Published in final edited form as:

J Am Stat Assoc. 2010 September 1; 105(491): 942–955. doi:10.1198/jasa.2010.ap09231.

Informative Retesting

Christopher R. Bilder,

Associate Professor of Statistics (chris@chrisbilder.com, Website: www.chrisbilder.com) at the University of Nebraska-Lincoln, Lincoln, NE 68583.

Joshua M. Tebbs, and

Associate Professor of Statistics at the University of South Carolina, Columbia, SC 29208.

Peng Chen

Statistician at Takeda Pharmaceuticals, Lake Forest, IL 60045.

Summary

In situations where individuals are screened for an infectious disease or other binary characteristic and where resources for testing are limited, group testing can offer substantial benefits. Group testing, where subjects are tested in groups (pools) initially, has been successfully applied to problems in blood bank screening, public health, drug discovery, genetics, and many other areas. In these applications, often the goal is to identify each individual as positive or negative using initial group tests and subsequent retests of individuals within positive groups. Many group testing identification procedures have been proposed; however, the vast majority of them fail to incorporate heterogeneity among the individuals being screened. In this paper, we present a new approach to identify positive individuals when covariate information is available on each. This covariate information is used to structure how retesting is implemented within positive groups; therefore, we call this new approach “informative retesting.” We derive closed-form expressions and implementation algorithms for the probability mass functions for the number of tests needed to decode positive groups. These informative retesting procedures are illustrated through a number of examples and are applied to chlamydia and gonorrhea testing in Nebraska for the Infertility Prevention Project. Overall, our work shows compelling evidence that informative retesting can dramatically decrease the number of tests while providing accuracy similar to established non-informative retesting procedures.

Keywords

Binary response; Chlamydia; Gonorrhea; Group testing; Identification; Pooled testing

1. INTRODUCTION

Chlamydia and gonorrhea are the two most prevalent bacteria-based sexually transmitted diseases in the United States, and \$4 billion is spent annually on these infections (Screening and Treatment Guidelines, IPP, Region VII, 2003). Infected persons are often asymptomatic, resulting in individuals being left untreated and others becoming infected unknowingly. Both diseases can lead to severe consequences including infertility and a higher susceptibility to HIV infection (Kacena et al. 1998a, 1998b). To address this public health problem, the Infertility Prevention Project (IPP), funded by the Centers for Disease Control and Prevention, has been implemented nationwide. Its purpose is to screen and to provide

treatment for chlamydia and gonorrhea in higher risk populations while monitoring disease prevalence. We will focus on the screening aspect in this paper, specifically its implementation in Nebraska. Screening is especially important in Nebraska because chlamydia and gonorrhea infections have been characterized as being at epidemic levels (Zagurski 2006). Currently, more than 30,000 tests are completed annually in Nebraska, and all testing is performed on *individual* specimens.

Given the large number of tests in Nebraska and the associated cost, it is important to find ways to reduce the amount of testing needed without screening fewer individuals. In similar situations where a large number of individuals are screened for infectious diseases, it has become standard practice to perform screening tests on *pools* or *groups* of individual specimens (e.g., blood, urine, etc.). Group testing, also known as pooled testing, was introduced by Dorfman (1943) to screen World War II soldiers for syphilis. Since this seminal work, the usefulness of pooling has been demonstrated in blood donation screening (Stramer et al. 2004), in screening individuals for drug use (Gastwirth and Johnson 1994), in preventing the potential spread of bioterrorist agents (Schmidt et al. 2005), and in other applications including genetics, plant pathology, veterinary, and drug discovery (Gastwirth 2000; Tebbs and Bilder 2004; Peck 2006; Remlinger et al. 2006). In general, group testing for case identification involves testing individuals first in groups. Positive groups are then “decoded” through algorithmic procedures to identify positive individuals. Perhaps due to its simplicity, Dorfman’s (1943) original procedure, where each individual in a positive pool is retested, is the most widely used. However, many other retesting strategies have been proposed; see Hughes-Oliver (2006) for a review.

Currently used decoding procedures do not account for the fact that individuals have different levels of risk; i.e., probabilities of being diseased. For example, every individual screened in the Nebraska IPP provides a set of covariates, including age, gender, number of sexual partners, and initial clinical observations that result in different individual risk levels. The presence of these and other covariates motivates our significantly new approach to the case identification problem. In the Nebraska IPP and in many other screening situations where group testing is used, binary regression models can be used to estimate individual risk probabilities (even if only group responses are available). We propose to use these probabilities to determine the order in which individuals are retested in positive groups. Conceptually, this is an intuitive idea, and one would naturally hypothesize the potential for substantial benefits. Our paper develops these new group testing procedures and shows when these substantial benefits occur. Because additional covariate information is used in the retesting protocol, we call the decoding of a positive group “informative retesting.”

There is a relatively small body of literature, including Hwang (1975) and Yao and Hwang (1988), where authors have allowed individuals to have different probabilities of disease positivity in a group testing classification setting. However, this previous work has assumed that individuals fall into just a few categories, risk probabilities are known, and diagnostic tests are error-free. More recently, the seminal works of Vansteelandt et al. (2000) and Xie (2001) have shown how to use regression methods to obtain estimates of individual probabilities based on only the observed group responses while allowing for testing error. In our paper, we combine these estimation methods with the group testing identification problem for the first time.

The outline of this paper is as follows. Section 2 presents the notation and assumptions used throughout. In Section 3, our new informative retesting procedures are proposed and probability mass functions for the number of tests are derived. We compare the proposed procedures to analogous “non-informative” procedures in Section 4 using sets of known individual probabilities. Section 5 applies the procedures to the Nebraska IPP setting to

show their substantial benefits. Finally, Section 6 discusses how informative retesting can be applied to other settings where group testing is used already. This section also presents additional research challenges in this new area.

2. NOTATION AND PRELIMINARIES

Suppose each individual to be screened is assigned to exactly one initial group. Let $Y_{ik} = 1$ if the i^{th} individual in the k^{th} initial group is diagnosed as positive; $Y_{ik} = 0$ otherwise for $i = 1, \dots, I_k$ and $k = 1, \dots, K$. We assume throughout that all of the Y_{ik} are independent Bernoulli random variables. Let Z_k denote the initial group response where $Z_k = 1$ (0) is a diagnosed positive (negative) for the k^{th} group. If the diagnostic test is perfect (no testing error), then

$Z_k = 1 \iff \sum_{i=1}^{I_k} Y_{ik} > 0$ and $Z_k = 0 \iff \sum_{i=1}^{I_k} Y_{ik} = 0$. Thus, the Y_{ik} 's are "observed" when $Z_k = 1$ with perfect testing; the Y_{ik} 's are initially unobserved otherwise. The goal of a retesting procedure is to determine all of the Y_{ik} responses when $Z_k = 1$.

Few diagnostic tests are perfect, so it is important to incorporate the effects of misclassification. Let $\tilde{Z}_k = 1$ if the k^{th} group is truly positive and $\tilde{Z}_k = 0$ otherwise. Similarly, the true values of Y_{ik} are denoted by $\tilde{Y}_{ik}$, where we define $p_{ik} = P(\tilde{Y}_{ik} = 1)$. Let the test sensitivity and specificity be denoted by $S_e = P(Z_k = 1 \mid \tilde{Z}_k = 1)$ and $S_p = P(Z_k = 0 \mid \tilde{Z}_k = 0)$, respectively, so that the probability group k tests positive is

$P(Z_k=1) = S_e + (1 - S_p - S_e) \prod_{i=1}^{I_k} (1 - p_{ik})$. We assume that S_e and S_p are known, are diagnostic test dependent, do not depend on covariates, and do not depend on the number of individuals per group I_k . Nearly all recent group testing research has followed these assumptions (e.g., Vansteelandt et al. 2000; Kim et al. 2007). Numerous empirical studies have shown the last assumption to be reasonable for a sensibly chosen I_k . For example, Litvak, Tu, and Pagano (1994) cites many studies which document $I_k = 15$ as an upper bound for enzyme linked immunosorbant assay (ELISA) tests for HIV. Pilcher et al. (2005) uses an upper bound of $I_k = 90$ with HIV assays based on a nucleic acid test (NAT). For a number of chlamydia and gonorrhea studies, NATs have high sensitivity and specificity for groups up to size $I_k = 10$ when pooling urine or swabs (Kacena et al. 1998a, 1998b; Butylkina et al. 2007; Shipitsyna et al. 2007).

