Static Risk-based Group Testing Schemes under Imperfectly Observable Risk

Hrayer Aprahamian[1*], Ebru K. Bish[2†], Douglas R. Bish[2‡]

[1]Department of Industrial and Systems Engineering, Texas A&M University
College Station, TX 77843, United States
[2]Grado Department of Industrial and Systems Engineering, Virginia Tech
Blacksburg, VA 24061-0118, United States

November 2018; Revised: November 2019; March 2020

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.

Keywords: Group testing; Dorfman testing; risk-based testing; robust optimization; combinatorial optimization; constrained shortest path; order statistics

*Corresponding Author: *hrayer@tamu.edu*
†*ebru@vt.edu*
‡*drb1@vt.edu*

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 multi-access communication networks, software testing, and compressed sensing, e.g., [3, 5, 9, 11, 37]. The origins of group testing date back to the 1940's, when Dorfman [11], 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 set-up 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; and 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 the human immunodeficiency virus (HIV), and hepatitis viruses [10]; 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 [29]. In both cases, specimens are loaded into automated testing machines in *batches* (typically between 40-200 specimens per batch depending on the equipment used [22]). 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 (US) are around seven times more likely to have an HIV infection than repeat donors [43]; various risk factors exist for sexually-transmitted diseases, including chylamydia [6]; vector-borne infections, such as babesiosis and Zika, are endemic in certain areas of the US [21]. Thus, subjects come from a *heterogeneous* population, with each sub-population having a potentially different risk. However, the process of estimating the subject-specific risk, informed by sub-population-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 sub-populations are inherently uncertain, and risk factors are often not well-understood, e.g., [2, 19]. 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, i.e., 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 and that satisfies certain limitations on group sizes; and that 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; and 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, i.e., there are no classification errors, a *homogeneous* population, i.e., the risk of the binary characteristic is identical across subjects, and infinite testing *batch* sizes (e.g., [11, 37, 38]). While several papers extend the analysis of Dorfman testing schemes to imperfect tests (e.g., [18, 24, 26, 29]), there is very limited work on Dorfman testing for a heterogeneous population, i.e., with subject-specific risk, and the few papers that consider a heterogeneous population (e.g., [1, 23, 29]) mainly do so under restrictive assumptions, including that subject risk is perfectly observable, or they determine testing schemes heuristically. In particular, Hwang [23] 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, i.e., 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. [29] extend the analysis in [23] 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 develop heuristics. More recently, focusing on dynamic testing schemes, Aprahamian et al. [1] 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.

While 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. [1] and Hwang [23] 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. While 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. [29] 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 non-decreasing 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 non-increasing for a risk-ordered batch, see [1]). 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 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. [1], to consider static testing schemes under uncertainty, and allows us to quantify the "price of implementability". Some key properties established in [1] for the deterministic model extend to the stochastic setting. Consequently, we take advantage of a reformulation technique presented in [1] in which the set partitioning problem is cast as a Constrained Shortest Path Problem. However, unlike the model in [1], 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. [1] 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 US. 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 two percent of the more complicated dynamic risk-based testing schemes, i.e., schemes that are customized for each batch [1, 23]. 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 lab 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 US. 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 upper-case letters to denote random variables, lower-case 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 or not the binary characteristic is present), or to the binary test outcome (i.e., to denote whether or not 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, e.g., the nucleic acid amplification testing machine [22, 41] 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 3-4 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, independently of other subjects. However, the value of P^m, i.e., 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, i.e., 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 $\tilde{P}^m, m = 1, \cdots, N$, are independent and identically distributed (iid), following an arbitrary continuous distribution with support in $[a, b]$, for $0 \leq a < b \leq 1$; random variables $\Xi^m, m = 1, \cdots, N$, are iid, following an arbitrary continuous distribution with support in $[-\delta, \delta]$, for some $\delta \geq 0$; and random vectors $\tilde{\boldsymbol{P}}$ and $\boldsymbol{\Xi}$ are independent[1]. Our modeling of subject risk, $\tilde{P}^m, m = 1, \cdots, N$, 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

[1]The independence assumption is common in the field of statistical analysis with measurement errors [17]. 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.

as, $P^m|\tilde{P}^m, \Xi^m = t\big(\tilde{P}^m, \Xi^m\big)$, for $m = 1, \cdots, N$, 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 non-decreasing in each of p and ξ; $\mathbf{E}_{\Xi^m}\big[t\big(\tilde{p}^m, \Xi^m\big)\big] = \tilde{p}^m$, that is, the expectation of the true risk equals the estimated risk; and $t\big(p, 0\big) = p$, for all p, that is, the true risk reduces to the estimated risk when the perturbation term is zero, i.e., 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, $\big[t\big(\tilde{P}^m, -\delta\big), t\big(\tilde{P}^m, \delta\big)\big] \subseteq [0, 1]$. Notice 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*, $S \equiv \{1, \cdots, N\}$, that follows a non-decreasing order of the estimated subject risk, i.e., $\tilde{p}^1 \leq \tilde{p}^2 \leq \cdots \leq \tilde{p}^N$. Then $\tilde{\boldsymbol{P}} = (\tilde{P}^{(1)}, \tilde{P}^{(2)}, \cdots, \tilde{P}^{(N)})$ denotes the random *ordered* estimated risk vector per batch, with $\tilde{P}^{(m)}$ denoting the m^{th} 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). While 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, $m = 1, \cdots, N$, 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*, i.e., a true-positive subject falsely classified as negative, with probability $Pr\big(D^m(1 - R^m(n)) = 1\big)$, and a *false positive classification*, i.e., a true-negative subject falsely classified as positive, with probability $Pr\big((1 - D^m)R^m(n) = 1\big)$. 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. While 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], i.e., $Sp \geq (1 - Se) \Rightarrow Se + Sp - 1 \in [0, 1]$.

[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.

