Published in final edited form as:

Stat Med. 2016 September 20; 35(21): 3851–3864. doi:10.1002/sim.6964.

Estimating the prevalence of multiple diseases from two-stage hierarchical pooling

Md S. Warasi^a, Joshua M. Tebbs^{a,*}, Christopher S. McMahan^b, and Christopher R. Bilder^c

^aDepartment of Statistics, University of South Carolina, Columbia, SC 29208, U.S.A.

^bDepartment of Mathematical Sciences, Clemson University, Clemson, SC 29634, U.S.A.

^cDepartment of Statistics, University of Nebraska-Lincoln, Lincoln, NE 68583, U.S.A.

Abstract

Testing protocols in large-scale sexually transmitted disease screening applications often involve pooling biospecimens (e.g., blood, urine, swabs, etc.) to lower costs and to increase the number of individuals who can be tested. With the recent development of assays that detect multiple diseases, it is now common to test biospecimen pools for multiple infections simultaneously. In a recent article, Tebbs, McMahan, and Bilder (*Biometrics*, 2013) developed an expectation-maximization algorithm to estimate the prevalence of two infections using a two-stage, Dorfman-type testing algorithm motivated by current screening practices for chlamydia and gonorrhea in the United States. In this article, we have the same goal but instead take a more flexible Bayesian approach. Doing so allows us to incorporate information about assay uncertainty during the testing process, which involves testing both pools and individuals, and also to update information as individuals are tested. Overall, our approach provides reliable inference for disease probabilities and accurately estimates assay sensitivity and specificity even when little or no information is provided in the prior distributions. We illustrate the performance of our estimation methods using simulation and by applying them to chlamydia and gonorrhea data collected in Nebraska.

Keywords

Bayesian estimation; group testing; Infertility Prevention Project; pooled testing; screening; sensitivity; specificity

1. Introduction

Testing biospecimens in pools, which is known as group testing (or pooled testing), is a cost-effective alternative to individual testing in a variety of disease screening applications. Originally proposed by Dorfman [1] to screen World War II soldiers for syphilis, group testing is now widely used to screen human populations for sexually transmitted diseases, including HIV [2], HBV and HCV [3, 4], and chlamydia and gonorrhea [5], and for other infectious diseases including West Nile virus [6], malaria [7], and influenza [8]. Group

*Correspondence to: Joshua M. Tebbs, 209G LeConte, Department of Statistics, University of South Carolina, Columbia, SC 29208, U.S.A.

†tebbs@stat.sc.edu

testing also arises in other applications, including drug discovery [9], genetics [10], animal disease testing [11], and food safety [12].

Because pooling has become so widespread, statistical research in group testing has also flourished. This research has generally followed two different paths. In the classification (case identification) problem, the goal is to classify each individual as positive or negative. This involves retesting individuals in pools that test positively; see [13] for a review. In the estimation problem, responses from pools provide enough information to estimate a population prevalence, at times, more efficiently than when individual testing is used [14, 15]. Recent work has focused on the development of regression methods to estimate subject-specific probabilities, either parametrically [16, 17], semi-parametrically [18], or non-parametrically [19].

This article is motivated by screening practices for chlamydia and gonorrhea (CT/NG) in the United States as part of a national program formerly known as the Infertility Prevention Project (IPP). Chlamydia and gonorrhea are two of the most common sexually transmitted diseases; together, there are approximately 1.5 million new infections reported each year in the United States [20]. The IPP was a federally funded program managed by the Centers for Disease Control and Prevention (CDC) and implemented in each of the 50 states during 1988–2013. After the Affordable Care Act (ACA) was passed in 2010 and implemented in 2014, screening for CT/NG has continued in each state but now testing centers rely on other sources of funding (e.g., federal health care plans, private insurance, etc.). To reduce costs while still screening the same number of individuals for CT/NG, Iowa's IPP program switched from individual testing to group testing in 1999. Doing so has led to millions of dollars in savings [21], and other states [22] have since adopted group testing as well. In light of new funding uncertainties created by the ACA [23], Iowa's application of group testing might serve as a model for how to perform CT/NG screening nationwide.

Estimating the prevalence of a single disease has received a large amount of attention in the group testing literature. However, testing procedures for CT/NG and other infections are now moving towards the use of assays which detect multiple infections at once [20]. In these instances, pools of individuals are tested for multiple infections using a single assay, and then pools are resolved (decoded) for each infection. Estimation in this situation is challenging, because the true infection statuses on the same individual are latent (due to inherent assay error) and are also correlated. Recently, Tebbs, McMahan, and Bilder [24] developed an expectation-maximization (EM) algorithm to jointly estimate the prevalence of CT/NG, motivated by screening practices in Iowa which use group testing (see Section 2). Their work, in the two-infection case, generalized the estimation approach in Hughes-Oliver and Rosenberger [25] to allow for assay error and also for the inclusion of retesting information on positive pools.

In this article, we have the same estimation goals as in [24], but we take a Bayesian approach instead. Doing so confers important advantages. First, it allows us to relax the potentially untrustworthy assumption that diagnostic test accuracy rates (i.e., sensitivity and specificity) are fixed and known. In practice, these rates are usually estimated on the basis of small pilot studies that manufacturers publish in their product literature. Ignoring the

variability in these estimates could compromise inference, especially if the estimates deviate substantially from the true accuracy rates and/or if assay performance varies according to other factors [26]. Second, a Bayesian approach is natural given the sequential manner in which screening data amass over time. For example, the State Hygienic Laboratory (SHL) in Iowa City has screened thousands of Iowa residents each year for CT/NG, dating back to 1992. This affords investigators ample information to construct sensible prior distributions as well as to periodically update information on disease prevalence and assay performance. Our work extends previous Bayesian group testing estimation approaches for single diseases [27, 28].

Subsequent sections of this article are organized as follows. In Section 2, we describe the screening algorithm for CT/NG used in Iowa. This two-stage algorithm was described in detail in [24], so we herein summarize only the salient aspects. In Section 3, we present our estimation methods and discuss prior model selection. In Section 4, we use simulation to assess estimation performance under a variety of prior models, including models which incorporate little or no information about disease prevalence and assay accuracy. In Section 5, we analyze IPP data in the same manner as in [24] to illustrate the advantages of estimation from a Bayesian point of view. In Section 6, we conclude with a brief summary discussion. We use the Web-based Supporting Materials to show how one could generalize our work to estimate probabilities for more than two infections if needed.

2. Two-Stage Pooling Algorithm

The estimation methods we develop in this article are motivated by the two-stage pooling algorithm described below. This algorithm is used to complete CT/NG testing at the SHL in Iowa City and is potentially applicable in other situations.

POOLING ALGORITHM

Stage 1: Individuals are randomly assigned to master pools. Each pool is tested for both infections using a single assay. A single assay detects both infections simultaneously.

Stage 2: Individuals in pools that

- test negatively for both infections are diagnosed as negative for both infections.
 - test positively for either infection are retested (individually) for both infections using the same assay in Stage 1. Diagnoses for both infections are made from the outcomes of the individual tests.
-

Tebbs *et al.* [24] describe various logistical issues of this pooling procedure (as it relates to implementation at the SHL) that we do not repeat here. The point worth emphasizing is that, for simplicity, the SHL uses one assay, the Aptima Combo 2 Assay (Hologic/Gen-Probe, Inc., San Diego) nucleic acid amplification test, for its CT/NG testing. This assay detects both infections simultaneously when it is applied to pools (in Stage 1) and to individuals (in Stage 2). In the infectious disease testing literature, such an assay is said to *discriminate* because it elicits a diagnosis for each infection separately. In this article, we assume that a discriminating assay is available and that it can be applied to both pooled and individual specimens. The literature is replete with examples of multiple-infection assays that are

discriminating in nature. For example, most assays based on nucleic acid amplification technology used for CT/NG detection discriminate between the two infections in both urine and swab specimens [20, 26]. Furthermore, the CDC recommends that nucleic acid amplification testing be used for laboratory-based CT/NG detection [26].

