http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 65, No. 9, September 2019, pp. 4365–4384
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Optimal Risk-Based Group Testing

Hrayer Aprahamian,[a] Douglas R. Bish,[b] Ebru K. Bish[b]

[a] Industrial and Systems Engineering, Texas A&M University, College Station, Texas 77843; [b] Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, Virginia 24061

Contact: hrayer@tamu.edu, http://orcid.org/0000-0002-8750-2366 (HA); drb1@vt.edu (DRB); ebru@vt.edu (EKB)

Received: April 25, 2017
Revised: March 4, 2018; May 3, 2018
Accepted: May 24, 2018
Published Online in Articles in Advance: May 8, 2019

https://doi.org/10.1287/mnsc.2018.3138

Abstract. Group testing (i.e., testing multiple subjects simultaneously with a single test) is essential for classifying a large population of subjects as *positive* or *negative* for a binary characteristic (e.g., presence of a disease). We study optimal group testing designs under subject-specific risk characteristics and imperfect tests, considering classification accuracy-, efficiency- and equity-based objectives, and characterize important structural properties of optimal testing designs. These properties allow us to model the testing design problems as partitioning problems, develop efficient algorithms, and derive insights on equity versus accuracy trade-off. One of our models reduces to a constrained shortest path problem, for a special case of which we develop a polynomial-time algorithm. We also show that determining an optimal risk-based Dorfman testing scheme that minimizes the expected number of tests is tractable, resolving an open conjecture. We demonstrate the value of optimal risk-based testing schemes with a case study of public health screening.

History: Accepted by Yinyu Ye, optimization.
Funding: This material is based on work supported in part by the National Science Foundation [Grant 1055360]. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2018.3138.

Keywords: group testing • Dorfman testing • risk-based testing • classification errors • equity • combinatorial optimization • constrained shortest path

1. Introduction and Motivation

Screening a population of subjects so as to classify each subject as *positive* or *negative* for a binary characteristic (e.g., presence of a disease or genetic disorder, a product defect, error in a computer code) is essential in many settings. Individually testing each subject is often very costly, and hence may not be a viable strategy for classification, especially when the prevalence of the binary characteristic in the population is low and the population size is large. Therefore, in 1943, Dorfman, an economist, proposed the concept of *group testing*, which involves testing multiple subjects simultaneously using a single test, for the purpose of screening military inductees for syphilis in an economical manner (Dorfman 1943). This so-called *Dorfman testing* scheme has two stages: in the first stage, subjects are tested in groups; if a group tests negative, then all subjects in the group are classified as negative; and if a group tests positive, then each subject in the group is individually tested and classified according to the outcome of their individual test. Dorfman testing is one of the most commonly utilized group testing schemes today; for example, in donated blood screening, Dorfman testing has become the standard practice in the United States and several European countries (American Red Cross 2017, European Blood Alliance 2017). However, many unrealistic assumptions were imposed in Dorfman's original model, as well as in most of the subsequent research on Dorfman testing. These include the assumptions of *perfect* tests (i.e., no classification errors), *homogeneous* (identical) subjects (i.e., the probability of having the binary characteristic is the same for all subjects), and an infinite testing population (in reality, the number of subjects is finite and known in each testing period).

Specifically, the decision problem is as follows: there is a finite set of subjects to be classified as positive or negative for a certain binary characteristic in each testing period; and there are risk factors that are known to increase a subject's probability of positivity (*risk*) for the binary characteristic in question (i.e., subjects come from a *heterogeneous* population). A screening test, which may be used on individual subjects or groups of subjects, is available to detect the binary characteristic, but the test is imperfect, leading to the possibility of false-positive or false-negative classifications. The challenge, then, is to design a *risk-based Dorfman testing scheme* (i.e., determining group sizes, and assigning subjects, with different risk, to the groups) so as to classify the set of subjects for the binary characteristic *accurately* (i.e., with minimum classification error), *equitably* (i.e., with a fair and even distribution of misclassification probability across subjects), and *efficiently* (i.e., with

minimum resources). Equity (fairness) is an important, and often over-looked, dimension of resource allocation problems (e.g., Luss 1999, Bertsimas et al. 2011, 2012), and considering the trade-off between accuracy and equity is especially important in public health screening (Wagstaff 1991, Brandeau et al. 2004). We identify important structural properties of optimal risk-based Dorfman testing schemes with imperfect tests; use these properties to develop efficient algorithms; and derive key insights through a case study that demonstrates an application of the proposed scheme within a public health screening setting.

Our decision problem applies in a wide variety of settings. As mentioned above, an important application of group testing arises in public health screening. For example, state health laboratories often have programs to screen the sexually active population in the state for sexually transmitted diseases (STDs) (Roche Diagnostics USA 2017), and blood banks screen all donated blood for a set of transfusion-transmittable diseases, including, for example, the human immunodeficiency virus (HIV) and hepatitis viruses (Weusten et al. 2011, Aprahamian et al. 2016, U.S. Food and Drug Administration 2017). Individually testing each subject, within a large population, for a disease is typically not feasible owing to limited resources. Another important consideration is that disease prevalence rates may vary, and sometimes substantially, with subject characteristics. For example, in the United States, subjects within the 15–24-year age group are 12 times more likely to be infected with chlamydia, one of the most prevalent STDs in the United States, than subjects from other age groups (Centers for Disease Control and Prevention 2014); first-time blood donors are seven times more likely to be infected with HIV than repeat donors (Zou et al. 2012). Group testing can also enable the use of more accurate, yet more expensive, tests that would have been too expensive to be used for individual testing (e.g., nucleic acid testing technology). An equitable testing scheme is especially important in public health screening. For example, in STD screening, the testing scheme that minimizes the classification errors may do so by placing the subjects in Dorfman groups with different sizes in the first stage of testing, based, in part, on how the demographics (e.g., age, race/ethnicity) impact risk. It is important that in doing so the solution does not unfairly increase the classification errors for a certain subset of subjects (e.g., subjects within a specific age group or belonging to a certain race/ethnicity). Group testing has also seen wide applications beyond healthcare systems. Consider, for example, a common communication channel (e.g., a satellite connection) that is shared by a large number of users; if multiple users attempt to transmit a signal during the same time slot, a collision occurs and must be resolved by identifying the "active" users (i.e., users that are transmitting a signal) during this time slot. Group testing can be utilized to identify these active users efficiently, significantly reducing the conflict resolution time in multiaccess communication networks (Berger et al. 1984). In this context the probability of being active varies depending on user characteristics; for example, users with a history of heavy signal transmission would have a higher probability of being active. Alternatively, in industrial quality control, a manufacturer can test a group of light bulbs for defects by arranging them in series and applying a voltage across the group; or test a group of airtight containers for leaks by filling them with an indicator gas and placing the group in a low-pressure chamber—the presence of the gas in the chamber indicates leakage in at least one of the containers (Sobel and Groll 1959). The probability of being defective can vary on the basis of, for example, properties of manufacturing machines, including their age and maintenance schedules. Other examples of group testing applications include, among others, software testing (Blass and Gurevich 2002), data compression (Hong and Ladner 2002), compressed sensing (Cormode and Muthukrishnan 2006), and DNA library screening (Pevzner and Lipshutz 1994).

Dorfman testing has been extensively studied but mostly under restrictive assumptions, such as perfect tests (e.g., Dorfman 1943, Sobel and Groll 1959, Hwang 1975, Saraniti 2006, Feng et al. 2010, Li et al. 2014), which leads to a focus on minimizing the number of tests (i.e., maximizing efficiency) rather than maximizing classification accuracy or equity; infinite populations; and with subjects having identical risk for the binary characteristic (i.e., the testing population is homogeneous) (e.g., Graff and Roeloffs 1972, Wein and Zenios 1996, Gupta and Malina 1999, Kim et al. 2007). One of the earliest works to incorporate subject-specific risk characteristics in group testing design is by Hwang (1975), who studies the problem of determining a risk-based Dorfman testing scheme so as to minimize the expected number of tests for perfect tests. More recent work (e.g., Bilder and Tebbs 2012, Black et al. 2012, McMahan et al. 2012, Tebbs et al. 2013, Black et al. 2015) extends the analysis to the realistic case of imperfect tests; these analyses, however, solely rely on heuristics that attempt to reduce the expected number of tests rather than maximize the classification accuracy or equity under a testing budget constraint, as we do in this paper; other recent work derives analytical expressions for the performance measures under continuous test outcomes (Wang et al. 2018), but this work does not consider optimization considering a general objective function comprising classification accuracy and expected number of tests. In particular, McMahan et al. (2012) state that determining an optimal risk-based Dorfman testing scheme that minimizes the expected number of tests (i.e., an extension of the model of Hwang 1975 to the case of imperfect tests) "appears to be intractable"; hence,

the paper develops various heuristics, which range from restricting all group sizes to be equal; to testing the low-risk subjects (determined via a preset risk threshold) in groups of equal size, while testing the high-risk subjects individually; to varying group sizes but under the assumption that group sizes reduce as subject risk increases. Although some of these properties may seem intuitive, we show in this paper that an optimal solution does not necessarily satisfy any of these properties; thus, none of the greedy heuristics in McMahan et al. (2012) necessarily converge to the optimal solution. Further, we show that the aforementioned extension of Hwang's model to the case of imperfect tests (i.e., the problem of determining a risk-based Dorfman testing scheme so as to minimize the expected number of tests under imperfect tests) is in fact tractable, resolving the conjecture in the relevant literature. Our model is quite general and is able to incorporate subject-specific risk characteristics into the testing design and relax the perfect test and infinite population assumptions, thus enabling us to consider all important dimensions of testing: classification accuracy, efficiency, and equity.

Specifically, we consider a number of objective functions, because in practice there are different goals depending on the context of the problem. In particular, we explore (i) minimizing a weighted sum of the expected number of false-negative and false-positive classifications and number of tests (i.e., *the system's problem* and the first best solution), a special case of which reduces to the minimization of the expected number of tests (i.e., our model extends the earlier works of Hwang 1975 and McMahan et al. 2012); and (ii) minimizing a weighted sum of the expected number of false-negative and false-positive classifications under a testing budget constraint (i.e., *the budget-constrained problem*). Further, we explore an equity-based formulation, which aims to capture the trade-off between classification accuracy and equity under a testing budget constraint. These formulations may arise in various settings. For instance, in STD screening, minimizing classification errors and maximizing equity, especially with respect to false-negative classifications, are important objectives. In this setting, false-positive classifications may only lead to further confirmatory testing, whereas false-negative classifications may lead to medical complications and further spread of the disease; and a testing design in which a subset of the subjects has a higher false-negative classification probability compared with the others is not desirable from a societal perspective. In industrial quality control, both false negatives (which result in defective products shipped to customers) and false positives (which result in unnecessary wastage) are of importance; thus, minimizing a weighted sum of false negatives, false positives, and testing costs would be an appropriate objective.

Our contributions in this paper are multifold. We formulate the aforementioned decision problems as *partitioning problems* and develop key structural properties. These properties allow us to reduce the system's problem to a shortest path problem and the budget-constrained problem to a constrained shortest path problem. Not surprisingly, the shortest path problem arises in various other contexts, including pricing and inventory management, transportation, supply chain management, and scheduling (e.g., Van Hoesel et al. 2005, Deng and Yano 2006, Verter and Kara 2008, Solyalı et al. 2015), and our work adds a novel application to this set. The constrained shortest problem is, in general, *NP*-hard (Garcia 2009), and the algorithms proposed in the literature are not polynomial for the general problem (e.g., Yen 1971, Handler and Zang 1980, Beasley and Christofides 1989, Irnich et al. 2010, and Engineer et al. 2011). Depending on the setting, the problem size can be quite large in our context; for example, in STD screening it is common for a state health laboratory to screen specimens taken from approximately 100 subjects every day (Lewis et al. 2012), hence the algorithms developed in the literature become computationally expensive for realistic instances. Toward this end, we utilize the structural properties of our decision problem to improve its tractability; and, for special cases of our decision problem, we develop a polynomial-time algorithm that can solve the corresponding constrained shortest path problem. Further, our study of an equity-based objective provides valuable insight; for example, in our budget-constrained problem, when the objective is to maximize the equity with respect to false-negative classifications, we show that there is no trade-off between classification accuracy and equity (i.e., the price of fairness is zero), and the testing design that minimizes the number of false-negative classifications is also the one that maximizes equity. Finally, we demonstrate the effectiveness of the proposed risk-based Dorfman testing scheme through a case study on chlamydia screening in the United States using published data. The proposed risk-based Dorfman scheme not only substantially reduces each of the expected number of false negatives, expected number of false positives, and expected number of tests but also significantly increases the equity of the testing scheme, over optimal non–risk-based schemes and current screening practices. Such improvements in all performance measures underscore the value of incorporating subject-specific risk characteristics into the testing design.