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 set up 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, i.e., 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, $\boldsymbol{\Omega} = (\Omega_i)_{i=1,\cdots,g}$, for some $g \in \mathbf{Z}^+$, such that $\bigcup_{i=1}^{g} \Omega_i = S$, and $\Omega_i \cap \Omega_j = \emptyset$, for all $i, j = 1, \cdots, g$: $i \neq j$. Letting $n_i \equiv |\Omega_i|$ denote the cardinality (size) of set $\Omega_i, i = 1, \cdots, g$, we refer to the corresponding vector, $\boldsymbol{n} = (n_i)_{i=1,\cdots,g} : \sum_i n_i = N$, as the *group size* vector. Thus, 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 m^{th} 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, $j = 1, \cdots, N$, equal 1 if at least one group of size j is utilized, and 0 otherwise, that is, for a given $\boldsymbol{\Omega}$, and $\forall j = 1, \cdots, N$, $y_j = 1$ only if there exists at least one group $i, i = 1, \cdots, g$, 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, $\mathbf{\Omega}$, follow:

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

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

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

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

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: While 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 $\mathbf{\Xi}$) 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, $\mathbf{\Xi}$, 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):

$$\underset{\boldsymbol{\Omega}=(\Omega_i)_{i=1,\cdots,g},\, g\in\mathbf{Z}^+}{\text{minimize}} \quad \mathbf{E}_{\tilde{\boldsymbol{P}}}\Big[\mathbf{E}_{\boldsymbol{\Xi}}\Big[\mathbf{E}[Q(\boldsymbol{\Omega})|\boldsymbol{\Xi},\tilde{\boldsymbol{P}}]\Big]\Big]$$

$$\text{subject to} \quad \Omega_i \cap \Omega_j = \emptyset, \quad \forall i,j = 1,\cdots,g\text{: } i \neq j \tag{4}$$

$$\bigcup_{i=1}^{g} \Omega_i = \{1,\cdots,N\} \tag{5}$$

$$||\boldsymbol{\Omega}|| \leq \gamma, \tag{6}$$

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

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

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 [7]. Hence, for realistic problem sizes in which the number of subjects in each batch, N, is typically in the order of hundreds, the problem quickly becomes computationally expensive. Second, the objective function is non-linear and non-separable, 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., [4,15,34,36]. 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):

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

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

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

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

Remark 1. When $\delta=0$, i.e., 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.

While a worst-case objective function, similar to the one used in Problem **RM**, is a conservative measure, e.g., [4, 12, 34], 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 [1] to the case where the true risk vector is stochastic and not perfectly observable. We refer the interested reader to [1] 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 $\boldsymbol{\Omega}$:

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

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

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, $\tilde{\boldsymbol{P}}$, the perturbation vector, $\boldsymbol{\Xi}$, we have, for a given $\boldsymbol{\Omega}$:

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

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

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, $\tilde{\boldsymbol{P}}$, the perturbation vector, $\boldsymbol{\Xi}$, we have, for a given $\boldsymbol{\Omega}$:

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

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

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 Eq.s (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 Eq.s (11) and (25) substantially complicates the analysis, as the contribution of a subject to the objective function depends on the set of subjects it is grouped with.

Notice that when $\boldsymbol{\lambda} = (1, 0)$ the objective function in Eq. (3) reduces to minimizing the expected number of false negative classifications only, while 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 Eq. (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., [11, 23, 29, 35]). Thus, **EM** and **RM** formulations expand the deterministic formulation in Aprahamian et al. [1], 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* $m = 1, \cdots, N$*.*

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* $m = 1, \cdots, N$*.*

The condition imposed in the second part of Theorem 1 is realistic, as the weight (cost) of a false negative in the objective function, i.e., λ_1, is typically much larger than the weight (cost) of a false positive, i.e., 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:

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

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 reduces the two-stage robust formulation in (8)-(9) to a single-stage optimization problem. Moreover, notice 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)

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

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:

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

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 $\mathbf{E}\big[FN(\boldsymbol{\Omega})|\boldsymbol{\Xi}=\boldsymbol{z},\tilde{\boldsymbol{P}}\big]$, $\mathbf{E}\big[FP(\boldsymbol{\Omega})|\boldsymbol{\Xi}=\boldsymbol{z},\tilde{\boldsymbol{P}}\big]$, and $\mathbf{E}\big[T(\boldsymbol{\Omega})|\boldsymbol{\Xi}=\boldsymbol{z},\tilde{\boldsymbol{P}}\big]$ are given by Eq.s (10), (11), and (25), respectively, and $f_{\tilde{P}^{(1)},\cdots,\tilde{P}^{(N)}}(.)$ denotes the joint probability density function of the ordered random variables $\tilde{P}^{(1)},\tilde{P}^{(2)},\cdots,\tilde{P}^{(N)}$.

As stated earlier, Problem **CM** is challenging due to two main reasons: First, it is at least as hard as the partitioning problem, which is NP-hard [7]; and 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**. Towards this end, consider the following definition.

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

This definition allows the representation of an ordered testing scheme $\mathbf{\Omega} = \left(\Omega_i\right)_i$ 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, 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. [1] 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: While **C-SP** is NP-hard [16], 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 contains products of some function of a set of order statistics (see Eq.s (11) and (25)). 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 $[a, b]$:$0 \leq a < b \leq 1$. Let $X^{(1)} \leq X^{(2)} \leq \cdots \leq X^{(N)}$ denote the order statistics, and let $f_{X^{(i)},\cdots,X^{(j)}}(\cdot)$ and $f_{X^{(i)},X^{(j)}}(\cdot)$ respectively represent the joint probability density functions of the ordered random variables, $X^{(i)} \leq \cdots \leq X^{(j)}$, and of $X^{(i)} \leq X^{(j)}$, $i < j$. Let $g(\cdot)$ denote any continuous function. Then, for all $N \geq 4$ and*