To acknowledge the heterogeneity in the population, we model p_{ik} using logistic regression, although other binary regression models could be used. Let $\mathbf{x}_{ik} = (1, x_{ik1}, \dots, x_{ikp})'$ denote the vector of covariates for the i^{th} subject in the k^{th} group, and let $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$ denote the corresponding vector of model parameters, so that $\log[p_{ik}/(1 - p_{ik})] = \boldsymbol{\beta}'\mathbf{x}_{ik}$. In some applications, individual data are readily available, such as through a training data set, and the model above can be fit to them. For example, the Nebraska IPP has hundreds of thousands of observations accessible that can be used to estimate p_{ik} for new individuals yet to be screened. When training data are not available, the model can be fit using an initial set of group responses. This can be done using the methodology developed by Vansteelandt et al. (2000) and Xie (2001). Vansteelandt et al. (2000) uses a likelihood function constructed

from the initial group responses, $\prod_{k=1}^K \theta_k^{z_k} (1 - \theta_k)^{1-z_k}$ where

$\theta_k = S_e + (1 - S_p - S_e) \prod_{i=1}^{I_k} (1 - p_{ik})$, and maximizes it to obtain the maximum likelihood estimator of $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}$. Xie (2001) treats individual responses as missing and uses the EM algorithm to find $\hat{\boldsymbol{\beta}}$. An advantage of using Xie's approach is that more elaborate testing protocols, such as matrix-based testing, can be incorporated. Our emphasis herein is on hierarchical procedures that begin with an initial set of groups being tested, so we use the approach in Vansteelandt et al. (2000). Once the model is fit, informative retesting uses estimates of $p_{1k}, \dots, p_{I_k k}$ for each group $k = 1, \dots, K$. Within a positive group, the estimates

can be ordered as $\hat{p}_{(1)k} \leq \dots \leq \hat{p}_{(I_k)k}$, where $\hat{p}_{(i)k}$ corresponds to the individual with the i^{th} smallest estimated probability. These ordered estimates provide information on which individuals are more likely to be positive and can be used to choose which individuals to retest in an algorithmic manner. Conversely, non-informative retesting procedures treat all individuals as homogeneous, i.e., $p_{1k} = \dots = p_{I_k k}$.

3. STERRETT PROCEDURES

3.1 Non-informative Sterrett procedures

For an initially diagnosed positive group k ($Z_k = 1$), the non-informative Sterrett (NIS) procedure proposed by Sterrett (1957) begins by randomly retesting individuals until the first positive is found. The remaining individuals are then pooled to form a new group. If this new group tests negative, decoding is complete, and its individuals are declared negative. If this new group tests positive, the process begins again by randomly retesting its individuals until a second positive is found. Once the second positive is found, the remaining individuals again are pooled together to determine if any positives remain. The process of testing individually and repooling is repeated until all individuals are classified as positive or negative. Overall, this procedure usually results in a smaller expected number of tests than Dorfman's (1943) procedure.

Combinations of Sterrett and Dorfman's procedures have been proposed. For instance, after the first positive individual is found using Sterrett's procedure, Dorfman's procedure can be applied to the new group if it is positive. This decoding procedure is called "one-stage Sterrett" (Johnson, Kotz, and Wu 1991). The two-stage Sterrett procedure continues retesting until the second positive is found and Dorfman retesting begins afterward. These variations are easier to implement than Sterrett's full procedure, and they can work well for groups sized to have a small probability of producing more than one or two positives.

3.2 Informative Sterrett procedures

For a positive group k with $p_{(1)k} \leq \dots \leq p_{(I_k)k}$, we modify the NIS procedure to retest individuals based on the largest probabilities instead of selecting individuals randomly. Our "one-stage informative Sterrett" (ISIS) procedure retests individuals in order of descending probability until the first positive is found. The remaining individuals are repooled to allow for Dorfman retesting if the new group tests positive. The entire ISIS procedure is outlined in the top binary tree of Figure 1 for a group of size four ($I_k = 4$). The nodes of the tree show which individuals are being tested with (i) denoting the individual with the i^{th} smallest probability and $[]$ denoting a group test involving individuals within the brackets. A test results in a negative $(-)$ or a positive $(+)$, which is given underneath the nodes because it governs what is tested next. The leaves of the tree are terminal nodes indicating the last tests to be completed.

Figure 1 (bottom) depicts our "two-stage informative Sterrett" (2SIS) procedure, also with $I_k = 4$. This procedure is identical to ISIS, except testing continues until the second positive individual is found. At that point, Dorfman retesting is implemented if the remaining individuals produce a positive group test. Finally, Figure 2 depicts our "full informative Sterrett" (FIS) procedure for $I_k = 3$ and $I_k = 4$. This is the full informative version of NIS.

3.3 Probability mass functions

For each of our informative retesting procedures, we derive the probability mass function (PMF) for the number of tests, T_k , needed to decode a positive initial group k . Denote by $\mathbf{p}_k = (p_{(1)k}, \dots, p_{(I_k)k})'$ the vector of ordered probabilities for the I_k individuals in group k . While $\mathbf{p}_k$ is unknown, our subsequent derivations show how one would expect the retesting

procedures to perform. Section 5 provides examples of how the informative Sterrett procedures are carried out using estimated probabilities.

We start the derivations with 1SIS and refer the reader to Figure 1 (top). Each leaf is reached with a certain probability and has a number of tests associated with it. Let $G_{[.]}$ denote a 0 or 1 group test result where the individuals being tested are represented in the subscript. For example, G_{123} denotes the outcome of testing [(1), (2), (3)] in a group, and G_3 denotes the outcome of testing (3) only. The probability of reaching the leaf second from the right is $P(G_{1234}=1, G_4=1, G_{123}=0)$, where there are $t_k=3$ tests. Without testing error, this probability is simply $p_{(4)k}(1-p_{(3)k})(1-p_{(2)k})(1-p_{(1)k})$. In the presence of testing error, suitable modification is needed. Let $\tilde{G}_{[.]}$ denote the true 0 or 1 status for a test. By the theorem of total probability,

$$\begin{aligned} P(G_{1234}=1, G_4=1, G_{123}=0) = & \\ & P(G_{1234}=1, G_4=1, G_{123}=0, \tilde{G}_{1234}=0, \tilde{G}_4=0, \tilde{G}_{123}=0) + \\ & P(G_{1234}=1, G_4=1, G_{123}=0, \tilde{G}_{1234}=1, \tilde{G}_4=0, \tilde{G}_{123}=1) + \\ & P(G_{1234}=1, G_4=1, G_{123}=0, \tilde{G}_{1234}=1, \tilde{G}_4=1, \tilde{G}_{123}=0) + \\ & P(G_{1234}=1, G_4=1, G_{123}=0, \tilde{G}_{1234}=1, \tilde{G}_4=1, \tilde{G}_{123}=1), \end{aligned}$$

which simplifies to

$$\begin{aligned} S_p(1-S_p)^2 \prod_{i=1}^4 (1-p_{(i)}) + S_e(1-S_e)(1-S_p)(1-p_{(4)}) \left[1 - \prod_{i=1}^3 (1-p_{(i)}) \right] \\ + S_e^2 S_p p_{(4)} \prod_{i=1}^3 (1-p_{(i)}) + S_e^2(1-S_e)p_{(4)} \left[1 - \prod_{i=1}^3 (1-p_{(i)}) \right]. \end{aligned}$$

Note that not all possible binary outcomes for $\tilde{G}_{1234}$, $\tilde{G}_4$, and $\tilde{G}_{123}$ can occur together due to the group hierarchy. Our derivation here uses the standard assumption that test outcomes are conditionally independent given the true outcomes; see Litvak et al. (1994) for discussion and empirical evidence that this assumption is reasonable in practice.

Noticing that some leaves in Figure 1 (top) lead to the same number of tests, the PMF for T_k with $I_k=4$ can be written as

$$P(T_k=t_k) = \begin{cases} P(G_{1234}=0), & \text{for } t_k=1 \\ P(G_{1234}=1, G_4=1, G_{123}=0), & \text{for } t_k=3 \\ P(G_{1234}=1, G_4=0, G_3=1, G_{12}=0), & \text{for } t_k=4 \\ P(G_{1234}=1, G_4=0, G_3=0), & \text{for } t_k=5 \\ P(G_{1234}=1, G_4=1, G_{123}=1) + P(G_{1234}=1, G_4=0, G_3=1, G_{12}=1), & \text{for } t_k=6. \end{cases} \quad (1)$$

We have chosen $I_k=4$ only because the corresponding tree is easy to depict in a figure. Generalizing the form of the PMF for any group size I_k involves recognizing patterns that emerge by comparing a tree of size I_k to a tree of size I_k-1 . In general, the 1SIS PMF for T_k can be written as

$$P(T_k=t_k)=\begin{cases} P(G_{1,\dots,I_k}=0), & \text{for } t_k=1 \\ P\left(G_{1,\dots,I_k}=1, \sum_{j=I_k+4-t_k}^{I_k} G_j=0, G_{I_k+3-I_k}=1, G_{1,\dots,I_k+2-I_k}=0\right), & \text{for } t_k=3, \dots, I_k \\ P\left(G_{1,\dots,I_k}=1, \sum_{j=3}^{I_k} G_j=0\right), & \text{for } t_k=I_k+1 \\ \sum_{m=4}^{I_k+1} P\left(G_{1,\dots,I_k}=1, \sum_{j=m}^{I_k} G_j=0, G_{m-1}=1, G_{1,\dots,m-2}=1\right), & \text{for } t_k=I_k+2 \end{cases} \quad (2)$$

for $I_k \geq 3$, where it is understood that events like $\left\{\sum_{j=r}^s G_j=0\right\}$ for $r > s$ are removed. Table 1 provides closed form expressions for these PMF probabilities written in terms of S_e , S_p , and $p_{(i)k}$.