The remainder of this paper is organized as follows. Section 2 presents the notation, decision problem, and formulations; and Section 3 provides derivations of the performance measures. Section 4 studies the optimal design of risk-based Dorfman testing schemes in different settings and derives important structural properties of

optimal solutions. Then, Section 5 discusses findings from the U.S. chlamydia screening case study. Finally, Section 6 summarizes our findings and provides directions for future research. To facilitate the presentation, all proofs are relegated to the online appendix.

2. Notation, Decision Problem, and Models

2.1. Notation and Decision Problem

Throughout, we denote random variables in uppercase letters, their realization in lowercase letters, and vectors in bold. We use indices m and i to respectively refer to a subject and to a group, and we use the subject index as a superscript and group index as a subscript (i.e., X^m versus X_i). Finally, we use the terms positive and negative to refer both to subjects (i.e., to respectively denote the presence or absence of the characteristic) and to binary test outcomes (i.e., to respectively denote the test outcomes that *indicate* the presence or absence of the characteristic).

Consider an ordered set, $S = \{1, \cdots, N\}$, of subjects, ordered with respect to their risk (probability of positivity) for a certain binary characteristic, with their corresponding risk vector given by $\boldsymbol{p} = (p^1, p^2, \cdots, p^N)$, where $p^1 \le p^2 \cdots \le p^N$. Each subject in set S needs to be classified as positive or negative for the binary characteristic through testing. The test is not perfectly reliable, with test *sensitivity* (i.e., true-positive probability) denoted by Se and test *specificity* (i.e., true-negative probability) denoted by Sp. Consequently, misclassification, that is, classifying a truly negative subject as positive (a false-positive classification) or classifying a truly positive subject as negative (a false-negative classification), is possible. We assume, without loss of generality, that the test's true-negative probability is higher than its false-negative probability[1] (i.e., $Sp/(1 - Se) \ge 1$). As such, we have that $Se + Sp - 1 \in [0, 1]$, with the special case of $Se + Sp - 1 = 1$ corresponding to the perfect test case (i.e., false positives and false negatives are not possible), that is, the case studied in the previous literature (Hwang 1975); see Section 1. Both individual testing and group testing are possible, and the test's sensitivity and specificity remain constant with group size. Our modeling also implies that testing outcomes are conditionally independent, given the true-positivity status of the subjects. This is a common assumption in the group testing literature and is mainly made for analytical tractability (e.g., Graff and Roeloffs 1972, Johnson et al. 1991, Kim et al. 2007). We discuss the implications of this assumption in Section 4.2. For the given set, S, of subjects to be classified, the decision maker needs to decide whether each subject is to be tested individually (i.e., with one test per subject and with the subject classified according to the individual test outcome) or in groups, and if the latter, then group sizes and assignments. Each group is to be tested following the Dorfman testing scheme: in the first stage the group is tested with one test; if the group test outcome is negative, then all subjects in the group are classified as negative; and if the group test outcome is positive, then all subjects in the group are individually tested and classified according to their individual test outcome.

Thus, the **decision problem** is to find a feasible *partition* of set S that is optimal with respect to a certain objective function (see Section 2.2). We represent a partition by a combination of mutually disjoint sets, $\boldsymbol{\Omega} = (\Omega_i)_{i=1,\ldots,g}$, each with cardinality $n_i \equiv |\Omega_i|$, for some $g \in \{1, \cdots, N\}$, such that $\bigcup_i \Omega_i = S$, $\Omega_i \cap \Omega_j = \emptyset$, for all $i, j \in \{1, \cdots, g\} : i \ne j$; and each subject in $\Omega_i : n_i = 1$ is individually tested, and each set of subjects in $\Omega_i : n_i > 1$ is tested according to a Dorfman testing scheme with a group size of n_i. We define $\Omega^I \equiv \bigcup_{i:n_i=1} \Omega_i$ and $\Omega^G \equiv \bigcup_{i:n_i>1} \Omega_i$ (i.e., the set of subjects to be tested individually and to be tested in groups, respectively).

Our work focuses on testing facilities that are capable of dynamically changing their testing scheme frequently (e.g., each day), according to the risk vector realization of the testing population in each period. This type of dynamic testing is feasible, for example, in settings that utilize automated testing machines. For instance, in public health screening, many testing facilities use automated molecular testing machines (e.g., HOLOGIC 2017 and Roche Diagnostics USA 2017) to conduct screening tests for various diseases (e.g., chlamydia, gonorrhea, HIV). These machines can perform both individual and group testing and are capable of performing such dynamic testing schemes.

The objective functions in Section 2.2 are based on the following random variables: number of false-positive classifications ($FP(\boldsymbol{\Omega})$), number of false-negative classifications ($FN(\boldsymbol{\Omega})$), and number of tests to be performed ($T(\boldsymbol{\Omega})$) for a partition $\boldsymbol{\Omega}$ of set S. Let I^m denote the indicator random variable corresponding to the true-positive status of subject $m \in S$; and for a partition $\boldsymbol{\Omega}$, let $FN^m(\boldsymbol{\Omega})$ and $FP^m(\boldsymbol{\Omega})$, $m \in S$, denote the indicator random variables, respectively corresponding to the false-negative classification and false-positive classification of subject m, that is, $I^m, FN^m(\boldsymbol{\Omega}), FP^m(\boldsymbol{\Omega}) = 1$, if subject m is, respectively, truly positive, classified falsely as negative, or classified falsely as positive; and 0 otherwise. Similarly, let $N_i^+(\Omega_i)$, $FN_i(\Omega_i)$, and $FP_i(\Omega_i)$ respectively denote the counterparts of these random variables for group $i, \forall i$ (i.e., number of true-positive subjects, number of false-negative classifications, and number of false-positive classifications in group i); that is,

$$N_i^+(\Omega_i) = \sum_{m \in \Omega_i} I^m, \quad FN_i(\Omega_i) = \sum_{m \in \Omega_i} FN^m(\Omega_i), \text{ and}$$
$$FP_i(\Omega_i) = \sum_{m \in \Omega_i} FP^m(\Omega_i), \ \forall i.$$

We also let $T_i(\Omega_i)$ denote the random number of tests performed for group $i, \forall i$. Then, the performance

measures, corresponding to a partition $\mathbf{\Omega}$ of set S, can be expressed as follows:

$$FN(\mathbf{\Omega}) = \sum_i FN_i(\Omega_i), \quad FP(\mathbf{\Omega}) = \sum_i FP_i(\Omega_i), \text{ and}$$
$$T(\mathbf{\Omega}) = \sum_i T_i(\Omega_i).$$

To simplify the subsequent notation, we drop the arguments in parentheses when clear from context.

2.2. Models

As discussed above, we consider the decision problem, of finding an optimal partition, $\mathbf{\Omega}^* = (\Omega_i^*)_i$, under different objective functions and in different settings; the latter is characterized by the presence or absence of a constraint on the testing budget.

2.2.1. System-Optimal Model. In the system-optimal model (**SM**), we consider the problem from a system's perspective, with an objective of minimizing the system-wide cost associated with the binary characteristic, that is, we consider a weight (cost) associated with each test, each false-negative classification (e.g., consequences of the disease when not detected), and each false-positive classification (e.g., follow-up testing cost or hassle, which is unnecessary), that is, the goal is to identify a partition that minimizes a weighted sum of both types of classification errors and the testing cost. Such an objective function has been analyzed before (e.g., Malinovsky et al. 2016) but not within a heterogeneous population framework. Problem **SM** applies in a wide variety of settings, for example in the context of a single-payer healthcare system or a centrally managed production system.

The **SM** is as follows:

$$\begin{aligned} \underset{\mathbf{\Omega}}{\text{minimize}} \quad & \lambda_1 \mathbb{E}\left[FN(\mathbf{\Omega})\right] + \lambda_2 \mathbb{E}\left[FP(\mathbf{\Omega})\right] \\ & + (1-\lambda_1-\lambda_2)\mathbb{E}[T(\mathbf{\Omega})], \end{aligned} \tag{1}$$

where parameters $\lambda_1, \lambda_2 \in [0,1]$ represent the weight the decision maker places on each objective, with special cases corresponding to the minimization of the expected number of false-negative classifications only ($\lambda_1 = 1$), expected number of false-positive classifications only ($\lambda_2 = 1$), and expected number of testing cost only ($\lambda_1 = \lambda_2 = 0$); as discussed in Section 1, the last case is the case most studied in the literature (e.g., Dorfman 1943, Hwang 1975, Saraniti 2006, and McMahan et al. 2012).

2.2.2. Budget-Constrained Model. As opposed to the setting above, in the budget-constrained model (**BM**) we consider a decision maker that must perform testing under a testing budget constraint, and the objective is to minimize a weighted sum of both types of classification errors. This applies, for example, in the context of a testing laboratory that is constrained by the available resources (e.g., a testing budget), which we represent in terms of parameter B, corresponding to the number of tests that can be conducted.

The **BM** is as follows:

$$\begin{aligned} \underset{\mathbf{\Omega}}{\text{minimize}} \quad & \lambda \mathbb{E}\left[FN(\mathbf{\Omega})\right] + (1-\lambda)\mathbb{E}\left[FP(\mathbf{\Omega})\right] \\ \text{subject to} \quad & \mathbb{E}[T(\mathbf{\Omega})] \le B, \end{aligned} \tag{2}$$

where parameter $\lambda \in [0,1]$ represents the weight the decision maker places on each type of classification error, with special cases corresponding to the minimization of the expected number of false negatives only ($\lambda = 1$) and expected number of false positives only ($\lambda = 0$).

Remark 1. In some settings, additional confirmatory testing is conducted on all subjects that test positive in the initial screening (see Section 5.2). This can be easily incorporated into the **BM** formulation by adjusting the budget constraint to include the additional (expected) cost of confirmatory testing; and all the subsequent results continue to hold under this new formulation.

In general, the partitioning problem, of determining $\mathbf{\Omega}$, under an arbitrary objective function is *NP*-hard (Chakravarty et al. 1982), and enumeration-based methods may lead to highly inefficient solution techniques even for small problem instances (e.g., when $N = 20$, the number of possible partitions is approximately 52 trillion, whereas realistic problem instances often have hundreds of subjects; e.g., see Section 5). Therefore, in the remainder of the paper we develop important structural properties for each optimization problem. These properties allow us to develop efficient algorithms and analyze their computational complexity.

3. Derivations of the Performance Measures

We first derive expressions for the performance measures, including the expected number of false negatives, false positives, and tests. Equity-based measures are discussed in Section 4.3. Recall that, for a partition $\mathbf{\Omega}$, Ω^I and Ω^G respectively correspond to the sets of subjects to be tested individually and in groups, and n_i denotes the size of group i, $\forall i$.

3.1. False-Negative Classifications

Recall that in individual testing, a truly positive subject is falsely classified as negative if the test outcome is negative, whereas in group testing, a truly positive subject is falsely classified as negative if (i) the group test outcome is negative, or (ii) the group test outcome is positive and the subject's subsequent individual test

outcome is negative. Then, given $\mathbf{\Omega}$, for any subject $m \in S$, we have

$$\begin{aligned}\mathbb{E}[FN^m] &= \mathbb{E}[FN^m|I^m = 1]P(I^m = 1) \\ &\quad + \mathbb{E}[FN^m|I^m = 0]P(I^m = 0) \\ &= \begin{cases} (1 - Se)p^m + 0, & \text{if } m \in \Omega^I, \\ (Se(1 - Se) + (1 - Se))p^m + 0, & \text{if } m \in \Omega^G, \end{cases}\end{aligned}$$

leading to: $$\mathbb{E}[FN^m] = \begin{cases} (1 - Se)p^m, & \text{if } m \in \Omega^I, \\ (1 - Se^2)p^m, & \text{if } m \in \Omega^G. \end{cases}$$

Then, the expected number of false-negative classifications for group i is given by

$$\mathbb{E}[FN_i(\Omega_i)] = \begin{cases} (1 - Se)\sum_{m\in\Omega_i} p^m, & \text{if } n_i = 1, \\ (1 - Se^2)\sum_{m\in\Omega_i} p^m, & \text{otherwise,} \end{cases}$$

and the expected number of false-negative classifications for all subjects in set S is given by

$$\begin{aligned}\mathbb{E}[FN(\mathbf{\Omega})] &= \sum_i \mathbb{E}[FN_i(\Omega_i)] \\ &= \sum_{i:n_i=1} \mathbb{E}[FN_i(\Omega_i)] + \sum_{i:n_i>1} \mathbb{E}[FN_i(\Omega_i)] \\ &= (1 - Se)\sum_{m\in\Omega^I} p^m + (1 - Se^2)\sum_{m\in\Omega^G} p^m. \end{aligned} \tag{3}$$

Interestingly, for a grouped subject, m, $\mathbb{E}[FN^m]$ is independent of the risk of the remaining subjects in the group. This behavior follows for two reasons: (i) subjects are independent of one another (i.e., knowledge of the true status of one subject does not alter the risk of another), and (ii) conditioned on subject m, in group i, being positive, the probability that group i tests positive is Se, regardless of the status of the remaining subjects in the group. This follows under our assumption that the test sensitivity and specificity (i.e., Se and Sp) are independent of the group size, which implies that, conditional on the true status of subject m, $m \in \Omega^G$, the group and individual test outcomes for subject m are independent of one another. This is a common assumption in the group testing literature; to relax this assumption, one can, for example, explicitly model the dilution effect of grouping (Wein and Zenios 1996). This is an interesting research direction but is beyond the scope of this paper.

