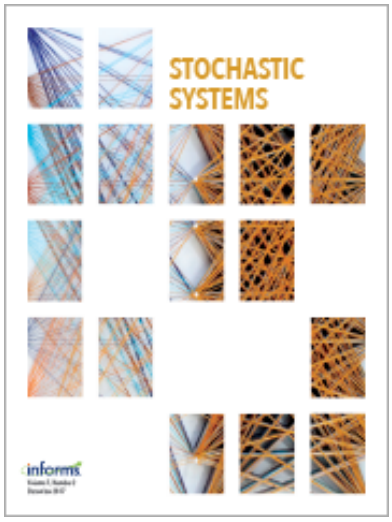


Stochastic Systems

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Static Risk-Based Group Testing Schemes Under Imperfectly Observable Risk

Hrayer Aprahamian, Ebru K. Bish, Douglas R. Bish

To cite this article:
Hrayer Aprahamian, Ebru K. Bish, Douglas R. Bish (2020) Static Risk-Based Group Testing Schemes Under Imperfectly Observable Risk. Stochastic Systems 10(4):361-390. https://doi.org/10.1287/stsy.2019.0059

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

STOCHASTIC SYSTEMS
Vol. 10, No. 4, December 2020, pp. 361–390
ISSN 1946-5238 (online)

Static Risk-Based Group Testing Schemes Under Imperfectly Observable Risk

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

[a] Department of 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, https://orcid.org/0000-0002-8750-2366 (HA); ebru@vt.edu (EKB); drb1@vt.edu (DRB)

Received: November 17, 2018
Revised: November 20, 2019; March 17, 2020
Accepted: March 27, 2020
Published Online in Articles in Advance: September 17, 2020

https://doi.org/10.1287/stsy.2019.0059

Abstract. Testing multiple subjects within a group, with a single test applied to the group (i.e., *group testing*), is an important tool for classifying populations as positive or negative for a specific binary characteristic in an efficient manner. We study the design of easily implementable, static group testing schemes that take into account operational constraints, heterogeneous populations, and uncertainty in subject risk, while considering classification accuracy- and robustness-based objectives. We derive key structural properties of optimal risk-based designs and show that the problem can be formulated as network flow problems. Our reformulation involves computationally expensive high-dimensional integrals. We develop an analytical expression that eliminates the need to compute high-dimensional integrals, drastically improving the tractability of constructing the underlying network. We demonstrate the impact through a case study on chlamydia screening, which leads to the following insights: (1) Risk-based designs are shown to be less expensive, more accurate, and more robust than current practices. (2) The performance of static risk-based schemes comprised of only two group sizes is comparable to those comprised of many group sizes. (3) Static risk-based schemes are an effective alternative to more complicated dynamic schemes. (4) An expectation-based formulation captures almost all benefits of a static risk-based scheme.

Funding: This material is based upon work supported in part by the National Science Foundation [Grant 1055360].

Keywords: group testing • Dorfman testing • risk-based testing • robust optimization • combinatorial optimization • constrained shortest path • order statistics

1. Introduction and Motivation

Classifying subjects within a large population as positive or negative for a certain binary characteristic (e.g., disease, product defect), through screening, is important in many settings. Testing each subject individually incurs high testing costs and is often not budget-feasible. Consequently, testing facilities often utilize a testing method known as *group testing*, wherein multiple subjects, or specimens from those subjects (e.g., blood or urine samples, genetic material), are grouped and tested together, with one test applied to each group. In this case, the test provides one outcome for the entire group, with a positive group test outcome suggesting the presence of the binary characteristic in at least one subject in the group and a negative test outcome suggesting that all subjects in the group are free of the binary characteristic. Subjects in positive-testing groups may undergo follow-up testing via new specimens collected from those subjects, so that they can be classified as positive or negative for the binary characteristic. Thus, group testing can offer substantial reductions in testing costs over individual testing, especially in the case of a binary characteristic with a low prevalence and is commonly utilized as an integral part of screening/testing schemes across various disciplines, including public health screening, industrial quality control, conflict resolution in multiaccess communication networks, software testing, and compressed sensing (e.g., Dorfman 1943, Sobel and Groll 1959, Berger et al. 1984, Blass and Gurevich 2002, Cormode and Muthukrishnan 2006). The origins of group testing date back to the 1940s, when Dorfman (1943), an economist, introduced this concept as a way to test military inductees for syphilis in an efficient manner. Dorfman proposed a simple two-stage testing scheme: in the first stage, a group of subjects are tested with a single test; if the group test outcome is negative, then testing stops and all subjects in

the group are classified as negative; if, on the other hand, the test outcome is positive, then testing proceeds to the second stage in which each subject is individually tested and classified according to the outcome of their individual test. Today, this so-called *Dorfman testing* is one of the most commonly adopted schemes in practice due to its simplicity and efficiency and is the focus of this paper.

In these settings, the tester needs to determine the various group sizes to be used in the first stage of Dorfman testing, along with the assignment of *heterogeneous* subjects, with different *risk* (probability of positivity) estimates for the binary characteristic, to the different groups. While doing so, various operational and technological constraints may need to be taken into account to ensure that the testing scheme is feasible and/or easily implementable for large populations. For example, it may not be practical to change the testing scheme (e.g., group sizes) frequently or to use a large number of different (distinct) group sizes, as doing so may incur high setup cost/time (e.g., for configuring the testing machine for different group sizes) and operational challenges or may simply be infeasible due to the capability of the testing equipment. Thus, a *static* testing scheme, which does not change over time, is highly desirable for practitioners and is the focus of this paper. In addition to ease of implementation, another advantage of a static testing scheme is that it is determined only once and can be integrated into the testing protocol, allowing for a clear and concise testing protocol. While determining a static testing scheme, it is essential to customize the scheme based on the characteristics of the testing population (e.g., risk distribution) and the testing facility (e.g., capabilities of the testing machines), and we develop optimization models that determine customized static testing schemes considering these characteristics.

What further complicates the testing scheme design is that the test is not perfectly reliable; consequently, subject misclassification is possible. This includes *false negative classifications*, that is, positive subjects falsely classified as negative, and *false positive classifications*, that is, negative subjects falsely classified as positive. This testing design problem arises in many settings, as discussed above. For example, in the context of public health screening, testing laboratories screen donated blood (via a specimen from each donation), received periodically (e.g., with each shipment), for a set of infectious diseases such as the human immunodeficiency virus (HIV) and hepatitis viruses (American Red Cross 2016); similarly, public health screening laboratories test segments of the state population for sexually transmitted diseases (STDs), such as chlamydia, via specimens from the subjects received periodically (McMahan et al. 2012). In both cases, specimens are loaded into automated testing machines in *batches* (typically between 40–200 specimens per batch depending on the equipment used; HOLOGIC 2017). Due to the large number of subjects that need to be tested, testing begins as soon as a sufficient number of subjects to form a complete batch arrive. As another example, consider industrial quality control in which a set of products, received periodically (e.g., with each shipment from the supplier, in each manufacturing shift), needs to be tested for defects; and the batch size may represent the shipment size or the capacity of the testing machine. Note that the batch size is a decision related, for example, to equipment acquisition, operational characteristics, and financial parameters (e.g., based on the trade-off between fixed costs and the delay in obtaining results), whereas the testing scheme (group sizes, which is the focus of our paper) is an operational decision based on the population risk distribution and test efficacy (sensitivity and specificity), and the batching decision constrains the grouping decision. We model this aspect of the problem by considering the batch size as exogeneous and studying the operational level grouping decision, constrained by the exogenous batch size. While determining an optimal batch size is an important problem, it is outside the scope of this paper.

The probability of having the binary characteristic (*risk*) may vary, sometimes substantially, with subject-specific characteristics (*risk factors*) that are often known prior to testing. For example, in donated blood screening, first-time blood donors in the United States (U.S.) are around seven times more likely to have an HIV infection than repeat donors (Zou et al. 2012); various risk factors exist for STDs, including chylamydia (Centers for Disease Control and Prevention Division of STD Prevention 2014); vector-borne infections, such as babesiosis and Zika, are endemic in certain areas of the U.S. (Herwaldt et al. 2011). Thus, subjects come from a *heterogeneous* population, with each subpopulation having a potentially different risk. However, the process of estimating the subject-specific risk, informed by subpopulation-specific risk estimates or based on established risk factors for the binary characteristic, is far from perfect. This is because risk estimates in the different subpopulations are inherently uncertain, and risk factors are often not well understood (e.g., Grassly et al. 2004, Arora et al. 2010). Consequently, the true risk of a subject is unobservable, and the tester needs to estimate the risk of each subject and construct an uncertainty set that contains the true risk with a high probability. Under such uncertainty in risk estimation, it is important to determine testing schemes that are not highly sensitive to perturbations in risk estimates, that is, robust testing schemes. Specifically, we study the problem of determining optimal *static risk-based Dorfman testing schemes*, comprised of a set of group sizes

and a policy to assign subjects, with different risk, to the mutually exclusive groups, under uncertainty on both subject characteristics (hence, the estimated risk) and the actual risk. Our goal is to identify a static testing scheme that is used repetitively for every batch, satisfies certain limitations on group sizes, and is *accurate* in terms of subject classification, *efficient* in terms of testing cost, and *robust* with respect to deviations from the estimated risk vector. Most literature on group testing considers the objective of minimizing the testing cost under perfect tests, with limited focus on misclassification; robustness is an often-overlooked dimension in group testing, as the relevant literature almost exclusively assumes that subject risk is perfectly observable, that is, deterministic and known.

More specifically, both Dorfman's original model and the majority of the subsequent research impose unrealistic assumptions, such as *perfect* tests, that is, there are no classification errors, a *homogeneous* population, that is, the risk of the binary characteristic is identical across subjects, and infinite testing *batch* sizes (e.g., Dorfman 1943, Sobel et al. 1959, Sterrett 1957). Although several papers extend the analysis of Dorfman testing schemes to imperfect tests (e.g., Graff and Roeloffs 1972, Johnson et al. 1991, Kim et al. 2007, McMahan et al. 2012), there is very limited work on Dorfman testing for a heterogeneous population, that is, with subject-specific risk, and the few papers that consider a heterogeneous population (e.g., Hwang 1975, McMahan et al. 2012, Aprahamian et al. 2019) mainly do so under restrictive assumptions, including that subject risk is perfectly observable, or they determine testing schemes heuristically. In particular, Hwang (1975) determines optimal risk-based Dorfman testing schemes for a heterogeneous population, but under the assumption that the test is perfect (hence, the objective is to minimize the number of tests) and the subject risk is perfectly observable. Moreover, Hwang's focus is on *dynamic* testing schemes, that is, group sizes and the subject-group assignment policy are allowed to change with each batch, that is, the group testing problem is a deterministic problem, solved after the risk of each batch of subjects is observed. McMahan et al. (2012) extend the analysis in Hwang (1975) to the case of imperfect tests, but conjecture that the problem of determining risk-based Dorfman testing schemes that minimize the expected number of tests, under imperfect tests and perfectly observable subject risk, is intractable and they develop heuristics. More recently, focusing on dynamic testing schemes, Aprahamian et al. (2019) demonstrate that the generalization of Hwang's model that takes into account imperfect tests, but with perfectly observable subject risk, is in fact tractable, resolving the conjecture in the literature; establish the equivalence of the dynamic testing design problem to a specific form of a network flow problem, namely, the constrained-shortest path problem; and develop exact algorithms to determine optimal *dynamic* risk-based Dorfman testing schemes.

Although the aforementioned studies have improved our understanding of optimal risk-based Dorfman testing for a heterogeneous population, they leave out other important aspects of the problem, such as *implementability* of the testing scheme and *uncertainty* in subject risk estimates. For example, both Aprahamian et al. (2019) and Hwang (1975) assume that the decision-maker can construct an optimal dynamic testing scheme, *customized* for each batch of subjects, and subject risk values are deterministic and perfectly observable by the tester. Although the first assumption may be justified in certain settings, in other settings the decision-maker may not have the flexibility to modify the testing scheme for every batch, as discussed above. McMahan et al. (2012) consider a static testing scheme (same group sizes and assignment policy are used for every batch), which partially resolves the implementability issue, but this is done by: (i) ordering the subjects in nondecreasing order of their risk and simply setting the risk of each subject to the expected risk of the corresponding order statistics and (ii) determining testing schemes that attempt to minimize the expected number of tests using heuristics, which are based on structures that are not guaranteed to exist in an optimal testing scheme (e.g., group sizes are nonincreasing for a risk-ordered batch; see Aprahamian et al. 2019). A lack of properties and algorithms for optimal static Dorfman testing schemes in this setting is not surprising, because various functions of order statistics (for batch sizes that are in the hundreds) arise in an exact formulation, substantially complicating the analysis. Our analysis of optimal static Dorfman testing schemes for heterogeneous populations resolves all of the aforementioned issues and, as a by-product, provides an analytical expression to compute the expectation of the product of some function of a set of consecutive order statistics in an efficient manner, without requiring high-dimensional integrals. In addition, we investigate the realistic situation in which the true risk of a subject is not known with certainty, but lies within a known uncertainty set, and this aspect of the problem gives rise to a novel *robust* formulation of the problem.

Our contributions in this paper are as follows: First, we model important aspects of group testing that are often overlooked in the literature, such as the implementability of the testing scheme and the uncertainty in subject risk, and we do so within a classification accuracy maximization framework (rather than the minimization of the expected number of tests that is prevalent in the literature). In this aspect, our formulation builds upon the deterministic formulation in Aprahamian et al. (2019) to consider static testing schemes under

uncertainty and allows us to quantify the "price of implementability." Some key properties established in Aprahamian et al. (2019) for the deterministic model extend to the stochastic setting. Consequently, we take advantage of a reformulation technique presented in Aprahamian et al. (2019) in which the set partitioning problem is cast as a constrained-shortest path problem. However, unlike the model in Aprahamian et al. (2019), this reformulation does not, by itself, lead to an efficient solution technique for the static grouping problem, and the structure of the optimal solution needs to be further analyzed. Specifically, constructing the underlying graph requires the computation of a large number of high-dimensional integrals, the number of which grows quadratically in the number of subjects. By exploiting the specific structure of the resulting integral for consecutive order statistics, we are able to provide an equivalent expression for these integrals, substantially reducing their dimension. This drastically improves the tractability of our approach and allows us to solve realistic problem sizes to optimality. Armed with these results, we explore novel expectation-based and robust formulations of this decision problem, show that both models reduce to a common form, which can be equivalently formulated as a network flow problem, and use this reformulation to solve the static testing design problem to optimality, expanding the previous result on dynamic testing schemes in Aprahamian et al. (2019) to static testing schemes under risk uncertainty. Analysis of the expectation-based and robust models further provides valuable insight on the trade-off between classification accuracy and robustness. We demonstrate the effectiveness of the proposed static risk-based Dorfman testing scheme through a case study on chlamydia screening, one of the most prevalent STDs in the U.S. The proposed testing schemes reduce the misclassification and testing costs substantially over optimal non-risk-based (*uniform*) schemes and current screening practices. Further, our numerical study suggests that the expected testing cost and misclassification cost of static risk-based testing schemes are within 2% of the more complicated dynamic risk-based testing schemes, that is, schemes that are customized for each batch (Hwang 1975, Aprahamian et al. 2019). Thus, restricting the testing scheme to a static scheme does not hinder the performance of screening in a significant way. We find that this price of implementability is especially low when the batch sizes are large or test accuracy is low. Our numerical results also indicate that the performance of static risk-based testing schemes comprised of only a small number of group sizes (only two in our setting) is comparable to more complicated static risk-based testing schemes comprised of many group sizes. These findings indicate that simple static schemes, with a small number of groups, can capture most benefits of risk-based testing, underscoring the value of static risk-based testing schemes studied in this paper.

Lastly, from a practical perspective, which motivated this paper, the decision-maker (e.g., a laboratory manager at a public health state laboratory) must develop testing protocols. Our optimization models develop customized static testing schemes that specify, for all batches, both the group sizes to be used and the subject assignment policy (i.e., based on a risk-ordering of the subjects), allowing for a clear and concise testing protocol. The testing protocol for a dynamic testing scheme, in which group sizes and subject-group assignment potentially change with each batch of subjects, are much more complex. We also show that for an optimal assignment of the heterogeneous subjects to testing groups, a risk-ordering of subjects is sufficient, that is, the tester does not need to estimate subject risk with high accuracy. The models developed in this paper provide the decision-maker with the necessary tools to identify an optimal static policy, which can then be included in their testing protocol.

The remainder of this paper is structured as follows. Section 2 presents the notation and the decision problem, and Section 3 discusses the expectation-based and robust formulations and provides expressions for the relevant metrics. Section 4 then analyzes the optimal design of static risk-based Dorfman testing schemes and unveils key properties of an optimal testing scheme. Section 5 presents our results and findings from our case study on chlamydia screening in the U.S. Finally, Section 6 concludes the paper and provides possible future research directions. To improve the presentation and flow of the paper, all proofs and derivations are relegated to the appendix.

2. The Notation and the Decision Problem

In this section, we present the notation and the decision problem. Throughout, we use uppercase letters to denote random variables, lowercase letters to denote their realization, and boldface to denote vectors. The terms *positive* and *negative* are used to refer to a subject's true status (i.e., to denote whether the binary characteristic is present) or to the binary test outcome (i.e., to denote whether the test outcome *indicates* the presence of the binary characteristic).

Consider a testing facility (e.g., laboratory) where subjects (or specimens collected from subjects) arrive throughout the day. Subjects are tested in batches of size N for a binary characteristic, where the batch size N is determined by the testing equipment and, hence, is considered exogenous in our models (see the discussion

in Section 1). Due to limited testing capacity and throughput requirements, we assume that testing begins when a sufficient number of subjects arrive to form a batch. For example, most public health screening laboratories have testing equipment dedicated to the screening of a certain condition (i.e., disease or genetic disorder), and most testing machines are highly automated, for example, the nucleic acid amplification testing machine (HOLOGIC 2017, Roche Diagnostics 2017) that we consider in our case study in Section 5. These testing machines are loaded in batches (e.g., with batch sizes, N, ranging from 40 to 200, depending on the testing machine), and testing of each batch typically takes three to four hours. Thus, testing starts as soon as a set of N subjects is received. The tester needs to place these N subjects in groups and classify each subject as positive or negative for the binary characteristic, so as to minimize the costs of misclassification and testing under imperfectly reliable tests and constrained by the given batch size.

