Published in final edited form as:
Biometrics. 2013 December ; 69(4): 1064–1073. doi:10.1111/biom.12080.

Two-stage hierarchical group testing for multiple infections with application to the Infertility Prevention Project

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

[1]Department of Statistics, University of South Carolina, Columbia, SC 29208, U.S.A

[2]Department of Mathematical Sciences, Clemson University, Clemson, SC 29634, U.S.A

[3]Department of Statistics, University of Nebraska-Lincoln, Lincoln, NE 68583, U.S.A

Summary

Screening for sexually transmitted diseases has benefited greatly from the use of group testing (pooled testing) to lower costs. With the development of assays that detect multiple infections, screening practices now involve testing pools of individuals for multiple infections simultaneously. Building on the research for single infection group testing procedures, we examine the performance of group testing for multiple infections. Our work is motivated by chlamydia and gonorrhea testing for the Infertility Prevention Project (IPP), a national program in the United States. We consider a two-stage pooling algorithm currently used to perform testing for the IPP. We first derive the operating characteristics of this algorithm for classification purposes (e.g., expected number of tests, misclassification probabilities, etc.) and identify pool sizes that minimize the expected number of tests. We then develop an expectation-maximization algorithm to estimate probabilities of infection using both group and individual retest responses. Our research shows that group testing can offer large cost savings when classifying individuals for multiple infections and can provide prevalence estimates that are actually more efficient than those from individual testing.

Keywords

Correlated binary response; Efficiency; EM algorithm; Pooled testing; Screening; Specimen pooling

1. Introduction

The Infertility Prevention Project (IPP) is a national program funded by the Centers for Disease Control and Prevention (CDC) and the Department of Health and Human Services (HHS). The primary goals of the IPP are to identify individuals who are infected with chlamydia and/or gonorrhea, to monitor trends in prevalence, and to provide treatment for those infected. The dangers of chlamydia and gonorrhea, both bacterial infections, are their potentially serious sequelae, which include pelvic inflammatory disease, ectopic pregnancy,

[*]tebbs@stat.sc.edu.

Supplementary Materials: The Web Appendices referenced in Sections 3–5 and 7 are available with this paper at the Biometrics website on Wiley Online Library.

sterility, and infertility (Lewis, Lockary, and Kobic, 2012). In addition, both are believed to facilitate the transmission of other sexually transmitted diseases (STDs), including HIV and human papillomavirus infection (Farley, Cohen, and Elkins, 2003; Samoff et al., 2005). In 2011 alone, there were approximately 1.7 million new cases of chlamydia and gonorrhea reported in the United States. However, because both infections are largely asymptomatic, it is believed that even a greater number of cases go unreported each year (CDC, 2012).

The IPP started as a CDC trial program in 1988, carried out initially in Alaska, Idaho, Oregon, and Washington. It soon after spread to all 50 states, operating within each of the 10 federal HHS regions. The manner in which individuals are tested for chlamydia and gonorrhea as part of the IPP has always varied from state to state. The state of Iowa currently uses group testing (pooled testing), where individual specimens are tested in pools for both infections simultaneously. Pools that test negative are declared to contain all negative individuals. Pools that test positive for either infection are resolved (or "decoded") by testing each specimen individually. The practical motivation for pooling is that it can offer substantial cost savings. For example, the Iowa IPP has reported over $2.2 million in savings since switching from individual testing to group testing in 1999 (Jirsa, 2008).

Dorfman (1943) first proposed group testing as a way to screen military inductees for syphilis during World War II. Since his seminal work, pooling has been used to test for a variety of STDs, including HIV and hepatitis B/C (Cardoso, Koerner, and Kubanek, 1998; Pilcher et al., 2005), elsewhere for chlamydia and gonorrhea (Lindan et al., 2005), and for other infections like West Nile Virus (Busch et al., 2005) and H1N1 influenza virus (Van et al., 2012). Statistical research in group testing generally splits into two areas: classification and estimation. The classification problem deals with case identification; i.e., the identification of each individual as being positive or negative (Kim et al., 2007). The estimation problem deals with estimating the overall probability of infection in a population (Hughes-Oliver and Swallow, 1994; Liu et al., 2012) or subject-specific probabilities using regression (Vansteelandt, Goetghebeur, and Verstraeten, 2000; Xie, 2001; Delaigle and Meister, 2011). Both problems are of interest to states as part of the IPP. Classification is needed to diagnose individuals for treatment purposes and estimation is needed to monitor disease prevalence.

While classification and estimation have received a large amount of attention in group testing, this research has been limited largely to a single disease. However, in addition to the IPP, the infectious disease literature is replete with applications where individuals are tested in pools for multiple infections simultaneously. For example, the American Red Cross (ARC) uses group testing to screen millions of blood donations each year for HIV, hepatitis B, and hepatitis C (ARC, 2013), as do Red Cross organizations in Japan and Germany (Mine et al., 2003; Hourfar et al., 2008). The only statistical research that has examined multiple infections in group testing is Hughes-Oliver and Rosenberger (2000), who investigate optimal design for estimation. However, this work assumes that a perfect assay test is available, which is not realistic in the IPP, and the authors do not consider classification, which is needed to begin treatment for infected individuals and to help prevent the spread of future infections.

In this paper, we examine both classification and estimation for group testing with multiple infections, motivated by current IPP screening practices described in Section 2. To the best of our knowledge, this is the first research paper that examines classification for multiple infections with group testing. The estimation part expands on the work in Hughes-Oliver and Rosenberger (2000), by allowing for imperfect testing and also for the inclusion of retest responses that arise naturally as part of the classification process. As laboratories across the United States continue to see their federal funding reduced, the group testing methodology outlined in this paper could allow them to reduce the cost of screening for multiple infections while maintaining their current testing loads.

In Section 2, we describe the pooling algorithm used by the state of Iowa as part of the IPP. In Section 3, we derive the operating characteristics of this algorithm, including expressions for the expected number of tests and classification accuracy measures, and we discuss optimal pool size selection. In Section 4, we develop an expectation-maximization (EM) algorithm to estimate multiple disease prevalences. This procedure is substantially different than the one outlined in Hughes-Oliver and Rosenberger (2000), which uses only the responses from initial pools. In Section 5, we use simulation to investigate small sample characteristics of the estimators in Section 4. In Section 6, we apply our methods to data from the IPP. In Section 7, we discuss extensions to more than two infections and future areas of research.

2. IPP Pooling Algorithm

We now describe the pooling protocol used by the University of Iowa Hygienic Laboratory (UIHL). As per IPP guidelines, individual specimens are collected from individuals across the state and are shipped to the UIHL to be tested for both chlamydia and gonorrhea. Upon arrival, individual specimens are cross-classified by gender and specimen type (urine or swab). The IPP pooling algorithm is outlined below.

IPP POOLING ALGORITHM

1. Individual specimens are randomly assigned to (master) pools of size $c > 1$.
2. Each pool is tested for both infections using a single assay (i.e., a single assay detects both infections simultaneously).
3.
   i. Individuals in pools that test negative for **both** infections are diagnosed as negative for both infections.

   ii. Individuals in pools that test positive for **either** infection are retested (individually) for both infections using the same assay in Step 2. Diagnoses for both infections are made from the outcomes of the individual tests.

Several comments are in order. First, it should be noted that only female swab specimens are tested using this procedure in Iowa; all other specimens are tested individually. Because males are more likely to be tested only when symptoms are present (e.g., painful urination, etc.), it is believed that the proportion of positive male specimens received by the UIHL is too large to make pooling worthwhile. On the other hand, females are tested routinely as part of annual checkups and pregnancy examinations (pooling female urine samples has been considered by the UIHL but has not been implemented). Other states use this pooling protocol for each of the four gender/specimen type strata; for example, see Lewis et al. (2012) for a description of IPP screening practices in Idaho.

Second, an obvious question arises regarding the master pool size in Step 1, namely, what is the "best" value for c? The UIHL uses master pools of size $c = 4$ because this is how the pooling algorithm was originally calibrated with the automated TECAN DTS platform currently in place. However, a different master pool size might reduce the total number tests needed and therefore be more cost efficient. Pool size selection has received a large amount of attention in the statistics literature for single infections (Hughes-Oliver and Swallow, 1994; Kim et al., 2007; Liu et al., 2012). In addition, Hughes-Oliver and Rosenberger (2000) consider pool size selection for multiple infections, but only for estimation.