The PMF for T_k in the 2SIS procedure can be derived using the 1SIS PMF. To see why, note that $P(G_{1234}=1, G_4=1, G_{123}=1)$ in Equation (1) is given as part of $P(T_k=6)$.

Decomposing this joint probability is the only change needed to obtain the 2SIS PMF.

Specifically, this probability is broken up into three separate pieces because $G_{123}=1$, which indicates a second positive individual is among individuals (1), (2), and (3). Thus,

$$\begin{aligned} P(G_{1234}=1, G_4=1, G_{123}=1) &= P(G_{1234} \\ &=1, G_4=1, G_{123}=1, G_3=1, G_{12}=0) + P(G_{1234} \\ &=1, G_4=1, G_{123}=1, G_3=0) + P(G_{1234} \\ &=1, G_4=1, G_{123}=1, G_3=1, G_{12}=1). \end{aligned}$$

Because these three pieces each correspond to a different number of tests, the 2SIS PMF with $I_k=4$ is

$$P(T_k=t_k)=\begin{cases} P(G_{1234}=0), & \text{for } t_k=1 \\ P(G_{1234}=1, G_4=1, G_{123}=0), & \text{for } t_k=3 \\ P(G_{1234}=1, G_4=0, G_3=1, G_{12}=0), & \text{for } t_k=4 \\ P(G_{1234}=1, G_4=0, G_3=0) + P(G_{1234}=1, G_4=1, G_{123}=1, G_3=1, G_{12}=0), & \text{for } t_k=5 \\ P(G_{1234}=1, G_4=0, G_3=1, G_{12}=1) + P(G_{1234}=1, G_4=1, G_{123}=1, G_3=0), & \text{for } t_k=6 \\ P(G_{1234}=1, G_4=1, G_{123}=1, G_3=1, G_{12}=1), & \text{for } t_k=7. \end{cases}$$

In Equation (1), note that the event $\{G_{1234}=1, G_4=0, G_3=1, G_{12}=1\}$ at $t_k=6$ for 1SIS also indicates a second positive exists; however, the number of tests corresponding to it is the same for 2SIS. In general, only probabilities from the event of

$\left\{G_{1,\dots,I_k}=1, \sum_{j=m}^{I_k} G_j=0, G_{m-1}=1, G_{1,\dots,m-2}=1\right\}$ for $m=5, \dots, I_k+1$ in Equation (2) need to be decomposed into additional pieces for 2SIS; Table 2 provides these probabilities for $I_k \geq 4$. The PMF for T_k under 2SIS then consists of the probabilities in Table 2 added to the probabilities in Equation (2) at the same t_k values, excluding the $m=5, \dots, I_k+1$ probabilities at $t_k=I_k+2$ in the 1SIS PMF.

Derivations are similar for a larger number of stages; however, the PMF probabilities become far more complicated for additional stages. To avoid the complications, we have developed a recursive algorithm to calculate the FIS PMF for T_k . In order to explain this

algorithm, examine the binary trees in Figure 2 for $I_k = 3$ and $I_k = 4$. Starting at the (3) nodes, one can see that this part of the $I_k = 3$ tree is embedded twice within the full $I_k = 4$ tree. For larger I_k , this embedding continues and leads to the development of recursive computational algorithms. We have discovered a number of ways to compute the FIS PMF by exploiting this pattern, but most encounter memory limitations when programmed into computational software. One algorithm, given in the Appendix, avoids these memory limitations. An R program that implements this algorithm is available on the Journal's supplementary file website and at www.chrisbilder.com/grouptesting.

4. COMPARISONS FOR A FIXED SET OF INDIVIDUAL PROBABILITIES

4.1 Dorfman and Sterrett procedures

To quantify the gains from our informative retesting procedures, we first compare them to the Dorfman and NIS procedures for a single group with fixed $\mathbf{p}_k$. In the group testing identification literature, various measures have been used for comparison purposes, including pooling sensitivity and specificity, positive (negative) predictive values, and the expected number of tests. In homogeneous populations where $p_{1k} = \dots = p_{I_k k}$, closed-form expressions can be derived for each of these measures (e.g., Kim et al. 2007). In heterogeneous populations where individual probabilities differ, this is generally not possible. Thus, we instead use the cumulative distribution function (CDF) of T_k with known $\mathbf{p}_k$ to illustrate differences among the procedures. Note that the CDF for Dorfman's procedure is $F(t_k) = 0$ for $t_k < 1$, $F(t_k) = 1 - \theta_k$ for $1 \leq t_k < I_k + 1$, and $F(t_k) = 1$ for $t_k \geq I_k + 1$. Convenient expressions for the NIS CDF do not exist.

We now present three simple examples. These examples are used to illustrate the potential usefulness of informative testing and are not meant to be exhaustive. However, each example is motivated by actual applications.

Example 1. Suppose that $I_k = 8$ and $\mathbf{p}_k = (0.002, 0.010, 0.031, 0.053, 0.100, 0.102, 0.150, 0.172)'$. These (ordered) probabilities represent a single random sample from a beta(1, 10) distribution. This distribution is right skewed and concentrated towards 0 making it a reasonable choice to describe a group testing setting. Figure 3 (upper left) shows the CDFs for T_k with $S_e = 0.95$ and $S_p = 0.9$, and its inset lists the mean and standard deviation for the number of tests. The CDFs provide a visual description of how probabilities are distributed over the support of T_k . In particular, those CDFs that rise more quickly to 1 correspond to the more efficient retesting procedures, and subtracting the means for two procedures gives the area difference between the corresponding CDFs. In this example, FIS provides the smallest mean number of tests and the smallest standard deviation. While 2SIS has slightly larger mean and standard deviation than FIS, its performance is comparable, which suggests that 2SIS may be preferred because of its easier implementation.

As we discussed in Section 3, the NIS procedure decodes positive groups by retesting individuals at random, so there are $I_k!$ different possible NIS implementations. In Figure 3 (upper right), we have sampled 100 implementations at random and have plotted the corresponding CDFs. These CDFs are found by permuting the elements in $\mathbf{p}_k$ and using the FIS CDF expression without reordering the individual probabilities. Essentially, our FIS procedure finds the most efficient NIS implementation, which is to retest individuals in order of descending probabilities. Thus, the FIS CDF is necessarily stochastically smaller than the NIS CDF for all I_k and $\mathbf{p}_k$.

Example 2. Suppose that $I_k = 10$, $S_e = S_p = 0.99$, $p_{(1)k} = \dots = p_{(9)k} = 0.01$, and $p_{(10)k} = 0.5$; that is, one individual is "high risk" and the others are "low risk." This choice for $p_{(10)k}$ is not unusual; for example, estimated probabilities this large occur in the Nebraska IPP

example examined in Section 5. For this configuration, Figure 3 (lower left) plots the CDFs and lists the means and standard deviations of T_k . All of the informative Sterrett procedures offer profound improvements over Dorfman retesting. In addition, because there is only one high risk individual, the three informative Sterrett procedures give similar results.

Example 3. Suppose that $I_k = 100$, $S_e = S_p = 0.99$, $p_{(1)k} = \dots = p_{(98)k} = 0.01$, and $p_{(99)k} = p_{(100)k} = 0.1$; that is, two individuals are at higher risk than the others. It is not unusual for group sizes this large to be used in practice. For example, Pilcher et al. (2005) uses NATs with initial groups of size 90 as a quality control check on individual blood donations, and Kennedy (2006) uses groups of size 100 to screen cattle for bovine viral diarrhea virus. Figure 3 (lower right) provides similar comparisons as before. Once again, the informative Sterrett procedures are far superior to Dorfman; the CDF separation is more pronounced in this case because of the larger group size. Overall, there is a 43.4% reduction in the expected number of tests for FIS when compared to Dorfman retesting.

4.2 Additional comparisons

While Dorfman's retesting procedure is most often used in practice, other hierarchical classification procedures have been proposed more recently. One commonly suggested approach involves successively splitting a positive group into smaller sub-groups until decoding has been completed. For example, Litvak et al. (1994) present a "halving" hierarchical algorithm for use with blood donation screening where positive groups are split into two equal-sized halves (or as close to equal as possible). Non-hierarchical procedures, such as matrix pooling (Phatarfod and Sudbury 1994; Kim et al. 2007), have found successful applications in genetics and in some disease screening settings. In the "SA1" testing protocol of Phatarfod and Sudbury (1994), specimens are arranged into a matrix-like grid. Specimens are pooled within each row and column separately. Individuals at the intersections of positive rows and positive columns are possible positives. These individuals are further tested individually to complete the decoding.

