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 the specimens individually. In the biostatistics literature, testing pools of specimens is 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 with multiplex assays motivated by the current testing practices for chlamydia and gonorrhea in the United States. 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 testing data collected at the State Hygienic Laboratory at University of Iowa to illustrate our work with two diseases. 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 his seminal approach, 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). Pooling specimens through group testing has also been identified as a useful strategy to increase testing loads for influenza (Van et al., 2012), including the most recent SARS CoV-2 (Abdalhamid et al., 2020; Hogan et al., 2020; Pilcher, Westreich, and Hudgens, 2020).

*Corresponding author: e-mail: msarker@radford.edu, Phone: +1 540-831-5026, Fax: +1 540-831-6452

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 (CT) and gonorrhea (NG), two of the most common sexually transmitted diseases in the United States and elsewhere. As part of a national program formerly known as the Infertility Prevention Project in the United States, state-run laboratories screen millions of individuals each year for CT/NG using multiplex assays, and some of these laboratories use group testing to reduce costs and/or to increase testing capacity (Lewis, Lockary, and Kobic, 2012; Tebbs, McMahan, and Bilder, 2013). Recent advances in technology have seen the development of multiplex assays which test for more than two diseases at once. 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), and, at the time of this writing, there is little doubt multiplex assays for influenza viruses which include SARS CoV-2 will soon be available for widespread use.

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 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, motivated by screening practices for CT/NG in Iowa and elsewhere. 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 estimation 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 biostatistics 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 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 University of Iowa uses the Aptima Combo 2 Assay (AC2A, Hologic, Marlborough, USA), a mutliplex assay which uses nucleic acid amplification to detect CT and NG simultaneously (Gaydos et al., 2003). 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 on 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 (Appendix E). 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; see Tebbs, McMahan, and Bilder (2013). 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 disease 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, $\widetilde{\mathbf{V}}_i|\widetilde{\mathbf{Y}}_{-i}, \mathbf{p}, \mathbf{Z}, \boldsymbol{\delta} \sim \text{multinomial}(p^*_{(00)i}, p^*_{(10)i}, p^*_{(01)i}, p^*_{(11)i})$, 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 $\widetilde{\mathbf{Y}}_i^{(0)} = (\widetilde{Y}_{i1}^{(0)}, \widetilde{Y}_{i2}^{(0)})'$, for $i = 1, 2, ..., N$. Initialize $\mathbf{p}^{(0)} = (p_{00}^{(0)}, p_{10}^{(0)}, p_{01}^{(0)}, p_{11}^{(0)})'$, where $\sum_{u=0}^1 \sum_{v=0}^1 p_{uv}^{(0)} = 1$. 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

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

   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. 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 $\pi_1 = p_{10} + p_{11}$ and $\pi_2 = p_{01} + p_{11}$ 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 $Q(\mathbf{p},\mathbf{p}^{(d)})$, 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 $\mathbf{p}^{(d)}$, 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 $\widetilde{\mathbf{Y}}_i^{(0)} = (\widetilde{Y}_{i1}^{(0)}, \widetilde{Y}_{i2}^{(0)})'$, for $i = 1, 2, ..., N$. Initialize $\mathbf{p}^{(0)} = (p_{00}^{(0)}, p_{10}^{(0)}, p_{01}^{(0)}, p_{11}^{(0)})'$, where $\sum_{u=0}^{1}\sum_{v=0}^{1} p_{uv}^{(0)} = 1$. Set $g = 1$. Set $d = 0$.
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 $\widetilde{\mathbf{V}}_i|\widetilde{\mathbf{Y}}_{-i}^{(g)}, \mathbf{p}^{(d)}, \mathbf{Z}, \boldsymbol{\delta} \sim$ 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 $E[(\widetilde{V}_{(00)i}, \widetilde{V}_{(10)i}, \widetilde{V}_{(01)i}, \widetilde{V}_{(11)i})'|\mathbf{Z}, \boldsymbol{\delta}; \mathbf{p}^{(d)}]$.
3. (M-Step): Calculate $\mathbf{p}^{(d+1)}$ using the solution in Appendix B in the Supporting Information; i.e., this maximizer depends on $E[(\widetilde{V}_{(00)i}, \widetilde{V}_{(10)i}, \widetilde{V}_{(01)i}, \widetilde{V}_{(11)i})'|\mathbf{Z}, \boldsymbol{\delta}; \mathbf{p}^{(d)}]$ and exists in closed form.
4. Set $g = g+1$, $d = d+1$, and repeat steps 2-4 until the maximum absolute difference in $\mathbf{p}^{(d+1)} - \mathbf{p}^{(d)}$ 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 $E[(\widetilde{V}_{(00)i}, \widetilde{V}_{(10)i}, \widetilde{V}_{(01)i}, \widetilde{V}_{(11)i})'|\mathbf{Z}, \boldsymbol{\delta}; \mathbf{p}^{(d)}]$, 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 disease 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 to incorporate the uncertainty in $\boldsymbol{\delta}$. Such an extension is potentially useful. 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, $S_{e:(l)k} \sim \text{beta}(a_{S_{e:(l)k}}, b_{S_{e:(l)k}})$ and $S_{p:(l)k} \sim \text{beta}(a_{S_{p:(l)k}}, b_{S_{p:(l)k}})$, 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 $S_{p:(l)k}$; see Section 6. 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 $S_{e:(l)k}|\mathbf{Z}, \widetilde{\mathbf{Y}} \sim \text{beta}(a^*_{S_{e:(l)k}}, b^*_{S_{e:(l)k}})$ and $S_{p:(l)k}|\mathbf{Z}, \widetilde{\mathbf{Y}} \sim \text{beta}(a^*_{S_{p:(l)k}}, b^*_{S_{p:(l)k}})$, where $a_{S^*_{e:(l)k}} = a_{S_{e:(l)k}} + \sum_{j \in \mathcal{M}(l)} Z_{jk}\widetilde{Z}_{jk}$, $b_{S^*_{e:(l)k}} = b_{S_{e:(l)k}} + \sum_{j \in \mathcal{M}(l)} (1 - Z_{jk})\widetilde{Z}_{jk}$, $a_{S^*_{p:(l)k}} = a_{S_{p:(l)k}} + \sum_{j \in \mathcal{M}(l)} (1 - Z_{jk})(1 - \widetilde{Z}_{jk})$, and $b_{S^*_{p:(l)k}} = b_{S_{p:(l)k}} + \sum_{j \in \mathcal{M}(l)} Z_{jk}(1 - \widetilde{Z}_{jk})$. 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 $Q(\mathbf{p}, \boldsymbol{\delta}, \mathbf{p}^{(d)}, \boldsymbol{\delta}^{(d)})$, 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 emulate our CT/NG data application in Section 6, 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. Finding these 'optimal designs in Table 1 was described in Hou et al. (2017) for hierarchical protocols and Hou et al. (2020) for AT.

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

Table 2 shows the results for estimating $\mathbf{p}$ under 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. Under the configurations, Table 3 shows the estimates of $\boldsymbol{\delta}$. That is, the results in Tables 2 and 3 are obtained from the same posterior distribution but split into two tables. When $\boldsymbol{\delta}$ is known (Section 3), which arises as a special case, a table consisting of the estimates of $\mathbf{p}$ is shown 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, $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4)$, $S_{e:(1)k} \sim \text{beta}(1, 1)$, and $S_{p:(1)k} \sim \text{beta}(1, 1)$, for $k = 1, 2$. 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 when $\boldsymbol{\delta}$ is known and should be approximately equal when $\boldsymbol{\delta}$ is unknown.