These observations lead to an important property, discussed in Remark 2.

Remark 2. For any partition $\mathbf{\Omega}$, the expected number of false-negative classifications corresponding to the set of subjects that are grouped (i.e., in set Ω^G) depends only on set Ω^G and not on *how* the subjects are grouped.

Remark 2 will allow us to develop a polynomial-time algorithm for a special case of the problem; see Section 4.2.

3.2. False-Positive Classifications

Recall that in individual testing, a truly negative subject is falsely classified as positive if the test outcome is positive, whereas in group testing, a truly negative subject is falsely classified as positive if the group test outcome is positive and the subject's subsequent individual test outcome is positive. Then, given a partition $\mathbf{\Omega}$, for any individually tested subject $m \in \Omega^I$, we can write

$$\begin{aligned}\mathbb{E}[FP^m] &= \mathbb{E}[FP^m|I^m = 1]P(I^m = 1) \\ &\quad + \mathbb{E}[FP^m|I^m = 0]P(I^m = 0) \\ &= 0 + (1 - Sp)(1 - p^m),\end{aligned}$$

and for any subject $m \in \Omega^G$ grouped in some set $\Omega_i : n_i > 1$, $i \in \{1, \cdots, g\}$ (i.e., $m \in \Omega_i$), we have

$$\begin{aligned}&\mathbb{E}[FP^m] \\ &= \mathbb{E}[FP^m|I^m = 1]P(I^m = 1) + \mathbb{E}[FP^m|I^m = 0]P(I^m = 0) \\ &= 0 + \Bigg[(1 - Sp)^2 \prod_{k\in\Omega_i\setminus\{m\}} (1 - p^k) \\ &\quad + Se(1 - Sp)\left(1 - \prod_{k\in\Omega_i\setminus\{m\}} (1 - p^k)\right)\Bigg](1 - p^m) \\ &= (1 - Sp)\left[Se - (Se + Sp - 1)\prod_{k\in\Omega_i\setminus\{m\}} (1 - p^k)\right](1 - p^m) \\ &= (1 - Sp)Se(1 - p^m) \\ &\quad - (1 - Sp)(Se + Sp - 1)\prod_{k\in\Omega_i} (1 - p^k),\end{aligned}$$

leading to

$$\mathbb{E}[FP^m] = \begin{cases} (1 - Sp)(1 - p^m), & \text{if } m \in \Omega^I, \\ (1 - Sp)Se(1 - p^m) - (1 - Sp) \\ \cdot (Se + Sp - 1)\prod_{k\in\Omega_i}(1 - p^k), & \text{if } m \in \Omega^G. \end{cases}$$

Then, the expected number of false-positive classifications for group i is given by

$$\mathbb{E}[FP_i(\Omega_i)] = \begin{cases} (1 - Sp)\sum_{m\in\Omega_i}(1 - p^m), & \text{if } n_i = 1, \\ (1 - Sp)Se\sum_{m\in\Omega_i}(1 - p^m) - n_i(1 - Sp) \\ \cdot (Se + Sp - 1)\prod_{m\in\Omega_i}(1 - p^m), & \text{otherwise,} \end{cases}$$

and the expected number of false-positive classifications for all subjects in set S is given by $\mathbb{E}[FP(\mathbf{\Omega})] = \sum_i \mathbb{E}[FP_i(\Omega_i)]$.

3.3. Number of Tests

Recall that in individual testing, the number of tests per subject is always one, whereas in group testing, the number of tests depends on the outcome of the group test: if the group test outcome is negative, then only one test is performed for the entire group, and if the group test outcome is positive, then an additional

individual test is performed for each subject in the group. Given a partition $\mathbf{\Omega}$, the expected number of tests for group i, $i = \{1, \cdots, g\}$, is 1 if $n_i = 1$ (i.e., individual testing), and if $n_i > 1$, we can write

$$\begin{aligned}\mathbb{E}[T_i(\Omega_i)] &= \sum_{k=0}^{n_i} \mathbb{E}[T_i(\Omega_i)|N_i^+(\Omega_i) = k]P(N_i^+(\Omega_i) = k)\\ &= \mathbb{E}[T_i(\Omega_i)|N_i^+(\Omega_i) = 0]P(N_i^+(\Omega_i) = 0)\\ &\quad + \sum_{k=1}^{n_i} \mathbb{E}[T_i(\Omega_i)|N_i^+(\Omega_i) = k]P(N_i^+(\Omega_i) = k)\\ &= (Sp + (1 - Sp)(1 + n_i))P(N_i^+(\Omega_i) = 0)\\ &\quad + \sum_{k=1}^{n_i}(1 - Se + Se(1 + n_i))P(N_i^+(\Omega_i) = k)\\ &= 1 + n_i\Big(Se - (Se + Sp - 1)\prod_{m\in\Omega_i}(1 - p^m)\Big).\end{aligned}$$

Thus, $\mathbb{E}[T_i(\Omega_i)]$

$$= \begin{cases} 1, & \text{if } n_i = 1,\\ 1 + n_i\Big(Se - (Se + Sp - 1)\prod_{m\in\Omega_i}(1 - p^m)\Big), & \\ & \text{otherwise,}\end{cases} \tag{4}$$

and the expected number of tests needed for all subjects in set S is given by $\mathbb{E}[T(\mathbf{\Omega})] = \sum_i \mathbb{E}[T_i(\Omega_i)]$.

4. Structural Properties and Algorithms

As discussed earlier, the partitioning problem under an arbitrary objective function is *NP*-hard (Chakravarty et al. 1982). Therefore, in what follows, we develop important structural properties of the two optimization problems, **SM** and **BM**, presented in Section 2.2. These properties allow us to reduce the partitioning problem into network flow problems and analyze their computational complexity. In many instances the resulting network flow problems can be solved with algorithms whose complexity is polynomial in problem size, that is, N, the number of subjects in set S.

Definition 1. A partition, $\mathbf{\Omega} = (\Omega_i)_{i=1,\cdots,g}$, is said to be an ordered partition if it follows the ordered set $S = \{1, 2, \cdots, N\}$, that is, $\Omega_1 = \{1, \cdots, n_1\}$, $\Omega_2 = \{n_1 + 1, \cdots, n_1 + n_2\}, \cdots, \Omega_g = \{\sum_{i=1}^{g-1} n_i + 1, \cdots, N\}$, for some $g \in \{1, \cdots, N\}$ and $n_i \in Z^+, i = 1, \cdots, g$.

By this definition, an ordered partition $\mathbf{\Omega} = (\Omega_i)_i$ can be equivalently expressed in terms of the group size vector, $\mathbf{n} = (n_i)_i$, as groups are constructed following the ordered set S. In the following, we first present our main results that hold for both **SM** and **BM** and then derive additional properties for each problem respectively in Sections 4.1 and 4.2. All proofs can be found in the online appendix.

Theorem 1. *For* **SM** *and* **BM**, *the following properties hold in an optimal solution:*

a. *There exists an optimal partition that is an ordered partition of S.*

b. *If in the optimal ordered partition, subject m, with risk p^m, is individually tested, then it is optimal to individually test all subjects having a risk higher than p^m.*

The first part of Theorem 1 allows us to reformulate the partitioning problem as a network flow problem defined on the network in Definition 2, whereas the second part of Theorem 1 enables us to improve the computational complexity of the proposed algorithms for certain special cases.

Definition 2. For a problem instance with N subjects ($N \in \mathbb{Z}^+$) in set S, let $G = (V, E)$ denote an acyclic directed graph with vertex set $V = \{1, \cdots, N + 1\}$ and edge set $E = \{(i, j) \in V : i < j\}$, with cardinality, $|E| = N(N + 1)/2$.

Figure 1 depicts an example of $G = (V, E)$ for $N = 10$, where the bold end of an edge represents its direction (i.e., the flow is directed toward the bold end).

Remark 3. For a problem instance with N subjects ($N \in \mathbb{Z}^+$) in set S, each path from vertex 1 to vertex $N + 1$ in network $G = (V, E)$ corresponds to an ordered partition of set S, and the number of unique paths is given by 2^{N-1}. Further, $G = (V, E)$ is a dense graph, with the degree of each vertex given by N (i.e., $deg(v) = N$ for all $v \in V$).

To derive the number of paths given in Remark 3, let $Path(N)$ denote the number of paths from vertex 1 to vertex $N + 1$ in $G = (V, E)$. We have that

$$Path(N) = 1 + Path(N - 1) + Path(N - 2) + \cdots + Path(1), \tag{5}$$

which follows because from vertex 1, one can directly go to vertex $N + 1$, or go to vertex 2 and then go to $N + 1$ (in the latter case the number of possible paths from vertex 2 to $N + 1$ equals $Path(N - 1)$), and so on. Rearranging Equation (5) and noting that $Path(1) = 1$, we have

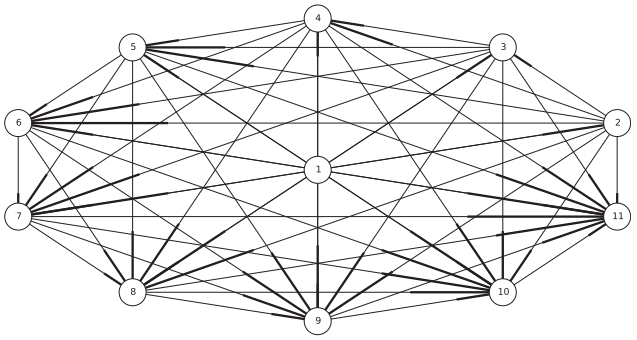
$$\begin{aligned}Path(N) &= Path(N - 1)\\ &\quad + (1 + Path(N - 2) + \cdots + Path(1))\\ &= 2Path(N - 1) = 2^{N-1}.\end{aligned}$$

The fact that $G = (V, E)$ is a dense graph follows by definition, that is, a graph is said to be dense if $\min_{v\in V}\{deg(v)\} \geq N/2$ (Gimbel et al. 1993); this will play an important role for the construction of an algorithm for a special case of the problem (see Section 4.2). Theorem 1 leads to the following results.

Property 1.

1. **SM** can be formulated as a shortest path (SP) problem defined on $G = (V, E)$, with edge costs given

Figure 1. $G = (V, E)$ when $N = 10$



by $c_{ij} = \lambda_1 \mathbb{E}[FN_i(\Omega_{i-j})] + \lambda_2 \mathbb{E}[FP_i(\Omega_{i-j})] + (1 - \lambda_1 - \lambda_2)\mathbb{E}[T_i(\Omega_{i-j})]$, where $\Omega_{i-j} = \{i, \cdots, j-1\}$, that is, c_{ij} is the cost of utilizing group Ω_{i-j}, for all $(i, j) \in E$.

2. **BM** can be formulated as a constrained-SP problem, having a single constraint, defined on $G = (V, E)$, with edge costs given by $c_{ij} = \lambda \mathbb{E}[FN_i(\Omega_{i-j})] + (1 - \lambda)\mathbb{E}[FP_i(\Omega_{i-j})]$, where $\Omega_{i-j} = \{i, \cdots, j-1\}$, that is, c_{ij} is the cost of utilizing group Ω_{i-j}, for all $(i, j) \in E$.

Remark 4.

1. The SP problem for an acyclic graph can be solved in polynomial time (e.g., via a topological sorting algorithm in $\mathbb{O}(|V| + |E|)$; Cormen 2009). As such, a topological sorting algorithm solves **SM** with N subjects in $\mathbb{O}(N^2)$.

2. The constrained-SP problem is, in general, *NP*-hard (Garcia 2009).

Thus, **SM** can be solved in polynomial time, resolving the intractability conjecture stated in the literature (McMahan et al. 2012). However, **BM** is a difficult problem. In the remainder of the paper we develop structural properties of **SM** and **BM** that enable us to improve their computational efficiency and to develop an algorithm that can solve an important special case of **BM** in polynomial time.

4.1. Analysis of SM

We next study structural properties for important special cases of **SM**, that is, those that seek to minimize a subset of the expected number of false classifications or to minimize solely the expected number of tests. Each of these objectives can be important for the decision maker, depending on the setting. For example, Bilder and Tebbs (2012), McMahan et al. (2012), and Black et al. (2015) exclusively focus on the special case of $\lambda_1 = \lambda_2 = 0$ (i.e., minimization of the expected number of tests) and develop various heuristics.