Third, the same assay is used (a) to test for both chlamydia and gonorrhea and (b) to test both master pools in Step 2 and individuals in Step 3(ii). Along with other states, the UIHL currently uses the GenProbe Aptima Combo 2 Assay nucleic acid amplification test (GenProbe, San Diego), which simultaneously detects the presence of chlamydia and gonorrhea in both pooled and individual samples. Therefore, instead of administering separate, infection-specific assays in Step 3(ii), UIHL officials have judged it simpler (and more cost efficient) to use this same dual-infection assay when retesting individuals-even when a master pool tests positive for only one infection. Specimens are carefully prepared by the UIHL to ensure that testing error rates are the same for both pooled and individual samples.

In this paper, we derive the classification characteristics of the pooling procedure described above, and we develop an EM algorithm to estimate population level prevalences with the observed data from the procedure. Unlike classification and estimation research with single infections, we deal with multiple disease statuses on the same individual, which are not necessarily observed (due to pooling and assay testing error) but are likely correlated. Our work is potentially applicable for any laboratory that tests for chlamydia and gonorrhea (as part of the IPP or otherwise) and for others that screen for multiple infections.

3. Classification

3.1. Notation and assumptions

Suppose N individuals are to be tested and that each individual is initially assigned to one master pool. Let $\tilde{\mathbf{Y}}_{ik} = (\tilde{Y}_{i1k}, \tilde{Y}_{i2k})'$ denote the vector of true individual statuses, where $\tilde{Y}_{ijk} = 1$ if the ith individual in the kth pool is positive for the jth infection, $\tilde{Y}_{ik} = 0$ otherwise, for $i = 1, 2, \ldots, c_k$, $j = 1, 2$, and $k = 1, 2, \ldots, K$. For generality, we allow the master pool sizes c_k to be different across the K pools; in addition, notation that indexes different pools will be helpful in Section 4 when we consider estimation. We assume throughout that the $\tilde{\mathbf{Y}}_{ik}$'s are independent and identically distributed random vectors, but we allow for correlation between the (latent) infection statuses on the same individual, $\tilde{Y}_{i1k}$ and $\tilde{Y}_{i2k}$, and write

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

for $\tilde{y}_1, \tilde{y}_2 \in \{0, 1\}$, where $p_{00} + p_{10} + p_{01} + p_{11} = 1$.

Let $\tilde{\boldsymbol{Z}}_k = \left(\tilde{Z}_{1k}, \tilde{Z}_{2k}\right)'$ denote the vector of true statuses for the kth master pool, where $\tilde{Z}_{jk} = I(\sum_{i=1}^{c_k} \tilde{Y}_{ijk} > 0)$ and $I(\cdot)$ is the indicator function; i.e., $\tilde{Z}_{jk} = 1$ if the kth pool contains at least one individual that is truly positive for the jth infection, $\tilde{Z}_{jk} = 0$ otherwise. To allow for misclassification, let $\mathbf{Z}_k = (Z_{1k}, Z_{2k})'$ denote the vector of testing responses for the kth master pool, where $Z_{jk} = 1$ if this pool tests positive for the jth infection, $Z_{jk} = 0$ otherwise. Define the sensitivity and the specificity for the jth infection by $S_{e:j} = \text{pr}(Z_{jk} = 1 | \tilde{Z}_{jk} = 1)$ and $S_{p:j} = \text{pr}(Z_{jk} = 0 | \tilde{Z}_{jk} = 0)$, respectively, for $j = 1, 2$. We assume that $S_{e:j}$ and $S_{p:j}$ are known and do not depend on the pool size c_k. This assumption is standard in the group testing literature for single infections and proper assay calibration is needed to ensure that this is reasonable in application. We also assume that testing responses are independent, conditional on the true status of the specimen being tested. This assumption is common for single infections in group testing (see, e.g., Kim et al., 2007) and is needed to derive closed-form expressions for the expected number of tests and probabilities of misclassification. Under these assumptions, we show in Web Appendix A that the probability mass function of the master pool testing response $\mathbf{Z}_k = (Z_{1k}, Z_{2k})'$, for $z_1, z_2 \in \{0, 1\}$, is given by

$$\text{pr}(Z_{1k} = z_1, Z_{2k} = z_2) = \sum_{\tilde{z}_1 = 0}^{1} \sum_{\tilde{z}_2 = 0}^{1} \theta_{\tilde{z}_1 \tilde{z}_2} \prod_{j=1}^{2} \left(S_{e:j}^{z_j} \overline{S}_{e:j}^{1-z_j}\right)^{\tilde{z}_j} \left(S_{p:j}^{1-z_j} \overline{S}_{p:j}^{z_j}\right)^{1-\tilde{z}_j},$$

where $\overline{S}_{e:j} = 1 - S_{e:j}, \overline{S}_{p:j} = 1 - S_{p:j}$, and $\theta_{\tilde{z}_1 \tilde{z}_2} = \text{pr}(\tilde{Z}_{1k} = \tilde{z}_1, \tilde{Z}_{2k} = \tilde{z}_2)$, for $\tilde{z}_1, \tilde{z}_2 \in \{0, 1\}$. Straightforward calculations show that $\theta_{00} = p_{00}^{c_k}$, $\theta_{10} = (p_{00} + p_{10})^{c_k} - p_{00}^{c_k}$, $\theta_{01} = (p_{00} + p_{01})^{c_k} - p_{00}^{c_k}$ and $\theta_{11} = 1 - \theta_{00} - \theta_{10} - \theta_{01}$.

3.2. Expected number of tests

An important characteristic of any group testing classification algorithm is the expected number of tests needed to complete it. Let T_k denote the number of tests needed to provide both infection diagnoses for all individuals in the kth master pool. For the pooling algorithm in Section 2, we show in Web Appendix A that

$$\begin{aligned} E(T_k) &= \text{pr}(Z_{1k} = 0, Z_{2k} = 0) + (1 + c_k)\left\{1 - \text{pr}(Z_{1k} = 0, Z_{2k} = 0)\right\} \\ &= 1 + c_k \left\{1 - \overline{S}_{e:1}\overline{S}_{e:2} - \gamma_1\gamma_2 p_{00}^{c_k} + \overline{S}_{e:1}\gamma_2 (p_{00} + p_{10})^{c_k} + \overline{S}_{e:2}\gamma_1 (p_{00} + p_{01})^{c_k}\right\}, \end{aligned} \quad (1)$$

where $\gamma_j = 1 - S_{e:j} - S_{p:j}$, for $j = 1, 2$. From Equation (1), one notes that the expected number of tests depends on the assay sensitivity and specificity for both infections, the pool size c_k, and the individual probabilities p_{00}, p_{10}, and p_{01}. The expected number of tests also depends on the correlation $\rho = \text{corr}(\tilde{Y}_{i1k}, \tilde{Y}_{i2k})$ through the values of p_{00}, p_{10}, and p_{01}. Because $\tilde{Y}_{i1k}$, and $\tilde{Y}_{i2k}$ are binary random variables, ρ satisfies

$$max\left\{-\sqrt{\frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)}}, -\sqrt{\frac{\pi_2(1-\pi_1)}{\pi_1(1-\pi_2)}}\right\} \le \rho \le min\left\{\sqrt{\frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)}}, \sqrt{\frac{\pi_2(1-\pi_1)}{\pi_1(1-\pi_2)}}\right\} \quad (2)$$

(see, e.g., Emrich and Piedmonte, 1991), where π_1 = pr($\tilde{Y}_{i1k}$ = 1) = p_{10} + p_{11} and π_2 = pr($\tilde{Y}_{i2k}$ = 1) = p_{01} + p_{11} are the marginal infection probabilities. That is, the correlation ρ is not unrestricted in [−1, 1] unless $\pi_1 = \pi_2$.