The population is *heterogeneous* with respect to *risk* (probability of positivity) for the binary characteristic due to subject-specific demographic and clinical factors. To model population heterogeneity, let D^m denote the true status of subject m for the binary characteristic, with a value of 1 if subject m is a true positive for the characteristic and 0 otherwise, that is, random variable D^m, unknown to the tester, follows a Bernoulli distribution with a subject-specific probability of positivity given by P^m, independent of other subjects. However, the value of P^m, that is, the *true* risk of subject m, is unobservable by the tester. Therefore, the tester *estimates* the risk of subject m, denoted by $\tilde{P}^m$, based on the subject's characteristics. We let Ξ^m denote the random perturbation (error) term for the risk of subject m, that is, the deviation of the estimated risk from the true risk. Thus, the true risk of subject m can be expressed as a function of the estimated risk and the random perturbation term. We assume that random variables $\tilde{P}^m, m = 1, \ldots, N$, are independent and identically distributed (iid), following an arbitrary continuous distribution with support in $[a, b]$, for $0 \le a < b \le 1$; random variables $\Xi^m, m = 1, \ldots, N$, are iid, following an arbitrary continuous distribution with support in $[-\delta, \delta]$, for some $\delta \ge 0$; and random vectors $\tilde{\boldsymbol{P}}$ and $\boldsymbol{\Xi}$ are independent.[1] Our modeling of subject risk, $\tilde{P}^m, m = 1, \ldots, N$, as a continuous random variable not only allows for a broad class of risk prediction models with continuous outcome (e.g., regression), but also significantly improves the computational tractability of our solution algorithm. In what follows, we also discuss the implications of a discrete (categorical) risk distribution. Then, the true risk of subject m, conditional on the estimated risk and perturbation term, can be written as $P^m|\tilde{P}^m, \Xi^m = t(\tilde{P}^m, \Xi^m)$, for $m = 1, \ldots, N$, where $t(\cdot)$ is some arbitrary continuous function in $[0, 1]$. We do not make any assumptions on function $t(p, \xi)$, other than that it is nondecreasing in each of p and ξ; $\mathbf{E}_{\Xi^m}[t(\tilde{p}^m, \Xi^m)] = \tilde{p}^m$, that is, the expectation of the true risk equals the estimated risk; and $t(p, 0) = p$, for all p, that is, the true risk reduces to the estimated risk when the perturbation term is zero, that is, the case of no estimation error. Then, $D^m|\tilde{P}^m$ follows a compound Bernoulli distribution with a probability of positivity of $P^m|\tilde{P}^m$, which lies within an uncertainty set, $[t(\tilde{P}^m, -\delta), t(\tilde{P}^m, \delta)] \subseteq [0, 1]$. Note that the size of the uncertainty set can differ among the subjects, as it is a function of the subject's estimated risk, $\tilde{P}^m, \forall m$; this modeling reflects the fact that the accuracy of a subject's estimated risk may depend on their specific characteristics. In addition, the size of the uncertainty set becomes larger for higher values of δ.

Without loss of generality, we represent the set of subjects in each batch as a *risk-ordered set*, $S \equiv \{1, \ldots, N\}$, that follows a nondecreasing order of the estimated subject risk, that is, $\tilde{p}^1 \le \tilde{p}^2 \le \cdots \le \tilde{p}^N$. Then $\tilde{\boldsymbol{P}} = (\tilde{P}^{(1)}, \tilde{P}^{(2)}, \ldots, \tilde{P}^{(N)})$ denotes the random *ordered* estimated risk vector per batch, with $\tilde{P}^{(m)}$ denoting the mth-order statistic of a random sample of size N.

On the testing side, the test can be used for both individual testing and group testing (i.e., with specimens from multiple subjects combined and tested as a group with a single test). Although one test per subject suffices for individual testing, group testing follows the *two-stage Dorfman* testing scheme: initially, the group is tested with a single test; if the group test outcome is negative, then testing stops and all subjects in the group are classified as negative. On the other hand, if the group test outcome is positive, then each subject in the group is tested separately and classified according to the outcome of their individual test. In either case, the test is not perfectly reliable, leading to the possibility of misclassification. Let $R^m(n)$ denote the random classification outcome for subject m, $m = 1, \ldots, N$, when tested within a group of size n in the first stage, with $R^m(n) = 1$ if subject m is classified as positive and 0 otherwise. Then, subject m will become a *false negative classification*, that is, a true-positive subject falsely classified as negative, with probability $\Pr(D^m(1 - R^m(n)) = 1)$, and a *false positive classification*, that is, a true-negative subject falsely classified as positive, with probability $\Pr((1 - D^m)R^m(n) = 1)$. Let Se and Sp respectively denote the test's *sensitivity* (true positive probability, i.e., the probability that the test outcome is positive, given that the group contains at least one true positive) and *specificity* (true negative probability, i.e., the probability that the test outcome is negative, given that the group contains all true negatives), and we assume that the test's sensitivity and specificity are not altered by group size. Although the assumption that a test's specificity is not altered by group size generally holds, in certain

settings the test's sensitivity may reduce as the group size increases, due to the *dilution effect* of grouping. We investigate the impact of dilution on the optimal group testing solution in Section 5.4. Without loss of generality, we consider that the test's true negative probability is higher than its false negative probability,[2] that is, $Sp \geq (1 - Se) \Rightarrow Se + Sp - 1 \in [0, 1]$.

The tester needs to determine a *testing scheme*, comprised of a *set of group sizes* to be used (with a group size of one indicating individual testing) and an *assignment policy* (i.e., a set of rules that specify how each of the N subjects, each with a given risk estimate, is to be assigned to one of the mutually exclusive groups in a batch). Our focus is on *static* testing schemes in which group sizes and the assignment policy remain the same for each batch; and we also consider the practical restriction that the tester is able to use a maximum of γ *distinct* group sizes, for some given $\gamma \in \mathbf{Z}^+$. As discussed in Section 1, these constraints may be dictated by the capability of the testing machine or the setup needed to configure the testing machine for the different group sizes.

Then, the *risk-based testing problem* is to determine an optimal static testing scheme, that is, a set of group sizes and an assignment policy, under uncertainty on both the estimated risk vector, $\tilde{\boldsymbol{P}}$, and the perturbation vector, $\boldsymbol{\Xi}$. We represent the decision variables as a collection of mutually exclusive sets, $\boldsymbol{\Omega} = (\Omega_i)_{i=1,\ldots,g}$, for some $g \in \mathbf{Z}^+$, such that $\cup_{i=1}^{g} \Omega_i = S$ and $\Omega_i \cap \Omega_j = \emptyset$, for all $i, j = 1, \ldots, g$: $i \neq j$. Letting $n_i \equiv |\Omega_i|$ denote the cardinality (size) of set $\Omega_i, i = 1, \ldots, g$, we refer to the corresponding vector, $\boldsymbol{n} = (n_i)_{i=1,\ldots,g} : \sum_i n_i = N$, as the *group size* vector. Thus, the set Ω_i is the index set of subjects assigned to group i, where a subject's index is determined based on the risk-ordered set, S, that is, index m denotes the mth-order statistic for a sample of size N. To represent the constraint on the maximum number of distinct group sizes allowable in any testing scheme, let y_j, $j = 1, \ldots, N$, equal 1 if at least one group of size j is utilized and 0 otherwise, that is, for a given $\boldsymbol{\Omega}$ and $\forall j = 1, \ldots, N$, $y_j = 1$ only if there exists at least one group $i, i = 1, \ldots, g$, such that $n_i = j$. Then, letting

$$\|\boldsymbol{\Omega}\| \equiv \sum_{j=1}^{N} y_j, \tag{1}$$

we have that $\|\boldsymbol{\Omega}\| \leq \gamma$.

In this setting, the risk vector, $\tilde{\boldsymbol{p}}$, for each subject in a batch is estimated, and testing is conducted following the assignment indicated by $\boldsymbol{\Omega}$, via groups of sizes $\boldsymbol{n} = (n_i)_{i=1,\cdots,g} : \sum_i n_i = N$. The objective is to minimize a function of misclassification and testing costs. To express the objective function, we define the following random variables.

For any testing scheme given by $\boldsymbol{\Omega}$, let $FN_i(\Omega_i)$, $FP_i(\Omega_i)$, and $T_i(\Omega_i)$, respectively, denote the number of false negative classifications, number of false positive classifications, and number of tests performed for group i, $\forall i$. Then, we have that

$$FN_i(\Omega_i) = \sum_{m \in \Omega_i} D^m(1 - R^m(n_i));\ FP_i(\Omega_i) = \sum_{m \in \Omega_i} (1 - D^m)R^m(n_i);\ T_i(\Omega_i) = \begin{cases} 1, & \text{if } n_i = 1, \\ 1 + n_i I(\Omega_i), & \text{if } n_i > 1, \end{cases}$$

where $I(\Omega_i) = 1$ if the test outcome for group i is positive and 0 otherwise. Then, the total number of false negative classifications, false positive classifications, and tests performed for the set of N subjects under a given testing scheme, $\boldsymbol{\Omega}$, follow:

$$FN(\boldsymbol{\Omega}) = \sum_{i=1}^{g} FN_i(\Omega_i), \quad FP(\boldsymbol{\Omega}) = \sum_{i=1}^{g} FP_i(\Omega_i), \quad \text{and } T(\boldsymbol{\Omega}) = \sum_{i=1}^{g} T_i(\Omega_i).$$

By using these expressions, the total cost for group $i, i = 1, \ldots, g$, and the total cost for the set of N subjects can be, respectively, written as

$$Q_i(\Omega_i) \equiv \lambda_1 FN_i(\Omega_i) + \lambda_2 FP_i(\Omega_i) + (1 - \lambda_1 - \lambda_2)T_i(\Omega_i), \text{ and} \tag{2}$$

$$Q(\boldsymbol{\Omega}) \equiv \sum_{i=1}^{g} Q_i(\Omega_i) = \sum_{i=1}^{g} \left[\lambda_1 FN_i(\Omega_i) + \lambda_2 FP_i(\Omega_i) + (1 - \lambda_1 - \lambda_2)T_i(\Omega_i)\right], \tag{3}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2) \in [0, 1]^2 : \lambda_1 + \lambda_2 \leq 1$ denotes a normalized weight (cost) vector. In practice, the cost of a false negative classification is typically much higher than the cost of a false positive classification. Although false positives are often detected during subsequent confirmatory testing, false negatives may lead to a missed diagnosis and, hence, to potentially severe negative outcomes. We discuss the choice of weight parameters in Section 5.

When clear from the context, we remove the arguments in parentheses to improve the presentation. All mathematical proofs and derivations can be found in the appendix.

3. Optimization Models and the Objective Function

We first present, in Section 3.1, expectation-based and robust formulations of the decision problem. Then, in Section 3.2, we provide analytical expressions of the various performance measures that contribute to the objective function.

3.1. Optimization Models

We use two different approaches for formulating the decision problem: (i) *an expectation-based optimization model* (**EM**) in which the objective is to minimize the expected value (under uncertainty over both $\tilde{\boldsymbol{P}}$ and $\boldsymbol{\Xi}$) of the objective function and (ii) *a robust optimization model* (**RM**) in which the objective is to minimize the expected worst-case value of the objective function, that is, for each possible realization of the estimated risk vector, $\tilde{\boldsymbol{P}}$, we determine a realization of the error vector, $\boldsymbol{\Xi}$, that provides the worst-case objective function value, and we minimize the expected worst-case value (under uncertainty over $\tilde{\boldsymbol{P}}$) of the objective function.

Expectation-Based Optimization Model (Problem EM).

$$\begin{aligned} \underset{\boldsymbol{\Omega}=(\Omega_i)_{i=1,\dots,g},\, g\in\boldsymbol{Z}^+}{\text{minimize}} \quad & \mathbf{E}_{\tilde{P}}\Big[\mathbf{E}_{\Xi}\big[\mathbf{E}\big[Q(\boldsymbol{\Omega})|\boldsymbol{\Xi},\tilde{\boldsymbol{P}}\big]\big]\Big] \\ \text{subject to} \quad & \Omega_i \cap \Omega_j = \emptyset, \quad \forall i,j = 1,\dots,g\colon i \neq j, && (4)\\ & \bigcup_{i=1}^{g} \Omega_i = \{1,\dots,N\}, && (5)\\ & \|\boldsymbol{\Omega}\| \leq \gamma, && (6) \end{aligned}$$

where $Q(\boldsymbol{\Omega})$ is as defined in Equation (3), $\gamma \in \boldsymbol{Z}^+$ represents the maximum number of distinct group sizes allowed, and the operator $\|\cdot\|$ is as defined in Equation (1).

The following lemma provides an equivalent expression of the **EM** objective function, and we use it throughout the paper.

Lemma 1. *Problem* **EM** *can be equivalently formulated as follows:*

$$\begin{aligned} \underset{\boldsymbol{\Omega}=(\Omega_i)_{i=1,\dots,g},\, g\in\boldsymbol{Z}^+}{\text{minimize}} \quad & \mathbf{E}_{\tilde{P}}\Big[\mathbf{E}\big[Q(\boldsymbol{\Omega})|\boldsymbol{\Xi}=\mathbf{0},\tilde{\boldsymbol{P}}\big]\Big] && (7)\\ \text{subject to} \quad & (4),(5),(6). \end{aligned}$$

Problem **EM** is challenging due to two main reasons. First, the problem of determining an optimal testing scheme, $\boldsymbol{\Omega}$, that minimizes the objective function reduces to a partitioning problem, which is *NP*-hard (Chakravarty et al. 1982). Hence, for realistic problem sizes in which the number of subjects in each batch, N, is typically on the order of hundreds, the problem quickly becomes computationally expensive. Second, the objective function is nonlinear and nonseparable, and further, even the evaluation of the objective function for a given solution, $\boldsymbol{\Omega}$, poses some difficulty, as it requires the computation of higher-dimensional integrations (see Sections 3.2 and 4.1).

We next formulate the robust optimization problem. Under uncertainty on the estimated risk vector, $\tilde{\boldsymbol{P}}$, the tester determines a *robust* static testing scheme that would perform *well* under most perturbations to a realized vector, $\tilde{\boldsymbol{p}}$. For this purpose, we consider a mini-max (worst-case) type objective function, commonly adopted in the robust optimization literature (e.g., Scarf et al. 1958, Gallego and Moon 1993, Perakis and Roels 2008, Bertsimas et al. 2011). Specifically, the objective is to determine a robust static testing scheme that minimizes the worst-case cost, which, for a given realization of the estimated risk vector $\tilde{\boldsymbol{p}}$ and a given testing scheme, is attained by a realization of the error vector, $\boldsymbol{\Xi}$, that maximizes the objective function. Then, the goal is to determine a static testing scheme that minimizes the expectation (under uncertainty over $\tilde{\boldsymbol{P}}$) of the worst-case cost. The formulation of the robust optimization problem follows.

Robust Optimization Model (RM).

$$\begin{aligned} \underset{\boldsymbol{\Omega}=(\Omega_i)_{i=1,\dots,g},\, g\in\boldsymbol{Z}^+}{\text{minimize}} \quad & \mathbf{E}_{\tilde{P}}\big[Q^*(\boldsymbol{\Omega})|\tilde{\boldsymbol{P}}\big] && (8)\\ \text{subject to} \quad & (4),(5),(6), \end{aligned}$$

where $Q^*(\boldsymbol{\Omega})|\tilde{\boldsymbol{P}}$ is the optimal solution to the following stage 2 problem:

$$\begin{aligned} Q^*(\boldsymbol{\Omega})|\tilde{\boldsymbol{P}} \equiv \underset{\boldsymbol{\xi}}{\text{maximize}} \quad & \mathbf{E}\big[Q(\boldsymbol{\Omega})|\boldsymbol{\Xi}=\boldsymbol{\xi},\tilde{\boldsymbol{P}}\big] \\ \text{subject to} \quad & -\delta \le \xi^m \le \delta, \quad \forall m = 1,\dots,N. \end{aligned} \tag{9}$$

We denote the optimal solution to the stage 2 problem by $\xi^{m*}(\boldsymbol{\Omega})|\tilde{\boldsymbol{P}}$, for $m = 1,\dots,N$.

Remark 1. When $\delta = 0$, that is, when $\boldsymbol{P} = \tilde{\boldsymbol{P}}$, problem **RM** reduces to problem **EM**.

Problem **RM** suffers from all the difficulties stated for Problem **EM**; in addition, problem **RM** faces yet another challenge: since $\tilde{\boldsymbol{P}}$ is a continuous random vector with an uncountable sample space, to evaluate the expectation in the objective function of (8), one needs to solve an infinite number of optimization problems in stage 2 (i.e., (9)) to obtain $Q^*(\boldsymbol{\Omega})|\tilde{\boldsymbol{P}}$, one for each possible realization of $\tilde{\boldsymbol{P}}$. In the subsequent sections, we characterize key properties of problem **RM** that will enable us to solve it to optimality.

Although a worst-case objective function, similar to the one used in problem **RM**, is a conservative measure (e.g., Perakis and Roels 2008, Bertsimas et al. 2011, El-Amine et al. 2018), one might reduce the value of δ in order to reduce the conservativeness of the solution.

3.2. The Objective Function

The objective function is a function of the expected number of false negatives, false positives, and tests. In the following, we provide analytical expressions on each of these performance measures, extending those in Aprahamian et al. (2019) to the case where the true risk vector is stochastic and not perfectly observable. We refer the interested reader to Aprahamian et al. (2019) and Appendix B for derivation details.

False Negative Classifications. In individual testing, a false negative classification occurs if a subject is positive and the test outcome is negative. In group testing, on the other hand, a false negative classification occurs if (i) the group contains positive subjects and the group test outcome is negative or (ii) the group contains positive subjects, the group test outcome is positive, and the individual test outcome of a positive subject is negative. Then, conditioned on the estimated risk vector, $\tilde{\boldsymbol{P}}$, and the perturbation vector, $\boldsymbol{\Xi}$, we can write, for a given $\boldsymbol{\Omega}$,

$$\mathbf{E}\big[FN_i(\Omega_i)|\boldsymbol{\Xi},\tilde{\boldsymbol{P}}\big] = \begin{cases} (1-Se)\sum_{m\in\Omega_i} t\big(\tilde{P}^{(m)},\Xi^m\big), & \text{if } n_i = 1, \\ \big(1-Se^2\big)\sum_{m\in\Omega_i} t\big(\tilde{P}^{(m)},\Xi^m\big), & \text{otherwise,} \end{cases} \tag{10}$$

$$\text{and } \mathbf{E}\big[FN(\boldsymbol{\Omega})|\boldsymbol{\Xi},\tilde{\boldsymbol{P}}\big] = \sum_{i=1}^{g} \mathbf{E}\big[FN_i(\Omega_i)|\boldsymbol{\Xi},\tilde{\boldsymbol{P}}\big].$$

False Positive Classifications. In individual testing, a false positive classification occurs if a subject is negative and the test outcome is positive. In group testing, on the other hand, a false positive classification occurs if the group contains negative subjects, the group test outcome is positive, and the individual test outcome of a negative subject is positive. Then, conditioned on the estimated risk vector, $\tilde{\boldsymbol{P}}$, and the perturbation vector, $\boldsymbol{\Xi}$, we have, for a given $\boldsymbol{\Omega}$,

$$\mathbf{E}\big[FP_i(\Omega_i)|\boldsymbol{\Xi},\tilde{\boldsymbol{P}}\big] = \begin{cases} (1-Sp)\sum_{m\in\Omega_i}\big(1-t\big(\tilde{P}^{(m)},\Xi^m\big)\big), & \text{if } n_i = 1, \\ (1-Sp)Se\sum_{m\in\Omega_i}\big(1-t\big(\tilde{P}^{(m)},\Xi^m\big)\big) & \\ -n_i(1-Sp)(Se+Sp-1)\prod_{m\in\Omega_i}\big(1-t\big(\tilde{P}^{(m)},\Xi^m\big)\big), & \text{otherwise,} \end{cases} \tag{11}$$

$$\text{and } \mathbf{E}\big[FP(\boldsymbol{\Omega})|\boldsymbol{\Xi},\tilde{\boldsymbol{P}}\big] = \sum_{i=1}^{g} \mathbf{E}\big[FP_i(\Omega_i)|\boldsymbol{\Xi},\tilde{\boldsymbol{P}}\big].$$

Number of Tests. In individual testing, the per subject number of tests is equal to one. In contrast, the number of tests in group testing relies on the test outcome of the first stage group test. Specifically, if the group test outcome is negative, then only a single test is needed, and if the group test outcome is positive, then additional

individual tests are needed to fully classify the subjects. Then, conditioned on the estimated risk vector, $\tilde{\boldsymbol{P}}$, and the perturbation vector, $\boldsymbol{\Xi}$, we have, for a given $\boldsymbol{\Omega}$,

$$\mathbf{E}\left[T_i(\Omega_i)|\boldsymbol{\Xi},\tilde{\boldsymbol{P}}\right] = \begin{cases} 1, & \text{if } n_i = 1, \\ 1 + n_i\left(Se - \left(Se + Sp - 1\right)\prod_{m\in\Omega_i}\left(1 - t\left(\tilde{P}^{(m)}, \Xi^m\right)\right)\right), & \text{otherwise.} \end{cases} \tag{12}$$

$$\text{and } \mathbf{E}\left[T(\boldsymbol{\Omega})|\boldsymbol{\Xi},\tilde{\boldsymbol{P}}\right] = \sum_{i=1}^{g}\mathbf{E}\left[T_i(\Omega_i)|\boldsymbol{\Xi},\tilde{\boldsymbol{P}}\right].$$

Then, for a given weight vector, $\boldsymbol{\lambda}$, and a testing scheme, $\boldsymbol{\Omega}$, one needs to use the law of total probability to compute the objective functions for each of **EM** and **RM** (see Equations (2) and (3) and the formulations in (7)–(9)). This, however, requires computations of higher-dimensional integrations (up to N-fold), as we discuss in detail in Section 4.1. Moreover, the multiplicative nature of the expressions in Equations (11) and (29) substantially complicates the analysis, as the contribution of a subject to the objective function depends on the set of subjects it is grouped with.