$i, j = 1 \cdots, N : i < j$, *we have:*

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

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

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

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

$$
\begin{aligned}
\underset{\substack{\boldsymbol{y}=(y_j)_{j=1,\cdots,N},\\ \boldsymbol{x}=(x_{ij})_{(i,j)\in E}}}{\text{minimize}} \quad & \sum_{(i,j)\in E} \mathbf{E}_{\tilde{\boldsymbol{P}}}\Big[\mathbf{E}\big[Q_i(S_{i-j})|\boldsymbol{\Xi}=\boldsymbol{z},\tilde{\boldsymbol{P}}\big]\Big]x_{ij} \\
\textit{subject to} \quad & \sum_{j\in V:j>i} x_{ij} - \sum_{j\in V:j<i} x_{ji} = \begin{cases} 1, & \textit{if } i=1 \\ -1, & \textit{if } i=N+1 \\ 0, & \textit{otherwise} \end{cases}, \qquad \forall i\in V \\
& \sum_{j\in V:j>i} x_{ij} \le 1, \qquad \forall i\in V \\
& x_{kl} \le y_{l-k}, \quad \forall (k,l)\in E \qquad (15.1) \\
& \sum_{j=1}^{N} y_j \le \gamma \qquad (15.2) \\
& y_j \le \sum_{(k,l)\in E:\ l-k=j} x_{kl}, \quad \forall j=1,\cdots,N \qquad (15.3) \\
& y_j \in \{0,1\}, \quad \forall j=1,\cdots,N \qquad (15.4) \\
& x_{ij}\in\{0,1\}, \quad \forall (i,j)\in E,
\end{aligned}
\tag{15}
$$

where $Q_i(\cdot)$ function is as defined in Eq. (2); $S_{i-j}=\{i,\cdots,j-1\}$, for all $(i,j)\in E$; y_j, $j=1,\cdots,N$, 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, i.e., the group, comprised of subjects $\{i,\cdots,j-1\}$, 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 $V=\{1,\cdots,N+1\}$ 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 $(i,j) \in E$, i.e., $\mathbf{E}[Q_i(S_{i-j})]$, 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 (15.1)-(15.4), 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 $\mathcal{O}(|V|+|E|) = \mathcal{O}(N^2)$ [8]. While 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 the **SP**, no longer holds with the addition of Constraints (15.1)-(15.4). 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 US [6], and existing screening practices differ substantially across states. For instance, certain states screen only a subset of the their population, e.g., many states follow the United States 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 [14]. 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 [31] utilizes individual testing for the USPSTF recommended population, while Iowa and Idaho utilize groups of size four [27, 30, 40]. 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 [30]. While 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, e.g., gender, age, and race/ethnicity.

Our objectives in this case study are multi-fold: **(1)** To quantify the benefits of risk-based testing, i.e., **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., [11, 18, 26]): Such uniform testing schemes assume that the population is homogeneous, and 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., [27]), 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, i.e., 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**), i.e., 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 [1]. 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, i.e., the price of implementability, and to gain insight into the conditions under which static schemes are expected to perform relatively well compared to their dynamic counter-parts, 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 [6]. The data set provides the number of chlamydia cases reported, along with the size of the corresponding sub-population 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 [13]; and the number of actual cases is estimated to be roughly three-fold the number of cases reported [20]. This underscores the importance of factoring in under-reporting, which we do by multiplying the number of reported cases for each sup-population 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 under-reporting rate for all sub-populations; this assumption is made due to a lack of under-reporting data for the different sub-populations.

Based on the histogram of the risk for the CDC data set (see Figure 4 in 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 [28], i.e., 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, i.e., $P|\tilde{P}, \Xi = t(\tilde{P}, \Xi) = \tilde{P}(1+\Xi)$, 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 [25], with a reported sensitivity of $Se = 0.95$ and a specificity of $Sp = 0.99$. We utilize the findings of [32, 42] 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] [32]), while the per test screening cost is set

[3] The cost of a single false negative is estimated based on quality-adjusted life-years (QALYs), so as to assess the economic value of medical treatment; see [32] for details.

to \$55 [32]. 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 [27].

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, i.e., 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, i.e., E-OF $\equiv \mathbf{E}_{\tilde{\boldsymbol{P}}}\Big[\mathbf{E}\big[Q|\boldsymbol{\Xi} = \mathbf{0}, \tilde{\boldsymbol{P}}\big]\Big]$, and W-OF denote the worst-case expected cost of a testing scheme, i.e., W-OF $\equiv \mathbf{E}_{\tilde{\boldsymbol{P}}}\Big[\mathbf{E}\big[Q|\boldsymbol{\Xi} = \boldsymbol{\delta}, \tilde{\boldsymbol{P}}\big]\Big]$. Similarly, we let CE-OF denote the expected cost of a testing scheme *conditioned* on a given realization of the estimated risk vector, i.e., CE-OF $\equiv \mathbf{E}\big[Q|\boldsymbol{\Xi} = \mathbf{0}, \tilde{\boldsymbol{P}}\big]$, and CW-OF denote the worst-case cost of a testing scheme *conditioned* on a given realization of the estimated risk vector, i.e., CW-OF $\equiv \mathbf{E}\big[Q|\boldsymbol{\Xi} = \boldsymbol{\delta}, \tilde{\boldsymbol{P}}\big]$.

5.2 Risk-based versus Non Risk-based (Uniform) Schemes

In this section, we compare the performance of **EM** and **RM** to non risk-based (uniform) testing schemes, i.e., solutions to Problem **UM**, for various scenarios, see Table 1. 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, while 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, i.e., the solutions converge to the solution with $\gamma = 5$ (and in some scenarios, this convergence happens faster, i.e., for $\gamma \geq 4$); see Table 1.

The results in Table 1 highlight several important properties. First, both **EM** and **RM** sub-

[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.