Unlike our informative Sterrett procedures, neither halving nor matrix pooling take into account the different risk levels of individuals; therefore, they would be classified as non-informative. However, it is still of interest to compare them to our new informative retesting algorithms. To do this, we use the derivations in Bilder et al. (2009) that give the expected number of tests when individual probabilities are unequal. Similar to NIS, each possible ordering of $p_{1k}, \dots, p_{I_k k}$ leads to a potentially different mean for the number of tests expended with these non-informative procedures. Thus, we calculate the expected number of tests for them by averaging $E(T_k)$ over a set of 100 randomly selected permutations of the individual probabilities.

In Example #1 of Section 4.1, non-informative halving leads to an average expected value of 4.23 and 4.26 for 3-stage, 3H, (test in groups of size 8, 4, and then individually) and 4-stage, 4H, (8, 4, 2, and individually) versions, respectively. When applied to a 2×4 grid, SA1 matrix pooling (MP) leads to an average expected number of tests of 6.90. Alternatively, applying MP to an 8×8 grid with 8 replicates of each probability (in order to use groups of size 8), the average expected number of tests is 4.17 per 8 individuals. In Example #2, halving leads to expected values of 4.90 and 4.72 for 3H (10, 5, and individually) and 4H (10, 5, 2 or 3, and individually), respectively. MP applied to a 2×5 grid leads to an expected value of 7.65. When MP is applied to a 10×10 grid with 10 replicates of each probability, the average expected number of tests is 4.22 per 10 individuals. In both examples, all three informative Sterrett procedures have smaller means (see Figure 3).

For Example #3, both halving and matrix pooling can provide an expected number of tests smaller than those of informative Sterrett. For halving, we have average expected values of

46.62 for 3H (100, 50, and individually), 29.29 for 4H (100, 50, 25, and individually), and 19.55 for 5-stage (100, 50, 25, 12 or 13, and individually). For MP in a 10×10 grid, we obtain a mean of 22.32. When applied to a 100×100 grid with 100 replicates of each probability, MP results in a mean of 49.86 per 100 individuals. With FIS, the mean number of tests is 39.80. Halving and matrix pooling can perform better here because they can more quickly remove large numbers of negative individuals. For example, halving can eliminate 50 individuals at once if one of the two initial halves tests negative.

5. NEBRASKA IPP

The proposed informative retesting procedures are applicable to any group testing setting where p_{ik} can be estimated. To illustrate this, we now apply our procedures to previously obtained individual diagnoses from the Nebraska IPP. There are more than 30,000 tests done annually (all at one laboratory in Omaha), and each test is performed currently on individual swab or urine specimens. With a cost of approximately \$11 (swab) or \$16 (urine) for each test, group testing could provide significant cost savings. For the sake of brevity, we will use 30,067 individuals tested in 2005. The numbers of individuals screened for chlamydia are 2,679 females using urine (10.2% were declared positive), 19,451 females using a swab (5.8%), 3,852 males using urine (8.3%), and 4,085 males using a swab (13.0%). The number of individuals screened for gonorrhea are 2,679 females using urine (2.0%), 19,450 females using a swab (1.0%), 3,852 males using urine (1.6%), and 4,086 males using a swab (4.9%).

To perform the identification, we treat the 2005 diagnoses as the true values, and we assign subjects to artificially constructed initial groups within each of the four gender and specimen combinations. The corresponding S_e and S_p levels are given in the plot insets of Figure 4 (to be discussed in Section 5.1). Simulated diagnoses for group and individual tests in the testing process are found using these S_e and S_p levels, which allows us to estimate how accurately a procedure performs in identifying positive or negative individuals.

5.1 Using training data to estimate the models

If informative retesting were applied in the Nebraska IPP setting, a training data set would be used to estimate $\text{logit}(p_{ik}) = \beta'x_{ik}$ based on a prior set of individual data. This is because testing is performed daily, and the positive or negative diagnoses are needed promptly for treatment purposes. We use the 2004 testing results as the training data here. The covariates for each individual are gender, age, race, clinic type, clinic location, reason for visit (family planning, prenatal, and STD screening), symptoms, initial clinical observations (cervical friability, pelvic inflammatory disease, cervicitis, and urethritis), and risk history (multiple partners in last 90 days, new partner in last 90 days, and contact to a STD); all are used as first-order terms in the model. Higher order terms could be included, but our informative retesting procedures work well already with the simpler model. For each disease, we fit separate models for each gender and specimen type combination and compute estimates for each p_{ik} among the 2005 individuals. Histograms of these estimates are available in the supplementary materials. These 2005 individuals are assigned then to initial groups in chronological order based on the specimen date, which is practical in this setting.

Figure 4 displays the number of tests expended for the testing procedures. Because test diagnoses for each infection are simulated and because of the random testing order inherent in the non-informative procedures, we perform each retesting procedure 10 times so that one can assess the variability in the results; lines connect the number of tests for each implementation in the plots. Group sizes of 5, 10, and 20 are used, where the last group formed may have a smaller size. For each gender and specimen combination, one can see in the plots that group testing significantly reduces the total number of tests needed. Dorfman's procedure most often results in the largest number of tests (MP is usually larger for $I_k = 5$),

and the informative procedures (1SIS, 2SIS, and FIS) almost always beat NIS, sometimes quite dramatically (it is possible for NIS to be lower due to the simulated diagnoses). For example, when $I_k = 20$, the average number of gonorrhea tests needed for the female swab aggregate is 3,251.6 for NIS, 2,350 for FIS, 2,598.6 for 1SIS, and 2,370.6 for 2SIS. FIS reduces the number of tests by 27.7% when compared to NIS, which, in turn, reduces the number of tests by 83.3% when compared with individual testing. It is also worth noting that the efficiency of 2SIS closely rivals that of FIS in almost all cases, which makes 2SIS attractive because it is easier to implement. When compared to 3H, 4H, and MP, FIS most often results in a smaller number of tests. For both diseases with female urine specimens (and at times with male urine when testing for gonorrhea), 3H, 4H, and MP can sometimes beat FIS, but their classification accuracy, to be discussed next, can be much lower.

The efficiency of a pooling procedure is typically measured by the number of tests needed to classify each individual as positive or not. However, it is also important to characterize how accurately a pooling procedure classifies. Two commonly used measures of correctness are pooling sensitivity (PS_e) and pooling specificity (PS_p). PS_e (PS_p) is defined as the probability an individual is diagnosed as positive (negative) through using a group testing procedure when the individual is actually positive (negative). Group testing identification procedures usually possess values of PS_e smaller than S_e if individuals are declared negative immediately when their group tests negative. For example, the pooling sensitivity for

Dorfman's procedure is S_e^2 (Johnson et al., 1991, p. 83). In contrast, PS_p is usually higher than S_p because individuals in positive groups are tested more than once. We find these same patterns in Table 3, where averaged values of PS_e and PS_p (across the 10 implementations) are given for groups of size 10. While some PS_e levels are low, the S_e values given in Figure 4 are low already, especially for female urine testing. Higher PS_e can be achieved by using a more sensitive test or by performing additional tests on individuals declared negative (Pilcher et al. 2005 use this latter strategy). In Table 3, we also observe that PS_p is almost always higher for the informative Sterrett procedures than for NIS. It is especially worth noting the lower PS_e levels of 3H, 4H, and MP for female urine when compared to the informative Sterrett procedures. Similar results occur with groups of size 5 and 20 (available in the supplementary materials). Therefore, while 3H, 4H, and MP may sometimes result in a smaller number of tests, it is at cost of having significantly lower accuracy in diagnosing positive individuals.

Two additional measures of correctness are the pooling positive predictive value (PPPV) and the pooling negative predictive value (PNPV). PPPV (PNPV) is defined as the probability an individual is truly positive (negative) when a group testing procedure has diagnosed the individual to be positive (negative). We also provide averaged estimates for these measures in Table 3. Informative retesting usually provides higher values for PPPV and PNPV than NIS and Dorfman retesting. When compared to 3H, 4H, and MP, there is not a clear winner. For example, 4H frequently has the largest PPPV overall, but it is often close to FIS and can be smaller (see the case $I_k = 20$ for males tested for gonorrhea using swabs).

5.2 Using the initial group responses to estimate the models

The Nebraska IPP example illustrates a situation where training data are easily accessible and can be used to fit a model to perform informative retesting. In other situations, training data may not be available, so there is a need to use the initial group testing observations themselves to find individual probability estimates. We illustrate this next using the 2005 data.