Note that when $\boldsymbol{\lambda} = (1, 0)$ the objective function in Equation (3) reduces to minimizing the expected number of false negative classifications only, whereas when $\boldsymbol{\lambda} = (0, 1)$ the objective function reduces to minimizing the expected number of false positive classifications only. Lastly, when $\boldsymbol{\lambda} = (0, 0)$ the objective function in Equation (3) reduces to minimizing the expected number of tests only, which, as discussed in Section 1, is what is studied in the vast majority of the literature (e.g., Dorfman 1943, Hwang 1975, Saraniti 2006, McMahan et al. 2012). Thus, **EM** and **RM** formulations expand the deterministic formulation in Aprahamian et al. (2019) to consider easily implementable, static testing schemes under risk uncertainty.

4. Structural Properties and Algorithms

Recall that, in its current form, problem **RM** is intractable, as it requires solutions to an infinite number of optimization problems in stage 2 (see (9)), one for each possible realization of the risk vector, $\tilde{\boldsymbol{P}}$, which is continuous. Thus, in what follows, we first provide an equivalent formulation for **RM**. Interestingly, this result also implies that problems **EM** and **RM** both reduce to an optimization problem with a common structure. Then, in the remainder of this section, we exploit this common structure to develop structural properties and effective solution algorithms for both **EM** and **RM**.

4.1. Equivalent Formulations for EM and RM

The following result is essential, as it reduces Problem **RM** from an intractable problem to a problem that is only as difficult as **EM**.

Theorem 1.

1. *For all* $\boldsymbol{\Omega}$ *and* $\tilde{\boldsymbol{P}}$, *there exists an optimal solution to* (9) *such that* $\xi^{m*}(\boldsymbol{\Omega})|\tilde{\boldsymbol{P}}$ *equals either* $-\delta$ *or* δ, *for all* $m = 1, \ldots, N$.
2. *If* $\lambda_1(1 - Se) \geq \lambda_2(1 - Sp)$, *then for all* $\boldsymbol{\Omega}$ *and* $\tilde{\boldsymbol{P}}$, *there exists an optimal solution to* (9) *such that* $\xi^{m*}(\boldsymbol{\Omega})|\tilde{\boldsymbol{P}}$ *equals* δ, *for all* $m = 1, \ldots, N$.

The condition imposed in the second part of Theorem 1 is realistic, as the weight (cost) of a false negative in the objective function, that is, λ_1, is typically much larger than the weight (cost) of a false positive, that is, λ_2, as discussed in Section 2. As such, in the remainder of the paper, we assume that the condition $\lambda_1(1 - Se) \geq \lambda_2(1 - Sp)$ is satisfied.

Corollary 1. If $\lambda_1(1 - Se) \geq \lambda_2(1 - Sp)$, then problem **RM** reduces to the following optimization problem:

$$\begin{aligned} \underset{\boldsymbol{\Omega}=(\Omega_i)_{i=1,\ldots,g},\, g\in\mathbf{Z}^+}{\text{minimize}} \quad & \mathbf{E}_{\tilde{P}}\left[\mathbf{E}\left[Q(\boldsymbol{\Omega})|\boldsymbol{\Xi} = \boldsymbol{\delta}, \tilde{\boldsymbol{P}}\right]\right] \\ \text{subject to} \quad & (4), (5), (6). \end{aligned} \tag{13}$$

Theorem 1 is significant, because it shows that an optimal solution does *not* depend on the distribution of the error term, but rather on its support, $[-\delta, \delta]$; further, it eliminates the need to solve an infinite number of optimization problems in (9) to determine an optimal solution to **RM** and it reduces the two-stage robust formulation in (8)–(9) to a single-stage optimization problem. Moreover, note that the equivalent formulations for problems **EM** and **RM** provided, respectively, in (7) and (13) (Lemma 1 and Corollary 1) have a common

structure, in that the random perturbation vector, $\boldsymbol{\Xi}$, is reduced to a constant vector in both cases: in **EM**, $\boldsymbol{\Xi} = \mathbf{0}$, and in **RM**, $\boldsymbol{\Xi} = \boldsymbol{\delta}$. Hence, both **EM** and **RM** reduce to the following form of an optimization problem.

Common-Form Optimization Model (CM).

$$\begin{aligned} \underset{\boldsymbol{\Omega}=(\Omega_i)_{i=1,\ldots,g},\, g\in\mathbf{Z}^+}{\text{minimize}} \quad & \mathbf{E}_{\tilde{\boldsymbol{P}}}\Big[\mathbf{E}\big[Q(\boldsymbol{\Omega})|\boldsymbol{\Xi}=\boldsymbol{z},\tilde{\boldsymbol{P}}\big]\Big] \\ \text{subject to} \quad & \Omega_i\cap\Omega_j=\emptyset, \quad \forall i,j=1,\ldots,g\colon i\neq j, \\ & \bigcup_{i=1}^{g}\Omega_i=\{1,\ldots,N\}, \\ & \|\boldsymbol{\Omega}\|\le\gamma, \end{aligned} \tag{14}$$

where $\boldsymbol{z}$ is a constant vector, which equals $\mathbf{0}$ for **EM** and $\boldsymbol{\delta}$ for **RM**, and the objective function is given by

$$\mathbf{E}_{\tilde{\boldsymbol{P}}}\Big[\mathbf{E}\big[Q(\boldsymbol{\Omega})|\boldsymbol{\Xi}=\boldsymbol{z},\tilde{\boldsymbol{P}}\big]\Big]=\int_a^b\int_{\tilde{p}_1}^b\cdots\int_{\tilde{p}^{N-2}}^b\int_{\tilde{p}^{N-1}}^b\mathbf{E}\big[Q(\boldsymbol{\Omega})|\boldsymbol{\Xi}=\boldsymbol{z},\tilde{\boldsymbol{P}}=\tilde{\boldsymbol{p}}\big]\,f_{\tilde{P}^{(1)},\ldots,\tilde{P}^{(N)}}\big(\tilde{p}^1,\ldots,\tilde{p}^N\big)d\tilde{p}^N\cdots d\tilde{p}^1,$$

where

$$\mathbf{E}\big[Q(\boldsymbol{\Omega})|\boldsymbol{\Xi}=\boldsymbol{z},\tilde{\boldsymbol{P}}\big]=\lambda_1\mathbf{E}\big[FN(\boldsymbol{\Omega})|\boldsymbol{\Xi}=\boldsymbol{z},\tilde{\boldsymbol{P}}\big]+\lambda_2\mathbf{E}\big[FP(\boldsymbol{\Omega})|\boldsymbol{\Xi}=\boldsymbol{z},\tilde{\boldsymbol{P}}\big]+(1-\lambda_1-\lambda_2)\mathbf{E}\big[T(\boldsymbol{\Omega})|\boldsymbol{\Xi}=\boldsymbol{z},\tilde{\boldsymbol{P}}\big],$$

and $\mathbf{E}[FN(\boldsymbol{\Omega})|\boldsymbol{\Xi}=\boldsymbol{z},\tilde{\boldsymbol{P}}]$, $\mathbf{E}[FP(\boldsymbol{\Omega})|\boldsymbol{\Xi}=\boldsymbol{z},\tilde{\boldsymbol{P}}]$, and $\mathbf{E}[T(\boldsymbol{\Omega})|\boldsymbol{\Xi}=\boldsymbol{z},\tilde{\boldsymbol{P}}]$ are given by Equations (10), (11), and (29), respectively, and $f_{\tilde{P}^{(1)},\ldots,\tilde{P}^{(N)}}(.)$ denotes the joint probability density function of the ordered random variables $\tilde{P}^{(1)},\tilde{P}^{(2)},\ldots,\tilde{P}^{(N)}$.

As stated earlier, problem **CM** is challenging for two main reasons. First, it is at least as hard as the partitioning problem, which is *NP*-hard (Chakravarty et al. 1982). Second, the evaluation of the objective function for a given solution, $\boldsymbol{\Omega}$, requires the computation of up to N-fold integrations, which are computationally expensive. Therefore, in this section, we explore important structural properties of **CM**. Toward this end, consider the following definition.

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

This definition allows the representation of an ordered testing scheme $\boldsymbol{\Omega}=(\Omega_i)_i$ in terms of a vector corresponding to the group size, $\boldsymbol{n}=(n_i)_i$, as groups are constructed (i.e., subjects in each batch are assigned to groups) following the risk-ordered set, S.

Our main results for **CM** are given in Theorems 2 and 3 and Lemma 2.

Theorem 2. *For all $N\in\mathbf{Z}^+$ and $\gamma\in\mathbf{Z}^+$, there exists an optimal solution to* **CM** *in which the testing scheme is an ordered testing scheme.*

Theorem 2 expands the previous results in Aprahamian et al. (2019) to static testing schemes under risk uncertainty. Theorem 2 can be proven by an interchange argument, which is used to transform any unordered partition into an ordered partition. By Theorem 2, to determine an optimal risk-based testing solution, it is sufficient to consider the ordered testing schemes. This result is important in two ways. First, it allows us to reformulate problem **CM** as a constrained-shortest path (**C-SP**) problem. Although **C-SP** is *NP*-hard (Garcia 2009), the equivalent **C-SP**-type formulation enables us to characterize important structural properties of the risk-based testing problem, allowing us to efficiently solve the problem for realistic problem sizes. Second, recall that the objective function in **CM** includes the expected number of false positive classifications and the number of tests, and the expressions for each term contain products of some function of a set of order statistics (see Eqs. (11) and (29)). However, Theorem 2 indicates that these expressions need to be evaluated for products of functions of *consecutive* order statistics and not any set of order statistics. This result turns out to be very useful, as we are able to exploit this property in Lemma 2 to reduce the higher-dimensional (up to N-fold) integrations required to compute those expectations into 3-fold integrations, substantially improving the efficiency with which the **CM** objective function can be evaluated.

As a side note, Theorem 2 highlights an additional benefit of optimal static risk-based testing schemes for practitioners: the tester does not need to evaluate the exact risk of each subject; rather it is sufficient to

determine a risk-ordering of the subjects. This greatly facilitates the implementation of static risk-based testing schemes.

Lemma 2. *Consider N iid continuous random variables, each with a continuous probability density function $f_X(\cdot)$, cumulative distribution function $F_X(\cdot)$, and support $[a,b]$:$0 \le a < b \le 1$. Let $X^{(1)} \le X^{(2)} \le \cdots \le X^{(N)}$ denote the order statistics, and let $f_{X^{(i)},\ldots,X^{(j)}}(\cdot)$ and $f_{X^{(i)},X^{(j)}}(\cdot)$, respectively, represent the joint probability density functions of the ordered random variables, $X^{(i)} \le \cdots \le X^{(j)}$, and of $X^{(i)} \le X^{(j)}$, $i < j$. Let $g(\cdot)$ denote any continuous function. Then, for all $N \ge 4$ and $i, j = 1 \ldots, N$:$i < j$, we have:*

$$\mathbf{E}\left[\prod_{m=i}^{j} g\left(X^{(m)}\right)\right] = \int_a^b \int_a^{x^j} \cdots \int_a^{x^{i+1}} g\left(x^i\right) g\left(x^{i+1}\right) \cdots g\left(x^j\right) f_{X^{(i)},\ldots,X^{(j)}}\left(x^i, \ldots, x^j\right) dx^i \cdots dx^j$$
$$= \int_a^b \int_a^{x^j} g\left(x^i\right) g\left(x^j\right) \left[\int_{x^i}^{x^j} \frac{g(x) f_X(x) dx}{F_X(x^j) - F_X(x^i)}\right]^{j-i-1} f_{X^{(i)},X^{(j)}}\left(x^i, x^j\right) dx^i dx^j.$$

Lemma 2 follows because, by conditioning on the values of the lowest and highest order statistics and by exploiting the structure of the integral, we are able to recursively reduce its dimensionality. Thus, the continuous distribution assumption of the risk random variable leads to the expressions in Lemma 2, greatly facilitating the evaluation of the **CM** objective function. In contrast, if one considers a discrete (categorical) risk distribution, then such integration-based manipulations will no longer be valid, and one has to keep track of all possible risk realizations for all N subjects, significantly increasing the difficulty of evaluating the objective function value. For example, even with only two risk categories, for example, low and high risk, the total number of risk possibilities increases exponentially with the number of subjects, N. In Appendix C, we demonstrate that a continuous risk distribution has the potential of accurately emulating the statistical behavior of a discrete risk distribution. Doing so enables the use of Lemma 2, which significantly simplifies the analysis.

In the following, we provide an equivalent, **C-SP**-type formulation for **CM**.

Remark 2. For a given $\boldsymbol{y} = (y_j)_{j=1,\ldots,N}$, the problem of finding a feasible decomposition, $\boldsymbol{\Omega} = (\Omega_i)_{i=1,\ldots,g}$, that corresponds to vector $\boldsymbol{y}$, that is, $\forall j = 1, \ldots, N$, $y_j = 1$ only if there exists at least one i, $i = 1, \ldots, g$, such that $n_i = j$, reduces to a shortest path (**SP**) problem defined on an acyclic directed graph $G = (V, E)$, with vertex set $V = \{1, \ldots, N+1\}$, edge set $E = \{(i,j) \in V : y_{j-i} = 1\}$, and edge costs given by

$$\begin{cases} \mathbf{E}_{\tilde{P}}\left[\mathbf{E}\left[Q_i\left(S_{i-j}\right) | \boldsymbol{\Xi} = \mathbf{0}, \tilde{\boldsymbol{P}}\right]\right], & \text{for Problem } \mathbf{EM}, \\ \mathbf{E}_{\tilde{P}}\left[\mathbf{E}\left[Q_i\left(S_{i-j}\right) | \boldsymbol{\Xi} = \boldsymbol{\delta}, \tilde{\boldsymbol{P}}\right]\right], & \text{for Problem } \mathbf{RM}, \end{cases}$$

Theorem 3. *Problem* **CM** *can be equivalently formulated as a* **C-SP** *problem as follows:*

$$\underset{\substack{\boldsymbol{y}=(y_j)_{j=1,\ldots,N,} \\ \boldsymbol{x}=(x_{ij})_{(i,j)\in E}}}{\text{minimize}} \quad \sum_{(i,j)\in E} \mathbf{E}_{\tilde{P}}\left[\mathbf{E}\left[Q_i\left(S_{i-j}\right) | \boldsymbol{\Xi} = \boldsymbol{z}, \tilde{\boldsymbol{P}}\right]\right] x_{ij}$$

$$\text{subject to} \quad \sum_{j\in V: j>i} x_{ij} - \sum_{j\in V: j<i} x_{ji} = \begin{cases} 1, & \text{if } i = 1, \\ -1, & \text{if } i = N+1, \\ 0, & \text{otherwise}, \end{cases} \qquad \forall i \in V,$$

$$\sum_{j\in V: j>i} x_{ij} \le 1, \qquad \forall i \in V, \tag{15}$$

$$x_{kl} \le y_{l-k}, \qquad \forall (k,l) \in E, \tag{16}$$

$$\sum_{j=1}^{N} y_j \le \gamma, \tag{17}$$

$$y_j \le \sum_{(k,l)\in E:\ l-k=j} x_{kl}, \quad \forall j = 1, \ldots, N, \tag{18}$$

$$y_j \in \{0,1\}, \qquad \forall j = 1, \ldots, N, \tag{19}$$

$$x_{ij} \in \{0,1\}, \qquad \forall (i,j) \in E,$$

where the function $Q_i(\cdot)$ is as defined in Equation (2); $S_{i-j} = \{i, \ldots, j-1\}$, for all $(i,j) \in E$; y_j, $j = 1, \ldots, N$, is 1 if a group of size j is utilized and 0 otherwise; x_{ij}, $(i,j) \in E$, is 1 if edge (i,j) is selected, that is, the group, comprised of subjects $\{i, \ldots, j-1\}$, is utilized, and 0 otherwise; and $\boldsymbol{z}$ is a constant vector, which equals $\mathbf{0}$ for **EM** *and $\boldsymbol{\delta}$ for* **RM**.

In summary, the equivalence between the testing/grouping model in **CM** and the network flow formulation provided in Theorem 3 holds in our setting due to two main properties: (1) an optimal solution corresponds to a risk-ordered testing scheme (Theorem 2); thus, one can, without loss of optimality, search over the set of risk-ordered testing schemes (rather than all possible testing schemes); and (2) the objective function is additive by group. In Theorem 3, we utilize these properties to represent the testing problem for N subjects as a network flow model, defined on graph $G = (V, E)$, with vertex set $V = \{1, \ldots, N+1\}$ and edge set $E = \{(i,j) \in V : i < j\}$. Each vertex in V (with the exception of the dummy sink vertex, $N+1$) corresponds to a subject, and an edge from vertex i to vertex j corresponds to a group comprised of subjects i to $j-1$. Note that, under this representation, each path from vertex 1 to $N+1$ corresponds to an ordered testing scheme, and the set of all paths from vertex 1 to vertex $N+1$ represents the set of all possible risk-ordered testing schemes. Consequently, by setting the weight of each edge in E as the contribution of its corresponding group to the objective function, one can obtain the optimal testing scheme by solving a constrained-shortest path problem on graph $G = (V, E)$, provided in Theorem 3.

