University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Dissertations and Theses in Statistics | Statistics, Department of

Summer 6-17-2013

Informative Retesting for Hierarchical Group Testing

Michael S. Black
University of Nebraska-Lincoln, my2k@huskers.unl.edu

Follow this and additional works at: https://digitalcommons.unl.edu/statisticsdiss

 Part of the Applied Statistics Commons

Black, Michael S., "Informative Retesting for Hierarchical Group Testing" (2013). *Dissertations and Theses in Statistics*. 10.
https://digitalcommons.unl.edu/statisticsdiss/10

This Article is brought to you for free and open access by the Statistics, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Dissertations and Theses in Statistics by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

INFORMATIVE RETESTING FOR HIERARCHICAL GROUP TESTING

by

Michael Black

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Statistics

Under the Supervision of Professor Christopher R. Bilder

Lincoln, Nebraska

June, 2013

INFORMATIVE RETESTING FOR HIERARCHICAL GROUP TESTING

Michael Black, Ph.D.
University of Nebraska, 2013

Advisor: Christopher R. Bilder

Group testing is the process of pooling samples (e.g., blood, chemical compounds) from multiple sources and testing the pooled material for some binary characteristic. It is used in pathogen screening for humans and animals, drug discovery studies, electrical systems testing, and many other applications. Group testing has traditionally been used for two main types of investigations: 1)the identification of positive specimens and 2) the estimation of a characteristic's prevalence in a population. This dissertation focuses on the identification process. We propose new identification procedures that exploit the heterogeneity among samples in order to reduce the number of tests needed to detect the binary characteristic. We first propose the "ordered halving" procedure which is shown to reduce the expected number of tests in comparison to current implementations of halving. Next, we generalize our proposals to a class of hierarchical group testing procedures. Our proposals result in significant reductions in the expected number of tests while also maintaining accuracy at levels similar to those procedures which do not account for heterogeneity.

ACKNOWLEDGEMENTS

Much appreciation for my advisor Dr. Christopher R. Bilder, who always showed a desire to make my papers perfect. I would also like to show my thanks for my committee members, Dr. Kent Eskridge, Dr. Shunpu Zhang, Dr. Joshua Tebbs, and Dr. Clayton Kelling. Also to my wife without whom I would never have finished, thank you.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Foundations of group testing

In the four years after the bombing of Pearl Harbor, the US military put almost 12 million men in uniform (Flynn 1993, page 53). As part of the build-up process, all new men needed to be screened to determine their fitness for combat. This screening process involved testing for various diseases, including syphilis. As detailed in Du and Hwang (2000), a group of government economists in 1942 or 1943 were disappointed at the wastefulness that came from individually testing every man for syphilis. These economists subsequently developed the idea of group testing to reduce the time and costs associated from testing. Among them, Robert Dorfman subsequently published this idea.

Dorfman (1943) proposed that blood specimens from a set of individuals could be combined into a single group specimen. If the group tested negative (syphilis was not present), all individuals within the group would be declared negative. If the group tested positive (syphilis was present), all individuals would be retested individually in order to decode the positives from the negatives. Because syphilis had a low prevalence, it was hoped that the number of tests resulting from group testing would be less than from testing each person individually.

Group testing ultimately was not used for the military induction screening (Du and Hwang, 2000). The group testing idea may have disappeared if not for a problem described in Feller (1950, p. 189; 1957, p. 225) which started others thinking about how group testing may be done more efficiently. We next give details of Dorfman's original proposal and of others' subsequent proposals.

1.1.1 Dorfman (1943)

Consider a group of I individuals where specimens are pooled. If the probability each individual is positive is denoted by p and we assume that individuals screened represent a random selection, the probability of no positive individuals in the group is $(1 - p)^I$. The probability that a group tests positive is then $1 - (1 - p)^I$. This leads to an expected number of tests of $1 + I\{1 - (1 - p)^I\}$ for one group, and an expected total number of tests across all groups formed of size I to be $N/I + N\{1 - (1 - p)^I\}$, where N is the number of individuals screened.

Using these formulas, Dorfman (1943) showed a significant cost savings from using his proposal rather than individual testing. Dorfman presented tables to provide the most efficient group size given a value for p and concluded that group testing is only effective if the characteristic of interest has a small prevalence. Watson (1961) later showed that the optimal group size is approximately $p^{-1/2}$.

1.1.2 Sterrett (1957)

Sterrett (1957) proposed a retesting algorithm for screening blood specimens that did not immediately test each individual in a positive group, but instead started testing individual specimens one at a time (chosen at random) until the first positive individual was found. Once this positive individual was found, the remaining individuals would be retested in a new group. If this new group tested negative, these remaining individuals would be declared negative. However, if this new group tested positive, the same one-by-one testing process would begin again with these remaining individuals to decode the positive individuals from the negative individuals. For situations with a reasonably chosen initial group size

and small p, this retesting procedure works well because there is most likely only one positive individual in a group.

1.1.3 Group splitting procedures

Sobel and Groll (1959) presented several industrial situations where group testing is useful as well, including the testing for a faulty device and the testing of electrical devices in series (e.g., Christmas tree lights). They also proposed a retesting algorithm that successively splits positive groups into two subgroups. If one of these subgroups tested positive, it can be split again and so on until only single units are left to be tested. Their algorithm used the knowledge of the prevalence to optimize the size of each retesting subgroup. If the prevalence was unknown, they suggested that the groups be split in half (two equally sized subgroups). This is what we refer to as halving.

Subsequent papers by others further expanded upon the group splitting proposals. These papers include: Sobel and Groll (1966) expanded on the idea of halving from their 1959 paper, Hwang (1972) examined a binary splitting algorithm of which halving could be considered a subset, and Litvak et al. (1994) considered halving in the presence of testing error with the possibility of confirmation testing considered.

1.1.4 Non-hierarchical procedures

Li (1962) categorized group testing procedures by the number of possible stages (i.e., steps, cycles) and by hierarchical (adaptive) or non-hierarchical (non-adaptive) approaches. Any procedure that uses information from a previous stage to determine the testing pattern on a subsequent stage is considered hierarchical. Dorfman's procedure is an example of a simple hierarchical, two-stage procedure because the results from the initial testing of groups are needed to decide if

individual testing is needed. Sterrett is an example of a hierarchical procedure with a non-fixed number of stages, because the number of stages is unknown at the start.

Non-hierarchical examples include standard individual testing and array (matrix) testing. Individual testing is a one-stage procedure because it takes just one step of testing to determine who is positive and who is negative. Array testing (Phataford and Sudbury, 1994) involves organizing individual specimens into a grid and pooling individuals for testing by both rows and columns. This leads to each individual being within two groups at the beginning. If an individual is in a row and column that both test positive, it could be classified as positive. Thus, there is only one stage of testing. However, when more than one row and more than one column test positive, ambiguities exist concerning which individual is positive or negative, so additional individual testing may be performed in these cases. Subsequent work in array testing after Phataford and Sudbury (1994) include: Berger et al. (2000) who proposed array testing in more than two directions without testing error, Kim et al. (2007) who incorporated testing error and a master array pool, and Kim and Hudgens (2009) who developed a three-dimensional array procedure with testing error.

1.1.5 Other approaches

Dorfman (1943) stated his proposal in probabilistic terms, "we have a population expressing a trait with probability (or prevalence rate) of p," and the efficacy of his procedure is determined by the expected number of tests needed to decode positive and negative individuals. Li (1962) and Katona (1973) looked at group testing in combinatorial terms by trying to optimally find a subset of d positives from an initial set of size n. For a detailed exploration of combinatorial group

testing, see Du and Hwang's (2000) textbook on the subject. We focus on probabilistic group testing, because we consider situations where individual probabilities of positivity may be different from an overall prevalence rate.

1.2 Testing error

While Dorfman (1943) did not consider testing error, he did mention the possibility of it in his paper. Most disease testing situations need to account for testing error because assays are usually not 100% correct. Watson (1961) is possibly the earliest paper to consider testing error in a group testing context. Watson looked at experimental error arising where significant factors in an experimental design were identified. Also, Graff and Roeloffs (1972) looked at the effect testing error could have on group testing in terms of the number of tests and associated costs. Incorporating testing error into an analysis is now standard for disease testing situations.

Testing errors occur when positive items are incorrectly identified as negative or vice versa. Altman and Bland (1994a, b) define the sensitivity (S_e) as the proportion of true positives identified by a test and the specificity (S_p) as the proportion of true negatives correctly identified. Typically, for disease diagnostic tests, the S_e and S_p values are given by an assay manufacturer or validated by a laboratory using the assay. In these settings, S_e and S_p are actually statistics summarizing a selected set of specimens with known positive and negative values. However, currently in most applications, S_e and S_p are treated as constants, and we will do the same within this dissertation.

Altman and Bland (1994a, b) also contend that it is desirable to know how well a diagnostic test does at correctly diagnosing subjects. Corresponding measures are the positive predictive value (PPV), which is the proportion of

positive test results that correctly indicate a true positive status, and the negative predictive value (NPV), which is the proportion of negative test results that correctly indicate a true negative status. PPV and NPV can then be calculated as $PPV = S_e p/[S_e p + (1 - S_p)(1 - p)]$ and $NPV = S_p(1 - p)/[(1 - S_e)p + S_p(1 - p)]$ by using Bayes' rule.

In the context of group testing, an assay may be applied multiple times to the same specimen (either in a group or individually) to determine its positive or negative status. The S_e and S_p values are often treated as the same for each testing stage (e.g., for Dorfman's procedure, S_e is the same for the initial group test and subsequent individual tests), and we will do the same within this dissertation. However, because a specimen may be tested multiple times, its probability of being correctly diagnosed through the group testing process is no longer S_e and S_p. Instead, Johnson et al. (1991) define the pooling sensitivity (PS_e) and pooling specificity (PS_p) to measure the correctness of the group testing procedure's classification of individuals. The pooling sensitivity is the probability that an individual is categorized as positive by a particular group testing procedure given that individual is truly positive. The pooling specificity is defined in a similar manner for the negative individuals. Kim et al. (2007) use PS_e and PS_p then to define the pooling positive predictive value ($PPPV$) and pooling negative predictive value ($PNPV$) for a given group testing procedure as

$$PPPV = \frac{pPS_e}{pPS_e + (1-p)(1-PS_p)}$$

and

$$PNPV = \frac{(1-p)PS_p}{(1-p)PS_p + p(1-PS_e)},$$

again through an application of Bayes' rule.

The assumption of identical S_e and S_p values at multiple stages of group testing is often used due to the work of papers such as Kline et al. (1989), Tu et al. (1995), and Soroka et al. (2003). These papers showed that for group sizes up to 15 or 20 there was negligible loss of S_e and S_p when screening for HIV with enzyme-linked immunosorbent assays (ELISA). Additionally, for nucleic acid tests (NATs), Kacena et al. (1998a, b) showed negligible loss for S_e and S_p as group sizes increase when screening for chlamydia and gonorrhea. Of course, the assumption of identical S_e and S_p values at multiple stages may not hold true if different diagnostic tests are used at multiple stages. For example, ELISA tests may be used to test the initial groups due to their lower costs and NATs may be used to decode positive groups due to their often higher sensitivity values. Also, the assumption may not hold true if a positive specimen is diluted enough by other specimens within a pool so that the group no longer tests positive. Ideally, before implementation of a group testing procedure, an assay should be calibrated to make sure this does not happen. If it does happen, there are a few group testing procedures that can account for a dilution effect. In particular, McMahan et al. (2013) showed that taking into account possible dilution effects can improve prevalence estimation.

1.3 Individual probabilities of positivity

In many cases where group testing is used, covariate information is available on the individual items being screened. This provides extra information that can be used to estimate covariate-specific probabilities of positivity for each individual. There are two ways these estimates can be obtained in a group testing context. First, a binary regression model can be estimated using a training data set of

known individual test results. This model can then be applied to the individuals being screened through group testing in order to obtain estimates of individual probabilities. Second, when a training data set is not available, the individual probabilities can be estimated through the methods of Vansteelandt et al. (2000) and Xie (2001). These papers discuss how to use initial group responses alone to estimate individual probabilities. For the purposes of this dissertation, we solely implement the training data approach.

These individual probabilities of positivity can be exploited in order to reduce the number of tests needed for group testing. Hwang (1975) is likely the first group testing paper that discusses how to take advantage of probabilities of positivity for each individual, and he does so in the context of a Dorfman-like procedure. However, this paper did not discuss how these probabilities can be estimated or how to incorporate testing error.

Surprisingly, there was little further group testing research on using individual probabilities until recently. Bilder et al. (2010) propose "informative retesting" as a way to take advantage of estimated individual probabilities when testing error is present. They used individual probability information to improve Sterrett's procedure by retesting individuals with the highest probabilities first within positive groups. McMahan et al. (2012a, b) further explored informative retesting procedures that looked at how Dorfman's procedure and array testing could take advantage of estimated individual probabilities. Closed form expressions for PS_e, PS_p, $PPPV$, and $PNPV$ were also derived by McMahan et al. (2012a, b) to assess the procedures. They showed that their procedures resulted in similar and sometimes even better accuracy than corresponding non-informative retesting procedures. Bilder and Tebbs (2012) present an overall comparison of all

informative retesting procedures, including those in Chapter 2 of this dissertation. They conclude that no single procedure is the overall best in terms of the expected number of tests and accuracy. Prevalence, assay accuracy, availability of covariate information, the underlying distribution of individual probabilities, and application ease can all affect which procedure is best for a given situation.

1.4 Integer programming

An integer programming (IP) problem is a type of linear or non-linear programming (LP) problem where all or part of the solution require integer values. IP often uses LP methods to find solutions, but integer programs are often more difficult to solve. IP begins by defining an objective function to be solved (e.g., minimize a mathematical function) subject to a set of constraints. A simple method to find the solution is the simplex algorithm (Dantzig 1963). It begins by finding a feasible starting point for the function and then moves from one feasible point to another that successively improves the objective function (e.g., find a smaller minimum for the objective function). This process continues until the optimal solution is found. The direction of movement at each step is the direction that improves the function the most, which is often referred to as the path of steepest descent (ascent) for minimization (maximization) problems.

The method of steepest descent (or ascent) is a well-known optimization strategy for convex functions. The direction of steepest descent can be found using partial derivatives when known. Alternatively, for IP problems, the direction is found by calculating the objective function value one step in each feasible direction to find the direction with the greatest improvement. Once the

direction of steepest descent is found, it can be followed until improvement in that direction stops. The process is continued until we reach the optimal solution.

In the context of this dissertation, the expected number of tests can be treated as an objective function for an IP. The method of steepest descent then can be used to find how best to split a group into subgroups (the number of subgroups and their sizes are integers) to minimize the expected number of tests.

1.5 Organization of the dissertation

The remainder of this dissertation is organized as follows. Chapter 2 is the first submission of a paper that was published in the *Journal of the Royal Statistical Society* (*Series C*) (a revision of the paper completed by my advisor was published). This paper shows how to use estimated individual probabilities to improve the halving group testing procedure. In Section 2.2, we derive formula for the mean, variance, and probability mass function (PMF) for the number of tests under halving where individuals may have different probabilities of positivity. We prove that the mean and variance for individuals from a heterogeneous population (individuals have different probabilities of positivity) assigned to subgroups at random are the same as assuming homogeneity in the population. In Section 2.3, we use the derivations from Section 2.2 to develop a new group testing procedure that we name "ordered halving". We prove that ordered halving results in a smaller number of tests than for other implementations of halving. In Section 2.4, we examine this reduction in tests in the context of a beta distribution assumption for the individual probabilities. In Section 2.5, we apply our new procedure to chlamydia and gonorrhea screening in Nebraska. Finally, Section 2.6 summarizes the improvements and discusses how

other retesting procedures may benefit by taking into account heterogeneity among individual probabilities.

We plan to submit Chapter 3 to the *Journal of the Royal Statistical Society* (*Series C*) as well. Section 3.2 generalizes the $E(T)$ formula from Section 2.2 for any fixed-stage hierarchical group testing procedure. We minimize $E(T)$ by finding the optimal configuration through looking at every possible configuration (subgroup sizes, members of each group) and using IP methods to find a possible solution. In Section 3.3, we examine the performance of our proposals in the context of beta distributions. In Section 3.4, we apply our procedures to the chlamydia and gonorrhea screening data as described for Section 2.5. Finally, in Section 3.5, we summarize our work and discuss future extensions.

Chapter 4 contains additional items examined during this research that were not included in the papers of Chapter 2 and Chapter 3. In Section 4.1, we look at conditions when halving will not be optimal and suggest modifications for improvement. In Section 4.2, we derive the PMF for the number of tests in a general hierarchical group testing procedure. In Section 4.3, we discuss in more detail the IP solution used to find possible optimal retesting configurations for more than three stages. In Section 4.4, we provide documentation on how to run new R functions that implement the research in this dissertation. In Section 4.5, we layout possible future research directions.

Because Chapters 2 and 3 are actual papers, their journal submission formats are preserved. For example, references are included within the chapters.

Chapter 2: Group testing in heterogeneous populations using halving algorithms

Abstract

Group (pooled) testing is often used to reduce the total number of tests needed to screen a large number of individuals for an infectious disease or some other binary characteristic. Traditionally, research in group testing has assumed each individual is independent with the same risk of positivity. More recently, there is 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. In this paper, we investigate the impact of acknowledging population heterogeneity on a commonly used group testing procedure known as "halving." For this procedure, positive groups are successively split into two equal sized halves until all groups test negative or until individual testing occurs. We show that heterogeneity does not affect the mean number of tests when individuals are randomly assigned to sub-groups. However, when individuals are assigned to sub-groups based on their risk probabilities, we show that our proposed procedures reduce the number of tests by taking advantage of the heterogeneity. This is illustrated using chlamydia and gonorrhea screening data from the state of Nebraska.

Key words: binary response; classification; identification; pooled testing; retesting; screening.

2.1 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 needed. Group testing, 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 pool (group) tests negative, all individuals within it are declared negative. If the pool tests positive, 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 to simply 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 has been well established in many areas, including blood donation screening (Dodd et al. 2002), opportunistic testing of individuals for chlamydia (Mund et al. 2008), Bovine Viral Diarrhea virus detection in cattle herds (Peck 2006), estimation of virus infection levels for carnations (Hepworth, 2009), West Nile virus detection among mosquitoes (Biggerstaff 2008), estimating transmission rates of pathogens from insects to plants (Tebbs and Bilder 2004), 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 is 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 modeled using the group testing regression methods of Vansteelandt et al. (2000) or Xie (2001). Bilder et al. (2010) show further how estimates of these individual probabilities can be used to retest individuals in a positive group, and they demonstrate how one can reduce significantly the number of tests needed through an extension of Sterrett's (1957) identification procedure.

Given these advances, it is now important to determine if accounting for population heterogeneity provides benefits with other retesting procedures used in practice. One widely used procedure involves successively splitting positive groups into smaller sub-groups 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 sub-groups 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 HIV. In our paper, we generalize the use of halving to heterogeneous population settings.

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

The paper is ordered as follows. In Section 2.2, we derive the mean, variance, and probability mass function for the number of tests under halving. When compared to a homogeneous population setting, we prove that the mean and variance do not change if individuals from a heterogeneous population are assigned to sub-groups at random. Using the derivations from Section 2.2, we propose a new halving procedure in Section 2.3 that exploits risk heterogeneity to reduce the expected number of tests, and we identify in Section 2.4 the situations where our new procedure performs best. In Section 2.5, we apply our new procedure to chlamydia and gonorrhea screening in Nebraska. Finally, Section 2.6 summarizes and discusses how other retesting procedures may benefit by taking risk heterogeneity into account.

2.2 Halving

2.2.1 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 2.3, we focus on one particular initial group of size I where individual i has risk probability p_i for $i = 1, \ldots, I$. 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 negative or until individual testing occurs. For example, 3-step halving for an initial group of size $I = 16$ begins by testing the entire group. If the group tests positive, the second step involves splitting it into two sub-groups of size 8. If either sub-group tests positive, a third and final step occurs where each individual in a positive sub-group is tested. A 4-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 $G_{s,j} = 1$ (0) denote a positive (negative) test result for the j^{th} ordered sub-group at the s^{th} step for $j = 1, \ldots, 2^{s-1}$ and $s = 1, \ldots, S$. 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 sub-group of size 8 halved from a positive initial group. In a 3-step setting, we can write the expected number of tests for an initial group of size I as

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

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

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

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

$$\begin{aligned} E(T \mid \boldsymbol{p}^{vec}) = 1 + 2\sum_{s=1}^{S-2}\sum_{j=1}^{2^{s-1}} P\left(\bigcap_{\{(s',j'):G_{s,j}=1\}} \{G_{s',j'} = 1\}\right) + \\ \sum_{j=1}^{2^{S-2}} I_{S-1,j} P\left(\bigcap_{\{(s',j'):G_{S-1,j}=1\}} \{G_{s',j'} = 1\}\right) \end{aligned} \tag{2.1}$$

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 4-step halving procedure with $I = 7$ can have an initial split with sub-

groups 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 $\tilde{G}_{s,j} = 1\,(0)$ denote a positive (negative) true status for the j^{th} ordered sub-group at the s^{th} step, and define the test sensitivity and specificity as $S_e = P(G_{s,j} = 1 \mid \tilde{G}_{s,j} = 1)$ and $S_p = P(G_{s,j} = 0 \mid \tilde{G}_{s,j} = 0)$, respectively. The probability that the initial group tests positive can be written as

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

where we make the standard assumption that the test outcomes are conditionally independent given the true statuses (see Litvak et al. 1994).

Probabilities involving groups for steps two 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 two, after the initial group tests positive, is

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

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

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

where $i\in B_{s,j}$ is understood to mean those individuals who belong to the j^{th} ordered group at the s^{th} step. These results can be generalized for $s>1$ to

$$P\left(\bigcap_{\{(s',j'):G_{s,j}=1\}}\{G_{s',j'}=1\}\right)=(1-S_p)^s\left\{\prod_{i=1}^{I}(1-p_i)\right\}+$$
$$\sum_{a=1}^{s-1}S_e^a(1-S_p)^{s-a}\left\{\prod_{i\in B_{a+1,\ell}}(1-p_i)\right\}\left\{1-\prod_{i\in \overline{B}_{a+1,\ell}}(1-p_i)\right\}+$$
$$S_e^s\left\{1-\prod_{i\in B_{s,j}}(1-p_i)\right\}, \qquad (2.2)$$

where $\ell=\left\lceil j/2^{s-1-a}\right\rceil$ and $i\in\overline{B}_{s,j}$ denotes the set of individuals within the parent group of $B_{s,j}$ excluding those in $B_{s,j}$ itself (e.g., $i\in\overline{B}_{3,3}$ 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 (2.2) into (2.1), 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 3-step procedure,

$$E(T^2\mid \boldsymbol{p}^{vec})=P(G_{1,1}=0)+3^2P(G_{1,1}=0\cap G_{2,1}=0\cap G_{2,2}=0)+$$
$$(3+I_{2,1})^2P(G_{1,1}=0\cap G_{2,1}=1\cap G_{2,2}=0)+$$
$$(3+I_{2,2})^2P(G_{1,1}=0\cap G_{2,1}=1\cap G_{2,2}=0)+$$
$$(3+I)^2P(G_{1,1}=0\cap G_{2,1}=1\cap G_{2,2}=1).$$

The four probability terms in this expression are found using Equation (2.2). For 4-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 probability mass function (PMF) of T by exploiting

the hierarchical nature inherent to the halving procedure. Combining the PMF with the standard variance formula leads to the desired result.

2.2.2 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 $i = 1, \ldots, I$. The overall expected number of tests is

$$\begin{aligned} E(T) &= E\left\{E(T \mid \boldsymbol{p}^{vec})\right\} \\ &= 1 + 2\sum_{s=1}^{S-2}\sum_{j=1}^{2^{s-1}} E\left\{P\left(\bigcap_{\{(s',j'):G_{s,j}=1\}} \{G_{s',j'} = 1\}\right)\right\} + \\ &\quad \sum_{j=1}^{2^{S-2}} I_{S-1,j} E\left\{P\left(\bigcap_{\{(s',j'):G_{S-1,j}=1\}} \{G_{s',j'} = 1\}\right)\right\}. \end{aligned} \tag{2.3}$$

The expectation of the joint probability in (2.1) is

$$\begin{aligned} E\left\{P\left(\bigcap_{\{(s',j'):G_{s,j}=1\}} \{G_{s',j'} = 1\}\right)\right\} &= (1-S_p)^s E\left\{\prod_{i=1}^{I}(1-p_i)\right\} + \\ &\quad \sum_{a=1}^{s-1} S_e^a (1-S_p)^{s-a} E\left\{\prod_{i\in B_{a+1,\ell}}(1-p_i)\right\}\left[1 - E\left\{\prod_{i\in B_{a+1,\ell}}(1-p_i)\right\}\right] + \\ &\quad S_e^s\left[1 - E\left\{\prod_{i\in B_{s,j}}(1-p_i)\right\}\right]. \end{aligned} \tag{2.4}$$

Because of independence among the individual probabilities, Equation (2.4) simplifies to

$$\begin{aligned} &E\left\{P\left(\bigcap_{\{(s',j'):G_{s,j}=1\}} \{G_{s',j'} = 1\}\right)\right\} \\ &= (1-S_p)^s(1-p)^{u_{1,1}} + \sum_{a=1}^{s-1} S_e^a (1-S_p)^{s-a}(1-p)^{u_{a+1,l}-l_{a+1,l}}\left\{1-(1-p)^{\overline{u}_{a+1,l}-\overline{l}_{a+1,l}}\right\} + \\ &\quad S_e^s\left\{1-(1-p)^{u_{s,j}-l_{s,j}}\right\}, \end{aligned} \tag{2.5}$$

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 the sub-group $B_{s,j}$ and $\overline{l}_{s,j}$ and $\overline{u}_{s,j}$ are the lowest and highest subscripts, respectively, for the individuals in the sub-

group $\overline{B}_{s,j}$. The expected number of tests $E(T)$ is found by substituting (2.5) into (2.3).

It is especially insightful to note that $P(\bigcap_{\{(s',j'):G_{s,j}=1\}}\{G_{s',j'}=1\})$ reduces to (2.5) when all individuals have a common risk probability p; this implies the unconditional means are the same for homogeneous or heterogeneous population assumptions. Furthermore, we show in Appendix B that $Var(T)$ also remains unchanged. Therefore, when individuals with different risks are assigned randomly to groups, neither $E(T)$ nor $Var(T)$ is affected. This is reassuring if the researcher is unable to 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) give the probability mass function for T in a "3-stage" informative Sterrett procedure. If one treats the individual risk probabilities as independent random variables, all of their $P(T=t)$ expressions rely on these simple products.

2.3 Ordered halving

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

After an initial group of size I tests positive, two sub-groups of equal size are created. Our goal is to maximize one sub-group's probability of testing positive and maximize the other sub-group's probability of testing negative. We show this type of sub-group construction allows for faster identification of positive individuals on average while allowing negative individuals to be classified sooner. Define a set of ordered risk probabilities for an initial group of size I as $\boldsymbol{p}^{ord} = (p_{(1)},\ldots,p_{(I)})'$ where $p_{(i)}$ denotes the i^t smallest probability within the group. The second step of "ordered halving" creates one sub-group of individuals with lower risks $p_{(1)},\ldots,p_{(I_{2,1})}$ and one sub-group of individuals with higher risks $p_{(I_{2,1}+1)},\ldots,p_{(I)}$. If one of these sub-groups tests positive and $S \geq 4$ (i.e., individual testing does not occur at step 3 for positive sub-groups), 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 sub-group sizes are equal, i.e., compare $E(T \mid \boldsymbol{p}^{ord})$ and $E(T \mid \boldsymbol{p}^{vec})$, we need only focus on $\sum_{j=1}^{2^{s-1}} P\left(\bigcap_{\{(s',j'):G_{s,j}=1\}} \{G_{s',j'}=1\}\right)$ for each step $s = 1, \ldots, S-1$ of Equation(2.1). This is true because ordering only changes expressions that are functions of the risk probabilities. To help with the comparison, note that

$$\begin{aligned}
&\sum_{j=1}^{2^{s-1}} P\left(\bigcap_{\{(s',j'):G_{s,j}=1\}} \{G_{s',j'}=1\}\right) \\
&= 2^{s-1}S_e^s + \Big\{(1-S_p) - S_e\Big\} \sum_{a=1}^{s} 2^{s-a} S_e^{a-1} (1-S_p)^{s-a} \sum_{j=1}^{2^{a-1}} \prod_{i \in B_{a,j}} (1-p_i). \qquad (2.6)
\end{aligned}$$

When $s = 1$, Equation (2.6) is the same regardless of whether sub-group 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 (2.6) is minimized under ordered assignment as long as $S_e > 1 - S_p$, which will be true for any diagnostic test used in application. This shows that

$E(T \mid \boldsymbol{p}^{ord}) \leq E(T \mid \boldsymbol{p}^{vec})$ whenever our ordered sub-group construction is used. To find $E(T) = E[E(T \mid \boldsymbol{p}^{ord})]$, we make use of Equations (2.1) and (2.4) again where the individual risk probabilities within (2.4) are properly ordered for the sub-groups. Because the expectations in (2.4) are now distribution dependent, a simple expression for $E(T)$ no longer exists. However, we can use the result in Junjiro (1962) to find the distribution of the ordered risk probabilities. This distribution is

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

where $p_{(l_{s,j})} \leq \ldots \leq p_{(u_{s,j})}$ are the ordered risk probabilities for individuals in group $B_{s,j}$ (see Section 2.2.2), $f(p_i)$ 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 (2.3). We examine values of $E(T)$ for specific distributions in Section 2.4.

2.4 Mean Comparisons

Group testing is used 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, and 0.10 and examine the number of tests performed. The distributions chosen for p_i are a beta(1, $1/p$ – 1), a uniform(0, $2p$), and a degenerate at p (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 $1 - p$. While 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 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 using halving with 2, 3, 4 and 5 steps for a number of different 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 10,000 individuals. We use the equations derived in Sections 2.2.2 and 2.3 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 using the probability mass function algorithm described in Appendix A.

For each level of overall risk and number of steps considered, Table 2.1gives 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. While perfect testing does not often occur in actual applications, we assume $S_p = S_e = 1$ because it provides a useful initial examination.

Table 2.1shows the degenerate case always results in the maximum expected number of tests among the four distributions. For 2-step, 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 3-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 can be pronounced 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., 4-step with group size of 10 when $p = 0.05$), because we choose to have the larger risk probabilities in the larger sub-group.

Figure 2.1 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, and 0.10 are available in Appendix C. Figure 2.1 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 among 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.

2.5 Application

The Infertility Prevention Program is a nationally implemented program whose goals are to assess and reduce the prevalence of chlamydia and gonorrhea in the United States. In Nebraska, urine and swab specimens are collected from individuals visiting health clinics throughout the state. These specimens are sent then to the Nebraska Public Health Laboratory (NPHL), where each specimen is

tested individually for both infections. Clinical, demographic, and risk-behavior information is recorded for each individual prior to 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 2004-year results are used as a training data set to estimate the probability of positivity for individuals tested in 2005. First-order logistic regression models are fit to the training data with the response variable as disease status and the explanatory variables of age, race, clinic type, clinic location, reason for visit, symptoms, initial clinical observations, and risk history. These models are fit separately by disease (chlamydia and gonorrhea), gender, and specimen type (swab or urine). The 2005-year individuals are ordered by specimen date and are placed into successive groups by disease/gender/specimen combination. Assuming 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 ten times to account for simulation variability, and we record the average number of tests.

Table 2.2 displays the average number of tests, and Table 2.3gives the percentage reduction in tests for ordered vs. unordered halving at specific group sizes. These tables provide the chlamydia screening results only. Similar results are found for gonorrhea screening, which are given in Appendix D. Overall, we find the chlamydia results to be similar to those found for the beta distribution cases in Section 2.4. This is not surprising because a beta distribution often fits

these individual probabilities well and overall prevalence ranges from 5.8% to 13.0% for each gender/specimen combination. Generally, improvements from ordering are 1% to 6%, where some improvements are larger for swab/male (up to 10.49%). Also, the benefits from ordered halving are more pronounced for larger group sizes and prevalences, which is consistent with our findings in Section 2.4.

We have also investigated how well ordered halving performs in terms of classification accuracy (e.g., pooling sensitivity, pooling positive predictive values). Complete results are in Appendix D. We found no discernible increases or decreases through ordered halving.

2.6 Conclusions

We have generalized the use of halving algorithms in group testing to heterogeneous population settings. Our results demonstrate that ordering risk probabilities reduces the number of tests needed to classify all individuals as positive or negative. The NPHL example shows a specific instance where ordered halving would reduce the testing load. Even when ordered halving provides a small percentage reduction in the number of tests, this can be magnified greatly in situations with very large numbers of individuals. For example, Kim and Hudgens (2009, p. 903) describe a HIV detection program in North Carolina where "slight improvements in efficiency can lead to substantial cost savings" because 120,000 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 reduction in the number of tests through ordering 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 $S_e = S_p = 1$. When there are no positives or only one positive within a group at step 1, ordered halving results in the same number of tests as without ordering. When there are two or more positives 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.

Our results from Sections 2.4 and 2.5 lead us to possible future research areas that can further improve halving. First, we showed that the variation in the risk probabilities was important, but its magnitude of importance changes when uneven sub-group sizes are needed. Future research should examine if there are optimal unequal sub-group 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. We see an informal application of this already at the Nebraska Veterinary Diagnostic Laboratory; however, research is needed to determine actual benefits. Second, group splitting could involve more than two sub-groups. For example, Pilcher et al. (2005) use an initial group size of 90 and subsequent splits into 9 groups of size 10 when the initial group tests positive. It would be of interest to determine how ordering can

further reduce the number of tests needed when multiple sub-groups are used. Choosing the optimal sub-group sizes and the number of sub-groups for a split are open research problems.

