Published in final edited form as:

J R Stat Soc Ser C Appl Stat. 2012 March 1; 61(2): 277–290. doi:10.1111/j.1467-9876.2011.01008.x.

Group testing in heterogeneous populations by using halving algorithms

Michael S. Black,

University of Nebraska—Lincoln, USA

Christopher R. Bilder, and

University of Nebraska—Lincoln, USA

Joshua M. Tebbs

University of South Carolina, Columbia, USA

Summary

Group (pooled) testing is often used to reduce the total number of tests that are needed to screen a large number of individuals for an infectious disease or some other binary characteristic. Traditionally, research in group testing has assumed that each individual is independent with the same risk of positivity. More recently, there has been a growing set of literature generalizing previous work in group testing to include heterogeneous populations so that each individual has a different risk of positivity. We investigate the effect of acknowledging population heterogeneity on a commonly used group testing procedure which is known as ‘halving’. For this procedure, positive groups are successively split into two equal-sized halves until all groups test negatively or until individual testing occurs. We show that heterogeneity does not affect the mean number of tests when individuals are randomly assigned to subgroups. However, when individuals are assigned to subgroups on the basis of their risk probabilities, we show that our proposed procedures reduce the number of tests by taking advantage of the heterogeneity. This is illustrated by using chlamydia and gonorrhoea screening data from the state of Nebraska.

Keywords

Binary response; Classification; Identification; Pooled testing; Retesting; Screening

1. Introduction

When a large number of individuals need to be screened for an infectious disease or some other binary characteristic, group testing is often used to reduce the total number of tests that

---

© 2011 Royal Statistical Society

Address for correspondence: Christopher R. Bilder, Department of Statistics, University of Nebraska—Lincoln, Lincoln, NE 68583, USA., chris@chrisbilder.com.

Supporting information

Additional ‘supporting information’ may be found in the on-line version of this article:

‘Supplementary materials for “Group testing in heterogeneous populations using halving algorithms” by Black, Bilder, and Tebbs’.

Please note: Wiley–Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors.

Any queries (other than missing material) should be directed to the author for correspondence for the article.

are needed. Group testing, which is also known as pooled testing, refers to the process of combining individual specimens (e.g. urine or blood) into a ‘pooled’ specimen for testing. If the group (pool) tests negatively, all individuals within it are declared negative. If the group tests positively, retesting is needed to decode the positive and negative individuals. This idea was introduced by Dorfman (1943) as a way to screen World War II soldiers for syphilis. For this situation, Dorfman proposed simply to retest all subjects individually within the positive groups. Other retesting procedures have been proposed since then, and many of them result in a smaller number of tests; see Gupta and Malina (1999) and Hughes-Oliver (2006) for a review. The usefulness of group testing to identify positive individuals has been well established in many areas; these areas include blood donation screening (Dodd *et al.*, 2002), opportunistic testing of individuals for chlamydia (Mund *et al.*, 2008), bovine viral diarrhoea virus detection in cattle herds (Kennedy *et al.*, 2006) and discovery of chemical compounds to use in new drugs (Remlinger *et al.*, 2006).

Traditionally, research in group testing has assumed each individual to be independent with the same risk of positivity p , i.e. a homogeneous population of independent individuals with an overall prevalence p . More recently, there has been an expanding set of literature that generalizes past work to include heterogeneous populations. In this setting, each individual has their own individual probability of positivity, and heterogeneity can be modelled by using the group testing regression methods of Vansteelandt *et al.* (2000) or Xie (2001). Bilder *et al.* (2010) showed further how estimates of these individual probabilities can be used to retest individuals in a positive group, and they demonstrated how one can reduce significantly the number of tests that are needed through an extension of Sterrett’s (1957) identification procedure.

Given these advances, it is now important to determine whether accounting for population heterogeneity provides benefits with other retesting procedures that are used in practice. One widely used procedure involves successively splitting positive groups into smaller subgroups until all positive and negative individuals have been identified (Sobel and Groll, 1959; Johnson *et al.*, 1991; Pilcher *et al.*, 2005; Kim *et al.*, 2007). A common example of this retesting approach is to form subgroups which are halves of larger groups; we refer to this as ‘halving’. Litvak *et al.* (1994) popularized the halving technique in the context of blood donation screening for human immunodeficiency virus. In our paper, we generalize the use of halving to heterogeneous population settings.

Our work is motivated by chlamydia and gonorrhoea screening performed by the Nebraska Public Health Laboratory (NPHL). In this setting, clinical, demographic and risk behaviour information is available on each individual being screened. Because these risk factors have a strong relationship with whether or not an individual is infected, it is natural to consider the screening population as being heterogeneous. Through exploiting this heterogeneity, we examine how well new halving procedures can reduce the total number of tests that are needed for screening.

The paper is ordered as follows. In Section 2, we derive the mean, variance and probability mass function (PMF) for the number of tests under halving. When compared with a homogeneous population setting, we prove that the mean and variance do not change if

individuals from a heterogeneous population are assigned to subgroups at random. Using the derivations from Section 2, we propose a new halving procedure in Section 3 that exploits risk heterogeneity to reduce the expected number of tests, and we identify in Section 4 the situations where our new procedure performs best. In Section 5, we apply our new procedure to chlamydia and gonorrhoea screening data from Nebraska. Finally, Section 6 summarizes our results and provides recommendations for application.

2. Halving

2.1. Moments for a fixed set of individual risk probabilities

We begin by assuming that each individual is assigned to exactly one initial group. For the remainder of this section and Section 3, we focus on one particular initial group of size I where individual i has risk probability p_i for $i = 1, \dots, I$. Later sections examine individuals across all initial groups.

Halving involves successively splitting positive groups into two equal-sized halves. Positive groups are halved until all groups test negatively or until individual testing occurs. For example, three-step halving for an initial group of size $I = 16$ begins by testing the entire group. If the group tests positively, the second step involves splitting it into two subgroups of size 8. If either subgroup tests positively, a third and final step occurs where each individual in a positive subgroup is tested. A four-step halving protocol with $I = 16$ would continue with halving into groups of size 4 before individual testing. A larger number of steps can be performed in a similar manner until only individuals remain.