With $E(Tk)$ in closed form, it is possible to determine the master pool size c_k that minimizes the expected number of tests on a per-individual basis. For the algorithm in Section 2, we define the "optimal" pool size to be $c_k^*=arg\ min_{c_k>1} c_k^{-1} E(T_k)$. Figure 1 identifies optimal master pool sizes for different values of π_1 and π_2 when $S_{e:1}$ = 0.942, $S_{p:1}$ = 0.976, $S_{e:2}$ = 0.992, and $S_{p:2}$ = 0.987. These are the values of the sensitivity and specificity (1 = chlamydia, 2 = gonorrhea) associated with the GenProbe Aptima Combo 2 Assay when testing female swab specimens. Figure 1 shows that the optimal pool size c_k^* depends largely on π_1 and π_2, as expected, but only mildly on the correlation ρ. For example, when $\pi_1 = \pi_2$ = 0.04, Figure 1 shows that $c_k^*=4$ when ρ = 0; this optimal size increases to $c_k^*=5$ when $\rho \approx$ 0.17 and again to $c_k^*=6$ only when $\rho \approx$ 0.91. Although the correlation does not largely affect the optimal pool size, it does play an important role in estimation; see Section 4.

3.3. Classification accuracy

In addition to the expected number of tests, it is also important to characterize an algorithm's classification accuracy. We define the *pooling sensitivity (pooling specificity)* for the *j*th infection, denoted by $\mathrm{PS}_{e:j}$ ($\mathrm{PS}_{p:j}$), to be the probability an individual is classified as positive (negative) for the *j*th infection given that the individual is truly positive (negative) for the *j*th infection, *j* = 1, 2. For the algorithm in Section 2, we show in Web Appendix A that

$$\begin{aligned}\mathrm{PS}_{e:1}&=S_{e:1}^2+S_{e:1}\overline{S}_{e:1}\left\{S_{e:2}+\gamma_2 p_{10}(p_{00}+p_{10})^{c_k-1}(p_{10}+p_{11})^{-1}\right\}\\ \mathrm{PS}_{e:2}&=S_{e:2}^2+S_{e:2}\overline{S}_{e:2}\left\{S_{e:1}+\gamma_1 p_{01}(p_{00}+p_{01})^{c_k-1}(p_{01}+p_{11})^{-1}\right\}.\end{aligned}$$

Unfortunately, the expressions for $\mathrm{PS}_{p:1}$ and $\mathrm{PS}_{p:2}$ are not nearly as friendly, but we have derived them to be in closed form; see Web Appendix A. We also define the *pooling positive predictive value* for the *j*th infection, PPV_j, as the probability an individual is truly positive for the *j*th infection, given that the individual has been classified as positive for the *j*th infection, *j* = 1, 2. The *pooling negative predictive value* for the *j*th infection, denoted by NPV_j, is defined similarly for negative individuals. By Bayes' Rule, for *j* = 1, 2,

$$\mathrm{PPV}_j=\frac{\pi_j\mathrm{PS}_{e:j}}{\pi_j\mathrm{PS}_{e:j}+(1-\pi_j)(1-\mathrm{PS}_{p:j})} \text{ and } \mathrm{NPV}_j=\frac{(1-\pi_j)\mathrm{PS}_{p:j}}{(1-\pi_j)\mathrm{PS}_{p:j}+\pi_j(1-\mathrm{PS}_{e:j})}.$$

4. Estimation

Having derived the salient characteristics of the IPP algorithm with respect to classification, we now turn our attention to estimation. Specifically, our goal is to estimate the population probabilities p_{00}, p_{10}, p_{01}, and p_{11} with the observed data from the algorithm.

The observed data consist of (a) the testing responses $\mathbf{Z}_k = (Z_{1k}, Z_{2k})'$ from the K master pools and (b) the additional c_k individual testing responses $\mathbf{Y}_{ik} = (Y_{i1k}, Y_{i2k})'$ from those pools which tested positive for either infection. 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 $\mathbf{Y}$. Because of the correlation between a pool's response and individual retesting responses on the same pool, it is difficult to write out a closed-form expression for the observed data likelihood. We therefore develop an EM algorithm by introducing the individuals' true statuses $\tilde{Y}_{ijk}$, as "missing data." This leads to the complete data likelihood

$$L_C(\boldsymbol{\vartheta}|\boldsymbol{Z},\boldsymbol{Y},\tilde{\boldsymbol{Y}})=\prod_{i=1}^{c_k}\prod_{k=1}^{K}p_{00}^{(1-\tilde{Y}_{i1k})(1-\tilde{Y}_{i2k})}p_{10}^{\tilde{Y}_{i1k}(1-\tilde{Y}_{i2k})}p_{01}^{(1-\tilde{Y}_{i1k})\tilde{Y}_{i2k}}(1-p_{00}-p_{10}-p_{01})^{\tilde{Y}_{i1k}\tilde{Y}_{i2k}}$$
$$\times\prod_{j=1}^{2}\prod_{k=1}^{K}\left(S_{e:j}^{Z_{jk}}\overline{S}_{e:j}^{1-Z_{jk}}\right)^{I\left(\sum_{i=1}^{c_k}\tilde{Y}_{ijk}>0\right)}\left(S_{p:j}^{1-Z_{jk}}\overline{S}_{p:j}^{Z_{jk}}\right)^{I\left(\sum_{i=1}^{c_k}\tilde{Y}_{ijk}=0\right)}$$
$$\times\left\{\prod_{i=1}^{c_k}S_{e:j}^{Y_{ijk}\tilde{Y}_{ijk}}\overline{S}_{e:j}^{(1-Y_{ijk})\tilde{Y}_{ijk}}S_{p:j}^{(1-Y_{ijk})(1-\tilde{Y}_{ijk})}\overline{S}_{p:j}^{Y_{ijk}(1-\tilde{Y}_{ijk})}\right\}^{I(Z_{1k}+Z_{2k}>0)},$$

where $\vartheta = (p_{00}, p_{10}, p_{01})'$ and where the vector $\tilde{\mathbf{Y}}$ contains all of the true statuses $\tilde{Y}_{ijk}$. Note that we write $p_{11} = 1 - p_{00} - p_{10} - p_{01}$ in $L_C(\vartheta|\mathbf{Z}, \mathbf{Y}, \tilde{\mathbf{Y}})$ to reduce the dimension of the parameter space and to avoid constrained optimization.

In the E-step, one calculates $Q(\vartheta, \vartheta^{(d)}) = E\{\log L_C(\vartheta|\mathbf{Z}, \mathbf{Y}, \tilde{\mathbf{Y}})|\mathbf{Z}, \mathbf{Y}; \vartheta^{(\mathrm{d})}\}$ at the current parameter estimate $\boldsymbol{\vartheta}^{(d)}=(p_{00}^{(d)},p_{10}^{(d)},p_{01}^{(d)})'$. In the M-step, one finds the value $\vartheta^{(d+1)}$ that maximizes $Q(\vartheta, \vartheta^{(d)})$; i.e., $\vartheta^{(d+1)} = \arg\max_{\vartheta} Q(\vartheta, \vartheta^{(d)}$. Setting $\partial Q(\vartheta, \vartheta^{(d)}/\partial\vartheta$ equal to zero and solving the resulting system leads to the first order critical point $\boldsymbol{\vartheta}^{(d+1)}=(p_{00}^{(d+1)},p_{10}^{(d+1)},p_{01}^{(d+1)})'$ whose individual components are given by $p_{00}^{(d+1)}=N^{-1}\sum_{i=1}^{c_k}\sum_{k=1}^{K}E(\tilde{W}_{ik1}|\boldsymbol{Z},\boldsymbol{Y};\boldsymbol{\vartheta}^{(d)})$, $p_{10}^{(d+1)}=N^{-1}\sum_{i=1}^{c_k}\sum_{k=1}^{K}E(\tilde{W}_{ik2}|\boldsymbol{Z},\boldsymbol{Y};\boldsymbol{\vartheta}^{(d)})$, and $p_{01}^{(d+1)}=N^{-1}\sum_{i=1}^{c_k}\sum_{k=1}^{K}E(\tilde{W}_{ik3}|\boldsymbol{Z},\boldsymbol{Y};\boldsymbol{\vartheta}^{(d)})$, where, $\tilde{W}_{ik1}=(1-\tilde{Y}_{i1k})(1-\tilde{Y}_{i2k})$, $\tilde{W}_{ik2}=\tilde{Y}_{i1k}(1-\tilde{Y}_{i2k})$, and $\tilde{W}_{ik3}=(1-\tilde{Y}_{i1k})\tilde{Y}_{i2k}$. It is easy to show that the Hessian of $Q(\vartheta, \vartheta^{(d)})$ is negative definite; i.e., that $(\vartheta^{(d+1)})$ maximizes $Q(\vartheta, \vartheta^{(d)})$.