3. Bayesian Estimation

3.1. Model formulation and inference

Suppose N individuals are to be tested for two infections (e.g., CT/NG, etc.) using the algorithm described in Section 2. Let $\tilde{\mathbf{Y}}_{ik} = (\tilde{Y}_{i1k}, \tilde{Y}_{i2k})'$ denote the vector of true individual

binary statuses, for $i = 1, 2, \dots, c_k$ and $k = 1, 2, \dots, K$, where $N = \sum_{k=1}^K c_k$. We call c_k the pool size for the k th master pool. The number of master pools formed at Stage 1 is K . We assume the $\tilde{\mathbf{Y}}_{ik}$'s, conditional on $\mathbf{p} = (p_{00}, p_{10}, p_{01}, p_{11})'$, are independent and identically distributed random vectors with probability mass function

$$\text{pr}(\tilde{Y}_{i1k} = \tilde{y}_1, \tilde{Y}_{i2k} = \tilde{y}_2 | \mathbf{p}) = p_{00}^{(1-\tilde{y}_1)(1-\tilde{y}_2)} p_{10}^{\tilde{y}_1(1-\tilde{y}_2)} p_{01}^{(1-\tilde{y}_1)\tilde{y}_2} p_{11}^{\tilde{y}_1\tilde{y}_2},$$

where $\tilde{y}_1, \tilde{y}_2 \in \{0, 1\}$ and $p_{00} + p_{10} + p_{01} + p_{11} = 1$. Note that because of inherent assay error, the $\tilde{\mathbf{Y}}_{ik}$'s are best regarded as latent.

Let $\tilde{\mathbf{Z}}_k = (\tilde{Z}_{1k}, \tilde{Z}_{2k})'$ denote the vector of true binary statuses for the k th master pool, where $\tilde{Z}_{jk} = I(\sum_{i=1}^{c_k} \tilde{Y}_{ijk} > 0)$, for $j = 1, 2$, and $I(\cdot)$ is the indicator function. In other words, $\tilde{Z}_{jk} = 1$ if at least one individual in the k th master pool is truly positive for the j th infection, $\tilde{Z}_{jk} = 0$ otherwise. Let $\mathbf{Z}_k = (Z_{1k}, Z_{2k})'$ denote the vector of testing responses observed for the k th master pool in Stage 1, where $Z_{jk} = 1$ if the k th master pool tests positively for the j th infection, $Z_{jk} = 0$ otherwise. If the k th master pool tests positively for at least one infection in Stage 1, let $\mathbf{Y}_{ik} = (Y_{i1k}, Y_{i2k})'$ denote the vector of individual testing responses observed for the i th individual, $i = 1, 2, \dots, c_k$, in Stage 2. We allow for pools (in Stage 1) and individuals (in Stage 2) to be misclassified and denote the assay sensitivity and specificity by

$$S_{e,j} = \text{pr}(Z_{jk} = 1 | \tilde{Z}_{jk} = 1) = \text{pr}(Y_{ijk} = 1 | \tilde{Y}_{ijk} = 1)$$

$$S_{p,j} = \text{pr}(Z_{jk} = 0 | \tilde{Z}_{jk} = 0) = \text{pr}(Y_{ijk} = 0 | \tilde{Y}_{ijk} = 0),$$

respectively, for $j = 1, 2$. We assume $S_{e,j}$ and $S_{p,j}$ do not depend on the pool size c_k in Stage 1 so that these probabilities also apply for individual tests performed in Stage 2. This assumption is common in group testing research for single infections [13]. For this to be reasonable in practice, assay detection thresholds and/or dilution ratios may need to be changed to accommodate both pooled and individual specimens; see [29] and the references therein.

The observed data from the pooling algorithm in Section 2 consist of (a) the testing responses $\mathbf{Z}_k = (Z_{1k}, Z_{2k})'$ from the K master pools in Stage 1 and (b) the additional c_k individual testing responses $\mathbf{Y}_{jk} = (Y_{1k}, Y_{2k})'$ from those pools which tested positively for either infection in Stage 1. For notational purposes, we aggregate all master pool testing responses into a vector denoted by $\mathbf{Z}$ and all individual testing responses into a vector denoted by $\mathbf{Y}$. Let $\boldsymbol{\theta} = (\mathbf{p}', \boldsymbol{\delta}')'$, where $\boldsymbol{\delta} = (S_{e,1}, S_{e,2}, S_{p,1}, S_{p,2})'$. Although it is possible to write out the observed data likelihood $\pi(\mathbf{Z}, \mathbf{Y}|\boldsymbol{\theta})$, its form is not easily amenable to performing a Bayesian analysis. Therefore, we use a data augmentation step that introduces the individuals' true infection statuses $\tilde{Y}_{jk}$ as latent random variables. Let $\tilde{\mathbf{Y}}$ denote the vector that aggregates all of the latent $\tilde{Y}_{jk}$'s. The joint distribution of the observed data $\{\mathbf{Z}, \mathbf{Y}\}$ and the latent data $\tilde{\mathbf{Y}}$, conditional on $\boldsymbol{\theta}$, can be expressed as

$$\begin{aligned} \pi(\mathbf{Z}, \mathbf{Y}, \tilde{\mathbf{Y}}|\boldsymbol{\theta}) = & \prod_{k=1}^K \prod_{i=1}^{c_k} p_{00}^{(1-\tilde{Y}_{i1k})(1-\tilde{Y}_{i2k})} p_{10}^{\tilde{Y}_{i1k}(1-\tilde{Y}_{i2k})} p_{01}^{(1-\tilde{Y}_{i1k})\tilde{Y}_{i2k}} p_{11}^{\tilde{Y}_{i1k}\tilde{Y}_{i2k}} \\ & \times \left[\prod_{j=1}^2 \prod_{k=1}^K \left(S_{e,j}^{Z_{jk}} \bar{S}_{e,j}^{1-Z_{jk}} \right)^{I(\sum_{i=1}^{c_k} \tilde{Y}_{ijk} > 0)} \left(S_{p,j}^{1-Z_{jk}} \bar{S}_{p,j}^{Z_{jk}} \right)^{I(\sum_{i=1}^{c_k} \tilde{Y}_{ijk} = 0)} \right. \\ & \left. \times \left\{ \prod_{i=1}^{c_k} S_{e,j}^{Y_{ijk}} \tilde{S}_{e,j}^{(1-Y_{ijk})\tilde{Y}_{ijk}} S_{p,j}^{(1-Y_{ijk})(1-\tilde{Y}_{ijk})} \bar{S}_{p,j}^{Y_{ijk}(1-\tilde{Y}_{ijk})} \right\}^{I(Z_{+k} > 0)} \right], \quad (1) \end{aligned}$$

where $Z_{+k} = Z_{1k} + Z_{2k}$, $\bar{S}_{e,j} = 1 - S_{e,j}$, and $\bar{S}_{p,j} = 1 - S_{p,j}$. The first line in Equation (1) represents the contribution of the individual latent statuses, while the part within the brackets describes the contributions from Stage 1 (master pool test results; second line) and Stage 2 (individual test results; third line). Note that if $\boldsymbol{\delta}$ were known, Equation (1) would be the same as the complete data likelihood in [24]. For further discussion on additional assumptions underpinning the construction of $\pi(\mathbf{Z}, \mathbf{Y}, \tilde{\mathbf{Y}}|\boldsymbol{\theta})$ in Equation (1), see Section 6.

To complete our Bayesian model specification, we elicit independent beta prior distributions for the assay test accuracies; i.e., $S_{e,j} \sim \text{beta}(a_{S_{e,j}}, b_{S_{e,j}})$ and $S_{p,j} \sim \text{beta}(a_{S_{p,j}}, b_{S_{p,j}})$, for $j = 1, 2$, where all hyperparameters are known. For the vector of infection status probabilities, we specify a Dirichlet prior; i.e.,

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

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. We assume the test accuracies $S_{e,j}$ and $S_{p,j}$ are both independent of $\mathbf{p}$, for $j = 1, 2$. These assumptions are analogous to the assumptions made in [27, 28] for single infections.