2.7 Acknowledgement

This research is supported by Grant R01 AI067373 from the National Institutes of Health.

2.8 References

Bilder, C., Tebbs, J. and Chen, P. (2010) Informative retesting. *Journal of the American Statistical Association*, 105, 942-955.

Biggerstaff, B. (2008) Confidence intervals for the difference of two proportions estimated from pooled samples. *Journal of Agricultural, Biological, and Environmental Statistics*, 13, 478-496.

Dorfman, R. (1943) The detection of defective members of large populations. *Annals of Mathematical Statistics*, 14, 436-440.

Dodd, R., Notari, E., and Stramer, S. (2002) Current prevalence and incidence of infectious disease markers and estimated window-period risk in the American Red Cross donor population. *Transfusion*, 42, 975-979.

Gupta, D. and Malina, R. (1999) Group testing in presence of classification errors. *Statistics in Medicine*, 18, 1049-1068.

Hepworth, G. and Watson, R. (2009) Debiased estimation of proportions in group testing. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58, 105-121.

Hughes-Oliver, J. (2006) Pooling experiments for blood screening and drug discovery. In *Screening: Methods for Experimentation in Industry, Drug Discovery, and Genetics*, (eds A. Dean and S. Lewis), New York: Springer.

Hwang, F. (1975) A generalized binomial group testing problem. *Journal of the American Statistical Association*, 70, 923-926.

Johnson, N., Kotz, S., and Wu, X. (1991) *Inspection Errors for Attributes in Quality Control.* New York: Chapman & Hall.

Junjiro, O. (1962) Distribution and moments of order statistics. In *Contributions to Order Statistics* (ed A. Sarhan), New York: John Wiley and Sons Inc.

Kim, H., Hudgens, M., Dreyfuss, J., Westreich, D., and Pilcher, C. (2007) Comparison of group testing algorithms for case identification in the presence of test error. *Biometrics*, 63, 1152-1163.

Kim, H., and Hudgens, M. (2009) Three dimensional array based group testing algorithms. *Biometrics*, 65, 903-910.

Litvak, E., Tu, X., and Pagano, M. (1994) Screening for the presence of a disease by pooling sera samples. *Journal of the American Statistical Association*, 89, 424-434.

Mund, M., Sander, G., Potthoff, P., Schicht, H., and Matthias, K. (2008) Introduction of *Chlamydia trachomatis* screening for young women in Germany. *Journal der Deutschen Dermatologischen Gesellschaft*, 6, 1032-1037.

Peck, C. (2006) Going after BVD. *Beef*, 42, 34-44.

Phatarfod, R. and Sudbury, A. (1994) The use of a square array scheme in blood testing. *Statistics in Medicine*, 13, 2337-2343.

Pilcher, C., Fiscus, S., Nguyen, T., Foust, E., Wolf, L., Williams, D., Ashby, R., O'Dowd, J., McPherson, J., Stalzer, B., Hightow, L., Miller, W., Eron, J., Cohen, M., and Leone, P. (2005) Detection of acute infections during HIV testing in North Carolina. *New England Journal of Medicine*, 352, 1873-1883.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2013.URL http://www.R-project.org.ISBN 3-900051-07-0.

Remlinger, K., Hughes-Oliver, J., Young, S., and Lam, R. (2006) Statistical design of pools using optimal coverage and minimal collision. *Technometrics*, 48, 133-143.

Stramer, S. L., Glynn, S. A., Kleinman, S. H., Strong, D. M., Caglioti, S., Wright, D. J., Dodd, R. Y., and Busch, M. P. (2004) Detection of HIV-1 and HCV infections among antibody-negative blood donors by nucleic acid–amplification testing. *New England Journal of Medicine* 351, 760-768.

Sterrett, A. (1957) On the detection of defective members of large populations.

Annals of Mathematical Statistics, 28, 1033-1036.

Tebbs, J. and Bilder, C. (2004) Confidence interval procedures for the probability of disease transmission in multiple-vector-transfer designs. *Journal of Agricultural, Biological, and Environmental Statistics*, 9, 75-90.

Vansteelandt, S., Goetghebeur, E., and Verstraeten, T. (2000) Regression models for disease prevalence with diagnostic tests on pools of serum samples. *Biometrics*, 56, 1126-1133.

Xie, M. (2001) Regression analysis of group testing samples. *Statistics in Medicine*, 20, 1957-1969.

			Expected number of tests				Max	Degenerate
p	Steps	Group size	Degenerate	Uniform	Beta	Extreme	difference	standard dev
0.10	2	4	5,938.00	5,938.00	5,938.00	5,938.00	-	95.00
	3	6	5,938.53	5,880.67	5,827.52	5,577.68	360.86	115.53
	4	8	6,293.33	6,210.41	6,129.94	5,618.29	674.44	133.33
	5	16	6,686.70	6,493.23	6,304.65	5,005.09	1,681.61	147.97
0.05	2	5	4,262.19	4,262.19	4,262.19	4,262.19	-	93.55
	3	8	3,946.39	3,916.30	3,886.74	3,774.42	171.96	109.92
	4	10	3,953.15	3,706.77	3,732.70	3,810.91	246.38	119.47
	5	20	4,094.67	3,781.16	3,739.71	3,403.54	691.13	137.72
0.01	2	10	1,956.18	1,956.18	1,956.18	1,956.18	-	93.00
	3	16	1,583.23	1,577.03	1,570.37	1,553.39	29.84	93.01
	4	20	1,363.43	1,358.51	1,352.43	1,319.42	44.01	83.52
	5	32	1,257.24	1,250.15	1,241.89	1,171.96	85.29	90.17
0.005	2	16	1,395.69	1,395.69	1,395.69	1,395.69	-	106.68
	3	20	1,084.29	1,081.71	1,078.98	1,072.34	11.95	81.19
	4	32	894.53	891.60	887.60	868.24	26.29	80.47
	5	48	785.46	782.20	778.08	742.95	42.51	79.13

Table 2.1. Mean number of tests for specific risk distributions and halving steps where $S_p = S_e = 1$. The group size chosen is the optimal size for the degenerate distribution case. Note that two-step halving is Dorfman's procedure.

Specimen/Gender	Group size	Dorfman	3-step Unordered	3-step Ordered	4-step Unordered	4-step Ordered	5-step Unordered	5-step Ordered
	8	1,557.7	1,238.9	1,236.3	1,228.5	1,207.8	NA	NA
	12	1,735.1	1,297.5	1,229.1	1,136.2	1,104.8	NA	NA
Urine/Female	16	1,925.6	1,405.6	1,352.0	1,146.4	1,102.5	1,132.2	1,100.8
	24	2,114.5	1,540.3	1,502.8	1,161.3	1,115.4	1,018.4	1,019.7
	32	2,186.5	1,632.1	1,541.3	1,211.9	1,085.4	1,007.1	961.3
	8	2,415.6	2,009.0	1,971.8	1,996.4	1,965.2	NA	NA
	12	2,701.8	2,109.2	2,076.8	1,908.0	1,884.2	NA	NA
Urine/Male	16	3,037.4	2,338.2	2,253.2	2,000.0	1,930.6	1,992.9	1,946.8
	24	3,277.4	2,598.4	2,519.2	2,092.5	2,022.0	1,935.3	1,857.6
	32	3,506.6	2,882.0	2,762.4	2,233.8	2,236.2	1,930.7	1,909.9
	8	9,492.6	7,833.0	7,705.8	7,804.2	7,731.2	NA	NA
	12	10,791.9	8,222.9	8,007.2	7,443.4	7,242.8	NA	NA
Swab/Female	16	12,341.1	9,035.2	8,759.5	7,569.4	7,404.9	7,533.7	7,387.7
	24	14,481.7	10,448.6	9,957.0	8,115.7	7,745.4	7,368.4	7,107.1
	32	15,691.0	11,711.6	11,124.1	8,771.6	8,173.4	7,378.6	7,103.1
	8	2,984.6	2,633.8	2,534.6	2,721.6	2,639.0	NA	NA
	12	3,357.8	2,840.0	2,680.4	2,666.1	2,546.3	NA	NA
Swab/Male	16	3,568.0	2,996.4	2,815.6	2,638.4	2,495.4	2,702.4	2,579.2
	24	3,819.0	3,325.0	2,998.6	2,832.2	2,607.4	2,616.2	2,471.8
	32	3,802.6	3,427.5	3,165.1	2,939.8	2,631.3	2,605.9	2,403.6

Table 2.2. Average number of tests for chlamydia screening. For urine/female, there are 2,679 individuals, $S_e = 0.805$, and $S_p = 0.96$. For urine/male, there are 3,852 individuals, $S_e = 0.930$, and $S_p = 0.95$. For swab/female, there are 19,451 individuals, $S_e = 0.928$, and $S_p = 0.96$. For swab/male, there are 4,085 individuals, $S_e = 0.925$, and $S_p = 0.95$.

Specimen/Gender	Group size	Percentage decrease 3-step	4-step	5-step
Urine/Female	8	0.21%	1.68%	NA
	12	5.27%	2.76%	NA
	16	3.81%	3.83%	2.77%
	24	2.43%	3.95%	-0.13%
	32	5.56%	10.44%	4.55%
Urine/Male	8	1.85%	1.56%	NA
	12	1.54%	1.25%	NA
	16	3.64%	3.47%	2.31%
	24	3.05%	3.37%	4.01%
	32	4.15%	-0.11%	1.08%
Swab/Female	8	1.62%	0.94%	NA
	12	2.62%	2.70%	NA
	16	3.05%	2.17%	1.94%
	24	4.70%	4.56%	3.55%
	32	5.02%	6.82%	3.73%
Swab/Male	8	3.77%	3.03%	NA
	12	5.62%	4.49%	NA
	16	6.03%	5.42%	4.56%
	24	9.82%	7.94%	5.52%
	32	7.66%	10.49%	7.76%

Table 2.3. Percentage reduction in tests for ordered vs. unordered halving at a specific number of steps and group size.

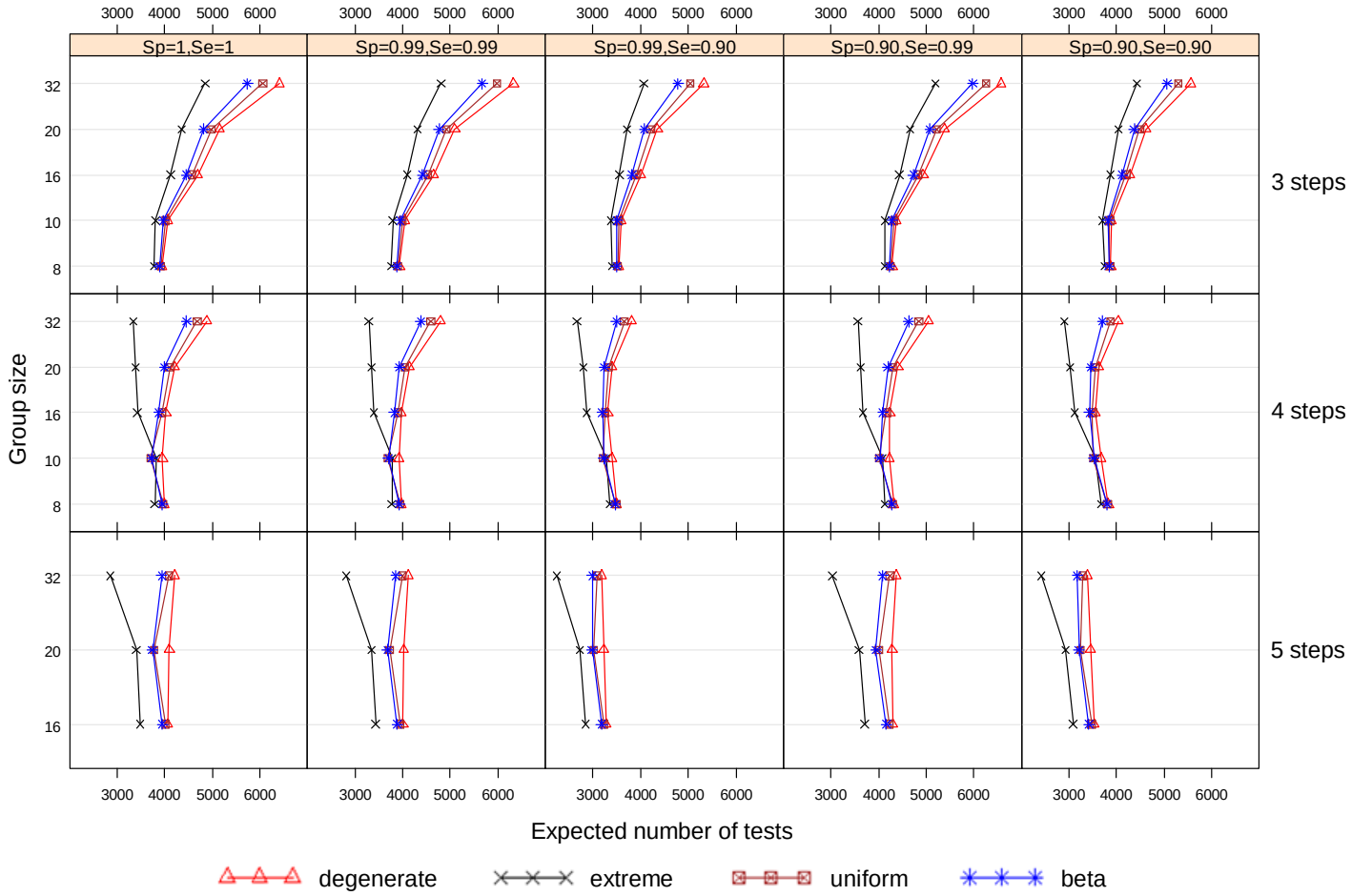

Figure 2.1. Mean number of tests when $p = 0.05$. Each row of panels corresponds to the number of halving steps. Each column of panels corresponds to specificity and sensitivity settings.

Chapter 3: Optimal retesting configurations for hierarchical group testing

Abstract

Hierarchical group testing is a widely implemented procedure used to efficiently screen individuals for infectious diseases and other binary characteristics. This screening protocol works by amalgamating individual specimens into groups for testing. Groups testing positive are successively divided into smaller subgroups and retested to decode positive individuals from negative individuals. In our paper, we propose a general procedure to incorporate risk factor information into the testing process by optimally selecting these subgroup configurations for the individuals. We derive the expected number of tests and classification accuracy measures for our proposals, and we show that our proposals can significantly reduce the number of tests needed and still maintain high classification accuracy. An added benefit is that our proposals can be much more easily applied than most other group testing procedures that take into account risk factor information. We apply our proposals to infectious disease screening which was performed as part of the Infertility Prevention Project in the United States.

Key words: classification; Infertility Prevention Project; informative retesting; pooled testing; retesting; screening

3.1 Introduction

Infectious disease screening frequently occurs through the process known concurrently as group testing and pooled testing. For this screening protocol, clinical specimens (e.g., blood, urine) from separate individuals are amalgamated into a single specimen. Individuals within negative testing groups are declared negative. Individuals within positive testing groups are retested in some predetermined manner to decode the positive individuals from the negative individuals. As long as group sizes are judiciously chosen and the overall disease prevalence is small, group testing significantly reduces the overall number of tests required while subsequently reducing costs. Due to these benefits and the high volume of clinical specimens that often occur for infectious disease screening, group testing is successfully being applied now for chlamydia and gonorrhea testing within the Infertility Prevention Project (Lewis et al. 2012); for HIV, hepatitis B, and hepatitis C screening of blood donations (American Red Cross 2013; Hourfar et al. 2008); and for HIV screening among known HIV-positive individuals to detect antiretroviral treatment failure (Tilghman et al. 2011, Smith et al. 2009; Mehta et al. 2011).

Group testing algorithms are generally divided into two categories: hierarchical and non-hierarchical. Hierarchical group testing involves dividing positive groups into two or more non-overlapping groups for retesting, where a group size may be as small as one. If any of these subsequent groupings test positive, additional stages of dividing occur until individual testing at the final possible stage S, where individuals are decoded as positive or negative. Throughout the process, the results from a previous stage are necessary before further retesting can be performed. In contrast, non-hierarchical group testing involves placing individuals into overlapping groups in the hope that positive testing groups quickly lead to the identification of positive individuals. In particular, two-

dimensional array (matrix) testing places specimens into an array where specimens within rows and columns are pooled. The intersections of positive testing groups correspond to those individuals who are potentially positive; additional individual retests can be completed within these intersections as necessary. For a further review of hierarchical and non-hierarchical group testing, see the excellent chapter given by Hughes-Oliver (2006).

Our paper focuses on hierarchical group testing by taking advantage of recent advances in group testing collectively known as "informative retesting" (Bilder et al. 2010, McMahan et al. 2012a, McMahan et al. 2012b, Black et al. 2012). Informative retesting incorporates individual probabilities of positivity into the decoding process. To obtain these probabilities, binary regression models are estimated using individual disease statuses along with individual risk factor covariates from a training data set. Individual probabilities of positivity are estimated with these models and used for retesting. By taking advantage of the heterogeneity among these probabilities, we propose in our paper new informative retesting procedures that significantly reduce the number of tests needed while also being easier to apply than most previously proposed methods. We achieve our gains in testing efficiency by optimally selecting the number of subgroups and their sizes at each stage. We achieve our ease in application by limiting the number of testing stages and ordering individuals by their probability of positivity.

While our methods can be applied to a large number of infectious disease settings, we focus our discussion on the high volume of clinical specimens evaluated each year for the Infertility Prevention Project (IPP). The IPP was a nationally implemented program in the United States for chlamydia and gonorrhea detection, and hundreds of thousands of specimens are screened each year at laboratories across the country. Due to this high volume, many states (e.g., Idaho, Iowa, New York, Oregon, Virginia, Washington, and

Wisconsin) now use group testing for their screening, and it is likely that many more states will need to in the future given the current fiscal environment. With each specimen tested, risk factors, such as gender, sexual history, and clinician observations, are available. This has prompted at least one state, Idaho, to implement the "threshold optimal Dorfman" informative retesting procedure of McMahan et al. (2012a), which takes advantage of known risk factors to reduce their number of tests (Lewis et al. 2012). It is of interest to determine if further reductions in tests can be obtained by taking advantage of this risk factor information. We will show later that this is the case, which in turn leads to lower screening costs.

An outline of our paper is as follows. In Section 3.2, we derive the expected number of tests and measures of classification accuracy for hierarchical group testing. Using these derivations, we develop new proposals that reduce the number of tests needed to decode positive from negative individuals within positive testing groups. In Section 3.3, we compare our proposals through the use of beta distributions. In Section 3.4, we apply our procedures to chlamydia and gonorrhea screening data from the IPP. Finally, in Section 3.5, we summarize our work and discuss future extensions.

3.2 Hierarchical group testing

3.2.1 Expected number of tests

Consider a group of I individuals that are to be screened for an infectious disease using group testing. Define G_{sj} as a binary random variable denoting the test status for a group (or subgroup) j at the s^{th} stage, where a 0 denotes a negative test result and a 1 denotes a positive test result. For example, G_{11} denotes the initial group's test outcome. The number of individuals screened within the group corresponding to G_{sj} is defined as I_{sj}, where $I_{11} \equiv I$. If $G_{sj} = 0$, all individuals within the corresponding group are declared negative. If $G_{sj} = 1$, individuals within the corresponding group are divided into m_{sj}

subgroups for the next stage of testing. Define c_s as the total possible subgroups tested at stage s, where $c_1 = 1$ and $c_s = \sum_{j=1}^{c_{s-1}} m_{s-1,j}$ for $s = 2, \ldots, S$.

To help explain this notation, consider the use of three-stage hierarchical group testing implemented by Pilcher et al. (2005) to screen blood donations. Specimens from blood donations were placed into groups of size $I = 90$. If $G_{11} = 0$, all individuals are diagnosed as negatives. If $G_{11} = 1$, the initial group is divided into $m_{11} = 9$ subgroups of size $I_{21} = \cdots = I_{29} = 10$ for stage 2 testing. For any of the subgroups at stage 2 that test negative, i.e., $G_{2j} = 0$, these corresponding individuals are declared negative. If $G_{2j} = 1$ for a particular subgroup, it is divided into $m_{2j} = 10$ subgroups of size one. Notice that the total number of possible subgroups tested at stage 3 is $c_3 = 90$ because individual testing would occur. We provide additional examples of more complicated testing protocols in Appendix E.

The expected number of tests for the group of I individuals is

$$E(T) = 1 + \sum_{s=1}^{S-1} \sum_{j=1}^{c_s} m_{sj} P\left(\bigcap_{\{(s'j'):G_{sj}=1\}} \{G_{s'j'} = 1\} \right) \tag{3.1}$$

over S stages. We see that $E(T)$ depends on the number of subgroups, subgroup sizes, and probabilities of groups testing positive. The probability expression within Equation (3.1) is a joint probability representing a succession of groups testing positive up to and including $G_{sj} = 1$. For example, to find $E(T)$ for the Pilcher et al. (2005) application, one probability that would be needed is

$$P\left(\bigcap_{\{(s'j'):G_{21}=1\}} \{G_{s'j'} = 1\} \right) = P(G_{11} = 1 \cap G_{21} = 1)$$

when $s = 2$ and $j = 1$.

To find the general probability expression within Equation (3.1), we need to reexpress it as a function of the true group statuses $\tilde{G}_{sj}$ to account for testing error that

can occur during screening. Define $S_e = P(G_{sj} = 1 \mid \tilde{G}_{sj} = 1)$ and $S_p = P(G_{sj} = 0 \mid \tilde{G}_{sj} = 0)$ as the sensitivity and specificity, respectively, of the assay. The joint probability is then

$$P\left(\bigcap_{\{(s'j'):G_{sj}=1\}} \{G_{s'j'} = 1\}\right) = (1-S_p)^s \left\{\prod_{i=1}^{I_{11}} (1-p_i)\right\} + \sum_{a=1}^{s-1} S_e^a (1-S_p)^{s-a} \left\{\prod_{i \in B_{a+1,j'}} (1-p_i)\right\} \left\{1 - \prod_{i \in \overline{B}_{a+1,j'}} (1-p_i)\right\} + S_e^s \left\{1 - \prod_{i \in B_{sj}} (1-p_i)\right\}, \quad (3.2)$$

where p_i is the probability that individual i is truly positive, $i \in B_{sj}$ is understood to mean those individuals who belong to the j^{th} ordered group at the s^{th} stage, and $i \in \overline{B}_{sj}$ denotes the set of individuals within the parent group of B_{sj} excluding those in B_{sj} itself. Equation (3.2) is written the same way as the expected number of tests formulas given in Black et al. (2012), which examined the special case of hierarchical group testing where positive groups are halved. This equivalence is simply due to the generality of the B_{sj} and $\overline{B}_{sj}$ notation, and the derivation of $E(T)$ is the same. We can now use Equation (3.2) in Equation (3.1) to fully define $E(T)$.

3.2.2 Optimal retesting configurations

For infectious disease screening settings, we want the number of tests to be as small as possible while also minimizing the number of stages. As a result, costs generally will be as low as possible and testing will be completed in a timely manner. Before an application of group testing begins, we will not necessarily know the best retesting configuration (i.e., group sizes, members of each group, ...) for positive testing groups. However, we can examine the expected number of tests among potential configurations before screening in order to choose one that is "optimal"; i.e., choose the procedure that minimizes the expected number of tests.

To find the best retesting configuration, we first order individuals by their probability of positivity within an initial group that tests positive. This helps to isolate

those individuals with small and large probabilities while also lowering the number of possible configurations that need to be examined (to be discussed more shortly). Define $p_{(i)}$ for $i = 1, \ldots, I$ as these ordered individual probabilities, where $p_{(i)} \leq p_{(i+1)}$. Whenever a group tests positive, we assume that individuals are put into groups successively by this ordering. For example, a group of size $I = 6$ could be divided into $m_{11} = 3$ groups of size $I_{21} = 3$, $I_{22} = 2$, and $I_{23} = 1$. This retesting configuration will contain the corresponding individuals with the following probabilities: subgroup #1 includes $p_{(1)}$, $p_{(2)}$, $p_{(3)}$, subgroup #2 includes $p_{(4)}$, $p_{(5)}$, and subgroup #3 includes $p_{(6)}$. Ordering in this manner is intuitive because it allows larger (smaller) subgroups to be formed among the low-probability (high-probability) individuals, which in turn leads to reductions in the number of tests needed for decoding. We define the optimal retesting configuration (ORC) as the configuration which minimizes $E(T)$ when the ordered individuals are successively put into subgroups of this form.

3.2.3 All possible configurations

The most direct approach to find the ORC is to calculate $E(T)$ for all possible configurations. For a three-stage procedure, it is easy to see that the number of configurations is the combination ${}_{I-1}C_1$ when two subgroups are formed at stage 2, ${}_{I-1}C_2$ when three subgroups are formed at stage 2, and so on. In general, there are $\sum_{i=0}^{I-1} {}_{I-1}C_i = 2^{I-1}$ configurations for a three-stage procedure. For example, for an initial group of size $I = 4$, this leads to $2^{4-1} = 8$ possible configurations of subgroups at stage 2 with sizes: [4], [3,1], [2,2], [1,3], [2,1,1], [1,2,1], [1,1,2] or [1,1,1,1], where we use $[\cdot]$ to denote each possible subgroup configuration of particular sizes. If needed, a third stage for positive testing subgroups of size two or more leads to individual testing. Note that this enumeration contains configurations that would not typically be implemented, such as [4] (retest the entire group again), and those that would not result in a stage 3, such

as [1,1,1,1] (further individual testing would not be needed). Similarly, one can show the number of subgroups for a four-stage procedure to be

$$\sum_{c_2=0}^{I-1} {}_{I-1}C_{c_2} \sum_{c_3=0}^{c_2-1} {}_{c_2-1}C_{c_3} = \sum_{c_2=0}^{I-1} {}_{I-1}C_{c_2} 2^{c_2} = 3^{I-1}.$$

For a 5-stage procedure, we have 4^{I-1} possible configurations. This pattern can be continued using binomial expansion, so that S stages have $(S-1)^{I-1}$ possible configurations.

The number of retesting configurations grows exponentially. With respect to calculation time for $E(T)$, consider the case of $I = 12$. A three-stage procedure results in 2,048 configurations. Using R 2.15.0 (R Development Core Team, 2013) and a single 2.40 GHZ core of a processor, the enumeration process takes less than one second. A four-stage procedure leads to 177,147 configurations that takes approximately 2.4 minutes to complete, and a five-stage procedure leads to 4,194,304 configurations that takes over one hour to complete. For at least up to four-stages, time is not too much of a concern. The five-stage configuration may be reasonable in some applications as long as it does not need to be repeated multiple times (see Section 3.4 for application examples). Furthermore, the enumeration of these configurations falls into a setting commonly known as "embarrassingly parallel", so additional time savings could be obtained by parallelizing the computations. Once $E(T)$ is found for each configuration, the optimal configuration is the one with the minimum expected value.

It is important to note that we have limited the possible configurations to those constructed sequentially with ordered individual probabilities. In addition to it being an intuitive approach, past research has shown that ordering is a preferred choice. For example, Black et al (2012) proved that for the special case of dividing a group into only two subgroups at each stage, ordering always produced an $E(T)$ as small as or smaller than unordered. McMahan et al. (2012a) also uses this methodology in his

application of a two-stage procedure. Finally, again for the case of a two-stage procedure, Hwang (1975) showed that groups with a larger number of individuals should always have smaller individual probabilities than groups with a fewer number of individuals.

3.2.4 Steepest descent search algorithm

For many applications, the maximum possible initial group size cannot be large due to the possibility of dilution effects. However, there are other applications where large initial group sizes can be used. For example, we discussed earlier the use of a three-stage procedure by Pilcher et al. (2005) where an initial group size of 90 is used. Other applications include Quinn et al. (2000) where a four-stage procedure is used with an initial group size of 100, subgroup sizes of 50 at stage 2, and subgroup sizes of 10 at stage 3 before individual testing at stage 4. When there are large group sizes, examining all possible configurations may not be computationally feasible without massively parallel computations performed by supercomputers. For these larger initial group size scenarios, we formulate the problem as an integer program and use the method of steepest descent to more quickly find a retesting configuration that we refer to as the candidate retesting configuration (CRC). Note that in most cases it will be the same as the ORC or lead to an expected number of tests very close to that of the ORC.

The method of steepest descent begins by first choosing a base configuration for a specified number of subgroups at each stage. We alter this base configuration by adding one member to a subgroup and subtracting one member from a different subgroup for each subgroup pair. We then choose a "better" retesting configuration that has the lowest $E(T)$ among the new ones created. We continue this same process, where the number of subgroups at each stage stays the same until no other configurations can be

found with a lower $E(T)$. We repeat this process for all other possible number of subgroups. The configuration that minimizes $E(T)$ overall is the CRC.

For example, consider an initial group of size 16 to be decoded over three stages where a base configuration at stage 2 is chosen to be [8,4,4]. We then compare $E(T)$ for that configuration to the expected number of tests for [9,4,3], [9,3,4], [8,3,5], [8,5,3], [7,5,4] and [7,4,5]. If [8,4,4] had the smallest $E(T)$ among these configurations, the process would stop. Otherwise, we choose the configuration with the smallest $E(T)$ and make this our new base configuration. We continue to perform the same add one and subtract one process to determine if there is a configuration with a smaller $E(T)$. The search algorithm ends when the base configuration has the smallest $E(T)$ within the constraints of the algorithm. In addition to three subgroups, the same process would be used for 2, 4, 5, …, 16 subgroups in order to find the CRC. For more stages, we would simultaneously change sizes of the subgroups at each stage.

There are a number of ways to choose the base configuration for the algorithm. For a three-stage procedure with initial group size of I and m_{11} subgroups to be formed from it, we arrange specimens at stage 2 so that the $I - m_{11} + 1$ individuals with the smallest probabilities are in the first subgroup. The remaining $m_{11} - 1$ subgroups each have one individual. We use base configurations of this type because subgroups should get smaller as individual probabilities increase. However, this base configuration structure, and the method of steepest descent overall, does not guarantee that the CRC found will minimize $E(T)$ overall unless there is upward convexity in $E(T)$ as a function of the number of subgroups and their sizes. In Appendix F, we give special cases where convexity fails. Despite the absence of convexity in general, we show in Sections 3.3 and 3.4 that the CRC does result in an $E(T)$ which is the same or very close to that

resulting from the ORC. Therefore, the CRC still provides a convenient alternative to the ORC when all possible configurations cannot be easily enumerated.

3.2.5 Accuracy measures

While the expected number of tests is of primary interest when evaluating a group testing procedure, the accuracy of correctly classifying truly positive and negative individuals is of great interest as well. Define $Y_i = 1(0)$ as the positive (negative) diagnosed status of the i^{th} individual ($i = 1, \ldots, I$), and define $\tilde{Y}_i = 1(0)$ in a similar manner as the true status. The probability of a correct positive diagnosis, the pooling sensitivity, is $PS_e^{(i)} = P(Y_i = 1 \mid \tilde{Y}_i = 1)$ for individual i. Similarly, the pooling specificity is $PS_p^{(i)} = P(Y_i = 0 \mid \tilde{Y}_i = 0)$ for a correct negative diagnosis of individual i. We can also define the pooling positive predictive value and the pooling negative predictive value for individual i as $PPPV^{(i)} = P(\tilde{Y}_i = 1 \mid Y_i = 1)$ and $PNPV^{(i)} = P(\tilde{Y}_i = 0 \mid Y_i = 0)$, respectively. These predictive values are useful accuracy measures once a diagnosis is made.

For the i^{th} individual to be diagnosed as positive ($Y_i = 1$), the initial group and all subsequent sub-groups containing the individual, including the last sub-group which contains only the i^{th} individual, need to test positive as well. Define subgroup j^* in stage L ($L \leq S$) as the last subgroup and stage where individual i could be tested. The individual pooling sensitivity is then

$$PS_e^{(i)} = P(Y_i = 1 \mid \tilde{Y}_i = 1) = P\left(\bigcap_{\{(s'j'):G_{Lj^*}=1\}} \{G_{s'j'} = 1\} \,\middle|\, \bigcap_{\{(s'j'):G_{Lj^*}=1\}} \{\tilde{G}_{s'j'} = 1\} \right) = S_e^L$$

where we use the assumption that test results are conditionally independent once the true status is known (see Litvak et al. 1994 for justification of this assumption). Thus, the pooling sensitivity is the same for each individual testing positive within L stages.

Note that this result is practically the same as given by Kim et al. (2007), where it was assumed that all individual probabilities of positivity were equal ($p_1 = \cdots = p_I$).

The remaining accuracy measures are found in a similar manner. The individual pooling specificity can be expressed as

$$\begin{aligned} PS_p^{(i)} &= 1 - P(Y_i = 1 \mid \tilde{Y}_i = 0) \\ &= 1 - \Big[P(Y_i = 1) - P(Y_i = 1 \cap \tilde{Y}_i = 1)\Big] \Big/ P(\tilde{Y}_i = 0) \\ &= 1 - \Big[P(Y_i = 1) - P(Y_i = 1 \mid \tilde{Y}_i = 1) P(\tilde{Y}_i = 1)\Big] \Big/ P(\tilde{Y}_i = 0) \\ &= \left[P\left(\bigcap_{\{(s'j'):G_{Lj^*}=1\}} \{G_{s'j'} = 1\}\right) - S_e^L p_i\right] \Big/ (1 - p_i) \end{aligned}$$

where $P\left(\bigcap_{\{(s'j'):G_{Lj^*}=1\}} \{G_{s'j'} = 1\}\right)$ is found from Equation (3.2). Notice that the individual pooling specificity is a function of the individual probabilities unlike what was found for $PS_e^{(i)}$. The predictive values are then found through applications of Bayes' rule:

$$PPPV^{(i)} = \frac{p_i PS_e^{(i)}}{p_i PS_e^{(i)} + (1 - p_i)(1 - PS_p^{(i)})} \text{ and } PNPV^{(i)} = \frac{(1 - p_i) PS_p^{(i)}}{(1 - p_i) PS_p^{(i)} + p_i (1 - PS_e^{(i)})}.$$

Again, we see these measures differ as functions of the individual probabilities.