Because of the dependence among the latent statuses $\tilde{Y}_{ijk}$ and the observed data $\{\mathbf{Z}, \mathbf{Y}\}$, calculating the expectations $E(\tilde{W}_{ik2}|\boldsymbol{Z},\boldsymbol{Y};\boldsymbol{\vartheta}^{(d)})$ in closed form is difficult. We therefore develop a Gibbs sampler to estimate them. Let $\tilde{\mathbf{W}}_{ik}=(\tilde{W}_{ik1},\tilde{W}_{ik2},\tilde{W}_{ik3},\tilde{W}_{ik4})'$, where $\tilde{W}_{ik1}$, $\tilde{W}_{ik2}$, and $\tilde{W}_{ik3}$ are as previously defined and where $\tilde{W}_{ik4}=\tilde{Y}_{i1k}\tilde{Y}_{i2k}$. At the current estimate $\vartheta^{(d)}$ and conditional on $\mathbf{Z}$, $\mathbf{Y}$, and $\tilde{\mathbf{Y}}_{-ik}$ (i.e., all true statuses in $\tilde{\mathbf{Y}}$ except those in $\tilde{\mathbf{Y}}_{ik}$), the random vector $\tilde{\mathbf{W}}_{ik}$ follows a multinomial distribution with "cell probabilities" $p_{00}^{ik(d)}=\zeta_1^{ik(d)}/\zeta_+^{ik(d)}$, $p_{10}^{ik(d)}=\zeta_2^{ik(d)}/\zeta_+^{ik(d)}$, $p_{01}^{ik(d)}=\zeta_3^{ik(d)}/\zeta_+^{ik(d)}$, and $p_{11}^{ik(d)}=\zeta_4^{ik(d)}/\zeta_+^{ik(d)}$, where

$$\zeta_1^{ik(d)}=p_{00}^{(d)}\prod_{j=1}^{2}(S_{e:j}^{Z_{jk}}\overline{S}_{e:j}^{1-Z_{jk}})^{\delta_{ijk}}(S_{p:j}^{1-Z_{jk}}\overline{S}_{p:j}^{Z_{jk}})^{1-\delta_{ijk}}(S_{p:j}^{1-Y_{ijk}}\overline{S}_{p:j}^{Y_{ijk}})^{I(Z_{1k}+Z_{2k}>0)}$$
$$\zeta_2^{ik(d)}=p_{10}^{(d)}S_{e:1}^{Z_{1k}}\overline{S}_{e:1}^{1-Z_{1k}}(S_{e:2}^{Z_{2k}}\overline{S}_{e:2}^{1-Z_{2k}})^{\delta_{i2k}}(S_{p:2}^{1-Z_{2k}}\overline{S}_{p:2}^{Z_{2k}})^{1-\delta_{i2k}}(S_{e:1}^{Y_{i1k}}\overline{S}_{e:1}^{1-Y_{i1k}}S_{p:2}^{1-Y_{i2k}}\overline{S}_{p:2}^{Y_{i2k}})^{I(Z_{1k}+Z_{2k}>0)}$$
$$\zeta_3^{ik(d)}=p_{01}^{(d)}S_{e:2}^{Z_{2k}}\overline{S}_{e:2}^{1-Z_{2k}}(S_{e:1}^{Z_{1k}}\overline{S}_{e:1}^{1-Z_{1k}})^{\delta_{i1k}}(S_{p:1}^{1-Z_{1k}}\overline{S}_{p:1}^{Z_{1k}})^{1-\delta_{i1k}}(S_{e:2}^{Y_{i2k}}\overline{S}_{e:2}^{1-Y_{i2k}}S_{p:1}^{1-Y_{i1k}}\overline{S}_{p:1}^{Y_{i1k}})^{I(Z_{1k}+Z_{2k}>0)}$$
$$\zeta_+^{ik(d)}=(1-p_{00}^{(d)}-p_{10}^{(d)}-p_{01}^{(d)})\prod_{j=1}^{2}S_{e:j}^{Z_{jk}}\overline{S}_{e:j}^{1-Z_{jk}}(S_{e:j}^{Y_{ijk}}\overline{S}_{e:j}^{1-Y_{ijk}})^{I(Z_{1k}+Z_{2k}>0)},$$

$\zeta_+^{ik(d)}=\sum_{l=1}^{4}\zeta_l^{ik(d)}$, and $\delta_{ijk}=I(\sum_{i'\neq i}\tilde{Y}_{i'jk}>0)$. Note that these cell probabilities depend on both the observed data $\{\mathbf{Z}, \mathbf{Y}\}$ and the unobserved statuses of other individuals in the kth pool, that is, $\tilde{Y}_{i'jk}$, for $i' = i$. In Web Appendix B, we show how the conditional distributions $\tilde{\mathbf{W}}_{ik}|\{\boldsymbol{Z},\boldsymbol{Y},\tilde{\boldsymbol{Y}}_{-ik};\boldsymbol{\vartheta}^{(d)}\}$ are used to estimate the expectations $E(\tilde{W}_{ik}|\boldsymbol{Z},\boldsymbol{Y};\boldsymbol{\vartheta}^{(d)})$ in the E-Step.

Let $\hat{\boldsymbol{\vartheta}}=(\hat{p}_{00},\hat{p}_{10},\hat{p}_{01},)'$ denote the estimate of ϑ at convergence of the EM algorithm. A direct appeal to the missing information principle and Louis's method (1982) gives the observed data information matrix

$$\mathscr{I}(\boldsymbol{\vartheta})=-E\left\{\frac{\partial^2 logL_C(\boldsymbol{\vartheta}|\boldsymbol{Z},\boldsymbol{Y},\tilde{\boldsymbol{Y}})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}|\boldsymbol{Z},\boldsymbol{Y};\boldsymbol{\vartheta}\right\}-cov\left\{\frac{\partial logL_C(\boldsymbol{\vartheta}|\boldsymbol{Z},\boldsymbol{Y},\tilde{\boldsymbol{Y}})}{\partial\boldsymbol{\vartheta}}|\boldsymbol{Z},\boldsymbol{Y};\boldsymbol{\vartheta}\right\}. \quad (3)$$

Estimated standard errors are obtained from $\mathscr{I}(\hat{\boldsymbol{\vartheta}})^{-1}$, making the construction of large sample Wald confidence intervals possible. Calculating the matrices on the right hand side of Equation (3) uses the same Gibbs sampler described previously; complete details are provided in Web Appendix B. Finally, note that p_{11}, the marginal probabilities $\pi_1 = p_{10} + p_{11}$ and $\pi_2 = p_{01} + p_{11}$, and $\rho = \text{corr}(\tilde{Y}_{i1k}, \tilde{Y}_{i2k})$ are each functions of $\vartheta = (p_{00}, p_{10}, p_{01})'$. Large sample Wald inference for these parameters is possible using the Delta Method, as shown in the next section.