Remark 3.

1. To construct graph $G = (V, E)$ for the **CM** formulation in Theorem 3, one needs to compute $N(N+1)/2$ edge costs, where the cost of each edge $(i,j) \in E$, that is, $\mathbf{E}[Q_i(S_{i-j})]$, requires $(j-i)$-fold integration. Thus, Lemma 2 greatly facilitates the construction of this graph by allowing all higher-dimensional integrations, with $j - i \geq 4$, to be computed via 3-dimensional integrations.
2. If constraints (16)–(19), which limit the number of allowable distinct group sizes, are relaxed, then **CM** reduces to an **SP** problem, which, for an acyclic graph, can be solved in polynomial time via, for example, a topological sorting algorithm in $\mathcal{O}(|V| + |E|) = \mathcal{O}(N^2)$ (Cormen et al. 2009). Although such an algorithm runs in quadratic time, one must still construct the graph by computing all edge costs, and Lemma 2 substantially reduces the computational effort required for constructing the graph.

We note here that the total unimodularity property, satisfied for **SP**, no longer holds with the addition of constraints (16)–(19). Nevertheless, in what follows, we show that the integrality constraint can still be relaxed for a large set of decision variables, while preserving the integrality of the optimal solution.

Lemma 3. *The integrality constraint for $\boldsymbol{x}$ in (15) can be relaxed without loss of optimality.*

As a result of Lemma 3, integrality constraints are needed only on the $\boldsymbol{y}$ variables, and hence, the number of binary decision variables in **CM** grows only linearly with problem size, improving the computational efficiency.

In the next section, we utilize the properties developed in this section to determine optimal testing schemes for our case study and discuss our findings.

5. Case Study: Chlamydia Screening in the United States

In this section, we demonstrate the value of static risk-based group testing schemes through a case study on chlamydia screening. Chlamydia is among the most prevalent STDs in the U.S. (Centers for Disease Control and Prevention Division of STD Prevention 2014), and existing screening practices differ substantially across states. For instance, certain states screen only a subset of their population, for example, many states follow the U.S. Preventive Services Task Force (USPSTF) recommendations to screen sexually active females of 24 years of age or younger and older females who are at high risk (U.S. Preventive Services Task Force 2020). This focus on females is partial because the consequences of chlamydia can be more severe for females, including infertility, and thus, this is a harm-mitigation strategy. North Carolina (North Carolina State Laboratory of Public Health 2016) utilizes individual testing for the USPSTF-recommended population, whereas Iowa and Idaho utilize groups of size four (see Lewis et al. 2012, Tebbs et al. 2013, Family Planning Council of Iowa 2020). It is worth noting that by screening we imply the testing of asymptomatic subjects for chlamydia. For instance, Iowa state's laboratory does individual testing for subjects that have chlamydia symptoms, but we consider this diagnostic testing, not screening. Further, whether testing is done through the state laboratory can depend on many factors, including insurance status (e.g., see Family Planning Council of Iowa 2020). Although these are interesting issues, worthy of research, we consider them out of scope and here just assume a laboratory receiving many daily samples that include some risk information, for example, gender, age, and race/ethnicity.

Our objectives in this case study are multifold:

1. To quantify the benefits of risk-based testing, that is, **EM** and **RM**, in the realistic setting where subject risk is not perfectly observable, over testing schemes that ignore the risk characteristics of the subjects. For the latter class of testing schemes, we consider uniform testing schemes (**UM**), commonly studied in the literature (e.g., Dorfman 1943, Graff and Roeloffs 1972, Kim et al. 2007). Such uniform testing schemes assume that the population is homogeneous, rely solely on the overall prevalence rate in the population for determining a *common* group size, and randomly assign subjects to testing groups. Owing to their simplicity, uniform testing schemes are commonly adopted in practice (e.g., Lewis et al. 2012), which is why we set them as the primary benchmark scheme. Following current practices, in **UM**, if N is not divisible by the common group size, then all leftover subjects are combined into a (smaller) group for testing. We also compare the proposed testing schemes with a testing scheme that does not utilize group testing and, as a result, restricts screening to only a certain group of the population, which can be considered a proxy for the current practices discussed above.
2. To compare the performance of the robust and expectation-based versions of risk-based testing, that is, solutions to **RM** and **EM**, and to quantify the price of robustness for **RM.**
3. To compare the performance of the static risk-based schemes, studied in this paper, to dynamic risk-based schemes (**DM**), that is, testing schemes that are customized (in terms of group sizes and subject assignment to groups) for each testing batch, based on the estimated risk vector for that particular batch (Aprahamian et al. 2019). Of course, due to the additional flexibility, dynamic testing schemes are expected to outperform static schemes, but this may come at a high operational complexity/cost; hence, our goal is to shed light on the potential loss in screening performance from using static schemes, that is, the price of implementability, and to gain insight into the conditions under which static schemes are expected to perform relatively well compared with their dynamic counterparts, that is, cases when the price of implementability is low.
4. To gain insight on the impact of the dilution of grouping (which was assumed negligible in our models) on the benefits of static testing schemes.

The remainder of this section is organized as follows. In Section 5.1, we calibrate our models and discuss the data sources. Then, in Sections 5.2, 5.3, and 5.4 we discuss the findings from our case study in terms of the aforementioned objectives.

5.1. Model Calibration and Data Sources

To derive the risk estimate distribution for chlamydia in the general population (i.e., random variable $\tilde{P}$), we use data from the Centers for Disease Control and Prevention (CDC) for the year 2014 (Centers for Disease Control and Prevention Division of STD Prevention 2014). The data set provides the number of chlamydia cases reported, along with the size of the corresponding subpopulation for each combination of gender, age group, and race/ethnicity group. (The data set contains two gender categories, seven age group categories, and five race/ethnicity categories, leading to a total of 70 categories.) Our analysis of this data set indicates that prevalence rates for chlamydia differ substantially among the 70 categories; hence, in our study we consider each category as a risk group, with risk corresponding to the prevalence rate of that group. In addition, studies show that a large proportion of chlamydia cases go undiagnosed or unreported (Centers for Disease Control and Prevention 2000), and the number of actual cases is estimated to be roughly three times the number of cases reported (Grimes et al. 2013). This underscores the importance of factoring in underreporting, which we do by multiplying the number of reported cases for each subpopulation by 3. This leads to a total prevalence rate that is in line with published rates. One potential drawback of this approach is that it assumes a common underreporting rate for all subpopulations; this assumption is made due to a lack of underreporting data for the different subpopulations.

Based on the histogram of the risk for the CDC data set (see Appendix C), we model the estimated risk, $\tilde{P}$, with a mixture distribution, comprised of two exponential distributions. This particular mixture distribution provides a good fit for the histogram, which closely resembles the probability density function of an exponential distribution, but with a higher coefficient of variation than that of a single exponential distribution (which is one). The probability density function of the mixture distribution follows:

$$f_{\tilde{P}}(p) = w\beta_1 e^{-\beta_1 p} + (1-w)\beta_2 e^{-\beta_2 p}, \quad \forall p \geq 0, \tag{16}$$

where $w \in [0,1]$ is a weight parameter, and $\beta_1, \beta_2 > 0$ are scale parameters corresponding to each exponential distribution. The parameters of the mixture distribution, w, β_1, β_2, are derived by matching the first two moments of the distribution to those of the data set so as to minimize the Kolmogorov-Smirnov statistic (Massey 1951), that is, to minimize the maximum distance between the empirical and the fitted cumulative distribution functions (see Appendix C for details). This method provides parameter values of $w = 0.235$, $\beta_1 = 25.708$, and $\beta_2 = 1{,}291.832$ for the mixture distribution.

We model the relationship between the true (unobservable) risk, P, and the estimated risk, $\tilde{P}$, through a multiplicative model, that is, $P|\tilde{P}, \Xi = t(\tilde{P}, \Xi) = \tilde{P}(1 + \Xi)$, where Ξ is a random error term with support $[-\delta, \delta]$, and δ, which represents the degree of uncertainty in the estimated risk with respect to the true risk, is set to 0.667. This value represents the largest possible value that δ can take while ensuring that all risk realizations are within the interval $[0, 1]$. Given that this choice is somewhat arbitrary, we also conduct a one-way sensitivity analysis on δ to determine how the degree of uncertainty impacts the optimal solutions to **EM** and **RM**.

On the testing side, we utilize a DNA-based assay, known as *Viper ProbeTec Chlamydia* Q^x. This assay is routinely used for the screening of chlamydia, and it can be administered either individually or within groups (Kapala et al. 2000), with a reported sensitivity of $Se = 0.95$ and a specificity of $Sp = 0.99$. We utilize the findings of Owusu-Edusei et al. (2015) and Van Der Pol et al. (2012) to estimate the cost parameter values. In particular, the cost of a single false negative is set to the expected cost of any consequences resulting from not appropriately treating a chlamydia patient (estimated as \$2,927;[3] Owusu-Edusei et al. 2015), while the per test screening cost is set to \$55 (Owusu-Edusei et al. 2015). Lastly, the cost of a single false positive is assumed to be equal to the cost of an additional confirmatory test, which we set to the cost of the screening test.[4] Normalizing these cost parameters leads to $\lambda_1 = 0.96$ and $\lambda_2 = 0.02$. (Hence, $1 - \lambda_1 - \lambda_2 = 0.02$.) In what follows, we illustrate the benefits of our model by considering a testing batch size, N, of 60, as this provides a realistic treatment of the problem (Lewis et al. 2012).

For each *scenario*, characterized by the maximum number of distinct group sizes allowed, γ, we determine the optimal solution for each of problems **UM**, **EM**, and **RM**. In uniform testing schemes, generated by **UM**, the population is assumed to be homogeneous, that is, the risk of each subject is the same and equals the mean prevalence rate of the population, given by 0.97%. For both **EM** and **RM**, we determine an optimal testing scheme that is an ordered testing scheme (see Theorem 2). Thus, for all models, we can represent the testing scheme in terms of its group size vector, $\boldsymbol{n} = (n_i)_i$, because in **EM** and **RM**, groups are constructed (i.e., subjects in each batch are assigned to groups) following the risk-ordered set, S; and in **UM**, groups are constructed in a random fashion (i.e., subjects, which are assumed identical, are assigned to groups randomly). To simplify the presentation of the group size vector, we use the notation $x_{(y)}$ to represent y groups of size x. We also let E-OF denote the expected cost of a testing scheme, that is, E-OF $\equiv \mathbf{E}_{\tilde{P}}[\mathbf{E}[Q|\boldsymbol{\Xi} = \mathbf{0}, \tilde{\boldsymbol{P}}]]$, and we let W-OF denote the worst-case expected cost of a testing scheme, that is, W-OF $\equiv \mathbf{E}_{\tilde{P}}[\mathbf{E}[Q|\boldsymbol{\Xi} = \boldsymbol{\delta}, \tilde{\boldsymbol{P}}]]$. Similarly, we let CE-OF denote the expected cost of a testing scheme *conditioned* on a given realization of the estimated risk vector, that is, CE-OF $\equiv \mathbf{E}[Q|\boldsymbol{\Xi} = \mathbf{0}, \tilde{\boldsymbol{P}}]$, and we let CW-OF denote the worst-case cost of a testing scheme *conditioned* on a given realization of the estimated risk vector, that is, CW-OF $\equiv \mathbf{E}[Q|\boldsymbol{\Xi} = \boldsymbol{\delta}, \tilde{\boldsymbol{P}}]$.

5.2. Risk-Based vs. Non-Risk-Based (Uniform) Schemes

In this section, we compare the performance of **EM** and **RM** to non-risk-based (uniform) testing schemes, that is, solutions to problem **UM**, for various scenarios (see Table 1).

Table 1. Performance Comparison of **UM**, **EM**, and **RM**

Problem **UM**		
$\boldsymbol{n}^* = (11_{(5)}, 5_{(1)})$	E-OF = 0.2976	W-OF = 0.4023
Problem EM		
γ	$\boldsymbol{n}^*$	E-OF (% change[†])	W-OF (% change[†])
1	$(12_{(5)})$	0.2729 (–8.30%)	0.3574 (–11.14%)
2	$(46_{(1)}, 7_{(2)})$	0.2212 (–25.69%)	0.3093 (–23.11%)
3	$(41_{(1)}, 11_{(1)}, 4_{(2)})$	0.2143 (–27.98%)	0.2908 (–27.71%)
≥ 4	$(38_{(1)}, 12_{(1)}, 6_{(1)}, 4_{(1)})$	0.2129 (–28.45%)	0.2881 (–28.38%)
Problem RM		
γ	$\boldsymbol{n}^*$	E-OF (% Change[†])	W-OF (% Change[†])
1	$(10_{(6)})$	0.2769 (–6.95%)	0.3557 (–11.57%)
2	$(24_{(2)}, 4_{(3)})$	0.2264 (–23.91%)	0.2981 (–25.90%)
3	$(23_{(2)}, 6_{(2)}, 1_{(2)})$	0.2278 (–23.45%)	0.2877 (–28.48%)
4	$(34_{(1)}, 14_{(1)}, 5_{(2)}, 1_{(2)})$	0.2236 (–24.87%)	0.2817 (–29.98%)
≥ 5	$(34_{(1)}, 14_{(1)}, 6_{(1)}, 4_{(1)}, 1_{(2)})$	0.2224 (–25.27%)	0.2802 (–30.35%)

[†]Percent change over **UM**.

First, observe that, for scenarios with $\gamma = 1$ (i.e., with only one group size allowed), the optimal group sizes in **EM** and **UM** are not necessarily equal (e.g., the single group size in **UM** is equal to 11, whereas the single group size in **EM**, when $\gamma = 1$, is 12). This difference in group sizes arises due to the ordering of the estimated risk vector in **EM**, that is, the optimal group size under a random assignment policy (in **UM**) differs from the optimal group size under an ordered assignment policy (in **EM**). Note that the number of distinct group sizes is strictly enforced for **EM** and **RM**, but not **UM**. Also, observe that, when the maximum number of distinct group sizes exceeds five (i.e., $\gamma \geq 5$), no additional benefits are realized in **EM** and **RM** solutions, that is, the solutions converge to the solution with $\gamma = 5$ (and in some scenarios, this convergence happens faster, that is, for $\gamma \geq 4$; see Table 1).

The results in Table 1 highlight several important properties. First, both **EM** and **RM** substantially reduce both the expectation and the worst case of the cost over **UM**. For example, for all $\gamma \geq 5$, comparing **EM** (**RM**) with **UM**, we observe substantial reductions in both the expected cost and the worst-case cost, respectively, by 28% (25%) and 28% (30%) over **UM**. Even for the most restrictive case of $\gamma = 1$, that is, with only one group size allowed, **EM** still reports reductions in the expected cost. Also, observe that both the optimal expected cost (optimal solution to **EM**) and the worst-case cost (optimal solution to **RM**) reduce as γ increases, but in both cases, the reductions exhibit diminishing returns, with a substantial reduction occurring when γ increases from one to two and with all subsequent reductions being much smaller in magnitude (see Figure 1). For example, if the solution of **EM** were used with two allowable group sizes, then costs are expected to reduce by around 23% over a uniform (non-risk-based) testing scheme. This finding has important implications, as schemes with only two group sizes are easier to implement in practice, making them especially appealing to practitioners. Following the recommendation of the USPSTF, we also consider a testing scheme in which only a subset of the population (in this case sexually active females of 24 years of age or younger) is tested individually, and all others, in the absence of testing, are classified as negative. This policy, when applied to the CDC data set, leads to an objective function value of 0.4874, which is 76%–129% higher than the proposed static testing schemes, and hence performs much worse than both the static and uniform testing schemes.

Next, we compare the costs incurred in **EM** and **RM** solutions so as to quantify the price of robustness and analyze the trade-off in expected classification accuracy versus solution robustness. According to Table 1, **RM** leads to a 3.9% increase in the expected cost over **EM**; equivalently, the price of robustness is 3.9%, whereas **EM** leads to a 2.1% increase in the worst-case expected cost over **RM**. Examining the **RM** and **EM** solutions in detail, we observe another advantage of **RM**, in that **RM** has a tendency to reduce the expected number of misclassifications (false positives plus false negatives) over **EM**. Specifically, **RM** reduces the expected number of misclassifications over **EM** by an average of 4.7%, but this comes at a cost, as **RM** increases the expected number of tests over **EM** by 9%. In conclusion, both **EM** and **RM** provide substantial benefits over non-risk-based testing schemes, such as **UM**. Depending on the setting, the added benefits of a robust testing solution may outweigh the observed increase in the expected cost produced by the **RM** solution.

Next, we study how the testing performance varies with δ, that is, the degree of uncertainty in the estimated risk with respect to the true risk. For this purpose, we conduct a one-way sensitivity analysis on δ and determine **UM**, **EM**, and **RM** optimal solutions for various values of δ in $\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$ (see Figure 2 for the case with $\gamma \geq 5$, i.e., the case where there is effectively no limit on the number of distinct group sizes; see the discussion above). Our results indicate that both **EM** and **RM** substantially improve both objective functions over **UM** for all values of δ. However, the performance of **EM** and **RM** turns out to be similar in this case study, suggesting that **EM** captures almost all benefits of a static risk-based testing scheme over **UM**, at least in this case study.

5.3. Static Risk-Based Schemes vs. Dynamic Risk-Based Schemes

Having quantified the value of risk-based testing, in this section we compare the performance of the static **EM** model to dynamic risk-based schemes, **DM** (see Aprahamian et al. 2019), in which the decision-maker

Figure 1. Expected Cost (E-OF) (Left) and Worst-Case Cost (W-OF) (Right) for **UM**, **EM**, and **RM**, as a Function of γ

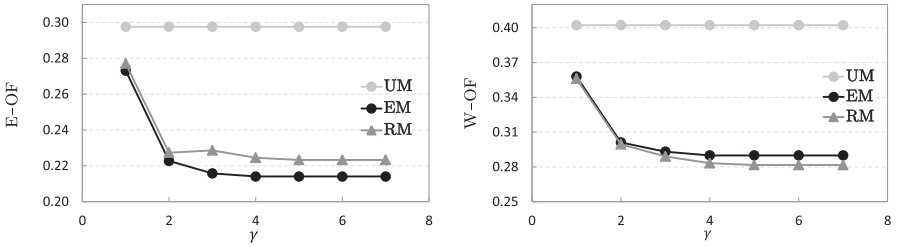
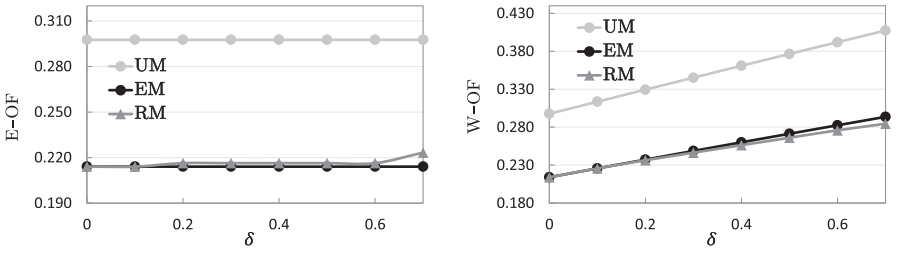


Figure 2. Expected Cost (E-OF) (Left) and Worst-Case Cost (W-OF) (Right) for **UM**, **EM**, and **RM**, as a Function of δ for All $\gamma \geq 5$



customizes the testing scheme to each batch, that is, in **DM**, the decision-maker first observes the estimated risk vector for the batch and then optimizes accordingly. The **DM** testing scheme will outperform the **EM** scheme; here we want to quantify the reduction in performance by using a static scheme, that is, to quantify the price of implementability. Toward this end, we run a Monte Carlo simulation with 10,000 replications. Although the optimal static **EM** solution is computed only once, prior to the simulations (as in the previous section), an optimal **DM** solution is computed for each batch, that is, in each replication, a random realization of the estimated risk vector for a batch is generated following the distribution in Equation (16), the optimal **DM** solution is determined for this specific estimated risk vector, and the resulting expected costs are computed for the optimal **EM** and **DM** solutions. Figure 3 depicts the histogram of the conditional expected cost, CE-OF, for **EM** and **DM** for all $\gamma \geq 5$. Figure 3 illustrates an interesting finding; the restriction to a static scheme has minor cost implications. In Figure 3, the increase in expected costs under **EM** is only 2.2%± 0.05% over **DM**, whereas **EM** is significantly easier to implement. This implies that a static risk-based testing scheme captures most of the benefits of dynamic risk-based testing, while greatly simplifying the implementation, making static testing schemes appealing for practitioners.