We now derive the operating characteristics of halving for a heterogeneous group. Let $G_{s,j} = 1$ and $G_{s,j} = 0$ respectively denote a positive and negative test result for the j th subgroup at the s th step for $j = 1, \dots, 2^{s-1}$ and $s = 1, \dots, S$. In the last example, $G_{1,1}$ represents the test result for the initial group of size 16, and $G_{2,1}$ represents the test result for the first subgroup of size 8 halved from a positive initial group. In a three-step setting, we can write the expected number of tests for an initial group of size I as

$$E(T|\mathbf{p}^{\text{vec}}) = 1 + 2P(G_{1,1}=1) + I_{2,1}P(G_{1,1}=1 \cap G_{2,1}=1) + I_{2,2}P(G_{1,1}=1 \cap G_{2,2}=1),$$

where T is the number of tests, $\mathbf{p}^{\text{vec}} = (p_1, \dots, p_I)'$ and $I_{s,j}$ is the number of individuals remaining in the j th subgroup at step s . Adding a fourth step leads to an expected number of tests

$$\begin{aligned} E(T|\mathbf{p}^{\text{vec}}) = & 1 + 2P(G_{1,1}=1) + 2P(G_{1,1}=1 \cap G_{2,1}=1) + 2P(G_{1,1}=1 \cap G_{2,2}=1) \\ & + I_{3,1}P(G_{1,1}=1 \cap G_{2,1}=1 \cap G_{3,1}=1) + I_{3,2}P(G_{1,1}=1 \cap G_{2,1}=1 \cap G_{3,2}=1) \\ & + I_{3,3}P(G_{1,1}=1 \cap G_{2,2}=1 \cap G_{3,3}=1) + I_{3,4}P(G_{1,1}=1 \cap G_{2,2}=1 \cap G_{3,4}=1). \end{aligned}$$

In general for an S -step halving procedure, it follows that

$$E(T|\mathbf{p}^{\text{vec}}) = 1 + 2 \sum_{s=1}^{S-2} \sum_{j=1}^{2^{s-1}} P \left(\left(\{(s', j') : G_{s,j}=1\} \right) \cap \{G_{s',j'}=1\} \right) + \sum_{j=1}^{2^{S-2}} I_{S-1,j} P \left(\left(\{(s', j') : G_{S-1,j}=1\} \right) \cap \{G_{s',j'}=1\} \right) \quad (1)$$

for an appropriate number of steps S given the initial group size. When an odd-sized group is halved, final step group sizes $I_{S,j}$ can be set equal to 0. For example, a four-step halving procedure with $I = 7$ can have an initial split with subgroups of size 4 and 3. The group of size 3 can be split further into groups of size 2 and 1. Because the ‘group’ of size 1 cannot be split again, we can set its $I_{4,j}$ equal to 0 so that its corresponding term is excluded from the mean calculation.

Each of the probabilities in the above expressions is found by taking into account the true group statuses. Let $\tilde{G}_{s,j} = 1$ and $\tilde{G}_{s,j} = 0$ respectively denote a positive and negative true status for the j th subgroup at the s th step, and define the test sensitivity and specificity as $S_e = P(G_{s,j} = 1 | \tilde{G}_{s,j} = 1)$ and $S_p = P(G_{s,j} = 0 | \tilde{G}_{s,j} = 0)$ respectively. The probability that the initial group tests positively can be written as

$$\begin{aligned} P(G_{1,1}=1) &= P(G_{1,1}=1 \cap \tilde{G}_{1,1}=0) + P(G_{1,1}=1 \cap \tilde{G}_{1,1}=1) \\ &= P(G_{1,1}=1 | \tilde{G}_{1,1}=0) P(\tilde{G}_{1,1}=0) + P(G_{1,1}=1 | \tilde{G}_{1,1}=1) P(\tilde{G}_{1,1}=1) \\ &= (1 - S_p) \prod_{i=1}^I (1 - p_i) + S_e \left\{ 1 - \prod_{i=1}^I (1 - p_i) \right\}. \end{aligned}$$

Probabilities involving groups for steps 2 and higher become more complicated to derive because past steps must be taken into account. For example, the probability of positivity for the first group at step 2, after the initial group has tested positively, is

$$\begin{aligned} P(G_{1,1}=1 \cap G_{2,1}=1) &= P(G_{1,1}=1 \cap G_{2,1}=1 \cap \tilde{G}_{1,1}=0 \cap \tilde{G}_{2,1}=0) \\ &\quad + P(G_{1,1}=1 \cap G_{2,1}=1 \cap \tilde{G}_{1,1}=1 \cap \tilde{G}_{2,1}=0) \\ &\quad + P(G_{1,1}=1 \cap G_{2,1}=1 \cap \tilde{G}_{1,1}=1 \cap \tilde{G}_{2,1}=1), \end{aligned}$$

which takes into account the three ways that $\{G_{1,1} = 1\} \cap \{G_{2,1} = 1\}$ may occur with respect to the true statuses. Continuing, we obtain

$$P(G_{1,1}=1 \cap G_{2,1}=1) = (1 - S_p)^2 \prod_{i=1}^I (1 - p_i) + S_e (1 - S_p) \left\{ \prod_{i \in B_{2,1}} (1 - p_i) \right\} \left\{ 1 - \prod_{i \in B_{2,2}} (1 - p_i) \right\} + S_e^2 \left\{ 1 - \prod_{i \in B_{2,1}} (1 - p_i) \right\},$$

where $i \in B_{s,j}$ is understood to mean those individuals who belong to the j th group at the s th step and where we make the standard assumption that the test outcomes are conditionally independent given the true statuses (see Litvak *et al.* (1994)). These results can be generalized for $s > 1$ to