5. Simulation Study

We used simulation to evaluate the performance of the EM algorithm in Section 4 and also to provide insight on how the pooling procedure in Section 2 compares to individual testing in terms of the number of tests expended. For a given $\vartheta = (p_{00}, p_{10}, p_{01})'$, we first simulated true statuses $\tilde{\mathbf{Y}}_{ik} = (\tilde{Y}_{i1k}, \tilde{Y}_{i2k})'$ corresponding to $N = 1000$ individuals. For given values of $S_{e:j}$ and $S_{p:j}$, we then randomly assigned individual statuses $\tilde{\mathbf{Y}}_{ik}$ to pools of optimal size c_k^* as described in Section 3.2. For $j = 1, 2$, the true pool statuses $\tilde{\boldsymbol{Z}}_k=(\tilde{Z}_{1k},\tilde{Z}_{2k})'$ were recorded using $\tilde{Z}_{jk}=I(\sum_{i=1}^{c_k^*}\hat{Y}_{ijk}>0)$ and diagnosed pool statuses $\mathbf{Z}_k = (Z_{1k}, Z_{2k})'$ were recorded by simulating Z_{jk} ~ Bernoulli $\{S_{e:j}\tilde{Z}_{jk}+(1-S_{p:j})(1-\tilde{Z}_{jk})\}$. For those pools which were diagnosed as positive for either infection, we also simulated individual diagnoses $\mathbf{Y}_{ik} = (Y_{i1k}, Y_{i2k})'$ according to Y_{ijk} ~ Bernoulli $\{S_{e:j}\ \tilde{Y}_{ijk} + (1 - S_{p:j})\ (1 - \tilde{Y}_{ijk})\}$. This entire process was repeated $B = 1000$ times at each configuration of ϑ, $S_{e:j}$, and $S_{p:j}$.

In our investigation, we let $S_{e:j} = 0.95$ and $S_{p:j} = 0.99$, for $j = 1, 2$, and also $S_{e:j} = 0.99$ and $S_{p:j} = 0.95$. The first case allows for an assay that is highly specific for both infections and the second for an assay that is highly sensitive. At all configurations of ϑ, $S_{e:j}$, and $S_{p:j}$, we used the initial value $\vartheta^{(0)} = (0.25, 0.25, 0.25)'$ for each data set; however, we found that the performance of our EM algorithm was largely invariant to the choice of $\vartheta^{(0)}$. When implementing the Gibbs sampler at each E-step and in the calculation of $\mathscr{I}(\hat{\boldsymbol{\vartheta}})^{-1}$, as described in Section 4, we used $G = 500$ iterates (which included $a = 100$ burn-in iterates); complete details are provided in Web Appendix B. Convergence was declared when the maximum value in $\vartheta^{(d+1)} - \vartheta^{(d)}$ was less than 0.001 in absolute value.

A subset of our results appears in Table 1 and complete results are given in Web Appendix C. Our choices for $\vartheta = (p_{00}, p_{10}, p_{01})'$ included four different values of p_{00}, the probability that an individual is not infected with either disease, with $p_{00} \in \{0.80, 0.85, 0.90, 0.95\}$. These choices were motivated by our IPP application data set in Section 6. For each value of p_{00}, the remaining probabilities were varied to include a broad range of configurations. We also include simulation results for p_{11}, $\pi_1 = p_{10} + p_{11}$, $\pi_2 = p_{01} + p_{11}$, and ρ. Note that each of these parameters can be written as $g(\vartheta)$ for a suitably chosen function $g : \mathbb{R}^3 \to \mathbb{R}$ with continuous first partial derivatives. Therefore, the large sample variance of $g(\hat{\boldsymbol{\vartheta}})$ can be approximated using the Delta Method, that is, var $\{g(\hat{\boldsymbol{\vartheta}})\} \approx \{\dot{g}(\hat{\boldsymbol{\vartheta}})\}'\mathscr{I}(\hat{\boldsymbol{\vartheta}})^{-1}\dot{g}(\boldsymbol{\vartheta})$, where $\dot{g}(\vartheta)$ is the gradient of g evaluated at ϑ. An approximate $100(1 - \alpha)\%$ confidence interval for $g(\vartheta)$ is given by $g(\hat{\boldsymbol{\vartheta}}) \pm z_{\alpha/2}[\hat{var}\{g(\hat{\boldsymbol{\vartheta}})\}]^{1/2}$, where $z_{\alpha}/2$ is the upper $\alpha/2$ quantile of the standard normal distribution and where $\hat{var}\{g(\hat{\boldsymbol{\vartheta}})\}$ is any consistent estimator of var $\{g(\hat{\boldsymbol{\vartheta}})\}$.

The results in Table 1 and Web Appendix C demonstrate that our EM algorithm performs well overall; mean estimates, averaged over $B = 1000$ data sets, are all extremely close to the true values. In addition, the sample standard deviation of the 1000 estimates (SD) and the averaged standard error (SE) are always in close agreement. This suggests that the large sample covariance matrix of $\hat{\boldsymbol{\vartheta}}$ is being estimated correctly on average and also that variances obtained from the Delta Method (for p_{11}, π_1, π_2, and ρ) are good approximations. A potential criticism arises from examining the estimated Wald coverage probabilities, which when calculated at the nominal 95% level, are sometimes anti-conservative. However, closer inspection reveals that when this occurs, it is usually when estimating a very small probability or the correlation parameter ρ. The former is not surprising even when individual testing is used; the latter results likely because ρ is a highly nonlinear function of ϑ.

At each configuration of ϑ, $S_{e:j}$, and $S_{p:j}$, we have also included the number of tests expended (averaged over $B = 1000$ data sets), denoted by $\overline{T}$. From Table 1 and the results in Web Appendix C, one notes that the number of tests depends largely on p_{00}, as expected, and that using more specific assays always leads to a smaller number of tests required. The latter finding is intuitive given the nature of the IPP algorithm; fewer false positives in Step 2 lead to fewer additional individual retests. Under all settings, the average number of tests expended is less than $N = 1000$, the number of tests used under individual testing.

We have also compared our EM algorithm approach to the estimation procedure outlined in Hughes-Oliver and Rosenberger (2000), which assumes that $S_{e:j} = S_{p:j} = 1$ and uses only the responses from master pools. This comparison reveals (at least under perfect testing) that adding the individual retesting responses improves estimation efficiency. Complete details are in Web Appendix C.

6. Application to IPP Data

Because the state of Iowa already implements the pooling algorithm in Section 2, we have decided to illustrate the potential benefits of the algorithm for a state (Nebraska) that currently uses individual testing as part of the IPP. Our data set from Nebraska consists of individual testing results for 23,146 subjects tested in 2008 and 27,551 subjects tested in 2009. Having individual test results affords us the flexibility to make informative comparisons, for example, comparing the IPP algorithm in Section 2 to separate pooling algorithms for individual infections and, of course, comparing the IPP algorithm to individual testing.

To perform our analysis, we first use the 2008 individual test outcomes as “training data” to calculate optimal pool sizes c_k^*. We then implement the algorithm in Section 2 with the 2009 individuals using optimally-sized pools (that we construct and decode ourselves) and estimate population prevalences using the methods in Section 4. For verisimilitude, this is done separately within each of the four gender/specimen type strata, acknowledging that assay characteristics vary across these strata. Individuals in 2009 are assigned to pools chronologically (within strata) based on the specimen’s date of arrival for testing. A complete summary of the 2008 training data, including stratum sample sizes, estimated prevalences and correlations, optimal pool sizes, and assay characteristics, is provided in Table 2. When constructing this table, we assumed that the 2008 individual outcomes were the true statuses.

To emulate how the 2009 diagnoses would be made using the pooling procedure in Section 2, we first treat the 2009 observed individual statuses as the true statuses. Then, in order to account for potential misclassification, we simulate both group and (where necessary) individual testing diagnoses using the values of $S_{e:j}$ and $S_{p:j}$, $j = 1, 2$, in Table 2. Doing this also enables us to characterize the accuracy of the algorithm by comparing simulated diagnoses with the “true” statuses. To average over the effect of simulating the observed 2009 diagnoses, we repeat this entire process $B = 1000$ times for each gender/specimen type combination. All values reported in Tables 3 and 4 are averages over these 1000 implementations.