Theorem 2. *Consider the following special cases of* **SM** *that respectively minimize the expected number of a weighted combination of both types of classification errors, or number of false negatives only, or number of false positives only.*

1. *For all* $\lambda_1, \lambda_2 \in [0, 1] : \lambda_1 + \lambda_2 = 1$ *(i.e., when minimizing a weighted sum of both types of classification errors),*

   a. *the optimal ordered partition does not contain a group having more than three subjects, that is, each group is comprised of one, two, or three subjects (i.e.,* $n_i^* \le 3$*, for all* $i = 1, \cdots, g$*, for some* $g \in \{1, \cdots, N\}$*);*

   b. *if* $p^N \le 1/3$*, then the group sizes of the optimal ordered partition are in nonincreasing order, that is,* $n_1^* \ge n_2^* \ge \cdots \ge n_g^*$*, for some* $g \in \{1, \cdots, N\}$*;*

2. *for* $\lambda_1 = 1$ *(i.e., when minimizing* $\mathbb{E}[FN]$*), the optimal partition is to individually test each subject, that is,* $n_i^* = 1$*,* $i = 1, \cdots, N$*;*

3. *for* $\lambda_2 = 1$ *(i.e., when minimizing* $\mathbb{E}[FP]$*), the optimal ordered partition can have at most one individual test, which, by the second part of Theorem 1, has to be for the highest risk subject (i.e., subject N).*

Remark 5. For all $\lambda_1, \lambda_2 \in [0, 1] : \lambda_1 + \lambda_2 = 1$, by Theorem 2, the number of edges in $G = (V, E)$, $|E|$, decreases from $N(N+1)/2$ to $3(N-1)$. As such, a topological sorting algorithm solves **SM** with N subjects in $\mathbb{O}(|V| + |E|) = \mathbb{O}(N)$, that is, in linear time.

Similarly, the following result characterizes properties of the optimal **SM** solution that minimizes the expected number of tests.

Theorem 3. *Consider a special case of* **SM** *that minimizes* $\mathbb{E}[T]$ *(i.e.,* $\lambda_1 = \lambda_2 = 0$*). If*

$$p^N \le 1 - \left(\frac{Se - 0.5}{Se + Sp - 1}\right)^{1/2},$$

then the optimal ordered partition can have at most one individual test, which, by the second part of Theorem 1, has to be for the highest risk subject (i.e., subject N).

We expect the condition imposed in Theorem 3 to be satisfied when, for example, the prevalence of the binary characteristic is low and the test specificity is high. As an example, in our case study (see Section 5), this condition reduces to $p^N \le 0.308$, which is satisfied by all subjects in the case study.

Remark 6. If $\lambda_1 = \lambda_2 = 0$ and $p^N \le 1 - ((\text{Se} - 0.5)/(\text{Se} + \text{Sp} - 1))^{(1/2)}$, then the number of edges in $G = (V, E)$, $|E|$, decreases from $N(N+1)/2$ to $1 + N(N-1)/2$, improving the computational complexity of the SP algorithm for **SM**.

Theorem 2 establishes that for a special case of **SM** that minimizes the expected number of misclassifications (i.e., $\lambda_1 + \lambda_2 = 1$), and under a certain condition (i.e., $p^N \le 1/3$), the group sizes for an optimal ordered partition are nonincreasing. This property may seem intuitive because it indicates that higher-risk subjects are placed in smaller groups than lower-risk subjects. Thus, the next question is whether the optimal group

sizes continue to be nonincreasing when this condition is not satisfied or when the objective in **SM** contains the expected number of tests (i.e., $\lambda_1 + \lambda_2 < 1$). The special case of $\lambda_1 = \lambda_2 = 0$ (i.e., the minimization of the expected number of tests) is the objective almost exclusively studied in the literature, and various heuristics are proposed that generate ordered partitions with nonincreasing group sizes (McMahan et al. 2012). The following counter-example, which is based on realistic problem parameters, indicates that this property does not necessarily hold in general.

Example 1. Consider **SM** with $\lambda_1 = \lambda_2 = 0$ (i.e., the objective is to minimize $\mathbb{E}[T]$). Consider a test with $Se = 0.90$ and $Sp = 0.95$, and a set, S, of 100 subjects (i.e., $N = 100$), with an ordered risk vector, $\boldsymbol{p}$, given by $p_{i+1} = p_i + \beta$, $i = 1, \cdots, 99$, where $\beta = 13/3{,}300$ and $p_1 = 0.01$. The optimal partition that minimizes $\mathbb{E}[T]$ is given by $\boldsymbol{n}^* = (7, 6, 5, 4, 4, 4, 4, 4, 4, 3, 3, 3, 3, 3, 3, 3, 3, 34)$ with $\mathbb{E}[T] = 74.48$, which does not satisfy the property of nonincreasing group sizes.

Example 1 serves two main purposes. First, although placing higher-risk subjects in smaller groups seems intuitive, Example 1 demonstrates that this is not always the case in an optimal solution; that is, an optimal solution does not necessarily follow the nonincreasing group size property. Second, Example 1 further shows that the minimization of the expected number of tests, by itself, is not an adequate objective function for this problem, because the imperfect test sensitivity and specificity parameters may lead to optimal solutions that may be impractical, or simply undesirable from a classification accuracy maximization perspective. To see this last point, notice that in the optimal solution of Example 1, the 34 highest-risk subjects are placed into a single group, mainly because the test sensitivity is 0.90. Thus, when the objective is to minimize the expected number of tests only (i.e., without any consideration of classification accuracy), grouping such high-risk subjects may be preferable owing to the proportion of time the high-risk group is expected to test negative (resulting in only one test for the group). Even when an upper limit is set on the maximum group size, we still observe this behavior, providing further evidence for the need to consider classification accuracy in the objective function.

4.2. Analysis of BM

Next we study structural properties of **BM**, in which the objective is to minimize a weighted sum of both types of classification errors under a testing budget constraint; see Equation (2). By Property 1, we formulate **BM** as a constrained-SP having a single constraint, which, by Remark 4, is *NP*-hard. Various methods are proposed in the literature to solve the constrained-SP problem, as we briefly discuss below (for a thorough review, see Garcia 2009).

One method to solve the constrained-SP problem is the *path ranking method*, in which the next unconstrained shortest path is successively generated until the first feasible path (i.e., that satisfies the constraints) is identified. The problem of generating the next shortest path is related to the k-shortest path problem, in which the objective is to generate the first k (unconstrained) shortest paths for a given $k \in Z^+$. When k is fixed, there exist polynomial-time algorithms that generate the k shortest paths (Garcia 2009). However, when the k-SP problem is used to solve the constrained-SP problem, k is no longer fixed, and hence there is no guarantee of a polynomial-time algorithm (Garcia 2009). This is especially true in our case, because the number of paths grows exponentially with N (see Remark 3), rendering this method inefficient and computationally expensive. Alternatively, one can utilize a Lagrangean relaxation-based method to solve the constrained-SP problem (e.g., Handler and Zang 1980); specifically, when the constraints are relaxed, the problem reduces to an SP problem, which can be solved in polynomial time. In particular, Handler and Zang (1980) solve the Lagrangean relaxation of the problem to determine upper and lower bounds, and then, using a k-SP algorithm (e.g., Yen 1971) closes the gap until optimality or δ-optimality is attained. Although such methods are, in general, more efficient than path-ranking methods, they may still require a large number of iterations to converge to the optimal solution, especially when the number of paths is large, as in our case.

In what follows, we first analyze an important special case of **BM** that minimizes the expected number of false-negative classifications under a testing budget constraint (i.e., $\lambda = 1$). When $\lambda = 1$, Theorem 1 and Remark 2 enable us to develop an algorithm that can solve **BM** in polynomial time. The algorithm is motivated by the properties by which, keeping all else the same, (i) testing any subject individually reduces the objective function value (Equation (3)), (ii) the objective function value is independent of how the subjects in set Ω^G are grouped (Remark 2), and (iii) by Theorem 1, the subjects in set Ω^I must correspond to the highest-risk subjects in set S. Therefore, the proposed algorithm determines the optimal ordered partition by identifying the largest feasible set of subjects that can be tested individually (set Ω^I) and by minimizing the expected number of tests for the remaining subjects (set Ω^G).

Theorem 4. *When $\lambda = 1$, the following algorithm solves* **BM** *for N subjects in $\mathbb{O}(N^3)$:*

Step 0: *If $B \geq N$, stop; the optimal solution is to individually test each subject, that is, $n_i^* = 1$, $i = 1, \cdots, N$.*

Step 1: *Let $\hat{N} = 2$, $S_1 = \{1, 2\}$, and $S_2 = S \backslash S_1$.*

Step 2: *Solve* **SM** *with* $S = S_1$ *and parameters* $\lambda_1 = \lambda_2 = 0$, *and let* $Z^*(S_1)$ *denote the optimal objective function value for set* S_1, *i.e.*, $Z^*(S_1, \mathbf{\Omega}^*(S_1)) = \min_{\mathbf{\Omega}} \{\mathbb{E}[T(S_1; \mathbf{\Omega})]\}$.

Step 3: *If* $Z^*(S_1) + |S_2| \le B$, *stop; the optimal solution is to test the subjects in* S_2 *individually and to test the subjects in* S_1 *according to the optimal ordered partition in Step* 2, *that is*, $\mathbf{\Omega}^*(S_1)$.

Step 4: *If* $\hat{N} = N$ *and* $Z^*(S_1) > B$, *stop; the problem is infeasible.*

Step 5: *Set* $\hat{N} = \hat{N} + 1$, $S_1 = S_1 \cup \{\hat{N}\}$, $S_2 = S \backslash S_1$, *and go to Step* 2.

The algorithm provided in Theorem 4 depends on Remark 2, which in turn depends on the assumption that testing responses are conditionally independent, given the true positivity status of the subjects. In some settings this assumption may not hold, because test outcomes performed on a truly positive subject may be positively correlated. As an example, consider a test that measures the concentration of a disease-related biomarker, which should be high for a truly positive subject (except for the initial phase of the disease, also known as the window period); then if the outcome of the first-stage Dorfman test is positive for a group that includes a truly positive subject, then it is likely that this truly positive subject is outside of the window period, and hence the probability that the outcome of the subsequent individual test conducted on this truly positive subject is positive should be higher. This is not necessarily the case for a truly negative subject, however. In the absence of the disease, the biomarker concentration should be negligible, and hence outcomes of the different tests will only be subject to a random testing error, which need not be correlated; thus, given that a subject is truly negative, we can expect the subsequent test outcomes to be independent. Recall also that in the Dorfman testing scheme, only the groups that test positive in the first-stage Dorfman testing undergo further testing. Then, the aforementioned conditional independence assumption does not impact the subjects belonging to negative-testing groups in the first stage, because a negative-testing group is tested only once. Consequently, we expect that this assumption should not impact (i) the expected number of tests, because the number of tests to be conducted depends only on the outcome of the first-stage group tests; and (ii) the expected number of false positives, because the conditional independence assumption should not impact test outcomes of truly negative subjects, as discussed above. However, the expected number of false negatives in our model is impacted by the conditional independence assumption, and our model provides an upper bound on the expected number of false negatives compared with the case in which this assumption is relaxed. This follows because, by assuming conditional independence, we are underestimating the probability that the outcome of the second-stage individual test is positive, given that the subject is truly positive and that the first-stage test outcome is positive for the group that includes this subject.

Remark 7. When $\lambda = 1$, the algorithm presented in Theorem 4 not only determines an optimal solution to **BM** (i.e., that minimizes $\mathbb{E}[FN]$) but also provides a solution with the minimum expected number of tests among multiple optimal solutions (if any). This property is not guaranteed by other algorithms, such as Handler's algorithm, which generate any one of the optimal solutions.

Having developed a polynomial-time algorithm for **BM** when $\lambda = 1$, we next explore solving **BM** when $\lambda < 1$. In particular, we formulate **BM** as a binary integer programming problem. Note that the total unimodularity property, present in the integer programming formulation of the unconstrained SP problem, no longer holds with the addition of the budget constraint. Therefore, in the following we exploit the structure of an optimal solution to **BM**. In particular, by Theorem 1(b), one can add a set of constraints to the integer programming formulation of **BM** that reduces the feasible region without cutting off the optimal solution, as stated in the following lemma.

Lemma 1. *By Theorem* 1, *the following set of constraints does not cut off the optimal solution to* **BM**:

$$x_{j,j+1} \ge x_{i,i+1}, \ \forall\, (i,j) \in E : j > i.$$

In light of Lemma 1, the integer programming formulation of **BM** follows:

$$\begin{aligned}
\underset{x}{\text{minimize}} \quad & \sum_{i=1}^{N} \sum_{j=i+1}^{N+1} \Big(\lambda \mathbb{E}\left[FN_i(\Omega_{i-j})\right] \\
& \qquad + (1-\lambda)\mathbb{E}\left[FP_i(\Omega_{i-j})\right] \Big) x_{ij} \\
\text{subject to} \quad & \sum_{j=i+1}^{N+1} x_{ij} - \sum_{j=1}^{i-1} x_{ji} \\
& \qquad = \begin{cases} 1, & \text{if } i = 1 \\ -1, & \text{if } i = N+1 \\ 0, & \text{otherwise} \end{cases} && \forall i \in V \\
& \sum_{j=i+1}^{N+1} x_{ij} \le 1 && \forall i \in V \\
& x_{j,j+1} \ge x_{i,i+1} && \forall\, (i,j) \in E : j > i \\
& \sum_{i=1}^{N} \sum_{j=i+1}^{N+1} \mathbb{E}[T_i(\Omega_{i-j})]\, x_{ij} \le B \\
& x_{ij} \in \{0,1\} && \forall (i,j) \in E,
\end{aligned} \tag{6}$$

where $\Omega_{i-j} = \{i, \cdots, j-1\}$, $\forall i$. Our numerical study shows that, in general, the branch and bound approach used to solve the formulation in Beasley and

Christofides (1989), which includes the additional constraints in Lemma 1, outperforms the algorithm of Handler and Zang (1980) in terms of efficiency, allowing us to solve considerably larger problem instances to optimality.