Table 1: Performance comparison of **UM**, **EM**, and **RM**

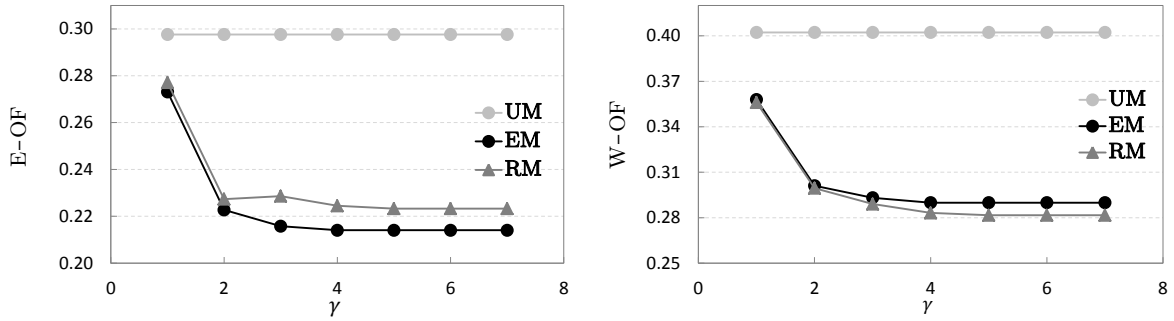
	Problem **UM**		
	$\boldsymbol{n}^*=(11_{(5)}, 5_{(1)})$	E-OF=0.2976	W-OF=0.4023
	Problem **EM**		
γ	$\boldsymbol{n}^*$	E-OF (%Change†)	W-OF (%Change†)
1	$(12_{(5)})$	0.2729 (-8.30%)	0.3574 (-11.14%)
2	$(46_{(1)}, 7_{(2)})$	0.2212 (-25.69%)	0.3093 (-23.11%)
3	$(41_{(1)}, 11_{(1)}, 4_{(2)})$	0.2143 (-27.98%)	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%)

$\dagger$ Percent change over **UM**

stantially 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 $\gamma = 1$, i.e., 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% to 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

Figure 1: Expected cost (E-OF) (left) and worst-case cost (W-OF) (right) for **UM**, **EM**, and **RM**, as a function of γ



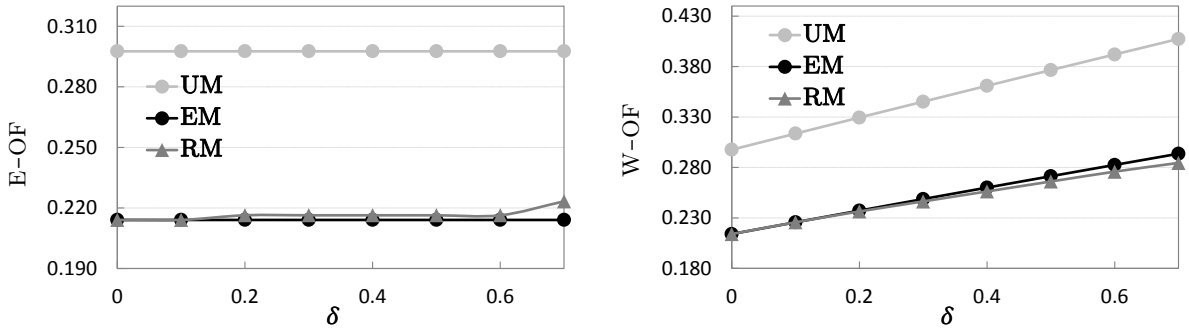
price of robustness is 3.9%, while **EM** leads to a 2.1% increase in the worst-case expected cost over **RM**. Examining the **RM** and **EM** solutions in detail, we observe another advantage of **RM**, in that **RM** has a tendency to reduce the expected number of misclassifications (false positives 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, i.e., 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 $\gamma \geq 5$, that is, 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** turn 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 versus 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 [1], in which the decision-maker *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, i.e., to quantify the price of implementability. Towards this end, we run a Monte Carlo simulation with 10,000 replications. While 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

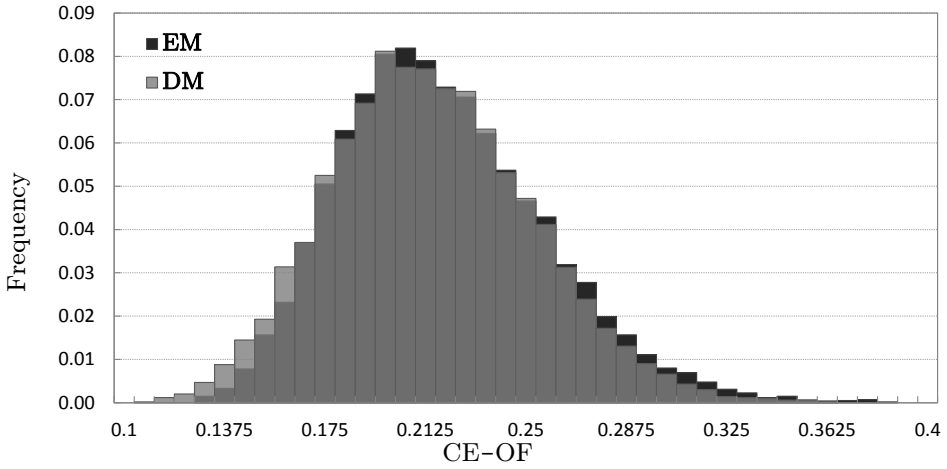
Figure 2: Expected cost (E-OF) (left) and worst-case cost (W-OF) (right) for **UM**, **EM**, and **RM**, as a function of δ for all $\gamma \geq 5$



batch is generated following the distribution in Eq. (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**, while **EM** is significantly easier to implement. This implies that a static risk-based testing scheme captures most benefits of dynamic risk-based testing, while greatly simplifying the implementation, making static testing schemes appealing for practitioners.

While 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 9 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 to 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 $N = 100$; and a similar trend can be observed for moderate and high accuracy tests. However, notice 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 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.

Figure 3: Histogram of the conditional expected cost (CE-OF) for **EM** and **DM** for all $\gamma \geq 5$



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 non-negligible 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., [29]), 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, i.e., $Se(n)$. 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 definition of the dilution effect, $Se(n)$ is non-increasing 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 [39]. For illustrative purposes, we use linear regression to model the

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 to 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%

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

test's sensitivity ($Se(n)$) as a linear function of group size (with a correlation coefficient of 0.98); see Figure 7 in the Appendix. 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. While 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. While 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 on 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 non-negligible 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 multi-stage hierarchical schemes or schemes that take advantage of overlapping groups (e.g., array-based grouping schemes [26]). While 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 (sub-populations) and what risk value should be assigned to each of these sub-populations.