To find overall measures of accuracy for all individuals screened, we need to modify our notation to designate a particular group for an individual. Define Y_{ik} and $\tilde{Y}_{ik}$ as before, but now for individual i in group k, where $i = 1, \ldots, I_k$ and $k = 1, \ldots, K$. In a similar manner, we add a k to the individual pooling sensitivity $PS_e^{(ik)}$, the individual pooling specificity $PS_p^{(ik)}$, and the individual probability of positivity $P(\tilde{Y}_{ik} = 1) = p_{ik}$ notation. Overall measures of pooling sensitivity and specificity across all individuals screened are

$$PS_e = \frac{\sum_{k=1}^{K} \sum_{i=1}^{I_k} P(\tilde{Y}_{ik} = 1 \cap Y_{ik} = 1)}{\sum_{k=1}^{K} \sum_{i=1}^{I_k} P\left(\tilde{Y}_{ik} = 1\right)} = \frac{\sum_{k=1}^{K} \sum_{i=1}^{I_k} p_{ik} PS_e^{(ik)}}{\sum_{k=1}^{K} \sum_{i=1}^{I_k} p_{ik}}$$

and

$$PS_p = \frac{\sum_{k=1}^{K}\sum_{i=1}^{I_k} P(\tilde{Y}_{ik} = 0 \cap Y_{ik} = 0)}{\sum_{k=1}^{K}\sum_{i=1}^{I_k} P(\tilde{Y}_{ik} = 0)} = \frac{\sum_{k=1}^{K}\sum_{i=1}^{I_k} (1 - p_{ik}) PS_p^{(ik)}}{\sum_{k=1}^{K}\sum_{i=1}^{I_k} (1 - p_{ik})}.$$

Overall measures of the predictive values are

$$PPPV = \frac{\sum_{k=1}^{K}\sum_{i=1}^{I_k} P(\tilde{Y}_{ik} = 1 \cap Y_{ik} = 1)}{\sum_{k=1}^{K}\sum_{i=1}^{I_k} P(Y_{ik} = 1)} = \frac{\sum_{k=1}^{K}\sum_{i=1}^{I_k} p_{ik} PS_e^{(ik)}}{\sum_{k=1}^{K}\sum_{i=1}^{I_k} [p_{ik} PS_e^{(ik)} + (1 - p_{ik})(1 - PS_p^{(ik)})]}$$

and

$$PNPV = \frac{\sum_{k=1}^{K}\sum_{i=1}^{I_k} P(\tilde{Y}_{ik} = 0 \cap Y_{ik} = 0)}{\sum_{k=1}^{K}\sum_{i=1}^{I_k} P(Y_{ik} = 0)} = \frac{\sum_{k=1}^{K}\sum_{i=1}^{I_k} (1 - p_{ik}) PS_p^{(ik)}}{\sum_{k=1}^{K}\sum_{i=1}^{I_k} [(1 - p_{ik}) PS_p^{(ik)} + p_{ik}(1 - PS_e^{(ik)})]}.$$

These expressions for the accuracy measures are logical extensions to those given in Altman and Bland (1994a, 1994b).

3.3 Mean and accuracy comparisons

In actual application, we would estimate the individual probabilities p_i. These estimates would then be used to develop the ORC or CRC and to assess overall measures of accuracy. In order to understand properties of our proposals first, it is useful to assume a particular probability distribution for p_i and examine how well our proposals work under specific situations. In particular, we use a beta distribution for p_i here because its support is between 0 and 1 and the distribution is flexible enough to emulate actual group testing applications.

Let $p_i \sim$ beta(α, $\alpha(1 - p)/p$) for $i = 1, \ldots, I$, where $E(p_i) = p$ and $Var(p_i) = p^2(1 - p)/(\alpha + p)$. For specific values of α and p, we will examine the expected performance of our proposed group testing procedures. Note that p represents the overall prevalence for a population. Also, notice that as $\alpha \to \infty$, the variance approaches 0 so that individual probabilities become homogeneous ($p_i = p$ for $i = 1, \ldots, I$). Conversely, as $\alpha \to 0$, the

variance grows, which induces more heterogeneity among the individual probabilities. In fact, one can show that the limiting distribution is Bernoulli with mean p (McMahan et al. 2012b); thus, $p_i = 1$ with probability p and $p_i = 0$ with probability $1 - p$. While this is unlikely to occur in application, it is useful to consider this situation because it maximizes the amount of heterogeneity within a group. We also choose additional values of α and p for our investigations that are motivated by the IPP data example of Section 3.4 (where maximum likelihood estimates are given in Table 3.1). For a given group size I, we calculate $E(p_{(i)})$, the expected value of an ordered individual probability, based on an α and p combination. We find the ORC and CRC using these expected values in place of their corresponding p_i to examine the corresponding expected number of tests and accuracy measures. We provide an example of a CRC for $I = 18$, $p = 0.05$, and $\alpha = 1$ in Appendix E.

We limit our subsequent investigations by the following. First, only three- and four-stage procedures are examined because a larger number of stages can be more difficult to implement in practice, while also resulting in low pooling sensitivity. Second, only groups of size 20 or less are used due to possible dilution effects in some applications and because of computational limitations. Finally, only results involving CRC are presented, because it found the same configuration as did the ORC in all three-stage cases and all four-stage cases with $I \leq 14$. For four-stage and $I > 14$, we did not compute the ORC due to excessive computation time.

Figure 3.1plots $E(T)/I$, the expected number of tests per individual, for $S_e = S_p = 0.95$. Appendix G lists the numerical values for $E(T)/I$ along with I_{sj}, the group sizes at each stage. As one would expect, Figure 3.1shows that a four-stage procedure has a smaller expected number of tests than a three-stage procedure for the same α and p, except when probabilities are homogeneous and p is large or the group size is small. Of

course, four-stage can be more difficult to apply. Also, we see that as the variability among the p_i values increases (α decreases), the expected number of tests decrease. This occurs because the retesting procedure can more adeptly take advantage of the information provided by the differences in probabilities as their spread increases. For example, in an extreme situation of $S_e = S_p = 1$ for the maximum heterogeneity case, each individual with $p_i = 1$ would be retested individually in subgroups of size 1 and each $p_i = 0$ would be retested in one subgroup, which perfectly optimizes the testing strategy.

Figure 3.2plots the overall accuracy measures for $p = 0.05$ with $S_e = S_p = 0.95$. We have included similar plots in Appendix H for $p = 0.01$, 0.10, and 0.15. Overall, as probabilities become more heterogeneous, the diagnostic values increase. This shows an important benefit gained by taking advantage of heterogeneity among the individual probabilities. At the same levels of heterogeneity (same levels of α), we see that PS_p and $PPPV$ are larger for four-stages than for three-stages, while the reverse is generally true for PS_e and $PNPV$. The reason for the PS_p and $PPPV$ result is that positive diagnoses occur only after multiple positive tests. Thus, a more stringent criterion is needed to be diagnosed as a positive with four-stages. Conversely, PS_e and $PNPV$ tend to be larger for three-stages because it takes only one negative test to produce a negative diagnosis for an individual. Note for the homogeneous case, $PS_e = S_e^S$ almost everywhere except in situations where subgroups contain single individuals before stage S.

3.4 Infertility Prevention Project

To assess how well ORC or CRC work in application, we examine a database of previously diagnosed individuals who were tested for the IPP. Our focus is on the tests performed in the state of Nebraska because they currently perform individual testing for their chlamydia and gonorrhea screening. This type of data, rather than group testing

data from other states, provides a clearer assessment because fewer assumptions need to be made through its use.

3.4.1 Background and methods of application

For each individual tested in Nebraska, information is available on clinical observations (any symptoms, cervical friability, pelvic inflammatory disease, cervicitis, and urethritis), demographic variables (age, race), and risk behavior (multiple partners, new partner in the last 90 days, contact with an individual who has a sexually transmitted disease). We treat these items as covariates (all are binary except for age) in a logistic regression model to estimate the individual probability of disease positivity. The observed disease diagnoses are treated as the true responses in fitting these models, so that we can assess the accuracy of our proposals. Separate models are found for each disease (chlamydia or gonorrhea), gender (male or female), and specimen type (urine or swab) combination. We estimate the models based on data from 23,146 individuals who were tested in 2008. These models are then applied to the 27,521 individuals tested in 2009 to obtain individual estimates of positivity, where Table 3.1summarizes these individuals by disease, gender, and specimen type combination. We apply our group testing proposals to these data from 2009.

In most applications, individuals would be placed into an initial group based on the order in which the testing laboratory received the specimen. We replicate the process here by placing individuals into groups in this same manner. Let T_k denote the number of tests needed to diagnoses individuals in the k^{th} group of size I_k. For K groups, our goal is to estimate the expected number of tests for the 2009 data using

$$\hat{E}(T_+) = \sum_{k=1}^{K} \hat{E}(T_k) = K + \sum_{s=1}^{S-1} \sum_{j=1}^{c_{sk}} m_{sjk} \hat{P}\left(\bigcap_{\{(s'j'k'):G_{sjk}=1\}} \{G_{s'j'k'} = 1\} \right),$$

where we have added a subscript k in the appropriate locations to denote the initial group number, similarly to what was done in Section 3.2.5. We calculate

$\hat{P}\left(\bigcap_{\{(s'j'):G_{Lj^*}=1\}}\{G_{s'j'}=1\}\right)$ by replacing p_i with $\hat{p}_{ik}$, which is the individual estimate of disease positivity given by our logistic regression model, in Equation (3.2).

In practice, there are two ways that an ORC or CRC can be found for a group. First, we can find a retesting configuration separately for each positive testing initial group through the use of the available estimated individual probabilities. Thus, we adaptively choose a retesting configuration using the actual individuals within a group, so we refer to this as an "adaptive" procedure (A-ORC, A-CRC). Second, a simpler method is to estimate one overall "best" retesting configuration using a training data set (the 2008 data here) and apply it to each positive testing initial group in 2009. We implement this approach by first finding $\hat{p}_{(1)k}$, ..., $\hat{p}_{(I)k}$, where the subscript "$(i)k$" allows us to denote ordered probabilities within group k, for all possible groups of size $I \leq 20$ in the 2008 data. We average these probabilities across the initial groups to form $\overline{\hat{p}}_{(1)}$, ..., $\overline{\hat{p}}_{(I)}$, and we find one retesting configuration with these probabilities. Because only one configuration is applied throughout 2009, we refer to this as a "non-adaptive" procedure (N-ORC, N-CRC). In the results that follow, we signify the maximum number of possible stages used by appending S to the end of the method, e.g., three-stage non-adaptive ORC would be N-ORC3.

3.4.2 Results

Tables 3.2 and 3.3provide summaries of our results. Note that we did not include N-ORC4 due to the excessive time it would take to complete its calculations. We have included additional results from a homogeneous probability assumption, where we find the ORC assuming equal probabilities (denoted by H3 for three-stage and H4 for four-stage in the tables), and from using ordered halving (Black et al. 2012) (denoted by OH3 for three-stage and OH4 for four-stage in the tables) for comparison purposes. For each procedure, we choose the optimal initial group size by minimizing the expected

number of tests in the 2008 data for group sizes 5 to 20 and then apply the resulting initial group size to the 2009 data.

Overall, N-CRC4 generally results in the smallest number of tests, where the reduction can be by as much as 30.6% when compared to H3 and 30.0% when compared to H4. Among the three-stage procedures, N-CRC3 and A-CRC3 always result in significant reductions in tests in comparison to H3 (up to 23.1%). Interestingly, while A-ORC3 is always less than N-ORC3, it is only by small amounts (the largest amount is 3.33%), which indicates that the much easier to implement non-adaptive approach that implements the CRC is likely to be preferred for actual application. Also, comparing N-ORC3 to N-CRC3 and A-ORC3 to A-CRC3, the biggest difference is 0.09%, so the CRC may be preferred to the ORC when the ORC takes an excessive time to compute.

With respect to accuracy, four-stage procedures are less accurate with respect to the PS_e and $PNPV$ than three-stage procedures, and vice versa for the PS_p and $PPPV$, as would be expected. When compared based on the same number of stages, the ORC and CRC procedures have similar levels of accuracy to the H3 and H4 procedures, where one procedure does not necessarily always have a larger or smaller accuracy. Also, the accuracy levels for the adaptive and non-adaptive approaches are similar.

Similar to Section 3.3, the greatest benefits to using the ORC or CRC occur when there is the most variability among the estimated individual probabilities. Table 3.1gives the variances by disease/gender/specimen combination. With respect to chlamydia, the male/swab combination has the largest variability and tends to have the largest reductions in $\hat{E}(T_+)$. With respect to gonorrhea, male/swab again has the largest variability and it has the largest or second largest reductions in $\hat{E}(T_+)$.

3.5 Discussion

In our paper, we show that finding the ORC or CRC can significantly reduce the expected number of tests during the decoding of positives from negatives, while also maintaining classification accuracy. We also show two ways that our proposals can be implemented, where the more easily applied non-adaptive approach achieved similar results to the more difficult to implement adaptive approach. This is especially important because implementation of the non-adaptive approach would allow practitioners to use only one retesting configuration throughout the screening process.

Finding an ORC over three stages is closely related to the pool-specific optimal Dorfman (PSOD) procedure proposed by McMahan et al. (2012a). In summary, PSOD is a two-stage procedure that tries to minimize the expected number of tests using a greedy search algorithm to place N individuals within an initial set of groups, where each individual has a different probability of positivity. The groups are tested, and positive testing groups are decoded through individual testing. If we let $N = I$ as in Section 3.2, we see that PSOD is very similar to the last two stages of our three-stage ORC procedure. The advantage that ORC has over PSOD is ORC's first stage can immediately diagnosis all I individuals as negative if the initial group tests negative. If the initial group tests positive for ORC, PSOD will always result in an expected number of tests greater than or equal to the remaining tests that ORC performs. This is due to ORC looking at all possible configurations and PSOD using a search algorithm that does not necessarily find the optimal configuration.

As we have discussed in this paper, hierarchical group testing has been applied in a number of areas. Our research shows that these applications could be improved upon if risk factor information can be incorporated into the testing process. To illustrate, consider again the Pilcher et al. (2005) example from Section 3.2. In their application,

the overall prevalence rate was $p = 0.00021$. Assuming no testing error, the expected number of tests for one group with their method is 1.3573. If subgroup sizes were selected based on this overall prevalence, we obtain $E(T) = 1.3572$ where 10 subgroups of size 9 would be used at stage 2, which is very similar to what was actually implemented. One would expect that individual probabilities of positivity would vary based on risk factors, but, unfortunately, this information is not available. However, if we suppose the individual probabilities had a beta distribution with $\alpha = 0.5$ and same p, then the expected number of tests for one group is reduced to $E(T) = 1.292$, which is almost a 5% decrease. If $\alpha = 0.1$ with the same p again, then $E(T) = 1.199$, which is a decrease of more than 11%. Given the number of individuals screened was 109,250, we can see a significant savings in tests could occur.

3.6 Acknowledgement

This research is supported by Grant R01 AI067373 from the National Institutes of Health.

3.7 References

Altman, D. and Bland, J. (1994a) Diagnostic tests 1: sensitivity and specificity. *BMJ*, 308, 1552.

Altman, D. and Bland, J. (1994b) Diagnostic tests 2: Predictive values. *BMJ*, 309, 102.

American Red Cross (2013) Blood testing. URL http://www.redcrossblood.org/learn-about-blood/what-happens-donated-blood/blood-testing, retrieved January 30, 2012.

Bilder, C., Tebbs, J. and Chen, P. (2010) Informative retesting. *Journal of the American Statistical Association*, 105, 942-955.

Black, M., Bilder, C. and Tebbs, J. (2012) Group testing in heterogeneous populations by using halving algorithms. *Journal of the Royal Statistical Society: Series C* (Applied Statistics), 61, 277-290.

Hourfar, M., Jork, C., Schottstedt, V., Weber-Schehl, M., Brixner, V., Busch, M., Geusendam, G., Gubbe, K., Mahnhardt, C., Mayr-Wohlfart, U., Pichl, L., Roth,

W., Schmidt, M., Seifried, E. and Wright, D. (2008) Experience of German Red Cross blood donor services with nucleic acid testing: results of screening more than 30 million blood donations for human immunodeficiency virus-1, hepatitis C virus, and hepatitis B virus. *Transfusion*, 48, 1558-1566.

Hughes-Oliver, J. (2006) Pooling experiments for blood screening and drug discovery. *In Screening: Methods for Experimentation in Industry, Drug Discovery, and Genetics*, (eds A. Dean and S. Lewis), New York: Springer.

Hwang, F. (1975) A generalized binomial group testing problem. *Journal of the American Statistical Association*, 70, 923-926.

Kim, H., Hudgens, M., Dreyfuss, J., Westreich, D. and Pilcher, C. (2007) Comparison of group testing algorithms for case identification in the presence of test error. *Biometrics*, 63, 1152-1163.

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

Litvak, E., Tu, X. and Pagano, M. (1994) Screening for the presence of a disease by pooling sera samples. Journal of the American Statistical Association, 89, 424-434.

McMahan, C., Tebbs, J. and Bilder, C. (2012a) Informative Dorfman screening. *Biometrics*, 68, 287-296.

McMahan, C., Tebbs, J. and Bilder, C. (2012b) Two-Dimensional Informative Array Testing. *Biometrics*, 68, 793-804.

Mehta, S., Nguyen, V., Osorio, G., Little, S. and Smith, D. (2011) Evaluation of pooled rapid HIV antibody screening of patients admitted to a San Diego hospital. *Journal of Virological Methods*, 174, 94-98.

Pilcher, C., Fiscus, S., Nguyen, T., Foust, E., Wolf, L., Williams, D., Ashby, R., O'Dowd, J., McPherson, T., Stalzer, B., Hightow, L., Miller, W., Eron, J., Cohen, M. and Leone, P. (2005) Detection of acute infections during HIV testing in North Carolina. *New England Journal of Medicine*, 352, 1873-1883.

Quinn, T., Brookmeyer, R., Kline, R., Shepherd, M., Paranjape, R., Mehendale, S., Gadkari, D. and Bollinger, R. (2000) Feasibility of pooling sera for HIV-1 rival RNA to diagnose acute primary HIV-1 infection and estimate HIV incidence. *AIDS*, 14, 2751-2757.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2013.URL http://www.R-project.org.ISBN 3-900051-07-0.

Smith, D., May, S., Perez-Santiago, J., Strain, M., Ignacio, C., Haubrich, R., Richman, D., Benson, C. and Little, S. (2009) The use of pooled viral load testing to identify antiretroviral treatment failure. *AIDS*, 23, 2151-2158.

Tilghman, M., Guerena, D., Licea, A., Perez-Santiago, J., Richman, D., May, S. and Smith, D. (2011) Pooled nucleic acid testing to detect antiretroviral treatment failure in Mexico. *Journal of Acquired Immune Deficiency Syndrome*, 56, e70-e74.

Disease	Gender	Specimen	Count	$\overline{p}$	$\overline{\hat{p}}$	S_e	S_p	MLE of α	MLE of p	$\hat{Var}(\hat{p}_{ik})$
Chlamydia	Female	Swab	14,503	0.069	0.066	92.8%	96.0%	2.5	0.067	0.001641
		Urine	4,970	0.080	0.083	80.5%	96.0%	1.1	0.087	0.005775
	Male	Swab	1,909	0.157	0.145	92.5%	95.0%	1.0	0.149	0.016491
		Urine	6,139	0.081	0.089	93.0%	95.0%	1.8	0.090	0.003864
Gonorrhea	Female	Swab	14,503	0.013	0.012	96.6%	98.0%	0.5	0.011	0.000236
		Urine	4,970	0.017	0.017	84.9%	98.0%	0.5	0.018	0.000613
	Male	Swab	1,909	0.070	0.068	98.5%	96.0%	0.4	0.077	0.011453
		Urine	6,139	0.021	0.017	97.0%	96.0%	0.2	0.014	0.000952

Table 3.1. Summary statistics for chlamydia and gonorrhea screening in 2009. The overall observed prevalence is denoted by $\overline{p}$. The average estimated individual probability is denoted by $\overline{\hat{p}}$. The maximum likelihood estimates (MLE) for α and p are found by fitting a beta distribution model to all estimated individual probabilities. The estimated variance of $\hat{p}_{ik}$ uses the beta distribution MLEs in its calculation.

Female/Urine								Male/Urine						
Method	I	$E(T)/N$	Change	PS_e	PS_p	$PPPV$	$PNPV$	I	$E(T)/N$	Change	PS_e	PS_p	$PPPV$	$PNPV$
H3	12	0.414	0.0%	52.19%	99.36%	88.11%	95.83%	9	0.522	0.0%	80.44%	99.17%	90.54%	98.10%
OH3	9	0.409	-1.2%	52.17%	99.35%	88.19%	95.83%	7	0.517	-0.9%	80.44%	99.10%	89.91%	98.10%
N-CRC3	16	0.394	-4.9%	52.19%	99.43%	89.49%	95.83%	8	0.497	-4.8%	81.19%	99.06%	89.94%	98.27%
N-ORC3	16	0.394	-4.9%	52.19%	99.43%	89.49%	95.83%	8	0.497	-4.8%	81.19%	99.06%	89.94%	98.27%
A-CRC3	16	0.391	-5.6%	52.56%	99.36%	88.78%	95.96%	9	0.496	-5.0%	81.04%	99.06%	89.93%	98.27%
A-ORC3	16	0.390	-5.8%	52.51%	99.37%	89.01%	95.95%	9	0.496	-5.0%	81.04%	99.06%	89.93%	98.27%
H4	20	0.370	-10.8%	42.05%	99.57%	89.90%	95.00%	18	0.519	-0.6%	74.84%	99.26%	90.85%	97.57%
OH4	19	0.354	-14.5%	42.05%	99.45%	87.93%	94.99%	13	0.506	-3.1%	74.81%	99.24%	90.91%	97.56%
N-CRC4	20	0.346	-16.4%	42.05%	99.62%	91.36%	95.00%	16	0.488	-6.5%	75.88%	99.22%	91.09%	97.79%
Female/Swab								Male/Swab						
Method	I	$E(T)/N$	Change	PS_e	PS_p	$PPPV$	$PNPV$	I	$E(T)/N$	Change	PS_e	PS_p	$PPPV$	$PNPV$
H3	9	0.431	0.0%	79.92%	99.51%	92.07%	98.59%	9	0.685	0.0%	79.17%	98.74%	91.44%	96.55%
OH3	7	0.434	0.6%	79.92%	99.46%	91.28%	98.59%	7	0.665	-2.9%	79.16%	98.66%	91.30%	96.55%
N-CRC3	8	0.420	-2.6%	80.70%	99.42%	91.01%	98.71%	9	0.630	-8.0%	80.59%	98.52%	91.08%	97.05%
N-ORC3	8	0.420	-2.6%	80.70%	99.42%	91.01%	98.71%	9	0.630	-8.0%	80.59%	98.52%	91.08%	97.05%
A-CRC3	9	0.418	-3.1%	80.28%	99.46%	91.61%	98.67%	8	0.625	-8.7%	80.55%	98.60%	91.58%	97.06%
A-ORC3	9	0.418	-3.1%	80.28%	99.46%	91.61%	98.67%	8	0.624	-8.8%	80.57%	98.60%	91.60%	97.08%
H4	18	0.422	-2.0%	74.17%	99.56%	92.35%	98.20%	20	0.682	-0.4%	73.25%	98.90%	91.90%	95.62%
OH4	13	0.417	-3.2%	74.17%	99.54%	92.11%	98.20%	17	0.639	-6.7%	73.23%	98.55%	90.24%	95.60%
N-CRC4	16	0.405	-5.9%	75.25%	99.53%	92.24%	98.35%	20	0.611	-10.8%	75.03%	98.81%	92.31%	96.19%

Table 3.2. Expected number of tests and accuracy measures for chlamydia screening, where N is the total number of individuals screened as given in Table 3.1. The "Change" column represents the percent improvement over homogeneous three-step (H3).

Female/Urine							
Method	I	$E(T)/N$	Change	PS_e	PS_p	$PPPV$	$PNPV$
H3	20	0.168	0.0%	61.22%	99.89%	91.11%	99.32%
OH3	13	0.180	7.4%	61.20%	99.87%	89.68%	99.32%
N-CRC3	20	0.155	-7.5%	61.22%	99.92%	93.10%	99.32%
N-ORC3	20	0.155	-7.5%	61.22%	99.92%	93.10%	99.32%
A-CRC3	20	0.153	-8.8%	61.38%	99.91%	92.88%	99.34%
A-ORC3	20	0.153	-8.8%	61.38%	99.91%	92.88%	99.34%
H4	20	0.149	-11.0%	52.01%	99.95%	94.47%	99.16%
OH4	19	0.148	-11.6%	52.01%	99.93%	92.81%	99.15%
N-CRC4	20	0.142	-15.0%	52.01%	99.96%	96.13%	99.16%

Male/Urine							
I	$E(T)/N$	Change	PS_e	PS_p	$PPPV$	$PNPV$
20	0.197	0.0%	91.28%	99.71%	84.73%	99.85%
15	0.206	4.4%	91.27%	99.63%	81.23%	99.85%
20	0.156	-20.6%	91.42%	99.81%	89.86%	99.87%
20	0.156	-20.6%	91.42%	99.81%	89.86%	99.87%
20	0.151	-23.1%	91.42%	99.80%	89.38%	99.88%
20	0.151	-23.1%	91.42%	99.80%	89.38%	99.88%
20	0.182	-7.7%	88.54%	99.84%	90.35%	99.80%
19	0.174	-11.9%	88.53%	99.80%	89.17%	99.80%
20	0.137	-30.6%	88.96%	99.89%	93.81%	99.85%

Female/Swab							
Method	I	$E(T)/N$	Change	PS_e	PS_p	$PPPV$	$PNPV$
H3	20	0.151	0.0%	90.14%	99.91%	92.05%	99.88%
OH3	15	0.165	9.5%	90.15%	99.87%	89.20%	99.88%
N-CRC3	20	0.133	-11.8%	90.30%	99.92%	93.24%	99.89%
N-ORC3	20	0.133	-11.8%	90.30%	99.92%	93.24%	99.89%
A-CRC3	20	0.132	-12.7%	90.26%	99.93%	93.81%	99.89%
A-ORC3	20	0.132	-12.7%	90.26%	99.93%	93.83%	99.89%
H4	20	0.141	-6.6%	87.08%	99.95%	95.13%	99.85%
OH4	19	0.140	-7.2%	87.08%	99.93%	93.80%	99.85%
N-CRC4	20	0.123	-18.9%	88.00%	99.94%	94.57%	99.89%

Male/Swab						
I	$E(T)/N$	Change	PS_e	PS_p	$PPPV$	$PNPV$
9	0.469	0.0%	95.57%	99.43%	92.55%	99.67%
7	0.443	-5.4%	95.57%	99.45%	93.30%	99.67%
13	0.371	-20.8%	95.80%	99.34%	92.46%	99.74%
13	0.371	-20.8%	95.80%	99.34%	92.46%	99.74%
13	0.362	-22.7%	95.77%	99.37%	92.82%	99.75%
13	0.362	-22.7%	95.76%	99.37%	92.84%	99.75%
9	0.475	1.3%	95.26%	99.49%	93.30%	99.65%
15	0.420	-10.4%	94.14%	99.45%	93.68%	99.57%
20	0.332	-29.1%	94.57%	99.49%	94.50%	99.71%

Table 3.3. Expected number of tests and accuracy measures for gonorrhea screening, where N is the total number of individuals screened as given in Table 3.1. The "Change" column represents the percent improvement over homogeneous three-step (H3).

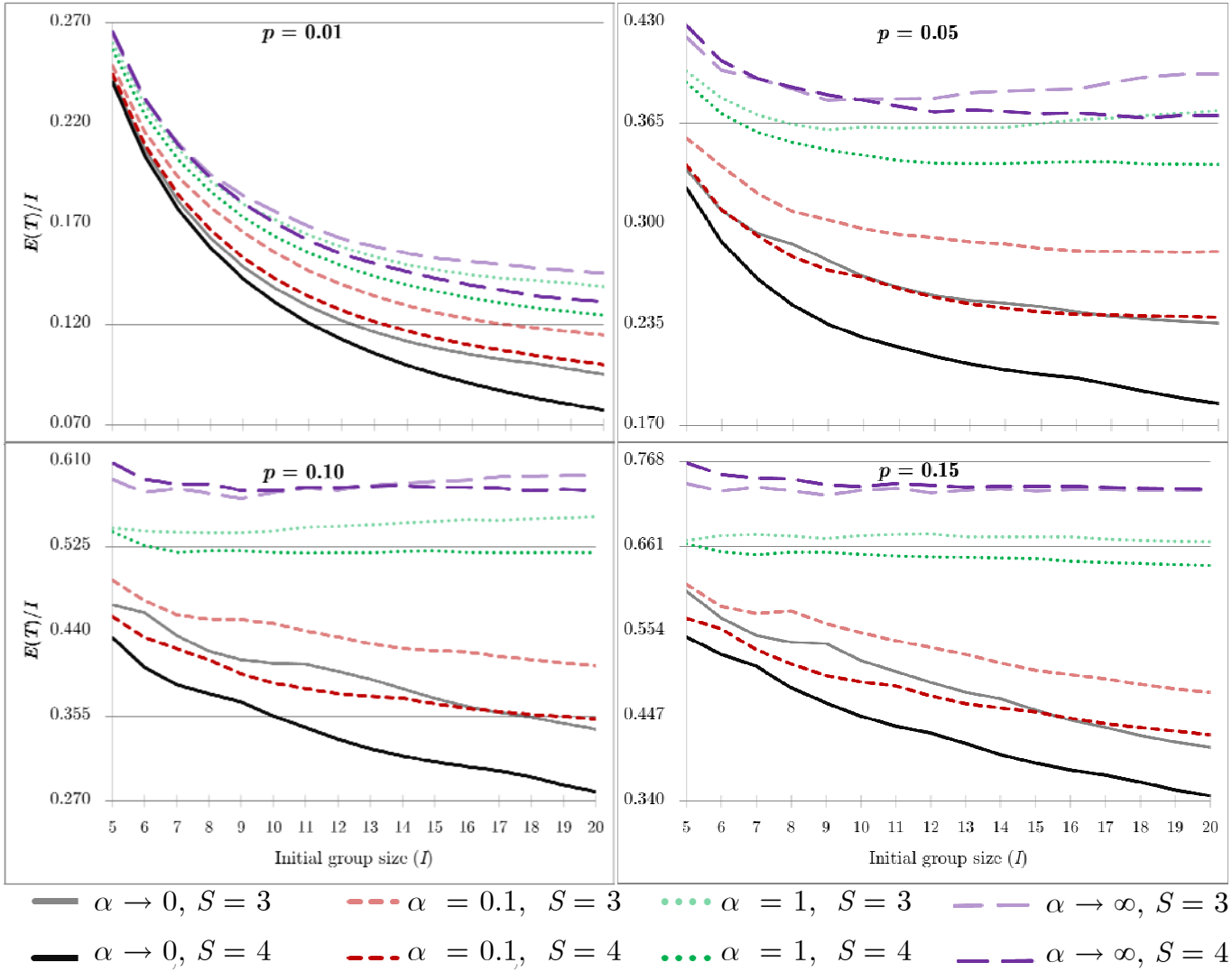

Figure 3.1. Expected number of tests per individual with $S_e = 0.95$ and $S_p = 0.95$. Each panel corresponds to a level of p.