With these prior choices and assumptions, the “full conditional” distributions can be easily derived from the augmented likelihood function $\pi(\mathbf{Z}, \mathbf{Y}, \tilde{\mathbf{Y}}|\boldsymbol{\theta})$. For the assay accuracies, these distributions are given by

$$S_{e:j}|\mathbf{Z}, \mathbf{Y}, \tilde{\mathbf{Y}} \sim \text{beta}(a_{S_{e:j}}^*, b_{S_{e:j}}^*)$$

$$S_{p:j}|\mathbf{Z}, \mathbf{Y}, \tilde{\mathbf{Y}} \sim \text{beta}(a_{S_{p:j}}^*, b_{S_{p:j}}^*),$$

for $j = 1, 2$, where

$$a_{S_{e:j}}^* = a_{S_{e:j}} + \sum_{k=1}^K \left\{ Z_{jk} \tilde{Z}_{jk} + I(Z_{+k} > 0) \sum_{i=1}^{c_k} Y_{ijk} \tilde{Y}_{ijk} \right\}$$

$$b_{S_{e:j}}^* = b_{S_{e:j}} + \sum_{k=1}^K \left\{ (1 - Z_{jk}) \tilde{Z}_{jk} + I(Z_{+k} > 0) \sum_{i=1}^{c_k} (1 - Y_{ijk}) \tilde{Y}_{ijk} \right\}$$

$$a_{S_{p:j}}^* = a_{S_{p:j}} + \sum_{k=1}^K \left\{ (1 - Z_{jk})(1 - \tilde{Z}_{jk}) + I(Z_{+k} > 0) \sum_{i=1}^{c_k} (1 - Y_{ijk})(1 - \tilde{Y}_{ijk}) \right\}$$

$$b_{S_{p:j}}^* = b_{S_{p:j}} + \sum_{k=1}^K \left\{ Z_{jk}(1 - \tilde{Z}_{jk}) + I(Z_{+k} > 0) \sum_{i=1}^{c_k} Y_{ijk}(1 - \tilde{Y}_{ijk}) \right\}$$

and $\tilde{Z}_{jk} = I(\sum_{i=1}^{c_k} \tilde{Y}_{ijk} > 0)$. For the prevalence parameter $\mathbf{p}$, the full conditional distribution is again Dirichlet; i.e., $\mathbf{p}|\tilde{\mathbf{Y}} \sim \text{Dirichlet}(\Psi)$, where

$$\Psi = \left(\alpha_{00} + \sum_{k=1}^K \sum_{i=1}^{c_k} \tilde{V}_{(00)ik}, \alpha_{10} + \sum_{k=1}^K \sum_{i=1}^{c_k} \tilde{V}_{(10)ik}, \alpha_{01} + \sum_{k=1}^K \sum_{i=1}^{c_k} \tilde{V}_{(01)ik}, \alpha_{11} + \sum_{k=1}^K \sum_{i=1}^{c_k} \tilde{V}_{(11)ik} \right)'$$

and the latent random variables $\tilde{V}_{(uv)ik} = \tilde{Y}_{i1k}^u (1 - \tilde{Y}_{i1k})^{1-u} \tilde{Y}_{i2k}^v (1 - \tilde{Y}_{i2k})^{1-v}$, for $u, v \in \{0, 1\}$.

If the latent data $\tilde{\mathbf{Y}}$ were observed, the posterior distributions for the prevalence parameters in $\mathbf{p}$ and the test accuracies in δ would be fully determined. However, because the true individual statuses in $\tilde{\mathbf{Y}}$ are not observed, we develop a Gibbs sampler to enable posterior inference. Let $\tilde{\mathbf{Y}}_{k(i)} = (\tilde{\mathbf{Y}}'_{1k}, \dots, \tilde{\mathbf{Y}}'_{i-1,k}, \tilde{\mathbf{Y}}'_{i+1,k}, \dots, \tilde{\mathbf{Y}}'_{c_k k})'$ denote the vector of latent responses in the k th group for all individuals except the i th one. The conditional distribution of $\tilde{\mathbf{Y}}_{ik} = (\tilde{V}_{(00)ik}, \tilde{V}_{(10)ik}, \tilde{V}_{(01)ik}, \tilde{V}_{(11)ik})'$ given $\{\tilde{\mathbf{Y}}_{k(i)}, \mathbf{p}, \delta, \mathbf{Y}, \mathbf{Z}\}$ is multinomial with cell probabilities $\zeta_{00}^{ik}/\zeta^{ik}, \zeta_{10}^{ik}/\zeta^{ik}, \zeta_{01}^{ik}/\zeta^{ik}$, and $\zeta_{11}^{ik}/\zeta^{ik}$, where

$$\zeta_{00}^{ik} = p_{00} \prod_{j=1}^2 \left(S_{e:j}^{Z_{jk}} \bar{S}_{e:j}^{1-Z_{jk}} \right)^{\gamma_{ijk}} \left(S_{p:j}^{1-Z_{jk}} \bar{S}_{p:j}^{Z_{jk}} \right)^{1-\gamma_{ijk}} \left(S_{p:j}^{1-Y_{ijk}} \bar{S}_{p:j}^{Y_{ijk}} \right)^{I(Z_{+k}>0)}$$

$$\zeta_{10}^{ik} = p_{10} S_{e:1}^{Z_{1k}} \bar{S}_{e:1}^{1-Z_{1k}} \left(S_{e:2}^{Z_{2k}} \bar{S}_{e:2}^{1-Z_{2k}} \right)^{\gamma_{i2k}} \left(S_{p:2}^{1-Z_{2k}} \bar{S}_{p:2}^{Z_{2k}} \right)^{1-\gamma_{i2k}} \left(S_{e:1}^{Y_{i1k}} \bar{S}_{e:1}^{1-Y_{i1k}} S_{p:2}^{1-Y_{i2k}} \bar{S}_{p:2}^{Y_{i2k}} \right)^{I(Z_{+k}>0)}$$

$$\zeta_{01}^{ik} = p_{01} S_{e:2}^{Z_{2k}} \bar{S}_{e:2}^{1-Z_{2k}} \left(S_{e:1}^{Z_{1k}} \bar{S}_{e:1}^{1-Z_{1k}} \right)^{\gamma_{i1k}} \left(S_{p:1}^{1-Z_{1k}} \bar{S}_{p:1}^{Z_{1k}} \right)^{1-\gamma_{i1k}} \left(S_{e:2}^{Y_{i2k}} \bar{S}_{e:2}^{1-Y_{i2k}} S_{p:1}^{1-Y_{i1k}} \bar{S}_{p:1}^{Y_{i1k}} \right)^{I(Z_{+k}>0)}$$

$$\zeta_{11}^{ik} = p_{11} \prod_{j=1}^2 S_{e:j}^{Z_{jk}} \bar{S}_{e:j}^{1-Z_{jk}} \left(S_{e:j}^{Y_{ijk}} \bar{S}_{e:j}^{1-Y_{ijk}} \right)^{I(Z_{+k}>0)},$$

$\zeta^{ik} = \sum_{u=0}^1 \sum_{v=0}^1 \zeta_{uv}^{ik}$, and $\gamma_{ijk} = I(\tilde{Y}_{ijk} > 0)$. Note that by sampling $\tilde{\mathbf{Y}}_{ik}$ from this conditional distribution, $\tilde{\mathbf{Y}}_{ik} = (\tilde{V}_{(10)ik} + \tilde{V}_{(11)ik}, \tilde{V}_{(01)ik} + \tilde{V}_{(11)ik})'$; in other words, the true individual statuses in $\tilde{\mathbf{Y}}_{ik}$ are uniquely determined.

Using the full conditional distributions of $\mathbf{p}$, δ , and $\tilde{\mathbf{Y}}_{ik}$ described above, we now outline our Gibbs sampler to implement a Bayesian analysis with the observed data from the pooling algorithm described in Section 2:

GIBBS SAMPLER

- 1 Initialize $\tilde{\mathbf{Y}}_{ik}^{(0)} = (\tilde{Y}_{i1k}^{(0)}, \tilde{Y}_{i2k}^{(0)})'$, for $i = 1, 2, \dots, c_k$ and $k = 1, 2, \dots, K$. Set $d = 1$.
- 2 Sample $\mathbf{p}^{(d)}$ from $\mathbf{p} | \tilde{\mathbf{Y}}^{(d-1)} \sim \text{Dirichlet}(\Psi)$, where $\tilde{\mathbf{Y}}^{(d-1)}$ is the collection of all $\tilde{\mathbf{Y}}_{ik}^{(d-1)}$'s.
- 3 Sample $S_{e:j}^{(d)}$ and $S_{p:j}^{(d)}$ from $S_{e:j} | \mathbf{Z}, \mathbf{Y}, \tilde{\mathbf{Y}}^{(d-1)} \sim \text{beta}(a_{S_{e:j}}^*, b_{S_{e:j}}^*)$ and $S_{p:j} | \mathbf{Z}, \mathbf{Y}, \tilde{\mathbf{Y}}^{(d-1)} \sim \text{beta}(a_{S_{p:j}}^*, b_{S_{p:j}}^*)$, for $j = 1, 2$, and set $\delta^{(d)} = (S_{e:1}^{(d)}, S_{e:2}^{(d)}, S_{p:1}^{(d)}, S_{p:2}^{(d)})'$.
- 4 For $i = 1, 2, \dots, c_k$ and $k = 1, 2, \dots, K$, sample $\tilde{\mathbf{Y}}_{ik}^{(d)} = (\tilde{V}_{(00)ik}^{(d)}, \tilde{V}_{(10)ik}^{(d)}, \tilde{V}_{(01)ik}^{(d)}, \tilde{V}_{(11)ik}^{(d)})'$ from $\tilde{\mathbf{Y}}_{ik} | \tilde{\mathbf{Y}}_{k(i)}^{(d)}, \mathbf{p}^{(d)}, \delta^{(d)}, \mathbf{Y}, \mathbf{Z} \sim \text{multinomial} \{1, (\zeta_{00}^{ik}/\zeta^{ik}, \zeta_{10}^{ik}/\zeta^{ik}, \zeta_{01}^{ik}/\zeta^{ik}, \zeta_{11}^{ik}/\zeta^{ik})'\}$, where $\tilde{\mathbf{Y}}_{k(i)}^{(d)} = (\tilde{\mathbf{Y}}_{1k}^{(d)}, \dots, \tilde{\mathbf{Y}}_{i-1,k}^{(d)}, \tilde{\mathbf{Y}}_{i+1,k}^{(d)}, \dots, \tilde{\mathbf{Y}}_{c_k,k}^{(d)})'$. Set $\tilde{\mathbf{Y}}_{ik}^{(d)} = (\tilde{V}_{(10)ik}^{(d)} + \tilde{V}_{(11)ik}^{(d)}, \tilde{V}_{(01)ik}^{(d)} + \tilde{V}_{(11)ik}^{(d)})'$.
- 5 Set $d = d + 1$.