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: Our special thanks go to Professor Shane Henderson for providing highly insightful feedback and excellent suggestions on this paper. We would also like to thank the Review Team for excellent comments and suggestions. Their suggestions have improved both the presentation and the analysis in the paper. We are grateful to 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. This material is based upon work supported in part by the National Science Foundation under Grant No. #1055360. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] H. Aprahamian, D. R. Bish, and E. K. Bish. Optimal risk-based group testing. *Management Science*, 65(9):4365–4384, 2019.

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

[3] T. Berger, N. Mehravari, D. Towsley, and J. Wolf. Random multiple-access communication and group testing. *IEEE Transactions on Communications*, 32(7):769–779, 1984.

[4] D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.

[5] A. Blass and Y. Gurevich. The logic in computer science column pairwise testing. *Bulletin of EATCS*, 78:100–132, 2002.

[6] Centers for Disease Control and Prevention Division of STD Prevention. *Sexually transmitted disease surveillance 2014*, accessed November, 2016. `http://www.cdc.gov/std/stats14/surv-2014-print.pdf`.

[7] A. K. Chakravarty, J. B. Orlin, and U. G. Rothblum. Technical note: A partitioning problem with additive objective with an application to optimal inventory groupings for joint replenishment. *Operations Research*, 30(5):1018–1022, 1982.

[8] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to algorithms*. MIT Press, 2009.

[9] G. Cormode and S. Muthukrishnan. Combinatorial algorithms for compressed sensing. In *International Colloquium on Structural Information and Communication Complexity*, pages 280–294. Springer, 2006.

[10] American Red Cross. *Details of tests performed for different infectious agents*, accessed July, 2016. `http://www.redcrossblood.org/learn-about-blood/what-happens-donated-blood/blood-testing`.

[11] R. Dorfman. The detection of defective members of large populations. *The Annals of Mathematical Statistics*, 14(4):436–440, 1943.

[12] H. El-Amine, E. K. Bish, and D. R. Bish. Robust postdonation blood screening under prevalence rate uncertainty. *Operations Research*, 2017. Published online, `https://doi.org/10.1287/opre.2017.1658`.

[13] Centers for Disease Control and Prevention. *Tracking the hidden epidemics, trends in STDs in the United States, 2000*, accessed October, 2015. `https://www.cdc.gov/std/trends2000/trends2000.pdf`.

[14] United States Preventive Services Task Force. *Final Recommendation Statement Chlamydia and Gonorrhea: Screening*, accessed February, 2020. `https://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/chlamydia-and-gonorrhea-screening`.

[15] G. Gallego and I. Moon. The distribution free newsboy problem: review and extensions. *Journal of the Operational Research Society*, 44:825–834, 1993.

[16] R. Garcia. *Resource constrained shortest paths and extensions*. Unpublished Doctoral Dissertation, Georgia Institute of Technology, School of Industrial and Systems Engineering, 2009.

[17] Y. Grace. Statistical analysis with measurement error or misclassification: Strategy, method and application. 2017.

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

[19] N. C. Grassly, M. Morgan, N. Walker, G. Garnett, K. A. Stanecki, J. Stover, T. Brown, and P. D. Ghys. Uncertainty in estimates of HIV/AIDS: the estimation and application of plausibility bounds. *Sexually Transmitted Infections*, 80(suppl 1):i31–i38, 2004.

[20] J. A. Grimes, L. A. Smith, and K. Fagerberg. *Sexually Transmitted Disease: An Encyclopedia of Diseases, Prevention, Treatment, and Issues [2 volumes]: An Encyclopedia of Diseases, Prevention, Treatment, and Issues*. ABC-CLIO, 2013.

[21] B. L. Herwaldt, J. V. Linden, E. Bosserman, C. Young, D. Olkowska, and M. Wilson. Transfusion-associated babesiosis in the United States: a description of cases. *Annals of Internal Medicine*, 155(8):509–519, 2011.

[22] HOLOGIC®. *Tigris DTS system*, accessed August, 2017. `http://www.hologic.com/products/clinical-diagnostics/instrument-systems/tigris-dts-system`.

[23] F. K. Hwang. A generalized binomial group testing problem. *Journal of the American Statistical Association*, 70(352):923–926, 1975.

[24] N. L. Johnson, S. Kotz, and X. Z. Wu. *Inspection errors for attributes in quality control*, Volume 44. CRC Press, 1991.

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

[26] H. Y. Kim, M. G. Hudgens, J. M. Dreyfuss, D. J. Westreich, and C. D. Pilcher. Comparison of group testing algorithms for case identification in the presence of test error. *Biometrics*, 63(4):1152–1163, 2007.

[27] J. L. Lewis, V. M. Lockary, and S. Kobic. Cost savings and increased efficiency using a stratified specimen pooling strategy for Chlamydia trachomatis and Neisseria gonorrhoeae. *Sexually Transmitted Diseases*, 39(1):46–48, 2012.

[28] F. J. Massey. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

[29] C. S. McMahan, J. M. Tebbs, and C. R. Bilder. Informative Dorfman screening. *Biometrics*, 68(1):287–296, 2012.

[30] Family Planning Council of Iowa. *Iowa Community-Based Screening Services Procedures Manual*, accessed March, 2020. `http://www.shl.uiowa.edu/dcd/iippmanual.pdf`.

[31] North Carolina State Laboratory of Public Health. *Virology/Serology: Chlamydia/Gonorrhea*, accessed November, 2016. `http://slph.ncpublichealth.com/virology-serology/chlamydia/default.asp`.

[32] K. Owusu-Edusei Jr, H. W. Chesson, T. L. Gift, R. C. Brunham, and G. Bolan. Cost-effectiveness of Chlamydia vaccination programs for young women. *Emerging Infectious Diseases*, 21(6):960, 2015.

[33] S. Park. The entropy of consecutive order statistics. *IEEE Transactions on Information Theory*, 41(6):2003–2007, 1995.

[34] G. Perakis and G. Roels. Regret in the newsvendor model with partial information. *Operations Research*, 56(1):188–203, 2008.

[35] B. A. Saraniti. Optimal pooled testing. *Health Care Management Science*, 9(2):143–149, 2006.