$$\begin{aligned}
P\left(\bigcap_{\{(s',j'):G_{s,j}=1\}}\{G_{s',j'}=1\}\right) &= (1-S_p)^s \prod_{i=1}^I (1-p_i) \\
+ \sum_{a=1}^{s-1} S_e^a (1-S_p)^{s-a} &\left\{ \prod_{i \in B_{a+1,m}} (1-p_i) \right\} \left\{ 1 - \prod_{i \in \bar{B}_{a+1,m}} (1-p_i) \right\} \\
&+ S_e^s \left\{ 1 - \prod_{i \in B_{s,j}} (1-p_i) \right\},
\end{aligned} \quad (2)$$

where $m = \lceil j/2^{s-1-a} \rceil$ and $i \in \bar{B}_{s,j}$ denotes the set of individuals within the parent group of $B_{s,j}$ excluding those in $B_{s,j}$ itself (for example, $i \in \bar{B}_{3,3}$ denotes all individuals in $B_{3,4}$ because $\{i \in B_{3,3}\} \cup \{i \in B_{3,4}\} = \{i \in B_{2,2}\}$). Substituting equation (2) into equation (1) gives the expected number of tests for a specific set of risk probabilities.

To find the variance, we need to calculate the second moment for T . For a three-step procedure,

$$\begin{aligned}
E(T^2 | \mathbf{p}^{\text{vec}}) &= P(G_{1,1}=0) + 3^2 P(G_{1,1}=0 \cap G_{2,1}=0 \cap G_{2,2}=0) \\
&+ (3+I_{2,1})^2 P(G_{1,1}=0 \cap G_{2,1}=1 \cap G_{2,2}=0) \\
&+ (3+I_{2,2})^2 P(G_{1,1}=0 \cap G_{2,1}=1 \cap G_{2,2}=0) \\
&+ (3+I)^2 P(G_{1,1}=0 \cap G_{2,1}=1 \cap G_{2,2}=1).
\end{aligned}$$

The four probability terms in this expression are found by using equation (2). For four-step and higher procedures, the number of terms grows very quickly, so we do not recommend direct evaluation. Instead, in Appendix A, we present a recursive algorithm to calculate the PMF of T by exploiting the hierarchical nature that is inherent to the halving procedure. Combining the PMF with the standard variance formula leads to the result desired.

2.2. Treating risk probabilities as random

Individual risk probabilities will vary from group to group. Therefore, in this subsection, we treat these probabilities as random and re-examine our moment calculations. Specifically, we now envision p_i as independent random variables with $E(p_i) = p$ for $i = 1, \dots, I$. The overall expected number of tests is

$$\begin{aligned}
E(T) &= E\{E(T | \mathbf{p}^{\text{vec}})\} \\
&= 1 + 2 \sum_{s=1}^{S-2} \sum_{j=1}^{2^{s-1}} E \left[P \left(\bigcap_{\{(s',j'):G_{s,j}=1\}} \{G_{s',j'}=1\} \right) \right] \\
&+ \sum_{j=1}^{2^{S-2}} I_{S-1,j} E \left[P \left(\bigcap_{\{(s',j'):G_{S-1,j}=1\}} \{G_{s',j'}=1\} \right) \right].
\end{aligned} \quad (3)$$

The expectation of the joint probability in equation (3) is

$$\begin{aligned}
& E \left[P \left(\bigcap_{\{(s',j'): G_{s,j}=1\}} \{G_{s',j'}=1\} \right) \right] \\
&= (1-S_p)^s E \left\{ \prod_{i=1}^I (1-p_i) \right\} \\
&+ \sum_{a=1}^{s-1} S_e^a (1-S_p)^{s-a} E \left\{ \prod_{i \in B_{a+1,m}} (1-p_i) \right\} \left[1 - E \left\{ \prod_{i \in \bar{B}_{a+1,m}} (1-p_i) \right\} \right] \quad (4) \\
&+ S_e^s \left[1 - E \left\{ \prod_{i \in B_{s,j}} (1-p_i) \right\} \right].
\end{aligned}$$

Because of independence among the individual probabilities, equation (4) simplifies to

$$\begin{aligned}
& E \left[P \left(\bigcap_{\{(s',j'): G_{s,j}=1\}} \{G_{s',j'}=1\} \right) \right] \\
&= (1-S_p)^s (1-p)^{u_{1,1}} + \sum_{a=1}^{s-1} S_e^a (1-S_p)^{s-a} (1-p)^{u_{a+1,m} - l_{a+1,m}} \{1 - (1-p)^{\bar{u}_{a+1,m} - \bar{l}_{a+1,m}}\} \quad (5) \\
&+ S_e^s \{1 - (1-p)^{u_{s,j} - l_{s,j}}\},
\end{aligned}$$

where $l_{s,j} = \sum_{i=1}^j I_{s,i} - I_{s,j} + 1$ and $u_{s,j} = \sum_{i=1}^j I_{s,i}$ are the lowest and highest subscripts respectively for the individuals in subgroup $B_{s,j}$ and $l_{s,j}$ and $u_{s,j}$ are the lowest and highest subscripts respectively for the individuals in subgroup $B_{s,j}$. The expected number of tests $E(T)$ is found by substituting equation (5) into equation (3).

It is especially insightful to note that $P(\cap_{\{(s',j'): G_{s,j}=1\}} \{G_{s',j'}=1\})$ reduces to equation (5) when all individuals have a common risk probability p ; this implies that the unconditional means are the same for homogeneous or heterogeneous population assumptions. Furthermore, we show in the on-line supplementary materials that $\text{var}(T)$ also remains unchanged. Therefore, when individuals with different risks are assigned randomly to groups, neither $E(T)$ nor $\text{var}(T)$ is affected. This is reassuring if the researcher cannot account for heterogeneity when implementing the halving procedure.