For each gender and specimen combination, we randomly assign individuals tested in 2005 to groups of size $I_k = 5, 10$, and 20 (with one group potentially smaller) and estimate the same first-order model $\text{logit}(p_{ik}) = \beta'x_{ik}$ as before, now using the methodology of Vansteelandt et al. (2000). Estimates of p_{ik} are computed from this model fit. In Table 4, we provide information on the number of tests and measures of identification correctness for $I_k = 10$ (averaged over 10 implementations). Similar results occur for $I_k = 5$ and 20 (available in the supplementary materials).

Overall, the informative Sterrett procedures perform better than the non-informative procedures, but the differences in this setting are sometimes not as dramatic as in Section 5.1. This is not unexpected because the number of group responses used to fit the models is about I_k times less than when using individual observations. This notwithstanding, the increase in efficiency still can be significant. For example, the maximum reduction in mean number of tests for $I_k = 10$ was 24.4% for FIS vs. NIS when screening males for gonorrhea using swabs. For this same situation, the reduction was 24.0% for FIS vs. 3H, 21.7% for FIS vs. 4H, and 18.0% for FIS vs. MP. In addition, measures of identification accuracy are similar to those in Section 5.1. In particular, where 3H, 4H, and MP provide a smaller number of tests than FIS, there is a decrease in classification accuracy for these non-informative procedures.

6. CONCLUSIONS

We have presented a formal introduction to a new area of group testing research, which we call informative retesting. In doing so, we have shown that decoding positive groups using available covariates can significantly reduce the number of tests needed and maintain or sometimes improve identification accuracy. While we focus on chlamydia and gonorrhea screening in Nebraska, the benefits from informative retesting would be realized in other applications where group testing is used and individual probabilities can be estimated. For example, from a small survey of the infectious disease literature where group testing has been used, informative retesting could have been utilized in screening blood donations by the American Red Cross (Stramer et al. 2004), in testing Kenyan women for HIV (Vansteelandt et al. 2000), and in assessing hepatitis C prevalence among Scottish health care workers (Thorburn et al. 2001). In these applications, differences among individual disease prevalences for covariate aggregates were cited, but these differences were not exploited to identify positive individuals. Overall, for informative retesting to be useful, the only requirements are that individual probabilities differ within the groups, otherwise they become their non-informative counterparts, and that the individual probability ordering, not necessarily the estimates themselves, can be well approximated by a binary regression model.

In addition to the number of tests used, ease of implementation is an important consideration. Dorfman's procedure is the easiest, which likely explains its frequent use. On the other hand, while using FIS often results in the smallest expected number of tests for the informative procedures, it is more difficult to perform in practice. In this light, we view 1SIS and 2SIS as worthy alternatives, because they confer most of the benefits of FIS and are easier to use. Robotic testing machines, already used for high throughput screening with matrix pooling, could perhaps be adapted to automate the informative Sterrett procedures. Our results show this would be a worthwhile endeavor.

Informative Sterrett procedures generally perform better than halving and matrix pooling for smaller group sizes. The reverse can be true when group sizes are very large or large matrices can be used as seen in Example #3 of Section 4.2. While these group configurations may be feasible in some applications, they are usually limited in others due to possible

dilution effects, past assay calibrations, or laboratory physical limitations. For example, we have yet to find an application in the chlamydia and gonorrhea literature where the group size was larger than $I_k = 10$. When larger group sizes can be used, we anticipate that informative versions of halving and matrix pooling may yield even larger benefits than their non-informative counterparts. Research into these new informative procedures is underway. In order to help researchers decide which procedure to use, we have made available an R function in the supplementary materials that provides moment information for a given p_k .

Group size selection is a topic frequently visited in the literature, but mainly in the homogeneous population setting. When individual probabilities are different, determining optimal group sizes becomes far more complex. For example, each group could potentially be of a different size in order to satisfy an optimality criterion. The difficulty of finding optimal group sizes is further exacerbated when training data are not available so that initial group responses are needed to fit the model, as in Section 5.2. When training data are available, an ad-hoc approach to find one overall optimal initial group size is to first

compute $\bar{p}_{(1)}, \dots, \bar{p}_{(I)}$, where $\bar{p}_{(i)} = K^{-1} \sum_{k=1}^K \hat{p}_{(i)k}$, on the training data for different group sizes I . Using these averaged, ordered individual probabilities, one can compute an estimate of the expected number of tests. The value of I that provides the minimum estimated expected number of tests with the training data could be used to form the groups. While this ad-hoc approach is sensible, we feel that optimal group size determination in a heterogeneous population merits further investigation.

In the Nebraska IPP data example and in many other settings, individuals are assigned to groups chronologically because testing is performed continuously over time. However, there are situations where covariates for each individual are known before testing begins (e.g., Thorburn et al. 2001). In this setting, a potentially important design consideration is to decide how individuals are assigned to groups. Vansteelandt et al. (2000) and Bilder and Tebb (2009) have demonstrated that forming homogeneous groups (in terms of the covariates) reduces variability in the group testing regression estimates. On the other hand, Remlinger et al. (2006) finds that groups with individuals as diverse as possible produces the smallest number of tests needed for identification. Additional research is needed to remedy these potentially conflicting criteria when group composition can be controlled.

We hope this paper will encourage researchers to think about using covariates to decode positive pools in group testing. As public health expenses continue to soar upward and as reform efforts are debated, implementing cost effective screening procedures is now more important than ever. Our work shows that procedures which exploit available covariate information can provide large savings when compared to those that ignore it.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Dr. Steven Hinrichs, Dr. Peter Iwen, and Philip Medina for their consultation on the Infertility Prevention Project in Nebraska. The authors also thank Michael Black, Chris McMahan, and Boan Zhang for their assistance and Owen Worley and Brandon Bookstaver for helpful discussion. Finally, the authors thank the Editor, Associate Editor, and two referees for their comments that led to a strengthened paper. This research is supported by Grant R01 AI067373 from the National Institutes of Health.

REFERENCES

- Bilder C, Tebbs J. Bias, efficiency, and agreement for group-testing regression models. *Journal of Statistical Computation and Simulation* 2009;79:67–80. [PubMed: 20046956]
- Bilder, C.; Black, M.; Tebbs, J. Technical Report. University of Nebraska-Lincoln: Department of Statistics; 2009. Expected number of tests for halving and matrix pooling in heterogeneous populations.
- Butylkina R, Juseviciute V, Kasparaviciene G, Vagoras A, Pagirskas E, Unemo M, Domeika M. Pooling of urine specimens allows accurate and cost-effective genetic detection of *Chlamydia trachomatis* in Lithuania and other low-resource countries. *Scandinavian Journal of Infectious Diseases* 2007;39:209–212. [PubMed: 17366049]
- Dorfman R. The detection of defective members of large populations. *Annals of Mathematical Statistics* 1943;14:436–440.
- Gastwirth J. The efficiency of pooling in the detection of rare mutations. *American Journal of Human Genetics* 2000;67:1036–1039. [PubMed: 10986050]
- Gastwirth J, Johnson W. Screening with cost-effective quality control: Potential applications to HIV and drug testing. *Journal of the American Statistical Association* 1994;89:972–981.
- Hughes-Oliver, J. Pooling experiments for blood screening and drug discovery. In: Dean, A.; Lewis, S., editors. *Screening: Methods for Experimentation in Industry, Drug Discovery, and Genetics*. New York: Springer; 2006.
- Hwang F. A generalized binomial group testing problem. *Journal of the American Statistical Association* 1975;70:923–926.
- Johnson, N.; Kotz, S.; Wu, X. *Inspection Errors for Attributes in Quality Control*. New York: Chapman & Hall; 1991.
- Kacena K, Quinn S, Hartman S, Quinn T, Gaydos C. Pooling of urine samples for screening for *Neisseria gonorrhoeae* by ligase chain reaction: accuracy and application. *Journal of Clinical Microbiology* 1998a;36:3624–3628. [PubMed: 9817885]
- Kacena K, Quinn S, Howell M, Madico G, Quinn T, Gaydos C. Pooling urine samples for ligase chain reaction screening for genital *Chlamydia trachomatis* infection in asymptomatic women. *Journal of Clinical Microbiology* 1998b;36:481–485. [PubMed: 9466763]
- Kennedy JA. Diagnostic efficacy of a reverse transcriptase–polymerase chain reaction assay to screen cattle for persistent bovine viral diarrhea virus infection. *Journal of the American Veterinary Medical Association* 2006;229:1472–1474. [PubMed: 17078812]
- Kim H, Hudgens M, Dreyfuss J, Westreich D, Pilcher C. Comparison of group testing algorithms for case identification in the presence of test error. *Biometrics* 2007;63:1152–1163. [PubMed: 17501946]
- Litvak E, Tu X, Pagano M. Screening for the presence of a disease by pooling sera samples. *Journal of the American Statistical Association* 1994;89:424–434.
- Peck C. Going after BVD. *Beef* 2006;42:34–44.
- Phatarfod R, Sudbury A. The use of a square array scheme in blood testing. *Statistics in Medicine* 1994;13:2337–2343. [PubMed: 7855467]
- 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]
- Remlinger K, Hughes-Oliver J, Young S, Lam R. Statistical design of pools using optimal coverage and minimal collision. *Technometrics* 2006;48:133–143.
- Schmidt M, Roth W, Meyer H, Seifried E, Hourfar M. Nucleic acid test screening of blood donors for orthopoxviruses can potentially prevent dispersion of viral agents in case of bioterrorism. *Transfusion* 2005;45:399–403. [PubMed: 15752158]
- Screening and Treatment Guidelines, Infertility Prevention Project, Region VII. 2003. Available at <http://www.devsys.org/html/ipp/index.html>