3.2. Prior elicitation

We now discuss prior model specification for the assay accuracies $S_{e,j}$ and $S_{p,j}$ and the prevalence parameter $\mathbf{p}$. A noninformative approach might specify flat priors for all parameters; i.e., $S_{e,j} \sim \text{beta}(1, 1)$, $S_{p,j} \sim \text{beta}(1, 1)$, and $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4)$. In fact, we demonstrate in Section 4 that posterior estimates of $\mathbf{p}$ and δ are close to the true values even when these priors are used. Of course, properly chosen informative prior models should increase posterior precision.

In most screening situations, there will be information readily available for researchers to formulate informative priors for $S_{e,j}$ and $S_{p,j}$. Before a diagnostic assay is introduced for commercial use, its performance is assessed in pilot studies involving known positive and known negative specimens. Data from these studies can be used to elicit sensible beta hyperparameters. For example, the most recent product literature for the Aptima Combo 2 Assay, which is available at <http://www.hologic.com>, summarizes a pilot study describing the assay's performance with urine and swab specimens from females and males. This literature documents that among 127 known NG-positive female urine samples, 116 tested positively with the assay, giving an estimated sensitivity of 0.913. Note that in the absence of additional information, the estimation methods in [24] might require one to essentially treat 0.913 as the “true” sensitivity, disregarding the fact that this is only an estimate. In practice, it is important to acknowledge that assay performance may depend on a variety of factors, including those related to implementation and perhaps even the population being tested [26]. As we show in the Web-based Supporting Materials, a beta(117,12) distribution is consistent with the pilot data described above.

To model the prevalence parameter $\mathbf{p}$, we use a class of distributions motivated by the power prior class described in Ibrahim *et al.* [30]. This class assumes the availability of “historical data” and amalgamates their likelihood function with a prior distribution assumed for them. Applying this idea to our problem, we specify $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4 + a_0 N_0 \mathbf{p}_0)$ as a prior distribution. The value $a_0 \in [0, 1]$ controls the amount of weight given to the historical data (the so-called “precision parameter”), N_0 is the historical data sample size, and $\mathbf{p}_0$ is an estimate of $\mathbf{p}$ obtained from the historical data. This family of priors is ideal for the large-scale screening applications we consider, where, for example, previous years' testing outcomes can be treated as historical data; see Section 5. Note that choosing $a_0 = 0$ gives $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4)$. As a_0 increases, the amount of weight given to the historical data increases.

4. Simulation Evidence

We performed simulation studies to assess the characteristics of our estimation procedure. We used different values of the prevalence parameter $\mathbf{p}$, different prior models for $S_{e,j}$ and $S_{p,j}$, and prior models for $\mathbf{p}$ which incorporate various levels of historical information. A subset of the results is given herein; complete results are in the Web-based Supporting Materials.

4.1. Simulation description

We generated individual true statuses from a multinomial distribution with cell probabilities in $\mathbf{p} = (p_{00}, p_{10}, p_{01}, p_{11})'$. We took $\mathbf{p} = (0.80, 0.10, 0.09, 0.01)$ to consider moderate-level infections and $\mathbf{p} = (0.95, 0.03, 0.01, 0.01)$ to allow for rarer infections. Note that the marginal infection probabilities are 11 and 10 percent for the first case and 4 and 2 percent for the second case. These probabilities are representative of our CT/NG data application in Section 5. In all cases, we used $N = 1000$ individuals. This sample size is much smaller than our application in Section 5 but is sufficient to illustrate our main findings.

To simulate the observed diagnoses from the algorithm in Section 2, we took the individual true statuses $\tilde{\mathbf{Y}}_{jk}$ and randomly assigned them to pools of size c_k^* , where c_k^* is the pool size that minimizes the expected number of tests needed to diagnose each individual using the algorithm in Section 2 (i.e., it is the most “cost-effective” choice); see [24]. Testing responses for pools in Stage 1 were then determined by simulating $Z_{jk} \sim \text{Bernoulli}\{S_{e,j} \tilde{Y}_{jk} + S_{p,j}(1 - \tilde{Y}_{jk})\}$, where $\tilde{Y}_{jk} = I(\sum_{i=1}^{c_k^*} \tilde{Y}_{ijk} > 0)$. In Stage 2, individual testing responses for those pools testing positively in Stage 1 (for either infection) were simulated as $Y_{ijk} \sim \text{Bernoulli}\{S_{e,j} \tilde{Y}_{ijk} + S_{p,j}(1 - \tilde{Y}_{ijk})\}$. Throughout our investigation, we took $S_{e,j} = 0.95$ and $S_{p,j} = 0.99$, for $j = 1, 2$. At each parameter configuration, this entire process was repeated $B = 500$ times.

We used different prior distributions for $\mathbf{p}$ and $\boldsymbol{\delta} = (S_{e,1}, S_{e,2}, S_{p,1}, S_{p,2})'$, including those which incorporated no information (i.e., flat priors) and those which were highly informative. For the prevalence parameter, we took $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4 + a_0 \mathbf{N}_0 \mathbf{p}_0)$, where $\mathbf{N}_0 = 1000$ is the size of the historical data set, $\mathbf{p}_0$ is the historical estimate, and $a_0 \in \{0, 0.1, 0.2, \dots, 1\}$. For the assay accuracies, we used

- Flat priors: $S_{e,j} \sim \text{beta}(1, 1)$ and $S_{p,j} \sim \text{beta}(1, 1)$, for $j = 1, 2$
- Informative priors: $S_{e,j} \sim \text{beta}(109.0, 6.7)$ and $S_{p,j} \sim \text{beta}(55.2, 1.6)$, for $j = 1, 2$.

The informative distributions have modes located at the true $S_{e,j} = 0.95$ and $S_{p,j} = 0.99$, 5th percentiles of 0.903 and 0.929, and 95th percentiles of 0.973 and 0.996, respectively.

Posterior sampling was done using the Gibbs sampler described in Section 3.1. In the initialization step, we specified the individual true statuses to be the diagnosed statuses; i.e.,

$\tilde{Y}_{ijk}^{(0)} = I(Z_{i1k} + Z_{i2k} > 0, Y_{ijk} = 1)$, for $j = 1, 2$, and later noted that the performance of our methods was invariant to the choice of initialization. At each configuration, we took $G = 10000$ Gibbs iterates after discarding the first 500. Posterior estimates were calculated from the full set of 10000 Gibbs iterates in each simulation setting (i.e., we did not thin the posterior samples).

4.2. Simulation results

We present results for the moderate-level infection case $\mathbf{p} = (0.80, 0.10, 0.09, 0.01)$ using flat priors for $S_{e,j}$ and $S_{p,j}$ in Figures 1 and 2; complete results for all settings described in Section 4.1 are given in the Web-based Supporting Materials.