An important generalization of these results is that they can be extended to other commonly used retesting algorithms, such as Dorfman's (1943) procedure and Sterrett's (1957) procedure, where moments can also be written in terms of $\prod_i (1-p_i)$. This is due to the underlying independence of the risk probabilities. For example, Bilder *et al.* (2010) gave the PMF for T in a 'three-stage' informative Sterrett procedure. If we treat the individual risk probabilities as independent random variables, all of their $P(T=t)$ expressions rely on these simple products.

3. Ordered halving

We have shown that the moment formulae for T do not depend on the individual risk probabilities when individuals are assigned to subgroups at random. Instead of random assignment, we now control how individuals are assigned to subgroups. Our overall goal is to assign individuals to subgroups in a manner that reduces the expected number of tests.

After an initial group of size I tests positively, two subgroups of equal size are created. Our goal is to maximize one subgroup's probability of testing positively and to maximize the other subgroup's probability of testing negatively. This type of subgroup construction attempts to keep the positive individuals in the same grouping to lessen the number of further retests. Define a set of ordered risk probabilities for an initial group of size I as $\mathbf{p}^{\text{ord}} = (p_{(1)}, \dots, p_{(I)})'$ where $p_{(i)}$ denotes the i th smallest probability within the group. The second step of 'ordered halving' creates one subgroup of individuals with lower risks $p_{(1)}, \dots, p_{(I/2,1)}$ and one subgroup of individuals with higher risks $p_{(I/2,1+1)}, \dots, p_{(I)}$. If one of these subgroups tests positively and $S \geq 4$ (i.e. individual testing does not occur at step 3 for positive subgroups), the process of halving groups by the ordered risks continues in a similar manner.

To compare the expected number of tests with and without ordering when subgroup sizes are equal, i.e. to compare $E(T|\mathbf{p}^{\text{ord}})$ and $E(T|\mathbf{p}^{\text{vec}})$, we need only to focus on

$$\sum_{j=1}^{2^{s-1}} P(\cap_{\{(s',j'): G_{s,j}=1\}} \{G_{s',j'}=1\})$$

for each step $s = 1, \dots, S-1$ of equation (1). This is true because ordering only changes expressions that are functions of the risk probabilities. To help with the comparison, note that

$$\sum_{j=1}^{2^{s-1}} P\left(\cap_{\{(s',j'): G_{s,j}=1\}} \{G_{s',j'}=1\}\right) = 2^{s-1} S_e^s + \{(1-S_p) - S_e\} \times \sum_{a=1}^s 2^{s-a} S_e^{a-1} (1-S_p)^{s-a} \sum_{j=1}^{2^{a-1}} \prod_{i \in B_{a,j}} (1-p_i). \quad (6)$$

When $s = 1$, equation (6) is the same regardless of whether subgroup assignment is ordered or random. However, for any step $a > 1$, one can show that ordering the individual risk

probabilities maximizes $\sum_{j=1}^{2^{a-1}} \prod_{i \in B_{a,j}} (1-p_i)$. Thus, equation (6) is minimized under ordered assignment as long as $S_e > 1 - S_p$, which will be true for any diagnostic test that is used in application. This shows that $E(T|\mathbf{p}^{\text{ord}}) \leq E(T|\mathbf{p}^{\text{vec}})$ whenever our subgroup construction is used.

To find $E(T) = E\{E(T|\mathbf{p}^{\text{ord}})\}$, we make use of equations (1) and (4) again where the individual risk probabilities within equation (4) are properly ordered for the subgroups. Because the expectations in equation (4) are now distribution dependent, a simple expression for $E(T)$ no longer exists. However, we can use a result from Junjiro (1962) to find the distribution of the ordered risk probabilities. This distribution is

$$f(p_{(l_{s,j})}, \dots, p_{(u_{s,j})}) = \frac{I!}{(l_{s,j}-1)!(I-u_{s,j})!} F(p_{(l_{s,j})})^{l_{s,j}-1} \left\{ \prod_{i=l_{s,j}}^{u_{s,j}} f(p_{(i)}) \right\} \{1-F(p_{(u_{s,j})})\}^{I-u_{s,j}},$$

where $p_{(l_{s,j})} \dots p_{(u_{s,j})}$ are the ordered risk probabilities for individuals in group $B_{s,j}$ (see Section 2.2), $f(p_i)$ is the probability density function for p_i and $F(p_i)$ is the cumulative distribution function for p_i . Using this distribution, moments for T can be found by substituting the expected values into equation (3). We examine values of $E(T)$ for specific distributions in Section 4.

4. Mean comparisons

Group testing is of most value in situations where the overall prevalence is small. To understand how well ordered halving works in practice, we take $p = 0.005, 0.01, 0.05, 0.10$ and examine the number of tests performed. The distributions that are chosen for p_i are a $\text{beta}(1, 1/p - 1)$, a $\text{uniform}(0, 2p)$ and a degenerate at p distribution (which corresponds to a homogeneous population of individuals). We also look at an ‘extreme case’ of $p_i = 1$ with probability p and $p_i = 0$ with probability $1 - p$. Although this last case is unrealistic, it is useful to examine because it maximizes the variance among the individual probabilities. For all distributions, the expected value of p_i is p , but the variances are different. For example, the variances are 0.048, 0.0023 and 0.0002 for the extreme, beta and uniform cases respectively, when $p = 0.05$, and this ordering among the distributions occurs for the other values of p as well.

We compare the expected number of tests for these different distributions by using halving with two, three, four and five steps for various group sizes. To make comparisons on a realistic numerical scale, we convert the expected number of tests for a single group into the expected number of tests in a population of 10000 individuals. We use the equations that were derived in Sections 2.2 and 3 to calculate the expected number of tests. For the beta distributions, it is necessary to estimate the expected values because of the difficulty in integrating over the distribution of the order statistics. For the degenerate case, the expected number of tests and the variance for the number of tests are calculated by using the PMF algorithm that is described in Appendix A.