Although the above analysis is certainly promising, it does not, however, convey any information as to how such observations extend to other contexts. In what follows, in an effort to have a better understanding of the impact of problem parameters on the price of implementability, we conduct a two-dimensional sensitivity analysis. Specifically, we repeat the aforementioned Monte Carlo simulation (of 10,000 replications) under nine distinct settings that are characterized by three test accuracy settings (low accuracy, $(Se, Sp) = (0.6, 0.6)$, moderate accuracy, $(Se, Sp) = (0.8, 0.8)$, and high accuracy, $(Se, Sp) = (0.95, 0.99)$), and three testing population sizes ($N = 20$, $N = 60$, and $N = 100$). Our choice of conducting a sensitivity analysis on parameters (Se, Sp) and N is based on our numerical experiments, which demonstrate that the price of implementability is most heavily impacted by (Se, Sp) and N. Table 2 displays the percent increase in expected cost resulting from optimal static testing schemes, compared with optimal dynamic testing schemes for each setting. The numerical results reveal that lower test accuracy leads to smaller differences between static and dynamic testing schemes. For example, when $N = 20$, the gap is equal to 4.5%±0.16% for high-accuracy tests and reduces to 1.4%±0.04% for low-accuracy tests; a similar trend can be observed for $N = 60$ and $N = 100$. Further, increasing the number of subjects reduces the gap between static and dynamic schemes. For example, for low-accuracy tests, the gap reduces from 1.4%±0.04% when $N = 20$ to only 0.4%±0.01% when $N = 100$; a similar trend can be observed for moderate- and high-accuracy tests. However, note that the largest gap across all settings is still relatively low at 4.5%±0.16%. In summary, this analysis demonstrates how the proposed static testing scheme

Figure 3. Histogram of the Conditional Expected Cost (CE-OF) for **EM** and **DM** for All $\gamma \geq 5$

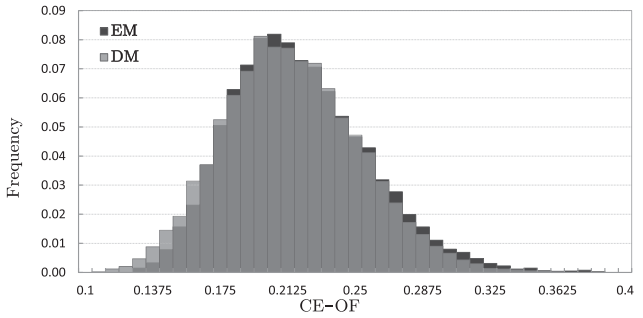


Table 2. Performance Comparison of Static and Dynamic Testing Schemes, in Terms of the Price of Implementability (Percent Increase in Expected Cost of Optimal Static Schemes Compared with Optimal Dynamic Schemes) for Various Settings

	$N = 20$	$N = 60$	$N = 100$
$(Se, Sp) = (0.6, 0.6)$	1.4% ± 0.04%	0.5% ± 0.01%	0.4% ± 0.01
$(Se, Sp) = (0.8, 0.8)$	4.4% ± 0.09%	1.0% ± 0.02%	0.6% ± 0.02
$(Se, Sp) = (0.95, 0.99)$	4.5% ± 0.16%	2.2% ± 0.05%	1.6% ± 0.03%

Note. The error terms represent the half widths of the 95% confidence intervals.

is not only simpler, but also an effective alternative to the more complicated dynamic testing scheme; this is especially true when the number of subjects is large and/or the test accuracy is low.

5.4. The Effect of Dilution

So far, we have assumed that the dilution effect of grouping is negligible, that is, the test sensitivity does not change with group size. In general, the magnitude of the dilution effect depends on the infection and the testing technology, and in certain settings, the dilution effect can be nonnegligible for sufficiently large group sizes, reducing the test's sensitivity. Therefore, we next investigate the impact of dilution on static testing schemes.

There are a number of ways dilution can be incorporated into our models. When dilution for group sizes below a certain threshold is negligible, as is the case for chlamydia screening via a DNA-based assay with group sizes up to 10 (e.g., McMahan et al. 2012), one can set an upper bound on the allowable group size. In our network flow reformulation (Theorem 3), this implies removing a subset of the edges from the graph representation of the problem. Specifically, since each edge corresponds to a group in our graph, all edges corresponding to groups violating the upper bound constraint can be removed. Our numerical experiments with an upper bound of 10 on group size indicate that the proposed static scheme still reduces the overall cost by more than 12% over uniform testing schemes, demonstrating that static testing schemes continue to offer significant benefits in the presence of dilution.

In the more general setting where the dilution effect impacts the test sensitivity for all group sizes, one can explicitly model the test's sensitivity as a function of group size, n, that is, $Se(n)$. All the analytical results in Section 4 continue to hold in this case, and as a result we can take advantage of both the problem reformulation and the elimination of high-dimensional integrals (Lemma 2 and Theorem 3) to solve the resulting problem, as long as $Se(n) + Sp \geq 1, \forall n$—since, by the definition of the dilution effect, $Se(n)$ is nonincreasing in n, this condition can be guaranteed by, for instance, setting an upper bound on group size. To demonstrate the impact of a group-dependent sensitivity function, we investigate our case study setting using published data on the sensitivity of DNA-based assays as a function of group size (Stramer et al. 2013). For illustrative purposes, we use linear regression to model the test's sensitivity ($Se(n)$) as a linear function of group size (with a correlation coefficient of 0.98; see Appendix D). Our numerical study with this linear sensitivity function leads to findings similar to the no-dilution case; for example, under the dilution effect, the proposed static testing scheme still reduces the overall cost by 34% over the uniform testing scheme. A main difference, however, is that the model that considers dilution tends to place lower risk subjects in larger groups and higher risk subjects in smaller groups. This is intuitive: due the reduced sensitivity under dilution, larger groups have less accurate test outcomes. Consequently, the optimal solution places risky subjects in smaller groups so as to have a high level of accuracy, and this comes at the expense of placing some lower risk subjects in larger groups. Although these are interesting observations, we do note, however, that the differences are subtle as the overall structures of the two solutions remain similar. In fact, evaluating the objective function value of the no-dilution model's optimal solution under the dilution effect reveals that the no-dilution optimal solution still performs exceptionally well and reduces the cost by 30% over the uniform testing scheme. This indicates that considering dilution does not lead to drastic differences in optimal grouping and that the no-dilution model continues to provide "good" solutions.

6. Conclusions and Suggestions for Future Research

We develop novel models for determining optimal static risk-based Dorfman testing schemes under imperfectly observable subject risk, with the objective of accurately and efficiently classifying a set of subjects as positive or negative for a binary characteristic. Our models take into account important test and population-level characteristics and generate easily implementable risk-based testing schemes. Although these problems

can be modeled as partitioning problems, we derive various key structural properties of their optimal solutions and reduce them into network flow problems; this allows us to obtain optimal risk-based testing schemes for realistic problem sizes. Further, our novel expression for the expected value of the product of a function of a set of consecutive order statistics enables us to substantially improve the efficiency with which the corresponding graph can be constructed. We also explore a novel robust formulation, an important special case of which we are able to solve to optimality. Our case study, on chlamydia screening in the US, demonstrates the effectiveness of static risk-based testing schemes, which substantially reduce the costs of misclassification and testing over current screening practices, while significantly improving the robustness of the solution. Such results highlight the importance of customizing testing procedures based on the characteristics of the population (e.g., risk distribution) and the characteristics of the testing equipment (e.g., batch sizes, level of automation).

There are several important extensions of this research effort. We consider a purely static testing scheme, comprised of group sizes and a subject assignment policy that is used repetitively for each testing batch. One might consider various partially dynamic testing schemes in which some components of the testing scheme may be customized for the specific batch. Further research directions may include improving the realism of the model. Our models assume that the dilution effect is negligible. The magnitude of the dilution effect is context dependent, varies substantially with the testing technology and the infection, and in certain settings, can be nonnegligible for sufficiently large group sizes, reducing the test's sensitivity. One may study the testing scheme optimization model under a test sensitivity function that varies with group size. One can also expand this work to consider other group testing schemes, such as multistage hierarchical schemes or schemes that take advantage of overlapping groups (e.g., array-based grouping schemes; Kim et al. 2007). Although such schemes may be more complicated to implement, they have the potential to outperform Dorfman testing schemes, and the complexity versus benefit trade-off needs to be studied. Another important direction is to determine an optimal batch size, which was considered exogeneous in our models, considering the trade-off between fixed costs and the delay in obtaining test outcomes. Finally, a promising research direction is to utilize group testing for the purpose of risk estimation, where important research questions arise on how the population should be clustered into different risk groups (subpopulations) and what risk value should be assigned to each of these subpopulations.

We hope that this work, which indicates that the benefits of static risk-based group testing schemes can be substantial, encourages academicians and practitioners to further study static risk-based testing schemes.

Acknowledgments

The authors thank Professor Shane Henderson for providing highly insightful feedback and excellent suggestions on this paper. They also thank the review team for excellent comments and suggestions. These suggestions have improved both the presentation and the analysis in the paper. The authors thank Dr. Scott J. Zimmerman, Director of the North Carolina State Laboratory of Public Health, for many valuable discussions and insights into public health screening practices. 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.

Appendix A. Mathematical Proofs

Proof of Lemma 1. Consider the **EM** objective function given by

$$\sum_{i=1}^{g} \mathbf{E}_{\tilde{P}}\Big[\mathbf{E}_{\Xi}\Big[\lambda_1 \mathbf{E}\big[FN_i(\Omega_i)|\Xi, \tilde{P}\big] + \lambda_2 \mathbf{E}\big[FP_i(\Omega_i)|\Xi, \tilde{P}\big] + (1-\lambda_1-\lambda_2)\mathbf{E}\big[T_i(\Omega_i)|\Xi, \tilde{P}\big]\Big]\Big].$$

Next, we show that each term in the objective function reduces to the case when $\Xi = \mathbf{0}$:

$$\begin{aligned}
\mathbf{E}_{\Xi}\big[\mathbf{E}\big[FN_i(\Omega_i)|\Xi, \tilde{P}\big]\big] &= \begin{cases} (1-Se)\sum_{m\in\Omega_i} \mathbf{E}_{\Xi^m}\big[t\big(\tilde{P}^{(m)}, \Xi^m\big)\big], & \text{if } n_i = 1, \\ \big(1-Se^2\big)\sum_{m\in\Omega_i} \mathbf{E}_{\Xi^m}\big[t\big(\tilde{P}^{(m)}, \Xi^m\big)\big], & \text{otherwise,} \end{cases} \\
&= \begin{cases} (1-Se)\sum_{m\in\Omega_i} \tilde{P}^{(m)}, & \text{if } n_i = 1\Big(\text{by assumption that } \mathbf{E}_{\Xi^m}\big[t\big(\tilde{P}^m, \Xi^m\big)\big] = \tilde{P}^m\Big), \\ \big(1-Se^2\big)\sum_{m\in\Omega_i} \tilde{P}^{(m)}, & \text{otherwise,} \end{cases} \\
&= \mathbf{E}\Big[FN_i(\Omega_i)|\Xi = \mathbf{0}, \tilde{P}\Big],
\end{aligned}$$

$$\mathbf{E}_{\Xi}\left[\mathbf{E}\left[FP_i(\Omega_i)|\Xi,\tilde{\boldsymbol{P}}\right]\right]=\begin{cases}(1-Sp)\sum_{m\in\Omega_i}\left(1-\mathbf{E}_{\Xi^m}\left[t\left(\tilde{P}^{(m)},\Xi^m\right)\right]\right), & \text{if } n_i=1,\\ (1-Sp)Se\sum_{m\in\Omega_i}\left(1-\mathbf{E}_{\Xi^m}\left[t\left(\tilde{P}^{(m)},\Xi^m\right)\right]\right) & \\ -n_i(1-Sp)(Se+Sp-1)\mathbf{E}_{\Xi^m}\left[\prod_{m\in\Omega_i}\left(1-t\left(\tilde{P}^{(m)},\Xi^m\right)\right)\right], & \text{otherwise.}\end{cases}$$

Noting that random variables Ξ^m, $m=1,\ldots,N$, are iid and that $\mathbf{E}_{\Xi^m}[t(\tilde{P}^m,\Xi^m)]=\tilde{P}^m$, we can write

$$\begin{aligned}\mathbf{E}_{\Xi}\left[\mathbf{E}\left[FP_i(\Omega_i)|\Xi,\tilde{\boldsymbol{P}}\right]\right]&=\begin{cases}(1-Sp)\sum_{m\in\Omega_i}\left(1-\tilde{P}^{(m)}\right), & \text{if } n_i=1,\\ (1-Sp)Se\sum_{m\in\Omega_i}\left(1-\tilde{P}^{(m)}\right)-n_i(1-Sp)(Se+Sp-1)\prod_{m\in\Omega_i}\left(1-\tilde{P}^{(m)}\right), & \text{otherwise,}\end{cases}\\ &=\mathbf{E}\left[FP_i(\Omega_i)|\boldsymbol{\Xi}=\mathbf{0},\tilde{\boldsymbol{P}}\right],\end{aligned}$$

Following a similar logic, it can be shown that $\mathbf{E}_{\Xi}[\mathbf{E}[T_i(\Omega_i)|\Xi,\tilde{\boldsymbol{P}}]]=\mathbf{E}[T_i(\Omega_i)|\boldsymbol{\Xi}=\mathbf{0},\tilde{\boldsymbol{P}}]$, thus concluding the proof. □

Proof of Theorem 1. Suppose, to the contrary, that in an optimal solution to (9), denoted by $\boldsymbol{\xi}^*$, there exists a subject, denoted by m, in group i, such that $-\delta<\xi^{m*}<\delta$. In what follows, we show that one can always improve the objective function value to (9) by either increasing ξ^{m*} to δ or decreasing ξ^{m*} to $-\delta$. Toward this end, consider an alternative solution, denoted by $\tilde{\boldsymbol{\xi}}$, which is identical to $\boldsymbol{\xi}^*$, with the only exception that $\tilde{\xi}^m=\xi^{m*}+\varepsilon$, for some $|\varepsilon|>0$.

Case I: $n_i=1$

By using the expressions in Section 3.2, the contribution of group i to the objective function in (9) is given by

$$\begin{aligned}Q_i\left(\boldsymbol{\xi}^*\right)&=t\left(\tilde{P}^{(m)},\xi^{m*}\right)\left[\lambda_1(1-Se)-\lambda_2(1-Sp)\right]+1-\lambda_1-\lambda_2+\lambda_2(1-Sp),\\ Q_i\left(\tilde{\boldsymbol{\xi}}\right)&=t\left(\tilde{P}^{(m)},\tilde{\xi}^m\right)\left[\lambda_1(1-Se)-\lambda_2(1-Sp)\right]+1-\lambda_1-\lambda_2+\lambda_2(1-Sp)\\ \Rightarrow Q_i\left(\tilde{\boldsymbol{\xi}}\right)-Q_i\left(\boldsymbol{\xi}^*\right)&=\left[t\left(\tilde{P}^{(m)},\tilde{\xi}^m\right)-t\left(\tilde{P}^{(m)},\xi^{m*}\right)\right]\left[\lambda_1(1-Se)-\lambda_2(1-Sp)\right].\end{aligned}$$

Subcase I: $\lambda_1(1-Se)\geq\lambda_2(1-Sp)$. Let $\varepsilon=\delta-\xi^{m*}>0\Rightarrow\tilde{\xi}^m=\delta$, and hence, we get that

$$Q_i\left(\tilde{\boldsymbol{\xi}}\right)-Q_i\left(\boldsymbol{\xi}^*\right)=\left[t\left(\tilde{P}^{(m)},\delta\right)-t\left(\tilde{P}^{(m)},\xi^{m*}\right)\right]\left[\lambda_1(1-Se)-\lambda_2(1-Sp)\right]\geq 0,$$

since $\lambda_1(1-Se)-\lambda_2(1-Sp)\geq 0$, $\xi^{m*}<\delta$, and $t(\tilde{p},\xi)$ is increasing in ξ by assumption.

Subcase II: $\lambda_1(1-Se)<\lambda_2(1-Sp)$. Let $\varepsilon=-\delta-\xi^{m*}<0\Rightarrow\tilde{\xi}^m=-\delta$, and hence, we get that

$$Q_i\left(\tilde{\boldsymbol{\xi}}\right)-Q_i\left(\boldsymbol{\xi}^*\right)=\left[t\left(\tilde{P}^{(m)},-\delta\right)-t\left(\tilde{P}^{(m)},\xi^{m*}\right)\right]\left[\lambda_1(1-Se)-\lambda_2(1-Sp)\right]\geq 0,$$

since $\lambda_1(1-Se)-\lambda_2(1-Sp)<0$, $\xi^{m*}>-\delta$, and $t(\tilde{p},\xi)$ is increasing in ξ by assumption.