Figure 1 provides a summary of the point estimates for $\mathbf{p}$ for all $a_0 \in \{0, 0.1, 0.2, \dots, 1\}$; specifically, the figure shows 5th, 25th, 50th, 75th, and 95th percentiles of the $B = 500$ posterior median estimates of $\mathbf{p}$. Figure 2 shows the same summary, but for estimates of the assay accuracy parameters in $\delta = (S_{e,1}, S_{e,2}, S_{p,1}, S_{p,2})'$. Recall that when $a_0 = 0$, the prior distribution for $\mathbf{p}$ is Dirichlet($\mathbf{1}_4$); i.e., no historical information is used. Even in this situation with flat priors for $S_{e,j}$ and $S_{p,j}$, median estimates for all prevalence parameters (Figure 1) and assay accuracies (Figure 2) are on target. In other words, providing no information in the prior distributions still allows for accurate inference on all model parameters. We find this phenomenon to be quite remarkable; not only can one estimate the prevalence parameter $\mathbf{p}$ but one can also accurately estimate the sensitivity and specificity parameters in δ . This desirable feature of our approach is clearly a byproduct of using a single discriminating assay in both stages; retesting individuals in Stage 2—whenever a master pool in Stage 1 tests positively for one infection or both—provides an abundance of information on the assay's accuracy for both infections. Note that, in single-infection group testing applications, assay accuracy parameters cannot be identified without retesting or making use of historical data [27, 28, 31].

Figures 1 and 2 also show the impact of adding historical information in the Dirichlet($\mathbf{1}_4 + a_0 N_0 \mathbf{p}_0$) prior distribution for $\mathbf{p}$. For the prevalence parameters in Figure 1, posterior distributions tighten noticeably as more information is added (i.e., as a_0 increases), while for the assay accuracies in Figure 2, posterior variability is largely unaffected. Figures 1 and 2 were constructed assuming that the historical estimate $\mathbf{p}_0$ matched the true value of $\mathbf{p}$. However, we obtained nearly identical results when $\mathbf{p}_0$ was misspecified, especially in situations with larger N like our data application in Section 5. Of course, the fact that prevalence parameters in $\mathbf{p}$ are estimated well in the absence of prior information ($a_0 = 0$) should comfort the researcher unwilling to place too much faith in a historical prevalence estimate or when no historical data are available. Following the recommendation of an anonymous referee, we also examined how misspecified prior distributions for the assay accuracies $S_{e,j}$ and $S_{p,j}$ could impact our Bayesian estimates and, simultaneously, how using incorrect values of $S_{e,j}$ and $S_{p,j}$ could affect the maximum likelihood estimates in [24]. Simulation results summarizing this examination are provided in the Web-based Supporting Materials. When applied to $N = 1000$ individuals as before, Bayesian estimates for both $\mathbf{p}$ and δ remain generally on target even under severe prior model misspecification, whereas the estimates of $\mathbf{p}$ from [24] calculated using incorrect values of $S_{e,j}$ and $S_{p,j}$ can be severely biased.

Simulation results for the other settings; i.e., for (i) moderate-level infections, informative assay priors; (ii) rare infections, flat assay priors; and (iii) rare infections, informative assay priors, are given in the Web-based Supporting Materials. Not surprisingly, posterior distributions for assay accuracies tighten when informative priors are used. However, one observation does warrant a remark. When the marginal disease probabilities $p_{10} + p_{11}$ and $p_{01} + p_{11}$ are both small, fewer pools will be resolved in Stage 2 and the sensitivity parameters can be underestimated. Naturally, fewer retests will curtail the information needed to estimate both $S_{e,j}$ and $S_{p,j}$ accurately, but this will have a more pronounced effect on assay sensitivity because fewer individuals will be truly positive. At the same time, it

should be noted that this “underestimation behavior” disappeared completely when we increased the sample size in our simulation to $N = 10000$ individuals (results not shown), a sample size more in line with our data application in Section 5.

5. Infertility Prevention Project Data

To illustrate our methodology with real data, we use CT/NG testing outcomes collected by the Nebraska Public Health Laboratory (NPHL) during 2008 and 2009. As part of their IPP screening program, there were 23,146 individuals tested in 2008 and 27,551 individuals tested in 2009. Our data set consists of individual diagnoses for each infection for both 2008 and 2009. We implement our Bayesian group testing estimation methods for the 2009 individuals, using the 2008 results as historical data.

To perform the 2009 analysis, we first cross-classified individuals in both years according to their gender and specimen type (swab or urine). This was done because operating characteristics for assays commonly used for CT/NG detection depend on these factors [20]. Within each of the four gender/specimen type strata, we analyzed the 2008 individual testing results to formulate a historical estimate $\mathbf{p}_0$ that accounts for potential misclassification. These historical estimates are shown in Table 1; for more details on how these were computed, see the Web-based Supporting Materials. Also given in Table 1 are informative beta prior distributions for $S_{e,j}$ and $S_{p,j}$ ($j = 1$ for CT; $j = 2$ for NG), which, for purposes of illustration, were chosen to be consistent with the pilot studies presented in the Aptima Combo 2 Assay product literature. We use the Web-based Supporting Materials to provide precise details on how these informative prior distributions were chosen. Finally, for the prevalence parameter $\mathbf{p}$ in each stratum, we took $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4 + a_0 N_0 \mathbf{p}_0)$, where N_0 is the 2008 stratum size, and examined the $a_0 = 0$, $a_0 = 0.5$, and $a_0 = 1$ cases separately. Note that taking $a_0 = 0$ ignores the 2008 information about the prevalence whereas the $a_0 = 1$ case weighs the 2008 information heavily. One might view taking $a_0 = 0.5$ as a compromise between the two extremes.

To emulate the use of group testing with the 2009 individuals, we assigned them to pools of size c_k^* which was determined using the 2008 historical information in Table 1. This was done within each gender/specimen type stratum and chronologically based on the individual specimen's arrival date at the NPHL. Using the 2009 individual testing results, acknowledging that these are potentially incorrect, we first simulated the individual true responses $\tilde{\mathbf{Y}}_{ik} = (\tilde{Y}_{i1k}, \tilde{Y}_{i2k})'$; a description of how this was done is given in the Web-based Supporting Materials. Observed diagnoses from the pooling algorithm in Section 2 were then simulated using $Z_{jk} \sim \text{Bernoulli}\{S_{e,j}Z_{jk} + S_{p,j}(1 - Z_{jk})\}$ and $Y_{ijk} \sim \text{Bernoulli}\{S_{e,j}\tilde{Y}_{ijk} + S_{p,j}(1 - \tilde{Y}_{ijk})\}$ using the Aptima Combo 2 Assay pilot study point estimates for $S_{e,j}$ and $S_{p,j}$. To average over the effect of simulation error, we repeated this exercise $B = 500$ times, leaving us with 500 simulated group testing data sets for each gender/specimen type stratum in 2009.

For each stratum and for each choice of $a_0 \in \{0, 0.5, 1\}$, Table 2 shows the 2009 estimation results for the population-level parameters

$$p_{00} = \text{proportion of individuals negative for both CT and NG}$$

p_{10} = proportion of individuals positive for CT, but negative for NG

p_{01} = proportion of individuals negative for CT, but positive for NG

p_{11} = proportion of individuals positive for both CT and NG.

We used our estimation approach in Section 3 to calculate posterior medians for each of the 500 data sets; the results shown in Table 2 are averages across the data sets. For comparison purposes, we also show the maximum likelihood (ML) results averaged over the same 500 data sets. Table 3 shows the 2009 (Bayesian) estimation results for $\delta = (S_{e,1}, S_{e,2}, S_{p,1}, S_{p,2})'$ under the same settings. Recall that for each data set, our Bayesian approach estimates $\mathbf{p}$ and δ simultaneously whereas the ML approach from Tebbs *et al.* [24] regards δ as fixed and known.

Table 2 shows that the averaged ML estimates and the averaged Bayesian posterior median estimates are very similar in most settings. For the male/swab and female/urine cohorts, the average posterior median estimates are slightly different than the average ML estimates when the 2008 historical information (Table 1) is weighed more heavily; i.e., the posterior estimates are more strongly attracted to the historical estimates when a_0 increases. The column in Table 2 labeled “SE” is the averaged standard error of the 500 ML estimates, and the column labeled “BSE” is the analogous Bayesian measure of variability which we calculated as follows. For each data set, the entire Gibbs chain included 3500 iterates; the first 500 were discarded (as in Section 4) and the sample standard deviation was calculated using every 6th iterate from the last 3000. Thinning was used to remove autocorrelation from the successive iterates and left us with a within-data set measure of variation based on 500 posterior draws. BSE was then calculated as the average of these standard deviations across the $B = 500$ data sets, a measure that can be compared with SE.