For each level of overall risk and number of steps considered, Table 1 gives the expected number of tests for a selected number of group sizes. The group sizes selected are those that minimize the expected number of tests in the degenerate case. For example, the expected number of tests for the degenerate case with $p = 0.05$ is the smallest for two-step halving (Dorfman’s procedure) when the group size is 5. It is common for other group sizes to exist where ordered halving has a smaller expected number of tests for the same S ; thus, the expected benefits from ordering will be no worse than those presented here. Although perfect testing does not often occur in actual applications, we assume that $S_p = S_e = 1$ because it provides a useful initial examination.

Table 1 shows that the degenerate case always results in the maximum expected number of tests among the four distributions. For a two-step procedure, there is no decrease in the expected number of tests from ordered halving; ordering risk probabilities has no advantage when the second step is individual testing. For three-steps and higher, ordered halving always leads to a decrease in the expected number of tests. This decrease can be limited for smaller p , but it is more substantial for larger p . We also note that, as the variance among the

risk probabilities increases, the expected number of tests decreases. This result is intuitive because the more diversity in information available (in terms of the risk probabilities) the easier it is for an ‘informative retesting’ procedure to find positive individuals. Exceptions can occur when the last halving step results in uneven group sizes (e.g. four-step with a group size of 10 when $p = 0.05$), because we choose to have the larger risk probabilities in the larger subgroup.

Fig. 1 plots the expected number of tests when $p = 0.05$ for a number of group sizes and levels of sensitivity and specificity. Additional plots for $p = 0.005, 0.01, 0.10$ are available in the on-line supplementary materials. Fig. 1 provides additional evidence that ordered halving reduces the expected number of tests, even in the presence of imperfect testing. In addition, we see that testing error does not change the relative ordering between the distribution cases. Furthermore, the group size that results in the smallest number of tests can be larger for ordering than for the degenerate case. The meaningfulness of this result may be tempered if dilution effects prevent the use of larger group sizes.

5. Application

The Infertility Prevention Project is a nationally implemented programme whose goals are to assess and reduce the prevalence of chlamydia and gonorrhoea in the USA. In Nebraska, urine and swab specimens are collected from individuals visiting health clinics throughout the state. These specimens are then sent to the NPHL, where each specimen is tested individually for both infections. Clinical, demographic and risk behaviour information is recorded for each individual before testing. Therefore, it is sensible to envision individuals as having different probabilities of positivity, which leads to a potential application of ordered halving.

To assess how well ordered halving would work in this application, we use previously diagnosed individual statuses from the NPHL in the following manner. The NPHL’s year 2004 results are used as a training data set to estimate the probability of positivity for individuals who were tested in 2005. First-order logistic regression models are fitted to the training data with the response variable as disease status and the explanatory variables of age, race, type of clinic, location of clinic, reason for visit, symptoms, initial clinical observations and risk history. These models are fitted separately by disease (chlamydia and gonorrhoea), gender and type of specimen (swab or urine), and summaries of the estimated individual probabilities are given in the on-line supplementary materials. The year 2005 individuals are ordered by date of specimen and are placed in successive groups by disease–gender–specimen combination. Assuming that the observed 2005 diagnoses are the true responses, we simulate the halving process for each group, where simulated test responses are generated with the S_e - and S_p -values provided by the NPHL. We repeat halving for each disease–gender–specimen combination 10 times to account for simulation variability, and we record the average number of tests.

Table 2 displays the average number of tests at specific group sizes. This table provides the chlamydia screening results only. Similar results are found for gonorrhoea screening, which are given in the on-line supplementary materials. Overall, we find that the chlamydia results

are similar to those found for the beta distribution cases in Section 4. This is not surprising because a beta distribution often fits these individual probabilities well and overall prevalences range from 5.8% to 13.0% for each gender–specimen combination (Table 2 lists the prevalences for 2005). Generally, improvements from ordering are from 1% to 6%, where some improvements are larger for the swab–male combination (up to 10.49%). Also, the benefits from ordered halving are more substantial for larger group sizes and prevalences, which is consistent with our findings in Section 4.

We also investigated four measures of classification accuracy. The pooling sensitivity and the pooling specificity are defined respectively as the proportion of true positives and true negatives that are diagnosed respectively as positive and negative through using group testing. The pooling positive and pooling negative predictive values are defined respectively as the proportion of individuals who test positively and negatively through using group testing who are respectively truly positive and truly negative. These summary measures are displayed in Table 3 for an initial group size of 16. Summary measures for other initial group sizes are given in the on-line supplementary materials. Overall, there are no discernible increases or decreases through using ordered halving when compared with unordered halving. One exception may occur with the pooling specificity and pooling positive predictive value, where ordering can provide slightly higher accuracy; however, this does not occur consistently across all the initial group sizes that were examined.

As the number of steps increases, the pooling sensitivity and pooling negative predictive values decrease, but the pooling specificity and pooling positive predictive values increase. These are not necessarily new discoveries; for example, see Kim *et al.* (2007), who documented similar occurrences in a homogeneous population setting. However, these findings illustrate the potential weaknesses and strengths of multiple-step group testing. For example, as evidenced by some of the lower pooling sensitivity levels, it is generally undesirable to apply the higher step procedures when S_e is small. We provide suggestions in Section 6 on how to improve the measures of accuracy in these cases.

6. Conclusions

We have generalized the use of halving algorithms in group testing to heterogeneous population settings. Our results show that ordering risk probabilities reduces the number of tests that are needed to classify all individuals as positive or negative, while maintaining similar levels of accuracy to those with unordered halving. This reduction in the number of tests increases as the variation in the risk probabilities increases. Also, the test reduction grows as the overall prevalence increases. An intuitive explanation for this occurrence comes through examining the possible number of tests with halving. For simplicity, assume that $S_e = S_p = 1$. When there are no positive or only one positive individual within a group at step 1, ordered halving results in the same number of tests as without ordering. When there are two or more positive individuals within a group at step 1, ordered halving pools the larger probability individuals together. This leads to a larger probability that all positive individuals are within one half rather than in both halves, which reduces the potential number of tests remaining. Thus, ordered halving on a group is beneficial only when there is

more than one positive individual within the group. This is why ordered halving can have larger optimal group sizes.

