mathbf{Y}}_{-i}$, and γ_{ijk} are as defined in Appendix A. When $K = 2$, ζ^i_ω provides the four cell probabilities in Appendix A. Using the full conditionals of $\mathbf{p}$ and $\widetilde{\mathbf{V}}_i$, one can perform posterior sampling exactly as shown in Section 3.1.

For MAP estimation, we find the closed-form solution $\mathbf{p}^{(t+1)}$ of components

$$p^{(t+1)}_\omega = \frac{1}{\breve{N}} \left[\alpha_\omega - 1 + \sum_{i=1}^{N} E(\widetilde{V}_{(\omega)i} | \mathbf{Z}, \boldsymbol{\delta}; \mathbf{p}^{(t)}) \right],$$

for $\omega \in \mathbf{\Omega}$, and $\breve{N} = N - 2^K + \sum_{\omega \in \Omega} \alpha_\omega$. Using $\mathbf{p}^{(t+1)}$ in the M-Step, we again find the EM algorithm as in Section 3.2. For the special case $K = 2$, $p^{(t+1)}_\omega$ yields the four cell probabilities shown in Appendix B.

Unknown assay accuracy probabilities: When $\boldsymbol{\delta}$ is unknown, the posterior sampling algorithm involves an additional step to sample $S_{e:(l)k}$ and $S_{p:(l)k}$ from the full conditionals

$$S_{e:(l)k}|\mathbf{Z}, \widetilde{\mathbf{Y}} \sim \text{beta}(a^*_{lk}, b^*_{lk})$$
$$S_{p:(l)k}|\mathbf{Z}, \widetilde{\mathbf{Y}} \sim \text{beta}(c^*_{lk}, d^*_{lk}),$$

for $k = 1, 2, ..., K$ and $l = 1, 2, ..., L$, where a^*_{lk}, b^*_{lk}, c^*_{lk}, and d^*_{lk} are provided in Section 4.

For MAP estimation, the solutions to be used in the EM algorithm are $\mathbf{p}^{(t+1)}$ and $\boldsymbol{\delta}^{(t+1)}$. The $\mathbf{p}^{(t+1)}$ of the $p_\omega^{(t+1)}$'s shown above with known $\boldsymbol{\delta}$ remains the same. The components $S^{(t+1)}_{e:(l)k}$ and $S^{(t+1)}_{p:(l)k}$ in $\boldsymbol{\delta}^{(t+1)}$ are shown in Appendix C where the subscripts are now $k = 1, 2, ..., K$ and $l = 1, 2, ..., L$. With these conditionals and solutions, the posterior sampling and EM algorithms for the $K \geqslant 2$ case can be written analogously as in Appendix C.