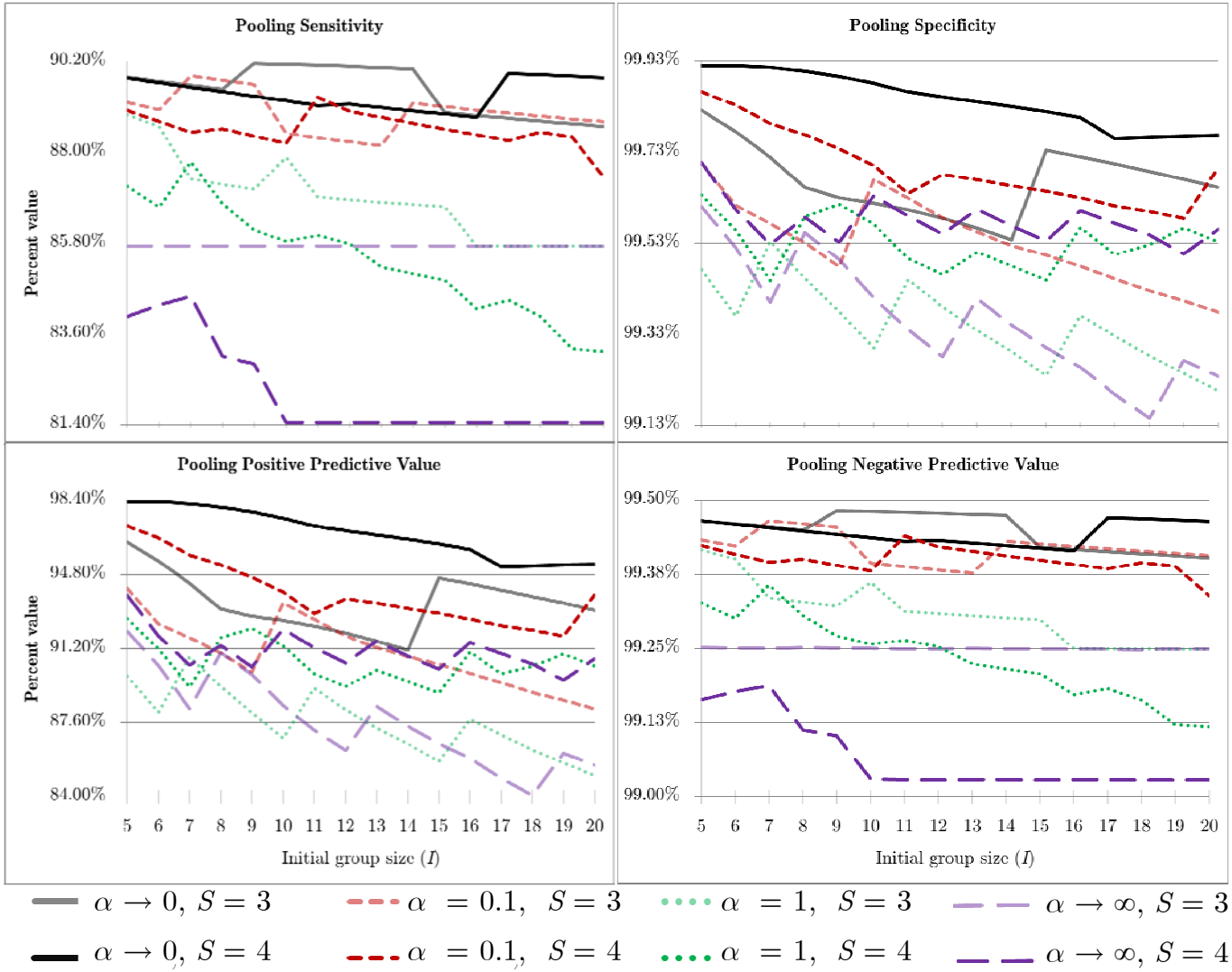

Figure 3.2. Accuracy measures when $p = 0.05$ and $S_e = S_p = 0.95$.

Chapter 4: Additional considerations

4.1 Additional investigations into halving

We present next a few interesting observations about halving. These observations show that an exact implementation of halving, while perhaps easy to implement, is not necessarily an optimal group testing procedure.

4.1.1 $I = 4$

Consider the case involving an initial group of size $I = 4$. If the initial group tests positive, the halving procedure splits the group into two subgroups of size 2. If either of these subgroups tests positive, individual testing is performed on those individuals within a positive subgroup. Interestingly, this procedure is always sub-optimal (in terms of the expected number of tests) to Dorfman's procedure when $S_e = S_p = 1$. To prove this result, let T_{Dorf} and $T_{halving}$ denote random variables for the number of tests performed under the Dorfman and halving procedures, respectively. The expected number of tests is

$$E(T_{Dorf}) = 1 + 4\left\{1 - \prod_{i=1}^{4}(1-p_i)\right\}$$

and

$$\begin{aligned} E(T_{halving}) &= 1 + 2\left\{1 - \prod_{i=1}^{4}(1-p_i)\right\} + 2\left\{1 - \prod_{i=1}^{2}(1-p_i)\right\} + 2\left\{1 - \prod_{i=3}^{4}(1-p_i)\right\} \\ &= 1 + 2 - 4\prod_{i=1}^{4}(1-p_i) + 2 - 2\prod_{i=1}^{2}(1-p_i) + 2 - 2\prod_{i=3}^{4}(1-p_i) \\ &\quad + 2\prod_{i=1}^{4}(1-p_i) \\ &= 1 + 4\left\{1 - \prod_{i=1}^{4}(1-p_i)\right\} + 2\left\{1 - \prod_{i=1}^{2}(1-p_i)\right\}\left\{1 - \prod_{i=3}^{4}(1-p_i)\right\} \\ &= E(T_{dorf}) + 2\left\{1 - \prod_{i=1}^{2}(1-p_i)\right\}\left\{1 - \prod_{i=3}^{4}(1-p_i)\right\}. \end{aligned}$$

Thus, $E(T_{halving}) > E(T_{Dorf})$ as long as p_1 or $p_2 > 0$ and p_3 or $p_4 > 0$, leading to Dorfman's procedure being preferred in application. Perhaps the more important

consequence of this result is for $I > 4$. For these larger initial group sizes, halving should never be completely applied to a point of individual testing. Rather, once a sub-group size of 4 or less is reached, Dorfman's procedure should be implemented to complete the decoding.

Of course, testing error (S_e and/or $S_p \neq 1$) typically occurs in most infectious disease applications. In Chapter 2, we saw that Dorfman's procedure is usually not better than a three-step halving procedure. For example, when S_e and S_p are different from 1 and $I = 4$, we have

$$E(T_{Dorf}) = 1 + 4S_e + 4(1 - S_p - S_e)\prod_{i=1}^{4}(1 - p_i),$$

and

$$\begin{aligned} E(T_{halving}) = {} & 1 + 2S_e + 2(1 - S_p - S_e)\prod_{i=1}^{4}(1 - p_i) \\ & + 4S_e^2 + 4(1 - S_p)(1 - S_p - S_e)\prod_{i=1}^{4}(1 - p_i) \\ & + 2S_e(1 - S_p - S_e)\left\{\prod_{i=1}^{2}(1 - p_i) + \prod_{i=1}^{4}(1 - p_i)\right\}. \end{aligned}$$

If $p_i = 0.05$ for all $i = 1, \ldots, 4$ and $S_e = S_p = 0.95$, then $E(T_{Dorf}) = 1.87$ and $\mathrm{E}(T_{halving}) = 1.81$.

4.1.2 *I*> 4

Consider again the setting with no testing error, but now $I > 4$. Also, consider the situation where at least three steps are used for decoding. One can show that removing the second halving step can lead to a smaller expected number of steps. For example, suppose $I = 16$. A four-step halving implementation would use subgroups of size 8 (step 2), size 4 (step 3), and size 1 (step 4). Instead, if a three-step procedure was implemented where four subgroups of size 4 were constructed from the initial group (immediately going to step 3 of the four-step

halving procedure), this new three-step procedure would have a smaller expected number of tests.

We can show that skipping a step leads to a smaller expected number of tests as follows. Consider a three-step procedure where step 2 has 4 groups the same size as the step 3 of a four-step halving procedure. Thus, we are essentially removing the second step from the four-step halving procedure to produce a new three-step procedure. Let $T_{3\text{-}step}$ and $T_{4\text{-}step}$ denote random variables for the number of tests performed under the three- and four-step procedures, respectively. The expected number of tests is

$$E(T_{3-step}) = 1 + 4P(G_{11} = 1) + \sum_{j=1}^{4} I_{2j} P\left(\bigcap_{\{(s'j'):G_{2j}=1\}} \{G_{s'j'} = 1\} \right)$$

and

$$\begin{aligned} E(T_{4-step}) = 1 &+ 2P(G_{11} = 1) + \sum_{j=1}^{2} 2P\left(\bigcap_{\{(s'j'):G_{2j}=1\}} \{G_{s'j'} = 1\} \right) \\ &+ \sum_{j=1}^{4} I_{3j} P\left(\bigcap_{\{(s'j'):G_{3j}=1\}} \{G_{s'j'} = 1\} \right). \end{aligned}$$

The last term in each equation is the same. Thus, by subtracting the two expressions, we obtain

$$\begin{aligned} E(T_{3-step}) - E(T_{4-step}) &= 2\left\{1 - \prod_{i=1}^{I}(1-p_i)\right\} - 2\left\{1 - \prod_{i=1}^{\lceil I/2 \rceil}(1-p_i)\right\} - 2\left\{1 - \prod_{i=\lceil I/2 \rceil+1}^{I}(1-p_i)\right\} \\ &= -2\left[1 - \prod_{i=1}^{\lceil I/2 \rceil}(1-p_i) - \prod_{i=\lceil I/2 \rceil+1}^{I}(1-p_i) + \left\{\prod_{i=1}^{\lceil I/2 \rceil}(1-p_i)\right\}\left\{\prod_{i=\lceil I/2 \rceil+1}^{I}(1-p_i)\right\}\right] \\ &= -2\left\{1 - \prod_{i=1}^{\lceil I/2 \rceil}(1-p_i)\right\}\left\{1 - \prod_{i=\lceil I/2 \rceil+1}^{I}(1-p_i)\right\} \\ &\leq 0. \end{aligned}$$

The inequality above occurs because

$$0 \leq \left\{1 - \prod_{i=1}^{\lceil I/2 \rceil}(1-p_i)\right\}\left\{1 - \prod_{i=\lceil I/2 \rceil+1}^{I}(1-p_i)\right\} \leq 1.$$

Thus, four-step halving always results in a larger expected number of tests when $p_i>0$ for at least one $i = 1, \ldots, \lfloor I/2 \rfloor$ and at least one $i = \lceil I/2 \rceil + 1, \ldots, I$. This same result occurs whenever there are two subgroups at one step that would normally be split into two additional subgroups in a next step for halving. Similar to what we saw in Section 4.1.1, the above results do not hold true when testing error is present.

4.2 PMF for generalized hierarchical retesting

This section shows how to calculate recursively the PMF for the number of tests using a generalized hierarchical retesting procedure. We use the notation from Section 3.2.1 and modify the process and notation from Appendix A of Chapter 2. As in Appendix A, we build from the second to last stage (s = S – 1), because this is the last time group sizes would be used. To begin, let $a = S_p$, $b = 1 - S_p$, $c = 1 - S_e$, and $d = S_e$. For every test, there are four possible combinations of test outcomes and true statuses, $G_{sj} = 0$ or 1 and $\tilde{G}_{sj} = 0$ or 1. In order to calculate the PMF for the number of tests, we begin by defining three sets of matrices for stage $S - 1$:

$$\boldsymbol{E}_{2j} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

a matrix of possible testing errors, $\boldsymbol{P}_{2j} = \left[P(\tilde{G}_{S-1,j} = 0), P(\tilde{G}_{S-1,j} = 1)\right]'$ a vector of probabilities for the true statuses, and $\boldsymbol{T}_{2j} = \left[0, m_{S-1,j}\right]'$ a vector for the number of tests, where $j = 1, \ldots, c_{S\text{-}1}$. The first subscript of "2" denotes the second to last level, or base level, as done in Appendix A. For a two-stage procedure, these defined matrices are the final values. For example, in the simple case of Dorfman's procedure, $c_1 = 1$ and the PMF for number of tests $\boldsymbol{T} = \boldsymbol{T}_{21} + \boldsymbol{j}_2$ is

$(\boldsymbol{P}_{21}'\boldsymbol{E}_{21})'$, where $\boldsymbol{j}_2$is a 2×1 vector of 1's added to account for the initial group test.

For a three-stage procedure, we need to build $\boldsymbol{E}_{31}$, $\boldsymbol{T}_{31}$, and $\boldsymbol{P}_{31}$ by combining the matrices for the subgroups from the penultimate stage. If for example we have a group with I individuals $m_{11} = 2$, m_{21}, and m_{22}, we have $\boldsymbol{T}_{21} = \left[0, m_{21}\right]'$, $\boldsymbol{T}_{22} = \left[0, m_{22}\right]'$, $\boldsymbol{P}_{21} = \left[P(\tilde{G}_{21} = 0), P(\tilde{G}_{21} = 1)\right]'$, $\boldsymbol{P}_{22} = \left[P(\tilde{G}_{22} = 0), P(\tilde{G}_{22} = 1)\right]'$, and $\boldsymbol{E}_{21} = \boldsymbol{E}_{22} = \boldsymbol{E}_{2j}$. Our final matrices become

$$\begin{aligned}\boldsymbol{T}_{31} &= [0, \boldsymbol{T}_{21}' \otimes \boldsymbol{j}_2' + \boldsymbol{j}_2' \otimes \boldsymbol{T}_{22}' + m_{11} * \boldsymbol{j}_4']' \\ &= [0, [0,0,m_{21},m_{21}] + [0,m_{22},0,m_{22}] + [2,2,2,2]]' \\ &= [0, 2, m_{22}+2, m_{21}+2, m_{22}+m_{22}+2]',\end{aligned}$$

$$\begin{aligned}\boldsymbol{P}_{31} = \boldsymbol{P}_{21} \otimes \boldsymbol{P}_{22} &= [P(\tilde{G}_{21} = 0), P(\tilde{G}_{21} = 1)]' \otimes [P(\tilde{G}_{22} = 0), P(\tilde{G}_{22} = 1)]' \\ &= [P(\tilde{G}_{21} = 0)P(\tilde{G}_{22} = 0), P(\tilde{G}_{21} = 0)P(\tilde{G}_{22} = 1), \\ &\quad P(\tilde{G}_{21} = 1)P(\tilde{G}_{22} = 0), P(\tilde{G}_{21} = 1)P(\tilde{G}_{22} = 1)]' \\ &= [P(\tilde{G}_{11} = 0), P(\tilde{G}_{11} = 1 \cap \tilde{G}_{21} = 0 \cap \tilde{G}_{22} = 1), \\ &\quad P(\tilde{G}_{11} = 1 \cap \tilde{G}_{21} = 1 \cap \tilde{G}_{22} = 0), P(\tilde{G}_{11} = 1 \cap \tilde{G}_{21} = 1 \cap \tilde{G}_{22} = 1)]',\end{aligned}$$

and

$$\boldsymbol{E}_{31} = \left[[a, c * \boldsymbol{j}_3']' \,||\, Diag([b, d * \boldsymbol{j}_3'])\times\left(\boldsymbol{E}_{21} \otimes \boldsymbol{E}_{22}\right)\right] = \begin{bmatrix} a & aab & abb & bab & bbb \\ c & acd & add & bcd & bdd \\ c & cad & cbd & dad & dbd \\ c & ccd & cdd & dcd & ddd \end{bmatrix}.$$

where $||$ denotes vertical concatenation, $*$ denotes scalar multiplication, $\boldsymbol{j}_m$ denotes a m×1 vector of 1's, and $\otimes$ denotes a Kronecker product. The corresponding probabilities for $\boldsymbol{T} = \boldsymbol{T}_{31} + \boldsymbol{j}_5$ can be found from $(\boldsymbol{P}_{31}'\boldsymbol{E}_{31})'$, where we see that

$$\begin{aligned}\boldsymbol{P}_{31}'\boldsymbol{E}_{31} = [&P(G_{11} = 0), P(G_{11} = 1 \cap G_{21} = 0 \cap G_{22} = 0), \\ &P(G_{11} = 1 \cap G_{21} = 0 \cap G_{22} = 1), P(G_{11} = 1 \cap G_{21} = 1 \cap G_{22} = 0), \\ &P(G_{11} = 1 \cap G_{21} = 1 \cap G_{22} = 1)].\end{aligned}$$

For the above example, we do not differ much from the halving method in Appendix A, mostly in where we add in testing from earlier stages. However, in a generalized hierarchical retesting procedure, we could have more than 2 subgroups at stage 2. For example, we could have $m_{11} = 3$, m_{21}, m_{22}, and m_{23}. Then our matrices become

$$\begin{aligned}
\boldsymbol{T}_{31} &= [0 \,||\, [\boldsymbol{T}_{21}{}' \otimes \boldsymbol{j}_2' + \boldsymbol{j}_2' \otimes \boldsymbol{T}_{22}{}'] \otimes \boldsymbol{j}_2' + \boldsymbol{j}_4' \boldsymbol{T}_{22}{}' + m_{11} * \boldsymbol{j}_8']' \\
&= [0 \,||\, [0, 0, m_{22}, m_{22}, m_{21}, m_{21}, m_{22} + m_{21}, m_{22} + m_{21}]' \\
&+ [0, m_{23}, 0, m_{23}, 0, m_{23}, 0, m_{23}] + [3, 3, 3, 3, 3, 3, 3, 3]]' \\
&= [0, 3, m_{23} + 3, m_{22} + 3, m_{23} + m_{22} + 3, m_{21} + 3, \\
&m_{23} + m_{21} + 3, m_{22} + m_{21} + 3, m_{23} + m_{22} + m_{21} + 3]',
\end{aligned}$$

$$\boldsymbol{P}_{31} = \boldsymbol{P}_{21} \otimes \boldsymbol{P}_{22} \otimes \boldsymbol{P}_{23}$$

and

$$\boldsymbol{E}_{31} = \Big[[a, c * \boldsymbol{j}_7']' \,||\, Diag([b, d * \boldsymbol{j}_7']) \times \Big(\boldsymbol{E}_{21} \otimes \boldsymbol{E}_{22} \otimes \boldsymbol{E}_{23}\Big)\Big].$$

Specifically, if $I = 10$, $m_{11} = 3$, $m_{21} = 5$, $m_{22} = 4$, and $m_{23} = 0$ our final

$$\boldsymbol{T} = [1, 4, 4, 8, 8, 9, 9, 13, 13]'.$$

We see repeats in the number of tests, because the 10^{th} individual has finished testing at stage 2 and so the number of tests is not changed by its testing status, positive or negative. To get the PMF for the unique set of possible number of tests, we would sum the corresponding probabilities for the identical number of tests.

For the general case, we build the PMF up from the penultimate step using the matrices

$$\begin{aligned}
\boldsymbol{T}_{sj} &= \Big[0, \oplus_{i=i^*+1}^{i^*+m_{S-s+1,j}} \boldsymbol{T}_{s-1,i}'\Big]' + \Big[0, m_{S-s+1,j} * \boldsymbol{j}_k'\Big]', \\
\boldsymbol{P}_{sj} &= \otimes_{i=i^*+1}^{i^*+m_{S-s+1,j}} \boldsymbol{P}_{s-1,i},
\end{aligned}$$

and

$$\boldsymbol{E}_{sj} = \left[[a, c * \boldsymbol{j}_k']' \;||\; Diag([b, d * \boldsymbol{j}_k']) \times \left(\otimes_{i=i^*+1}^{i^*+m_{S-s+1,j}} \boldsymbol{E}_{s-1,i} \right) \right],$$

for $s = 3, \ldots, S$ and $j = 1, \ldots, c_{S-s+1}$, k is the value that makes parts conformable and can vary in value based on the number of subgroups at each stage and the number of individuals who finish testing early, i^* is the last subscript for the subgroups split from the j^{th} subgroup of the $S - s + 1$ stage, and we use $\oplus$ to denote the operation

$$\boldsymbol{A} \oplus \boldsymbol{B} = \boldsymbol{A} \otimes \boldsymbol{1}_B + \boldsymbol{1}_A \otimes \boldsymbol{B},$$

where $\boldsymbol{1}_B$ and $\boldsymbol{1}_A$ are matrices of 1's with the dimensions of matrices $\boldsymbol{B}$ and $\boldsymbol{A}$ respectively. These differ from the halving functions, because for multiple subgroups at each stage we have to include all the splits. As we move up through the stages if there is individual testing that occurred at the earlier stage, we need to add additional

$$\boldsymbol{E}_{sj} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \; \boldsymbol{P}_{sj} = \left[P(\tilde{G}_{S-s+1,j} = 0), P(\tilde{G}_{S-s+1,j} = 1) \right]', \text{ and } \boldsymbol{T}_{sj} = [0, 0]$$

for each j at the $S - s + 1$ stage such that $m_{S-s+1,j} = 0$.

Our final resulting matrices will be $\boldsymbol{T}_{S1}$, $\boldsymbol{P}_{S1}$, and $\boldsymbol{E}_{S1}$. The corresponding probabilities for $\boldsymbol{T} = \boldsymbol{T}_{S1} + \boldsymbol{j}_k$ can be found from $(\boldsymbol{P}_{S1}' \boldsymbol{E}_{S1})'$. Because we may have repeats for the number of tests, we sum the probabilities over the corresponding unique number of tests in $\boldsymbol{T}$ to obtain the PMF.

These operations can be continued for any number of steps, and our algorithm is designed to allow for any combination of final sub-group sizes. Large matrices result when S is not small (e.g., $S \geq 6$) and when c_s the number of subgroups at each stage is large. This causes memory problems from using the matrix methods in R's base package (R Development Core Team, 2013). However, this is not too limiting because S and number of subgroups are usually small in practice.

4.3 The CRC computational process

To find the CRC by iterating through possible retesting configurations, we need to modify our notation from Chapter 3 in order to track all individuals until the final stage. In Chapter 3 if an individual had finished testing before the final stage S, a corresponding sub-group could no longer be split, which led to a $m_{sj} = 0$ to signify individual testing had occurred (e.g., see Figures E.1 and E.2). For computational purposes, we now need to have all individual testing occur at the final stage. Therefore, we replace our m_{sj} notation with $\breve{m}_{sj}$ as the number of subgroups a group would be split into. To illustrate, Figure 4.1 diagrams the five-stage example given in Figure E.1. A yellow box indicates testing occurs while a pink box indicates no testing occurs. We can see that $\breve{m}_{sj} = 1$ for these pink boxes because the sub-group is not being split. The $S = 5$ final stage has the individual testing for all $I = 10$ individuals. Note that by removing the pink boxes and shifting the remaining subgroups up in the diagram, we obtain Figure E.1. To distinguish the differences between Chapter 3 and here, we call a diagram like in Figure 4.1 a "computational configuration" as opposed to a retesting configuration in Chapter 3.

With these computational configurations, the expected number of tests given in Equation 3.1 becomes

$$E(T) = 1 + \sum_{s=1}^{S-1} \sum_{j=1}^{c_s} t_{sj} \breve{m}_{sj} P\left(\bigcap_{\{(s'j'):G_{sj}=1\}} \{G_{s'j'} = 1\} \right),$$

where we include the indicator function t_{sj} to account for subgroups that are not actually tested. The value of t_{sj} is 1 if $\breve{m}_{sj} > 1$, which allows us to include the $\breve{m}_{sj} P\left(\bigcap_{\{(s'j'):G_{sj}=1\}} \{G_{s'j'} = 1\} \right)$ contribution to $E(T)$. The value of t_{sj} is 0 when $\breve{m}_{sj} = 1$, because either no test is performed or individual testing is performed at stage S.

Figure 4.1 shows t_{sj} in the context of the five-stage example. With this formulation, note that $\sum_{j=1}^{c_s} I_{sj} = I$ for all $s = 1, \ldots, S$.

All possible computational configurations can be defined by finding $\breve{m}_{sj}$ for $s = 2, \ldots, S-1$, and $j = 1, \ldots, c_s$ so that the following conditions are met

Condition 1. $\quad 1 \le \breve{m}_{sj} \le c_{s+1} - c_s + 1$

Condition 2. $\quad \sum_{j=1}^{c_s} \breve{m}_{sj} = c_{s+1}$,

where $c_S = I$. Condition 1 is simply saying that each subgroup must contain at least one subgroup for the next stage. The upper limit follows because if each must contain at least one, then no single subgroup could contain more than the total subgroups for the next stage minus the number of other subgroups at the given stage. For $\breve{m}_{sj}$ at either extreme of Condition 1, we say that we are at an edge. Additionally, each set of $\breve{m}_{sj}$ defines the number of individuals in each subgroup at each stage for the computational configuration, as can be seen in Figure 4.1.

To facilitate iterating through the computational configurations when finding the CRC, define $\breve{\boldsymbol{m}}_s = (\breve{m}_{s1}, \ldots, \breve{m}_{sc_s})'$ as a vector indicating the number of subgroups that a positive group at stage s will be split into for all c_s subgroups. For example, Figure 4.2 illustrates a computational configuration corresponding to a retesting configuration first discussed in Chapter 3 (Figure E.2) where $\breve{\boldsymbol{m}}_1 = 3$, $\breve{\boldsymbol{m}}_2 = (3,3,1)'$ and $\breve{\boldsymbol{m}}_2 = (5,3,3,2,1,1,3)'$. For a given number of subgroups c_s, with $s = 2, \ldots, S-1$, we perform the following:

1. Start with the minimum number of subgroups at each stage. For a three-stage procedure, this means we start with $c_2 = 2$. For a four-stage procedure, this means we start with $c_2 = 2$ and $c_3 = 3$, because the third stage must have at least one more subgroup than the second.

2. Choose an initial configuration of $\breve{\boldsymbol{m}}_s$ with $s = 2, \ldots, S-1$ for the given number of subgroups and calculate $E(T)$.
3. Move one vertex in each plausible direction from the initial configuration and calculate the change in $E(T)$ at these points (i.e. each possible configuration is represented by a vertex in $\mathbb{N}^n$ space, connected to every other possible configuration by simple transformations, where $n = \sum_{s=2}^{S-1} c_s$).
4. If there is no decrease in $E(T)$, then our initial configuration is the best for this number of subgroups. Otherwise, find the direction of largest decrease in $E(T)$ and continue moving to vertices in the direction of largest decrease until reaching an edge or until $E(T)$ stops decreasing in that direction. The configuration with this new smallest $E(T)$ has applied the same process outlined in 3.

This process is repeated for other number of subgroups c_s, where the c_s values are increased each time in an iterative manner. The configuration with the smallest $E(T)$ for all possible number of subgroups is the CRC. Alternatively, after two increases in c_{S-1} with no decrease in the minimum $E(T)$ during the iterative procedure, the process is stopped and the smallest $E(T)$ found is treated as its smallest possible value. The corresponding configuration would then be taken as the CRC. Through observing a number of examples, this alternative method appears to work just as well as finding $E(T)$ for all possible numbers of subgroups and it substantially reduces run times. The computational configuration corresponding to the smallest $E(T)$ is used as the CRC.

As an example of how we move between vertices, suppose for a four-stage procedure we had $c_2 = 3$ and $c_3 = 6$ as the number of subgroups, an initial

configuration of $\breve{\boldsymbol{m}}_1 = 3$, $\breve{\boldsymbol{m}}_2 = (3, 2, 1)'$ and $\breve{\boldsymbol{m}}_3 = (3, 2, 1, 2, 1, 1)'$, and an initial group size of $I = 10$. To move one vertex, we change individual elements in $\breve{\boldsymbol{m}}_2$ and/or $\breve{\boldsymbol{m}}_3$ by +1 for one subgroup and -1 for another subgroup. For example, $\breve{\boldsymbol{m}}_2$ might change to $(2, 2, 2)'$ resulting in, say, $\nabla\breve{\boldsymbol{m}}_2 = (-1, 0, 1)'$. Also, $\breve{\boldsymbol{m}}_3$ might change to $(4, 2, 1, 1, 1, 1)'$ resulting in $\nabla\breve{\boldsymbol{m}}_3 = (1, 0, 0, -1, 0, 0)'$. Note that both $\breve{\boldsymbol{m}}_2$ and $\breve{\boldsymbol{m}}_3$ can change simultaneously.

Finding the CRC requires an initial starting configuration. For a three-stage procedure with a given number of subgroups at stage 2, this is simply done by setting $\breve{m}_{2j} = 1$ for $j = 2, \ldots, c_2$ and $\breve{m}_{21} = I - c_2 + 1$; that is $\breve{\boldsymbol{m}}_2 = (I - c_2 + 1, 1, \ldots, 1)'$. Starting with this configuration is intuitive because individuals are ordered by their probabilities of positivity. Thus, individuals with the largest probabilities are tested individually, and individuals with the lowest probabilities are tested as a group. When we have a high level of heterogeneity among the individual probabilities, we have found that this method is very effective and quick at finding the optimal configuration. However, when there are groups with low heterogeneity among individual probabilities, the CRC takes much longer to find its configuration. This is because the ORC would have subgroups with individuals more evenly divided between them than when a high level of heterogeneity exists. Despite the CRC taking a longer time, we would like to point out that the CRC always matches ORC for our beta distribution examples in Chapter 3.

If we pick the initial computational configuration for the four-stage procedures similarly to how we did for the three-stage procedures with $\breve{\boldsymbol{m}}_2 = (c_3 - c_2 + 1, 1, \ldots, 1)'$, and $\breve{\boldsymbol{m}}_3 = (I - c_3 + 1, 1, \ldots, 1)'$, the CRC generally finds the ORC for highly heterogeneous groupings, but does poorly as heterogeneity

decreases. If however, we choose $\breve{\boldsymbol{m}}_2 \approx (c_3/c_2, \ldots, c_3/c_2)'$, and $\breve{\boldsymbol{m}}_3 \approx (I/c_3, \ldots, I/c_3)'$, where we use "$\approx$" to mean approximately because the fractions may not be integers, the CRC finds the ORC for low heterogeneity relatively well, but does poorly when there is a high level of heterogeneity. Because of this, we start at both configurations for a four-stage procedure to better our chances of finding (or coming close to) the ORC. While these are not the only possible starting configurations, our results in Chapter 3 indicate that our choices end up working very well.

4.4 R functions

In this section, we show how to use four functions that can be used to implement the methods described in Chapters 2 and 3. First, `beta.dist()` computes the expected ordered probabilities from a beta distribution. Second, `halving()` computes the PMF for the halving procedure. Third, `hierarchical.desc()` returns a variety of information for hierarchical group testing, including the expected number of tests and accuracy measures. Finally, `get.CRC()` is the general hierarchical optimization function that finds the CRC or ORC for a given set of individual probabilities. Appendix I contains the R documentation for these functions.

4.4.1 Expected value of order statistics from a beta distribution

In Chapters 2 and 3, we use the order statistics from beta distributions to examine how well our proposed group testing procedures work. Our function `beta.dist()` computes these order statistics. The argument `grp.sz` is used to specify the initial group size and the arguments `p` and `alpha` are used to specify the distribution using the parameterization given in Section 3.3. Alternatively, the arguments `alpha` and `beta` can be used to parameterize the distribution in

the more conventional form as given by Casella and Berger (2002, p. 623). For example, suppose $E(p_i) = 0.05$ for $i = 1, \ldots, 16$ and $\alpha = 1$. Below are the corresponding expected values of the order statistics.

```
> p1 <- beta.dist(p = 0.05, alpha = 1, grp.sz = 16, plot = TRUE)
>round(p1, 4)
 [1] 0.0033 0.0068 0.0105 0.0145 0.0188 0.0234 0.0286 0.0342
 [9] 0.0405 0.0477 0.0560 0.0658 0.0779 0.0938 0.1171 0.1612
```

A plot of the distribution with the expected values of the order statistics is created by using the `plot = TRUE` argument. The plot for `p1` is given in Figure 4.3.

4.4.2 PMF for the halving procedure

The `halving()` function returns the PMF, $E(T)$, and $Var(T)$ for the halving procedure up to five steps. This function has the required argument `p` representing a vector of individual probabilities. The optional arguments include `se` for the sensitivity (default is `se = 1`), `sp` for the specificity (default is `sp = 1`), `stages` for the number of steps (default is `stages = 2`), and `order.p` for whether or not to order the individual probabilities within p (default is `order.p = TRUE`)

As a first example of using the function, consider a homogeneous set of individual probabilities where $p_i = 0.05$ for $i = 1 \ldots 16$, $S_e = S_p = 1$, and 3 steps:

```
> ex1 <- halving(p = rep(x = 0.05, times = 16), stages = 3)
> ex1
$pmf
num.testsprob.tests
1           1     0.4401
2           3     0.0000
3          11     0.4466
4          19     0.1133
$et
[1] 7.50502
$vt
[1] 39.04807
```

Notice that the number of tests t can only be $t = 1$, 11, and 19 in this setting, where $t = 3$ is not possible because there is no testing error. Two additional examples where $\mathrm{E}(p_i) = 0.05$ are given below:

```
> ex2 <- halving(p = rep(x = 0.05, times = 16), stages = 3, se =
         0.95, sp = 1)
> p1 <- beta.dist(p = 0.05, alpha = 1, grp.sz = 16)
> ex3 <- halving(p = p1, stages = 3, se = 0.95, sp = 1)

>data.frame(t = ex1$pmf[,1], ex1 = ex1$pmf[,2], ex2 = ex2$pmf[,2],
          ex3 = ex3$pmf[,2], row.names = NULL)  #PMF
t    ex1    ex2    ex3
1  1 0.4401 0.4681 0.4612
2  3 0.0000 0.0215 0.0240
3 11 0.4466 0.4133 0.4581
4 19 0.1133 0.0971 0.0568
>data.frame(ex1 = ex1$et, ex2 = ex2$et, ex3 = ex3$et) #E(T)
      ex1      ex2      ex3
1 7.50502 6.923968 6.651044
>data.frame(ex1 = ex1$vt, ex2 = ex2$vt, ex3 = ex3$vt) #Var(T)
       ex1      ex2      ex3
1 39.04807 37.78906 32.36854
```

The second example is for a homogenous case with $S_e = 0.95$ and $S_p = 1$, and the third example uses ordered halving where the expected order statistics from a beta distribution with $p = 0.05$ and $\alpha = 1$ are used. Comparing the two homogeneous cases, we see that the expected number of tests decreases for the case with testing error due to now the possibility of false negative groups or subgroups. A further reduction in the expected number of tests occurs for ordered halving illustrating the benefits extolled in Chapter 2.