Although ordered halving reduces the expected number of tests, the reduction may not be enough to warrant its use in some lower prevalence cases. However, this reduction can be magnified greatly for high volume and/or large assay cost situations. For example, Kim and Hudgens (2009), page 903, described a detection programme for human immunodeficiency virus in North Carolina where ‘slight improvements in efficiency can lead to substantial cost savings’ because 120000 specimens are screened per year. In addition, the American Red Cross screens millions of blood donations for multiple diseases per year by group testing (Stramer *et al.*, 2004; Dodd *et al.*, 2002), so even small improvements can translate to a large number of tests saved.

The direct application of multiple-step group testing procedures may be undesirable in situations where an assay’s sensitivity is low. This was evident in testing female urine for chlamydia in Section 5, where pooling sensitivity levels dropped significantly with each additional step. There has been only limited research on how to increase this accuracy. For the homogeneous population setting, a proposal by Litvak *et al.* (1994) includes a second test for any group that tests negatively. We would expect this confirmatory practice also to improve the accuracy of ordered halving by a similar amount. Future research should investigate how heterogeneity among individuals can be exploited to determine whether more benefits are possible.

When compared with unordered halving, the added complexity that arises from using our ordered halving procedure is minimal, where estimating the risk probability for each individual is the main addition. For simplicity, estimation could be done informally; for example, one could use covariate values that are labelled on a specimen’s storage tube (or that are obviously identified by seeing the specimen) and quickly categorize each individual as ‘low risk’, ‘medium risk’ or ‘high risk’, on the basis of the values of these covariates. Subgroups could then be formed by using these categorizations as needed. More generally, risk probabilities can be estimated per individual by using regression methods. This is possible if covariate information is readily available on each individual and has been entered into a database for future analysis, as in the case of the NPHL.

Our results from Sections 4 and 5 lead us to future research areas that could further improve halving. First, we showed that the variation in the risk probabilities was important, but its magnitude of importance changes when uneven subgroup sizes are needed. Future research should examine whether there are optimal unequal subgroup sizes that could be chosen at each step of the group splitting process. Variations on this idea include immediate individual testing for those individuals with a large positive probability. In fact, we see an informal application of this already at the Nebraska Veterinary Diagnostic Laboratory, where the presence of something ‘unusual’ in a preputial wash (e.g. pus or blood) instigates individual testing for trichomonas in bulls. Although this is an intuitive idea, research is needed to determine actual benefits. Second, group splitting could involve more than two subgroups. For example, Pilcher *et al.* (2005) used an initial group size of 90 and subsequent splits into nine groups of size 10 when the initial group tests positively. It would be of interest to

determine how ordering can further reduce the number of tests that are needed when multiple subgroups are used. Choosing the optimal subgroup sizes and the number of subgroups for a split are open research problems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the Joint Editor, the Associate Editor and the two referees for their comments that led to an improved paper. The authors also thank Dr Peter Iwen, Dr Steven Hinrichs and Philip Medina for their consultation on the Infertility Prevention Project in Nebraska. This research is supported by grant R01 AI067373 from the National Institutes of Health.

References

- Bilder C, Tebbs J, Chen P. Informative retesting. *J Am Statist Ass.* 2010; 105:942–955.
- Dodd R, Notari E, Stramer S. Current prevalence and incidence of infectious disease markers and estimated window-period risk in the American Red Cross donor population. *Transfusion.* 2002; 42:975–979. [PubMed: 12385406]
- Dorfman R. The detection of defective members of large populations. *Ann Math Statist.* 1943; 14:436–440.
- Gupta D, Malina R. Group testing in presence of classification errors. *Statist Med.* 1999; 18:1049–1068.
- 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.
- Johnson, N.; Kotz, S.; Wu, X. *Inspection Errors for Attributes in Quality Control.* New York: Chapman and Hall; 1991.
- Junjiro, O. Distribution and moments of order statistics. In: Sarhan, A., editor. *Contributions to Order Statistics.* New York: Wiley; 1962.
- Kennedy J, Mortimer R, Powers B. Reverse transcription-polymerase chain reaction on pooled samples to detect bovine viral diarrhea virus by using fresh ear-notch-sample supernatants. *J Veter Diag Investgn.* 2006; 18:89–93.
- Kim H, Hudgens M. Three dimensional array based group testing algorithms. *Biometrics.* 2009; 65:903–910. [PubMed: 19053996]
- 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. *J Am Statist Ass.* 1994; 89:424–434.
- Mund M, Sander G, Potthoff P, Schicht H, Matthias K. Introduction of *Chlamydia trachomatis* screening for young women in Germany. *J Deuts Derm Gesell.* 2008; 6:1032–1037.
- 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 Engl J Med.* 2005; 352:1873–1883. [PubMed: 15872202]
- R Development Core Team. *R: a Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing; 2010.
- Remlinger K, Hughes-Oliver J, Young S, Lam R. Statistical design of pools using optimal coverage and minimal collision. *Technometrics.* 2006; 48:133–143.
- Sobel M, Groll P. Group testing to eliminate efficiently all defectives in a binomial sample. *Bell Syst Tech J.* 1959; 38:1179–1252.

- Stramer SL, Glynn SA, Kleinman SH, Strong DM, Caglioti S, Wright DJ, Dodd RY, Busch MP. Detection of HIV-1 and HCV infections among antibody-negative blood donors by nucleic acid–amplification testing. *New Engl J Med*. 2004; 351:760–768. [PubMed: 15317889]
- Sterrett A. On the detection of defective members of large populations. *Ann Math Statist*. 1957; 28:1033–1036.
- 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. *Statist Med*. 2001; 20:1957–1969.