Table 3 summarizes the 2009 classification results. To evaluate the benefit of screening for two infections simultaneously, we also implemented Dorfman (1943) retesting separately for each infection. In Dorfman’s procedure, pools that test positive are decoded by retesting each individual. To ensure the fairest comparison, we used optimal pool sizes in each algorithm; Dorfman's optimal pool size was determined for each infection using Equation (2) in Kim et al. (2007). In terms of the number of tests expended, the results in Table 3 illustrate the benefits of group testing for multiple infections. For example, with $N = 14,530$ female swab specimens, the IPP algorithm uses 7709.8 tests on average; this is a 46.9% reduction in the number of tests when compared to individual testing and a 19.3% reduction when compared to separate Dorfman retesting (which requires 6381.5 + 3168.1 = 9549.6 tests on average). In terms of accuracy, there is mild evidence that the IPP algorithm provides slightly higher PS_e than Dorfman's procedure implemented separately, but somewhat lower values of PPV for gonorrhea. The latter occurs because gonorrhea is much less prevalent than chlamydia (in Nebraska and nationwide also); more individuals will be re-diagnosed for gonorrhea when the master pool tests positive for chlamydia only. Comparing the classification accuracy results to individual testing (see Table 2), one observes the IPP algorithm provides slightly lower sensitivity ($\mathrm{PS}_e < S_e$) but slightly higher specificity ($\mathrm{PS}_p > S_p$). These findings are consistent with the group testing literature for single infections (Kim et al., 2007).

Table 4 summarizes the 2009 estimation results. Using the methods in Section 4, we provide estimates and standard errors for $\vartheta = (p_{00}, p_{10}, p_{01})'$ and also for p_{11}, π_1, π_2, and ρ. To implement the EM algorithm, we used the same starting values, Gibbs sampler specifications, and convergence criteria described in Section 5. We also include in Table 4 the corresponding estimates and standard errors from individual testing, using the same assay error rates in Table 2. One will note that the point estimates from the IPP algorithm and individual testing are nearly identical. In addition, one finds that the group testing estimates are actually more efficient than those from individual testing. While this latter finding might seem counterintuitive, similar behavior was observed by Tu, Litvak, and Pagano (1995) and Liu et al. (2012) for single infection group testing procedures in the presence of testing error. Therefore, for states that might be considering adopting the IPP pooling protocol, our analysis suggests that doing so could be beneficial. There are potentially large cost savings available and prevalence estimates are as good or better than those from individual testing.

7. Discussion

We have examined group testing for classification and estimation with multiple infections, motivated by chlamydia and gonorrhea screening practices as part of the nationally-implemented IPP. When compared to individual testing, pooling for multiple infections reduces the number of tests and improves estimation efficiency. We have also observed that multiple infection pooling confers these same benefits when compared to group testing for single infections. To disseminate our work, the web site www.chrisbilder.com/grouptesting contains R programs that implement the methodology described in this paper.

This research could be extended in several ways. First, we have focused on the IPP algorithm in Section 2 explicitly because it is already in use by Iowa and by other states. However, other pooling algorithms could be formulated for classification purposes. The IPP algorithm is a hierarchical, (two-stage) Dorfman-type algorithm; i.e., individuals are assigned to non-overlapping pools and positive pools are decoded using individual testing. Kim et al. (2007) present a comprehensive evaluation of other pooling algorithms for single infections, including hierarchical algorithms that utilize a larger number of stages for decoding and array-based testing. We believe that these other algorithms could be generalized to classify individuals for multiple infections. In terms of estimation, our EM algorithm framework in Section 4 would generalize immediately to handle other pooling algorithms. As long as one can write out a complete data likelihood, the only change is that different conditional expectations would be estimated in the E-step. For the IPP algorithm in Section 2, we have observed (in separate investigations) that our joint modeling approach in Section 4 improves estimation efficiency when compared to modeling the infection statuses $\tilde{Y}_{i1k}$, and $\tilde{Y}_{i2k}$ separately. For other pooling algorithms, especially those involving a larger number of stages, we conjecture that these efficiency gains could be even larger.

Another extension would be to recast the IPP algorithm, or possibly other algorithms mentioned in the last paragraph, more generally for $J \geq 2$ infections. This extension might be of interest for the American Red Cross, for example, who uses pooling to simultaneously screen blood donations for HIV, hepatitis B, and hepatitis C (ARC, 2013). For the IPP algorithm with $J > 2$, derivations for the classification operating characteristics could be carried out in the same way as those for the $J = 2$ case, outlined in Web Appendix A, and obtaining prevalence estimates using the EM algorithm would also follow similarly; see Web Appendix B. We have formulated this extension with $J = 3$ using a multinomial distribution with $2^3 = 8$ cell probabilities, but the calculations involved are far more tedious than the $J = 2$ derivations shown in the Web Appendices.

Finally, a much more ambitious extension would be to acknowledge that individuals being tested have different risk factors (e.g., gender, race, number of sexual partners, etc.) and incorporate population heterogeneity into the classification and estimation procedures for multiple infections. Classification techniques in the presence of heterogeneity have been proposed recently for a single infection (Bilder, Tebbs, and Chen, 2010; McMahan, Tebbs, and Bilder, 2012). Generalizing this work for $J > 1$ infections would be more complex. First, a key question would arise as it pertains to classification, namely, is it desired to produce diagnoses for each infection or does it suffice to classify an individual as positive for at least one infection? The former would be of interest in public health settings, such as the IPP, where individuals require treatment for multiple diseases; the latter might be of interest for blood screening purposes, where it is critical to keep all infected units out of the blood supply. Second, depending on whether complete identification or purely negative identification is the goal, the next challenge would be to formulate exactly how multiple infection algorithms would exploit the different levels of risk. For example, using the algorithm in Section 2, positive pools could be resolved by retesting individuals in order according to their maximum risk probability or their combined risk probability in the same spirit as the “informative” Dorfman-type algorithms in McMahan et al. (2012) for single infections.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the Editor, the Associate Editor, and an anonymous referee for their helpful comments. We also thank Sandra Jirsa and Kristofer Eveland for their consultation on the IPP in Iowa, Drs. Peter Iwen and Steven Hinrichs for their consultation on the IPP in Nebraska, and Dr. Elizabeth Torrone at the Centers for Disease Control and Prevention. This research was supported by Grant R01 AI067373 from the National Institutes of Health.

References

American Red Cross. Blood testing. 2013. Available at http://www.redcrossblood.org. Last accessed June 15, 2013

Bilder C, Tebbs J, Chen P. Informative retesting. Journal of the American Statistical Association. 2010; 105:942–955. [PubMed: 21113353]

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]

Cardoso M, Koerner K, Kubanek B. Mini-pool screening by nucleic acid testing for hepatitis B virus, hepatitis C virus, and HIV: Preliminary results. Transfusion. 1998; 38:905–907. [PubMed: 9767739]

Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance 2011. 2012. Available at www.cdc.gov. Last accessed June 15, 2013

Delaigle A, Meister A. Nonparametric regression analysis for group testing data. Journal of the American Statistical Association. 2011; 106:640–650.

Dorfman R. The detection of defective members of large populations. Annals of Mathematical Statistics. 1943; 14:436–440.

Emrich L, Piedmonte M. A method for generating high-dimensional multivariate binary variates. American Statistician. 1991; 45:302–303.

Farley T, Cohen D, Elkins W. Asymptomatic sexually transmitted diseases: The case for screening. Preventative Medicine. 2003; 36:502–509.

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]

Hughes-Oliver J, Rosenberger W. Efficient estimation of the prevalence of multiple rare traits. Biometrika. 2000; 87:315–327.

Hughes-Oliver J, Swallow W. A two-stage adaptive group-testing procedure for estimating small proportions. Journal of the American Statistical Association. 1994; 89:982–993.

Jirsa, S. National STD Prevention Conference. Chicago, IL: 2008. Pooling specimens: A decade of successful cost savings. 2008

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]

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]

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]

Liu A, Liu C, Zhang Z, Albert P. Optimality of group testing in the presence of misclassification. Biometrika. 2012; 99:245–251. [PubMed: 23049137]

Louis T. Finding observed information using the EM algorithm. Journal of the Royal Statistical Society, Series B. 1982; 44:98–130.