4.4.3 Descriptive information for generalized hierarchical group testing procedure

The `hierarchical.desc()` function computes the expected number of tests and measurements of accuracy for two-, three-, or four-stage hierarchical group testing. This function has the required argument `p` representing a vector of individual probabilities. The optional arguments include `se` for the sensitivity (default is `se = 1`), `sp` for the specificity (default is `sp = 1`), `stages` for the

number of steps (default is `stages = 2`), and `order.p` for whether or not to order the individual probabilities within p (default is `order.p = TRUE`). The arguments `I2` and `I3` define the number of individuals at each subgroup in stages 2 and 3 and are used to define the retesting procedure considered. The default retesting procedure is two-stage (Dorfman), which is equivalent to when `I2 = NULL` and `I3 = NULL`. To get three-stage results, `I2` needs a vector with the number of individuals in each subgroup at stage 2 and `I3 = NULL`. To get four-stage results, `I2` and `I3` need vectors with the number of individuals in each subgroup at stage 2 and stage 3, respectively. For example, with $S_e = 0.95$ and $S_p = 1$, and the expected order statistics from a beta distribution with $p = 0.05$ and $\alpha = 1$, we get the same $E(T)$ found for three-stage halving by using `I2 = c(8, 8)` and `I3 = NULL`.

```
> p1 <- beta.dist(p = 0.05, alpha = 1, beta = NULL, grp.sz = 16)
> ex4 <- hierarchical.desc(p = p1, se = 0.95, sp = 1, I2 = c(8, 8),
          I3 = NULL)
> ex4
[1] "Three-stage procedure considered"
$ET
[1] 6.651044
$stages
[1] 3
$group.size
[1] 16
$I2
[1] 8 8
$I3
[1] "individual testing"
$m1
[1] 2
$m2
[1] 8 8
$m3
[1] "individual testing"
$individual.testerror
    p           pse.vec   psp.vec pppv.vec pnpv.vec
1  0.003278689 0.857375       1        1  0.9995311
<Output edited>
16 0.161210126 0.857375       1        1  0.9733197
$group.testerror
PSe         PSp       PPPV      PNPV
0.8573750 1.0000000 1.0000000 0.9925493
```

```
$individual.probabilities
 [1] 0.003278689 <Output edited> 0.161210126
```

The output produced provides descriptive information about the retesting procedure. The testing errors are given for the individuals in the group and for the group as a whole; these measures are calculated using the formulas defined in Section 3.2.3. The function also provides `m1`, `m2`, and `m3` as the numbers of subgroups that each positive testing group will split into at the next stage.

4.4.4 Optimal or candidate retesting configuration

The `find.CRC()` function finds the ORC or CRC to use with hierarchical group testing. Its arguments include p as a vector of individual probabilities, `se` for the sensitivity (default is `se = 1`), `sp` for the specificity (default is `sp = 1`), `stages` for the number of stages (default is `stages = 2`), and `every case` for the ORC (`= TRUE`) or the CRC (`= FALSE`). The function finds the CRC by default. If the ORC is chosen instead, a warning is provided that it may take a long period of time to complete the calculations. The function uses `hierarchical.desc()` to produce the expected number of tests and accuracy measures for the configuration chosen. For example, we show below how to find the CRC for a three-stage procedure, with $S_e = 0.95$, $S_p = 1$ and probabilities that are the expected order statistics from a beta distribution with $p = 0.05$ and $\alpha = 1$.

```
> p1 <- beta.dist(p = 0.05, a = 1, b = NULL, grp.sz = 16)
> ex5 <- get.CRC(p = p1, se = .95, sp = 1, stages = 3)
[1] "Three-stage procedure considered"
> ex5
$ET
[1] 5.460507
$stages
[1] 3
$group.size
[1] 16
$I2
[1] 7 4 3 2
$I3
[1] "individual testing"
```

```
$m1
[1] 4
$m2
[1] 7 4 3 2
$m3
[1] "individual testing"
$individual.testerror
     p          pse.vec     psp.vec     pppv.vec    pnpv.vec
1  0.003278689 0.857375       1           1         0.9995311
<Output edited>
16 0.161210126 0.857375       1           1         0.9733197
$group.testerror
PSe        PSp        PPPV      PNPV
0.8573750 1.0000000 1.0000000 0.9925493
$individual.probabilities
[1] 0.003278689 <Output edited> 0.161210126
```

Notice that the expected number of tests for this configuration is smaller than as shown for halving earlier with the same number of stages and testing error.

4.5 Future research

We have repeatedly shown in this dissertation that taking advantage of covariate information can decrease the number of tests needed for specific group testing procedures. We describe next extensions to our proposed methods and other ideas for future exploration. First, we discuss how group testing procedures can be used under specific laboratory conditions. Second, many screening procedures have some type of confirmation testing; i.e., after an initial screening, a more accurate test is used to double check if any positive individuals have been missed. It may be useful to apply the extra covariate information to these confirmation situations as well. Finally, the computational process for finding the CRC and ORC could be improved by using parallel processing.

4.5.1 High throughput testing situations

Many high throughput testing systems are available for diagnostic and chemical screening situations. High throughput systems allow for multiple simultaneous and automated testing to increase how quickly tests can be performed. One type of test used in these settings is an enzyme-linked immunosorbent assay (ELISA).

Equipment for ELISA testing is available in many forms. For example, ELISA tests can be done on 96 well plates (8 by 12), which can be processed and tested in single-channel (one well at a time), multichannel (rows of 8 or 12 at a time), semi-automated (one strip or plate at a time), or fully automated (multiple plates at a time) systems (IDEXX 2010). Which system is used depends on the type of screening being performed, training of personnel, and financial considerations. Ultrahigh throughput chemical compound screening systems can even screen up to 100,000 compounds per day (Remlinger, 2006). Understanding how retesting procedures can by utilized in these processes is very important to decide if a retesting procedure should be used and which one.

For example, one 96 well ELISA plate performs 96 evaluations. This could be used to immediately screen 96 individuals or alternatively 96 pooled specimens. With an initial size for groups of $I = 10$ and an overall prevalence of $p = 0.01$, one plate could provide the initial screening of 960 individuals, where approximately 810 to 920 individuals would be classified as negative. This approximation comes from using a normal approximation to a binomial ($96\left[\theta \pm Z_{0.975}\sqrt{\theta(1-\theta)\,/\,96}\right]$ where $\theta = 1-(1-0.01)^{10}$). The remaining individuals would need to be retested to diagnose the positive individuals. A Dorfman procedure would likely need only 1 or 2 more plates to complete identification of all positive individuals. This would save approximately 714 to 824 tests or 7 to 8 plates for classifying the 960 individuals. While this is a good improvement, the number of tests saved could be further reduced with more sophisticated retesting procedures. However, depending on how we order the tests, retesting procedures may not improve the number of plates used.

To evaluate how an overall process is improved by some retesting method, we will need to determine how well it can save time and reduce costs while improving or maintaining accuracy of results. Some factors that may affect how an overall process performs in comparison to others are the total number of tests used, the total number of plates used, the ease of implementation/automation, the efficient use of testing material, etc. Two-stage procedures such as Dorfman or multi-array testing would be easiest to implement and measure possible improvements. Multiple stage procedures such as Sterrett would be more difficult to implement and evaluate.

The retesting methods we have looked at treat an initial group as a contained testing unit. In a high throughput testing environment with multiple simultaneous tests being performed, it may be useful to look at retesting procedures that combine individuals across positive groups in the retesting phase (Sobel and Groll, 1958). For example, if we have two positive groups of size ten, it may be useful to retest the five lowest probability individuals from each group in one new group, while retesting the five highest probabilities from each group individually.

4.5.2 Changing assays

In many situations, there are two types of assays used. First, a less expensive assay with greater testing error is used, and second, a confirmatory assay that is more expensive but with higher accuracy is used. Additionally, a commonly used assumption in group testing is that sensitivity and specificity are the same for different sized groups that use the same assay; however, some research has shown they can vary for different group sizes and taking possible variations into account can improve estimation (McMahan, 2013). The research in this dissertation

assumes the same assay or, at least, the same sensitivity and specificity is used at each stage of testing. Future research could incorporate initial (lower accuracy) and confirmatory (higher accuracy) assays into the testing protocol.

One possible method to investigate involves immediately placing in the higher accuracy testing path any individuals whose probability of a false positive from the lower accuracy assay is large. How large and at what stage of retesting would be of future research interest.

A second method might be to stop testing individuals with the lower accuracy tests if a maximum expected number of positive individuals is reached, either for the entire population being screened or group by group. Since individuals being positive or negative are a series of Bernoulli trials, finding a PMF for the number of positive individuals can be found and used to find an upper bound for the likely total number of positive individuals expected in either an entire population or group by group. Instead of looking at the maximum number of individuals who are positive, we could alternatively look at the maximum number testing positive. Equation 3.2 can be used to find the probability each individual tests positive within the framework of a specific retesting procedure. These probabilities could be used in place of the individual probabilities of being positive to find a likely upper bound for the number of individuals testing positive.

4.5.3 Parallel processing

With the release of R 2.14.0, the parallel package for parallel processing was made part of the default installation for R (R Development Core Team, 2013). Parallel processing allows one to split-up computations to multiple cores of a single processor or across several processors. The end result is potentially large

reductions in computation time to complete a task. The nature of a search for an ORC would be well suited for improvement through using parallel processing techniques, because it could repetitiously iterate through possible configurations. Running simultaneous iterations of separate collections of possible configurations could greatly reduce the computation time making larger group sizes computationally possible.

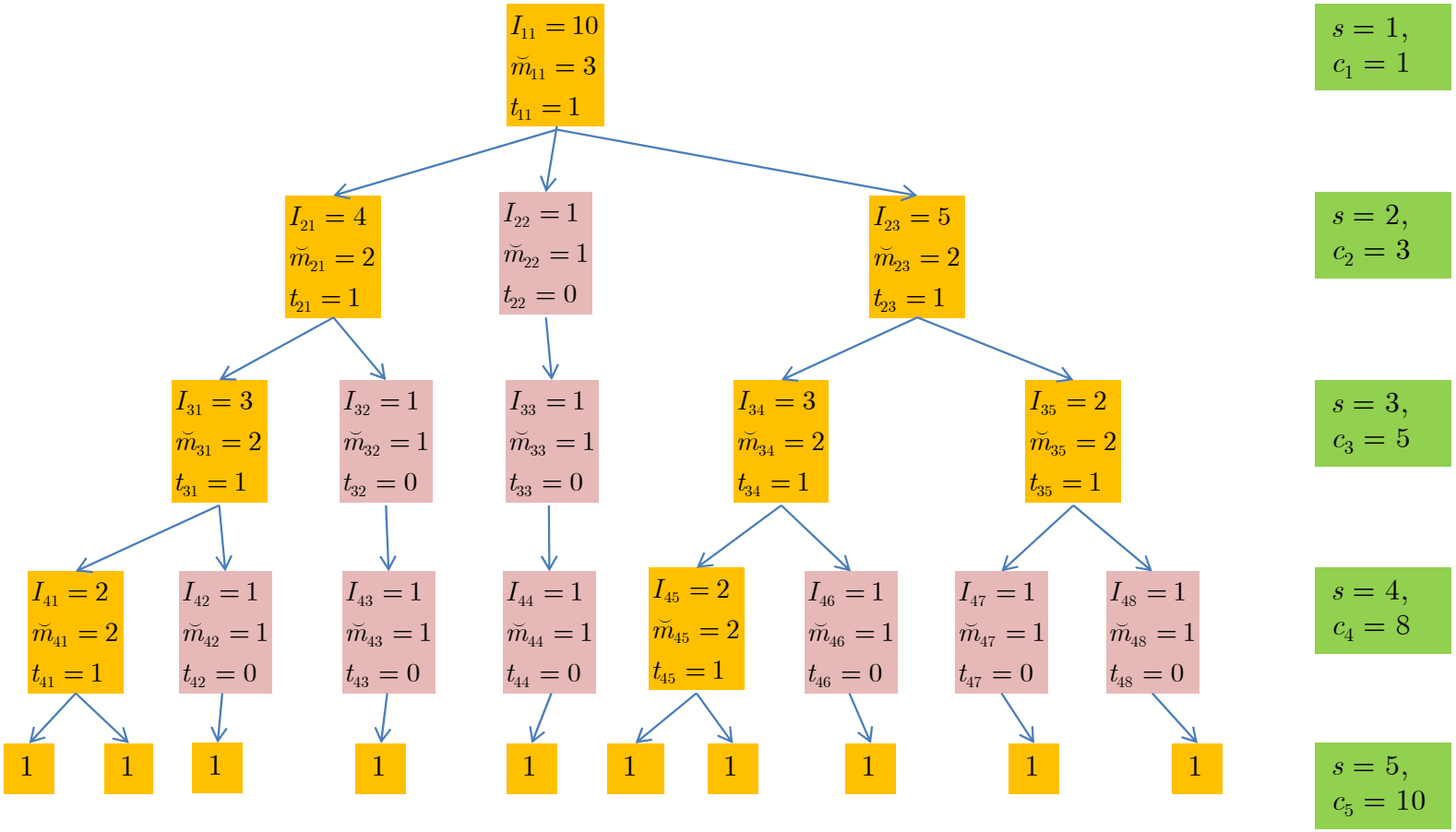

Figure 4.1. Possible, not necessarily optimal, computational configuration for a group size of 10, using notation for iterating through possible configurations. Figure E.1 shows the retesting configuration associated with this computational configuration.

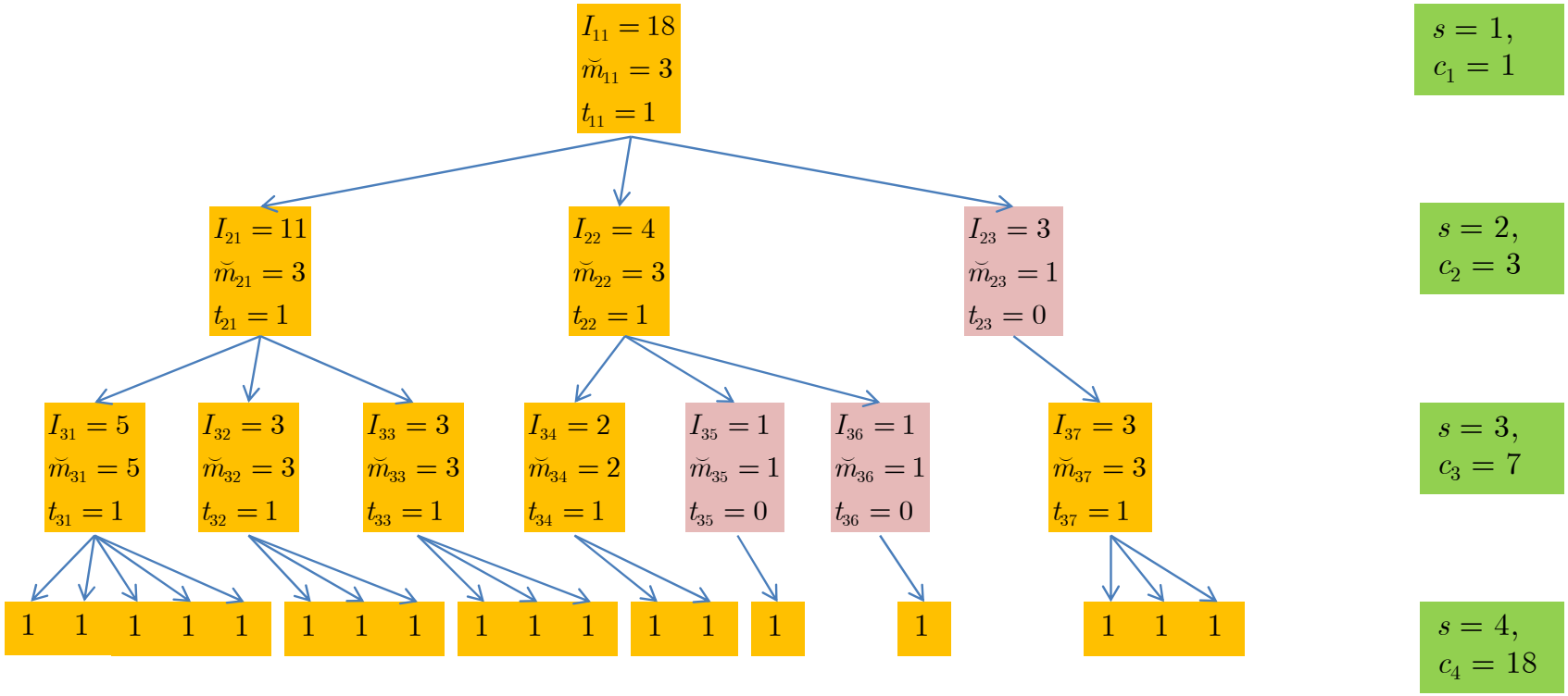

Figure 4.2. CRC computational configuration using notation for iterating through possible configurations for an expected ordered beta distribution with $I = 18$, $p = 0.05$, and $\alpha = 1$, also $S_e = 0.95$ and $S_p = 0.95$. Figure E.2 shows the retesting configuration used by this computational configuration.

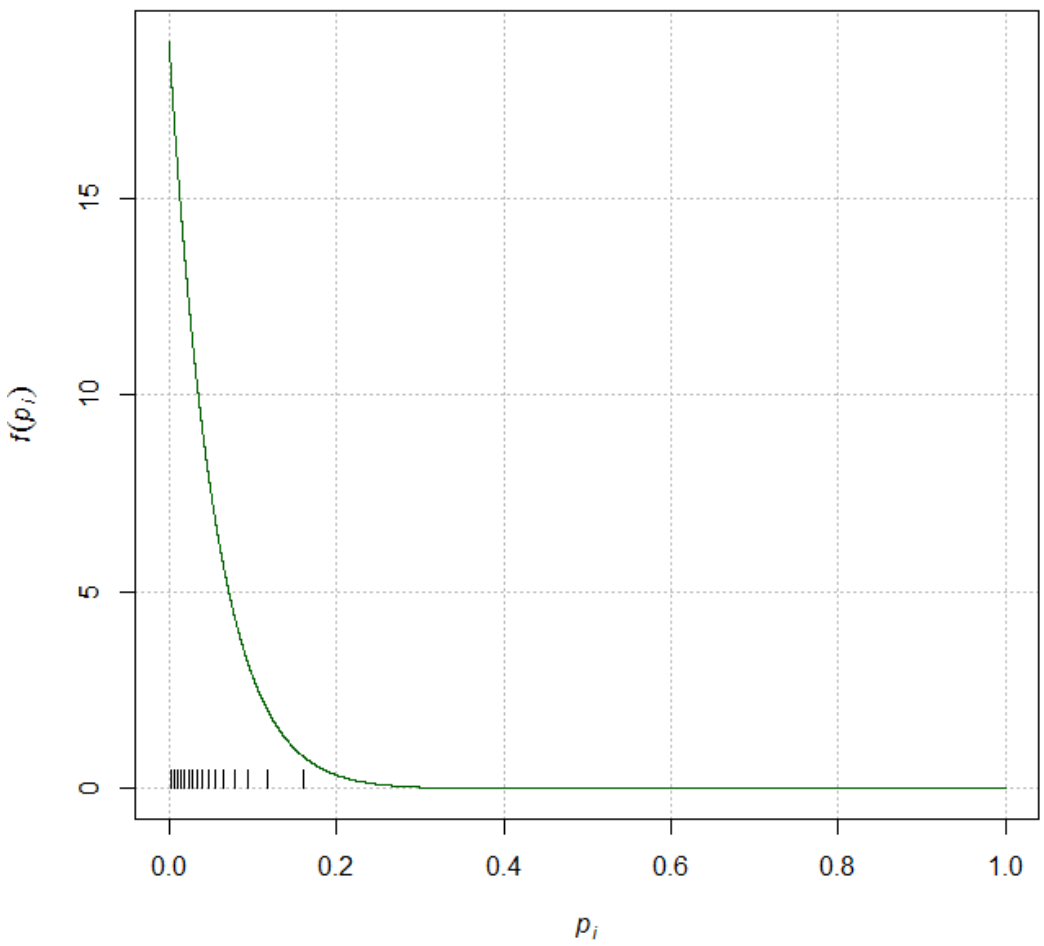

Figure 4.3. Output plot from `p1 <- beta.dist(p = 0.05, a = 1, b = NULL, grp.sz = 16, plot = TRUE)`. The numerical values for the expectations are denoted by the short vertical lines.

References

Altman, D. and Bland, J. (1994a) Diagnostic tests 1: sensitivity and specificity. *BMJ*, 308, 1552.

Altman, D. and Bland, J. (1994b) Diagnostic tests 2: Predictive values. *BMJ*, 309, 102.

American Red Cross (2013) Blood testing. URL http://www.redcrossblood.org/learn-about-blood/what-happens-donated-blood/blood-testing, retrieved April 22, 2013.

Barrodale, I. and Roberts, F. (1974) Solution of an over determined system of equations in the L_1norm.*Communications of the ACM,***17**, 319-320.

Berger, T., Mandell, J. and Subrahmanya, P. (2000) Maximally efficient two-stage screening. *Biometrics,***56**, 833-840.

Biggerstaff, B. (2008) Confidence intervals for the difference of two proportions estimated from pooled samples. *Journal of Agricultural, Biological, and Environmental Statistics*, 13, 478-496.

Bilder, C. and Tebbs, J. (2009) Bias, efficiency, and agreement for group-testing regression models. *Journal of Statistical Computation and Simulation,***79**, 67-80.

Bilder, C., Tebbs, J. and Chen, P. (2010) Informative retesting. *Journal of the American Statistical Association*, 105, 942-955.

Bilder, C. and Tebbs, J. (2012) Pooled-testing procedures for screening high volume clinical specimens in heterogeneous populations, *Statistics in Medicine,***31**, 27, 3261-3268.

Black, M., Bilder, C. and Tebbs, J. (2012) Group testing in heterogeneous populations by using halving algorithms. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61, 277-290.

Casella, G., Berger, L. (2002) *Statistical Inference, Second Edition.* Pacific Grove, CA: Wadsworth Group.

Centers for Disease Control and Prevention (2013) *Infertility Prevention Project*

*(IPP).*Retrieved April 22, 2013 from http://www.cdc.gov/std/infertility/ipa.htm.

Chen, C. and Swallow, W. (1990) Using group testing to estimate a proportion, and to test the binomial model. *Biometrics,***46**, 1035–1046.

Dantzig, G. (1963) *Linear programming and extensions.* Princeton, NJ: Princeton University Press.

Dorfman, R. (1943) The detection of defective members of large populations. *Annals of Mathematical Statistics*, 14, 436-440.

Dodd, R., Notari, E., and Stramer, S. (2002) Current prevalence and incidence of infectious disease markers and estimated window-period risk in the American Red Cross donor population. *Transfusion*, 42, 975-979.

Du, D. and Hwang, F. (2000) *Combinatorial group testing and its applications.* River Edge, NJ: Word Scientific Publishing Co. Pte. Ltd.

Feller, W. (1950) *An introduction probability theory and its applications*. New York: Wiley.

Feller, W. (1957) *An introduction probability theory and its applications, Second edition.* New York: Wiley.

Flynn, G. (1993) *The draft, 1940-1973.* Lawrence, KS: University Press of Kansas.

Graff, L. and Roeloffs, R. (1974) A group testing procedure in the presence of test error. *Journal of the American Statistical Association*, **69**, 159-163.

Gupta, D. and Malina, R. (1999) Group testing in presence of classification errors. *Statistics in Medicine*, 18, 1049-1068.

Hepworth, G. and Watson, R. (2009) Debiased estimation of proportions in group testing. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58, 105-121.

Hourfar, M., Jork, C., Schottstedt, V., Weber-Schehl, M., Brixner, V., Busch, M., Geusendam, G., Gubbe, K., Mahnhardt, C., Mayr-Wohlfart, U., Pichl, L., Roth, W., Schmidt, M., Seifried, E. and Wright, D. (2008) Experience of German Red Cross blood donor services with nucleic acid testing: results of screening more than 30 million blood donations for human immunodeficiency virus-1, hepatitis C virus, and hepatitis B virus. *Transfusion*, 48, 1558-1566.

Hughes-Oliver, J. (2006) Pooling experiments for blood screening and drug discovery. In *Screening: Methods for Experimentation in Industry, Drug Discovery, and Genetics*, (eds A. Dean and S. Lewis), New York: Springer.

Hwang, F. (1972) A method for detecting all defective members in a population by group testing. *Journal of the American Statistical Association,* **67**, 605-608.

Hwang, F. (1975) A generalized binomial group testing problem. *Journal of the American Statistical Association*,70, 923-926.

IDEXX (2010). ELISA Technical Guide. *IDEXX Laboratories Inc.*, 1-30.

Johnson, N., Kotz, S., and Wu, X. (1991) *Inspection Errors for Attributes in Quality Control.* New York: Chapman & Hall.

Junjiro, O. (1962) Distribution and moments of order statistics. In *Contributions to Order Statistics* (ed A. Sarhan), New York: John Wiley and Sons Inc.

Kacena, K., Quinn, S., Hartman, S., Quinn, T. and Gaydos, C. (1998a) Pooling of urine samples for screening for Neisseria gonorrhea by ligase chain reaction: accuracy and application. *Journal of Clinical Microbiology,***36**, 3624-8.

Kacena, K., Quinn, S., Howell, M., Madico, G., Quinn, T. and Gaydos, C. (1998b) Pooling urine samples for ligase chain reaction screening for genital Chlamydia Trachomatis infection in asymptomatic women. *Journal of Clinical Microbiology,***36**, 481-5.

Katona, G. (1973) Combinatorial search problem, in *A Survey of Combinatorial Theory*, Ed. J. N. Srivastava et al., Amsterdam: North Holland.

Kim, H., Hudgens, M., Dreyfuss, J., Westreich, D., and Pilcher, C. (2007) Comparison of group testing algorithms for case identification in the presence of test error. *Biometrics*, 63, 1152-1163.

Kim, H., and Hudgens, M. (2009) Three dimensional array based group testing algorithms. *Biometrics*, 65, 903-910.

Kline, R., Brothers, T., Brookmeyer, R., Zeger, S. and Quinn, T. (1989) Evaluation of HIV seroprevalence in population surveys using pooled sera. *Journal of Clinical Microbiology,***27**, 1449-1452.

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

Li, C. (1962) A sequential method for screening experimental variables. *Journal of the American Statistical Association,* **57**, 455-477.

Litvak, E., Tu, X., and Pagano, M. (1994) Screening for the presence of a disease by pooling sera samples. *Journal of the American Statistical Association*, 89, 424-434.

McMahan, C., Tebbs, J. and Bilder, C. (2012a) Informative Dorfman screening. *Biometrics*, 68, 287-296.

McMahan, C., Tebbs, J. and Bilder, C. (2012b) Two-Dimensional Informative Array Testing. *Biometrics*, 68, 793-804.

McMahan, C., Tebbs, J. and Bilder, C. (2013) Regression models for group testing data with pool dilution effects. *Biostatistics*, 14, 284-298

Mehta, S., Nguyen, V., Osorio, G., Little, S. and Smith, D. (2011) Evaluation of pooled rapid HIV antibody screening of patients admitted to a San Diego hospital. *Journal of Virological Methods*, 174, 94-98.

Mund, M., Sander, G., Potthoff, P., Schicht, H., and Matthias, K. (2008) Introduction of *Chlamydia trachomatis* screening for young women in Germany. *Journal der Deutschen Dermatologischen Gesellschaft*, 6, 1032-1037.

Peck, C. (2006) Going after BVD. *Beef*, 42, 34-44.

Phatarfod, R. and Sudbury, A. (1994) The use of a square array scheme in blood testing. *Statistics in Medicine*, 13, 2337-2343.

Pilcher, C., Fiscus, S., Nguyen, T., Foust, E., Wolf, L., Williams, D., Ashby, R., O'Dowd, J., McPherson, J., Stalzer, B., Hightow, L., Miller, W., Eron, J., Cohen, M., and Leone, P. (2005) Detection of acute infections during HIV testing in North Carolina. *New England Journal of Medicine*, 352, 1873-1883.

Quinn, T., Brookmeyer, R., Kline, R., Shepherd, M., Paranjape, R., Mehendale, S., Gadkari, D. and Bollinger, R. (2000) Feasibility of pooling sera for HIV-1 rival RNA to diagnose acute primary HIV-1 infection and estimate HIV incidence. *AIDS*, 14, 2751-2757.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2013.URL http://www.R-project.org.ISBN 3-900051-07-0.

Remlinger, K., Hughes-Oliver, J., Young, S., and Lam, R. (2006) Statistical design of pools using optimal coverage and minimal collision. *Technometrics*, 48, 133-143.

Smith, D., May, S., Perez-Santiago, J., Strain, M., Ignacio, C., Haubrich, R., Richman, D., Benson, C. and Little, S. (2009) The use of pooled viral load testing to identify antiretroviral treatment failure. *AIDS*, 23, 2151-2158.

Sobel, M. and Groll, P. (1959) Group testing to eliminate efficiently all defectives in a binomial sample. *The Bell System Technical Journal,***38**, 1179-1252.

Sobel, M. and Groll, P. (1966) Binomial group testing with an unknown proportion of defectives. *Technometrics*, **8**, 631-656.

Soroka, S., Granade, T., Phillips, S. and Parekh, B. (2003) The use of simple, rapid tests to detect antibodies to human immunodeficiency virus types 1 and 2 in pooled serum specimens. *Journal of Clinical Virology,***27**, 90-96.

Stramer, S. L., Glynn, S. A., Kleinman, S. H., Strong, D. M., Caglioti, S., Wright, D. J., Dodd, R. Y., and Busch, M. P. (2004) Detection of HIV-1 and HCV infections among antibody-negative blood donors by nucleic acid–amplification testing. *New England Journal of Medicine* 351, 760-768.

Sterrett, A. (1957) On the detection of defective members of large populations. *Annals of Mathematical Statistics*, 28, 1033-1036.

Tebbs, J. and Bilder, C. (2004) Confidence interval procedures for the probability of disease transmission in multiple-vector-transfer designs. *Journal of Agricultural, Biological, and Environmental Statistics*, 9, 75-90.

Tilghman, M., Guerena, D., Licea, A., Perez-Santiago, J., Richman, D., May, S. and Smith, D. (2011) Pooled nucleic acid testing to detect antiretroviral treatment failure in Mexico. *Journal of Acquired Immune Deficiency Syndrome*, 56, e70-e74.

Tu, X. Litvak, E. and Pagano, M. (1995) On the informativeness and accuracy of pooled testing in estimating prevalence of a rare disease: Application to HIV

screening. *Biometrika,***82**, 287-297.

Vansteelandt, S., Goetghebeur, E., and Verstraeten, T. (2000) Regression models for disease prevalence with diagnostic tests on pools of serum samples. *Biometrics*, 56, 1126-1133.

Watson, G. (1961) A study of the group screening method. *Technometics,* **3**, 371-388.

Xie, M. (2001) Regression analysis of group testing samples. *Statistics in Medicine*, 20, 1957-1969.

Appendix A: PMF for halving

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 1 and $\tilde{G}_{1,1} = 0$ or 1. In order to find the PMF for the number of tests with $S = 2$, let

$$\boldsymbol{E}_2 = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

be a matrix of possible testing errors, $\boldsymbol{P}_{2,I_{1,1}} = \left[P(\tilde{G}_{1,1} = 0), P(\tilde{G}_{1,1} = 1)\right]'$ be a vector of probabilities for the true statuses, and $\boldsymbol{T}_{2,I_{1,1}} = \left[1, 1 + I_{1,1}\right]'$ 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 $\boldsymbol{T}_{2,I_{1,1}}$ is $\boldsymbol{P}'_{2,I_{1,1}}\boldsymbol{E}_2$.

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

$$\boldsymbol{E}_3 = \left[[a,c,c,c]' \;||\; Diag([b,d,d,d]) \times \left(\boldsymbol{E}_2 \otimes \boldsymbol{E}_2\right)\right] = \begin{bmatrix} a & aab & abb & bab & bbb \\ c & acd & add & bcd & bdd \\ c & cad & cbd & dad & dbd \\ c & ccd & cdd & dcd & ddd \end{bmatrix},$$

$$\begin{aligned} \boldsymbol{P}_{3,I_{1,1}} &= \boldsymbol{P}_{2,I_{2,1}} \otimes \boldsymbol{P}_{2,I_{2,2}} \\ &= [P(\tilde{G}_{1,1} = 0), P(\tilde{G}_{1,1} = 1 \cap \tilde{G}_{2,1} = 0 \cap \tilde{G}_{2,2} = 1), \\ &\quad P(\tilde{G}_{1,1} = 1 \cap \tilde{G}_{2,1} = 1 \cap \tilde{G}_{2,2} = 0), P(\tilde{G}_{1,1} = 1 \cap \tilde{G}_{2,1} = 1 \cap \tilde{G}_{2,2} = 1)]', \end{aligned}$$

and

$$\boldsymbol{T}_{3,I_{1,1}} = \left[1, \boldsymbol{j}_2' \otimes \boldsymbol{T}_{2,I_{2,1}} + \boldsymbol{T}_{2,I_{2,2}} \otimes \boldsymbol{j}_2' + \boldsymbol{j}_4'\right]' = \left[1, 3, 3 + I_{2,1}, 3 + I_{2,2}, 3 + I_{1,1}\right]',$$

where || denotes vertical concatenation, $\boldsymbol{j}_m$ denotes a $m \times 1$ vector of 1's, and $\otimes$ denotes a Kronecker product. The corresponding probabilities for $\boldsymbol{T}_{3,I_{1,1}}$ can be found from $\boldsymbol{P}_{3,I_{1,1}}' \boldsymbol{E}_3$. After summing these probabilities over the same number of tests in $\boldsymbol{T}_{3,I_{1,1}}$, 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 $\boldsymbol{T}_{2,I_{S-1,j}} = \left[1, 1 + I_{S-1,j}\right]'$, $\boldsymbol{P}_{2,I_{S-1,j}} = \left[P(\tilde{G}_{S-1,j} = 0), P(\tilde{G}_{S-1,j} = 1)\right]'$ for $j = 1, \ldots, 2^{S-2}$, and $\boldsymbol{E}_2$ be the same as before. In reverse order from how the testing is actually done, we successively build new matrices

$$\boldsymbol{T}_{s,I_{S-s+1,j}} = \left[1, \boldsymbol{j}_k' \otimes \boldsymbol{T}_{s-1,I_{S-s+2,2j-1}} + \boldsymbol{T}_{s-1,I_{S-s+2,2j}} \otimes \boldsymbol{j}_k' + \boldsymbol{j}_{k^2}'\right]',$$
$$\boldsymbol{P}_{s,I_{S-s+1,j}} = \boldsymbol{P}_{s-1,I_{S-s+2,2j-1}} \otimes \boldsymbol{P}_{s-1,I_{S-s+2,2j}},$$

and

$$\boldsymbol{E}_s = \left[\left[a, c\boldsymbol{j}_m'\right]' \,||\, Diag([b, d\boldsymbol{j}_m']) \times \left(\boldsymbol{E}_{s-1} \otimes \boldsymbol{E}_{s-1}\right)\right]$$

for $s = 3, \ldots, S$ and $j = 1, \ldots, 2^{S-s}$, where m is 1 less than the number of rows in $\boldsymbol{E}_{s-1} \otimes \boldsymbol{E}_{s-1}$ and k is the number of rows in $\boldsymbol{T}_{s-1,I_{S-s+2,2j}}$. Our final resulting matrices will be $\boldsymbol{T}_{S,I_{1,1}}$, $\boldsymbol{P}_{S,I_{1,1}}$, and $\boldsymbol{E}_S$. The corresponding probabilities for $\boldsymbol{T}_{S,I_{1,1}}$ can be found from $\boldsymbol{P}_{S,I_{1,1}}' \boldsymbol{E}_S$. We sum these probabilities over the same number of tests in $\boldsymbol{T}_{S,I_{1,1}}$ 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 sub-group sizes. Large matrices result when S is not small (e.g., $S = 6$, $\boldsymbol{E}_6$ requires a 65,536 x 458,329 matrix) causing memory problems from using the matrix methods in R's base package (R

Development Core Team, 2013). However, this is not too limiting because S is usually small in practice.