- Shipitsyna E, Shalepo K, Savicheva A, Unemo M, Domeika M. Pooling samples: The key to sensitive, specific and cost-effective genetic diagnosis of *Chlamydia trachomatis* in low-resource countries. *Acta Dermato-Venereologica* 2007;87:140–143. [PubMed: 17340020]
- Stramer S, Glynn S, Kleinman S, Strong D, Caglioti S, Wright D, Dodd R, Busch M. Detection of HIV-1 and HCV infections among antibody-negative blood donors by nucleic acid–amplification testing. *New England Journal of Medicine* 2004;351:760–768. [PubMed: 15317889]
- Sterrett A. On the detection of defective members of large populations. *Annals of Mathematical Statistics* 1957;28:1033–1036.
- Tebbs J, Bilder C. Confidence interval procedures for the probability of disease transmission in multiple-vector-transfer designs. *Journal of Agricultural, Biological, and Environmental Statistics* 2004;9:75–90.
- Thorburn D, Dundas D, McCrudden E, Cameron S, Goldberg D, Symington I, Kirk A, Mills P. A study of hepatitis C prevalence in healthcare workers in the west of Scotland. *Gut* 2001;48:116–120. [PubMed: 11115832]
- 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]
- Yao Y, Hwang F. Individual testing of independent items in optimal group testing. *Probability in the Engineering and Informational Sciences* 1988;2:23–29.
- Zagurski K. Douglas County rates B+ on meetings its health goals, but Dr. Adi Pour says there's 'a lot of work to be done' on reducing STDs. *Omaha World Herald* 2006 Feb. 2;:08B.

APPENDIX: FIS PMF RECURSIVE ALGORITHM

For a given $(p_{(1)k}, \dots, p_{(j)k})'$ with $j \leq I_k$, define

$S^{(j)} = [P(T_k^{(j)}=1), P(T_k^{(j)}=3), \dots, P(T_k^{(j)}=2j-1)]'$ as a vectorized PMF for $T_k^{(j)}$, the number of tests using FIS that are needed to decode the individuals with the j smallest probabilities. We omit the subscript k on $S^{(j)}$ for brevity. These PMFs will be found in the order of $j = 3, \dots, I_k$ to obtain the PMF for $T_k^{(I_k)} \equiv T_k$. For convenience, define

$W^{(j)} = [P(T_k^{(j)}=1 | \tilde{\mathbf{y}}_k=\mathbf{0}), P(T_k^{(j)}=3 | \tilde{\mathbf{y}}_k=\mathbf{0}), \dots, P(T_k^{(j)}=2j-1 | \tilde{\mathbf{y}}_k=\mathbf{0})]'$ to be a vector of probabilities conditional on all negative true values for $\tilde{\mathbf{y}}_k = (\tilde{y}_{1k}, \dots, \tilde{y}_{I_k k})'$. Also, for a general vector $\mathbf{v}$, denote the i^{th} element as $\mathbf{v}[i]$ and the i through m elements by $\mathbf{v}[i:m]$ where $i < m$.

The recursive algorithm begins by finding $S^{(3)}$. Let $\mathbf{q}^{(3)}$ be a vector of the true probabilities for $\tilde{y}_{(3)k} = 0$ or 1 and at least one positive or all negatives among $\tilde{y}_{(1)k}$ and $\tilde{y}_{(2)k}$, where the subscript parentheses denotes ordering by individual probabilities within a group. This vector gives the probabilities ordered as follows:

$$\begin{aligned} \mathbf{q}^{(3)} &= [P(\tilde{G}_{123}=0), P(\tilde{G}_{123}=1, \tilde{G}_3=0), P(\tilde{G}_{123}=1, \tilde{G}_3=1, \tilde{G}_{12}=0), P(\tilde{G}_{123}=1, \tilde{G}_3=1, \tilde{G}_{12}=1)]' \\ &= [(1 - p_{(3)k}), p_{(3)k}]' \otimes \left[\prod_{i=1}^2 (1 - p_{(i)k}), \left(1 - \prod_{i=1}^2 (1 - p_{(i)k}) \right) \right]' \end{aligned}$$

These probabilities together form a PMF for $T_k^{(j)}$ (ordered $t_k^{(j)}=1, 4, 3, 5$ for algorithmic reasons) when $S_e = S_p = 1$. To incorporate testing errors, define

$$\mathbf{A}_{4 \times 4}^{(3)} = \begin{bmatrix} S_p & S_p(1-S_p)^2 & S_p(1-S_p) & (1-S_p)^3 \\ 1-S_e & S_e(1-S_p)(1-S_e) & S_e S_p & S_e^2(1-S_p) \\ 1-S_e & S_e^2 S_p & S_e(1-S_e) & S_e^2(1-S_p) \\ 1-S_e & S_e^2(1-S_e) & S_e(1-S_e) & S_e^3 \end{bmatrix}$$

so that $\mathbf{S}^{(3)} = \mathbf{A}^{(3)'} \mathbf{q}^{(3)}$. Notice that $\mathbf{W}^{(3)} = \mathbf{A}^{(3)}[1,]'$, the first row of $\mathbf{A}^{(3)}$.

To find $\mathbf{S}^{(4)}$, notice the $I_k = 3$ sub-tree starting with (3) in Figure 2 is represented once on the left and once on the right in the $I_k = 4$ tree. We can find probabilities for the number of tests needed to decode separately on both sides, taking into account the outcomes for G_{1234} and G_4 , and then sum the probabilities that correspond to the same number of tests. This leads to computing $\mathbf{S}^{(4)}$. An illustration of this process is shown in the supplementary materials. Similar successive calculations are performed for larger group sizes to obtain $\mathbf{S}^{(I_k)}$.

Starting with $\mathbf{S}^{(3)}$ and $\mathbf{W}^{(3)}$, the general computational algorithm successively finds $\mathbf{S}^{(j)}$ and $\mathbf{W}^{(j)}$ for $j = 4, \dots, I_k$ as follows:

1. Compute $\mathbf{W}^{(j)} = \mathbf{A}^{(j)}[1,]'$ through the following steps:
 - a. The contribution to the first row of the $\mathbf{A}^{(j)}$ matrix when $G_{1,\dots,j} = 0$ or $\{G_{1,\dots,j} = 1, G_j = 0\}$ is $\mathbf{W}_1^{(j)} = [S_p, S_p \mathbf{W}^{(j-1)}[2:(2j-4)]]'$
 - b. The contribution to the first row of the $\mathbf{A}^{(j)}$ matrix when $\{G_{1,\dots,j} = 1, G_j = 1\}$ is $\mathbf{W}_2^{(j)} = (1-S_p)^2 \mathbf{W}^{(j-1)}$
 - c. Combining $\mathbf{W}_1^{(j)}$ and $\mathbf{W}_2^{(j)}$ in the correct order for $t_k^{(j)} = 1, 3, \dots, 2j-1$ results in the first row of $\mathbf{A}^{(j)}$:

$$\mathbf{W}^{(j)} = [\mathbf{W}_1^{(j)}[1], \mathbf{W}_2^{(j)}[1], \mathbf{W}_1^{(j)}[2], \mathbf{W}_1^{(j)}[3:(2j-4)]' + \mathbf{W}_2^{(j)}[2:(2j-5)]', \mathbf{W}_2^{(j)}[2j-4]]'.$$

2. Compute $\mathbf{S}^{(j)}$ through the following steps:

- a. The probability for $G_{1,\dots,j} = 0$ is
$$S_0^{(j)} = S_p \prod_{i=1}^j (1 - p_{(i)k}) + (1 - S_e) \left[1 - \prod_{i=1}^j (1 - p_{(i)k}) \right]$$
- b. The probabilities for $\{G_{1,\dots,j} = 1, G_j = 0\}$ are

$$\mathbf{S}_1^{(j)} = [S_p(1 - p_{(j)k}) + (1 - S_e)p_{(j)k}] \mathbf{S}^{(j-1)}[2:(2j-4)] + \left(\frac{S_e}{1 - S_p} - 1 \right) (1 - S_e)p_{(j)k} \prod_{i=1}^{j-1} (1 - p_{(i)k}) \mathbf{W}^{(j-1)}[2:(2j-4)] \text{ for } S_p \neq 1$$