One of the surprising discoveries from Tebbs *et al.* [24] was that, despite the reduction in the number of tests needed, ML estimates of $\mathbf{p}$ from using group testing were more efficient than those from individual testing; see [14] for discussion on this same phenomenon in single disease settings. Comparing the SE and BSE columns in Table 2, one notes that further efficiency gains are possible using our Bayesian approach when historical information on the prevalence parameter is utilized (i.e., when $a_0 = 0.5$ and when $a_0 = 1$). When the 2008 information is ignored ($a_0 = 0$), it is not surprising that the Bayesian estimates are less precise than the ML estimates. However, they are nearly as precise and this is despite the fact that the assay accuracy parameters in δ are estimated simultaneously along with the prevalence parameters in $\mathbf{p}$ (see Table 3). One might conjecture that the informative priors we used for δ in Table 1 may be partly responsible for the increased efficiency in estimation when historical information about $\mathbf{p}$ is incorporated. However, we have repeated this same analysis using flat priors for $S_{e,j}$ and $S_{p,j}$ and have discovered largely the same findings. The tables summarizing this analysis are shown in the Web-based Supporting Materials.

6. Discussion

We have presented a Bayesian approach to estimate the prevalence of multiple diseases from group testing data, motivated by nationwide CT/NG screening activities to identify infected

individuals and to estimate the prevalence of each infection. Our methods allow researchers to incorporate uncertainty in diagnostic assays and also information from previous periods of screening. When compared to the estimation methods in Tebbs *et al.* [24], our approach is more flexible; one can obtain estimates of the prevalence parameters while simultaneously estimating assay accuracy. Furthermore, our Bayesian estimates are nearly as precise as ML estimates when no historical information is provided and are potentially more precise otherwise. The web site www.chrisbilder.com/grouptesting/ contains R programs that implement our estimation techniques for the algorithm in Section 2. Our programs determine appropriate beta prior distributions for δ based on pilot studies like those described in Section 3.2. The Web-based Supporting Materials show explicitly how these prior distributions can be constructed.

Our augmented likelihood function $\pi(\mathbf{Z}, \mathbf{Y}, \tilde{\mathbf{Y}}|\theta)$ in Equation (1) is constructed by making two simplifying assumptions. First, we assume that all testing outcomes associated with the same master pool (including any individual retest responses from it) are conditionally independent given the true individual statuses in the pool. This assumption is pervasive in the group testing literature for single infections and may be reasonable when misclassification is driven primarily by factors related to test implementation. Second, at both the master pool (Stage 1) and individual testing (Stage 2) levels, we assume that $S_{e,j}$ and $S_{p,j}$ for one infection do not depend on the true status of the other infection. Future research in group testing could investigate ways to avoid making either or both assumptions. For single infections, Wang et al. [32] relax the conditional independence assumption by relating binary testing outcomes to latent biomarker levels. Relaxing the second assumption with multiple infections might be possible by borrowing ideas from the recent causal inference literature [33].

Although we have focused on screening for CT/NG, the group testing algorithm in Section 2 could be implemented in other situations when a discriminating assay is used to test for two or more infections. For example, the CDC has recently proposed that a discriminating assay for HIV-1 and HIV-2 replace the more traditional Western Blot-type assay to improve the detection of acute HIV infections [34, 35]. Furthermore, discriminating multiplex assays, such as the Procleix Ultrio Plus Assay (Hologic/Gen-Probe, Inc., San Diego), are currently available to simultaneously detect HBV, HCV, and HIV in pooled and individual specimens. Generalizing our techniques to estimate the joint prevalence of more than two infections using the algorithm in Section 2 is straightforward; see the Web-based Supporting Materials.

Our work has assumed that a discriminating multiplex assay is available and that it can be applied to both pools of specimens and individual specimens. In other situations, especially those involving screening donated blood for HBV, HCV, and HIV [4, 36, 37, 38], a non-discriminating multiplex assay (i.e., an assay whose result indicates the presence of at least one infection or the absence of all infections) is typically used to test pools in the first stage. Positive pools are resolved in the next stage by applying the same multiplex assay to individuals (to determine which individuals are positive for at least one infection) and then confirmatory, infection-specific assays to positive individuals to diagnose the presence/absence of each infection. In our review of the infectious disease testing literature, we have found that there are numerous testing algorithms currently used with multiplex assays

—discriminating and not—for two or more infections. An advantage of the Bayesian framework outlined in Section 3 is that it can be suitably adapted to handle different algorithms with little added difficulty.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the Associate Editor and three anonymous referees for their insightful comments on an earlier version of this article. Implementing their suggestions improved our presentation significantly. We also thank Lucy DesJardin, Kristofer Eveland, Jeff Benfer, Peter Iwen, and Steven Hinrichs for their consultation on CT/NG testing in Iowa and Nebraska and Elizabeth Torrone at the Centers for Disease Control and Prevention for her consultation on CT/NG screening programs nationwide. Timothy Hanson offered helpful advice throughout the process of writing this article for which we are very appreciative. This research was supported by Grant R01 AI121351 from the National Institutes of Health.

References

1. Dorfman R. The detection of defective members of large populations. *Annals of Mathematical Statistics*. 1943; 14:436–440.
2. 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, Leone P. Detection of acute infections during HIV testing in North Carolina. *New England Journal of Medicine*. 2005; 352:1873–1883. [PubMed: 15872202]
3. 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, Wright D. 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*. 2008; 48:1558–1566. [PubMed: 18466173]
4. Stramer S, Notari E, Krysztof D, Dodd R. Hepatitis B virus testing by minipool nucleic acid testing: Does it improve blood safety? *Transfusion*. 2013; 53:2449–2458. [PubMed: 23607261]
5. Lindan C, Mathur M, Kumta S, Jerajani H, Gogate A, Schachter J, Moncada J. Utility of pooled urine specimens for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in men attending public sexually transmitted infection clinics in Mumbai, India, by PCR. *Journal of Clinical Microbiology*. 2005; 43:1674–1677. [PubMed: 15814983]
6. Busch M, Caglioti S, Robertson E, McAuley J, Tobler L, Kamel H, Linnen J, Shyamala V, Tomasulo P, Kleinman S. Screening the blood supply for West Nile virus RNA by nucleic acid amplification testing. *New England Journal of Medicine*. 2005; 353:460–467. [PubMed: 16079369]
7. Wang B, Han S, Cho C, Han J, Cheng Y, Lee S, Galappaththy G, Thimasarn K, Soe M, Oo H, Kyaw M, Han E. Comparison of microscopy, nested PCR, and real-time PCR assays using high-throughput screening of pooled samples for diagnosis of malaria in asymptomatic carriers from areas of endemicity in Myanmar. *Journal of Clinical Microbiology*. 2014; 52:1838–1845. [PubMed: 24648557]
8. Van T, Miller J, Warshauer D, Reisdorf E, Jerrigan D, Humes R, Shult P. Pooling nasopharyngeal/throat swab specimens to increase testing capacity for influenza viruses by PCR. *Journal of Clinical Microbiology*. 2012; 50:891–896. [PubMed: 22205820]
9. Remlinger K, Hughes-Oliver J, Young S, Lam R. Statistical design of pools using optimal coverage and minimal collision. *Technometrics*. 2006; 48:133–143.
10. Chi X, Lou X, Yang M, Shu Q. An optimal DNA pooling strategy for progressive fine mapping. *Genetica*. 2009; 135:267–281. [PubMed: 18506582]