[Table 2 here.]

In Tables 2 and 3, we show the sample mean ("Est") and the sample standard deviation (SD) of the posterior mean (Mean) and MAP estimates calculated from $B = 500$ independent data sets, along with the averaged estimated posterior standard deviation (SE) for each parameter in $\mathbf{p}$. 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. Under each of the 500 data sets, the mean estimates are calculated from a posterior sample of 2000 realizations, obtained by retaining every 5th of the $G = 10000$ Gibbs iterates from the posterior distribution. For MAP estimation, $G = 10000$ Gibbs iterates are sampled from the multinomial multinomial$(p^*_{(00)i}, p^*_{(10)i}, p^*_{(01)i}, p^*_{(11)i})$ in the E-Step (Step 2) to approximate $E[(\widetilde{V}_{(00)i}, \widetilde{V}_{(10)i}, \widetilde{V}_{(01)i}, \widetilde{V}_{(11)i})'|\mathbf{Z}, \boldsymbol{\delta}; \mathbf{p}^{(d)}]$ without thinning. In both estimation scenarios, 2000 initial iterates are discarded as a burn-in period before using the $G = 10000$ iterates.

[Table 3 here.]

As demonstrated in Tables 2-3, both the mean and MAP estimates of $\mathbf{p}$ are accurate under each prevalence configuration. The posterior standard deviation is also estimated well (SE/SD is close to 1), and the estimation efficiency (SE) does not degrade even though the higher-stage protocols (H3, H4, and AT) expend substantially fewer tests. For example, the H3 protocol, when compared to H2, estimates the parameters with 36% fewer tests for Configuration II, without losing accuracy or efficiency in estimation. For the assay accuracy probabilities, estimation by using the higher-stage protocols is even more fruitful. Now, the accuracy is improved and a considerable amount of efficiency gain is realized (40-60% decrease in SE under Configuration II). One of the aims in this simulation is to discover how the MAP procedure performs, as this has not been studied before in the literature for multiple-infection problems. Interestingly, the MAP estimates appear to be more accurate in all cases of the simulation. This is not surprising because the posterior distribution in such low-prevalence settings is skewed, making the mode a better candidate than the mean (or median) for point estimation.

To examine the performance of the estimation techniques in more challenging situations, we repeated the above simulation with two assays, $L = 2$, in which the pooled specimens were tested by an assay (e.g., screening assay) and the individual specimens in the final stage were tested by another assay (e.g., confirmatory assay). The models were again fit using flat priors when the assay accuracy probabilities are unknown. Overall, the pattern in estimation accuracy and efficiency that one observes with the $L = 1$ case is retained, although four extra parameters in $\boldsymbol{\delta}$ are now involved. When the prevalence rates are smaller, we expect that the advantages of using a higher-stage protocol will be even more pronounced.

6 Iowa chlamydia and gonorrhea data application

Chlamydia and gonorrhea (CT/NG) are common sexually transmitted diseases, caused by bacteria and can lead to serious health-related complications to patients, such as pelvic inflammatory disease, infertility, and ectopic pregnancy, if not treated in a timely manner. To prevent the transmission of CT/NG, the Centers for Disease Control and Prevention (CDC) conducts nationwide surveillance and screening for these infections in all 50 states. In Iowa, the state residents are tested for CT/NG at the State Hygienic Laboratory (SHL) at the University of Iowa. For efficient screening, the SHL currently uses the two-stage hierarchical protocol (H2). Hou et al. (2020) provided a comparative analysis of the screening accuracy and efficiency of the H2, H3, and AT protocols for the Iowa population. Their investigation shows that the H2 protocol, when

compared to individual testing, provides a great reduction in testing cost. They also found that even a further substantial reduction can be realized when H2 is extended to H3 and AT. In this section, the aim is to illustrate our estimation techniques using CT/NG data and to examine how a higher-stage protocol performs in estimating the prevalence of CT/NG at once.

We have a data set that consists of CT/NG diagnostic outcomes from 14450 females tested at the SHL in 2014. The diagnoses were obtained from 4402 urine specimens and 10048 swab specimens of the females. For research and comparison purposes, the SHL used both individual testing and the H2 pooling; i.e., (a) individual testing on all specimens, and (b) the H2 protocol where pooled testing in stage 1 and individual testing in stage 2. For all of the testing, the Aptima Combo 2 Assay was used. Though both data sets are available to us, we only use the individual testing data set so that a fair comparison can be made by artificially creating data for different pooling strategies. To this end, we first cross-classify the individuals into urine and swab strata, because the Aptima Combo 2 Assay accuracies are different for these strata. Doing this leads us to have two data sets of individual CT/NG outcomes (one for urine and the other one for swab), which are used to generate pooling data based on the H2, H3, and AT protocols. Since the CT/NG prevalence in our data set is not very small, we do not use H4 or other higher-stage protocols.

[Table 4 here.]

To generate group testing data, we first treat the individual test outcomes as the true statuses of the individuals. Using these true statuses and the Aptima Combo 2 Assay accuracies, we calculate the pooling configurations that minimize the expected number of tests. The optimal configurations and the Aptima Combo 2 Assay accuracies (obtained from the assay product literature) are reported in Table 4. Next, we assign the individuals to the initial pools and then simulate group testing data, similarly as described in Section 5. To be consistent with the SHL testing policy, we use one diagnostic assay ($L = 1$) for testing in all stages. To average out the Monte Carlo errors, we generate 500 data sets by repeating the above steps under each specimen stratum. The mean and MAP estimates are calculated from the data sets under the assumption that the assay accuracy probabilities are unknown. The model fitting and estimation is executed again with flat priors on both $\mathbf{p}$ and $\boldsymbol{\delta}$. We recorded the computing time for one data set in an Intel 3.6GHz 32GB Memory machine. The mean estimation takes about 46 and 89 seconds for the urine and swab strata, respectively, while the MAP estimation takes about 94 and 439 seconds, with $G = 10000$ Gibbs iterates after discarding the initial 2000 iterates.

[Table 5 here.]

Table 5 shows the estimates of $\mathbf{p} = (p_{00}, p_{10}, p_{01}, p_{11})'$, where p_{00} is the proportion of females negative for both chlamydia and gonorrhea, p_{10} is the proportion of females positive for chlamydia but negative for gonorrhea, and so on. Table 6 depicts the estimates of $\boldsymbol{\delta} = (S_{e:(1)1}, S_{e:(1)2}, S_{p:(1)1}, S_{p:(1)2})'$, where $S_{e:(1)1}$ and $S_{e:(1)2}$ are the sensitivities of chlamydia and gonorrhea, respectively, and $S_{p:(1)1}$ and $S_{p:(1)2}$ are specificities of these infections. The overall finding observed in the simulation study (Section 5) is revealed again with the CT/NG data; i.e., the accuracy probability estimates are somewhat improved (smaller SE), but the prevalence estimates are virtually unaffected by the pooling protocols. Again, the primary gain from a higher-stage method is enjoyed from the reduction in testing cost. For example, when compared to H2, the H3 protocol for the CT/NG screening requires 6.3% fewer tests in the urine stratum and 6.9% fewer tests in the swab stratum; i.e., the H3 protocol uses approximately 560 fewer tests in total (approximately $20,702 reduction at $37 cost per test) but still produces (nearly) identical estimates. Because CT/NG screening is also conducted in other states, the monetary benefits by adopting a three-stage hierarchical (or AT) procedure likely increase by a substantial amount.