Case II: $n_i>1$

Similarly, by using the expressions in Section 3.2, the contribution of group i to the objective function in (9) is given by

$$\begin{aligned}Q_i\left(\boldsymbol{\xi}^*\right)=&\left[\lambda_2(1-Sp)Se-\lambda_1\left(1-Se^2\right)\right]\sum_{l\in\Omega_i}\left(1-t\left(\tilde{P}^{(l)},\xi^{l*}\right)\right)\\ &-n_i(Se+Sp-1)\left[1-\lambda_1-\lambda_2+\lambda_2(1-Sp)\right]\prod_{l\in\Omega_i}\left(1-t\left(\tilde{P}^{(l)},\xi^{l*}\right)\right)\\ &+(1-\lambda_1-\lambda_2)(1+n_iSe)+\lambda_1\left(1-Se^2\right)n_i,\\ Q_i\left(\tilde{\boldsymbol{\xi}}\right)=&\left[\lambda_2(1-Sp)Se-\lambda_1\left(1-Se^2\right)\right]\sum_{l\in\Omega_i}\left(1-t\left(\tilde{P}^{(l)},\tilde{\xi}^{l}\right)\right)\\ &-n_i(Se+Sp-1)\left[1-\lambda_1-\lambda_2+\lambda_2(1-Sp)\right]\prod_{l\in\Omega_i}\left(1-t\left(\tilde{P}^{(l)},\tilde{\xi}^{l}\right)\right)\\ &+(1-\lambda_1-\lambda_2)(1+n_iSe)+\lambda_1\left(1-Se^2\right)n_i\\ \Rightarrow Q_i\left(\tilde{\boldsymbol{\xi}}\right)-Q_i\left(\boldsymbol{\xi}^*\right)=&\,h\left(\boldsymbol{\xi}^*,\tilde{\boldsymbol{\xi}}\right)\left[t\left(\tilde{P}^{(m)},\xi^{m*}\right)-t\left(\tilde{P}^{(m)},\tilde{\xi}^m\right)\right],\end{aligned}$$

where

$$h\left(\boldsymbol{\xi}^*, \tilde{\boldsymbol{\xi}}\right) = \lambda_2(1-Sp)Se - \lambda_1(1-Se^2) - n_i(Se+Sp-1)\left[1-\lambda_1-\lambda_2+\lambda_2(1-Sp)\right]\prod_{\substack{l\in\Omega_i\\ l\neq m}}\left(1-t\left(\tilde{P}^{(l)}, \tilde{\xi}^l\right)\right).$$

Note that $h(\boldsymbol{\xi}^*, \tilde{\boldsymbol{\xi}})$ is independent of both ξ^{m*} and $\tilde{\xi}^m$.

Subcase I: $h(\boldsymbol{\xi}^*, \tilde{\boldsymbol{\xi}}) \leq 0$. Let $\varepsilon = \delta - \xi^{m*} > 0 \Rightarrow \tilde{\xi}^m = \delta$, and hence we get that

$$Q_i\left(\tilde{\boldsymbol{\xi}}\right) - Q_i\left(\boldsymbol{\xi}^*\right) = h\left(\boldsymbol{\xi}^*, \tilde{\boldsymbol{\xi}}\right)\left[t\left(\tilde{P}^{(m)}, \xi^{m*}\right) - t\left(\tilde{P}^{(m)}, \delta\right)\right] \geq 0,$$

since $h(\boldsymbol{\xi}^*, \tilde{\boldsymbol{\xi}}) \leq 0$, $\xi^{m*} < \delta$, and $t(\tilde{p}, \xi)$ is increasing in ξ by assumption.

Subcase II: $h(\boldsymbol{\xi}^*, \tilde{\boldsymbol{\xi}}) > 0$. Let $\varepsilon = -\delta - \xi^{m*} < 0 \Rightarrow \tilde{\xi}^m = -\delta$, and hence, we get that

$$Q_i\left(\tilde{\boldsymbol{\xi}}\right) - Q_i\left(\boldsymbol{\xi}^*\right) = h\left(\boldsymbol{\xi}^*, \tilde{\boldsymbol{\xi}}\right)\left[t\left(\tilde{P}^{(m)}, \xi^{m*}\right) - t\left(\tilde{P}^{(m)}, -\delta\right)\right] \geq 0,$$

since $h(\boldsymbol{\xi}^*, \tilde{\boldsymbol{\xi}}) > 0$, $\xi^{m*} > -\delta$, and $t(\tilde{p}, \xi)$ is increasing in ξ by assumption.

Hence, in all possible cases, the objective function has been maintained or improved, thus concluding the proof. □

Proof of Theorem 2. The proof follows similarly to that of Theorem 1. However, note that if $\lambda_1(1-Se) \geq \lambda_2(1-Sp)$, then, when $n_i = 1$, subcase I is satisfied and the optimal solution is attained at δ. On the other hand, if $n_i > 1$, we have that

$$\begin{aligned}\lambda_2(1-Sp)Se - \lambda_1(1-Se^2) &= \lambda_2(1-Sp)Se - \lambda_1(1-Se)(1+Se)\\ &\leq \lambda_2(1-Sp)Se - \lambda_1(1-Se)Se\\ &= \left[\lambda_2(1-Sp) - \lambda_1(1-Se)\right]Se \leq 0,\end{aligned}$$

and since

$$n_i(Se+Sp-1)\left[1-\lambda_1-\lambda_2+\lambda_2(1-Sp)\right]\prod_{\substack{l\in\Omega_i\\ l\neq m}}\left(1-t\left(P^{(l)}, \tilde{\xi}^l\right)\right) \geq 0,$$

we get that $h(\boldsymbol{\xi}^*, \tilde{\boldsymbol{\xi}}) \leq 0$. Hence, subcase I is satisfied, and the optimal solution is attained at δ, concluding the proof. □

Proof of Theorem 2. We prove the result by showing that, for any risk vector realization, any unordered testing scheme can be converted into an ordered testing scheme while reducing or maintaining the values of all three performance measures in the objective function. We only prove the result for problem **EM**, as the proof for problem **RM** follows similarly, with the only difference being that the entire risk vector is multiplied by $1+\delta$. Toward this end, consider an estimated risk vector realization, $\tilde{\boldsymbol{p}}$, and suppose, to the contrary, that the optimal testing scheme, $\boldsymbol{\Omega}^* = \{\Omega_1^*, \ldots, \Omega_g^*\}$, for some $g = 2, \ldots, N$,[5] is not an ordered testing scheme. Then, there must exist two groups, Ω_i^* and Ω_j^*, $i, j = 1, \ldots, g : i \neq j$, such that

i. $\min_{m\in\Omega_i^*} \tilde{p}^m < \max_{m\in\Omega_j^*} \tilde{p}^m$, and
ii. $\max_{m\in\Omega_i^*} \tilde{p}^m > \min_{m\in\Omega_j^*} \tilde{p}^m$.

Assume, without loss of generality, that $n_i \leq n_j$.

Case I: $n_i = 1$

Since $n_i = 1$, it must be true that $n_j > 1$.[6] Due to conditions (i) and (ii) the single subject in group Ω_i^*, denoted with index k_i, has a lower risk than the subject with the maximum risk in group Ω_j^*, denoted with index k_j (i.e., $\tilde{p}^{k_i} < \tilde{p}^{k_j}$). Let $\Psi_i = \{k_i\}$, let $\Psi_j = \{k_j\}$, and define a new testing scheme, $\hat{\boldsymbol{\Omega}} = \{\hat{\Omega}_1, \ldots, \hat{\Omega}_g\}$, where subjects in Ψ_i are interchanged with subjects in Ψ_j, that is, $\hat{\Omega}_i = (\Omega_i^* \setminus \Psi_i) \cup \Psi_j$, $\hat{\Omega}_j = (\Omega_j^* \setminus \Psi_j) \cup \Psi_i$, and $\hat{\Omega}_l = \Omega_l^*$ for all $l = 1, \ldots, g : l \neq i, j$. As such, we have that

$$\prod_{m\in\Omega_j^*}(1-\tilde{p}^m) < \prod_{m\in\hat{\Omega}_j}(1-\tilde{p}^m) \Rightarrow n_j \prod_{m\in\Omega_j^*}(1-\tilde{p}^m) < n_j \prod_{m\in\hat{\Omega}_j}(1-\tilde{p}^m).$$

In what follows, we will show that $\hat{\boldsymbol{\Omega}}$ reduces or maintains the value of all performance measures.

a. Expected number of false negatives. We have that

$$\begin{aligned}\mathbf{E}\left[FN(\boldsymbol{\Omega}^*)\right] &= \sum_{l:l\neq i,j}\mathbf{E}[FN_l] + (1-Se)\tilde{p}^{k_i} + \left(1-Se^2\right)\sum_{m\in\Omega_j^*}\tilde{p}^m, \text{ and}\\ \mathbf{E}\left[FN(\hat{\boldsymbol{\Omega}})\right] &= \sum_{l:l\neq i,j}\mathbf{E}[FN_l] + (1-Se)\tilde{p}^{k_j} + \left(1-Se^2\right)\sum_{m\in\hat{\Omega}_j}\tilde{p}^m\\ \Rightarrow \mathbf{E}\left[FN(\boldsymbol{\Omega}^*)\right] - \mathbf{E}\left[FN(\hat{\boldsymbol{\Omega}})\right] &= -(1-Se)\left(\tilde{p}^{k_j} - \tilde{p}^{k_i}\right) + \left(1-Se^2\right)\left(\tilde{p}^{k_j} - \tilde{p}^{k_i}\right)\\ &= Se(1-Se)\left(\tilde{p}^{k_j} - \tilde{p}^{k_i}\right) \geq 0.\end{aligned}$$

As such, $\mathbf{E}[FN(\hat{\boldsymbol{\Omega}})] \leq \mathbf{E}[FN(\boldsymbol{\Omega}^*)]$.

b. Expected number of false positives. We have that

$$\mathbf{E}\left[FP(\boldsymbol{\Omega}^*)\right] = \sum_{l:l\neq i,j} \mathbf{E}[FP_l] + (1-Sp)\left(1-\tilde{p}^{k_i}\right) + (1-Sp)Se \sum_{m\in\Omega_j^*}(1-\tilde{p}^m) - n_j(1-Sp)(Se+Sp-1)\prod_{m\in\Omega_j^*}(1-\tilde{p}^m), \text{ and}$$

$$\mathbf{E}\left[FP(\hat{\boldsymbol{\Omega}})\right] = \sum_{l:l\neq i,j} \mathbf{E}[FP_l] + (1-Sp)\left(1-\tilde{p}^{k_j}\right) + (1-Sp)Se \sum_{m\in\hat{\Omega}_j}(1-\tilde{p}^m) - n_j(1-Sp)(Se+Sp-1)\prod_{m\in\hat{\Omega}_j}(1-\tilde{p}^m)$$

$$\Rightarrow \mathbf{E}\left[FP(\boldsymbol{\Omega}^*)\right] - \mathbf{E}\left[FP(\hat{\boldsymbol{\Omega}})\right] = (1-Sp)(1-Se)\left(\tilde{p}^{k_j}-\tilde{p}^{k_i}\right) + n_j(1-Sp)(Se+Sp-1)\left[\prod_{m\in\hat{\Omega}_j}(1-\tilde{p}^m) - \prod_{m\in\Omega_j^*}(1-\tilde{p}^m)\right] > 0.$$

As such, $\mathbf{E}[FP(\hat{\boldsymbol{\Omega}})] \leq \mathbf{E}[FP(\boldsymbol{\Omega}^*)]$.

c. Expected number of tests. We have that

$$\mathbf{E}\left[T(\boldsymbol{\Omega}^*)\right] = \sum_{l:l\neq i,j} \mathbf{E}[T_l] + 2 + n_j\left(Se - (Se+Sp-1)\prod_{m\in\Omega_j^*}(1-\tilde{p}^m)\right), \text{ and}$$

$$\mathbf{E}\left[T(\hat{\boldsymbol{\Omega}})\right] = \sum_{l:l\neq i,j} \mathbf{E}[T_l] + 2 + n_j\left(Se - (Se+Sp-1)\prod_{m\in\hat{\Omega}_j}(1-\tilde{p}^m)\right)$$

$$\Rightarrow \mathbf{E}\left[T(\boldsymbol{\Omega}^*)\right] - \mathbf{E}\left[T(\hat{\boldsymbol{\Omega}})\right] = n_j(Se+Sp-1)\left[\prod_{m\in\hat{\Omega}_j}(1-\tilde{p}^m) - \prod_{m\in\Omega_j^*}(1-\tilde{p}^m)\right] > 0.$$

As such, $\mathbf{E}[T(\hat{\boldsymbol{\Omega}})] \leq \mathbf{E}[T(\boldsymbol{\Omega}^*)]$.

Thus, by converting groups i and j into an ordered testing scheme, all measures are either maintained or reduced, implying that there exists an optimal partition, which is ordered.

Case II: $n_i > 1$

Note that, when the two group sizes are greater than one, the expected number of false negatives resulting from these groups is constant. As such, one can convert any unordered testing scheme into an ordered one without impacting the expected number of false negatives. Thus, we proceed by showing that the remaining performance measures (i.e., $\mathbf{E}[FP]$ and $\mathbf{E}[T]$) are reduced or maintained. By conditions (i) and (ii), there exist $\emptyset \subset \Psi_i \subseteq \Omega_i^*$ and $\emptyset \subset \Psi_j \subseteq \Omega_j^*$ such that $|\Psi_i| = |\Psi_j|$ and when Ψ_i and Ψ_j are interchanged the resulting set of groups will follow an ordered testing scheme in which the group with the smaller size contains the lowest risk subjects, while the group with the larger size contains the highest risk subjects. We have that

$$\prod_{m\in\Psi_j}(1-\tilde{p}^m) - \prod_{m\in\Psi_i}(1-\tilde{p}^m) > 0. \tag{21}$$

Subcase I: $n_i \prod_{m\in\Omega_i\setminus\Psi_i}(1-\tilde{p}^m) > n_j \prod_{m\in\Omega_j\setminus\Psi_j}(1-\tilde{p}^m)$.

Define a new testing scheme, $\hat{\boldsymbol{\Omega}} = \{\hat{\Omega}_1, \ldots, \hat{\Omega}_g\}$, where subjects in Ψ_i are interchanged with subjects in Ψ_j, that is, $\hat{\Omega}_i = (\Omega_i^* \setminus \Psi_i) \cup \Psi_j$, $\hat{\Omega}_j = (\Omega_j^* \setminus \Psi_j) \cup \Psi_i$, and $\hat{\Omega}_l = \Omega_l^*$ for all $l = 1, \ldots, g : l \neq i, j$. In what follows, we will show that partition $\hat{\boldsymbol{\Omega}}$ reduces or maintains the value of all performance measures. Multiplying the condition imposed in the subcase, that is,

$$n_i \prod_{m\in\Omega_i\setminus\Psi_i}(1-\tilde{p}^m) > n_j \prod_{m\in\Omega_j\setminus\Psi_j}(1-\tilde{p}^m),$$

by Equation (21), expanding, and rearranging gives

$$n_i \prod_{m\in\hat{\Omega}_i}(1-\tilde{p}^m) + n_j \prod_{m\in\hat{\Omega}_j}(1-\tilde{p}^m) > n_i \prod_{m\in\Omega_i^*}(1-\tilde{p}^m) + n_j \prod_{m\in\Omega_j^*}(1-\tilde{p}^m).$$

a. Expected number of false positives ($\mathbf{E}[FP]$). We have that

$$\mathbf{E}\left[FP(\boldsymbol{\Omega}^*)\right] = \sum_{l:l\neq i,j} \mathbf{E}[FP_l] + (1-Sp)Se \sum_{m\in\Omega_i^*}(1-\tilde{p}^m) - n_i(1-Sp)(Se+Sp-1)\prod_{m\in\Omega_i^*}(1-\tilde{p}^m)$$
$$+ (1-Sp)Se \sum_{m\in\Omega_j^*}(1-\tilde{p}^m) - n_j(1-Sp)(Se+Sp-1)\prod_{m\in\Omega_j^*}(1-\tilde{p}^m), \text{ and}$$
$$\mathbf{E}\left[FP(\hat{\boldsymbol{\Omega}})\right] = \sum_{l:l\neq i,j} \mathbf{E}[FP_l] + (1-Sp)Se \sum_{m\in\hat{\Omega}_i}(1-\tilde{p}^m) - n_i(1-Sp)(Se+Sp-1)\prod_{m\in\hat{\Omega}_i}(1-\tilde{p}^m)$$
$$+ (1-Sp)Se \sum_{m\in\hat{\Omega}_j}(1-\tilde{p}^m) - n_j(1-Sp)(Se+Sp-1)\prod_{m\in\hat{\Omega}_j}(1-\tilde{p}^m).$$

Noting that

$$\sum_{m\in\Omega_i^*\cup\Omega_j^*}(1-\tilde{p}^m) = \sum_{m\in\hat{\Omega}_i\cup\hat{\Omega}_j}(1-\tilde{p}^m),$$

and subtracting the two gives

$$\frac{\mathbf{E}\left[FP(\boldsymbol{\Omega}^*)\right] - \mathbf{E}\left[FP(\hat{\boldsymbol{\Omega}})\right]}{(1-Sp)(Se+Sp-1)} = n_i\prod_{m\in\hat{\Omega}_i}(1-\tilde{p}^m) + n_j\prod_{m\in\hat{\Omega}_j}(1-\tilde{p}^m) - n_i\prod_{m\in\Omega_i^*}(1-\tilde{p}^m) - n_j\prod_{m\in\Omega_j^*}(1-\tilde{p}^m) > 0.$$

As such, $\mathbf{E}[FP(\hat{\boldsymbol{\Omega}})] \le \mathbf{E}[FP(\boldsymbol{\Omega}^*)]$.

b. Expected number of tests ($\mathbf{E}[T]$). We have that

$$\mathbf{E}\left[T(\boldsymbol{\Omega}^*)\right] = \sum_{l:l\neq i,j}\mathbf{E}[T_l] + 2 + n_i\left(Se - (Se+Sp-1)\prod_{m\in\Omega_i^*}(1-\tilde{p}^m)\right)$$
$$+ n_j\left(Se - (Se+Sp-1)\prod_{m\in\Omega_j^*}(1-\tilde{p}^m)\right), \text{ and}$$
$$\mathbf{E}\left[T(\hat{\boldsymbol{\Omega}})\right] = \sum_{l:l\neq i,j}\mathbf{E}[T_l] + 2 + n_i\left(Se - (Se+Sp-1)\prod_{m\in\hat{\Omega}_i}(1-\tilde{p}^m)\right)$$
$$+ n_j\left(Se - (Se+Sp-1)\prod_{m\in\hat{\Omega}_j}(1-\tilde{p}^m)\right).$$

Subtracting the two gives

$$\frac{\mathbf{E}\left[T(\boldsymbol{\Omega}^*)\right] - \mathbf{E}\left[T(\hat{\boldsymbol{\Omega}})\right]}{(Se+Sp-1)} = n_i\prod_{m\in\hat{\Omega}_i}(1-\tilde{p}^m) + n_j\prod_{m\in\hat{\Omega}_j}(1-\tilde{p}^m) - n_i\prod_{m\in\Omega_i^*}(1-\tilde{p}^m) - n_j\prod_{m\in\Omega_j^*}(1-\tilde{p}^m) > 0.$$

As such, $\mathbf{E}[T(\hat{\boldsymbol{\Omega}})] \le \mathbf{E}[T(\boldsymbol{\Omega}^*)]$.

Thus, by converting groups i and j into an ordered testing scheme, all measures are either maintained or reduced, implying that there exists an optimal partition, which is ordered.

Subcase II: $n_i\prod_{m\in\Omega_i\setminus\Psi_i}(1-\tilde{p}^m) \le n_j\prod_{m\in\Omega_j\setminus\Psi_j}(1-\tilde{p}^m)$.