11. Dhand N, Johnson W, Toribio J. A Bayesian approach to estimate OJD prevalence from pooled fecal samples of variable pool size. *Journal of Agricultural, Biological, and Environmental Statistics*. 2010; 15:452–473.
12. Fahey J, Ourisson P, Degnan F. Pathogen detection, testing, and control in fresh broccoli sprouts. *Nutrition Journal*. 2006; 5:13. [PubMed: 16630354]
13. Kim H, Hudgens M, Dreyfuss J, Westreich D, Pilcher C. Comparison of group testing algorithms for case identification in the presence of testing error. *Biometrics*. 2007; 63:1152–1163. [PubMed: 17501946]
14. Liu A, Liu C, Zhang Z, Albert P. Optimality of group testing in the presence of misclassification. *Biometrika*. 2012; 99:245–251. [PubMed: 23049137]
15. Zhang B, Bilder C, Tebbs J. Group testing regression model estimation when case identification is a goal. *Biometrical Journal*. 2013; 55:173–189. [PubMed: 23401252]
16. Vansteelandt S, Goetghebeur E, Verstraeten T. Regression models for disease prevalence with diagnostic tests on pools of serum samples. *Biometrics*. 2000; 56:1126–1133. [PubMed: 11129470]
17. Chen P, Tebbs J, Bilder C. Group testing regression models with fixed and random effects. *Biometrics*. 2009; 65:1270–1278. [PubMed: 19210734]
18. Wang D, McMahan C, Gallagher C, Kulasekera K. Semiparametric group testing regression models. *Biometrika*. 2014; 101:587–598.
19. Delaigle A, Meister A. Nonparametric regression analysis for group testing data. *Journal of the American Statistical Association*. 2011; 106:640–650.
20. Gaydos C, Cartwright C, Colaninno P, Welsch J, Holden J, Yo S, Webb E, Anderson C, Bertuzis R, Zhang L, Miller T, Leckie G, Abravaya K, Robinson J. Performance of the Abbott RealTime CT/NG for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae*. *Journal of Clinical Microbiology*. 2010; 48:3236–3243. [PubMed: 20668135]
21. Jirsa, S. Pooling specimens: A decade of successful cost savings. National STD Prevention Conference; 2008; Chicago, IL.
22. Lewis J, Lockary V, Kobic S. Cost savings and increased efficiency using a stratified specimen pooling strategy for *Chlamydia trachomatis* and *Neisseria gonorrhoeae*. *Sexually Transmitted Diseases*. 2012; 39:46–48. [PubMed: 22183846]
23. JSI Research & Training Institute, Inc./Denver. The Future of Infertility Prevention Project Health Impact Assessment: Policy Implications and Recommendations in Light of Passage of the Patient Protection and Affordable Care Act. 2012 Jul 25. [Last accessed December 15, 2015] Available at <http://www.jsi.com>.
24. Tebbs J, McMahan C, Bilder C. Two-stage hierarchical group testing for multiple infections with application to the Infertility Prevention Project. *Biometrics*. 2013; 69:1064–1073. [PubMed: 24117173]
25. Hughes-Oliver J, Rosenberger W. Efficient estimation of the prevalence of multiple rare traits. *Biometrika*. 2000; 87:315–327.
26. Centers for Disease Control and Prevention. [Last accessed December 15, 2015] Recommendations for the Laboratory-Based Detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae*. *Morbidity and Mortality Weekly Report*. 2014 Mar 14. Available at <http://www.cdc.gov/mmwr>
27. Johnson W, Pearson L. Dual screening. *Biometrics*. 1999; 55:867–873. [PubMed: 11315019]
28. Hanson T, Johnson W, Gastwirth J. Bayesian inference for prevalence and diagnostic test accuracy based on dual-pooled screening. *Biostatistics*. 2006; 7:41–57. [PubMed: 15947009]
29. McMahan C, Tebbs J, Bilder C. Regression models for group testing data with pool dilution effects. *Biostatistics*. 2013; 14:284–298. [PubMed: 23197382]
30. Ibrahim J, Chen M, Gwon Y, Chen F. The power prior: Theory and applications. *Statistics in Medicine*. 2015; 34:3724–3749. [PubMed: 26346180]
31. Zhang Z, Liu C, Kim S, Liu A. Prevalence estimation subject to misclassification: The mis-substitution bias and some remedies. *Statistics in Medicine*. 2014; 33:4482–4500. [PubMed: 25043925]
32. Wang D, McMahan C, Gallagher C. A general parametric regression framework for group testing data with dilution effects. *Statistics in Medicine*. 2015; 34:3606–3621. [PubMed: 26173957]

33. Hudgens M, Halloran M. Towards causal inference with interference. *Journal of the American Statistical Association*. 2008; 103:832–842. [PubMed: 19081744]
34. Branson B, Mermin J. Establishing the diagnosis of HIV infection: New tests and a new algorithm for the United States. *Journal of Clinical Virology*. 2011; 52:S3–S4. [PubMed: 21993308]
35. Krajden M, Cook D, Mak A, Chu K, Chahil N, Steinberg M, Rekart M, Gilbert M. Pooled nucleic acid testing increases the diagnostic yield of acute HIV infections in a high-risk population compared to 3rd and 4th generation HIV enzyme immunoassays. *Journal of Clinical Virology*. 2014; 61:132–137. DOI: <http://dx.doi.org/10.1016/j.jcv.2014.06.024>. [PubMed: 25037533]
36. O'Brien S, Yi Q, Fan W, Scalia V, Fearon M, Allain J. Current incidence and residual risk of HIV, HBV and HCV at Canadian Blood Services. *Vox Sanguinis*. 2012; 103:83–86. [PubMed: 22289147]
37. Schmidt M, Pichl L, Jork C, Hourfar M, Schottstedt V, Wagner F, Seifried E, Müller T, Bux J, Saldanha J. Blood donor screening with cobas s 201/cobas TaqScreen MPX under routine conditions at German Red Cross institutes. *Vox Sanguinis*. 2010; 98:37–46. [PubMed: 19682348]
38. Stramer S, Wend U, Candotti D, Foster G, Hollinger F, Dodd R, Allain J, Gerlich W. Nucleic acid testing to detect HBV infection in blood donors. *New England Journal of Medicine*. 2011; 364:236–247. [PubMed: 21247314]

Figure 1.

Simulation results with $N = 1000$ individuals, $\mathbf{p} = (0.80, 0.10, 0.09, 0.01)$, $S_{e,j} = 0.95$, and $S_{p,j} = 0.99$, for $j = 1, 2$. The 5th (bottom), 25th, 50th (median), 75th, and 95th (top) percentiles of the $B = 500$ posterior median estimates of $\mathbf{p}$ are provided. Flat priors for $S_{e,j}$ and $S_{p,j}$ are used; i.e., $S_{e,j} \sim \text{beta}(1, 1)$ and $S_{p,j} \sim \text{beta}(1, 1)$. The precision parameter a_0 increases from 0 (no historical information about $\mathbf{p}$ provided) to 1 by increments of 0.1.

Figure 2. Simulation results with $N = 1000$ individuals, $\mathbf{p} = (0.80, 0.10, 0.09, 0.01)$, $S_{e,j} = 0.95$, and $S_{p,j} = 0.99$, for $j = 1, 2$. The 5th (bottom), 25th, 50th (median), 75th, and 95th (top) percentiles of the $B = 500$ posterior median estimates of $S_{e,j}$ and $S_{p,j}$ are provided. Flat priors for $S_{e,j}$ and $S_{p,j}$ are used; i.e., $S_{e,j} \sim \text{beta}(1, 1)$ and $S_{p,j} \sim \text{beta}(1, 1)$. The precision parameter a_0 increases from 0 (no historical information about $\mathbf{p}$ provided) to 1 by increments of 0.1.

Nebraska 2008 historical information for CT/NG. The historical estimate $\mathbf{p}_0 = (p_{00(0)}, p_{10(0)}, p_{01(0)}, p_{11(0)})'$ was calculated using the 2008 individual diagnoses (accounting for possible misclassification; see the Web-based Supporting Materials). Stratum sample sizes N_0 are given. Priors for $S_{e,j}$ and $S_{p,j}$ were determined using pilot data from the Aptima Combo 2 Assay product literature (see the Web-based Supporting Materials). The master pool size c_k^* minimizes the expected number of tests per individual as described in [24].

Table 1

Stratum	CT	NG	$\mathbf{p}_0$	Prior for S_e	Prior for S_p	Pool size
Male/Urine $N_0 = 3541$	-	-	$p_{00(0)} = 0.929$			
	+	-	$p_{10(0)} = 0.061$	CT: beta(277, 7)	CT: beta(802, 13)	$c_k^* = 4$
	-	+	$p_{01(0)} = 0.005$	NG: beta(325, 6)	NG: beta(803, 4)	
	+	+	$p_{11(0)} = 0.004$			
Male/Swab $N_0 = 2826$	-	-	$p_{00(0)} = 0.848$			
	+	-	$p_{10(0)} = 0.103$	CT: beta(261, 12)	CT: beta(775, 21)	$c_k^* = 3$
	-	+	$p_{01(0)} = 0.032$	NG: beta(320, 4)	NG: beta(765, 18)	
	+	+	$p_{11(0)} = 0.016$			
Female/Urine $N_0 = 2338$	-	-	$p_{00(0)} = 0.907$			
	+	-	$p_{10(0)} = 0.074$	CT: beta(198, 12)	CT: beta(1171, 14)	$c_k^* = 4$
	-	+	$p_{01(0)} = 0.006$	NG: beta(117, 12)	NG: beta(1348, 11)	
	+	+	$p_{11(0)} = 0.013$			
Female/Swab $N_0 = 14441$	-	-	$p_{00(0)} = 0.948$			
	+	-	$p_{10(0)} = 0.047$	CT: beta(196, 13)	CT: beta(1155, 29)	$c_k^* = 5$
	-	+	$p_{01(0)} = 0.001$	NG: beta(127, 2)	NG: beta(1336, 18)	
	+	+	$p_{11(0)} = 0.005$			

Table 2

Nebraska CT/NG prevalence estimation results for 2009. Bayesian estimates (Bayes) are posterior medians averaged over $B = 500$ data sets; BSE is the average of the standard deviations calculated from posterior samples of the $B = 500$ data sets. Values of $a_0 = 0$, $a_0 = 0.5$, and $a_0 = 1$ are used to incorporate different amounts of historical information for $\mathbf{p}$ as described in Section 5. Prior distributions for $S_{e,j}$ and $S_{p,j}$ where $j = 1$ for CT and $j = 2$ for NG, are given in Table 1. Maximum likelihood estimates, calculated from [24], are averaged over the same 500 data sets; the entries under SE are the averaged standard errors. Stratum sample sizes N are given.