[Table 6 here.]

7 Discussion

We have developed a general Bayesian framework to model group testing data involving multiplex assays. The work, motivated by the CT/NG testing practices in the United States, generalizes the two-stage Bayesian approach in Warasi et al. (2016). With the special case of flat priors, the MAP estimation is equivalent to the frequentist's likelihood-based estimation in Tebbs, McMahan, and Bilder (2013). Overall, the estimates produced by our work are accurate. A higher-stage hierarchical or array protocol confers similar precision in the prevalence estimates but better precision in the assay accuracy probability estimates, all of which is achieved with a substantially smaller testing cost in low-incidence populations. 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 cost-effective strategy for the detection and monitoring of CT/NG and other infectious diseases. To support practitioners implementing our estimation techniques, we present a user-friendly R function in Appendix F in the Supporting Information. The R programs with illustrative examples are provided at `https://github.com/mswarasi`.

This article requires that the multiplex assays that can detect multiple infections simultaneously from each specimen testing are used for diagnosis. With the recent development in assay technologies, this is not a strong requirement. There are many multiplex assays currently available and approved by the U.S. Food and Drug Administration. While focused on the CT/NG application in Iowa with the Aptima Combo 2 Assay, this work can be implemented elsewhere whenever a multiplex discriminatory assay is available. For example, our estimation techniques can be used in the HIV and HCV surveillance applications with the Procleix HIV/HCV Assay. To address the situation where more than two infections are involved, we present a general-version of our work with $K \geq 2$ infections in Appendix E in the Supporting Information. The extension can be especially useful to the American Red Cross and other organizations that perform blood testing for HIV, HBV, and HCV.

We have assumed that there is no pooled dilution effect; i.e., false negative and false positive test outcomes do not occur due to pooling. This assumption is reasonable when the pool sizes are not too large for the assay used in testing. Another aspect in our modeling that needs to be carefully handled is the use of multiple diagnostic assays, $L \geq 1$, (e.g., screening assay and confirmatory assay). Unfortunately, when L is large, the number of unknown assay accuracy parameters, $S_{e:(l)k}$ and $S_{p:(l)k}$, to be estimated along with the prevalence parameter $\mathbf{p}$ grows rapidly. In such instances, concentrated prior distributions may be needed or/and the sample size, N, should be large to have the estimation techniques working properly. This difficulty would be alleviated if individual covariate information (e.g., age, symptoms, number of partners, etc.) were incorporated; i.e., a regression approach for covariate-adjusted estimation. Because covariate information is recorded in public health applications, in addition to testing, the future research that explores this extension should be of high practical value.

Acknowledgements We thank Jeffrey Benfer and Kristofer Eveland at the State Hygienic Laboratory at the University of Iowa. We also thank Elizabeth Torrone at the Centers for Disease Control and Prevention (Atlanta, GA, USA) for her insight on CT/NG testing practices in the United States. This research was funded by Grant R01 AI121351 from the National Institutes of Health.

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.

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.

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.

Gaydos, C., Quinn, T., Willis, D., Weissfeld, A., Hook, E., Martin, D., Ferrero, D., and Schachter, J. (2003). Performance of the APTIMA Combo 2 Assay for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in female urine and endocervical swab specimens. *Journal of Clinical Microbiology* **41**, 304–309.

Hogan, C., Sahoo, M., and Pinsky, B. (2020). Sample pooling as a strategy to detect community transmission of SARS-CoV-2. *Journal of the American Medical Association* **323**, 1967–1969.

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.

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.

Lewis, J., Lockary, V., and Kobic S. (2012). Cost savings and increased efficiency using a stratified specimen pooling strategy for *Chlamydia trachomatis* and *Neisseria gonorrhoeae*. *Sexually Transmitted Diseases* **39**, 46–48.

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.

Pilcher, C., Westreich, D., and Hudgens M. (2020). Group testing for SARS-CoV-2 to enable rapid scale-up of testing and real-time surveillance of incidence. *Journal of Infectious Diseases* **222**, 903–909.

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.

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.

Van, T., Miller, J., Warshauer, D., Reisdorf, E., Jernigan, D., Humes, R., and Shult, P. (2012). Pooling nasopharyngeal/throat swab specimens to increase testing capacity for influenza viruses by PCR. *Journal of Clinical Virology* **50**, 891–896.

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 study. 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. Values of SD and SE presented below have been multiplied by 10^2. 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	0.35	0.39	0.949	0.33	0.34	0.949	0.34	0.33	0.950	0.34	0.34
			0.021	0.24	0.26	0.020	0.22	0.22	0.020	0.21	0.22	0.020	0.23	0.22
			0.021	0.25	0.26	0.020	0.21	0.22	0.020	0.23	0.22	0.020	0.22	0.22
			0.010	0.14	0.15	0.010	0.15	0.15	0.010	0.14	0.15	0.010	0.15	0.15
		MAP	0.950	0.35	0.39	0.950	0.32	0.34	0.950	0.34	0.33	0.950	0.33	0.34
			0.020	0.24	0.26	0.020	0.22	0.22	0.020	0.21	0.22	0.020	0.22	0.22
			0.020	0.24	0.26	0.020	0.21	0.22	0.020	0.22	0.22	0.020	0.22	0.22
			0.010	0.14	0.15	0.010	0.15	0.15	0.010	0.14	0.15	0.010	0.15	0.15
	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	0.18	0.21	0.989	0.16	0.16	0.989	0.15	0.16	0.989	0.14	0.16
			0.005	0.12	0.13	0.004	0.10	0.11	0.004	0.10	0.10	0.004	0.09	0.10
			0.005	0.12	0.13	0.004	0.10	0.11	0.004	0.10	0.10	0.004	0.09	0.10
			0.003	0.07	0.08	0.002	0.06	0.07	0.002	0.06	0.07	0.002	0.06	0.07
		MAP	0.990	0.18	0.21	0.990	0.16	0.16	0.990	0.15	0.16	0.990	0.15	0.16
			0.004	0.11	0.13	0.004	0.10	0.11	0.004	0.10	0.10	0.004	0.10	0.10
			0.004	0.12	0.13	0.004	0.10	0.11	0.004	0.10	0.10	0.004	0.09	0.10
			0.002	0.07	0.08	0.002	0.06	0.07	0.002	0.06	0.07	0.002	0.06	0.07
	Avg. number of tests		1054.3			671.0 (36.4%)			581.4 (44.9%)			751.4 (28.7%)		