4.3. An Equity-Based Objective and Properties

As mentioned in Section 1, an important consideration in resource allocation problems, especially in public health screening, is the trade-off between classification accuracy and equity. For example, does the most accurate solution unfairly increase the misclassification probability for certain subjects? If so, this solution would be unfair because certain subjects, on the basis of their demographics, which may impact risk, would not benefit from the improved solution but instead would be more likely to be misclassified. Therefore, we would like to understand whether such inequity occurs in the proposed solutions, and if so, to what degree (i.e., what is the trade-off between accuracy and equity). Toward this end, we capture the trade-off between accuracy and equity by adopting the *α-fairness* measure, commonly used in the literature (e.g., Atkinson 1970, Barr 1998, Bertsimas et al. 2012). The α-fairness measure is a function of parameter $\alpha \geq 0$, known as the *inequality aversion parameter*, that measures the tendency of the model to produce a more equitable solution (over a more accurate solution); see, for example, Bertsimas et al. (2012). The equity-based version of **BM**, which we refer to as **BM-E**(α), is as follows:

$$\begin{aligned} \underset{\boldsymbol{\Omega}}{\text{maximize}} \quad & \frac{1}{(1-\alpha)} \sum_{m \in S} \Big(1 - \lambda \mathbb{E}\left[FN^m(\boldsymbol{\Omega})\right] \\ & \qquad - (1-\lambda)\mathbb{E}\left[FP^m(\boldsymbol{\Omega})\right]\Big)^{1-\alpha} \\ \text{subject to} \quad & \mathbb{E}[T(\boldsymbol{\Omega})] \leq B. \end{aligned} \tag{7}$$

Remark 8. Consider the objective function in **BM-E**(α):

1. When $\alpha = 0$, the objective function reduces to the minimization of a weighted sum of both types of classification errors (i.e., the most accurate solution), that is, **BM-E**(α) reduces to **BM**.
2. As α increases, the objective function assigns more weight to equity (Barr 1998, Lan et al 2010).
3. As $\alpha \to \infty$, the objective function reduces to the most equitable function (Kalai and Smorodinsky 1975), that is, of minimizing the worst-case equity outcome given by,

$$\underset{\boldsymbol{\Omega}}{\text{minimize}} \quad \max_{m \in S}\{\lambda \mathbb{E}\left[FN^m(\boldsymbol{\Omega})\right] + (1-\lambda)\mathbb{E}\left[FP^m(\boldsymbol{\Omega})\right]\}.$$

Definition 3. Following Bertismas et al. (2011), we define the *Price of Fairness*, denoted by $PoF(\alpha)$, as the relative increase in the weighted sum of classification errors under the optimal equitable solution, with fairness level α, compared with the most accurate solution (i.e., when $\alpha = 0$), that is

$$PoF(\alpha) = \frac{\begin{array}{c}\lambda\Big(\mathbb{E}\left[FN(\boldsymbol{\Omega}^*(\alpha))\right] - \mathbb{E}\left[FN(\boldsymbol{\Omega}^*(0))\right]\Big) \\ + (1-\lambda)\big(\mathbb{E}\left[FP(\boldsymbol{\Omega}^*(\alpha))\right] - \mathbb{E}\left[FP(\boldsymbol{\Omega}^*(0))\right]\Big)\end{array}}{\lambda \mathbb{E}\left[FN(\boldsymbol{\Omega}^*(0))\right] + (1-\lambda)\mathbb{E}\left[FP(\boldsymbol{\Omega}^*(0))\right]}.$$

By Remark 8, **BM-E**($\alpha = 0$) reduces to **BM**, and all results of Section 4.2 follow. Therefore, in what follows we discuss the case in which $\alpha > 0$. The following example shows that when $\alpha > 0$ and $\lambda < 1$, that is, the objective is to maximize an equity-based objective with respect to either both false negatives and false positives, or false positives only, the optimal partition to **BM-E**(α) need not be an ordered partition, that is, Theorem 1 does not necessarily hold under the equity-based objective.

Example 2. Consider **BM-E**(α) with $\alpha \to \infty$ and $\lambda = 0$. Suppose that $B = 5$, and consider a test with $Se = 0.90$ and $Sp = 0.95$, and a set, S, of five subjects with the following risk vector:

$$\boldsymbol{p} = (0.10, 0.28, 0.30, 0.40, 0.45).$$

The optimal partition is to have two groups, with the first group containing the lowest- and highest-risk subjects (i.e., subjects 1 and 5) and the second group containing the remaining three subjects (i.e., $\Omega^* = \{\{1,5\},\{2,3,4\}\}$). Clearly, the optimal partition is not ordered. Table 1 reports the performance of the optimal partition with two ordered partitions that have the same group sizes as the optimal partition (i.e., $\boldsymbol{n} = (2,3)$ and $\boldsymbol{n} = (3,2)$), with $\boldsymbol{n} = (3,2)$ corresponding to the best solution among all ordered partitions.

Example 2 demonstrates that the optimal partition need not be an ordered partition. To explain this, first note the following observation.

Table 1. Performance of Various Partitions in Example 2

	$\mathbb{E}[FP^m]$						
Partition ($\boldsymbol{\Omega}$)	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	$\max_{m \in S}\{\mathbb{E}[FP^m]\}$	$\mathbb{E}[FP]$
{1,5},{2,3,4}	0.01946	0.01955[a]	0.01865	0.01415	0.00372	0.01955	0.07552
{1,2},{3,4,5}	0.01296	0.00486	0.02168[a]	0.01718	0.01493	0.02168	0.07162
{1,2,3},{4,5}	0.02122[a]	0.01312	0.01222	0.01298	0.01073	0.02122	0.07027

[a]The subject with the highest probability of being falsely classified as positive.

Remark 9. Consider **BM-E**(α) with $\alpha \to \infty$ and $\lambda = 0$. For any given partition $\mathbf{\Omega}$, the highest false-positive probability in each group is determined by the lowest-risk subject in that group. Therefore, $\max_{m \in S}\{\mathbb{E}[FP^m]\}$ corresponds to the false-positive probability of the lowest-risk subject of one of the groups.

As such, this example demonstrates how the optimal partition balances the maximum value of the false-positive probability by appropriately grouping the subjects, and the optimal grouping does not need to follow an ordered partition.

Remark 10. Because **BM-E**(α) no longer has the property that ensures the existence of an optimal partition that is ordered, it is *NP*-hard (Chakravarty et al. 1982).

However, for an important case of **BM-E**(α) with $\lambda = 1$ and $\alpha \geq 0$ (i.e., the equity-based objective applies only to false-negative classifications), we have the following important result.

Theorem 5. *For* **BM-E**(α) *with* $\lambda = 1$ *and* $\alpha \geq 0$, *there exists an optimal ordered partition that is independent of* α.

When $\lambda = 1$, Theorem 5 follows because, for all $\alpha \geq 0$, (i) testing any subject individually reduces the objective function value (Equation (3)), (ii) the objective function value is independent of how the subjects in set Ω^G are grouped (Remark 2), and (iii) by Theorem 1, the subjects in set Ω^I must correspond to the highest-risk subjects in set S. As such, for all $\alpha \geq 0$, the objective is to determine the largest feasible set of subjects that can be tested individually (set Ω^I), and because the constraint is independent of α, this set will be identical for all α levels. Theorem 5 has important implications, as stated in the following result.

Corollary 1. *Consider* **BM-E**(α) *with* $\lambda = 1$ *and* $\alpha \geq 0$ *(i.e., the equity-based objective applies only to false-negative classifications).*

1. $PoF(\alpha) = 0$ *for all* $\alpha \geq 0$, *that is, in terms of the false-negative classifications, the partition that is the most accurate is also the most equitable.*
2. *For all* $\alpha \geq 0$, *an optimal partition can be obtained by solving BM, for which a polynomial-time algorithm exists (see Theorem 4).*

In the next section, we perform a case study to illustrate the effectiveness of the proposed risk-based testing scheme over optimal non–risk-based schemes and current screening practices.

5. Case Study: Chlamydia Screening in the United States

In this section we perform a case study on chlamydia screening. Chlamydia is one of the most prevalent STDs in the United States (Centers for Disease Control and Prevention Division of STD Prevention 2014), and most chlamydia screening occurs at the state level via public health laboratories. There are no nationwide guidelines on screening practices for chlamydia, and as a result, screening practices differ significantly among states; for example, North Carolina individually screens high-risk female subjects only (North Carolina State Laboratory of Public Health 2016), whereas Idaho uses group testing on all subjects in group sizes of four, with the exception of individual testing of subjects who are exposed to chlamydia or who need to be tested after treatment (Lewis et al. 2012). A study conducted by the Centers for Disease Control and Prevention indicates how the positivity probability (risk) of a subject for chlamydia can vary substantially by gender, race/ethnicity, and age (Centers for Disease Control and Prevention Division of STD Prevention 2014). Consequently, in this case study, we decompose the U.S. population by gender, three race/ethnicity groups (black [B], hispanic [H], and other[2] [O]), and two age groups (15–24 years and other), leading to a total of 12 risk subpopulations. Studies also show that a large percentage of chlamydia cases goes undiagnosed and/or unreported (e.g., 75% of females and 50% of males with chlamydia show no symptoms and are likely to be unreported; Centers for Disease Control and Prevention 2000); and the actual number of cases is estimated to be at least three times the number of reported cases (Grimes et al. 2013). In Table 2, we report the risk (prevalence rate) for chlamydia in the United States and the proportion, in the general population, of each risk subpopulation according to data in Centers for Disease Control and Prevention (2014) for the year 2014 and using an under-reporting factor, denoted by *UP*, of three. In addition, we conduct a one-way sensitivity analysis on the parameter *UP* and also investigate the cases with $UP = 4$ and $UP = 5$. Hence, the mean overall prevalence rate, μ_p, which we use in non–risk-based schemes, is equal to 0.97%, 1.29%, and 1.62%, respectively corresponding to $UP = 3, 4$, and 5.

We consider an amplified DNA assay for chlamydia (Viper ProbeTec Chlamydia Q^x), a commonly used chlamydia screening test that can be utilized for both individual and grouped testing (Kapala et al. 2000). A number of nucleic acid tests are available, with varying sensitivity and specificity values, hence in this case study we explore a set of sensitivity and specificity values ranging from 0.93 to 0.97. We use testing and cost data from Van Der Pol et al. (2012) and Owusu-Edusei et al. (2015). Specifically, we set the cost of a false negative to the average cost of sequelae (i.e., any complications resulting from not treating a chlamydia patient), estimated as \$2,927; the screening cost, per test (either individual or grouped), to \$55; and the cost of a false

Table 2. Risk for Chlamydia and Proportion in Population by Gender, Age, and Race/Ethnicity (Centers for Disease Control and Prevention 2014)

Gender	Race/ethnicity	Age group (years)	Risk (prevalence) (%)	Proportion in general population (%)
Female	Hispanic	15–24	6.54	1.41
		Other	0.65	7.01
	Black	15–24	19.19	1.07
		Other	1.22	5.67
	Other	15–24	4.38	4.29
		Other	0.25	31.31
Male	Hispanic	15–24	1.78	1.53
		Other	0.36	7.16
	Black	15–24	7.45	1.09
		Other	1.05	5.08
	Other	15–24	1.20	4.51
		Other	0.17	29.87

positive to the cost of an additional confirmatory test, which we assume equals the cost of the initial screening test.

To illustrate the benefits of the proposed risk-based Dorfman testing, we perform a Monte Carlo simulation. In particular, we set the number of subjects that need to be tested in a given period (day) to 100 (i.e., $N = 100$); this provides a realistic representation of the problem (Lewis et al. 2012), and for each day, we generate a realization of the random risk vector following the discrete distribution presented in Table 2. We perform 3,000 simulation replications for each scenario, characterized by μ_p, Se, and S_p, and determine the sample mean and sample variance for 3,000 replications for each performance measure. All simulation results in the tables are presented in the form, the point estimate ± the half width of the 95% confidence interval.