- c. The probabilities for $\{G_{1,\dots,j} = 1, G_j = 1\}$ are

$$\mathbf{S}_2^{(j)} = [S_e(1 - S_p)(1 - p_{(j)k}) + S_e^2 p_{(j)k}] \mathbf{S}^{(j-1)} + [(1 - S_p)^2 - S_e(1 - S_p)] \prod_{i=1}^j (1 - p_{(i)k}) \mathbf{W}^{(j-1)}$$

- d. Combining $S_0^{(j)}$, $S_1^{(j)}$, and $S_2^{(j)}$ in the correct order for $l_k^{(j)}=1, 3, \dots, 2j-1$ results in

$$S^{(j)} = [S_0^{(j)}, S_2^{(j)}[1], S_1^{(j)}[1], S_1^{(j)}[2:(2j-5)]' + S_2^{(j)}[2:(2j-5)]', S_2^{(j)}[2j-4]']'$$

3. Repeat steps 1) and 2) until $j = I_k$.

Alternatively, one could compute a full version of $A^{(j)}$ and find $S^{(j)} = A^{(j)'} q^{(j)}$ for $q^{(j)} = [(1 - p_{(j)k}), p_{(j)k}]' \otimes q^{(j-1)}$; however, this will result in very large matrices when I_k is not small, which leads to surpassing memory limits for commonly used software. Due to the simple structure of $A^{(j)}$ once $A^{(3)}$ is found, these limitations are avoided using our algorithm.

Figure 1.

Binary trees for 1SIS (top) and 2SIS (bottom) depicting the order in which tests are performed. The individuals being tested are given by (i) , which denotes the individual with the i^{th} smallest probability. Groups of individuals being tested are enclosed within brackets. Negative or positive test results are indicated beneath a node. “Ind. test” indicates that individual testing would be performed on individuals in the previous positive group. “Done” indicates no more testing is necessary.

Figure 2.

Binary trees for FIS depicting the order in which tests are performed. The top tree is for $I_k = 3$ and the bottom tree is for $I_k = 4$. The individuals being tested are given by (i) , which denotes the individual with the i^{th} smallest probability. Groups of individuals being tested are enclosed within brackets. Negative or positive test results are indicated beneath a node. “Done” indicates no more testing is necessary.

Figure 3.
CDFs for retesting procedures. Note that CDF lines may overlap (lines drawn in the legend order). “SD” denotes standard deviation.

Figure 4.

Chlamydia and gonorrhea testing results using the 2004 data as a training data set. The group sizes are given on the left for each of the 10 implementations, and S_e and S_p values are given as insets within each plot. “Dorf” denotes Dorfman’s procedure. Note that 4H can not be performed fully for $I_k = 5$.

Table 1

PMF for T_k with 1SIS ($I_k \geq 3$). For brevity, the subscript k on I_k , $p_{(i)k}$, and t_k has been omitted. It is understood that product expressions like $\prod_{j=r}^s$ omit those terms where $r > s$.

t	$P(T=t)$
1	$S_p \prod_{i=1}^I (1 - p_{(i)}) + (1 - S_e) \left[1 - \prod_{i=1}^I (1 - p_{(i)}) \right]$
3, ..., I	$(1 - S_p) S_p^{t-2} (1 - S_p - S_e) \prod_{i=1}^I (1 - p_{(i)}) + S_e \left\{ \prod_{i=I+4-t}^I [S_p (1 - p_{(i)}) + (1 - S_e) p_{(i)}] \right\} \\ \times \left[(1 - S_p) (1 - p_{(I+3-t)}) + S_e p_{(I+3-t)} \right] \left\{ S_p \prod_{i=1}^{I+2-t} (1 - p_{(i)}) + (1 - S_e) \left[1 - \prod_{i=1}^{I+2-t} (1 - p_{(i)}) \right] \right\}$
I + 1	$(1 - S_p) S_p^{I-2} \prod_{i=1}^I (1 - p_{(i)}) + S_e \left\{ \prod_{i=3}^I [S_p (1 - p_{(i)}) + (1 - S_e) p_{(i)}] + S_p^{I-2} \prod_{i=1}^I (1 - p_{(i)}) \right\}$
I + 2	$\sum_{m=4}^{I+1} (1 - S_p - S_e) S_p^{I+1-m} \prod_{i=1}^I (1 - p_{(i)}) + S_e \left\{ \prod_{i=m}^I [S_p (1 - p_{(i)}) + (1 - S_e) p_{(i)}] \right\} \\ \times \left[(1 - S_p) (1 - p_{(m-1)}) + S_e p_{(m-1)} \right] \left\{ (1 - S_p) \prod_{i=1}^{m-2} (1 - p_{(i)}) + S_e \left[1 - \prod_{i=1}^{m-2} (1 - p_{(i)}) \right] \right\}$

Table 2

Probabilities resulting from the second individual positive that are needed for the 2SIS PMF calculation ($I_k \geq 4$). For brevity, the subscript k on $I_k, p_{(i)}$ and t_k has been omitted. It is understood that product expressions like $\prod_{j=r}^s$ omit those terms where $r > s$.

t	Probability
$5, \dots, I+1$	$\sum_{m=1}^{t-4} P \left(G_{1,\dots,t} = 1, \sum_{j=I+2-m}^t G_j = 0, G_{t-m+1} = 1, G_{1,\dots,t-m} = 1, \sum_{j=I+5-t}^{I-m} G_j = 0, G_{t+1-t} = 1, G_{1,\dots,t+3-t} = 0 \right)$ $= \sum_{m=1}^{t-4} (1 - S_p - S_e)(1 - S_p)^3 S_p^{t-4} \prod_{i=1}^t (1 - p_{(i)}) + S_e \left[\prod_{j=I+2-m}^t S_p(1 - p_{(j)}) + (1 - S_e) p_{(j)} \right] \left[(1 - S_p)(1 - p_{(t-m+1)}) + S_e p_{(t-m+1)} \right] \times$ $\left\{ (1 - S_p - S_e)(1 - S_p) S_p^{t-m-3} \prod_{i=1}^{I-m} (1 - p_{(i)}) + S_e \left[\prod_{j=I+5-t}^{I-m} S_p(1 - p_{(j)}) + (1 - S_e) p_{(j)} \right] \times \right.$ $\left. \left[(1 - S_p)(1 - p_{(t+1-t)}) + S_e p_{(t+1-t)} \right] \times \left[S_p \prod_{i=1}^{I+3-t} (1 - p_{(i)}) + (1 - S_e) \left(1 - \prod_{i=1}^{I+3-t} (1 - p_{(i)}) \right) \right] \right\}$
$I+2$	$\sum_{m=5}^{I+1} P \left(G_{1,\dots,t} = 1, \sum_{j=m}^t G_j = 0, G_{m-1} = 1, G_{1,\dots,m-2} = 1, \sum_{j=3}^{m-2} G_j = 0 \right)$ $= \sum_{m=5}^{I+1} (1 - S_p - S_e)(1 - S_p)^2 S_p^{I-3} \prod_{i=1}^t (1 - p_{(i)}) + S_e \left[\prod_{j=m}^t S_p(1 - p_{(j)}) + (1 - S_e) p_{(j)} \right] \times$ $\left[(1 - S_p)(1 - p_{(m-1)}) + S_e p_{(m-1)} \right] \times$ $\left\{ (1 - S_p - S_e) S_p^{m-4} \prod_{i=1}^{m-2} (1 - p_{(i)}) + S_e \left[\prod_{j=3}^{m-2} S_p(1 - p_{(j)}) + (1 - S_e) p_{(j)} \right] \right\}$

t	Probability
$I+3$	$ \begin{aligned} & \sum_{m=1}^{I-3} \sum_{r=1}^{I-m-2} P \left(G_{1,\dots,I} = 1, \sum_{j=I+2-m}^I G_j = 0, G_{I-m+1} = 1, G_{1,\dots,I-m} = 1, \sum_{j=r+3}^{I-m} G_j = 0, G_{r+2} = 1, G_{1,\dots,r+1} = 1 \right) \\ &= \sum_{m=1}^{I-3} \sum_{r=1}^{I-m-2} (1 - S_p - S_e)(1 - S_p)^4 S_p^{I-r-3} \prod_{i=1}^I (1 - p_{(i)}) \\ & \quad \times \left[\prod_{j=I+2-m}^I S_p (1 - p_{(j)}) + (1 - S_e) p_{(j)} \right] \left[(1 - S_p)(1 - p_{(I-m+1)}) + S_e p_{(I-m+1)} \right] \times \\ & \quad \left\{ (1 - S_p - S_e)(1 - S_p)^2 S_p^{I-m-r-2} \prod_{i=1}^{I-m} (1 - p_{(i)}) \right. \\ & \quad \left. + S_e \left[\prod_{j=r+3}^{I-m} S_p (1 - p_{(j)}) + (1 - S_e) p_{(j)} \right] \left[(1 - S_p)(1 - p_{(r+2)}) + S_e p_{(r+2)} \right] \times \right. \\ & \quad \left. \left[(1 - S_p - S_e)(1 - p_{(r+1)}) + S_e p_{(r+1)} \right] \prod_{i=1}^r (1 - p_{(i)}) \right] \end{aligned} $

Table 3

Identification summary measures for $I_k = 10$ when using the 2004 training data to estimate the regression models. These are averaged over 10 implementations.