Appendix B: Mean and variance for number of tests

This appendix shows how to derive the mean and variance for the number of tests used in a halving procedure. We initially assume there are no testing errors and three steps only. First, we rewrite $E(T \mid \boldsymbol{p}^{vec})$ and $E(T^2 \mid \boldsymbol{p}^{vec})$ for a group of size $I \equiv I_{1,1}$ as

$$E(T \mid \boldsymbol{p}^{vec}) = 1 + 2\left\{1 - \prod_{i=1}^{I_{1,1}}(1-p_i)\right\} + I_{2,1}\left\{1 - \prod_{i=1}^{I_{2,1}}(1-p_i)\right\} + I_{2,2}\left\{1 - \prod_{i=I_{2,1}+1}^{I_{1,1}}(1-p_i)\right\}$$

and

$$\begin{aligned} E(T^2 \mid \boldsymbol{p}^{vec}) = 1 &+ (6I_{2,1} + I_{2,1}^2)\left\{1 - \prod_{i=1}^{I_{2,1}}(1-p_i)\right\} + (6I_{2,2} + I_{2,2}^2)\left\{1 - \prod_{i=I_{2,1}+1}^{I_{1,1}}(1-p_i)\right\} + \\ &2I_{2,1}I_{2,2}\left\{1 - \prod_{i=1}^{I_{2,1}}(1-p_i) - \prod_{i=I_{2,1}+1}^{I_{1,1}}(1-p_i) + \prod_{i=1}^{I_{1,1}}(1-p_i)\right\} + \\ &8\left\{1 - \prod_{i=1}^{I_{1,1}}(1-p_i)\right\}. \end{aligned}$$

These expressions are used to find

$$\begin{aligned} Var(T \mid \boldsymbol{p}^{vec}) = &4\prod_{i=1}^{I_{1,1}}(1-p_i)\left\{1 - \prod_{i=1}^{I_{1,1}}(1-p_i)\right\} + 4I_{2,1}\prod_{i=1}^{I_{1,1}}(1-p_i)\left\{1 - \prod_{i=1}^{I_{2,1}}(1-p_i)\right\} + \\ &4I_{2,2}\prod_{i=1}^{I_{1,1}}(1-p_i)\left\{1 - \prod_{i=I_{2,1}+1}^{I_{1,1}}(1-p_i)\right\} + I_{2,1}^2\prod_{i=1}^{I_{2,1}}(1-p_i)\left\{1 - \prod_{i=1}^{I_{2,1}}(1-p_i)\right\} + \\ &I_{2,2}^2\prod_{i=I_{2,1}+1}^{I_{1,1}}(1-p_i)\left\{1 - \prod_{i=I_{2,1}+1}^{I_{1,1}}(1-p_i)\right\}. \end{aligned}$$

When all of the individual probabilities are equal, we obtain

$$E(T \mid \boldsymbol{p}) = 1 + 2\{1-(1-p)^I\} + I_{2,1}\{1-(1-p)^{I_{2,1}}\} + I_{2,2}\{1-(1-p)^{I_{2,2}}\}$$

and

$$\begin{aligned} Var(T \mid \boldsymbol{p}) = &\ 4(1-p)^I\{1-(1-p)^I\} + 4I_{2,1}(1-p)^{I_{2,2}}(1-p)^{I_{2,1}}\{1-(1-p)^{I_{2,1}}\} + \\ &4I_{2,2}(1-p)^{I_{2,1}}(1-p)^{I_{2,2}}\{1-(1-p)^{I_{2,2}}\} + \\ &I_{2,1}^2(1-p)^{I_{2,1}}\{1-(1-p)^{I_{2,1}}\} + I_{2,2}^2(1-p)^{I_{2,2}}\{1-(1-p)^{I_{2,2}}\}, \end{aligned}$$

where $\boldsymbol{p}$ is a $I\times 1$ vector of equal individual probabilities (i.e., $p_i = p$ for i =1, ..., I).

When p_i are independent random variables with $E(p_i) = p$ for $i = 1, \ldots, I$, we obtain

$$\begin{aligned} E(T) &= E\{E(T \mid \boldsymbol{p}^{vec})\} = 1 + 2\{1-(1-p)^{I}\} + \\ &\quad I_{2,1}\{1-(1-p)^{I_{2,1}}\} + I_{2,2}\{1-(1-p)^{I_{2,2}}\} \\ &= E(T \mid \boldsymbol{p}). \end{aligned}$$

Therefore, the expected number of tests is the same as when the individual probabilities are all equal (i.e., a homogeneous population).

To find the variance of the number of tests (still with p_i as independent random variables and $E(p_i) = p$ for $i = 1, \ldots, I$), we use the expression $Var(T) = E\{Var(T \mid \boldsymbol{p}^{vec})\} + Var\{E(T \mid \boldsymbol{p}^{vec})\}$. The first term is

$$\begin{aligned} E\{Var(T \mid \boldsymbol{p}^{vec})\} = {} & 4\left[(1-p)^{I} - Var\left\{\prod_{i=1}^{I_{1,1}}(1-p_i)\right\} - (1-p)^{2I}\right] + \\ & 4I_{2,1}(1-p)^{I_{2,2}}\left[(1-p)^{I_{2,1}} - Var\left\{\prod_{i=1}^{I_{2,1}}(1-p_i)\right\} - (1-p)^{2I_{2,1}}\right] + \\ & 4I_{2,2}(1-p)^{I_{2,1}}\left[(1-p)^{I_{2,2}} - Var\left\{\prod_{i=I_{2,1}+1}^{I_{1,1}}(1-p_i)\right\} - (1-p)^{2I_{2,2}}\right] + \\ & I_{2,1}^2\left[(1-p)^{I_{2,1}} - Var\left\{\prod_{i=1}^{I_{2,1}}(1-p_i)\right\} - (1-p)^{2I_{2,1}}\right] + \\ & I_{2,2}^2\left[(1-p)^{I_{2,2}} - Var\left\{\prod_{i=I_{2,1}+1}^{I_{1,1}}(1-p_i)\right\} - (1-p)^{2I_{2,2}}\right], \end{aligned}$$

and the second term is

$$Var\{E(T \mid \boldsymbol{p}^{vec})\} = 4Var\left\{\prod_{i=1}^{I_{1,1}}(1-p_i)\right\} + I_{2,1}^2 Var\left\{\prod_{i=1}^{I_{2,1}}(1-p_i)\right\} + I_{2,2}^2 Var\left\{\prod_{i=I_{2,1}+1}^{I_{1,1}}(1-p_i)\right\} +$$
$$4I_{2,1}Cov\left\{\prod_{i=1}^{I_{1,1}}(1-p_i), \prod_{i=1}^{I_{2,1}}(1-p_i)\right\} + 4I_{2,2}Cov\left\{\prod_{i=1}^{I_{1,1}}(1-p_i), \prod_{i=I_{2,1}+1}^{I_{1,1}}(1-p_i)\right\} +$$
$$2I_{2,1}I_{2,2}Cov\left\{\prod_{i=1}^{I_{2,1}}(1-p_i), \prod_{i=I_{2,1}+1}^{I_{1,1}}(1-p_i)\right\}.$$

In order to simplify the above expressions, one can show

$$Var\left\{\prod_{i=1}^{I_{1,1}}(1-p_i)\right\} = \{Var(p_i) + (1-p)^2\}^I - (1-p)^{2I},$$
$$Cov\left\{\prod_{i=1}^{I_{2,1}}(1-p_i), \prod_{i=I_{2,1}+1}^{I_{1,1}}(1-p_i)\right\} = 0,$$
$$Cov\left\{\prod_{i=1}^{I_{1,1}}(1-p_i), \prod_{i=1}^{I_{2,1}}(1-p_i)\right\} = (1-p)^{I_{2,2}} Var\left\{\prod_{i=1}^{I_{2,1}}(1-p_i)\right\}$$
$$= (1-p)^{I_{2,2}}[\{Var(p_i) + (1-p)^2\}^{I_{2,1}} - (1-p)^{2I_{2,1}}]$$
$$Cov\left\{\prod_{i=1}^{I_{1,1}}(1-p_i), \prod_{i=I_{2,1}+1}^{I_{1,1}}(1-p_i)\right\} = (1-p)^{I_{2,1}} Var\left\{\prod_{i=I_{2,1}+1}^{I_{1,1}}(1-p_i)\right\}$$
$$= (1-p)^{I_{2,1}}[\{Var(p_i) + (1-p)^2\}^{I_{2,2}} - (1-p)^{2I_{2,2}}],$$

and the variances for $\prod_{i=1}^{I_{2,1}}(1-p_i)$and $\prod_{i=I_{2,1}+1}^{I_{1,1}}(1-p_i)$ are similar. Substituting into $E\{Var(T \mid \boldsymbol{p}^{vec})\}$ and $Var\{E(T \mid \boldsymbol{p}^{vec})\}$ gives us

$$E\{Var(T \mid \boldsymbol{p}^{vec})\} = 4[(1-p)^I - \{Var(p_i) + (1-p)^2\}^I] +$$
$$4I_{2,1}(1-p)^{I_{2,2}}[(1-p)^{I_{2,1}} - \{Var(p_i) + (1-p)^2\}^{I_{2,1}}] +$$
$$4I_{2,2}(1-p)^{I_{2,1}}[(1-p)^{I_{2,2}} - \{Var(p_i) + (1-p)^2\}^{I_{2,2}}] +$$
$$I_{2,1}^2[(1-p)^{I_{2,1}} - \{Var(p_i) + (1-p)^2\}^{I_{2,1}}] +$$
$$I_{2,2}^2[(1-p)^{I_{2,2}} - \{Var(p_i) + (1-p)^2\}^{I_{2,2}}]$$

and

$$Var\{E(T \mid \boldsymbol{p}^{vec})\} = 4[\{Var(p_i) + (1-p)^2\}^I - (1-p)^{2I}] +$$
$$I_{2,1}^2[\{Var(p_i) + (1-p)^2\}^{I_{2,1}} - (1-p)^{2I_{2,1}}] +$$
$$I_{2,2}^2[\{Var(p_i) + (1-p)^2\}^{I_{2,2}} - (1-p)^{2I_{2,2}}] +$$

$$4I_{2,1}(1-p)^{I_{2,2}}[\{Var(p_i)+(1-p)^2\}^{I_{2,1}}-(1-p)^{2I_{2,1}}]+$$
$$4I_{2,2}(1-p)^{I_{2,1}}[\{Var(p_i)+(1-p)^2\}^{I_{2,2}}-(1-p)^{2I_{2,2}}].$$

Finally, we calculate

$$\begin{aligned}Var(T) &= 4(1-p)^I\{1-(1-p)^I\}+4I_{2,1}(1-p)^{I_{2,2}}(1-p)^{I_{2,1}}\{1-(1-p)^{I_{2,1}}\}+\\ &\quad 4I_{2,2}(1-p)^{I_{2,1}}(1-p)^{I_{2,2}}\{1-(1-p)^{I_{2,2}}\}+\\ &\quad I_{2,1}^2(1-p)^{I_{2,1}}\{1-(1-p)^{I_{2,1}}\}+I_{2,2}^2(1-p)^{I_{2,2}}\{1-(1-p)^{I_{2,2}}\}\\ &= Var(T\mid \boldsymbol{p}).\end{aligned}$$

Similar to the results with the first moment, the variance for the number of tests is the same as when the individual probabilities are all equal.

When more steps are used or when testing error is present, our conclusion is the same due to the independence of the p_i random variables. Presenting the general argument only adds a considerable number of terms that greatly increase the algebraic complexity. Therefore, we have chosen not to show the most general case here.

Appendix C: Expected number of tests when $p = 0.10$, 0.01, and 0.005

This appendix includes the figures for the $p = 0.10$, 0.01, and 0.005 cases discussed in Section 2.4.

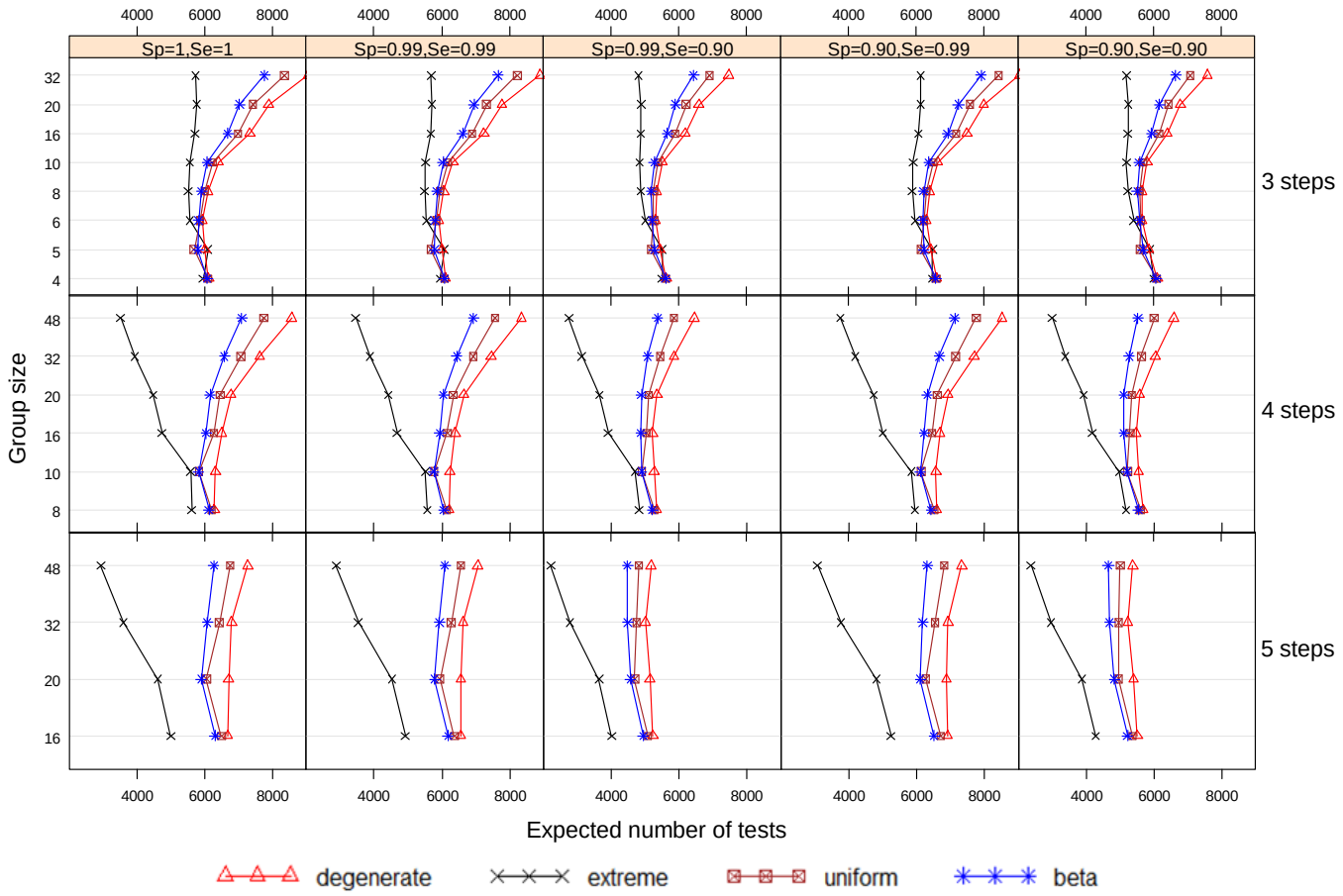

Figure C.1. Mean number of tests for $p = 0.1$. Each row of panels corresponds to the number of halving steps. Each column of panels corresponds to specificity and sensitivity settings.

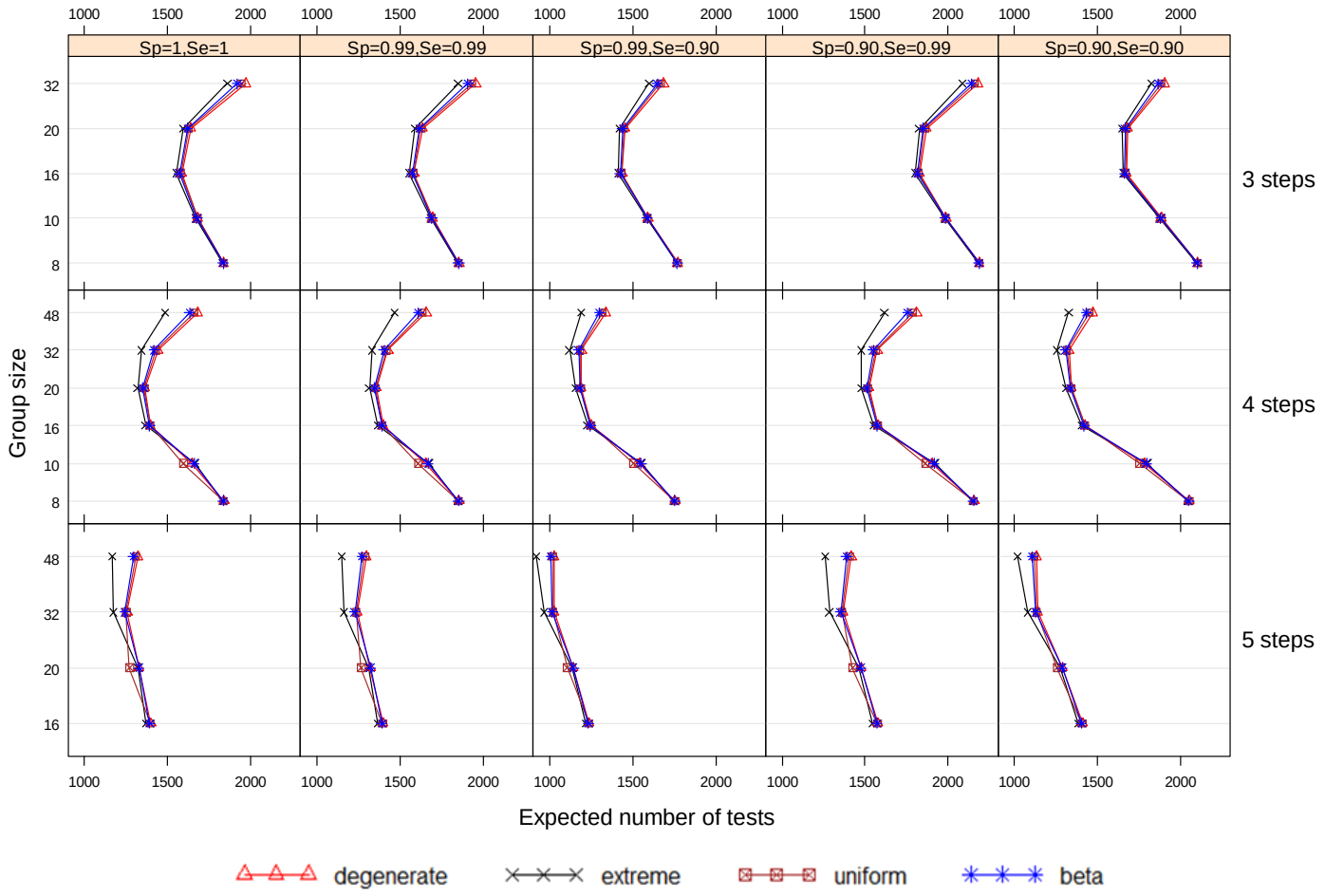

Figure C.2. Mean number of tests for $p = 0.01$. Each row of panels corresponds to the number of halving steps. Each column of panels corresponds to specificity and sensitivity settings.

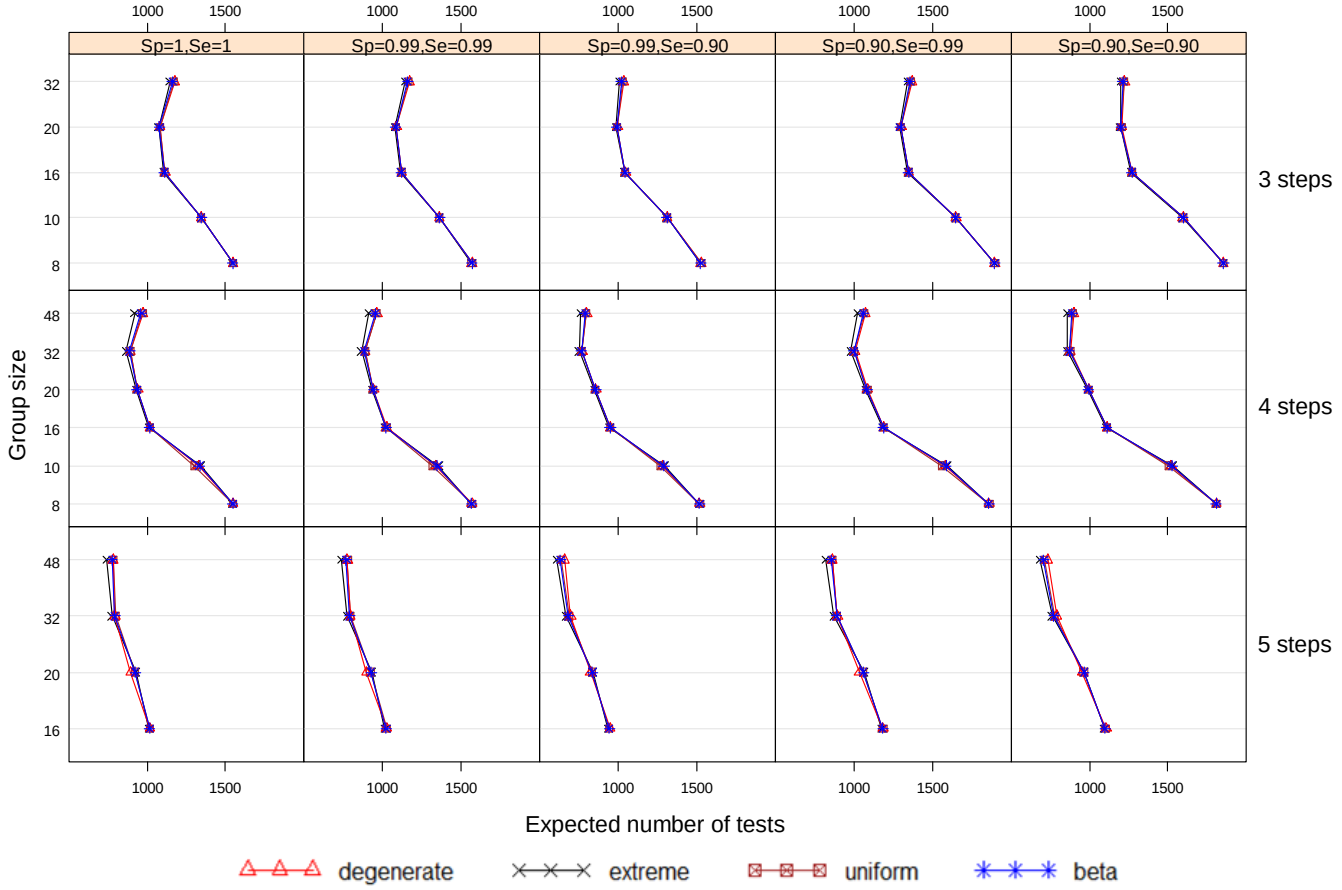

Figure C.3. Mean number of tests for $p = 0.005$. Each row of panels corresponds to the number of halving steps. Each column of panels corresponds to specificity and sensitivity settings.

Appendix D: Chlamydia and gonorrhea expected tests and accuracy results using halving

This appendix includes the additional results discussed in Section 2.5. Table D.1gives the sensitivities and specificities for each gender and specimen combination for gonorrhea. Tables D.2 to D.6show results for groups of size 8, 12, 16, 24, and 32. Within each of these tables, the following are displayed:

1. Number of tests,
2. Pooling sensitivity (PS_e) – Proportion of true positives that test positive through halving,
3. Pooling specificity (PS_p) – Proportion of true negatives that test negative through halving,
4. Pooling positive predictive value (PPPV) – Proportion of individuals that test positive through halving who are truly positive, and
5. Pooling negative predictive value (PNPV) – Proportion of individuals that test negative through halving who are truly negative.

We average the above measures across ten implementations of halving to account for simulation variability that arises due to imperfect diagnostic tests. There are a few instances where unordered halving has slightly fewer tests than ordered halving. These are due to the simulation variability.

		S_e	S_p	Observed prevalence	Number of individuals screened
Female	Urine	0.849	0.98	2.0%	2,679
	Swab	0.966	0.98	1.0%	19,450
Male	Urine	0.970	0.96	1.6%	3,852
	Swab	0.985	0.96	4.9%	4,086

Table D.1. Information about gonorrhea testing.

		Chlamydia					Gonorrhea				
		Tests	PS_e	PS_p	PPPV	PNPV	Tests	PS_e	PS_p	PPPV	PNPV
Urine Female	Dorfman	1,557.7	0.65	0.98	0.82	0.96	708.1	0.73	1.00	0.84	0.99
	3-step unordered	1,238.9	0.51	0.99	0.89	0.95	579.9	0.62	1.00	0.95	0.99
	3-step ordered	1,236.3	0.53	0.99	0.89	0.95	580.0	0.62	1.00	0.95	0.99
	4-step unordered	1,228.5	0.44	1.00	0.96	0.94	564.6	0.52	1.00	0.98	0.99
	4-step ordered	1,207.8	0.43	1.00	0.96	0.94	568.4	0.52	1.00	0.96	0.99
Urine Male	Dorfman	2,415.6	0.87	0.98	0.78	0.99	1,050.8	0.93	0.99	0.74	1.00
	3-step unordered	2,009.0	0.81	0.99	0.87	0.98	862.2	0.91	1.00	0.91	1.00
	3-step ordered	1,971.8	0.81	0.99	0.87	0.98	862.6	0.91	1.00	0.91	1.00
	4-step unordered	1,996.4	0.75	1.00	0.95	0.98	857.2	0.87	1.00	0.97	1.00
	4-step ordered	1,965.2	0.74	1.00	0.95	0.98	859.2	0.87	1.00	0.95	1.00
Swab Female	Dorfman	9,492.6	0.86	0.99	0.80	0.99	4,164.8	0.94	1.00	0.85	1.00
	3-step unordered	7,833.0	0.80	0.99	0.90	0.99	3,560.0	0.90	1.00	0.94	1.00
	3-step ordered	7,705.8	0.80	0.99	0.90	0.99	3,548.4	0.91	1.00	0.94	1.00
	4-step unordered	7,804.2	0.74	1.00	0.95	0.98	3,550.4	0.88	1.00	0.98	1.00
	4-step ordered	7,731.2	0.74	1.00	0.96	0.98	3,545.0	0.88	1.00	0.98	1.00
Swab Male	Dorfman	2,984.6	0.87	0.97	0.83	0.98	1,915.0	0.97	0.99	0.79	1.00
	3-step unordered	2,633.8	0.79	0.99	0.89	0.97	1,600.8	0.95	0.99	0.89	1.00
	3-step ordered	2,534.6	0.79	0.99	0.90	0.97	1,572.4	0.96	0.99	0.91	1.00
	4-step unordered	2,721.6	0.75	0.99	0.95	0.96	1,609.4	0.94	1.00	0.96	1.00
	4-step ordered	2,639.0	0.74	0.99	0.95	0.96	1,596.2	0.94	1.00	0.96	1.00

Table D.2. Additional results for the 2005 NPHL data with a group size of 8.

		Chlamydia					Gonorrhea				
		Tests	PS_e	PS_p	PPPV	PNPV	Tests	PS_e	PS_p	PPPV	PNPV
Urine Female	Dorfman	1,735.1	0.66	0.98	0.78	0.96	749.6	0.74	1.00	0.81	0.99
	3-step unordered	1,297.5	0.52	0.99	0.84	0.95	530.0	0.62	1.00	0.90	0.99
	3-step ordered	1,229.1	0.50	0.99	0.86	0.95	528.2	0.60	1.00	0.87	0.99
	4-step unordered	1,136.2	0.42	1.00	0.93	0.94	486.0	0.53	1.00	0.96	0.99
	4-step ordered	1,104.8	0.42	1.00	0.93	0.94	480.7	0.52	1.00	0.95	0.99
Urine Male	Dorfman	2,701.8	0.86	0.97	0.72	0.99	1091.4	0.95	0.99	0.66	1.00
	3-step unordered	2,109.2	0.80	0.98	0.81	0.98	787.2	0.92	1.00	0.82	1.00
	3-step ordered	2,076.8	0.80	0.98	0.82	0.98	792.6	0.93	1.00	0.85	1.00
	4-step unordered	1,908.0	0.74	0.99	0.91	0.98	736.0	0.89	1.00	0.90	1.00
	4-step ordered	1,884.2	0.73	0.99	0.91	0.98	726.5	0.87	1.00	0.93	1.00
Swab Female	Dorfman	10,791.9	0.86	0.98	0.75	0.99	3986.8	0.93	1.00	0.79	1.00
	3-step unordered	8,222.9	0.80	0.99	0.85	0.99	3058.1	0.91	1.00	0.91	1.00
	3-step ordered	8,007.2	0.79	0.99	0.85	0.99	3026.9	0.89	1.00	0.90	1.00
	4-step unordered	7,443.4	0.74	1.00	0.92	0.98	2874.3	0.88	1.00	0.96	1.00
	4-step ordered	7,242.8	0.73	1.00	0.93	0.98	2859.5	0.88	1.00	0.97	1.00
Swab Male	Dorfman	3,357.8	0.85	0.96	0.78	0.98	2219.0	0.97	0.98	0.75	1.00
	3-step unordered	2,840.0	0.79	0.98	0.85	0.97	1702.2	0.96	0.99	0.85	1.00
	3-step ordered	2,680.4	0.79	0.98	0.86	0.97	1675.8	0.96	0.99	0.86	1.00
	4-step unordered	2,666.1	0.74	0.99	0.92	0.96	1569.6	0.94	1.00	0.93	1.00
	4-step ordered	2,546.3	0.74	0.99	0.93	0.96	1526.1	0.93	1.00	0.93	1.00

Table D.3. Additional results for the 2005 NPHL data with a group size of 12.

		Chlamydia					Gonorrhea				
		Tests	PS_e	PS_p	PPPV	PNPV	Tests	PS_e	PS_p	PPPV	PNPV
Urine Female	Dorfman	1,925.6	0.65	0.98	0.76	0.96	816.6	0.72	0.99	0.74	0.99
	3-step unordered	1,405.6	0.54	0.99	0.81	0.95	540.5	0.62	1.00	0.84	0.99
	3-step ordered	1,352.0	0.53	0.99	0.83	0.95	529.6	0.60	1.00	0.88	0.99
	4-step unordered	1,146.4	0.43	0.99	0.90	0.94	452.9	0.53	1.00	0.93	0.99
	4-step ordered	1,102.5	0.42	0.99	0.90	0.94	457.0	0.56	1.00	0.93	0.99
	5-step unordered	1,132.2	0.35	1.00	0.95	0.93	434.4	0.38	1.00	0.98	0.99
	5-step ordered	1,100.8	0.35	1.00	0.95	0.93	443.0	0.45	1.00	0.98	0.99
Urine Male	Dorfman	3,037.4	0.86	0.96	0.68	0.99	1152.2	0.94	0.99	0.65	1.00
	3-step unordered	2,338.2	0.80	0.98	0.78	0.98	796.6	0.92	1.00	0.80	1.00
	3-step ordered	2,253.2	0.81	0.98	0.78	0.98	798.4	0.91	1.00	0.81	1.00
	4-step unordered	2,000.0	0.76	0.99	0.88	0.98	690.2	0.89	1.00	0.91	1.00
	4-step ordered	1,930.6	0.75	0.99	0.88	0.98	689.8	0.89	1.00	0.91	1.00
	5-step unordered	1,992.9	0.69	1.00	0.95	0.97	687.7	0.85	1.00	0.94	1.00
	5-step ordered	1,946.8	0.70	1.00	0.95	0.97	692.4	0.86	1.00	0.96	1.00
Swab Female	Dorfman	12,341.1	0.86	0.98	0.70	0.99	4237.2	0.93	1.00	0.75	1.00
	3-step unordered	9,035.2	0.80	0.99	0.80	0.99	2938.6	0.89	1.00	0.89	1.00
	3-step ordered	8,759.5	0.80	0.99	0.81	0.99	2917.3	0.90	1.00	0.87	1.00
	4-step unordered	7,569.4	0.74	0.99	0.89	0.98	2612.2	0.87	1.00	0.94	1.00
	4-step ordered	7,404.9	0.75	0.99	0.90	0.98	2580.0	0.86	1.00	0.94	1.00
	5-step unordered	7,533.7	0.67	1.00	0.96	0.98	2591.4	0.83	1.00	0.98	1.00
	5-step ordered	7,387.7	0.69	1.00	0.96	0.98	2572.4	0.84	1.00	0.98	1.00
Swab Male	Dorfman	3,568.0	0.86	0.96	0.76	0.98	2542.4	0.97	0.98	0.71	1.00
	3-step unordered	2,996.4	0.79	0.97	0.82	0.97	1895.4	0.96	0.99	0.80	1.00
	3-step ordered	2,815.6	0.80	0.98	0.84	0.97	1794.6	0.96	0.99	0.82	1.00
	4-step unordered	2,638.4	0.72	0.99	0.89	0.96	1616.6	0.95	0.99	0.90	1.00
	4-step ordered	2,495.4	0.73	0.99	0.90	0.96	1545.4	0.94	0.99	0.91	1.00
	5-step unordered	2,702.4	0.67	1.00	0.95	0.95	1627.2	0.94	1.00	0.96	1.00
	5-step ordered	2,579.2	0.67	1.00	0.96	0.95	1583.8	0.94	1.00	0.96	1.00

Table D.4. Additional results for the 2005 NPHL data with a group size of 16.