[36] H. Scarf, K. Arrow, and S. Karlin. A min-max solution of an inventory problem. *Studies in the Mathematical Theory of Inventory and Production*, 10(2):201, 1958.

[37] M. Sobel and P. A. Groll. Group testing to eliminate efficiently all defectives in a binomial sample. *Bell System Technical Journal*, 38(5):1179–1252, 1959.

[38] A. Sterrett. On the detection of defective members of large populations. *The Annals of Mathematical Statistics*, 28(4):1033–1036, 1957.

[39] S. L. Stramer, D. E. Krysztof, J. P. Brodsky, T. A. Fickett, B. Reynolds, R. Y. Dodd, and S. H. Kleinman. Comparative analysis of triplex nucleic acid test assays in United States blood donors. *Transfusion*, 53:2525–2537, 2013.

[40] J. M. Tebbs, C. S. McMahan, and C. R. Bilder. Two-stage hierarchical group testing for multiple infections with application to the infertility prevention project. *Biometrics*, 69(4):1064–1073, 2013.

[41] Roche Diagnostics USA. *Cobas® system*, accessed August, 2017. `https://usdiagnostics.roche.com/en/molecular/systems-automation/cobas-6800-8800-systems.html#overview`.

[42] B. Van Der Pol, O. Liesenfeld, J. A. Williams, S. N. Taylor, R. A. Lillis, B. A. Body, M. Nye, C. Eisenhut, and E. W. Hook. Performance of the cobas CT/NG test compared to the Aptima AC2 and Viper CTQ/GCQ assays for detection of Chlamydia trachomatis and Neisseria gonorrhoeae. *Journal of Clinical Microbiology*, 50(7):2244–2249, 2012.

[43] S. Zou, S. L. Stramer, and R. Y. Dodd. Donor testing and risk: current prevalence, incidence, and residual risk of transfusion-transmissible agents in US allogeneic donations. *Transfusion Medicine Reviews*, 26(2):119–128, 2012.

A Mathematical Proofs

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

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

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

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

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

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

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

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

Proof of Theorem 1. Part 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 $-\delta$. Towards 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:

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

$$\Rightarrow Q_i(\tilde{\boldsymbol{\xi}}) - Q_i(\boldsymbol{\xi}^*) = \Big[t\big(\tilde{P}^{(m)}, \tilde{\xi}^{m}\big) - t\big(\tilde{P}^{(m)}, \xi^{m*}\big)\Big]\big[\lambda_1(1-Se) - \lambda_2(1-Sp)\big].$$

Sub-case 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:

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

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.

Sub-case 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:

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

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:

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

$$\Rightarrow Q_i(\tilde{\boldsymbol{\xi}}) - Q_i(\boldsymbol{\xi}^*) = h\big(\boldsymbol{\xi}^*, \tilde{\boldsymbol{\xi}}\big)\Big[t\big(\tilde{P}^{(m)}, \xi^{m*}\big) - t\big(\tilde{P}^{(m)}, \tilde{\xi}^{m}\big)\Big],$$

where

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

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

Sub-case I: $h\big(\boldsymbol{\xi}^*, \tilde{\boldsymbol{\xi}}\big) \leq 0$

Let $\varepsilon = \delta - \xi^{m*} > 0 \Rightarrow \tilde{\xi}^m = \delta$, and hence we get that:

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

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

Sub-case II: $h\big(\boldsymbol{\xi}^*, \tilde{\boldsymbol{\xi}}\big) > 0$

Let $\varepsilon = -\delta - \xi^{m*} < 0 \Rightarrow \tilde{\xi}^m = -\delta$, and hence we get that:

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

since $h\big(\boldsymbol{\xi}^*, \tilde{\boldsymbol{\xi}}\big) > 0$, $\xi^{m*} > -\delta$, 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 1. Part 2.) The proof follows similarly to that of part 1.). However, notice that if $\lambda_1(1 - Se) \geq \lambda_2(1 - Sp)$, then, when $n_i = 1$, Sub-case I is satisfied and the optimal solution is attained at δ. On the other hand, if $n_i > 1$, we have that:

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

and since

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

we get that $h\big(\boldsymbol{\xi}^*, \tilde{\boldsymbol{\xi}}\big) \leq 0$. Hence, Sub-case 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 $1+\delta$. Towards this end, consider an estimated risk vector realization, $\tilde{\boldsymbol{p}}$, and suppose, to the contrary, that the optimal testing scheme, $\boldsymbol{\Omega}^* = \{\Omega_1^*, \cdots, \Omega_g^*\}$, for some $g = 2, \cdots, N$ [5], is not an ordered testing scheme. Then, there must exist two groups, Ω_i^* and Ω_j^*, $i, j = 1, \cdots, g : i \neq j$, such that:

(i) $\min\limits_{m \in \Omega_i^*} \tilde{p}^m < \max\limits_{m \in \Omega_j^*} \tilde{p}^m$, and **(ii)** $\max\limits_{m \in \Omega_i^*} \tilde{p}^m > \min\limits_{m \in \Omega_j^*} \tilde{p}^m$.

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

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

Case I: $n_i = 1$:

Since $n_i = 1$, then 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 $\Psi_i = \{k_i\}$ and let $\Psi_j = \{k_j\}$, and define a new testing scheme, $\hat{\mathbf{\Omega}} = \{\hat{\Omega}_1, \cdots, \hat{\Omega}_g\}$, 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 $l = 1, \cdots, g : l \neq i, j$. 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 $\hat{\mathbf{\Omega}}$ reduces or maintains the value of all performance measures.

(a) **Expected number of false negatives**

We have that:

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

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

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

(b) **Expected number of false positives**

We have that:

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

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

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

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

(c) **Expected number of tests**

We have that:

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

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

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

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

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:

$$\prod_{m\in\Psi_j}(1-\tilde{p}^m)-\prod_{m\in\Psi_i}(1-\tilde{p}^m)>0 \tag{17}$$

Sub-case 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, $\hat{\mathbf{\Omega}}=\{\hat{\Omega}_1,\cdots,\hat{\Omega}_g\}$, 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 $l=1,\cdots,g:l\neq i,j$. In what follows, we will show that partition $\hat{\mathbf{\Omega}}$ reduces or maintains the value of all performance measures. Multiplying the condition imposed in the sub-case, i.e.,

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

by Eq. (17), and 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:

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

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:

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

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

(b) **Expected number of tests ($\mathbf{E}[T]$):**

We have that:

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

Subtracting the two gives:

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

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

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.