Table 3 Simulation study. 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. Values of SD and SE presented below have been multiplied by 10^1. 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	0.23	0.27	0.946	0.15	0.15	0.947	0.12	0.12	0.947	0.14	0.15
			0.942	0.25	0.27	0.946	0.14	0.15	0.948	0.11	0.12	0.948	0.15	0.15
			0.990	0.04	0.04	0.989	0.03	0.04	0.989	0.03	0.04	0.989	0.04	0.05
			0.990	0.04	0.04	0.989	0.03	0.04	0.990	0.04	0.04	0.988	0.05	0.05
		MAP	0.950	0.23	0.27	0.950	0.14	0.15	0.950	0.12	0.12	0.950	0.14	0.15
			0.951	0.25	0.27	0.950	0.14	0.15	0.950	0.11	0.12	0.951	0.14	0.15
			0.990	0.04	0.04	0.990	0.03	0.04	0.990	0.03	0.04	0.990	0.04	0.05
			0.990	0.04	0.04	0.990	0.03	0.04	0.990	0.03	0.04	0.990	0.04	0.05
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	0.57	0.67	0.927	0.34	0.37	0.937	0.26	0.28	0.931	0.31	0.37
			0.889	0.55	0.67	0.928	0.33	0.37	0.938	0.26	0.28	0.933	0.31	0.36
			0.990	0.04	0.05	0.989	0.05	0.05	0.988	0.05	0.06	0.988	0.05	0.05
			0.989	0.04	0.05	0.988	0.04	0.05	0.988	0.05	0.06	0.988	0.05	0.06
		MAP	0.941	0.58	0.67	0.947	0.37	0.37	0.950	0.26	0.28	0.949	0.34	0.37
			0.940	0.58	0.67	0.948	0.36	0.37	0.950	0.27	0.28	0.951	0.34	0.36
			0.990	0.04	0.05	0.990	0.05	0.05	0.990	0.05	0.06	0.990	0.06	0.05
			0.990	0.04	0.05	0.990	0.05	0.05	0.990	0.05	0.06	0.990	0.06	0.06

Table 4 Iowa CT/NG data application. 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. The Aptima Combo 2 Assay accuracy probabilities are also reported.

	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 Iowa CT/NG data application. 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. Values of SE presented below are the same for both Mean and MAP, and have been multiplied by 10^2. 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 in each stratum, N, is reported.

				H2		H3		AT	
Stratum		CT	NG	Est	SE	Est	SE	Est	SE
Urine $N = 4402$	Mean	$-$	$-$	$\widehat{p}_{00} = 0.907$	0.65	$\widehat{p}_{00} = 0.908$	0.55	$\widehat{p}_{00} = 0.909$	0.54
		$+$	$-$	$\widehat{p}_{10} = 0.081$	0.62	$\widehat{p}_{10} = 0.081$	0.53	$\widehat{p}_{10} = 0.080$	0.51
		$-$	$+$	$\widehat{p}_{01} = 0.007$	0.18	$\widehat{p}_{01} = 0.006$	0.13	$\widehat{p}_{01} = 0.006$	0.13
		$+$	$+$	$\widehat{p}_{11} = 0.006$	0.13	$\widehat{p}_{11} = 0.005$	0.12	$\widehat{p}_{11} = 0.005$	0.11
	MAP	$-$	$-$	$\widehat{p}_{00} = 0.908$	0.65	$\widehat{p}_{00} = 0.908$	0.55	$\widehat{p}_{00} = 0.908$	0.54
		$+$	$-$	$\widehat{p}_{10} = 0.081$	0.62	$\widehat{p}_{10} = 0.081$	0.53	$\widehat{p}_{10} = 0.081$	0.51
		$-$	$+$	$\widehat{p}_{01} = 0.006$	0.18	$\widehat{p}_{01} = 0.006$	0.13	$\widehat{p}_{01} = 0.006$	0.13
		$+$	$+$	$\widehat{p}_{11} = 0.005$	0.13	$\widehat{p}_{11} = 0.005$	0.12	$\widehat{p}_{11} = 0.005$	0.11
	Avg. number of tests			2489.6		2332.9 (6.3%)		2333.3 (6.3%)	
Swab $N = 10048$	Mean	$-$	$-$	$\widehat{p}_{00} = 0.907$	0.52	$\widehat{p}_{00} = 0.908$	0.39	$\widehat{p}_{00} = 0.908$	0.38
		$+$	$-$	$\widehat{p}_{10} = 0.081$	0.51	$\widehat{p}_{10} = 0.082$	0.38	$\widehat{p}_{10} = 0.081$	0.37
		$-$	$+$	$\widehat{p}_{01} = 0.006$	0.11	$\widehat{p}_{01} = 0.006$	0.08	$\widehat{p}_{01} = 0.006$	0.08
		$+$	$+$	$\widehat{p}_{11} = 0.005$	0.09	$\widehat{p}_{11} = 0.005$	0.07	$\widehat{p}_{11} = 0.005$	0.07
	MAP	$-$	$-$	$\widehat{p}_{00} = 0.908$	0.52	$\widehat{p}_{00} = 0.909$	0.39	$\widehat{p}_{00} = 0.909$	0.38
		$+$	$-$	$\widehat{p}_{10} = 0.081$	0.51	$\widehat{p}_{10} = 0.081$	0.38	$\widehat{p}_{10} = 0.081$	0.37
		$-$	$+$	$\widehat{p}_{01} = 0.006$	0.11	$\widehat{p}_{01} = 0.005$	0.08	$\widehat{p}_{01} = 0.005$	0.08
		$+$	$+$	$\widehat{p}_{11} = 0.005$	0.09	$\widehat{p}_{11} = 0.005$	0.07	$\widehat{p}_{11} = 0.005$	0.07
	Avg. number of tests			$\overline{T} = 5802.8$		$\overline{T} = 5400.0(6.9)$		$\overline{T} = 5354.7(7.7)$	

Table 6 Iowa CT/NG data application. 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, and "SE" is the estimated posterior standard deviation. Values of SE presented below are the same for both Mean and MAP, and have been multiplied by 10^1. 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 in each stratum, N, is reported.

			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	$\widehat{S}_{e:(1)1} = 0.948$	0.22	$\widehat{S}_{e:(1)1} = 0.946$	0.12	$\widehat{S}_{e:(1)1} = 0.948$	0.11
			$\widehat{S}_{e:(1)2} = 0.875$	0.66	$\widehat{S}_{e:(1)2} = 0.899$	0.34	$\widehat{S}_{e:(1)2} = 0.904$	0.32
			$\widehat{S}_{p:(1)1} = 0.987$	0.07	$\widehat{S}_{p:(1)1} = 0.987$	0.06	$\widehat{S}_{p:(1)1} = 0.985$	0.08
			$\widehat{S}_{p:(1)2} = 0.993$	0.03	$\widehat{S}_{p:(1)2} = 0.993$	0.02	$\widehat{S}_{p:(1)2} = 0.993$	0.02
		MAP	$\widehat{S}_{e:(1)1} = 0.948$	0.22	$\widehat{S}_{e:(1)1} = 0.947$	0.12	$\widehat{S}_{e:(1)1} = 0.948$	0.11
			$\widehat{S}_{e:(1)2} = 0.911$	0.66	$\widehat{S}_{e:(1)2} = 0.911$	0.34	$\widehat{S}_{e:(1)2} = 0.914$	0.32
			$\widehat{S}_{p:(1)1} = 0.989$	0.07	$\widehat{S}_{p:(1)1} = 0.989$	0.06	$\widehat{S}_{p:(1)1} = 0.989$	0.08
			$\widehat{S}_{p:(1)2} = 0.993$	0.03	$\widehat{S}_{p:(1)2} = 0.993$	0.02	$\widehat{S}_{p:(1)2} = 0.993$	0.02
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	$\widehat{S}_{e:(1)1} = 0.940$	0.19	$\widehat{S}_{e:(1)1} = 0.941$	0.09	$\widehat{S}_{e:(1)1} = 0.941$	0.09
			$\widehat{S}_{e:(1)2} = 0.936$	0.40	$\widehat{S}_{e:(1)2} = 0.985$	0.09	$\widehat{S}_{e:(1)2} = 0.986$	0.09
			$\widehat{S}_{p:(1)1} = 0.976$	0.06	$\widehat{S}_{p:(1)1} = 0.976$	0.05	$\widehat{S}_{p:(1)1} = 0.975$	0.07
			$\widehat{S}_{p:(1)2} = 0.988$	0.02	$\widehat{S}_{p:(1)2} = 0.987$	0.02	$\widehat{S}_{p:(1)2} = 0.987$	0.02
		MAP	$\widehat{S}_{e:(1)1} = 0.942$	0.19	$\widehat{S}_{e:(1)1} = 0.942$	0.09	$\widehat{S}_{e:(1)1} = 0.942$	0.09
			$\widehat{S}_{e:(1)2} = 0.984$	0.40	$\widehat{S}_{e:(1)2} = 0.991$	0.09	$\widehat{S}_{e:(1)2} = 0.991$	0.09
			$\widehat{S}_{p:(1)1} = 0.976$	0.06	$\widehat{S}_{p:(1)1} = 0.976$	0.05	$\widehat{S}_{p:(1)1} = 0.976$	0.07
			$\widehat{S}_{p:(1)2} = 0.987$	0.02	$\widehat{S}_{p:(1)2} = 0.987$	0.02	$\widehat{S}_{p:(1)2} = 0.987$	0.02