McMahan C, Tebbs J, Bilder C. Informative Dorfman screening. Biometrics. 2012; 68:287–296. [PubMed: 21762119]

Mine H, Emura H, Miyamoto M, Tomono T, Minegishi K, Murokawa H, Yamanaka R, Yoshikawa A, Nishioka K. High throughput screening of 16 million serologically negative blood donors for hepatitis B virus, hepatitis C virus, and human immunodeficiency virus type-1 by nucleic acid amplification testing with specific and sensitive multiplex reagent in Japan. Journal of Virological Methods. 2003; 112:145–151. [PubMed: 12951223]

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]

Samoff E, Koumans E, Markowitz L, Sternberg M, Sawyer M, Swan D, Papp J, Black C, Unger E. Association of Chlamydia trachomatis with persistence of high-risk types of human papillomavirus in a cohort of female adolescents. American Journal of Epidemiology. 2005; 162:668–675. [PubMed: 16120706]

Tu X, Litvak E, Pagano M. On the informativeness and accuracy of pooled testing in estimating the prevalence of a rare disease: Application to HIV screening. Biometrika. 1995; 82:287–297.

Van T, Miller J, Warshauer D, Reisdorf E, Jerrigan D, Humes R, Shult P. Pooling nasopharyngeal/throat swab speciments to increase testing capacity for influenza viruses by PCR. Journal of Clinical Microbiology. 2012; 50:891–896. [PubMed: 22205820]

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]

Xie M. Regression analysis of group testing samples. Statistics in Medicine. 2001; 20:1957–1969. [PubMed: 11427952]

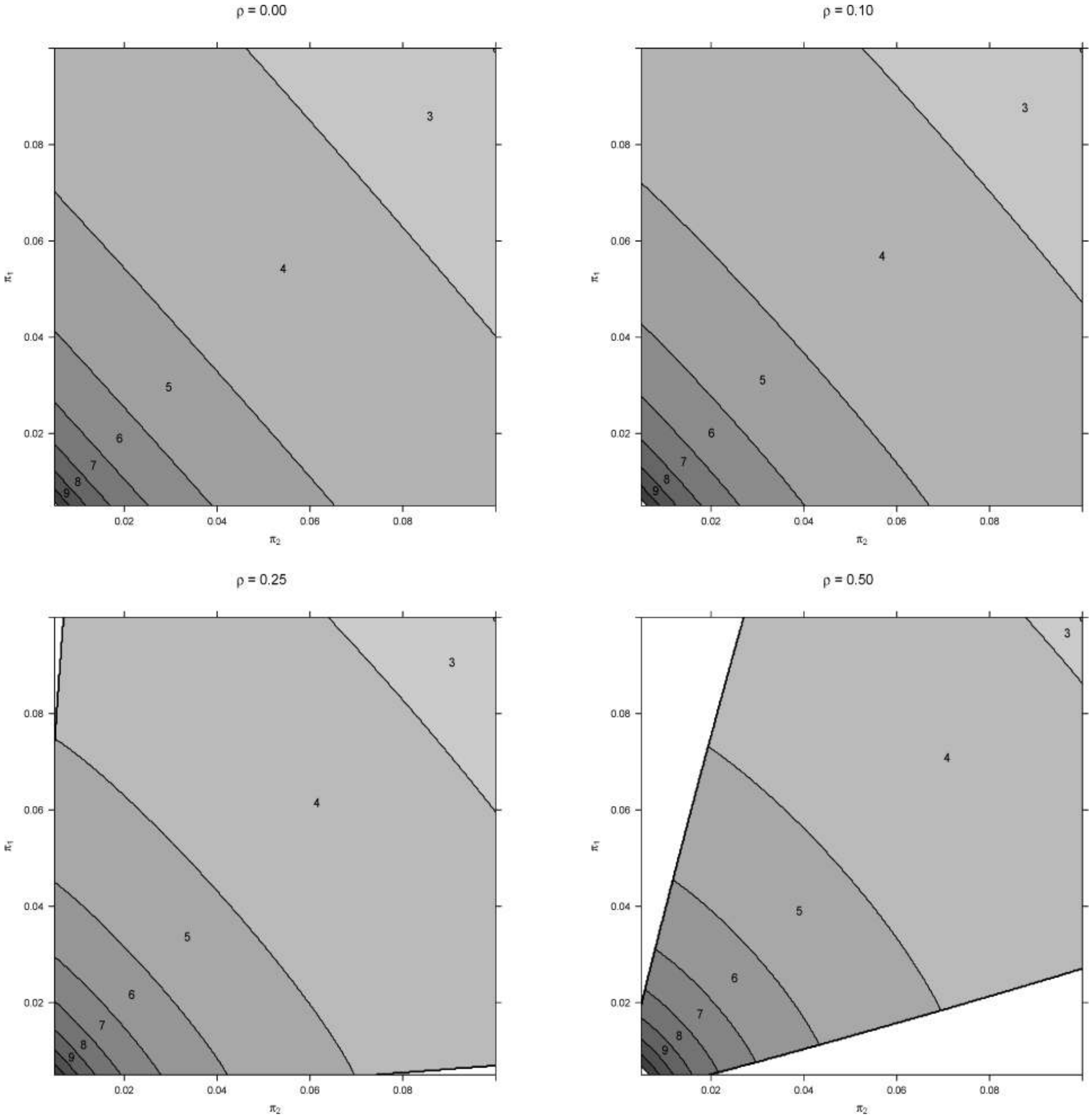

Figure 1.

IPP pooling algorithm. Optimal pool sizes c_k^* for different values of the correlation ρ when $S_{e:1} = 0{:}942$, $S_{p:1} = 0{:}976$, $S_{e:2} = 0{:}992$, and $S_{p:2} = 0{:}987$. In the lower left corner of each subfigure, we did not show the optimal pool sizes larger than $c_k^* = 9$ to avoid crowding. In the $\rho = 0{:}25$ and $\rho = 0{:}50$ subfigures, values of $(\pi_1; \pi_2)'$ in the white regions are not possible.

Table 1

IPP algorithm simulation results with $S_{e:j} = 0.95$ and $S_{p:j} = 0.99$, for $j = 1, 2$. Mean and standard deviation (SD) calculated from $B = 1000$ data sets, each with $N = 1000$ individuals. Averaged standard error (SE) and estimated coverage probability (Cov) associated with nominal 95% Wald confidence intervals are also included. The margin of error for the coverage probability estimates, assuming a 99% confidence level, is 0.018. Also included are the optimal pool size c_k^* and $\overline{T}$, the average number of tests expended.

	True	Mean	SD	SE	Cov		True	Mean	SD	SE	Cov
p_{00}	0.80	0.800	0.013	0.013	0.955	π_1	0.11	0.110	0.010	0.010	0.953
p_{10}	0.10	0.100	0.010	0.010	0.947	π_2	0.10	0.100	0.010	0.010	0.943
p_{01}	0.09	0.090	0.010	0.010	0.943	ρ	−0.01	−0.010	0.035	0.034	0.925
p_{11}	0.01	0.010	0.004	0.003	0.919		$\overline{T}$=811.0			c_k^*=3	
p_{00}	0.85	0.850	0.012	0.012	0.953	π_1	0.11	0.110	0.010	0.010	0.957
p_{10}	0.10	0.100	0.010	0.010	0.950	π_2	0.05	0.050	0.007	0.007	0.951
p_{01}	0.04	0.040	0.006	0.007	0.950	ρ	0.07	0.065	0.044	0.043	0.935
p_{11}	0.01	0.010	0.003	0.003	0.920		$\overline{T}$=714.7			c_k^*=3	
p_{00}	0.90	0.900	0.010	0.010	0.943	π_1	0.06	0.060	0.008	0.008	0.945
p_{10}	0.05	0.050	0.007	0.007	0.942	π_2	0.05	0.050	0.008	0.007	0.931
p_{01}	0.04	0.040	0.007	0.007	0.935	ρ	0.14	0.135	0.058	0.055	0.927
p_{11}	0.01	0.010	0.004	0.003	0.916		$\overline{T}$=593.8			c_k^*=4	
p_{00}	0.95	0.950	0.007	0.007	0.953	π_1	0.04	0.040	0.006	0.007	0.947
p_{10}	0.03	0.030	0.006	0.006	0.943	π^2	0.02	0.020	0.005	0.005	0.945
p_{01}	0.01	0.010	0.004	0.003	0.928	ρ	0.34	0.330	0.089	0.085	0.921
p_{11}	0.01	0.010	0.003	0.003	0.918		$\overline{T}$=432.5			c_k^*=5	

Table 2

IPP training data summary. All estimates are based on the individual testing results in 2008 (assuming no testing error). Stratum sample sizes N are also given. These estimates and values of S$_{e:j}$ and S$_{p:j}$, j = 1 (chlamydia, C), 2 (gonorrhea, G), are used to determine optimal pool sizes c_k^* for the 2009 individuals within each stratum.