In Sections 5.1 and 5.2, we compare each risk-based model, **SM** and **BM**, with a corresponding **base-case (BC)** model in which, following the common treatment of the group testing design in the literature (e.g., Dorfman 1943 and Kim et al. 2007), we assume that the testing population is homogeneous, with mean risk μ_p, and the population size is infinite. Thus, the base-case model generates a *static* group testing design, which is used repetitively every period, whereas the proposed risk-based policies generate *dynamic* testing designs, that is, they produce a potentially different testing design each period based on the observed risk vector for the N subjects. Following current practices, in the base-case, if N is not a multiple of the group size, then the remaining subjects form a (smaller) group for testing, and subjects are randomly assigned to the groups. Then in Section 5.3, we compare **SM** with the three risk-based heuristic testing designs proposed by McMahan et al. (2012), in which the objective is to minimize the expected number of tests. We denote the three heuristics by **MC1**, **MC2**, and **MC3**, respectively corresponding to the *Optimal Dorfman*, *Thresholding*, and *Pool-Specific Optimal Dorfman* Heuristics discussed in McMahan et al. (2012).

5.1. System-Optimal Model (SM)

We first consider the system's problem, of generating a testing design for N subjects so as to minimize the total cost of false-negative and false-positive classifications and testing. In our risk-based testing policy, this corresponds to **SM**, with the objective of minimizing $\lambda_1 \mathbb{E}[FN] + \lambda_2 \mathbb{E}[FP] + (1 - \lambda_1 - \lambda_2)\mathbb{E}[T]$, with weights $\lambda_1 = 0.96$ and $\lambda_2 = 0.02$, which are normalized on the basis of the cost data discussed above. Specifically, $\lambda_1 = \$2,927/\$3,037 \approx 0.96$ and $\lambda_2 = \$55/\$3,037 \approx 0.02$ (hence, $1 - \lambda_1 - \lambda_2 = 0.02$).[3] The corresponding base-case, **BC**, is obtained by minimizing the same objective function, but under the homogeneous and infinite population assumptions (i.e., minimizing the objective function per subject). Table 3 reports the performance measures for **BC** and **SM** for a range of parameter values.

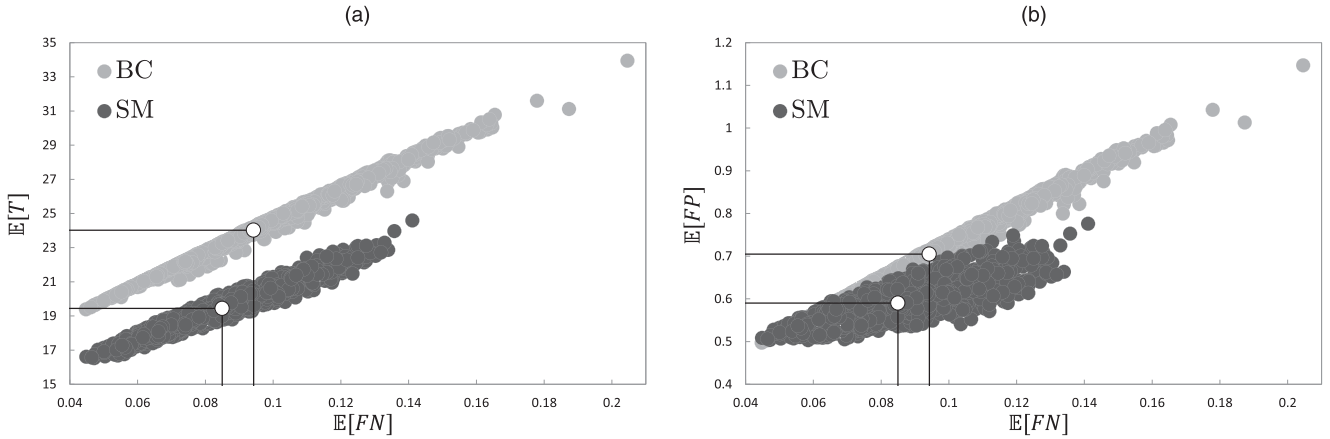
The results in Table 3 indicate the substantial reductions under **SM** for all performance measures over the non–risk-based scheme **BC**. Specifically, **SM** reduces each of $\mathbb{E}[FN]$, $\max_{m \in S}\{\mathbb{E}[FN^m]\}$, $\mathbb{E}[FP]$, and $\mathbb{E}[T]$ over **BC** by an average of 10%, 41%, 17%, and 19%, respectively. Moreover, the objective function of **SM** is reduced by an average of 18% over **BC**. The one-way sensitivity analyses reveal that the reductions in all performance measures are consistent among the different settings, indicating how risk-based testing can provide substantial benefits for a range of parameter values. Figure 2 also plots (a) $\mathbb{E}[T]$ and (b) $\mathbb{E}[FP]$ as a function of $\mathbb{E}[FN]$ for each of **BC** and **SM**; each point in the figure represents one of the 3,000 simulation replications of the random risk vector. In addition to the reduction of all performance measures under **SM**, interestingly, Figure 2 also reveals that **SM** substantially reduces the variance, that is, the sample variance corresponding to 3,000 simulation replications, of these measures as well, specifically, by 51%, 80%, and 66% for $\mathbb{E}[FN]$, $\mathbb{E}[FP]$, and $\mathbb{E}[T]$, respectively.

Table 3. Performance Measures (Point Estimate ± Half Width) for **BC** and **SM**

Model	$\mathbb{E}[FN]$	$\max_{m\in S}\{\mathbb{E}[FN^m]\}$	$\mathbb{E}[FP]$	$\mathbb{E}[T]$	OF[a]
			$UP = 3\,(\mu_p = 0.97\%)$, $Se = 0.95$, $Sp = 0.95$		
BC	0.0942 ± 0.0008	0.0146 ± 0.0002	0.7048 ± 0.0034	24.0196 ± 0.0753	0.5850 ± 0.0024
SM	0.0850 ± 0.0006	0.0086 ± 0.0001	0.5901 ± 0.0015	19.4419 ± 0.0441	0.4822 ± 0.0014
%Change	–10%	–41%	–16%	–19%	–18%
			$UP = 4\,(\mu_p = 1.29\%)$, $Se = 0.95$, $Sp = 0.95$		
BC	0.1266 ± 0.0011	0.0197 ± 0.0003	0.7400 ± 0.0039	26.0331 ± 0.0874	0.6570 ± 0.0028
SM	0.1138 ± 0.0008	0.0116 ± 0.0001	0.6259 ± 0.0014	21.4702 ± 0.0460	0.5512 ± 0.0017
%Change	–10%	–41%	–15%	–18%	–19%
			$UP = 5\,(\mu_p = 1.62\%)$, $Se = 0.95$, $Sp = 0.95$		
BC	0.1570 ± 0.0014	0.0241 ± 0.0003	0.8284 ± 0.0043	29.1062 ± 0.0994	0.7494 ± 0.0034
SM	0.1414 ± 0.0010	0.0143 ± 0.0001	0.6553 ± 0.0014	23.2094 ± 0.0495	0.6131 ± 0.0019
%Change	–10%	–41%	–21%	–20%	–18%
			$UP = 3\,(\mu_p = 0.97\%)$, $Se = 0.93$, $Sp = 0.95$		
BC	0.1302 ± 0.0011	0.0202 ± 0.0003	0.6944 ± 0.0032	23.7900 ± 0.0715	0.6147 ± 0.0025
SM	0.1179 ± 0.0008	0.0121 ± 0.0001	0.5863 ± 0.0015	19.2801 ± 0.0422	0.5105 ± 0.0016
%Change	–9%	–40%	–16%	–19%	–17%
			$UP = 3\,(\mu_p = 0.97\%)$, $Se = 0.97$, $Sp = 0.95$		
BC	0.0574 ± 0.0005	0.0090 ± 0.0001	0.7158 ± 0.0034	24.2618 ± 0.0753	0.5546 ± 0.0020
SM	0.0516 ± 0.0003	0.0052 ± 0.0001	0.5937 ± 0.0015	19.6080 ± 0.0436	0.4536 ± 0.0012
%Change	–10%	–42%	–17%	–19%	–18%
			$UP = 3\,(\mu_p = 0.97\%)$, $Se = 0.95$, $Sp = 0.93$		
BC	0.0947 ± 0.0008	0.0147 ± 0.0002	1.1135 ± 0.0046	25.8354 ± 0.0738	0.6299 ± 0.0024
SM	0.0851 ± 0.0006	0.0087 ± 0.0001	0.9564 ± 0.0021	21.2879 ± 0.0434	0.5265 ± 0.0014
%Change	–10%	–41%	–14%	–18%	–16%
			$UP = 3\,(\mu_p = 0.97\%)$, $Se = 0.95$, $Sp = 0.97$		
BC	0.0939 ± 0.0008	0.0146 ± 0.0002	0.3688 ± 0.0020	22.2124 ± 0.0756	0.5418 ± 0.0023
SM	0.0844 ± 0.0006	0.0086 ± 0.0001	0.2987 ± 0.0009	17.5682 ± 0.0435	0.4384 ± 0.0014
%Change	–10%	–41%	–19%	–21%	–19%

[a]OF is the objective function of **SM** (i.e., $OF = \lambda_1\mathbb{E}[FN] + \lambda_2\mathbb{E}[FP] + (1 - \lambda_1 - \lambda_2)\mathbb{E}[T]$).

Figure 2. Performance Comparison of **BC** and **SM** with Respect to $\mathbb{E}[FN]$, $\mathbb{E}[FP]$, and $\mathbb{E}[T]$, when $UP = 3$ ($\mu_p = 0.97\%$), $Se = 0.95$, and $Sp = 0.95$



5.2. Budget-Constrained Model (BM)

We next consider the budget-constrained testing problem, of generating a testing design for N subjects so as to minimize the number of false-negative classifications under a testing budget constraint. In our risk-based testing policy, this corresponds to **BM**, with $\lambda = 1$. Following common testing practices, we consider that all positive-testing subjects in the initial screening undergo additional confirmatory testing. Consequently, we modify the testing budget constraint in (2) to also include the cost of false positives in the initial screening, $\mathbb{E}[T] + \gamma\mathbb{E}[FP] \le B$, with $\gamma = 1$ (i.e., the cost of a false positive is equal to the testing cost). By Remark 1, this modification does not impact the analytical results.

In the corresponding base-case, **BC**, under the homogeneous and infinite population assumptions, the per-subject expected number of false negatives becomes independent of the group size (Aprahamian et al. 2017). As such, for the corresponding base-case, rather than minimizing the per-subject expected number of false negatives, we minimize the left-hand side of the budget constraint (i.e., $\mathbb{E}[T] + \gamma\mathbb{E}[FP]$), because this will determine the least costly scheme. In this setting, the optimal group size in **BC** is equal to 11, 10, or 9 corresponding to $\mu_p = 0.97\%, 1.29\%,$ and 1.62%, respectively. As stated above, if N is not a multiple of the group size, then the remaining subjects form a (smaller) group for testing; for example, for $\mu_p = 0.97\%$, all subjects are tested in group sizes of 11, except for one subject (randomly selected) who is tested individually (because $N = 100$). The subjects are randomly assigned to groups.

The per-period budgets of **BM** are set to the corresponding testing costs under the **BC** policy; that is, given a risk vector realization in a period, the budget for the risk-based scheme **BM** is set to $B = \mathbb{E}[T] + \mathbb{E}[FP]$ of **BC**. Doing so ensures that the cost of the risk-based scheme does not exceed that of **BC**. Table 4 reports the performance measures of **BM** and **BC** for a range of parameter values and indicates the substantial reductions in all performance measures under **BM**. Specifically, **BM** respectively reduces $\mathbb{E}[FN]$, $\max_{m \in S}\{\mathbb{E}[FN^m]\}$, and $\mathbb{E}[T] + \mathbb{E}[FP]$ by 28%, 48%, and 1% over **BC**. Hence, **BM** generates testing schemes that are substantially more accurate and equitable, in terms of false negatives, than current testing schemes while being cheaper to implement than non–risk-based testing schemes. Moreover, the one-way sensitivity analyses reveal that such reductions are consistent over a range of parameter values. Figure 3 plots $\mathbb{E}[T] + \mathbb{E}[FP]$ as a function of $\mathbb{E}[FN]$, with each point representing one of the 3,000 realizations of the random risk vector. In addition to the reduction of all performance measures under **BM**, interestingly, Figure 3 also reveals that **BM** substantially reduces the variance (i.e., the sample variance corresponding to the 3,000 simulation replications) of $\mathbb{E}[FN]$ and $\max_{m \in S}\{\mathbb{E}[FN^m]\}$, specifically by 72% and 77% over **BC**, respectively.