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*: msarker@radford.edu

Web Appendix A. *Multinomial distribution in Section 3.*

The full conditional of $\widetilde{\mathbf{V}}_i$ as discussed in Section 3 is multinomial$(p^*_{(00)i}, p^*_{(10)i}, p^*_{(01)i}, p^*_{(11)i})$, where the cell probabilities $p^*_{(uv)i} = \zeta^i_{uv}/\zeta^i_+$, for $u, v \in \{0, 1\}$, are given by

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

$\zeta^i_+ = \sum_{u=0}^{1} \sum_{v=0}^{1} \zeta^i_{uv}$, $\overline{S}_{e:jk} = 1 - S_{e:jk}$, $\overline{S}_{p:jk} = 1 - S_{p:jk}$, $\mathcal{A}_i = \{j : i \in \mathcal{P}_j\}$, $\mathcal{P}_{j(i)} = \mathcal{P}_j \setminus \{i\}$, 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 in Section 3.2.*

The MAP estimation for $\mathbf{p}$ with known $\boldsymbol{\delta}$ is discussed in Section 3.2. The solution, $\mathbf{p}^{(d+1)}$, which is used in the M-Step is given by

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

where $\widetilde{N} = N - 4 + \sum_{u=0}^{1}\sum_{u=0}^{1} \alpha_{uv}$, $\widetilde{V}_{(00)i} = (1 - \widetilde{Y}_{i1})(1 - \widetilde{Y}_{i2})$, $\widetilde{V}_{(10)i} = \widetilde{Y}_{i1}(1 - \widetilde{Y}_{i2})$, $\widetilde{V}_{(01)i} = (1 - \widetilde{Y}_{i1})\widetilde{Y}_{i2}$, and $\widetilde{V}_{(11)i} = \widetilde{Y}_{i1}\widetilde{Y}_{i2}$. As one can see, setting $\alpha_{00} = \alpha_{10} = \alpha_{01} = \alpha_{11} = 1$ (i.e., for flat Dirichlet prior) reduces $\mathbf{p}^{(d+1)}$ to the maximum likelihood estimate (Tebbs et al., 2013).

Web Appendix C. *Algorithms for the estimation techniques in Section 4.*

In Section 4, we briefly described how posterior sampling and MAP estimation can be performed when both $\mathbf{p}$ and $\boldsymbol{\delta}$ are unknown. For this estimation scenario, we show the sampling algorithm and the EM algorithm below by generalizing the work presented in Section 3.

POSTERIOR SAMPLING ALGORITHM

1. Initialize $\widetilde{\mathbf{Y}}_i^{(0)} = (\widetilde{Y}_{i1}^{(0)}, \widetilde{Y}_{i2}^{(0)})'$, $\mathbf{p}^{(0)} = (p_{00}^{(0)}, p_{10}^{(0)}, p_{01}^{(0)}, p_{11}^{(0)})'$, $S_{e:(l)k}^{(0)}$, and $S_{p:(l)k}^{(0)}$, for $i = 1, 2, ..., N$, $l = 1, 2, ..., L$, $k = 1, 2$. Accumulate $S_{e:(l)k}^{(0)}$'s and $S_{p:(l)k}^{(0)}$'s 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 $S_{e:(l)k}|\mathbf{Z}, \widetilde{\mathbf{Y}}^{(g)}$ and $S_{p:(l)k}^{(g)}$ from $S_{p:(l)k}|\mathbf{Z}, \widetilde{\mathbf{Y}}^{(g)}$, for $l = 1, 2, ..., L$ and $k = 1, 2$. Accumulate $S_{e:(l)k}^{(g)}$'s and $S_{p:(l)k}^{(g)}$'s into $\boldsymbol{\delta}^{(g)}$.

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

Before describing the EM algorithm, we provide the solution $(\mathbf{p}'^{(d+1)}, \boldsymbol{\delta}'^{(d+1)})'$, which is found by maximizing $Q(\mathbf{p}, \boldsymbol{\delta}, \mathbf{p}^{(d)}, \boldsymbol{\delta}^{(d)})$ at the current estimate $(\mathbf{p}'^{(d)}, \boldsymbol{\delta}'^{(d)})'$. Note that $\mathbf{p}^{(d+1)} = (p_{00}^{(d+1)}, p_{10}^{(d+1)}, p_{01}^{(d+1)}, p_{11}^{(d+1)})'$ remains the same as in Web Appendix B where $\boldsymbol{\delta} = \boldsymbol{\delta}^{(d)}$, and $\boldsymbol{\delta}^{(d+1)}$ consists of

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

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

for $l = 1, 2, ..., L$ and $k = 1, 2$, where $\widetilde{J}_{S_{e:(l)k}} = a_{S_{e:(l)k}} + b_{S_{e:(l)k}} - 2 + \sum_{j \in \mathcal{M}(l)} E(\widetilde{Z}_{jk} | \mathbf{Z}; \boldsymbol{\delta}^{(d)}, \mathbf{p}^{(d)})$ and $\widetilde{J}_{S_{p:(l)k}} = a_{S_{p:(l)k}} + b_{S_{p:(l)k}} - 2 + \sum_{j \in \mathcal{M}(l)} \{1 - E(\widetilde{Z}_{jk} | \mathbf{Z}; \boldsymbol{\delta}^{(d)}, \mathbf{p}^{(d)})\}$. The expectation is approximated empirically, as shown in the following E-Step.

MAP ESTIMATION VIA EM ALGORITHM

1. Initialize $\widetilde{\mathbf{Y}}_i^{(0)} = (\widetilde{Y}_{i1}^{(0)}, \widetilde{Y}_{i2}^{(0)})'$, $\mathbf{p}^{(0)} = (p_{00}^{(0)}, p_{10}^{(0)}, p_{01}^{(0)}, p_{11}^{(0)})'$, $S_{e:(l)k}^{(0)}$, and $S_{p:(l)k}^{(0)}$, for $i = 1, 2, ..., N$, $l = 1, 2, ..., L$, $k = 1, 2$. Accumulate $S_{e:(l)k}^{(0)}$'s and $S_{p:(l)k}^{(0)}$'s into $\boldsymbol{\delta}^{(0)}$. Set $g = 1$. Set $d = 0$.