Due to conditions (i) and (ii), there exist $\emptyset \subset Z_i \subseteq \Omega_i$ and $\emptyset \subset Z_j \subseteq \Omega_j$ such that $|Z_i| = |Z_j|$, and when Z_i and Z_j are interchanged, the resulting set of groups will follow an ordered testing scheme in which the group with the smaller size contains the highest risk subjects, whereas the group with the larger size contains the lowest risk subjects. Define a new testing scheme, $\tilde{\boldsymbol{\Omega}} = \{\tilde{\Omega}_1, \ldots, \tilde{\Omega}_g\}$, where subjects in Z_i are interchanged with subjects in Z_j, that is, $\tilde{\Omega}_i = (\Omega_i^* \setminus Z_i) \cup Z_j$, $\tilde{\Omega}_j = (\Omega_j^* \setminus Z_j) \cup Z_i$, and $\tilde{\Omega}_l = \Omega_l^*$ for all $l = 1, \ldots, g : l \neq i, j$. In what follows, we will show that partition $\tilde{\boldsymbol{\Omega}}$ reduces or maintains the value of all performance measures. By the condition imposed in the subcase, that is,

$$n_i\prod_{m\in\Omega_i\setminus\Psi_i}(1-\tilde{p}^m) \le n_j\prod_{m\in\Omega_j\setminus\Psi_j}(1-\tilde{p}^m),$$

and Equation (21) we get

$$n_i\prod_{m\in\Omega_i}(1-\tilde{p}^m) \le n_j\prod_{m\in\Omega_j}(1-\tilde{p}^m). \tag{22}$$

By the definitions of Z_i and Z_j, we have that

$$\prod_{m\in Z_i}(1-\tilde{p}^m)-\prod_{m\in Z_j}(1-\tilde{p}^m)>0. \tag{23}$$

From Equation (22) we have that

$$n_i\prod_{m\in\Omega_i\setminus Z_i}(1-\tilde{p}^m)\prod_{m\in Z_i}(1-\tilde{p}^m)\le n_j\prod_{m\in\Omega_j\setminus Z_j}(1-\tilde{p}^m)\prod_{m\in Z_j}(1-\tilde{p}^m). \tag{24}$$

Then, by Equations (23) and (24), it must be true that

$$n_i\prod_{m\in\Omega_i\setminus Z_i}(1-\tilde{p}^m)<n_j\prod_{m\in\Omega_j\setminus Z_j}(1-\tilde{p}^m). \tag{25}$$

Multiplying Equation (25) by Equation (23), expanding, and rearranging gives

$$n_i\prod_{m\in\tilde{\Omega}_i}(1-\tilde{p}^m)+n_j\prod_{m\in\tilde{\Omega}_j}(1-\tilde{p}^m)>n_i\prod_{m\in\Omega_i}(1-\tilde{p}^m)+n_j\prod_{m\in\Omega_j}(1-\tilde{p}^m).$$

Following a similar methodology to that of subcase I, one can show that $\mathbf{E}[FP(\tilde{\mathbf{\Omega}})]\le\mathbf{E}[FP(\mathbf{\Omega}^*)]$ and $\mathbf{E}[T(\tilde{\mathbf{\Omega}})]\le\mathbf{E}[T(\mathbf{\Omega}^*)]$. As such, for all cases, we are always able to construct an ordered testing scheme that reduces or maintains the values of all performance measures, concluding the proof. □

Proof of Lemma 2. We have that

$$\begin{aligned}\mathbf{E}\left[\prod_{m=i}^{j}g(X^{(m)})\right]&=\int_a^b\int_a^{x^j}\mathbf{E}\left[g(X^{(i)})\cdots g\left(X^{(j)}\right)|X^{(i)}=x^i,X^{(j)}=x^j\right]f_{X^{(i)},X^{(j)}}(x^i,x^j)dx^idx^j\\&=\int_a^b\int_a^{x^j}g(x^i)g(x^j)\mathbf{E}\left[g(X^{(i+1)})\cdots g\left(X^{(j-1)}\right)|X^{(i)}=x^i,X^{(j)}=x^j\right]f_{X^{(i)},X^{(j)}}(x^i,x^j)dx^idx^j,\end{aligned} \tag{26}$$

where

$$\begin{aligned}&\mathbf{E}\left[g(X^{(i+1)})\cdots g\left(X^{(j-1)}\right)|X^{(i)}=x^i,X^{(j)}=x^j\right]\\&=\int_{x^i}^{x^j}\int_{x^{i+1}}^{x^j}\cdots\int_{x^{j-2}}^{x^j}g(x^{i+1})\cdots g(x^{j-1})f_{X^{(i+1)},\ldots,X^{(j-1)}|X^{(i)}=x^i,X^{(j)}=x^j}\left(x^{i+1},\ldots,x^{j-1}\right)dx^{j-1}\cdots dx^{i+2}dx^{i+1},\end{aligned} \tag{27}$$

where

$$f_{X^{(i+1)},\ldots,X^{(j-1)}|X^{(i)}=x^i,X^{(j)}=x^j}\left(x^{i+1},\ldots,x^{j-1}\right)=\frac{f_{X^{(i)},\ldots,X^{(j)}}(x^i,\ldots,x^j)}{f_{X^{(i)},X^{(j)}}(x^i,x^j)}.$$

From Park (1995), we have that

$$f_{X^{(i)},\ldots,X^{(j)}}(x^i,\ldots,x^j)=\frac{N!}{(i-1)!(N-j)!}F_X(x^i)^{i-1}f_X(x^i)\cdots f_X(x^j)(1-F_X(x^j))^{N-j},$$

and

$$f_{X^{(i)},X^{(j)}}(x^i,x^j)=\frac{N!}{(i-1)!(j-i-1)!(N-j)!}f_X(x^i)f_X(x^j)F_X(x^i)^{i-1}\left(F_X(x^j)-F_X(x^i)\right)^{j-i-1}\left(1-F_X(x^j)\right)^{N-j}.$$

As such, we have that

$$f_{X^{(i+1)},\ldots,X^{(j-1)}|X^{(i)}=x^i,X^{(j)}=x^j}\left(x^{i+1},\ldots,x^{j-1}\right)=(j-i-1)!\frac{f_X(x^{i+1})\cdots f_X(x^{j-1})}{(F_X(x^j)-F_X(x^i))^{j-i-1}}. \tag{28}$$

Substituting Equation (28) into Equation (27) gives

$$\begin{aligned}&\mathbf{E}\left[g(X^{(i+1)})\cdots g\left(X^{(j-1)}\right)|X^{(i)}=x^i,X^{(j)}=x^j\right]\\&=\frac{(j-i-1)!}{(F_X(x^j)-F_X(x^i))^{j-i-1}}\int_{x^i}^{x^j}\int_{x^{i+1}}^{x^j}\cdots\int_{x^{j-2}}^{x^j}g(x^{i+1})\cdots g(x^{j-1})f_X(x^{i+1})\cdots f_X(x^{j-1})dx^{j-1}\cdots dx^{i+2}dx^{i+1}.\end{aligned}$$

Define $h(t)$ by

$$h(t) \equiv \int_t^{x^j} g(x) f_X(x) dx.$$

Note that $h(t)$ exists since $g(x)$ and $f_X(x)$ are both continuous (imposed in the theorem), and $h(x^j) = 0$ and $dh(t) = -g(t)f_X(t)dt$. Then Equation (27) can be written as

$$\begin{aligned}
&\mathbf{E}\left[g(X^{(i+1)}) \cdots g(X^{(j-1)}) | X^{(i)} = x^i, X^{(j)} = x^j\right] \\
&= \frac{(j-i-1)!}{(F_X(x^j) - F_X(x^i))^{j-i-1}} \int_{x^i}^{x^j} \cdots \int_{x^{j-3}}^{x^j} \left[\int_{x^{j-2}}^{x^j} g\left(x^{j-1}\right) f_X\left(x^{j-1}\right) dx^{j-1} \right] g\left(x^{i+1}\right) \cdots g\left(x^{j-2}\right) f_X\left(x^{i+1}\right) \cdots f_X\left(x^{j-2}\right) dx^{j-2} \cdots dx^{i+1} \\
&\quad \frac{(j-i-1)!}{(F_X(x^j) - F_X(x^i))^{j-i-1}} \int_{x^i}^{x^j} \cdots \int_{x^{j-3}}^{x^j} h\left(x^{j-2}\right) g\left(x^{i+1}\right) \cdots g\left(x^{j-2}\right) f_X\left(x^{i+1}\right) \cdots f_X\left(x^{j-2}\right) dx^{j-2} \cdots dx^{i+1} \\
&\quad \frac{(j-i-1)!}{(F_X(x^j) - F_X(x^i))^{j-i-1}} \int_{x^i}^{x^j} \cdots \int_{x^{j-4}}^{x^j} \left[\int_{x^{j-3}}^{x^j} g\left(x^{j-2}\right) f_X\left(x^{j-2}\right) h\left(x^{j-2}\right) dx^{j-2} \right] g\left(x^{i+1}\right) \cdots g\left(x^{j-3}\right) f_X\left(x^{i+1}\right) \cdots f_X\left(x^{j-3}\right) dx^{j-3} \cdots dx^{i+1}.
\end{aligned}$$

For the integral in brackets, we perform a change of variable $u = h(x^{j-2})$ with $du = -g(x^{j-2})f_X(x^{j-2})dx^{j-2}$. This gives

$$\begin{aligned}
&\mathbf{E}\left[g(X^{(i+1)}) \cdots g(X^{(j-1)}) | X^{(i)} = x^i, X^{(j)} = x^j\right] \\
&= \frac{(j-i-1)!}{(F_X(x^j) - F_X(x^i))^{j-i-1}} \int_{x^i}^{x^j} \cdots \int_{x^{j-5}}^{x^i} \left[\int_{x^{j-4}}^{x^j} g\left(x^{j-3}\right) f_X\left(x^{j-3}\right) \frac{h\left(x^{j-3}\right)^2}{2} dx^{j-3} \right] g\left(x^{i+1}\right) \cdots g\left(x^{j-4}\right) f_X\left(x^{i+1}\right) \cdots f_X\left(x^{j-4}\right) dx^{j-4} \cdots dx^{i+1}.
\end{aligned}$$

Continuing in this manner gives

$$\begin{aligned}
\mathbf{E}\left[g(X^{(i+1)}) \cdots g(X^{(j-1)}) | X^{(i)} = x^i, X^{(j)} = x^j\right] &= \frac{(j-i-1)!}{(F_X(x^j) - F_X(x^i))^{j-i-1}} \frac{h(x^i)^{j-i-1}}{(j-i-1)!} \\
&= \left[\int_{x^i}^{x^j} \frac{g(x) f_X(x) dx}{F_X(x^j) - F_X(x^i)} \right]^{j-i-1}.
\end{aligned}$$

Substituting the latter into Equation (26) provides the result. □

Proof of Lemma 3. The result follows by Remark 2, which states that, for a given vector $\boldsymbol{y}$, problem **CM** reduces to an **SP** problem, for which the constraint set possesses the total unimodularity property. As such, the optimal solution, corresponding to the specific $\boldsymbol{y}$, will be integral, and hence, integrality constraints are not required for the $\boldsymbol{x}$ variables. □

Proof of Theorem 3. The result directly follows from Theorem 2 and Remark 2, which state that the partitioning problem of **CM** reduces to a constrained-shortest path problem. □

Appendix B. Derivation of Performance Measures

All the subsequent derivations primarily rely on two assumptions: (i) subjects are assumed to be independent of one another, that is, knowledge of the true status of one subject does not impact the risk of another, and (ii) the test efficacy values (Se and Sp) are independent of the group size. Throughout, let D^m denote the indicator random variable corresponding to the true positive status of subject $m \in S$, and let $N_i^+(\Omega_i)$ denote the number of true positive subjects in group i that is comprised of subjects belonging in Ω_i, that is, $N_i^+(\Omega_i) = \sum_{m \in \Omega_i} D^m$.

B.1. False Negative Classifications

Conditioned on the estimated risk vector, $\tilde{\boldsymbol{P}}$, and the perturbation vector, $\boldsymbol{\Xi}$, we can write, for a given subject $m \in \{1, 2, \ldots, N\}$ and $\boldsymbol{\Omega}$,

$$\begin{aligned}
\mathbf{E}\left[FN^m | \boldsymbol{\Xi}, \tilde{\boldsymbol{P}}\right] &= \mathbf{E}\left[FN^m | D^m = 1, \boldsymbol{\Xi}, \tilde{\boldsymbol{P}}\right] P\left(D^m = 1 | \boldsymbol{\Xi}, \tilde{\boldsymbol{P}}\right) + \mathbf{E}\left[FN^m | D^m = 0, \boldsymbol{\Xi}, \tilde{\boldsymbol{P}}\right] P\left(D^m = 0 | \boldsymbol{\Xi}, \tilde{\boldsymbol{P}}\right) \\
&= \begin{cases} (1 - Se) t\left(\tilde{P}^{(m)}, \Xi^m\right) + 0, & \text{if } m \text{ is individually tested,} \\ \left(Se(1 - Se) + (1 - Se)\right) t\left(\tilde{P}^{(m)}, \Xi^m\right) + 0, & \text{otherwise,} \end{cases}
\end{aligned}$$

leading to:

$$\mathbf{E}\left[FN^m | \boldsymbol{\Xi}, \tilde{\boldsymbol{P}}\right] = \begin{cases} (1 - Se) t\left(\tilde{P}^{(m)}, \Xi^m\right), & \text{if } m \text{ is individually tested,} \\ \left(1 - Se^2\right) t\left(\tilde{P}^{(m)}, \Xi^m\right), & \text{otherwise.} \end{cases}$$

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

$$\mathbf{E}\left[FN_i(\Omega_i)|\Xi,\tilde{\boldsymbol{P}}\right]=\begin{cases}(1-Se)\sum_{m\in\Omega_i}t\left(\tilde{P}^{(m)},\Xi^m\right), & \text{if } n_i=1,\\ (1-Se^2)\sum_{m\in\Omega_i}t\left(\tilde{P}^{(m)},\Xi^m\right), & \text{otherwise,}\end{cases}$$

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

$$\mathbf{E}\left[FN(\boldsymbol{\Omega})|\Xi,\tilde{\boldsymbol{P}}\right]=\sum_{i=1}^{g}\mathbf{E}\left[FN_i(\Omega_i)|\Xi,\tilde{\boldsymbol{P}}\right].$$

B.2. False Positive Classifications

Conditioned on the estimated risk vector, $\tilde{\boldsymbol{P}}$, and the perturbation vector, Ξ, we can write, for a given $\boldsymbol{\Omega}$ and a subject m that is individually tested,

$$\begin{aligned}\mathbf{E}\left[FP^m|\Xi,\tilde{\boldsymbol{P}}\right]&=\mathbf{E}\left[FP^m|D^m=1,\Xi,\tilde{\boldsymbol{P}}\right]P\left(D^m=1|\Xi,\tilde{\boldsymbol{P}}\right)+\mathbf{E}\left[FP^m|D^m=0,\Xi,\tilde{\boldsymbol{P}}\right]P\left(D^m=0,\Xi,\tilde{\boldsymbol{P}}\right)\\&=0+\left(1-Sp\right)\left(1-t\left(\tilde{P}^{(m)},\Xi^m\right)\right),\end{aligned}$$

and for any subject $m\in\Omega^G$ grouped in some set Ω_i:$n_i>1$, $i\in\{1,\ldots,g\}$, that is, $m\in\Omega_i$, we have

$$\begin{aligned}\mathbf{E}\left[FP^m|\Xi,\tilde{\boldsymbol{P}}\right]&=\mathbf{E}\left[FP^m|D^m=1,\Xi,\tilde{\boldsymbol{P}}\right]P\left(D^m=1|\Xi,\tilde{\boldsymbol{P}}\right)+\mathbf{E}\left[FP^m|D^m=0,\Xi,\tilde{\boldsymbol{P}}\right]P\left(D^m=0|\Xi,\tilde{\boldsymbol{P}}\right)\\&=0+\left[\left(1-Sp\right)^2\prod_{k\in\Omega_i\setminus\{m\}}\left(1-t\left(\tilde{P}^{(k)},\Xi^k\right)\right)+Se\left(1-Sp\right)\left(1-\prod_{k\in\Omega_i\setminus\{m\}}\left(1-t\left(\tilde{P}^{(k)},\Xi^k\right)\right)\right)\right]\left(1-t\left(\tilde{P}^{(m)},\Xi^m\right)\right)\\&=\left(1-Sp\right)\left[Se-\left(Se+Sp-1\right)\prod_{k\in\Omega_i\setminus\{m\}}\left(1-t\left(\tilde{P}^{(k)},\Xi^k\right)\right)\right]\left(1-t\left(\tilde{P}^{(m)},\Xi^m\right)\right)\\&=\left(1-Sp\right)Se\left(1-t\left(\tilde{P}^{(m)},\Xi^m\right)\right)-\left(1-Sp\right)\left(Se+Sp-1\right)\prod_{k\in\Omega_i}\left(1-t\left(\tilde{P}^{(k)},\Xi^k\right)\right),\end{aligned}$$

leading to

$$\mathbf{E}\left[FP^m|\Xi,\tilde{\boldsymbol{P}}\right]=\begin{cases}\left(1-Sp\right)\left(1-t\left(\tilde{P}^{(m)},\Xi^m\right)\right), & \text{if } m \text{ is individually tested,}\\ \left(1-Sp\right)Se\left(1-t\left(\tilde{P}^{(m)},\Xi^m\right)\right)-\left(1-Sp\right)\left(Se+Sp-1\right)\prod_{k\in\Omega_i}\left(1-t\left(\tilde{P}^{(k)},\Xi^k\right)\right), & \text{otherwise.}\end{cases}$$

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

$$\mathbf{E}\left[FP_i(\Omega_i)|\Xi,\tilde{\boldsymbol{P}}\right]=\begin{cases}\left(1-Sp\right)\sum_{m\in\Omega_i}\left(1-t\left(\tilde{P}^{(m)},\Xi^m\right)\right), & \text{if } n_i=1,\\ \left(1-Sp\right)Se\sum_{m\in\Omega_i}\left(1-t\left(\tilde{P}^{(m)},\Xi^m\right)\right)-n_i\left(1-Sp\right)\left(Se+Sp-1\right)\prod_{m\in\Omega_i}\left(1-t\left(\tilde{P}^{(m)},\Xi^m\right)\right), & \text{otherwise,}\end{cases}$$

and the expected number of false positive classifications for all subjects in set S is given by $\mathbf{E}[FP(\boldsymbol{\Omega})|\Xi,\tilde{\boldsymbol{P}}]=\sum_{i=1}^{g}\mathbf{E}[FP_i(\Omega_i)|\Xi,\tilde{\boldsymbol{P}}]$.