Stratum	C	G	Prevalence	Correlation	Sensitivity	Specificity
Male/Urine (N = 3541)	−	−	$\hat{p}_{00}$=0.914			
	+	−	$\hat{p}_{10}$=0.073	$\hat{\rho}$=0.101	$S_{e:1}$ = 0.979	$S_{p:1}$ = 0.985
	−	+	$\hat{p}_{01}$=0.009		$S_{e:2}$ = 0.985	$S_{p:2}$ = 0.996
	+	+	$\hat{p}_{11}$=0.004			
Male/Swab (N = 2826)	−	−	$\hat{p}_{00}$=0.814			
	+	−	$\hat{p}_{10}$=0.118	$\hat{\rho}$=0.120	$S_{e:1}$ = 0.959	$S_{p:1}$ = 0.975
	−	+	$\hat{p}_{01}$=0.012		$S_{e:2}$ = 0.991	$S_{p:2}$ = 0.978
	+	+	$\hat{p}_{11}$=0.019			
Female/Urine (N = 2338)	−	−	$\hat{p}_{00}$=0.897			
	+	−	$\hat{p}_{10}$=0.080	$\hat{\rho}$=0.125	$S_{e:1}$ = 0.947	$S_{p:1}$ = 0.989
	−	+	$\hat{p}_{01}$=0. 012		$S_{e:2}$ = 0.913	$S_{p:2}$ = 0.993
	+	+	$\hat{p}_{11}$=0.011			
Female/Swab (N = 14441)	−	−	$\hat{p}_{00}$=0.920			
	+	−	$\hat{p}_{10}$=0.067	$\hat{\rho}$=0.148	$S_{e:1}$ = 0.942	$S_{p:1}$ = 0.976
	−	+	$\hat{p}_{01}$=0.008		$S_{e:2}$ = 0.992	$S_{p:2}$ = 0.987

Stratum	C	G	Prevalence	Correlation	Sensitivity	Specificity
	+	+	$\hat{p}_{11}=0.005$			

Table 3

IPP 2009 classification results. Mean number of tests $(\overline{T})$ and accuracy measures (PS$_e$, PS$_P$, PPV, and NPV), averaged over B = 1000 implementations, for four gender/specimen type strata using the algorithm in Section 2. Stratum sample sizes N are also given. The Dorfman procedure is carried out separately for each infection. Optimal pool sizes are given in parentheses.

		Male				**Female**			
		Urine		**Swab**		**Urine**		**Swab**	
		C	**G**	**C**	**G**	**C**	**G**	**C**	**G**
	N	6139		1910		4972		14530	
IPP	$\overline{T}$	3564.2 (4)		1607.1 (3)		2778.5 (4)		7709.8 (4)	
Dorfman		2867.7 (4)	1678.3 (10)	1172.0 (4)	880.4 (5)	2279.0 (4)	1176.1 (7)	6381.5 (5)	3168.1 (9)
IPP	PS$_e$	0.961	0.977	0.928	0.987	0.904	0.876	0.895	0.988
Dorfman		0.958	0.969	0.919	0.982	0.897	0.832	0.887	0.984
IPP	PS$_p$	0.996	0.999	0.990	0.990	0.997	0.998	0.994	0.996
Dorfman		0.997	0.999	0.990	0.994	0.998	0.999	0.994	0.999
IPP	PPV	0.954	0.945	0.943	0.879	0.966	0.878	0.923	0.787
Dorfman		0.961	0.968	0.945	0.923	0.970	0.954	0.916	0.903
IPP	NPV	0.997	0.999	0.987	0.999	0.992	0.998	0.992	0.999
Dorfman		0.996	0.999	0.985	0.999	0.991	0.997	0.992	0.999

Table 4

IPP 2009 estimation results. Parameter estimates and standard errors (SE), averaged over B = 1000 implementations, using the methods in Section 4. Stratum sample sizes N are also given. The results from individual testing are also included for comparison purposes. All point estimates are rounded to three digits; slight rounding error is possible.

			IPP algorithm				Individual testing			
Stratum	**C**	**G**	**Estimate**	**SE**	**Estimate**	**SE**	**Estimate**	**SE**	**Estimate**	**SE**
Male/Urine (N = 6139)	−	−	$\hat{p}_{00}$=0.905	0.0039	$\hat{\pi}_1$=0.081	0.0036	$\hat{p}_{00}$=0.905	0.0041	$\hat{\pi}_1$=0.081	0.0038
	+	−	$\hat{p}_{10}$=0.073	0.0035	$\hat{\pi}_2$=0.021	0.0019	$\hat{p}_{10}$=0.073	0.0037	$\hat{\pi}_2$=0.021	0.0020
	−	+	$\hat{p}_{01}$=0.014	0.0015	$\hat{\rho}$=0.152	0.0237	$\hat{p}_{01}$=0.014	0.0017	$\hat{\rho}$=0.151	0.0246
	+	+	$\hat{p}_{11}$=0.008	0.0011			$\hat{p}_{11}$=0.008	0.0012		
Male/Swab (N= 1910)	−	−	$\hat{p}_{00}$=0.971	0.0098	$\hat{\pi}_1$=0.156	0.0089	$\hat{p}_{00}$=0.791	0.0106	$\hat{\pi}_1$=0.157	0.0092
	+	−	$\hat{p}_{10}$=0.139	0.0085	$\hat{\pi}_2$=0.070	0.0059	$\hat{p}_{10}$=0.139	0.0089	$\hat{\pi}_2$=0.070	0.0067
	−	+	$\hat{p}_{01}$=0.051	0.0053	$\hat{\rho}$=0.071	0.0291	$\hat{p}_{01}$=0.052	0.0061	$\hat{\rho}$=0.070	0.0327
	+	+	$\hat{p}_{11}$=0.017	0.0031			$\hat{p}_{11}$=0.017	0.0035		
Female/Urine (N = 4972)	−	−	$\hat{p}_{00}$=0.911	0.0043	$\hat{\pi}_1$=0. 080	0.0041	$\hat{p}_{00}$=0.911	0.0046	$\hat{\pi}_1$=0. 080	0.0043
	+	−	$\hat{p}_{10}$=0.072	0.0039	$\hat{\pi}_2$=0.017	0.0019	$\hat{p}_{10}$=0.072	0.0041	$\hat{\pi}_2$=0.017	0.0023
	−	+	$\hat{p}_{01}$=0. 009	0.0015	$\hat{\rho}$=0.198	0.0299	$\hat{p}_{01}$=0. 008	0.0019	$\hat{\rho}$=0.200	0.0331
	+	+	$\hat{p}_{11}$=0.008	0.0014			$\hat{p}_{11}$=0.008	0.0014		
Female/Swab (N = 14530)	−	−	$\hat{p}_{00}$=0.924	0.0024	$\hat{\pi}_1$=0.069	0.0023	$\hat{p}_{00}$=0.924	0.0028	$\hat{\pi}_1$=0.069	0.0025
	+	−	$\hat{p}_{10}$=0.063	0.0022	$\hat{\pi}_2$=0.013	0.0009	$\hat{p}_{10}$=0.063	0.0025	$\hat{\pi}_2$=0.013	0.0013

Stratum	C	G	IPP algorithm				Individual testing			
			Estimate	**SE**	**Estimate**	**SE**	**Estimate**	**SE**	**Estimate**	**SE**
	−	+	$\hat{p}_{01}$=0.007	0.0007	$\hat{\rho}$=0.171	0.0182	$\hat{p}_{01}$=0.007	0.0012	$\hat{\rho}$=0.1 68	0.0217
	+	+	$\hat{p}_{11}$=0.006	0.0006			$\hat{p}_{11}$=0.006	0.0007		