Sub-case II: $n_i\prod_{m\in\Omega_i\setminus\Psi_i}(1-\tilde{p}^m) \leq n_j\prod_{m\in\Omega_j\setminus\Psi_j}(1-\tilde{p}^m)$

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, while the group with the larger size contains the lowest risk subjects. Define a new testing scheme, $\tilde{\mathbf{\Omega}} = \{\tilde{\Omega}_1, \cdots, \tilde{\Omega}_g\}$, 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 $l = 1, \cdots, g : l \neq i, j$. In what follows, we will show that partition $\tilde{\mathbf{\Omega}}$ reduces or maintains the value of all performance measures. By the condition imposed in the sub-case, i.e.,

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

and Eq. (17) we get:

$$n_i\prod_{m\in\Omega_i}(1-\tilde{p}^m) \leq n_j\prod_{m\in\Omega_j}(1-\tilde{p}^m). \tag{18}$$

By definitions of Z_i and Z_j, we have that:

$$\prod_{m\in Z_i}(1-\tilde{p}^m) - \prod_{m\in Z_j}(1-\tilde{p}^m) > 0. \tag{19}$$

From Eq. (18) we have that:

$$n_i\prod_{m\in\Omega_i\setminus Z_i}(1-\tilde{p}^m)\prod_{m\in Z_i}(1-\tilde{p}^m) \leq n_j\prod_{m\in\Omega_j\setminus Z_j}(1-\tilde{p}^m)\prod_{m\in Z_j}(1-\tilde{p}^m). \tag{20}$$

Then, by Eq.s (19) and (20), it must be true that:

$$n_i \prod_{m\in\Omega_i\setminus Z_i}(1-\tilde{p}^m) < n_j \prod_{m\in\Omega_j\setminus Z_j}(1-\tilde{p}^m). \tag{21}$$

Multiplying Eq. (21) by Eq. (19), 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 Sub-case I, one can show that $\mathbf{E}[FP(\tilde{\mathbf{\Omega}})] \leq \mathbf{E}[FP(\mathbf{\Omega}^*)]$ and $\mathbf{E}[T(\tilde{\mathbf{\Omega}})] \leq \mathbf{E}[T(\mathbf{\Omega}^*)]$. 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:

$$\begin{aligned}\mathbf{E}\left[\prod_{m=i}^{j} g(X^{(m)})\right] &= \int_a^b \int_a^{x^j} \mathbf{E}\big[g(X^{(i)})\cdots g(X^{(j)})|X^{(i)}=x^i, X^{(j)}=x^j\big] f_{X^{(i)},X^{(j)}}(x^i,x^j)dx^i dx^j \\ &= \int_a^b \int_a^{x^j} g(x^i)g(x^j)\mathbf{E}\big[g(X^{(i+1)})\cdots g(X^{(j-1)})|X^{(i)}=x^i, X^{(j)}=x^j\big] f_{X^{(i)},X^{(j)}}(x^i,x^j)dx^i dx^j, \end{aligned} \tag{22}$$

where

$$\begin{aligned}&\mathbf{E}\big[g(X^{(i+1)})\cdots g(X^{(j-1)})|X^{(i)}=x^i, X^{(j)}=x^j\big] = \\ &\int_{x^i}^{x^j}\int_{x^{i+1}}^{x^j}\cdots\int_{x^{j-2}}^{x^j} g(x^{i+1})\cdots g(x^{j-1}) f_{X^{(i+1)},\cdots,X^{(j-1)}|X^{(i)}=x^i,X^{(j)}=x^j}(x^{i+1},\cdots,x^{j-1})dx^{j-1}\cdots dx^{i+2}dx^{i+1},\end{aligned} \tag{23}$$

where

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

From [33], we have that:

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

and

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

As such,we have that:

$$f_{X^{(i+1)},\cdots,X^{(j-1)}|X^{(i)}=x^i,X^{(j)}=x^j}(x^{i+1},\cdots,x^{j-1}) = (j-i-1)!\frac{f_X(x^{i+1})\cdots f_X(x^{j-1})}{\big(F_X(x^j)-F_X(x^i)\big)^{j-i-1}}. \tag{24}$$

Substituting Eq. (24) into Eq. (23) gives:

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

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 Eq. (23) can be written as:

$$\mathbf{E}\big[g(X^{(i+1)}) \cdots g(X^{(j-1)}) | X^{(i)} = x^i, X^{(j)} = x^j\big] =$$

$$\frac{(j-i-1)!}{\big(F_X(x^j) - F_X(x^i)\big)^{j-i-1}} \int_{x^i}^{x^j} \cdots \int_{x^{j-3}}^{x^j} \left[\int_{x^{j-2}}^{x^j} g(x^{j-1}) f_X(x^{j-1}) dx^{j-1} \right] g(x^{i+1}) \cdots g(x^{j-2}) f_X(x^{i+1}) \cdots f_X(x^{j-2}) dx^{j-2} \cdots dx^{i+1}$$

$$\frac{(j-i-1)!}{\big(F_X(x^j) - F_X(x^i)\big)^{j-i-1}} \int_{x^i}^{x^j} \cdots \int_{x^{j-3}}^{x^j} h(x^{j-2}) g(x^{i+1}) \cdots g(x^{j-2}) f_X(x^{i+1}) \cdots f_X(x^{j-2}) dx^{j-2} \cdots dx^{i+1}$$

$$\frac{(j-i-1)!}{\big(F_X(x^j) - F_X(x^i)\big)^{j-i-1}} \int_{x^i}^{x^j} \cdots \int_{x^{j-4}}^{x^j} \left[\int_{x^{j-3}}^{x^j} g(x^{j-2}) f_X(x^{j-2}) h(x^{j-2}) dx^{j-2} \right] g(x^{i+1}) \cdots g(x^{j-3}) f_X(x^{i+1}) \cdots f_X(x^{j-3}) dx^{j-3} \cdots dx^{i+1}$$