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 $\widetilde{\mathbf{V}}_i | \widetilde{\mathbf{Y}}_{-i}^{(g)}, \mathbf{p}^{(d)}, \boldsymbol{\delta}^{(d)}, \mathbf{Z} \sim$ 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 $E[(\widetilde{V}_{(00)i}, \widetilde{V}_{(10)i}, \widetilde{V}_{(01)i}, \widetilde{V}_{(11)i})' | \mathbf{Z}; \boldsymbol{\delta}^{(d)}, \mathbf{p}^{(d)}]$;
   - calculate the sample mean $G^{-1} \sum_{g=1}^{G} \widetilde{Z}_{jk}^{(g)}$ as an estimate of $E(\widetilde{Z}_{jk} | \mathbf{Z}; \boldsymbol{\delta}^{(d)}, \mathbf{p}^{(d)})$, 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 $\mathbf{p}^{(d)}$ and $\boldsymbol{\delta}^{(d)}$ using the solution above.

4. Set $g = g + 1$, $d = d + 1$, and repeat steps 2-4 until the maximum absolute difference in $(\mathbf{p}'^{(d+1)}, \boldsymbol{\delta}'^{(d+1)})' - (\mathbf{p}'^{(d)}, \boldsymbol{\delta}'^{(d)})'$ is less than ϵ, where ϵ is small.

Web Appendix D. *Additional simulation results from Section 3.*

In Tables 2 and 3 of the article, we presented simulation estimates of $\mathbf{p}$ and $\boldsymbol{\delta}$ with unknown accuracy probabilities. We herein show the rest of the simulation evidence.

- Table D.1: **Known** accuracy probabilities. **One assay**, $L = 1$, is used for both pooled and individual testing.
  - Estimates of $\mathbf{p}$ for Configurations I and II.
- Table D.2: The pool sizes used for the simulations in Tables D.3-D.5.
- Tables D.3-D.5: **Unknown** accuracy probabilities. **Two assays**, $L = 2$, are used; one assay is for pooled testing, and the other one is for individual testing in the final stage.
  - Table D.3: Estimates of $\mathbf{p}$ for Configurations I and II.
  - Table D.4: Estimates of $\boldsymbol{\delta}$ for Configuration I.
  - Table D.5: Estimates of $\boldsymbol{\delta}$ for Configuration II.

Table D.1: Simulation study. Characteristics of posterior mean and MAP estimates when assay accuracy probabilities are **known**. Estimates ("Est") are averages over $B = 500$ Monte Carlo data sets, "SD" denotes the sample standard deviation of the 500 estimates, and "SE" is the estimated posterior standard deviation. Note that the Dirichlet flat prior has been used for $\mathbf{p}$; i.e., $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4)$. In addition, values of SD and SE presented below have been multiplied by 10^2. The average number of tests is also shown; the percent reduction (in parentheses) is compared to the two-stage hierarchical protocol H2 discussed 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. The values used for the assay accuracies are $S_{e:(1)k} = 0.95$ and $S_{p:(1)k} = 0.99$, for $k = 1, 2$.

			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	0.33	0.32	0.949	0.32	0.31	0.949	0.33	0.33	0.949	0.32	0.32
			0.020	0.21	0.21	0.020	0.21	0.21	0.020	0.21	0.21	0.020	0.21	0.22
			0.020	0.21	0.20	0.020	0.21	0.21	0.020	0.21	0.22	0.020	0.21	0.22
			0.010	0.15	0.14	0.010	0.14	0.14	0.010	0.14	0.14	0.010	0.15	0.15
		MAP	0.950	0.33	0.31	0.950	0.32	0.32	0.950	0.33	0.33	0.950	0.32	0.31
			0.020	0.21	0.21	0.020	0.21	0.21	0.020	0.21	0.20	0.020	0.21	0.21
			0.020	0.21	0.21	0.020	0.21	0.20	0.020	0.21	0.22	0.020	0.21	0.21
			0.010	0.15	0.14	0.010	0.14	0.14	0.010	0.14	0.14	0.010	0.15	0.15
	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	0.15	0.16	0.989	0.15	0.15	0.990	0.15	0.15	0.989	0.15	0.14
			0.004	0.10	0.10	0.004	0.10	0.10	0.004	0.10	0.09	0.004	0.10	0.09
			0.004	0.10	0.10	0.004	0.10	0.10	0.004	0.10	0.10	0.004	0.10	0.09
			0.002	0.07	0.07	0.002	0.07	0.06	0.002	0.07	0.06	0.002	0.07	0.06
		MAP	0.990	0.15	0.15	0.990	0.15	0.15	0.990	0.15	0.15	0.990	0.15	0.14
			0.004	0.10	0.09	0.004	0.10	0.09	0.004	0.10	0.09	0.004	0.10	0.09
			0.004	0.10	0.09	0.004	0.10	0.09	0.004	0.10	0.10	0.004	0.10	0.08
			0.002	0.07	0.06	0.002	0.07	0.06	0.002	0.07	0.06	0.002	0.07	0.06
	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). "Configuration I" uses $\mathbf{p} = (0.95, 0.02, 0.02, 0.01)'$ and "Configuration II" uses $\mathbf{p} = (0.990, 0.004, 0.004, 0.002)'$. For H2, H3, and H4, the configurations listed below minimize the expected number of tests per individual specimen. The values used for the assay accuracies are $S_{e:(1)k} = 0.95$, $S_{p:(1)k} = 0.98$, $S_{e:(2)k} = 0.98$, and $S_{p:(2)k} = 0.99$, for $k = 1, 2$.

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 multiple assays, the software that calculates the optimal pool sizes is currently unavailable. Thus, we used the same pool sizes, 29×29, reported in Table 1.

Table D.3: Simulation study with $L = 2$ assays. 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. Values of SD and SE presented below have been multiplied by 10^2. Flat priors have been used for all parameters; i.e., $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4)$, $S_{e:(l)k} \sim \text{beta}(1,1)$, and $S_{p:(l)k} \sim \text{beta}(1,1)$, for $l = 1, 2$ and $k = 1, 2$. 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 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	0.35	0.42	0.949	0.34	0.35	0.949	0.33	0.34	0.948	0.33	0.36
			0.021	0.25	0.27	0.021	0.24	0.23	0.020	0.22	0.22	0.021	0.23	0.24
			0.022	0.25	0.28	0.021	0.22	0.23	0.020	0.22	0.22	0.021	0.21	0.24
			0.011	0.15	0.16	0.010	0.15	0.15	0.010	0.15	0.14	0.010	0.15	0.15
		MAP	0.949	0.35	0.42	0.950	0.34	0.35	0.950	0.33	0.34	0.950	0.33	0.36
			0.020	0.24	0.27	0.020	0.23	0.23	0.020	0.22	0.22	0.020	0.23	0.24
			0.020	0.23	0.28	0.020	0.22	0.23	0.020	0.21	0.22	0.020	0.22	0.24
			0.010	0.15	0.16	0.010	0.15	0.15	0.010	0.15	0.14	0.010	0.15	0.15
	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	0.25	0.33	0.988	0.16	0.18	0.989	0.15	0.16	0.989	0.15	0.18
			0.006	0.16	0.21	0.005	0.11	0.12	0.004	0.09	0.10	0.004	0.10	0.11
			0.006	0.18	0.21	0.005	0.12	0.12	0.004	0.11	0.11	0.005	0.10	0.11
			0.003	0.06	0.09	0.002	0.06	0.07	0.002	0.06	0.07	0.002	0.06	0.07
		MAP	0.990	0.17	0.33	0.990	0.17	0.18	0.990	0.15	0.16	0.990	0.16	0.18
			0.004	0.11	0.21	0.004	0.11	0.12	0.004	0.09	0.10	0.004	0.10	0.11
			0.004	0.12	0.21	0.004	0.12	0.12	0.004	0.11	0.11	0.004	0.11	0.11
			0.002	0.06	0.09	0.002	0.06	0.07	0.002	0.06	0.07	0.002	0.06	0.07
	Avg. number of tests		1130.8			713.2 (36.9%)			610.8 (45.9%)			780.1 (31.0%)		