	Chlamydia						Gonorrhea					
	Tests	PS _e	PS _p	PPV	PNPV		Tests	PS _e	PS _p	PPV	PNPV	
Dorfman	1,651.3	0.668	0.982	0.807	0.963		737.0	0.728	0.997	0.827	0.995	
NIS	1,343.9	0.581	0.987	0.835	0.954		608.5	0.717	0.998	0.886	0.994	
FIS	1,249.8	0.585	0.990	0.869	0.955		575.1	0.743	0.999	0.909	0.995	
Urine	1,361.7	0.595	0.987	0.840	0.956		595.3	0.708	0.998	0.855	0.994	
Female	1,288.6	0.603	0.988	0.855	0.957		587.9	0.709	0.998	0.898	0.994	
2SIS	1,263.8	0.537	0.991	0.865	0.950		559.2	0.653	0.998	0.895	0.993	
3H	1,174.1	0.422	0.997	0.935	0.938		530.2	0.540	1.000	0.958	0.991	
4H												
MP	1,300.1	0.518	0.990	0.856	0.948		639.2	0.606	0.999	0.959	0.992	
Dorfman	2,521.2	0.863	0.975	0.755	0.987		1,082.0	0.948	0.993	0.691	0.999	
NIS	2,073.6	0.833	0.983	0.821	0.985		901.7	0.939	0.996	0.811	0.999	
FIS	1,887.4	0.832	0.987	0.854	0.985		810.6	0.934	0.997	0.823	0.999	
Urine	2,059.7	0.843	0.983	0.819	0.986		822.7	0.939	0.996	0.803	0.999	
2SIS	1,920.9	0.841	0.986	0.843	0.986		818.9	0.949	0.997	0.819	0.999	
3H	2,022.7	0.808	0.987	0.853	0.983		821.7	0.925	0.997	0.846	0.999	
4H	1,944.5	0.751	0.994	0.918	0.978		793.8	0.877	0.999	0.941	0.998	
MP	2,040.6	0.809	0.986	0.841	0.983		942.9	0.908	0.999	0.922	0.999	
Dorfman	10,192.0	0.868	0.984	0.774	0.992		4,001.0	0.942	0.998	0.823	0.999	
NIS	8,245.3	0.832	0.990	0.839	0.990		3,427.4	0.940	0.999	0.887	0.999	
FIS	7,142.9	0.842	0.993	0.886	0.990		2,994.0	0.942	0.999	0.923	0.999	
Swab	7,787.4	0.850	0.991	0.851	0.991		3,059.4	0.938	0.999	0.923	0.999	
2SIS	7,217.1	0.843	0.993	0.879	0.990		2,987.7	0.939	0.999	0.928	0.999	
3H	7,976.7	0.804	0.993	0.870	0.988		3,231.7	0.917	0.999	0.914	0.999	
4H	7,662.1	0.742	0.997	0.936	0.984		3,138.9	0.871	1.000	0.972	0.999	
MP	7,835.2	0.810	0.993	0.884	0.988		4,267.8	0.901	1.000	0.978	0.999	
Dorfman	3,171.0	0.853	0.966	0.790	0.978		2,108.0	0.972	0.985	0.765	0.999	

	Chlamydia					Gonorrhea				
	Tests	PS _e	PS _p	PPV	NPV	Tests	PS _e	PS _p	PPV	NPV
NIS	2,744.4	0.809	0.979	0.853	0.972	1,666.9	0.964	0.991	0.850	0.998
FIS	2,291.7	0.807	0.986	0.894	0.972	1,243.8	0.960	0.996	0.918	0.998
Swab	1SIS	2,674.4	0.818	0.978	0.848	1,356.0	0.967	0.994	0.886	0.998
	2SIS	2,383.2	0.812	0.984	0.886	1,256.7	0.966	0.995	0.913	0.998
Male	3H	2,735.4	0.800	0.982	0.867	1,656.8	0.956	0.992	0.858	0.998
	4H	2,671.1	0.729	0.992	0.930	1,607.7	0.946	0.997	0.941	0.997
MP	2,854.7	0.801	0.978	0.844	0.970	1,578.4	0.953	0.994	0.896	0.998

Identification summary measures for $I_k = 10$ when using the initial groups to estimate the models. These are averaged over 10 implementations.

Table 4

	Chlamydia						Gonorrhea					
	Tests	PS _e	PS _p	PPV	PNPV		Tests	PS _e	PS _p	PPV	PNPV	
Dorfman	1,718.5	0.648	0.981	0.790	0.961		728.7	0.708	0.997	0.820	0.994	
NIS	1,411.5	0.583	0.986	0.822	0.954		615.4	0.692	0.998	0.899	0.994	
FIS	1,333.7	0.596	0.988	0.846	0.956		577.0	0.709	0.999	0.912	0.994	
Urine	1,416.2	0.603	0.987	0.843	0.957		589.6	0.709	0.998	0.893	0.994	
Female	1,361.9	0.582	0.988	0.849	0.954		574.9	0.709	0.999	0.906	0.994	
3H	1,308.0	0.521	0.991	0.864	0.948		556.1	0.613	0.999	0.931	0.992	
4H	1,203.6	0.420	0.996	0.929	0.938		519.5	0.492	1.000	0.963	0.990	
MP	1,303.5	0.526	0.990	0.860	0.949		640.3	0.628	1.000	0.985	0.993	
Dorfman	2,515.4	0.855	0.974	0.746	0.987		1,068.0	0.930	0.993	0.688	0.999	
NIS	2,062.3	0.832	0.984	0.826	0.985		867.2	0.925	0.996	0.786	0.999	
FIS	1,965.9	0.829	0.985	0.834	0.985		796.7	0.931	0.997	0.829	0.999	
Urine	2,085.0	0.839	0.982	0.807	0.985		824.8	0.931	0.997	0.816	0.999	
Male	1,977.3	0.837	0.986	0.841	0.985		804.4	0.916	0.997	0.832	0.999	
3H	1,996.7	0.803	0.985	0.833	0.982		810.4	0.908	0.998	0.865	0.999	
4H	1,924.6	0.736	0.995	0.932	0.976		781.7	0.862	0.999	0.942	0.998	
MP	2,038.5	0.797	0.986	0.838	0.982		948.8	0.910	0.999	0.911	0.999	
Dorfman	10,522.0	0.861	0.984	0.765	0.991		4,028.0	0.938	0.998	0.815	0.999	
NIS	8,453.0	0.839	0.990	0.842	0.990		3,442.5	0.932	0.999	0.874	0.999	
FIS	7,454.6	0.841	0.993	0.877	0.990		3,122.1	0.932	0.999	0.909	0.999	
Swab	8,042.8	0.851	0.990	0.843	0.991		3,181.2	0.926	0.999	0.915	0.999	
Female	7,554.4	0.841	0.992	0.866	0.990		3,139.1	0.925	0.999	0.917	0.999	
3H	8,158.7	0.807	0.992	0.863	0.988		3,234.6	0.898	0.999	0.914	0.999	
4H	7,779.1	0.742	0.997	0.939	0.984		3,144.0	0.877	1.000	0.970	0.999	
MP	7,830.5	0.807	0.994	0.887	0.988		4,268.9	0.908	1.000	0.980	0.999	
Dorfman	3,295.5	0.850	0.968	0.800	0.977		2,113.8	0.971	0.984	0.760	0.998	

	Chlamydia					Gonorrhea				
	Tests	PS _e	PS _p	PPV	NPV	Tests	PS _e	PS _p	PPV	NPV
NIS	2,804.2	0.803	0.979	0.849	0.971	1,683.0	0.966	0.991	0.848	0.998
FIS	2,522.6	0.808	0.982	0.868	0.972	1,272.6	0.964	0.995	0.910	0.998
Swab	1SIS	2,822.9	0.825	0.977	0.840	1,419.6	0.967	0.994	0.885	0.998
	2SIS	2,599.0	0.815	0.980	0.859	1,289.0	0.972	0.996	0.918	0.999
Male	3H	2,757.4	0.781	0.981	0.862	1,674.8	0.959	0.993	0.871	0.998
	4H	2,715.8	0.734	0.992	0.932	1,625.6	0.945	0.996	0.932	0.997
MP	2,828.5	0.796	0.979	0.850	0.970	1,551.2	0.947	0.995	0.900	0.997