Stratum	CT	NG	Maximum likelihood		Bayes ($a_0 = 0$)		Bayes ($a_0 = 0.5$)		Bayes ($a_0 = 1$)	
			Estimate	SE	Estimate	BSE	Estimate	BSE	Estimate	BSE
Male/Urine $N = 6139$	-	-	$\hat{p}_{00} = 0.924$	0.0035	$\hat{p}_{00} = 0.923$	0.0037	$\hat{p}_{00} = 0.926$	0.0029	$\hat{p}_{00} = 0.927$	0.0025
	+	-	$\hat{p}_{10} = 0.061$	0.0032	$\hat{p}_{10} = 0.061$	0.0034	$\hat{p}_{10} = 0.061$	0.0027	$\hat{p}_{10} = 0.061$	0.0023
	-	+	$\hat{p}_{01} = 0.008$	0.0012	$\hat{p}_{01} = 0.008$	0.0012	$\hat{p}_{01} = 0.007$	0.0009	$\hat{p}_{01} = 0.006$	0.0007
	+	+	$\hat{p}_{11} = 0.007$	0.0011	$\hat{p}_{11} = 0.007$	0.0011	$\hat{p}_{11} = 0.006$	0.0008	$\hat{p}_{11} = 0.006$	0.0007
Male/Swab $N = 1910$	-	-	$\hat{p}_{00} = 0.831$	0.0091	$\hat{p}_{00} = 0.831$	0.0096	$\hat{p}_{00} = 0.837$	0.0074	$\hat{p}_{00} = 0.841$	0.0062
	+	-	$\hat{p}_{10} = 0.119$	0.0079	$\hat{p}_{10} = 0.119$	0.0085	$\hat{p}_{10} = 0.113$	0.0064	$\hat{p}_{10} = 0.111$	0.0054
	-	+	$\hat{p}_{01} = 0.034$	0.0043	$\hat{p}_{01} = 0.034$	0.0044	$\hat{p}_{01} = 0.033$	0.0035	$\hat{p}_{01} = 0.033$	0.0030
	+	+	$\hat{p}_{11} = 0.015$	0.0029	$\hat{p}_{11} = 0.015$	0.0030	$\hat{p}_{11} = 0.016$	0.0024	$\hat{p}_{11} = 0.016$	0.0021
Female/Urine $N = 4972$	-	-	$\hat{p}_{00} = 0.920$	0.0041	$\hat{p}_{00} = 0.919$	0.0044	$\hat{p}_{00} = 0.915$	0.0035	$\hat{p}_{00} = 0.913$	0.0030
	+	-	$\hat{p}_{10} = 0.066$	0.0038	$\hat{p}_{10} = 0.067$	0.0041	$\hat{p}_{10} = 0.069$	0.0032	$\hat{p}_{10} = 0.071$	0.0028
	-	+	$\hat{p}_{01} = 0.004$	0.0010	$\hat{p}_{01} = 0.004$	0.0011	$\hat{p}_{01} = 0.005$	0.0009	$\hat{p}_{01} = 0.005$	0.0008
	+	+	$\hat{p}_{11} = 0.009$	0.0014	$\hat{p}_{11} = 0.009$	0.0015	$\hat{p}_{11} = 0.011$	0.0012	$\hat{p}_{11} = 0.011$	0.0011
Female/Swab $N = 14530$	-	-	$\hat{p}_{00} = 0.949$	0.0019	$\hat{p}_{00} = 0.949$	0.0023	$\hat{p}_{00} = 0.949$	0.0017	$\hat{p}_{00} = 0.948$	0.0015
	+	-	$\hat{p}_{10} = 0.045$	0.0019	$\hat{p}_{10} = 0.045$	0.0023	$\hat{p}_{10} = 0.046$	0.0017	$\hat{p}_{10} = 0.047$	0.0014
	-	+	$\hat{p}_{01} = 0.001$	0.0001	$\hat{p}_{01} = 0.001$	0.0002	$\hat{p}_{01} = 0.001$	0.0001	$\hat{p}_{01} = 0.001$	0.0001
	+	+	$\hat{p}_{11} = 0.005$	0.0006	$\hat{p}_{11} = 0.005$	0.0006	$\hat{p}_{11} = 0.005$	0.0005	$\hat{p}_{11} = 0.005$	0.0004

Table 3

Bayesian assay accuracy estimates from 2009. Bayesian estimates (Bayes) are posterior medians averaged over $B = 500$ data sets; BSE is the average of the standard deviations calculated from posterior samples of the $B = 500$ data sets. Values of $a_0 = 0$, $a_0 = 0.5$, and $a_0 = 1$ are used to incorporate different amounts of historical information for $\mathbf{p}$ as described in Section 5. Prior distributions for $S_{e,j}$ and $S_{p,j}$ where $j = 1$ for CT and $j = 2$ for NG, are given in Table 1. Stratum sample sizes N are given.

Stratum	Accuracy	Bayes ($a_0 = 0$)		Bayes ($a_0 = 0.5$)		Bayes ($a_0 = 1$)	
		Estimate	BSE	Estimate	BSE	Estimate	BSE
Male/Urine	S_{e1}	0.979	0.0076	$\hat{S}_{e1} = 0.978$	0.0074	$\hat{S}_{e1} = 0.978$	0.0073
	S_{e2}	0.985	0.0069	$\hat{S}_{e2} = 0.984$	0.0066	$\hat{S}_{e2} = 0.984$	0.0065
	S_{p1}	0.985	0.0030	$\hat{S}_{p1} = 0.985$	0.0030	$\hat{S}_{p1} = 0.984$	0.0030
	S_{p2}	0.996	0.0012	$\hat{S}_{p2} = 0.996$	0.0012	$\hat{S}_{p2} = 0.996$	0.0012
Male/Swab	S_{e1}	0.959	0.0107	$\hat{S}_{e1} = 0.960$	0.0102	$\hat{S}_{e1} = 0.961$	0.0099
	S_{e2}	0.991	0.0061	$\hat{S}_{e2} = 0.989$	0.0060	$\hat{S}_{e2} = 0.989$	0.0060
	S_{p1}	0.975	0.0048	$\hat{S}_{p1} = 0.973$	0.0049	$\hat{S}_{p1} = 0.973$	0.0049
	S_{p2}	0.978	0.0034	$\hat{S}_{p2} = 0.979$	0.0035	$\hat{S}_{p2} = 0.979$	0.0035
Female/Urine	S_{e1}	0.947	0.0115	$\hat{S}_{e1} = 0.944$	0.0114	$\hat{S}_{e1} = 0.942$	0.0115
	S_{e2}	0.913	0.0217	$\hat{S}_{e2} = 0.909$	0.0220	$\hat{S}_{e2} = 0.904$	0.0222
	S_{p1}	0.989	0.0026	$\hat{S}_{p1} = 0.989$	0.0025	$\hat{S}_{p1} = 0.989$	0.0025
	S_{p2}	0.993	0.0016	$\hat{S}_{p2} = 0.993$	0.0015	$\hat{S}_{p2} = 0.993$	0.0015
Female/Swab	S_{e1}	0.942	0.0112	$\hat{S}_{e1} = 0.939$	0.0107	$\hat{S}_{e1} = 0.938$	0.0106
	S_{e2}	0.992	0.0104	$\hat{S}_{e2} = 0.986$	0.0098	$\hat{S}_{e2} = 0.987$	0.0097
	S_{p1}	0.976	0.0029	$\hat{S}_{p1} = 0.976$	0.0028	$\hat{S}_{p1} = 0.976$	0.0028
	S_{p2}	0.987	0.0013	$\hat{S}_{p2} = 0.987$	0.0014	$\hat{S}_{p2} = 0.987$	0.0013