5.3. Comparison with Existing Heuristics

In this section we compare the performance of **SM** with the following heuristics proposed by McMahan et al. (2012):

- Optimal Dorfman Heuristic (**MC1**): Group sizes are restricted to be equal, and subjects are assigned to groups following a nonincreasing ordering of their risk. The "common" group size is determined by enumerating over a range of possible group sizes and selecting the one that yields the smallest expected number of tests.
- Thresholding Heuristic (**MC2**): A risk threshold is computed, and all subjects having a risk higher than the risk threshold ("high-risk" subjects) are individually tested, whereas all other subjects ("low-risk" subjects) are tested in groups of a common size, which is determined by the **MC1** Heuristic considering only the low-risk subjects. The risk threshold is determined by the following procedure: (i) the **MC1** Heuristic is used to determine the common group size for all N subjects; (ii) using the common group size determined in step (i), groups are constructed starting with the

Table 4. Performance Measures (Point Estimate ± Half Width) for **BC** and **BM**

Model	$\mathbb{E}[FN]$	$\max_{m\in S}\{\mathbb{E}[FN^m]\}$	$\mathbb{E}[FP]+\mathbb{E}[T]$
	$UP = 3\,(\mu_p = 0.97\%)$, $Se = 0.95$, $Sp = 0.95$		
BC	0.0943 ± 0.0008	0.0147 ± 0.0002	24.7396 ± 0.0756
BM	0.0683 ± 0.0004	0.0077 ± 0.0001	24.3924 ± 0.0759
%Change	–28%	–48%	–1%
	$UP = 4\,(\mu_p = 1.29\%)$, $Se = 0.95$, $Sp = 0.95$		
BC	0.1267 ± 0.0011	0.0196 ± 0.0003	26.7660 ± 0.0935
BM	0.0902 ± 0.0005	0.0102 ± 0.0001	26.4263 ± 0.0933
%Change	–29%	–48%	–1%
	$UP = 5\,(\mu_p = 1.62\%)$, $Se = 0.95$, $Sp = 0.95$		
BC	0.1565 ± 0.0013	0.0242 ± 0.0003	29.9104 ± 0.1006
BM	0.1086 ± 0.0006	0.0125 ± 0.0002	29.5492 ± 0.1011
%Change	–31%	–49%	–1%
	$UP = 3\,(\mu_p = 0.97\%)$, $Se = 0.93$, $Sp = 0.95$		
BC	0.1307 ± 0.0011	0.0203 ± 0.0003	24.5217 ± 0.0775
BM	0.0952 ± 0.0006	0.0107 ± 0.0001	24.1705 ± 0.0779
%Change	–27%	–47%	–1%
	$UP = 3\,(\mu_p = 0.97\%)$, $Se = 0.97$, $Sp = 0.95$		
BC	0.0566 ± 0.0005	0.0086 ± 0.0001	24.8495 ± 0.0789
BM	0.0409 ± 0.0002	0.0045 ± 0.0001	24.4928 ± 0.0792
%Change	–28%	–48%	–1%
	$UP = 3\,(\mu_p = 0.97\%)$, $Se = 0.95$, $Sp = 0.93$		
BC	0.0941 ± 0.0008	0.0144 ± 0.0002	26.8905 ± 0.0784
BM	0.0683 ± 0.0004	0.0075 ± 0.0001	26.5379 ± 0.0788
%Change	–27%	–48%	–1%
	$UP = 3\,(\mu_p = 0.97\%)$, $Se = 0.95$, $Sp = 0.97$		
BC	0.0942 ± 0.0008	0.0147 ± 0.0002	22.6159 ± 0.0796
BM	0.0681 ± 0.0004	0.0077 ± 0.0001	22.2615 ± 0.0797
%Change	–28%	–48%	–2%

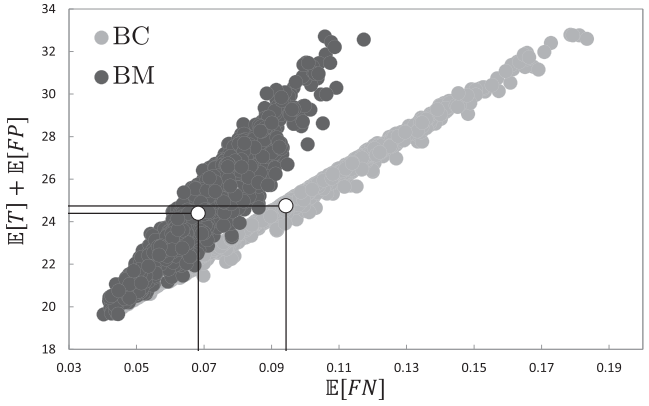
highest-risk subjects. The first group, for which the expected number of tests is lower than the case when the subjects in the group are individually tested, is determined and denoted by group i; (iii) the risk threshold is the average of the highest-risk subject in group i and the lowest-risk subject in group $i+1$.

- Pool-Specific Optimal Dorfman Heuristic (**MC3**): Starting with the lowest-risk subject, groups are constructed in a greedy fashion as follows: the lowest-risk subject that is not yet assigned to a group (i.e., the "lowest unassigned" subject) is assigned to the "current" group only if this leads to a reduction in the expected number of tests per subject; otherwise, the lowest unassigned subject is assigned to a new group, which becomes the "current" group, and the same procedure is repeated until all subjects are assigned to groups.

Thus, whereas **MC1** uses one common group size for all subjects, **MC2** uses individual testing for all high-risk subjects and a common group size for all low-risk subjects (but the common group size can be different from that in **MC1**), and **MC3** can potentially use a different size for each group. Note that even for perfect tests (i.e., $Se = 1$ and $Sp = 1$), the aforementioned heuristics do not necessarily provide the global optimal solution, as they are based on properties that do not necessarily hold in an optimal solution. This is not the case for **SM**, which provides the global optimal solution for both perfect and imperfect tests. Note that all three heuristics, **MC1**, **MC2**, and **MC3**, aim to (heuristically) minimize the expected number of tests. Therefore, for comparison purposes, in **SM** we set λ_1 and λ_2 (i.e., the weights of false negatives and false positives in the **SM** objective function) to zero, to also minimize the expected number of tests in **SM**.

To study how the testing schemes obtained by the **MC1**, **MC2**, and **MC3** Heuristics compare with the optimal solution obtained by **SM**, we perform a Monte Carlo simulation with 10,000 replications. In each replication we generate a risk vector realization and determine the testing schemes by **SM**, **MC1**, **MC2**, and **MC3**. In the following, we demonstrate our results for the case in which $UP = 3\,(\mu_p = 0.97\%)$, $Se = 0.95$, and $Sp = 0.95$. (Our extensive numerical study, with various parameter values, yields similar findings.) Figure 4 plots

Figure 3. Performance Comparison of **BM** and **BC** with Respect to $\mathbb{E}[FN]$, $\mathbb{E}[FP]$, and $\mathbb{E}[T]$, when $UP = 3$ ($\mu_p = 0.97\%$), $Se = 0.95$, and $Sp = 0.95$



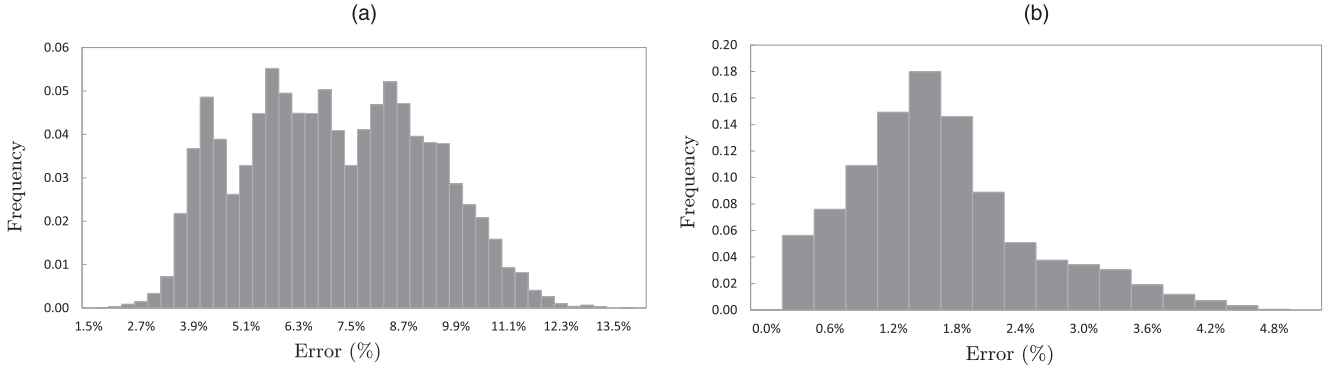
the histograms of the "error" (in percentage) (i.e., the percent deviation of the expected number of tests obtained by the heuristic solution from the optimal expected number of tests obtained by **SM**) for (a) **MC1** and **MC2** and (b) **MC3**. We note here that **MC1** and **MC2** Heuristics provided identical solutions for all problem instances considered in our numerical experiments of this section. This is because, when implementing the **MC2** Heuristic, all subjects were identified as low-risk subjects in all problem instances, effectively reducing the **MC2** Heuristic to the **MC1** Heuristic. According to Figure 4, out of the 10,000 Monte Carlo replications, **MC1** and **MC2** never attained optimality and deviated from optimality by up to 14%. On the other hand, **MC3** attained optimality only once out of the 10,000 replications and deviated from optimality by up to 5%. In terms of classification error, **MC1** and **MC2** reported, on average, a 12% increase in the total classification error over **SM**, whereas **MC3** reported an average reduction of 4.5% in the total classification error over **SM**. However, if classification error is of concern, then one should not utilize the minimization of the expected number of tests as an objective in the first place, and therefore, our models are designed so that they have the capability to incorporate the classification error into the modeling framework. Furthermore, our numerical studies with upper limits on group sizes reveal similar findings to those discussed here for this case study.

This case study underscores the substantial benefits of the proposed optimal risk-based policies in both settings (i.e., **SM** and **BM**); risk-based policies substantially reduce the classification errors and improve efficiency and equity over non–risk-based policies. Additionally, we demonstrate that heuristic solutions are rarely optimal and can deviate, sometimes substantially, from the optimal solution.

6. Conclusions and Future Research Directions

We study the problem of designing an optimal risk-based Dorfman testing scheme to accurately and equitably classify a set of subjects in an efficient manner, while taking into account imperfect tests. Our analytical results enable us to reduce the *NP*-hard partitioning problems into an SP problem (for **SM**) or a constrained-SP problem (for **BM**). Further, for special cases of **BM**, we develop highly efficient algorithms that exploit the structure of the problem and that are able to solve the constrained-SP problem in polynomial time. Our case study demonstrates the effectiveness of risk-based testing, producing solutions that substantially reduce all performance measures when compared with static, non–risk-based models. Our findings (i) demonstrate the drawbacks of formulations in which the only goal is to minimize the expected number of tests; (ii) highlight the importance of considering objective functions that can consider the different dimensions of this problem, including the classification accuracy, equity, and budget constraints; and (iii) underscore the importance of incorporating subject-specific risk characteristics into the modeling framework, because failing to do so can lead to higher classification errors and more costly and less equitable testing schemes.

Figure 4. The Deviation (in Percentage) from the Optimal Expected Number of Tests Obtained by **SM** for (a) **MC1** and **MC2** and (b) **MC3** when $UP = 3$ ($\mu_p = 0.97\%$), $Se = 0.95$, and $Sp = 0.95$



This research can be expanded in several important directions. In reality, subject risk values are not perfectly observable, hence the decision maker needs to estimate the risk of each subject given their characteristics. Therefore, an important future research direction would be to consider robust testing schemes that perform well under risk estimation errors or to consider adaptive strategies that study the exploration (e.g., initial testing for risk estimation) versus exploitation (e.g., testing for classification) trade-off, especially studying the decision of how to allocate a given budget between the efforts of exploration versus exploitation to maximize the classification accuracy.

Another important aspect in group testing is the dilution effect of group testing; that is, for some tests the accuracy of the test for detecting the positive subjects (i.e., the test's sensitivity) may decrease as group size increases. Various studies show, however, that the effect of dilution can be considered negligible up to certain group sizes (Shipitsyna et al. 2007, McMahan et al. 2012). Consequently, one possible way to incorporate the dilution effect into our models is to place upper limits on group sizes, as we have done in parts of our numerical study. This can be easily attained by eliminating all edges from the underlying graph of the partitioning problem that correspond to group sizes larger than the acceptable limit. Alternatively, a more accurate, yet a more complex, approach is to explicitly model the dilution effect (i.e., the test sensitivity becomes a function of the group size) and incorporate it into the modeling framework. Our analysis also depends on the assumption that test outcomes performed on the same subject are conditionally independent, given the actual positivity status of the subject. This assumption does not always hold in practice, especially for infection screening tests that measure infection-related biomarkers, because if a subject is infected with the infection in question, then concentrations of various infection-related biomarkers will be higher than in the infection-free subjects, and tests that measure the related biomarkers will have a tendency to produce positive test outcomes. Relaxing such assumptions on the test's sensitivity and specificity will increase model realism and may produce better testing schemes.