		Chlamydia					Gonorrhea				
		Tests	PS_e	PS_p	PPPV	PNPV	Tests	PS_e	PS_p	PPPV	PNPV
Urine Female	Dorfman	2,114.5	0.66	0.97	0.72	0.96	1,018.3	0.70	0.99	0.68	0.99
	3-step unordered	1,540.3	0.55	0.98	0.78	0.95	597.9	0.63	1.00	0.83	0.99
	3-step ordered	1,502.8	0.54	0.98	0.80	0.95	598.5	0.59	1.00	0.81	0.99
	4-step unordered	1,161.3	0.43	0.99	0.84	0.94	434.6	0.49	1.00	0.90	0.99
	4-step ordered	1,115.4	0.43	0.99	0.86	0.94	436.0	0.49	1.00	0.92	0.99
	5-step unordered	1,018.4	0.33	1.00	0.92	0.93	401.9	0.45	1.00	0.96	0.99
	5-step ordered	1,019.7	0.35	1.00	0.91	0.93	397.6	0.43	1.00	0.96	0.99
Urine Male	Dorfman	3,277.4	0.87	0.96	0.67	0.99	1,411.4	0.94	0.99	0.55	1.00
	3-step unordered	2,598.4	0.81	0.97	0.74	0.98	907.0	0.91	0.99	0.69	1.00
	3-step ordered	2,519.2	0.80	0.97	0.74	0.98	902.8	0.90	0.99	0.71	1.00
	4-step unordered	2,092.5	0.75	0.99	0.82	0.98	701.2	0.87	1.00	0.83	1.00
	4-step ordered	2,022.0	0.76	0.99	0.82	0.98	698.2	0.89	1.00	0.84	1.00
	5-step unordered	1,935.3	0.70	0.99	0.91	0.97	650.3	0.86	1.00	0.92	1.00
	5-step ordered	1,857.6	0.71	0.99	0.91	0.97	645.9	0.89	1.00	0.94	1.00
Swab Female	Dorfman	14,481.7	0.86	0.97	0.65	0.99	5,000.8	0.93	1.00	0.68	1.00
	3-step unordered	10,448.6	0.79	0.98	0.74	0.99	3,166.1	0.90	1.00	0.82	1.00
	3-step ordered	9,957.0	0.80	0.98	0.76	0.99	3,093.4	0.90	1.00	0.82	1.00
	4-step unordered	8,115.7	0.74	0.99	0.85	0.98	2,488.5	0.88	1.00	0.92	1.00
	4-step ordered	7,745.4	0.74	0.99	0.86	0.98	2,425.0	0.86	1.00	0.91	1.00
	5-step unordered	7,368.4	0.69	1.00	0.92	0.98	2,315.5	0.85	1.00	0.96	1.00
	5-step ordered	7,107.1	0.69	1.00	0.93	0.98	2,278.5	0.85	1.00	0.97	1.00
Swab Male	Dorfman	3,819.0	0.86	0.96	0.74	0.98	3,058.2	0.97	0.97	0.64	1.00
	3-step unordered	3,325.0	0.80	0.97	0.78	0.97	2,252.4	0.96	0.98	0.74	1.00
	3-step ordered	2,998.6	0.79	0.97	0.80	0.97	2,061.6	0.96	0.98	0.77	1.00
	4-step unordered	2,832.2	0.74	0.98	0.86	0.96	1,757.8	0.93	0.99	0.85	1.00
	4-step ordered	2,607.4	0.75	0.98	0.86	0.96	1,608.6	0.94	0.99	0.88	1.00
	5-step unordered	2,616.2	0.68	0.99	0.92	0.95	1,618.3	0.93	1.00	0.92	1.00
	5-step ordered	2,471.8	0.69	0.99	0.93	0.95	1,514.6	0.93	1.00	0.93	1.00

Table D.5. Additional results for the 2005 NPHL data with a group size of 24.

		Chlamydia					Gonorrhea				
		Tests	PS_e	PS_p	PPPV	PNPV	Tests	PS_e	PS_p	PPPV	PNPV
Urine Female	Dorfman	2,186.5	0.64	0.97	0.70	0.96	1,158.4	0.70	0.99	0.67	0.99
	3-step unordered	1,632.1	0.52	0.98	0.74	0.95	669.5	0.58	1.00	0.76	0.99
	3-step ordered	1,541.3	0.54	0.98	0.78	0.95	637.2	0.60	1.00	0.76	0.99
	4-step unordered	1,211.9	0.42	0.99	0.80	0.94	446.5	0.48	1.00	0.86	0.99
	4-step ordered	1,085.4	0.41	0.99	0.84	0.94	427.2	0.48	1.00	0.88	0.99
	5-step unordered	1,007.1	0.34	1.00	0.90	0.93	376.0	0.42	1.00	0.96	0.99
	5-step ordered	961.3	0.35	1.00	0.90	0.93	376.5	0.43	1.00	0.94	0.99
Urine Male	Dorfman	3,506.6	0.87	0.96	0.64	0.99	1,723.4	0.94	0.98	0.48	1.00
	3-step unordered	2,882.0	0.80	0.97	0.70	0.98	1,078.0	0.92	0.99	0.65	1.00
	3-step ordered	2,762.4	0.80	0.97	0.70	0.98	1,058.2	0.93	0.99	0.65	1.00
	4-step unordered	2,233.8	0.74	0.98	0.79	0.98	771.6	0.89	1.00	0.77	1.00
	4-step ordered	2,236.2	0.75	0.98	0.78	0.98	773.2	0.89	1.00	0.79	1.00
	5-step unordered	1,930.7	0.69	0.99	0.86	0.97	657.4	0.87	1.00	0.90	1.00
	5-step ordered	1,909.9	0.70	0.99	0.88	0.97	665.8	0.89	1.00	0.90	1.00
Swab Female	Dorfman	15,691.0	0.86	0.97	0.63	0.99	5,773.2	0.93	0.99	0.64	1.00
	3-step unordered	11,711.6	0.80	0.98	0.71	0.99	3,564.5	0.89	1.00	0.77	1.00
	3-step ordered	11,124.1	0.80	0.98	0.72	0.99	3,414.6	0.90	1.00	0.79	1.00
	4-step unordered	8,771.6	0.74	0.99	0.80	0.98	2,566.2	0.87	1.00	0.88	1.00
	4-step ordered	8,173.4	0.73	0.99	0.82	0.98	2,499.2	0.88	1.00	0.89	1.00
	5-step unordered	7,378.6	0.68	1.00	0.90	0.98	2,240.7	0.84	1.00	0.94	1.00
	5-step ordered	7,103.1	0.68	1.00	0.90	0.98	2,181.0	0.83	1.00	0.95	1.00
Swab Male	Dorfman	3,802.6	0.86	0.96	0.75	0.98	3,390.8	0.97	0.97	0.61	1.00
	3-step unordered	3,427.5	0.79	0.96	0.77	0.97	2,555.5	0.95	0.98	0.70	1.00
	3-step ordered	3,165.1	0.79	0.97	0.77	0.97	2,222.7	0.95	0.98	0.75	1.00
	4-step unordered	2,939.8	0.74	0.98	0.83	0.96	1,939.4	0.94	0.99	0.79	1.00
	4-step ordered	2,631.3	0.73	0.98	0.84	0.96	1,706.2	0.94	0.99	0.84	1.00
	5-step unordered	2,605.9	0.69	0.99	0.90	0.95	1,675.8	0.93	0.99	0.88	1.00
	5-step ordered	2,403.6	0.67	0.99	0.90	0.95	1,512.4	0.93	1.00	0.91	1.00

Table D.6. Additional results for the 2005 NPHL data with a group size of 32.

Appendix E: Hierarchical group testing examples for Sections 3.2 and 3.3

Figure E.1provides an example of a possible five-stage hierarchical group testing procedure. There are many other possible hierarchical group testing procedures, and this is not meant to be an optimal configuration.

At stage 1 within Figure, we have $I_{11} = 10$ individuals. If the initial group tests positive, $c_2 = m_{11} = 3$ subgroups are formed with $I_{21} = 4$, $I_{22} = 1$, and $I_{23} = 5$ individuals within each. Groups that test positive at stage 2 are further split into $m_{21} = 2$, $m_{22} = 0$ (because individual testing has already occurred), and $m_{23} = 2$ sub-groups. This leads to at most $c_3 = 4$ sub-groups that are tested with sizes $I_{31} = 3$, $I_{32} = 1$, $I_{33} = 3$, and $I_{34} = 2$. Groups that test positive at stage 3 are split into $m_{31} = 2$, $m_{32} = 0$, $m_{33} = 2$, and $m_{34} = 2$ sub-groups. This leads to at most $c_4 = 6$ sub-groups of size $I_{41} = 2$, $I_{42} = 1$, $I_{43} = 2$, $I_{44} = 1$, $I_{45} = 1$, and $I_{46} = 1$. Groups that test positive at stage 4 are split into $m_{41} = 2$, $m_{42} = 0$, $m_{43} = 2$, $m_{44} = 0$, $m_{45} = 0$, and $m_{46} = 0$ sub-groups. This leads to at most $c_5 = 4$ sub-groups of size $I_{51} = 1$, $I_{52} = 1$, $I_{53} = 1$, and $I_{54} = 1$.

Figure E.2provides an example of a CRC for a case where $I = 18$. To find this configuration, we used individual probabilities that were the expected values of order statistics from a beta distribution with $p = 0.05$ and $\alpha = 0.10$, which produces the set of $E(p_{(i)})$s (0.0029, 0.0060, 0.0093, 0.0127, 0.0164, 0.0204, 0.0293, 0.0344, 0.0400, 0.0463, 0.0534, 0.0616, 0.0714, 0.0835, 0.0993, 0.1224, 0.1662)(see Section 3.3).

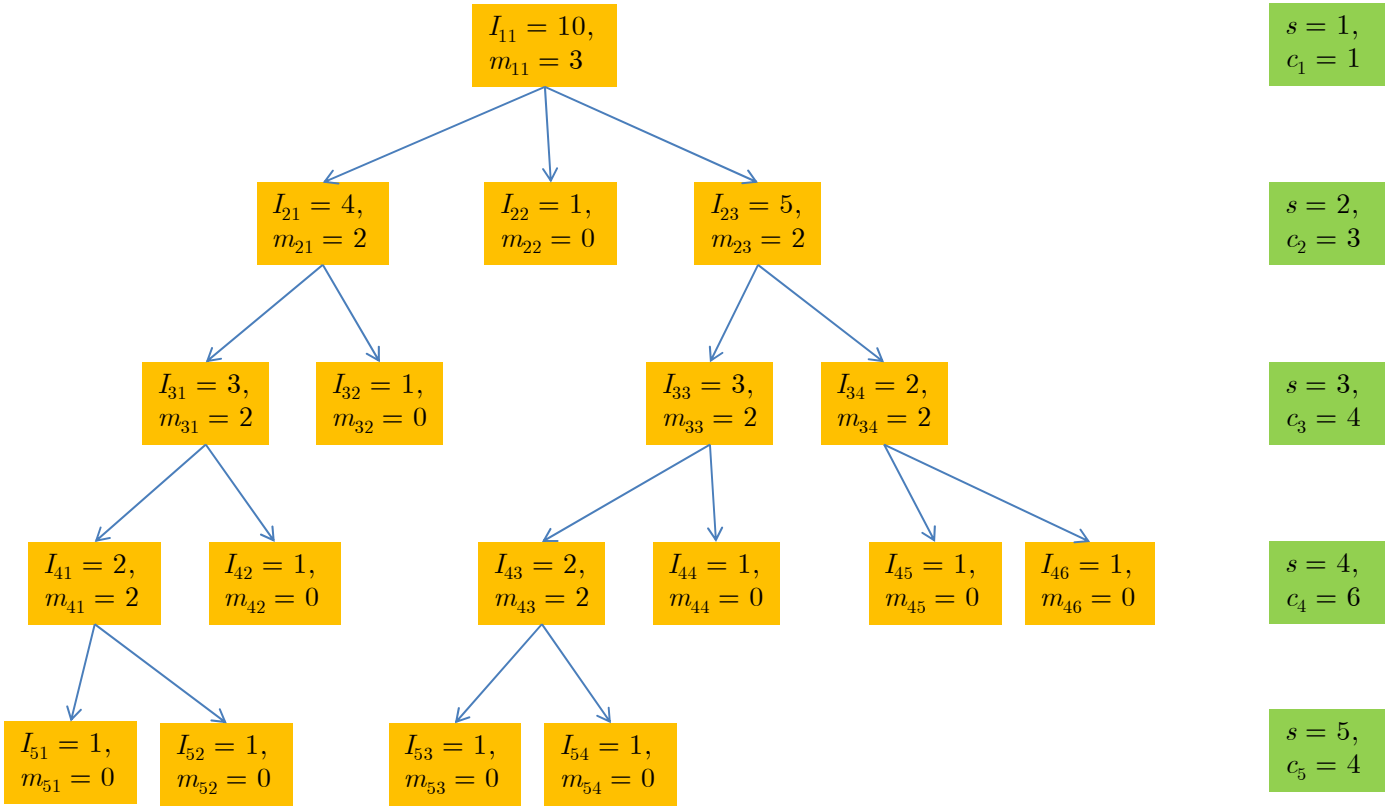

Figure E.1. Five-stage hierarchical group testing example.

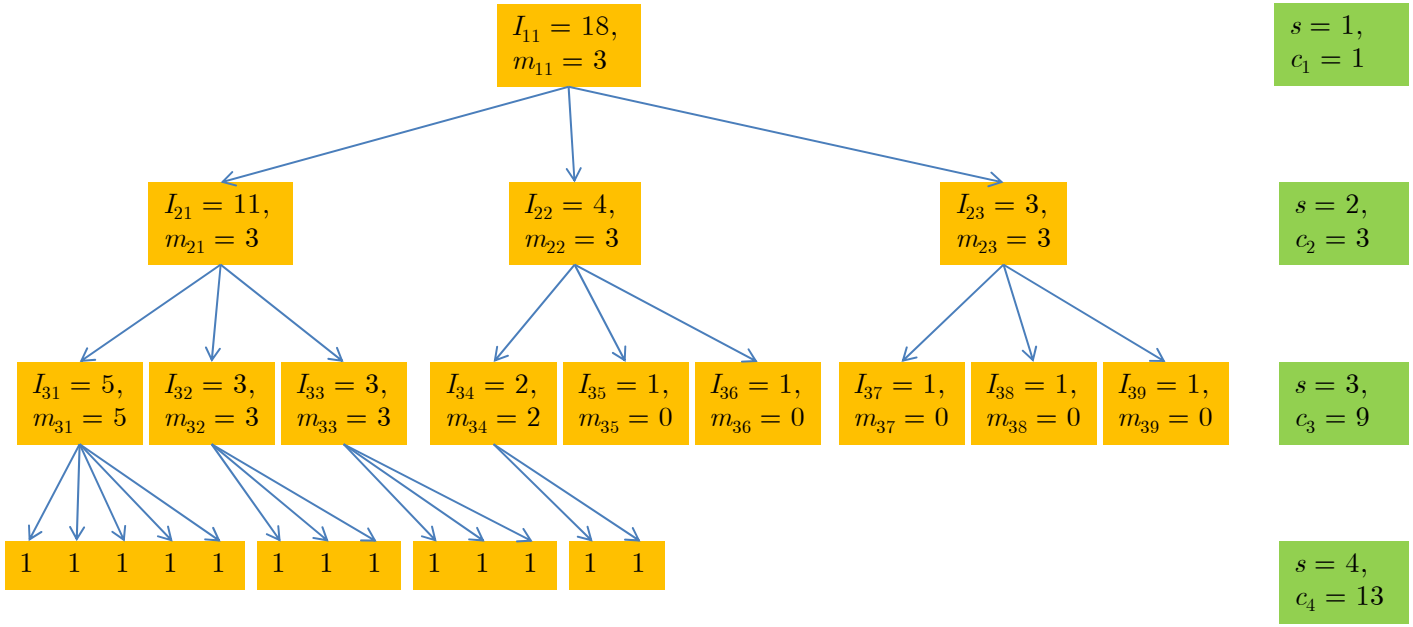

Figure E.2. CRC with $I = 18$, $p = 0.05$, and $\alpha = 1$. Due to the lack of space, the "1" in the last stage indicates $I_{4j} = 1$ for $j = 1, \ldots, 13$. Note that $m_{4j} = 0$ for $j = 1, \ldots, 13$.

Appendix F: Convexity of Equation (3.1)

The purpose of this appendix is to show there are situations when Equation (3.2) is convex upward as a function of the number of subgroups and their sizes.

Figure F.1 displays a simple example for three separate situations where $I = 35$, $S_e = S_p = 1$, and $c_2 = 2$. Case 1 involves $p_i = 0.05$ for all i and convexity holds. Case 2 involves $p_i = 0.15$ for all i, and convexity fails. Case 3 involves the $E(p_{(i)})$s from a beta distribution where $p = 0.15$ and $\alpha = 1.2$, and convexity fails.

Overall, we have found problems with convexity occur in extreme situations where the initial group is very likely to test positive, because the initial group size is large and/or the overall prevalence is high. For example, the probability that the initial group tests positive for case 2 is $1-(1-0.15)^{35} = 0.9966$. These extreme situations generally did not occur in the comparisons of Section 3 and in the IPP example of Section 4, which led to the CRC almost always being the same as the ORC. Unfortunately, due to the large number of factors – such as initial group size, sensitivity and specificity levels, number of sub-groups, and potentially I different individual probabilities – we found developing general conditions for convexity to be intractable.

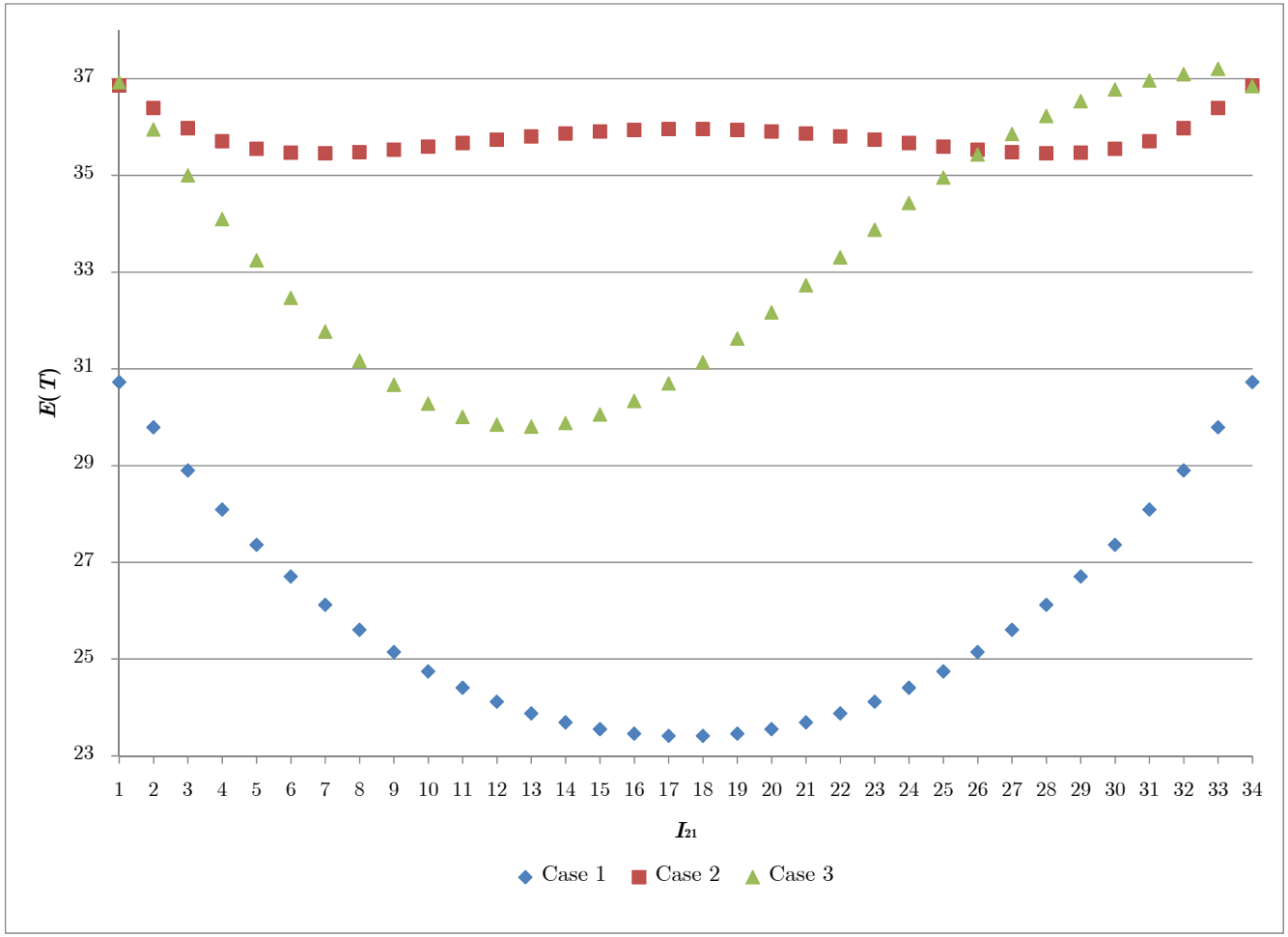

Figure F.1. $E(T)$ for the examples in Appendix F. Case 1 involves $p_i = 0.05$ for all i, case 2 involves $p_i = 0.15$ for all i, and case 3 involves the $E(p_{(i)})$s from a beta distribution where $p = 0.15$ and $\alpha = 1.2$.

Appendix G: Candidate retesting configurations for Section 3.3

Candidate retesting configurations for Section 3.3, Tables G.1 to G.4provide the three-stage CRCs and Tables G.5 to G.14provide the four-stage CRCs.

	$\alpha \to 0$				$\alpha = 0.10$				$\alpha = 1$					$\alpha \to \infty$				
I	$E(T)/I$	I_{21}	I_{22}	I_{23}	$E(T)/I$	I_{21}	I_{22}	I_{23}	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}
5	0.2425	4	1		0.2484	4	1		0.2599	4	1			0.2644	3	2		
6	0.2067	5	1		0.2155	5	1		0.2290	4	2			0.2314	3	3		
7	0.1815	6	1		0.1933	6	1		0.2074	5	2			0.2103	4	3		
8	0.1628	7	1		0.1781	7	1		0.1912	5	3			0.1945	4	4		
9	0.1486	8	1		0.1659	7	1	1	0.1797	6	3			0.1842	5	4		
10	0.1376	9	1		0.1554	7	2	1	0.1717	6	3	1		0.1759	5	5		
11	0.1288	10	1		0.1467	8	2	1	0.1645	6	3	2		0.1689	4	4	3	
12	0.1218	11	1		0.1398	9	2	1	0.1585	6	4	2		0.1627	4	4	4	
13	0.1161	12	1		0.1341	9	3	1	0.1536	6	4	3		0.1588	5	4	4	
14	0.1115	13	1		0.1293	10	3	1	0.1496	7	4	3		0.1554	5	5	4	
15	0.1078	14	1		0.1254	11	3	1	0.1468	8	4	3		0.1526	5	5	5	
16	0.1048	15	1		0.1223	12	3	1	0.1445	8	5	3		0.1511	6	5	5	
17	0.1024	16	1		0.1199	13	3	1	0.1429	9	5	3		0.1495	5	4	4	4
18	0.1004	16	1	1	0.1181	14	3	1	0.1417	9	5	4		0.1480	5	5	4	4
19	0.0976	17	1	1	0.1163	14	4	1	0.1404	7	5	4	3	0.1466	5	5	5	4
20	0.0951	18	1	1	0.1148	15	4	1	0.1389	8	5	4	3	0.1454	5	5	5	5

Table G.1. Group sizes for three-stage CRCs for $p = 0.01$.

	$\alpha \to 0$				$\alpha = 0.10$					$\alpha = 1$					$\alpha \to \infty$					
I	$E(T)/I$	I_{21}	I_{22}	I_{23}	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}	I_{25}
5	0.3351	4	1		0.3555	4	1			0.3984	3	1	1		0.4206	3	2			
6	0.3085	5	1		0.3374	5	1			0.3815	4	1	1		0.3992	3	3			
7	0.2937	6	1		0.3197	5	1	1		0.3703	4	2	1		0.3937	4	3			
8	0.2869	7	1		0.3080	6	1	1		0.3641	4	3	1		0.3873	3	3	2		
9	0.2764	7	1	1	0.3025	7	1	1		0.3608	5	3	1		0.3798	3	3	3		
10	0.2663	8	1	1	0.2968	7	2	1		0.3626	5	3	1	1	0.3805	4	3	3		
11	0.2589	9	1	1	0.2931	8	2	1		0.3620	5	3	2	1	0.3809	4	3	4		
12	0.2537	10	1	1	0.2913	8	3	1		0.3622	5	3	3	1	0.3810	4	4	4		
13	0.2504	11	1	1	0.2885	9	3	1		0.3623	6	3	3	1	0.3847	4	3	3	3	
14	0.2487	12	1	1	0.2870	9	3	1	1	0.3624	6	4	3	1	0.3859	4	4	3	3	
15	0.2464	12	2	1	0.2842	10	3	1	1	0.3646	7	4	3	1	0.3867	4	4	4	3	
16	0.2432	13	2	1	0.2826	11	3	1	1	0.3671	7	4	3	2	0.3873	4	4	4	4	
17	0.2405	14	2	1	0.2822	12	3	1	1	0.3685	7	4	3	3	0.3913	5	4	4	4	
18	0.2385	15	2	1	0.2821	12	4	1	1	0.3701	7	5	3	3	0.3948	5	5	4	4	
19	0.2370	16	2	1	0.2817	13	4	1	1	0.3713	7	5	4	3	0.3966	4	4	4	4	3
20	0.2361	17	2	1	0.2820	14	4	1	1	0.3729	8	5	4	3	0.3968	4	4	4	4	4

Table G.2. Group sizes for three-stage CRCs for $p = 0.05$.

	$\alpha \to 0$					$\alpha = 0.10$						$\alpha = 1$							$\alpha \to \infty$					
I	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}	I_{25}	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}	I_{25}	I_{26}	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}	I_{25}
5	0.4665	4	1			0.4912	4	1				0.5434	3	1	1				0.5924	3	2			
6	0.4590	5	1			0.4710	4	1	1			0.5406	4	1	1				0.5792	3	3			
7	0.4363	5	1	1		0.4562	5	1	1			0.5392	4	2	1				0.5836	3	2	2		
8	0.4204	6	1	1		0.4522	6	1	1			0.5386	4	3	1				0.5785	3	3	2		
9	0.4114	7	1	1		0.4516	6	2	1			0.5389	4	3	1	1			0.5733	3	3	3		
10	0.4083	8	1	1		0.4478	7	2	1			0.5407	5	3	1	1			0.5793	4	3	3		
11	0.4075	8	2	1		0.4407	7	2	1	1		0.5444	5	3	2	1			0.5834	4	4	3		
12	0.4002	9	2	1		0.4345	8	2	1	1		0.5453	5	3	3	1			0.5818	3	3	3	3	
13	0.3921	10	1	1	1	0.4279	8	3	1	1		0.5471	5	4	3	1			0.5858	4	3	3	3	
14	0.3826	10	2	1	1	0.4228	9	3	1	1		0.5487	6	4	3	1			0.5886	4	4	3	3	
15	0.3730	11	2	1	1	0.4202	10	3	1	1		0.5503	6	4	3	1	1		0.5904	4	4	4	3	
16	0.3650	12	2	1	1	0.4196	10	3	1	1	1	0.5523	6	4	3	2	1		0.5915	4	4	4	4	
17	0.3589	13	2	1	1	0.4150	11	3	1	1	1	0.5513	6	4	3	3	1		0.5951	4	4	3	3	3
18	0.3542	13	3	1	1	0.4119	11	4	1	1	1	0.5530	7	4	3	3	1		0.5959	4	4	3	4	3
19	0.3477	14	3	1	1	0.4082	12	4	1	1	1	0.5540	7	4	4	3	1		0.5963	4	4	4	4	3
20	0.3422	15	3	1	1	0.4060	13	4	1	1	1	0.5553	7	4	4	3	1	1	0.5962	4	4	4	4	4

Table G.3. Group sizes for three-stage CRCs for $p = 0.10$.

	$\alpha \to 0$							$\alpha = 0.10$							$\alpha = 1$								$\alpha \to \infty$						
I	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}	I_{25}	I_{26}	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}	I_{25}	I_{26}	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}	I_{25}	I_{26}	I_{27}	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}	I_{25}	I_{26}
5	0.6047	4	1					0.6137	3	1	1				0.6685	3	1	1					0.7407	3	2				
6	0.5707	4	1	1				0.5855	4	1	1				0.6745	4	1	1					0.7306	3	3				
7	0.5488	5	1	1				0.5762	5	1	1				0.6761	4	1	1	1				0.7364	3	2	2			
8	0.5403	6	1	1				0.5793	5	1	1	1			0.6743	4	2	1	1				0.7313	3	3	2			
9	0.5386	6	1	1	1			0.5640	6	1	1	1			0.6710	4	3	1	1				0.7253	3	3	3			
10	0.5173	7	1	1	1			0.5527	6	2	1	1			0.6746	5	3	1	1				0.7319	4	3	3			
11	0.5039	8	1	1	1			0.5423	7	2	1	1			0.6763	5	3	1	1	1			0.7346	3	3	3	2		
12	0.4897	8	2	1	1			0.5335	7	3	1	1			0.6768	4	3	3	1	1			0.7283	3	3	3	3		
13	0.4772	9	2	1	1			0.5252	8	3	1	1			0.6733	5	3	3	1	1			0.7320	4	3	3	3		
14	0.4693	10	2	1	1			0.5141	8	3	1	1	1		0.6731	5	4	3	1	1			0.7342	4	4	3	3		
15	0.4551	10	2	1	1	1		0.5049	9	3	1	1	1		0.6731	5	4	3	1	1	1		0.7306	3	3	3	3	3	
16	0.4422	11	2	1	1	1		0.4997	10	3	1	1	1		0.6730	6	4	3	1	1	1		0.7323	4	3	3	3	3	
17	0.4328	11	3	1	1	1		0.4940	10	3	1	1	1	1	0.6704	5	4	3	3	1	1		0.7332	4	4	3	3	3	
18	0.4227	12	3	1	1	1		0.4874	11	3	1	1	1	1	0.6687	6	4	3	3	1	1		0.7311	3	3	3	3	3	3
19	0.4147	13	3	1	1	1		0.4812	11	4	1	1	1	1	0.6674	6	4	3	3	1	1	1	0.7316	4	3	3	3	3	3
20	0.4079	13	3	1	1	1	1	0.4767	12	4	1	1	1	1	0.6668	6	4	4	3	1	1	1	0.7317	4	4	3	3	3	3

Table G.4. Group sizes for three-stage CRCs for $p = 0.15$.

	$\alpha \to 0$					$\alpha = 0.1$							
I	$E(T)/I$	I_{21}	I_{22}	I_{31}	I_{32}	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{31}	I_{32}	I_{33}	I_{34}
5	0.2404	4	1	3	1	0.2445	4	1		3	1		
6	0.2037	5	1	4	1	0.2092	5	1		4	1		
7	0.1775	6	1	5	1	0.1848	6	1		5	1		
8	0.1579	7	1	6	1	0.1669	7	1		5	1	1	
9	0.1427	8	1	7	1	0.1530	8	1		6	1	1	
10	0.1306	9	1	8	1	0.1424	9	1		7	1	1	
11	0.1207	10	1	9	1	0.1338	10	1		7	2	1	
12	0.1125	11	1	10	1	0.1269	11	1		8	2	1	
13	0.1056	12	1	11	1	0.1213	12	1		9	2	1	
14	0.0997	13	1	12	1	0.1166	13	1		9	3	1	
15	0.0946	14	1	13	1	0.1127	14	1		10	3	1	
16	0.0902	15	1	14	1	0.1095	15	1		11	3	1	
17	0.0864	16	1	15	1	0.1069	16	1		12	3	1	
18	0.0830	17	1	16	1	0.1044	16	1	1	12	3	1	
19	0.0800	18	1	17	1	0.1022	17	1	1	13	3	1	
20	0.0773	19	1	18	1	0.1002	18	1	1	13	3	1	1

Table G.5. Group sizes for four-stage CRC with $p = 0.01$ for $\alpha \to 0$ and $\alpha = 0.1$.