Table D.4: Simulation study with $L = 2$ assays. Characteristics of posterior mean (Mean) and MAP estimates of $\boldsymbol{\delta} = (S_{e:(1)1}, S_{e:(1)2}, S_{p:(1)1}, S_{p:(1)2}, S_{e:(2)1}, S_{e:(2)2}, S_{p:(2)1}, S_{p:(2)2})'$ for **Configuration I**, which uses $\mathbf{p} = (0.95, 0.02, 0.02, 0.01)'$. 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. Values of SD and SE presented below have been multiplied by 10^1. Flat priors have been used for all parameters; i.e., $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4)$, $S_{e:(l)k} \sim \text{beta}(1,1)$, and $S_{p:(l)k} \sim \text{beta}(1,1)$, for $l = 1, 2$ and $k = 1, 2$. 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	S3	Est	SD	SE
		0.924	0.33	0.42	0.943	0.20	0.21	0.946	0.14	0.15	0.941	0.21	0.23
		0.924	0.34	0.42	0.943	0.21	0.21	0.946	0.14	0.15	0.943	0.21	0.23
		0.983	0.05	0.07	0.979	0.06	0.06	0.979	0.06	0.06	0.981	0.07	0.09
		0.983	0.05	0.07	0.980	0.06	0.06	0.979	0.06	0.06	0.981	0.07	0.09
	Mean	0.944	0.21	0.37	0.971	0.14	0.17	0.971	0.14	0.16	0.965	0.17	0.22
$S_{e:(1)1} = 0.95$		0.944	0.22	0.36	0.971	0.14	0.17	0.971	0.13	0.16	0.965	0.16	0.22
$S_{e:(1)2} = 0.95$		0.989	0.04	0.05	0.988	0.05	0.06	0.987	0.05	0.06	0.989	0.05	0.05
$S_{p:(1)1} = 0.98$		0.990	0.04	0.04	0.988	0.05	0.05	0.988	0.05	0.06	0.988	0.05	0.05
$S_{p:(1)2} = 0.98$													
$S_{e:(2)1} = 0.98$		0.950	0.33	0.42	0.950	0.20	0.21	0.950	0.13	0.15	0.949	0.22	0.23
$S_{e:(2)2} = 0.98$		0.948	0.35	0.42	0.950	0.21	0.21	0.950	0.14	0.15	0.951	0.22	0.23
$S_{p:(2)1} = 0.99$		0.981	0.06	0.07	0.980	0.06	0.06	0.980	0.06	0.06	0.980	0.08	0.09
$S_{p:(2)2} = 0.99$		0.981	0.06	0.07	0.980	0.06	0.06	0.980	0.06	0.06	0.980	0.09	0.09
	MAP	0.973	0.26	0.37	0.980	0.15	0.17	0.980	0.14	0.16	0.979	0.19	0.22
		0.974	0.27	0.36	0.980	0.15	0.17	0.980	0.13	0.16	0.978	0.19	0.22
		0.990	0.04	0.05	0.990	0.05	0.06	0.989	0.06	0.06	0.991	0.05	0.05
		0.990	0.04	0.04	0.990	0.06	0.05	0.990	0.05	0.06	0.990	0.05	0.05

Table D.5: Simulation study with $L = 2$ assays. Characteristics of posterior mean (Mean) and MAP estimates of $\boldsymbol{\delta} = (S_{e:(1)1}, S_{e:(1)2}, S_{p:(1)1}, S_{p:(1)2}, S_{e:(2)1}, S_{e:(2)2}, S_{p:(2)1}, S_{p:(2)2})'$ for **Configuration II**, which uses $\mathbf{p} = (0.990, 0.004, 0.004, 0.002)'$. 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. Values of SD and SE presented below have been multiplied by 10^1. Flat priors have been used for all parameters; i.e., $\mathbf{p} \sim$ Dirichlet($\mathbf{1}_4$), $S_{e:(l)k} \sim$ beta$(1,1)$, and $S_{p:(l)k} \sim$ beta$(1,1)$, for $l = 1, 2$ and $k = 1, 2$. 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	S3	Est	SD	SE
		0.805	0.87	1.26	0.904	0.48	0.55	0.933	0.31	0.35	0.911	0.38	0.56
		0.806	0.89	1.25	0.907	0.44	0.55	0.930	0.30	0.36	0.908	0.41	0.56
		0.984	0.06	0.09	0.981	0.07	0.09	0.979	0.08	0.09	0.982	0.07	0.11
		0.983	0.06	0.09	0.980	0.08	0.09	0.978	0.08	0.09	0.982	0.07	0.12
	Mean	0.865	0.48	0.92	0.933	0.30	0.51	0.940	0.26	0.46	0.923	0.29	0.59
$S_{e:(1)1} = 0.95$		0.864	0.47	0.93	0.934	0.31	0.50	0.940	0.29	0.45	0.923	0.32	0.59
$S_{e:(1)2} = 0.95$		0.989	0.05	0.05	0.986	0.07	0.08	0.983	0.08	0.10	0.987	0.06	0.06
$S_{p:(1)1} = 0.98$		0.989	0.05	0.05	0.986	0.07	0.08	0.984	0.08	0.10	0.987	0.06	0.06
$S_{p:(1)2} = 0.98$													
$S_{e:(2)1} = 0.98$		0.940	0.74	1.26	0.944	0.55	0.55	0.952	0.33	0.35	0.952	0.52	0.56
$S_{e:(2)2} = 0.98$		0.940	0.79	1.25	0.948	0.51	0.55	0.949	0.32	0.36	0.947	0.53	0.56
$S_{p:(2)1} = 0.99$		0.982	0.08	0.09	0.982	0.09	0.09	0.981	0.09	0.09	0.981	0.12	0.11
$S_{p:(2)2} = 0.99$		0.982	0.08	0.09	0.981	0.09	0.09	0.981	0.09	0.09	0.981	0.12	0.12
	MAP	0.961	0.61	0.92	0.977	0.32	0.51	0.980	0.28	0.46	0.978	0.39	0.59
		0.961	0.60	0.93	0.977	0.33	0.50	0.979	0.31	0.45	0.979	0.41	0.59
		0.990	0.05	0.05	0.990	0.07	0.08	0.990	0.08	0.10	0.990	0.06	0.06
		0.990	0.05	0.05	0.990	0.07	0.08	0.990	0.08	0.10	0.990	0.06	0.06

Web Appendix E. *Extensions to include $K \geq 2$ infections.*

The estimation techniques described in the article are now extended to accommodate group testing data with $K \geq 2$ infections; i.e., we herein assume the assays used in the multistage protocols, such as H2, H3, and AT, can test each specimen for $K \geq 2$ infections simultaneously. The extensions are briefly described, without repeating the conditional distributions.

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 the $2^K \times 1$ 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 the $2^K \times 1$ vector of ω_k's. When $K = 2$, one has $\mathbf{\Omega} = \{00, 10, 01, 11\}$ and $\mathbf{p} = (p_{00}, p_{10}, p_{01}, p_{11})'$, the same notation and parameter introduced in Section 2.