B.3. Number of Tests

Given a partition $\boldsymbol{\Omega}$, the expected number of tests for group i, $i=\{1,\ldots,g\}$, is 1 if $n_i=1$ (i.e., individual testing). In contrast, if $n_i>1$, then, conditioned on the estimated risk vector, $\tilde{\boldsymbol{P}}$, and the perturbation vector, Ξ, we can write

$$\begin{aligned}\mathbf{E}\left[T_i(\Omega_i)|\Xi,\tilde{\boldsymbol{P}}\right]&=\sum_{k=0}^{n_i}\mathbf{E}\left[T_i(\Omega_i)|N_i^+(\Omega_i)=k,\Xi,\tilde{\boldsymbol{P}}\right]P\left(N_i^+(\Omega_i)=k|\Xi,\tilde{\boldsymbol{P}}\right)\\&=\mathbf{E}\left[T_i(\Omega_i)|N_i^+(\Omega_i)=0,\Xi,\tilde{\boldsymbol{P}}\right]P\left(N_i^+(\Omega_i)=0|\Xi,\tilde{\boldsymbol{P}}\right)+\sum_{k=1}^{n_i}\mathbf{E}\left[T_i(\Omega_i)|N_i^+(\Omega_i)=k,\Xi,\tilde{\boldsymbol{P}}\right]P\left(N_i^+(\Omega_i)=k|\Xi,\tilde{\boldsymbol{P}}\right)\\&=\left(Sp+\left(1-Sp\right)(1+n_i)\right)P\left(N_i^+(\Omega_i)=0|\Xi,\tilde{\boldsymbol{P}}\right)+\sum_{k=1}^{n_i}(1-Se+Se(1+n_i))P\left(N_i^+(\Omega_i)=k|\Xi,\tilde{\boldsymbol{P}}\right)\\&=1+n_i\left(Se-\left(Se+Sp-1\right)\prod_{m\in\Omega_i}\left(1-t\left(\tilde{P}^{(m)},\Xi^m\right)\right)\right).\end{aligned}$$

$$\text{Thus, }\mathbf{E}\left[T_i(\Omega_i)|\Xi,\tilde{\boldsymbol{P}}\right]=\begin{cases}1, & \text{if } n_i=1,\\ 1+n_i\left(Se-\left(Se+Sp-1\right)\prod_{m\in\Omega_i}\left(1-t\left(\tilde{P}^{(m)},\Xi^m\right)\right)\right), & \text{otherwise,}\end{cases}\tag{29}$$

and the expected number of tests needed for all subjects in set S is given by $\mathbf{E}[T(\boldsymbol{\Omega})|\Xi,\tilde{\boldsymbol{P}}]=\sum_{i=1}^{g}\mathbf{E}[T_i(\Omega_i)|\Xi,\tilde{\boldsymbol{P}}]$.

Appendix C. Case Study: Fitting Parameters of the Estimated Risk Distribution

As discussed in Section 5, the CDC data set reports the number of chlamydia cases diagnosed and the size of the corresponding population for the year 2014 for each combination of gender, age group, and race/ethnicity group. (Two gender categories, seven age group categories, and five race/ethnicity categories are considered in the data set, leading to 70 categories (see Centers for Disease Control and Prevention Division of STD Prevention 2014).) Based on studies, we use an underreporting factor of 3 for chlamydia (Grimes et al. 2013). Figure 4 depicts the histogram of the estimated risk obtained from this data set.

Let $\hat{\mu}_p$, $\hat{\sigma}_p$, and $\hat{CV}$, respectively, denote the mean, standard deviation, and coefficient of variation of the estimated risk obtained from the data set. Our objective is to estimate the parameters of the mixture distribution, w, β_1, and β_2 such that $w \in [0,1]$, $\beta_1 > 0$, and $\beta_2 > 0$ (see Eq. (20)), by matching the first two moments of the distribution to those of the data set so as to minimize the Kolmogorov-Smirnov statistic (Massey 1951), that is, to minimize the maximum distance between the empirical and the fitted cumulative distribution functions. Toward this end, we first solve the following system of equations for a given w, that is, derive expressions for parameters β_1 and β_2 as a function of w:

$$\frac{w}{\beta_1} + \frac{1-w}{\beta_2} = \hat{\mu}_p, \tag{30}$$

$$\frac{w}{\beta_1^2} + \frac{1-w}{\beta_2^2} = \frac{\hat{\mu}_p^2 + \hat{\sigma}_p^2}{2}, \tag{31}$$

where $\beta_1, \beta_2 > 0$. From Equation (30), we have that

$$\frac{1}{\beta_2} = \frac{\beta_1 \hat{\mu}_p - w}{\beta_1(1-w)},$$

which, when substituted into Equation (31), leads to the following quadratic equation:

$$\left(\frac{w}{1-w}\right)\left(\frac{1}{\beta_1}\right)^2 - \left(\frac{2\hat{\mu}_p w}{1-w}\right)\left(\frac{1}{\beta_1}\right) + \left(\frac{\hat{\mu}_p^2}{1-w} - \frac{\hat{\mu}_p^2 + \hat{\sigma}_p^2}{2}\right) = 0. \tag{32}$$

There are three cases.

Case I. $\hat{CV} < 1$

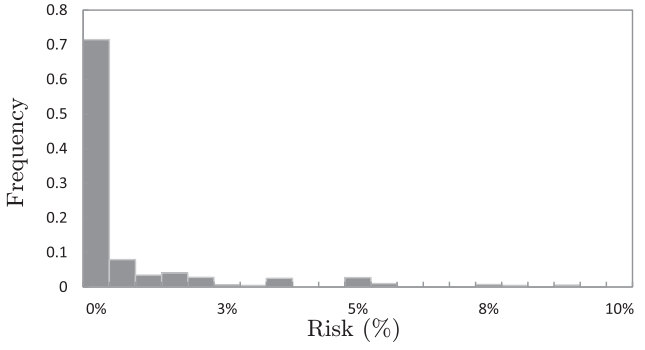
In this case, Equation (32) has no real root. This implies that a mixture distribution, comprised of two exponential distributions, is not a good representation of a data set having a coefficient of variation less than one, as one cannot match both the mean and the standard deviation to the data set.

Case II. $1 \le \hat{CV} < \sqrt{3}$

In this case, the system of equations given in Equations (30)–(31) has two solutions for all w such that

$$0 \le 1 - \frac{2}{\hat{CV}^2 + 1} < w < \frac{2}{\hat{CV}^2 + 1} \le 1. \tag{33}$$

Figure 4. Histogram of the Estimated Risk Based on the CDC Data Set in Centers for Disease Control and Prevention Division of STD Prevention (2014)



If w does not satisfy Equation (33), then β_1 or β_2 will not be positive. For all w that satisfy Equation (33), the two solutions are given by

$$\frac{1}{\beta_1} = \hat{\mu}_p\left(1 + \sqrt{\frac{(\hat{CV}^2 - 1)(1-w)}{2w}}\right), \quad \text{or} \quad \frac{1}{\beta_2} = \hat{\mu}_p\left(1 - \sqrt{\frac{(\hat{CV}^2 - 1)w}{2(1-w)}}\right),$$
$$\frac{1}{\beta_1} = \hat{\mu}_p\left(1 - \sqrt{\frac{(\hat{CV}^2 - 1)(1-w)}{2w}}\right), \qquad \frac{1}{\beta_2} = \hat{\mu}_p\left(1 + \sqrt{\frac{(\hat{CV}^2 - 1)w}{2(1-w)}}\right).$$

Case III. $\hat{CV} \geq \sqrt{3}$

In this case, the system of equations given in Equations (30)–(31) has a unique solution for all w such that

$$w < \frac{2}{\hat{CV}^2 + 1} \quad \text{or} \quad w > 1 - \frac{2}{\hat{CV}^2 + 1}. \tag{34}$$

Note that due to the condition imposed in case II, that is, $\hat{CV} \geq \sqrt{3}$, w can satisfy at most one of the two inequalities in Equation (34). If w does not satisfy any of the inequalities in Equation (34), then β_1 or β_2 will not be positive. For all w that satisfy Equation (34), the unique solution is given by

$$\frac{1}{\beta_1} = \begin{cases} \hat{\mu}_p\left(1 + \sqrt{\frac{(\hat{CV}^2-1)(1-w)}{2w}}\right), & \text{if} \quad w < \frac{2}{\hat{CV}^2+1}, \\ \hat{\mu}_p\left(1 - \sqrt{\frac{(\hat{CV}^2-1)(1-w)}{2w}}\right), & \text{if} \quad w > 1 - \frac{2}{\hat{CV}^2+1}, \end{cases} \quad \frac{1}{\beta_2} = \begin{cases} \hat{\mu}_p\left(1 - \sqrt{\frac{(\hat{CV}^2-1)w}{2(1-w)}}\right), & \text{if} \quad w < \frac{2}{\hat{CV}^2+1}, \\ \hat{\mu}_p\left(1 + \sqrt{\frac{(\hat{CV}^2-1)w}{2(1-w)}}\right), & \text{if} \quad w > 1 - \frac{2}{\hat{CV}^2+1}. \end{cases}$$

Due to symmetry, it is sufficient to study only one of these cases, as the other case can be obtained by interchanging w and $1 - w$ and interchanging β_1 and β_2. As such, by limiting the case to

$$w < \frac{2}{\hat{CV}^2 + 1},$$

the unique solution reduces to

$$\frac{1}{\beta_1} = \hat{\mu}_p\left(1 + \sqrt{\frac{(\hat{CV}^2 - 1)(1-w)}{2w}}\right), \tag{35}$$

$$\frac{1}{\beta_2} = \hat{\mu}_p\left(1 - \sqrt{\frac{(\hat{CV}^2 - 1)w}{2(1-w)}}\right). \tag{36}$$

The data set used in the case study in Section 5 has a mean of $\hat{\mu}_p = 0.0097$ and a standard deviation of $\hat{\sigma}_p = 0.0248$, leading to a coefficient of variation of $\hat{CV} \approx 2.55 \geq \sqrt{3}$. As such, by the above analysis, for all $w \in [0, 0.2659)$ β_1 and β_2 are obtained by Equations (35) and (36), respectively.

Figure 5. Kolmogorov-Smirnov Test Statistic as a Function of w

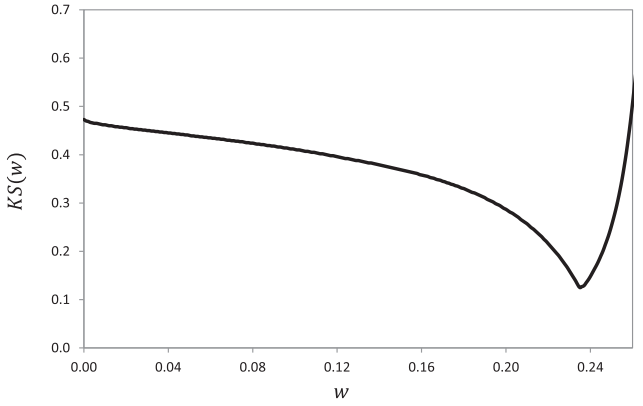
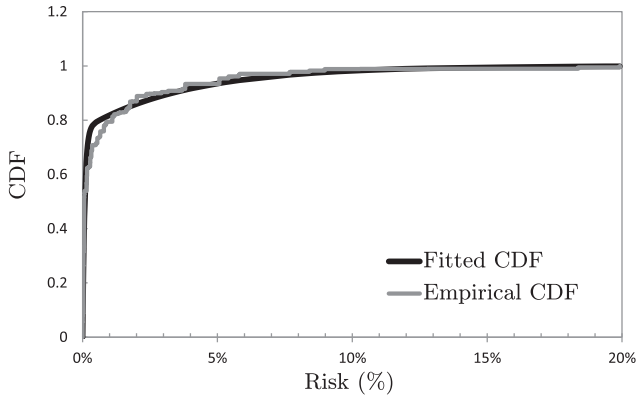


Figure 6. Empirical and Fitted Cumulative Distribution Functions (CDF), When $w = 0.235$, $\beta_1 = 25.708$, and $\beta_2 = 1,291.832$



As mentioned above, w is chosen so as to minimize the Kolmogorov-Smirnov test statistic (KS; Massey 1951), that is, the maximum distance between the empirical and fitted cumulative distribution functions. That is, w is the optimal solution to the following optimization problem:

$$\begin{aligned} &\underset{w}{\text{minimize}} && KS(w) \\ &\text{subject to} && w \in \left[0, \frac{2}{C\hat{V}^2 + 1}\right). \end{aligned}$$

Figure 5 plots $KS(w)$ as a function of w for the data set used in the case study and shows that the function $KS(w)$ is unimodal in w, with a global minimizer, $w = 0.235$, leading to $KS(0.235) = 0.125$. Then, from Equations (35) and (36), we obtain $\beta_1 = 25.708$ and $\beta_2 = 1{,}291.832$. Next, we study the goodness of fit of this mixture distribution with these parameter values through statistical hypothesis testing, with the following null and alternative hypotheses:

H_0: The data follow a mixture distribution comprised of two exponential distributions with parameters $w = 0.235$, $\beta_1 = 25.708$, and $\beta_2 = 1{,}291.832$.

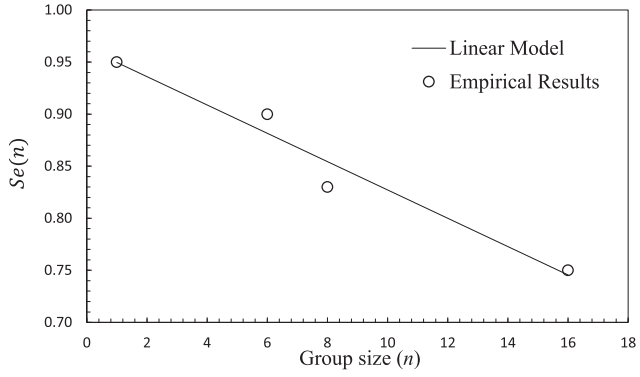
H_a: The data do not follow a mixture distribution comprised of two exponential distributions with parameters $w = 0.235$, $\beta_1 = 25.708$ and $\beta_2 = 1{,}291.832$.

Since we have 70 data points, the critical value for a significance level of $\alpha = 0.05$ equals 0.163 (Massey 1951). Since $KS(0.235) = 0.125 < 0.163$, we conclude that there is insufficient statistical evidence to reject the null hypothesis.

Figure 6 plots both the empirical and fitted cumulative distribution functions when $w = 0.235$, $\beta_1 = 25.708$, and $\beta_2 = 1{,}291.832$, further indicating that the fitted mixture distribution, comprised of two exponential distributions, provides a good fit for the data set.

Appendix D. The Effect of Dilution

Figure 7. Fitted Linear Sensitivity Function vs. Empirical Results (Stramer et al. 2013) as a Function of Group Size



Endnotes

[1] The independence assumption is common in the field of statistical analysis with measurement errors (Grace 2017). The main idea is that such measurement errors are due to exogenous factors that do not depend on the risk value. For example, within our context, the error terms may arise due to biased data sets or omission or misrepresentation of certain risk factors in risk estimation.

[2] The outcome of any test that does not satisfy this assumption can be transformed in a way that satisfies this assumption; this can be done by interpreting a positive (negative) test outcome as a negative (positive) test outcome.

[3] The cost of a single false negative is estimated based on quality-adjusted life-years, so as to assess the economic value of medical treatment (see Owusu-Edusei et al. 2015 for details).

[4] The DNA-based assay considered in this case study is highly accurate and can thus be considered as a gold standard test. However, owing to its high cost, individually testing all subjects using this gold standard test is not budget-feasible, which is why group testing was implemented in the first place.

[5] If $g = 1$, then all subjects are in one group, and hence, it is an ordered testing scheme.

[6] If both groups are of size 1, then they will follow an ordered testing scheme.

References

American Red Cross (2016) Details of tests performed for different infectious agents. Accessed July 3, 2016, http://www.redcrossblood.org/learn-about-blood/what-happens-donated-blood/blood-testing.

Aprahamian H, Bish DR, Bish EK (2019) Optimal risk-based group testing. *Management Sci.* 65(9):4365–4384.

Arora D, Arora B, Khetarpal A (2010) Seroprevalence of HIV, HBV, HCV and syphilis in blood donors in Southern Haryana. *Indian J. Pathology Microbiology* 53(2):308–309.

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, Brown DB, Caramanis C (2011) Theory and applications of robust optimization. *SIAM Rev.* 53(3):464–501.

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

Centers for Disease Control and Prevention (2000) Tracking the hidden epidemics, trends in STDs in the United States. Accessed October 27, 2015, https://www.cdc.gov/std/trends2000/trends2000.pdf.

Centers for Disease Control and Prevention Division of STD Prevention (2014) Sexually transmitted disease surveillance. Accessed November 13, 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, Leiserson CE, Rivest RL, Stein C (2009) *Introduction to Algorithms* (MIT Press, Cambridge, MA).

Cormode G, Muthukrishnan S (2006) Combinatorial algorithms for compressed sensing. *International Colloquium on Structural Information and Communication Complexity* (Springer, New York), 280–294.

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

El-Amine H, Bish EK, Bish DR (2018) Robust postdonation blood screening under prevalence rate uncertainty. *Oper. Res.*, 66(1):1–17.

Family Planning Council of Iowa (2020) Iowa community-based screening services procedures manual. Accessed March 8, 2020, http://www.shl.uiowa.edu/dcd/iippmanual.pdf.

Gallego G, Moon I (1993) The distribution free newsboy problem: Review and extensions. *J. Oper. Res. Soc.* 44(8):825–834.

Garcia R (2009) Resource constrained shortest paths and extensions. Unpublished doctoral dissertation, Georgia Institute of Technology, Atlanta.

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

Grassly NC, Morgan M, Walker N, Garnett G, Stanecki KA, Stover J, Brown T, Ghys PD (2004) Uncertainty in estimates of HIV/AIDS: the estimation and application of plausibility bounds. *Sexually Transmitted Infections* 80(suppl 1):i31–i38.

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

Herwaldt BL, Linden JV, Bosserman E, Young C, Olkowska D, Wilson M (2011) Transfusion-associated babesiosis in the United States: A description of cases. *Ann. Internal Medicine* 155(8):509–519.

HOLOGIC (2017) Tigris DTS system. Accessed August 6, 2017, http://www.hologic.com/products/clinical-diagnostics/instrument-systems/tigris-dts-system.

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

Johnson NL, Kotz S, Wu XZ (1991) *Inspection Errors for Attributes in Quality Control* (CRC Press, Boca Raton, FL).

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 Microbiology* 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.

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.

Massey FJ (1951) The Kolmogorov-Smirnov test for goodness of fit. *J. Amer. Statist. Assoc.* 46(253):68–78.

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 11, 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.

Park S (1995) The entropy of consecutive order statistics. *IEEE Trans. Inform. Theory* 41(6):2003–2007.

Perakis G, Roels G (2008) Regret in the newsvendor model with partial information. *Oper. Res.* 56(1):188–203.

Roche Diagnostics (2017) Cobas® system. Accessed August 12, 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.

Scarf H, Arrow K, Karlin S (1958) A min-max solution of an inventory problem. *Stud. Math. Theory Inventory Production* 10(2):201–209.

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

Sterrett A (1957) On the detection of defective members of large populations. *Ann. Math. Statist.* 28(4):1033–1036.

Stramer SL, Krysztof DE, Brodsky JP, Fickett TA, Reynolds B, Dodd RY, Kleinman SH (2013) Comparative analysis of triplex nucleic acid test assays in United States blood donors. *Transfusion* 53(10, Part 2):2525–2537.

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. Preventive Services Task Force (2020) Final recommendation statement chlamydia and gonorrhea: Screening. Accessed February 22, 2020, https://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/chlamydia-and-gonorrhea-screening.

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 with the Aptima AC2 and Viper CTQ/GCQ assays for detection of Chlamydia trachomatis and Neisseria gonorrhoeae. *J. Clinical Microbiology* 50(7):2244–2249.

Yi GY (2017) *Statistical Analysis with Measurement Error or Misclassification: Strategy, Method and Application* (Springer, New York).

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 Medical Rev.* 26(2):119–128.