Appendix A

This appendix shows how to calculate recursively the PMF for the number of tests using the halving procedure. To begin, consider the case of $S = 2$ (Dorfman's procedure), and let $a = S_p$, $b = 1 - S_p$, $c = 1 - S_e$ and $d = S_e$. There are four possible combinations of test outcomes and true statuses for this situation because $G_{1,1} = 0$ or $G_{1,1} = 1$ and $\tilde{G}_{1,1} = 0$ or $\tilde{G}_{1,1} = 1$. To find the PMF for the number of tests with $S = 2$, let

$$\mathbf{E}_2 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

be a matrix of possible testing errors, $\mathbf{P}_{2,I_{1,1}} = (P(\tilde{G}_{1,1} = 0), P(\tilde{G}_{1,1} = 1))'$ be a vector of probabilities for the true statuses and $\mathbf{T}_{2,I_{1,1}} = (1, 1 + I_{1,1})'$ be a vector for the number of tests. Note that the subscript $I \equiv I_{1,1}$ is used to denote the number of individuals in the top node of the group testing procedure. The PMF for $\mathbf{T}_{2,I_{1,1}}$, where positive, is $\mathbf{P}_{2,I_{1,1}}' \mathbf{E}_2$.

For $S = 3$, we use the fact that the two subgroups containing $I_{2,1}$ and $I_{2,2}$ individuals can be tested as separate two-step procedures. This leads to the matrices

$$\mathbf{E}_3 = ((a, c, c, c)' \parallel \text{diag}\{(b, d, d, d)\} \times (\mathbf{E}_2 \otimes \mathbf{E}_2)) = \begin{pmatrix} a & aab & abb & bab & bbb \\ c & acd & add & bcd & bdd \\ c & cad & cbd & dad & dbd \\ c & ccd & ccd & dcd & ddd \end{pmatrix},$$

$$\mathbf{P}_{3,I_{1,1}} = \mathbf{P}_{2,I_{2,1}} \otimes \mathbf{P}_{2,I_{2,2}}$$

$$= (P(\tilde{G}_{1,1}=0), P(\tilde{G}_{1,1}=1 \cap \tilde{G}_{2,1}=0 \cap \tilde{G}_{2,2}=1), P(\tilde{G}_{1,1}=1 \cap \tilde{G}_{2,1}=1 \cap \tilde{G}_{2,2}=0), P(\tilde{G}_{1,1}=1 \cap \tilde{G}_{2,1}=1 \cap \tilde{G}_{2,2}=1))',$$

$$\mathbf{T}_{3,I_{1,1}} = (1, \mathbf{j}_2' \otimes \mathbf{T}_{2,I_{2,1}} + \mathbf{T}_{2,I_{2,2}} \otimes \mathbf{j}_2' + \mathbf{j}_4')' = (1, 3, 3 + I_{2,1}, 3 + I_{2,2}, 3 + I_{1,1})',$$

where ' $\parallel$ ' denotes vertical concatenation, $\mathbf{j}_m$ denotes an $m \times 1$ vector of 1s and ' $\otimes$ ' denotes a Kronecker product. The corresponding probabilities for $\mathbf{T}_{3,I_{1,1}}$ can be found from $\mathbf{P}_{3,I_{1,1}}' \mathbf{E}_3$. After summing these probabilities over the same number of tests in $\mathbf{T}_{3,I_{1,1}}$, we obtain the PMF for the unique number of tests.

To generalize for any number of steps S , we start with the last step before individual testing and build up. Let $\mathbf{T}_{2,I_{S-1,j}} = (1, 1 + I_{S-1,j})'$, $\mathbf{P}_{2,I_{S-1,j}} = (P(\tilde{G}_{S-1,j} = 0), P(\tilde{G}_{S-1,j} = 1))'$ for $j = 1$,

$\dots, 2^{S-2}$ and $\mathbf{E}_2$ be the same as before. In reverse order from how the testing is actually done, we successively build new matrices

$$\begin{aligned}\mathbf{T}_{s,I_{S-s+1,j}} &= (\mathbf{1}, \mathbf{j}'_k \otimes \mathbf{T}_{s-1,I_{S-s+2,2j-1}} + \mathbf{T}_{s-1,I_{S-s+2,2j}} \otimes \mathbf{j}'_k + \mathbf{j}'_{k^2})', \\ \mathbf{P}_{s,I_{S-s+1,j}} &= \mathbf{P}_{s-1,I_{S-s+2,2j-1}} \otimes \mathbf{P}_{s-1,I_{S-s+2,2j}},\end{aligned}$$

and $\mathbf{E}_s = ((a, c\mathbf{j}'_m)' \parallel \text{diag}\{(b, d\mathbf{j}'_m)\}) \times (\mathbf{E}_{s-1} \otimes \mathbf{E}_{s-1})$ for $s = 3, \dots, S$ and $j = 1, \dots, 2^{S-s}$, where m is 1 less than the number of rows in $\mathbf{E}_{s-1} \otimes \mathbf{E}_{s-1}$ and k is the number of rows in $\mathbf{T}_{s-1,I_{S-s+2,2j}}$. Our final resulting matrices will be $\mathbf{T}_{S,I_{1,1}}$, $\mathbf{P}_{S,I_{1,1}}$ and $\mathbf{E}_S$. The corresponding probabilities for $\mathbf{T}_{S,I_{1,1}}$ can be found from $\mathbf{P}'_{S,I_{1,1}} \mathbf{E}_S$. We sum these probabilities over the same number of tests in $\mathbf{T}_{S,I_{1,1}}$ to obtain the PMF for the unique number of tests.

These operations can be continued for any number of steps, and our algorithm is designed to allow for any combination of final subgroup sizes. Large matrices result when S is not small (for example, for $S = 6$, $\mathbf{E}_6$ requires a $65\,536 \times 458\,329$ matrix) causing memory problems from using the matrix methods in R's base package (R Development Core Team, 2010). However, this is not too limiting because S is usually small in practice.

Fig. 1.