Let $\widetilde{\mathbf{Z}}_j = (\widetilde{Z}_{j1}, \widetilde{Z}_{j2}, ..., \widetilde{Z}_{jK})'$ denote the vector of true statuses of pool $j \in \{1, 2, ..., J\}$, where $\widetilde{Z}_{jk} = I(\sum_{i \in \mathcal{P}_j} \widetilde{Y}_{ik} > 0)$. Denote by $\mathbf{Z}_j = (Z_{j1}, Z_{j2}, ..., Z_{jK})'$ the vector of testing responses of 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 $\text{pr}(Z_{jk} = 0 | \widetilde{Z}_{jk} = 0) = S_{p:jk}$. Let $\mathbf{Z} = (\mathbf{Z}_1{}' \mathbf{Z}_2{}' \cdots \mathbf{Z}_J{}')'$ denote the vector of test responses and $\boldsymbol{\delta}$ denote the collection of $S_{e:jk}$'s and $S_{p:jk}$'s.

Using the same prior structure and assumptions, the posterior sampling algorithm and the EM algorithm can be described analogously as in Sections 3-4. For brevity, we show only the conditional distributions and the solutions involved in these algorithms.

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, $\alpha^*_\omega = \alpha_\omega + \sum_{i=1}^N \sum_{\omega \in \Omega} \widetilde{V}_{(\omega)i}$, and α_ω, for $\omega \in \mathbf{\Omega}$, are the Dirichlet hyperparameters. Let $\widetilde{\mathbf{V}}_i$ be the $2^K \times 1$ vector of $\{\widetilde{V}_{(\omega)i} : \omega \in \Omega\}$. Then the full conditional of $\widetilde{\mathbf{V}}_i$ is $\widetilde{\mathbf{V}}_i | \widetilde{\mathbf{Y}}_{-i}, \mathbf{p}, \boldsymbol{\delta}, \mathbf{Z} \sim \text{Multinomial}(\mathbf{p}^*_i)$, where $\mathbf{p}^*_i$ is a $2^K \times 1$ vector of $p^*_{(\omega)i}$'s, $p^*_{(\omega)i} = \zeta^i_\omega / \zeta^i_+$,

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

$\zeta^i_+ = \sum_{\omega \in \Omega} \zeta^i_\omega$, and $\mathcal{A}_i$, $\widetilde{\mathbf{Y}}_{-i}$, and γ_{ijk} are as defined in Web Appendix A. Using these conditionals, one can sample posterior realizations as in Section 3.

For MAP estimation, the solution used in the E-Step of the EM algorithm is $\mathbf{p}^{(d+1)}$, which has the components

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

for $\omega \in \mathbf{\Omega}$, and $\widetilde{N} = N - 2^K + \sum_{\omega \in \Omega} \alpha_\omega$.

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 conditionals

$$S_{e:(l)k}|\mathbf{Z}, \widetilde{\mathbf{Y}} \sim \text{beta}(a_{S^*_{e:(l)k}}, b_{S^*_{e:(l)k}})$$
$$S_{p:(l)k}|\mathbf{Z}, \widetilde{\mathbf{Y}} \sim \text{beta}(a_{S^*_{p:(l)k}}, b_{S^*_{p:(l)k}}),$$

where $a_{S^*_{e:(l)k}} = a_{S_{e:(l)k}} + \sum_{j\in\mathcal{M}(l)} Z_{jk}\widetilde{Z}_{jk}$, $b_{S^*_{e:(l)k}} = b_{S_{e:(l)k}} + \sum_{j\in\mathcal{M}(l)}(1-Z_{jk})\widetilde{Z}_{jk}$, $a_{S^*_{p:(l)k}} = a_{S_{p:(l)k}} + \sum_{j\in\mathcal{M}(l)}(1-Z_{jk})(1-\widetilde{Z}_{jk})$, $b_{S^*_{p:(l)k}} = b_{S_{p:(l)k}} + \sum_{j\in\mathcal{M}(l)} Z_{jk}(1-\widetilde{Z}_{jk})$, and $\mathcal{M}(l)$ is defined in Section 4, for $k = 1, 2, ..., K$ and $l = 1, 2, ..., L$.

For MAP estimation with unknown $\boldsymbol{\delta}$, the additional solutions to be used in the M-step of the EM algorithm are

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

where $\widetilde{J}_{S_{e:(l)k}} = a_{S_{e:(l)k}} + b_{S_{e:(l)k}} - K + \sum_{j\in\mathcal{M}(l)} E(\widetilde{Z}_{jk}|\mathbf{Z}; \boldsymbol{\delta}^{(d)}, \mathbf{p}^{(d)})$ and $\widetilde{J}_{S_{p:(l)k}} = a_{S_{p:(l)k}} + b_{S_{p:(l)k}} - 2 + \sum_{j\in\mathcal{M}(l)}\{1 - E(\widetilde{Z}_{jk}|\mathbf{Z}; \boldsymbol{\delta}^{(d)}, \mathbf{p}^{(d)})\}$.

Web Appendix F. *R function for the Bayesian estimation.*

An R function is provided to implement the estimation techniques with $K = 2$ diseases. The R code with documentation and examples are available at `https://github.com/mswarasi`.

Usage

```
multDiseaseBayes(p0=c(.90,.06,.03,.01),delta0=c(.95,.95,.98,.98),Z,
                 Yt=matrix(0,N,2),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=6000,
                 emGit=6000,emburn=1000,emmaxit=200,emtol=1e-03,
                 method=c("MAP","Mean"),accuracy=c("unknown","known"))
```

Arguments

`p0`	The initial value of $\mathbf{p} = (p_{00}, p_{10}, p_{01}, p_{11})'$, an estimate from the historical data.
`delta0`	The initial value of $\boldsymbol{\delta} = (S_{e:(1)1}, S_{e:(1)2}, S_{p:(1)1}, S_{p:(1)2})'$, an estimate from the assay product literature. Used only when the assay accuracies are unknown.
`Z`	A matrix of the observed group testing data, Z.
`Yt`	A $N \times 2$ matrix of the individual true binary statuses.
`N`	The number of individuals tested (i.e., the sample size).
`S`	The maximum number of times an individual is tested in different pools.
`p.pr`	Dirichlet prior for $\mathbf{p}$.
`Se1.pr`	Beta prior for $S_{e:(1)1}$.
`Se2.pr`	Beta prior for $S_{e:(1)2}$.
`Sp1.pr`	Beta prior for $S_{p:(1)1}$.
`Sp2.pr`	Beta prior for $S_{p:(1)2}$.
`postGit`	The number of Gibbs samples, G, to be drawn from the posterior dist.
`emGit`	The number of Gibbs samples, G, to be used in the EM algorithm.
`emburn`	The initial Gibbs samples to be discarded in the EM algorithm.
`emmaxit`	The maximum number of iterations the EM algorithm can run.
`emtol`	The convergence tolerance used in the EM algorithm.
`method`	The estimation method to be used: "MAP" or "Mean". Defaults to "MAP".
`accuracy`	Whether the assay accuracies are known or unknown. Defaults to "unknown".

Value

A list of components:

`prevalence`	An estimate of p, a point estimate for MAP but MCMC sample for Mean.
`accuracy`	An estimate of delta, a point estimate for MAP but MCMC sample for Mean.
`convergence`	A binary 0 or 1 indicator, where 0 indicates successful completion and 1 indicates the EM algorithm reaches the max iteration. For Mean, convergence is always 0.