Another important aspect is to consider the operational challenges of changing the testing scheme in each testing period, as is done in this paper. In some cases, modifying the testing scheme on a frequent basis may not be feasible, or simply not desirable. In this case, the decision maker may be interested in determining an optimal *static* policy that does not change from period to period (i.e., a policy characterized by static group sizes, or static risk thresholds with which to partition the subjects into groups once their risk vector is observed); that is, the problem is one of identifying an optimal static policy under uncertainty on the subject risk vector. Finally, our decision problem considers only Dorfman-type testing schemes. There are other, albeit more complicated, group testing schemes, and it will be interesting to study the problem of jointly determining an optimal partitioning and testing scheme. This is a challenging research direction.

It is our hope that this work builds the foundation for more complex risk-based testing schemes and drives future research in any of the aforementioned directions. As we show here, the benefits of risk-based group testing schemes can be substantial, and we hope our work also motivates the practitioners to consider implementing such risk-based testing schemes.

Acknowledgments

The authors thank the associate editor and two referees for excellent comments that improved the analysis and presentation of the paper; and Dr. Scott J. Zimmerman, director of the North Carolina State Laboratory of Public Health, for many valuable discussions and for offering important insights into public health screening practices.

Endnotes

[1] This follows because any test not satisfying this assumption can be transformed into one that satisfies it by interpreting the test outcome in the opposite way.

[2] The "other" category includes the following: white, American Indian or Alaska native, and Asian or Pacific Islander.

[3] $3,037 represents the total cost, that is, the sum of the cost of a false negative ($2,927), cost of a false positive ($55), and testing cost ($55), as discussed at the beginning of Section 5.

References

American Red Cross (2017) Details of tests performed for different infectious agents. Accessed February 3, 2017, http://www.redcrossblood.org/learn-about-blood/what-happens-donated-blood/blood-testing.

Aprahamian H, Bish DR, Bish EK (2016) Residual risk and waste in donated blood with pooled nucleic acid testing. *Statist. Medicine* 35(28):5283–5301.

Aprahamian H, Bish EK, Bish DR (2017) An analytical approach to the Dorfman grouping scheme. Working paper, Virginia Tech, Blacksburg.

Atkinson AB (1970) On the measurement of inequality. *J. Econom. Theory* 2(3):244–263.

Barr NA (1998) *The Economics of the Welfare State* (Stanford University Press, Stanford, CA).

Beasley JE, Christofides N (1989) An algorithm for the resource constrained shortest path problem. *Networks* 19(4):379–394.

Berger T, Mehravari N, Towsley D, Wolf J (1984) Random multiple-access communication and group testing. *IEEE Trans. Comm.* 32(7):769–779.

Bertsimas D, Farias VF, Trichakis N (2011) The price of fairness. *Oper. Res.* 59(1):17–31.

Bertsimas D, Farias VF, Trichakis N (2012) On the efficiency-fairness trade-off. *Management Sci.* 58(12):2234–2250.

Bilder CR, Tebbs JM (2012) Pooled-testing procedures for screening high volume clinical specimens in heterogeneous populations. *Statist. Medicine* 31(27):3261–3268.

Black MS, Bilder CR, Tebbs JM (2012) Group testing in heterogeneous populations by using halving algorithms. *J. Roy. Statist. Soc. Ser. C* 61(2):277–290.

Black MS, Bilder CR, Tebbs JM (2015) Optimal retesting configurations for hierarchical group testing. *J. Roy. Statist. Soc. Ser. C* 64(4): 693–710.

Blass A, Gurevich Y (2002) The logic in computer science column pairwise testing. *Bull. EATCS* 78:100–132.

Brandeau ML, Sainfort F, Pierskalla WP (2004) *Operations Research and Health Care: A Handbook of Methods and Applications*, International Series in Operations Research & Management Science, vol. 70 (Kluwer Academic Publishers, Dordrecht, Netherlands).

Centers for Disease Control and Prevention (2000) Tracking the hidden epidemics, trends in STDs in the United States, 2000. Accessed February 16, 2018, https://www.cdc.gov/std/trends2000/trends2000.pdf.

Centers for Disease Control and Prevention (2014) Selected STDs by age, race/ethnicity, and gender, 1996-2014 request. Accessed February 16, 2018, http://wonder.cdc.gov/std-race-age.html.

Centers for Disease Control and Prevention Division of STD Prevention (2014) Sexually transmitted disease surveillance 2014. Accessed November 21, 2016, http://www.cdc.gov/std/stats14/surv-2014-print.pdf.

Chakravarty AK, Orlin JB, Rothblum UG (1982) Technical note: A partitioning problem with additive objective with an application to optimal inventory groupings for joint replenishment. *Oper. Res.* 30(5):1018–1022.

Cormen TH (2009) *Introduction to Algorithms* (MIT Press, Cambridge, MA).

Cormode G, Muthukrishnan S (2006) Combinatorial algorithms for compressed sensing. Prencipe G, Zaks S, eds. *Internat. Colloquium Structural Inform. Comm. Complexity* (Springer, New York), 280–294.

Deng S, Yano CA (2006) Joint production and pricing decisions with setup costs and capacity constraints. *Management Sci.* 52(5): 741–756.

Dorfman R (1943) The detection of defective members of large populations. *Ann. Math. Statist.* 14(4):436–440.

Engineer FG, Nemhauser GL, Savelsbergh MW (2011) Dynamic programming-based column generation on time-expanded networks: Application to the dial-a-flight problem. *INFORMS J. Comput.* 23(1):105–119.

European Blood Alliance (2017) Blood, tissues and cells from human origin. Accessed February 13, 2017, http://www.europeanbloodalliance.eu/wp-content/uploads/2013/04/eba_online.pdf.

Feng J, Liu L, Parlar M (2010) An efficient dynamic optimization method for sequential identification of group-testable items. *IIE Trans.* 43(2):69–83.

Garcia R (2009) Resource constrained shortest paths and extensions. PhD thesis, Georgia Institute of Technology, School of Industrial and Systems Engineering, Atlanta.

Gimbel J, Kennedy JW, Quintas LV (1993) *Quo Vadis, Graph Theory?: A Source Book for Challenges and Directions*, Annals of Discrete Mathematics, vol. 55 (Elsevier, New York).

Graff LE, Roeloffs R (1972) Group testing in the presence of test error; an extension of the Dorfman procedure. *Technometrics* 14(1): 113–122.

Grimes JA, Smith LA, Fagerberg K (2013) *Sexually Transmitted Disease: An Encyclopedia of Diseases, Prevention, Treatment, and Issues* (ABC-CLIO, Santa Barbara, CA).

Gupta D, Malina R (1999) Group testing in presence of classification errors. *Statist. Medicine* 18(9):1049–1068.

Handler GY, Zang I (1980) A dual algorithm for the constrained shortest path problem. *Networks* 10(4):293–309.

HOLOGIC (2017) Tigris DTS system. Accessed August 5, 2017, http://www.hologic.com/products/clinical-diagnostics/instrument-systems/tigris-dts-system.

Hong ES, Ladner RE (2002) Group testing for image compression. *IEEE Trans. Image Processing* 11(8):901–911.

Hwang FK (1975) A generalized binomial group testing problem. *J. Amer. Statist. Assoc.* 70(352):923–926.

Irnich S, Desaulniers D, Desrosiers J, Hadjar A (2010) Path-reduced costs for eliminating arcs in routing and scheduling. *INFORMS J. Comput.* 22(2):297–313.

Johnson NL, Kotz S, Wu XZ (1991) *Inspection Errors for Attributes in Quality Control*, Monographs on Statistics and Applied Probability, vol. 44 (CRC Press, Boca Raton, FL).

Kalai E, Smorodinsky M (1975) Other solutions to Nash's bargaining problem. *Econometrica J. Econometric Soc.* 13(3):513–518.

Kapala J, Copes D, Sproston A, Patel J, Jang D, Petrich A, Mahony J, Biers K, Chernesky M (2000) Pooling cervical swabs and testing by ligase chain reaction are accurate and cost-saving strategies for diagnosis of *Chlamydia trachomatis*. *J. Clinical Microbiol.* 38(7): 2480–2483.

Kim HY, Hudgens MG, Dreyfuss JM, Westreich DJ, Pilcher CD (2007) Comparison of group testing algorithms for case identification in the presence of test error. *Biometrics* 63(4):1152–1163.

Lan T, Kao D, Chiang M, Sabharwal A (2010) An axiomatic theory of fairness in network resource allocation. *Proc. 29th Conf. Inform. Comm.* (IEEE, Piscataway, NJ), 1343–1351.

Lewis JL, Lockary VM, Kobic S (2012) Cost savings and increased efficiency using a stratified specimen pooling strategy for *Chlamydia trachomatis* and *Neisseria gonorrhoeae*. *Sexually Transmitted Diseases* 39(1):46–48.

Li T, Chan CL, Huang W, Kaced T, Jaggi S (2014) Group testing with prior statistics. *2014 IEEE Internat. Sympos. Inform. Theory* (IEEE, New York), 2346–2350.

Luss H (1999) On equitable resource allocation problems: A lexicographic minimax approach. *Oper. Res.* 47(3):361–378.

Malinovsky Y, Albert PS, Roy A (2016) Reader reaction: A note on the evaluation of group testing algorithms in the presence of misclassification. *Biometrics* 72(1):299–302.

McMahan CS, Tebbs JM, Bilder CR (2012) Informative Dorfman screening. *Biometrics* 68(1):287–296.

North Carolina State Laboratory of Public Health (2016) Virology/serology: Chlamydia/gonorrhea. Accessed November 21, 2016, http://slph.ncpublichealth.com/virology-serology/chlamydia/default.asp.

Owusu-Edusei K Jr, Chesson HW, Gift TL, Brunham RC, Bolan G (2015) Cost-effectiveness of Chlamydia vaccination programs for young women. *Emerging Infectious Diseases* 21(6):960–968.

Pevzner PA, Lipshutz RJ (1994) Towards DNA sequencing chips. *Internat. Symposium Math. Foundations Comput. Sci.* (Springer, New York), 143–158.

Roche Diagnostics USA (2017) Cobas system. Accessed August 11, 2017, https://usdiagnostics.roche.com/en/molecular/systems-automation/cobas-6800-8800-systems.html#overview.

Saraniti BA (2006) Optimal pooled testing. *Health Care Management Sci.* 9(2):143–149.

Shipitsyna E, Shalepo K, Savicheva A, Unemo M, Domeika M (2007) Pooling samples: The key to sensitive, specific and cost-effective genetic diagnosis of chlamydia trachomatis in low-resource countries. *Acta Dermato-Venereologica* 87(2):140–143.

Sobel M, Groll PA (1959) Group testing to eliminate efficiently all defectives in a binomial sample. *Bell System Tech. J.* 38(5): 1179–1252.

Solyalı O, Cordeau JF, Laporte G (2015) The impact of modeling on robust inventory management under demand uncertainty. *Management Sci.* 62(4):1188–1201.

Tebbs JM, McMahan CS, Bilder CR (2013) Two-stage hierarchical group testing for multiple infections with application to the infertility prevention project. *Biometrics* 69(4):1064–1073.

U.S. Food and Drug Administration (2017) FDA, complete list of donor screening assays for infectious agents and HIV

diagnostic assays. Accessed March 8, 2017, http://www.fda.gov/BiologicsBloodVaccines/BloodBloodProducts/ApprovedProducts/LicensedProductsBLAs/BloodDonorScreening/InfectiousDisease/ucm080466.htm#TOC.

Van Der Pol B, Liesenfeld O, Williams JA, Taylor SN, Lillis RA, Body BA, Nye M, Eisenhut C, Hook EW (2012) Performance of the cobas CT/NG test compared to the Aptima AC2 and Viper CTQ/GCQ assays for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae*. *J. Clinical Microbiol.* 50(7):2244–2249.

Van Hoesel S, Romeijn HE, Morales DR, Wagelmans AP (2005) Integrated lot sizing in serial supply chains with production capacities. *Management Sci.* 51(11):1706–1719.

Verter V, Kara BY (2008) A path-based approach for HAZMAT transport network design. *Management Sci.* 54(1):29–40.

Wagstaff A (1991) QALYs and the equity-efficiency trade-off. *J. Health Econom.* 10(1):21–41.

Wang D, McMahan CS, Tebbs JM, Bilder CR (2018) Group testing case identification with biomarker information. *Comput. Statist. Data Anal.* 122(June):156–166.

Wein LM, Zenios SA (1996) Pooled testing for HIV screening: Capturing the dilution effect. *Oper. Res.* 44(4):543–569.

Weusten J, Vermeulen M, van Drimmelen H, Lelie N (2011) Refinement of a viral transmission risk model for blood donations in seroconversion window phase screened by nucleic acid testing in different pool sizes and repeat test algorithms. *Transfusion* 51(1):203–215.

Yen JY (1971) Finding the k shortest loopless paths in a network. *Management Sci.* 17(11):712–716.

Zou S, Stramer SL, Dodd RY (2012) Donor testing and risk: Current prevalence, incidence, and residual risk of transfusion-transmissible agents in US allogeneic donations. *Transfusion Medicine Rev.* 26(2):119–128.