Mean number of tests when $p = 0.05$ (Δ , degenerate; $\times$, extreme; $\square$, uniform; $*$, beta): first row, three halving steps; second row, four halving steps; third row, five halving steps

Table 1

Mean number of tests for specific risk distributions and halving steps where $S_p = S_e = 1^{\dagger}$

p	Steps	Group size	Mean number of tests			Maximum differences	Degenerate standard deviation
			Degenerate	Uniform	Beta		
0.10	2	4	5938	5938	5938	—	95.0
	3	6	5939	5881	5828	360.9	115.5
	4	8	6293	6210	6130	674.4	133.3
	5	16	6687	6493	6305	1681.6	148.0
0.05	2	5	4262	4262	4262	—	93.6
	3	8	3946	3916	3887	172.0	109.9
	4	10	3953	3707	3733	246.4	119.5
	5	20	4095	3781	3740	691.1	137.7
0.01	2	10	1956	1956	1956	—	93.0
	3	16	1583	1577	1570	29.8	93.0
	4	20	1363	1359	1352	44.0	83.5
	5	32	1257	1250	1242	85.3	90.2
0.005	2	16	1396	1396	1396	—	106.7
	3	20	1084	1082	1079	12.0	81.2
	4	32	895	892	888	26.3	80.5
	5	48	785	782	778	42.5	79.1

[†]The group size chosen is the optimal size for the degenerate distribution case. The 'Maximum difference' column gives the greatest reduction in tests from using ordered halving for a particular group size and number of steps. Note that two-step halving is Dorfman's procedure.

Table 2

Mean number of tests for chlamydia screening[†]

Specimen-gender combination	Group size	Dorfman	Results for 3-step halving			Results for 4-step halving			Results for 5-step halving		
			Unordered	Ordered	Decrease (%)	Unordered	Ordered	Decrease (%)	Unordered	Ordered	Decrease (%)
Urine-female (prevalence = 0.102)	8	1558	1239	1236	0.21	1229	1208	1.68	†	†	†
	12	1735	1298	1229	5.27	1136	1105	2.76	†	†	†
	16	1926	1406	1352	3.81	1146	1103	3.83	1132	1101	2.77
	24	2115	1540	1503	2.43	1161	1115	3.95	1018	1020	-0.13
	32	2187	1632	1541	5.56	1212	1085	10.44	1007	961	4.55
Urine-male (prevalence = 0.083)	8	2416	2009	1972	1.85	1996	1965	1.56	†	†	†
	12	2702	2109	2077	1.54	1908	1884	1.25	†	†	†
	16	3037	2338	2253	3.64	2000	1931	3.47	1993	1947	2.31
	24	3277	2598	2519	3.05	2093	2022	3.37	1935	1858	4.01
	32	3507	2882	2762	4.15	2234	2236	-0.11	1931	1910	1.08
Swab-female (prevalence = 0.058)	8	9493	7833	7706	1.62	7804	7731	0.94	†	†	†
	12	10792	8223	8007	2.62	7443	7243	2.70	†	†	†
	16	12341	9035	8760	3.05	7569	7405	2.17	7534	7388	1.94
	24	14482	10449	9957	4.70	8116	7745	4.56	7368	7107	3.55
	32	15691	11712	11124	5.02	8772	8173	6.82	7379	7103	3.73
Swab-male (prevalence = 0.130)	8	2985	2634	2534	3.77	2722	2639	3.03	†	†	†
	12	3358	2840	2680	5.62	2666	2546	4.49	†	†	†
	16	3568	2996	2816	6.03	2638	2495	5.42	2702	2579	4.56
	24	3819	3325	2999	9.82	2832	2607	7.94	2616	2472	5.52
	32	3803	3428	3165	7.66	2940	2631	10.49	2606	2404	7.76

[†] For the urine-female combination, there are 2679 individuals with $S_e = 0.805$ and $S_p = 0.96$. For the urine-male combination, there are 3852 individuals with $S_e = 0.93$ and $S_p = 0.95$. For the swab-female combination, there are 19451 individuals with $S_e = 0.928$ and $S_p = 0.96$. For the swab-male combination, there are 4081 individuals with $S_e = 0.925$ and $S_p = 0.95$.

[‡] Not applicable.

Table 3

Accuracy measures for chlamydia screening by using $I = 16^{\dagger}$

Process	Results for urine-female combination					Results for swab-female combination				
	Tests	PS _e	PS _p	PPPV	PNPV	Tests	PS _e	PS _p	PPPV	PNPV
Dorfman	1926	0.655	0.977	0.760	0.962	12341	0.858	0.978	0.705	0.991
3-step unordered	1406	0.537	0.986	0.815	0.950	9035	0.798	0.988	0.802	0.988
3-step ordered	1352	0.535	0.987	0.828	0.949	8760	0.799	0.989	0.812	0.988
4-step unordered	1146	0.432	0.994	0.899	0.939	7569	0.739	0.995	0.893	0.984
4-step ordered	1103	0.420	0.995	0.903	0.938	7405	0.745	0.995	0.896	0.985
5-step unordered	1132	0.354	0.998	0.950	0.932	7534	0.675	0.998	0.960	0.980
5-step ordered	1101	0.353	0.998	0.950	0.932	7388	0.687	0.998	0.962	0.981
<i>Results for urine-male combination</i>										
Dorfman	3037	0.864	0.963	0.682	0.987	3568	0.857	0.960	0.762	0.978
3-step unordered	2338	0.802	0.979	0.782	0.982	2996	0.788	0.974	0.820	0.968
3-step ordered	2253	0.807	0.979	0.779	0.982	2816	0.798	0.977	0.838	0.970
4-step unordered	2000	0.760	0.990	0.879	0.979	2638	0.724	0.987	0.893	0.960
4-step ordered	1931	0.747	0.991	0.879	0.977	2495	0.729	0.988	0.898	0.961
5-step unordered	1993	0.691	0.997	0.950	0.973	2702	0.668	0.995	0.953	0.952
5-step ordered	1947	0.695	0.997	0.948	0.973	2579	0.672	0.996	0.957	0.953

[†] PS_e, pooling sensitivity; PS_p, pooling specificity; PPPV, pooling positive predictive value; PNPV, pooling negative predictive value.