	$\alpha = 1$									$\alpha \to \infty$								
I	$E(T)/I$	I_{21}	I_{22}	I_{31}	I_{32}	I_{33}	I_{34}	I_{35}	I_{36}	$E(T)/I$	I_{21}	I_{22}	I_{31}	I_{32}	I_{33}	I_{34}	I_{35}	I_{36}
5	0.2566	4	1	3	1					0.2654	3	2	2	1	1	1		
6	0.2241	5	1	3	1	1				0.2322	3	3	1	1	1	2	1	
7	0.2022	6	1	3	2	1				0.2094	3	4	1	1	1	2	2	
8	0.1859	7	1	4	2	1				0.1929	4	4	2	2	2	2		
9	0.1733	7	2	4	2	1	1	1		0.1802	4	5	2	2	3	2		
10	0.1635	8	2	4	3	1	1	1		0.1701	5	5	3	2	3	2		
11	0.1557	8	3	4	3	1	1	1	1	0.1618	6	5	3	3	3	2		
12	0.1492	9	3	4	3	2	1	1	1	0.1549	6	6	3	3	3	3		
13	0.1439	10	3	5	3	2	1	1	1	0.1503	7	6	3	4	3	3		
14	0.1396	10	4	5	3	2	2	1	1	0.1461	6	8	3	3	3	3	2	
15	0.1359	11	4	5	3	3	2	1	1	0.1423	6	9	3	3	3	3	3	
16	0.1329	12	4	6	3	3	2	1	1	0.1394	7	9	4	3	3	3	3	
17	0.1303	13	4	6	4	3	2	1	1	0.1366	8	9	2	3	3	3	3	3
18	0.1280	13	5	6	4	3	3	1	1	0.1339	9	9	3	3	3	3	3	3
19	0.1262	13	6	6	4	3	3	2	1	0.1324	9	10	3	3	3	3	4	3
20	0.1246	14	6	7	4	3	3	2	1	0.1310	10	10	3	4	3	3	4	3

Table G.6. Group sizes for four-stage CRC with $p = 0.01$ for$\alpha = 1$ and $\alpha \to \infty$.

	$\alpha \to 0$							$\alpha = 0.1$									
I	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{31}	I_{32}	I_{33}	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{31}	I_{32}	I_{33}	I_{34}	I_{35}	I_{36}
5	0.3232	4	1		3	1		0.3381	4	1		3	1				
6	0.2885	5	1		4	1		0.3093	5	1		4	1				
7	0.2645	6	1		5	1		0.2924	6	1		5	1				
8	0.2475	7	1		6	1		0.2788	7	1		5	1	1			
9	0.2353	8	1		7	1		0.2702	8	1		6	1	1			
10	0.2265	9	1		8	1		0.2654	9	1		7	1	1			
11	0.2205	10	1		9	1		0.2584	9	1	1	7	1	1			
12	0.2147	11	1		9	1	1	0.2526	10	1	1	7	2	1			
13	0.2097	12	1		10	1	1	0.2484	11	1	1	8	2	1			
14	0.2058	13	1		11	1	1	0.2456	12	1	1	8	3	1			
15	0.2028	14	1		12	1	1	0.2432	13	1	1	9	3	1			
16	0.2006	15	1		13	1	1	0.2418	14	1	1	10	3	1			
17	0.1960	15	1	1	13	1	1	0.2413	15	1	1	11	3	1			
18	0.1915	16	1	1	14	1	1	0.2405	16	1	1	11	3	1	1		
19	0.1876	17	1	1	15	1	1	0.2403	17	1	1	12	3	1	1		
20	0.1842	18	1	1	16	1	1	0.2399	17	2	1	12	3	1	1	1	1

Table G.7. Group sizes for four-stage CRC with $p = 0.05$ for $\alpha \to 0$ and $\alpha = 0.1$.

	$\alpha = 1$													$\alpha \to \infty$											
I	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{31}	I_{32}	I_{33}	I_{34}	I_{35}	I_{36}	I_{37}	I_{38}	I_{39}	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{31}	I_{32}	I_{33}	I_{34}	I_{35}	I_{36}	I_{37}
5	0.3914	4	1		2	1	1							0.4281	3	2		2	1	1	1			
6	0.3712	5	1		3	1	1							0.4051	3	3		1	1	1	2	1		
7	0.3594	5	1	1	3	1	1							0.3938	4	3		2	1	1	1	1	1	
8	0.3527	5	2	1	3	1	1	1	1					0.3882	5	3		3	2	1	1	1		
9	0.3475	6	2	1	3	2	1	1	1					0.3832	3	6		1	1	1	3	3		
10	0.3442	7	2	1	4	2	1	1	1					0.3796	4	6		2	2	3	3			
11	0.3409	7	3	1	4	2	1	1	1	1				0.3757	6	5		3	3	2	3			
12	0.3393	8	3	1	4	3	1	1	1	1				0.3722	6	6		3	3	3	3			
13	0.3389	9	3	1	4	3	2	1	1	1				0.3737	7	6		2	3	2	3	3		
14	0.3390	10	3	1	5	3	2	1	1	1				0.3726	8	6		3	3	2	3	3		
15	0.3398	11	3	1	5	3	3	1	1	1				0.3712	6	9		3	3	3	3	3		
16	0.3402	11	3	2	5	3	3	1	1	1	1	1		0.3717	6	4	6	3	3	2	2	3	3	
17	0.3401	11	3	3	5	3	3	1	1	1	1	1	1	0.3700	6	5	6	3	3	3	2	3	3	
18	0.3389	11	4	3	5	3	3	2	1	1	1	1	1	0.3683	6	6	6	3	3	3	3	3	3	
19	0.3386	11	5	3	5	3	3	3	2	1	1	1		0.3701	6	7	6	3	3	3	4	3	3	
20	0.3385	11	6	3	5	3	3	3	3	1	1	1		0.3698	6	8	6	3	3	3	3	2	3	3

Table G.8. Group sizes for four-stage CRC with $p = 0.05$ for$\alpha = 1$ and $\alpha \to \infty$.

	$\alpha \to 0$								$\alpha = 0.1$										
I	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}	I_{31}	I_{32}	I_{33}	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}	I_{31}	I_{32}	I_{33}	I_{34}	I_{35}	I_{36}
5	0.4339	4	1			3	1		0.4549	4	1			3	1				
6	0.4042	5	1			4	1		0.4341	5	1			4	1				
7	0.3866	6	1			5	1		0.4224	6	1			4	1	1			
8	0.3775	7	1			6	1		0.4112	6	1	1		5	1				
9	0.3693	8	1			6	1	1	0.3975	7	1	1		5	1	1			
10	0.3553	8	1	1		7	1		0.3881	8	1	1		6	1	1			
11	0.3437	9	1	1		8	1		0.3825	9	1	1		6	2	1			
12	0.3322	10	1	1		8	1	1	0.3775	10	1	1		7	2	1			
13	0.3227	11	1	1		9	1	1	0.3750	11	1	1		8	2	1			
14	0.3151	12	1	1		10	1	1	0.3731	11	1	1	1	8	2	1			
15	0.3095	13	1	1		11	1	1	0.3675	12	1	1	1	8	3	1			
16	0.3045	14	1	1		11	2	1	0.3628	13	1	1	1	9	3	1			
17	0.3003	15	1	1		12	2	1	0.3599	14	1	1	1	10	3	1			
18	0.2942	15	1	1	1	12	2	1	0.3569	15	1	1	1	10	3	1	1		
19	0.2863	16	1	1	1	13	2	1	0.3549	16	1	1	1	11	3	1	1		
20	0.2796	17	1	1	1	14	2	1	0.3524	16	2	1	1	11	3	1	1	1	1

Table G.9. Group sizes for four-stage CRC with $p = 0.10$ for $\alpha \to 0$ and $\alpha = 0.1$.

	$\alpha = 1$															
I	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}	I_{31}	I_{32}	I_{33}	I_{34}	I_{35}	I_{36}	I_{37}	I_{38}	I_{39}	$I_{3,10}$
5	0.5395	3	1	1		2	1								
6	0.5261	4	1	1		2	1	1							
7	0.5193	5	1	1		3	1	1							
8	0.5207	5	2	1		3	1	1	1	1					
9	0.5210	6	2	1		3	2	1	1	1					
10	0.5194	6	3	1		3	2	1	1	1	1				
11	0.5187	7	3	1		4	2	1	1	1	1				
12	0.5190	8	3	1		4	3	1	1	1	1				
13	0.5191	8	3	1	1	4	3	1	1	1	1				
14	0.5203	9	3	1	1	4	3	2	1	1	1				
15	0.5206	8	3	3	1	4	3	1	1	1	1	1	1	1	
16	0.5195	9	3	3	1	4	3	2	1	1	1	1	1	1	
17	0.5192	10	3	3	1	5	3	2	1	1	1	1	1	1	
18	0.5190	11	3	3	1	5	3	3	1	1	1	1	1	1	
19	0.5194	11	4	3	1	5	3	3	1	1	1	1	1	1	1
20	0.5190	10	6	3	1	5	3	2	3	3	1	1	1		

Table G.10. Group sizes for four-stage CRC with $p = 0.10$ for$\alpha = 1$.

	$\alpha \to \infty$												
I	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{31}	I_{32}	I_{33}	I_{34}	I_{35}	I_{36}	I_{37}	I_{38}	I_{39}
5	0.6086	3	2		2	1	1	1					
6	0.5923	3	3		1	2	1	1	1				
7	0.5875	3	4		1	1	1	2	1	1			
8	0.5878	3	3	2	1	1	1	2	1	1	1		
9	0.5814	3	3	3	1	1	1	2	1	1	1	1	
10	0.5808	3	4	3	1	1	1	2	1	1	1	1	1
11	0.5841	3	5	3	1	1	1	3	2	1	1	1	
12	0.5840	3	6	3	1	1	1	3	3	1	1	1	
13	0.5856	3	4	6	1	1	1	1	1	1	1	3	3
14	0.5862	3	5	6	1	1	1	3	2	3	3		
15	0.5843	6	6	3	3	3	3	3	1	1	1		
16	0.5842	6	6	4	3	3	3	3	1	1	1	1	
17	0.5832	6	5	6	3	3	3	2	3	3			
18	0.5807	6	6	6	3	3	3	3	3	3			
19	0.5827	6	7	6	3	3	3	2	2	3	3		
20	0.5814	6	8	6	3	3	3	3	2	3	3		

Table G.11. Group sizes for four-stage CRC with $p = 0.10$ for$\alpha \to \infty$.

	$\alpha \to 0$									$\alpha = 0.1$											
I	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}	I_{25}	I_{31}	I_{32}	I_{33}	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}	I_{25}	I_{31}	I_{32}	I_{33}	I_{34}	I_{35}	I_{36}
5	0.5474	4	1				3	1		0.5701	4	1				3	1				
6	0.5249	5	1				4	1		0.5574	5	1				4	1				
7	0.5103	5	1	1			4	1		0.5307	5	1	1			4	1				
8	0.4828	6	1	1			5	1		0.5132	6	1	1			4	1	1			
9	0.4642	7	1	1			6	1		0.4983	7	1	1			5	1	1			
10	0.4475	8	1	1			6	1	1	0.4904	8	1	1			6	1	1			
11	0.4343	9	1	1			7	1	1	0.4850	8	1	1	1		6	1	1			
12	0.4255	10	1	1			8	1	1	0.4728	9	1	1	1		6	2	1			
13	0.4128	10	1	1	1		8	1	1	0.4631	10	1	1	1		7	2	1			
14	0.3985	11	1	1	1		9	1	1	0.4573	11	1	1	1		8	2	1			
15	0.3880	12	1	1	1		9	2	1	0.4526	12	1	1	1		8	3	1			
16	0.3791	13	1	1	1		10	2	1	0.4445	12	1	1	1	1	8	3	1			
17	0.3728	14	1	1	1		11	2	1	0.4377	13	1	1	1	1	9	3	1			
18	0.3641	14	1	1	1	1	11	2	1	0.4325	14	1	1	1	1	9	3	1	1		
19	0.3544	15	1	1	1	1	12	2	1	0.4284	15	1	1	1	1	10	3	1	1		
20	0.3471	16	1	1	1	1	13	2	1	0.4235	14	3	1	1	1	10	3	1	1	1	1

Table G.12. Group sizes for four-stage CRC with $p = 0.15$ for $\alpha \to 0$ and $\alpha = 0.1$.

	$\alpha = 1$																
I	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}	I_{25}	I_{26}	I_{31}	I_{32}	I_{33}	I_{34}	I_{35}	I_{36}	I_{37}	I_{38}	I_{39}
5	0.6642	3	1	1				2	1							
6	0.6547	4	1	1				2	1	1						
7	0.6510	5	1	1				3	1	1						
8	0.6538	5	1	1	1			3	1	1						
9	0.6539	5	2	1	1			3	1	1	1	1				
10	0.6513	5	3	1	1			3	1	1	1	1	1			
11	0.6490	6	3	1	1			3	2	1	1	1	1			
12	0.6479	7	3	1	1			4	2	1	1	1	1			
13	0.6474	8	3	1	1			4	3	1	1	1	1			
14	0.6465	8	3	1	1	1		4	3	1	1	1	1			
15	0.6460	7	3	3	1	1		4	2	1	1	1	1	1	1	1
16	0.6425	8	3	3	1	1		4	3	1	1	1	1	1	1	1
17	0.6409	9	3	3	1	1		4	3	2	1	1	1	1	1	1
18	0.6399	9	3	3	1	1	1	4	3	2	1	1	1	1	1	1
19	0.6386	10	3	3	1	1	1	4	3	3	1	1	1	1	1	1
20	0.6371	11	3	3	1	1	1	5	3	3	1	1	1	1	1	1

Table G.13. Group sizes for four-stage CRC with $p = 0.15$ for$\alpha = 1$.

$\alpha \to \infty$

I	$E(T)/I$	I_{21}	I_{22}	I_{23}	I_{24}	I_{25}	I_{31}	I_{32}	I_{33}	I_{34}	I_{35}	I_{36}	I_{37}	I_{38}	I_{39}	$I_{3,10}$	$I_{3,11}$	$I_{3,12}$	$I_{3,13}$	$I_{3,14}$	$I_{3,15}$	$I_{3,16}$
5	0.7665	3	2				2	1	1	1												
6	0.7517	3	3				1	1	1	2	1											
7	0.7475	4	3				2	1	1	1	1	1										
8	0.7465	3	3	2			1	1	1	2	1	1	1									
9	0.7387	3	3	3			1	1	1	2	1	1	1	1								
10	0.7367	3	4	3			1	1	1	2	1	1	1	1	1							
11	0.7404	4	4	3			1	1	1	1	2	1	1	1	1	1						
12	0.7381	3	3	3	3		1	1	1	2	1	1	1	1	1	1	1					
13	0.7354	3	4	3	3		1	1	1	1	1	2	1	1	1	1	1	1				
14	0.7374	3	3	4	4		1	1	1	1	1	1	1	1	2	1	1	1	1			
15	0.7368	3	6	3	3		1	1	1	3	3	1	1	1	1	1	1					
16	0.7372	3	6	4	3		1	1	1	3	3	1	1	1	1	1	1	1				
17	0.7357	3	4	3	4	3	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1
18	0.7343	3	6	3	3	3	1	1	1	3	3	1	1	1	1	1	1	1	1	1		
19	0.7341	3	9	3	4		1	1	1	3	3	3	1	1	1	1	1	1	1			
20	0.7334	3	9	4	4		1	1	1	3	3	3	1	1	1	1	1	1	1	1		

Table G.14. Group sizes for four-stage CRC with $p = 0.15$ for$\alpha \to \infty$.

Appendix H: Additional accuracy summaries for Section 3.3

Corresponding to the discussion in Section 3.3, Figures H.1, H.2 and H.3give the accuracy measures with $p = 0.01$, 0.10, and 0.15, respectively.

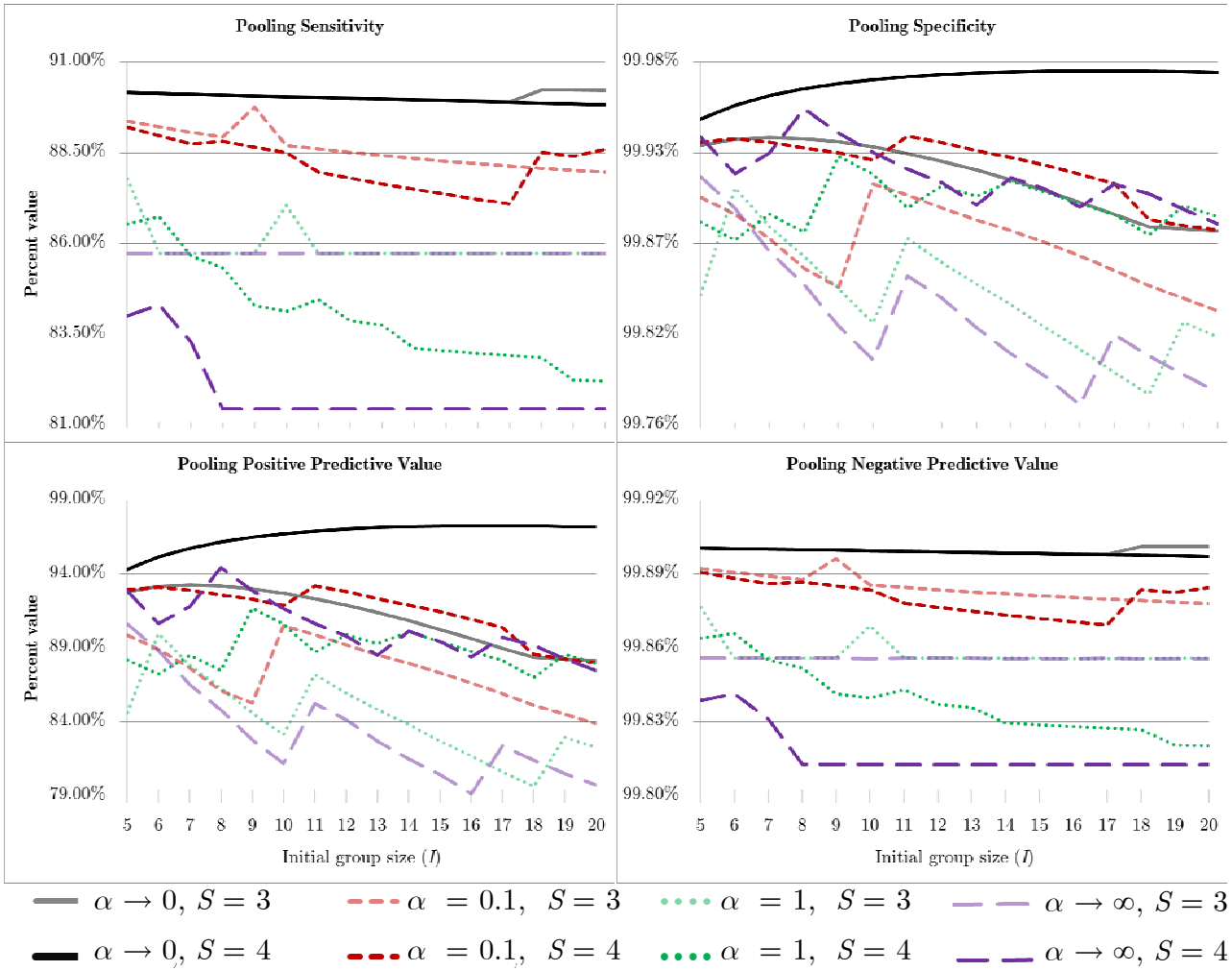

Figure H.1. Accuracy measures when $p = 0.01$ and $S_e = S_p = 0.95$.

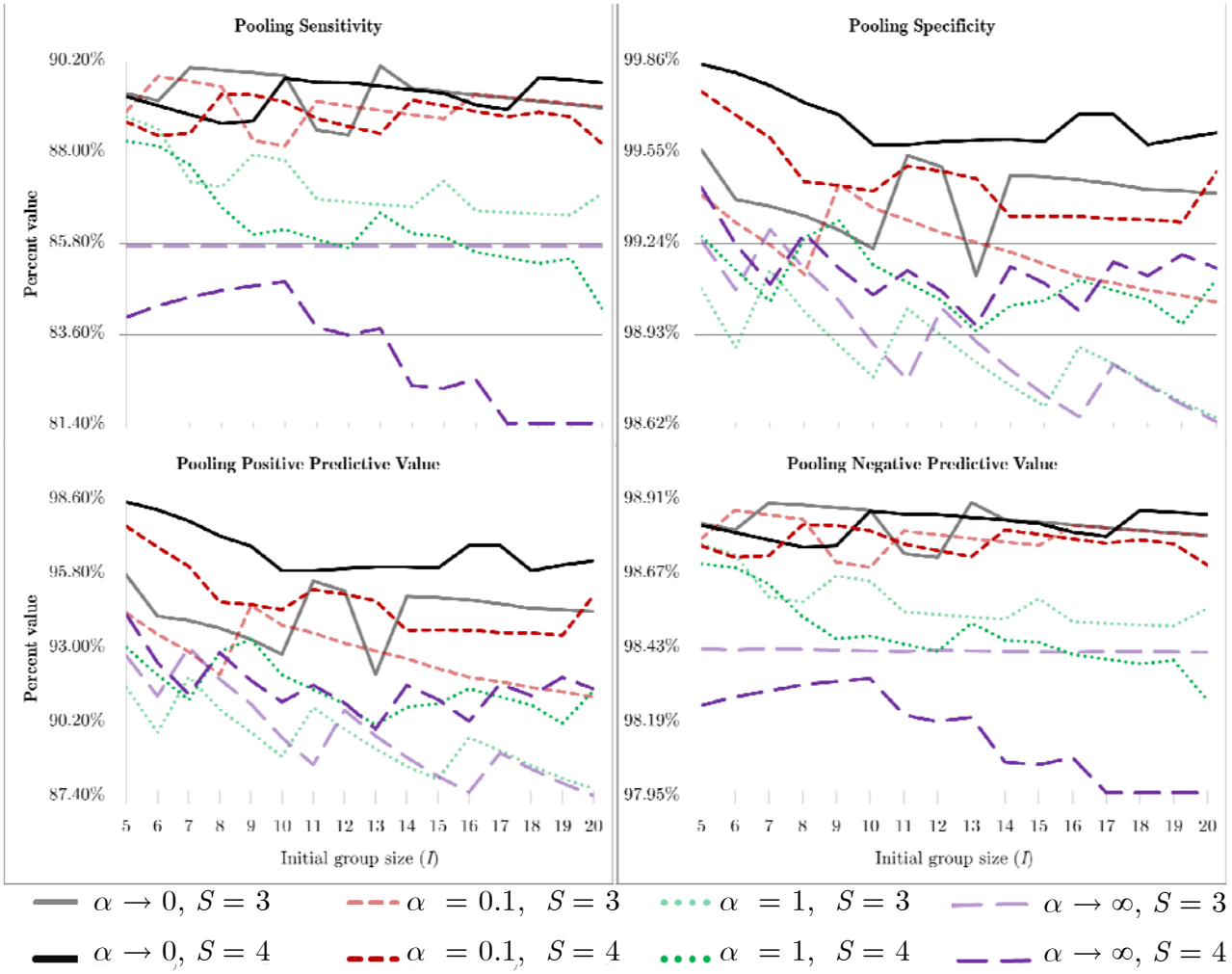

Figure H.2. Accuracy measures when $p = 0.10$ and $S_e = S_p = 0.95$.

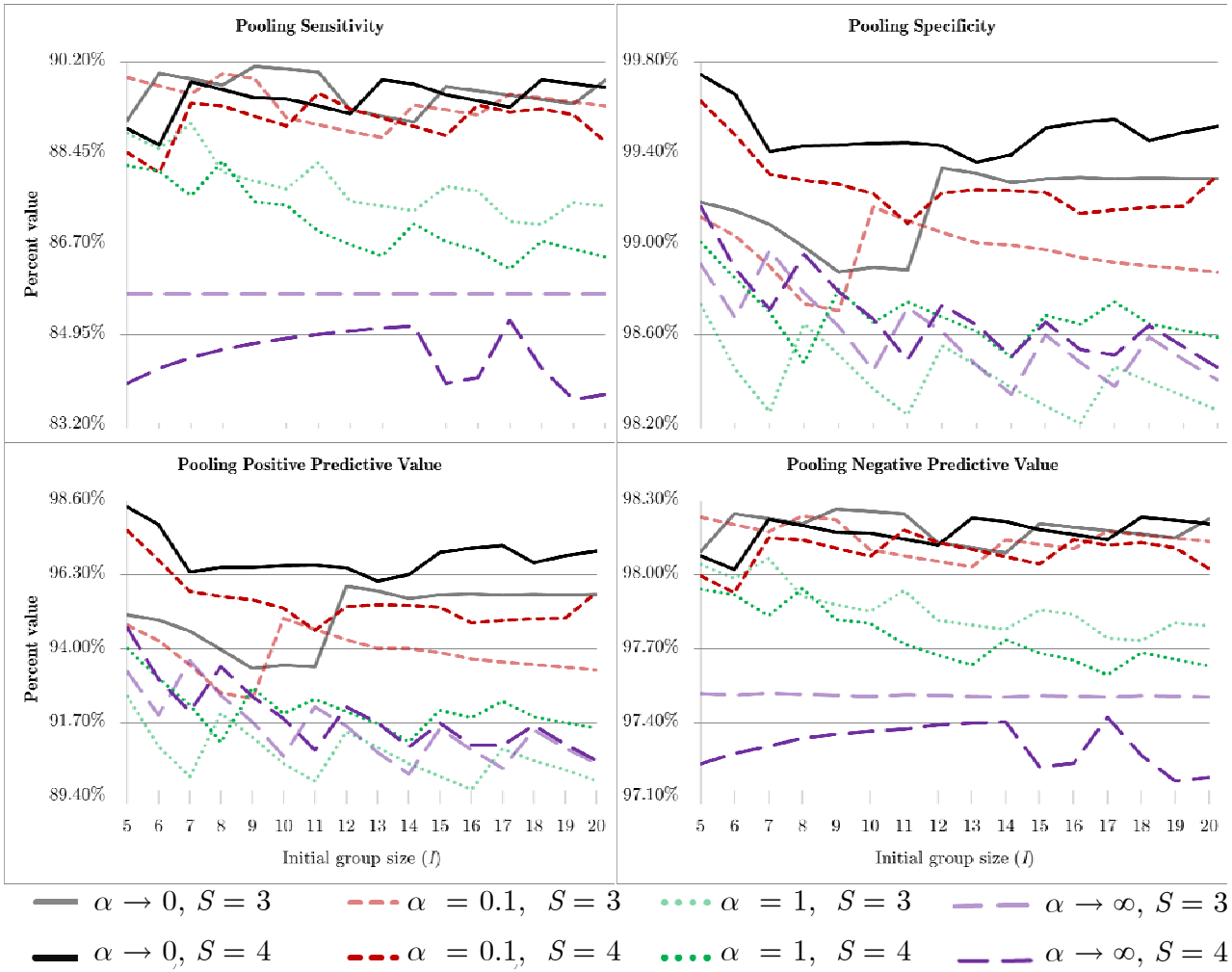

Figure H.3. Accuracy measures when $p = 0.15$ and $S_e = S_p = 0.95$.

Appendix I: R function documentation

Vector of probabilities produced from a beta distribution.

Description

Produces a vector of expected ordered statistics for a random sample of a specified size from a beta distribution

Usage

```
beta.dist(p, alpha = 1, beta = NULL, grp.sz = 10, simul = FALSE,
      rel.tol = ifelse(alpha < 1, .Machine$double.eps^0.1,
      .Machine$double.eps^0.25),  plot = FALSE)
```

Arguments

`p`	a probability value; if the value is between 0 and 1, it is used as mean of the beta random variable
`alpha`	value for alpha in the beta distribution. If `alpha = "hom"` or `"inf"`, a homogeneous vector of probabilities is produced. If `alpha = 0`, the Bernoulli distribution is used.
`beta`	if p = NULL or not a value between 0 and 1, b is the beta value for the beta distribution
`grp.sz`	number of probabilities to produce
`simul`	finds the expected order statistics through simulation instead of using the PDFs
`rel.tol`	used in `integrate()` function; may need to change for different values of `alpha`
`plot`	returns a plot of the associated beta distribution with expected ordered values indicated, only works for the beta distributions

Details

Produces a vector of probabilities

Value

`p.vec` vector of probabilities

Author(s)

Examples

```
p1 <- beta.dist(p = 0.05, alpha = 1, grp.sz = 16, plot = TRUE)
          round(p1, 4)
```

Descriptive information for the halving group testing procedure up to 5 stages.

Description

Obtains PMF for the possible number of tests to identify all positive individuals in a group using the "halving" group testing procedure. Returns the PMF, expected value, and variance.

Usage

```
halving(p, se = 1, sp = 1, stages = 2, order.p = TRUE)
```

Arguments

`p`	a vector of individual probabilities
`se, sp`	sensitivity and specificity, respectively
`stages`	number of stages for hierarchical group testing
`order.p`	default is `TRUE` indicating the values in `p` are sorted; `FALSE` leaves the individual probabilities in their given order

Details

The PMF is produced for a vector of individual probabilities.

Value

`PMF`	data frame with number of tests and associated probabilities
`ET`	expected number of tests
`VT`	variance of number of tests

Author(s)

Michael Black, Christopher R Bilder

Examples

```
ex1 <- halving(p = rep(x = 0.05, times = 16), stages = 3)
ex1
ex2 <- halving(p = rep(x = 0.05, times = 16), stages = 3, se = 0.95,
          sp = 1)
p1 <- beta.dist(p = 0.05, a = 1, b = NULL, grp.sz = 16)
ex3 <- halving(p = p1, stages = 3, se = 0.95, sp = 1, order.p = TRUE)
```

```
data.frame(t = ex1$pmf[,1], ex1 = ex1$pmf[,2], ex2 = ex2$pmf[,2], ex3
          = ex3$pmf[,2], row.names = NULL)  #PMF
data.frame(ex1 = ex1$et, ex2 = ex2$et, ex3 = ex3$et) #E(T)
data.frame(ex1 = ex1$vt, ex2 = ex2$vt, ex3 = ex3$vt) #Var(T)
```

Diagnostic information for generalized hierarchical group testing procedure.

Description

Returns diagnostic information for a given vector of probabilities for 2, 3, or 4 stages.

Usage

```
hierarchical.desc(p, se = 1, sp = 1, I2 = NULL, I3 = NULL, order.p =
            TRUE)
```

Arguments

`p`	a vector of individual probabilities
`se, sp`	sensitivity and specificity, respectively
`stages`	number of stages for hierarchical group testing
`I2`	vector of stage 2 subgroup sizes
`I3`	vector of stage 3 subgroup sizes
`order.p`	default is `TRUE` indicating informative retesting; `FALSE` leaves the individual probabilities in their given order

Details

Produces the diagnostic information. Pooling diagnostic information is for the given group. If the group is part of a larger retesting process, individual diagnostic values should be used for the entire process to calculate the pooling diagnostic values.

Value

`ET`	expected number of tests
`group.size I2, I3`	vectors with number of individuals in each subgroup
`m1, m2, m3`	vectors with number of subgroups a group splits into
`individual.testerror`	table of individual pooled accuracy measures
`group.testerror`	vector of pooled accuracy measures for the group

Author(s)

Michael Black, Christopher R Bilder

Examples

```
p1 <- beta.dist(p = 0.05, a = 1, b = NULL, grp.sz = 16)
ex4 <- hierarchical.desc(p = p1, se = 0.95, sp = 1, I2 = c(8, 8), I3
    = NULL)
ex4
```

Optimal or candidate retesting configuration.

Description

Finds the optimal configuration (ORC) or a possible optimal configuration (CRC) for retesting over 2, 3 or 4 stages

Usage

```
get.CRC(p, se = 1, sp = 1, stages = 2, order.p = TRUE, everycase =
            FALSE, init.config = "hom")
```

Arguments

`P`	a vector of individual probabilities
`se, sp`	sensitivity and specificity, respectively
`stages`	number of stages for hierarchical group testing
`order.p`	default is `TRUE` indicating informative retesting; `FALSE` leaves the individual probabilities in their given order
`everycase`	default is `FALSE` indicating the CRC is found; `TRUE` finds ORC, warning of possible excessive time.
`init.config`	default is `"hom"` which uses an initial configuration with subgroups of approximately equal sizes; `"ord"` uses an initial configuration with 1 individual in each subgroup except the first; `"both"` uses both `"hom"` and `"ord"` methods

Details

ORC can require an excessive amount of running time for large groups (>18 for three stages, >13 for four stages). If a group has fairly homogeneous probabilities then `init.config = "hom"` should be used. If a group has very heterogeneous probabilities `init.config = "ord"` should be used.

Value

`ORC` Returned if everycase = TRUE

`CRC` Returned if everycase = FALSE

`Desc` Provides diagnostic values from hierarchical.desc

Author(s)

Michael Black, Christopher R Bilder

Examples

```
p1 <- beta.dist(p = 0.05, a = 1, b = NULL, grp.sz = 16)
ex5 <- get.CRC(p = p1, se = .95,sp = 1, stages = 3)
ex5
```