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:

$$\mathbf{E}\big[g(X^{(i+1)}) \cdots g(X^{(j-1)}) | X^{(i)} = x^i, X^{(j)} = x^j\big] =$$

$$\frac{(j-i-1)!}{\big(F_X(x^j) - F_X(x^i)\big)^{j-i-1}} \int_{x^i}^{x^j} \cdots \int_{x^{j-5}}^{x^i} \left[\int_{x^{j-4}}^{x^j} g(x^{j-3}) f_X(x^{j-3}) \frac{h(x^{j-3})^2}{2} dx^{j-3} \right] g(x^{i+1}) \cdots g(x^{j-4}) f_X(x^{i+1}) \cdots f_X(x^{j-4}) dx^{j-4} \cdots dx^{i+1}$$

Continuing in this manner gives:

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

Substituting the latter in Eq.(22) 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. □

B Derivation of Performance Measures

All the subsequent derivations primarily rely on two assumptions: (i) Subjects are assumed to be independent of one another, i.e., 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, i.e., $N_i^+(\Omega_i) = \sum_{m \in \Omega_i} D^m$.

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 $m \in \{1, 2, \cdots, N\}$ and $\boldsymbol{\Omega}$:

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

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

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

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

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

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

B.2 False Positive Classifications

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

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

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

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

leading to:

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

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

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

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

B.3 Number of Tests

Given a partition $\boldsymbol{\Omega}$, the expected number of tests for group i, $i = \{1,\cdots,g\}$, 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, $\boldsymbol{\Xi}$, we can write:

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

Thus,
$$\mathbf{E}[T_i(\Omega_i)|\boldsymbol{\Xi},\tilde{\boldsymbol{P}}] = \begin{cases} 1, & \text{if } n_i = 1, \\ 1+n_i\Big(Se-(Se+Sp-1)\prod_{m\in\Omega_i}(1-t(\tilde{P}^{(m)},\Xi^m))\Big), & \text{otherwise,} \end{cases} \tag{25}$$

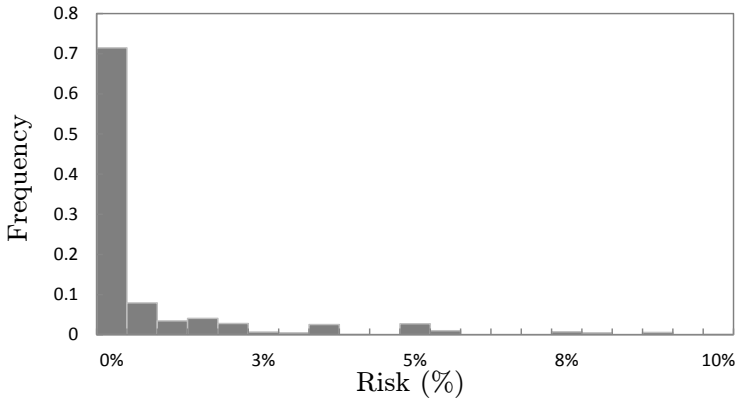
and the expected number of tests needed for all subjects in set S is given by $\mathbf{E}[T(\boldsymbol{\Omega})|\boldsymbol{\Xi},\tilde{\boldsymbol{P}}] = \sum_{i=1}^{g}\mathbf{E}[T_i(\Omega_i)|\boldsymbol{\Xi},\tilde{\boldsymbol{P}}]$.

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 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) [6]. Based on studies, we use an under-reporting factor of 3 for chlamydia [20]. 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. 16), by matching the first two moments of the distribution to those of the data set so as to minimize the

Figure 4: Histogram of the estimated risk based on the CDC data set in [6]



Kolmogorov-Smirnov statistic [28], i.e., to minimize the maximum distance between the empirical and the fitted cumulative distribution functions. Towards this end, we first solve the following system of equations for a given w , i.e., derive expressions for parameters β_1 and β_2 as a function of w:

$$\frac{w}{\beta_1} + \frac{1-w}{\beta_2} = \hat{\mu}_p, \tag{26}$$

$$\frac{w}{\beta_1^2} + \frac{1-w}{\beta_2^2} = \frac{\hat{\mu}_p^2 + \hat{\sigma}_p^2}{2}, \tag{27}$$

where $\beta_1, \beta_2 > 0$. From Eq. (26), we have that:

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

which, when substituted into Eq. (27), leads to the following quadratic equation:

$$\left(\frac{w}{1-w}\right)\left(\frac{1}{\beta_1}\right)^2 - \left(\frac{2\hat{\mu}_p w}{1-w}\right)\left(\frac{1}{\beta_1}\right) + \left(\frac{\hat{\mu}_p^2}{1-w} - \frac{\hat{\mu}_p^2 + \hat{\sigma}_p^2}{2}\right) = 0. \tag{28}$$

There are three cases:

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

In this case, Eq. (28) 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.

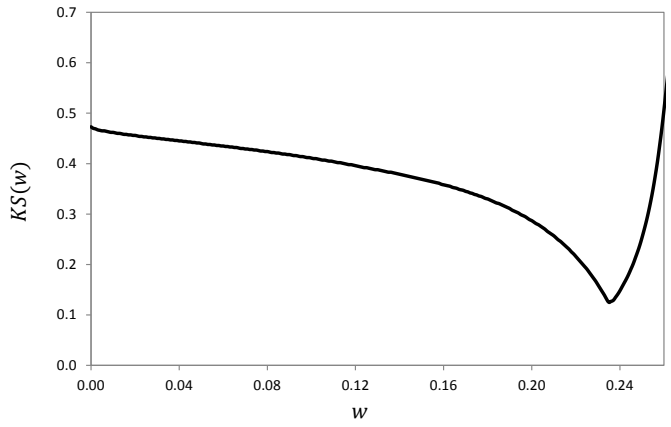
Case II: $1 \leq \hat{CV} < \sqrt{3}$:

In this case, the system of equations given in Eq.s (26)-(27) has two solutions for all w such that:

$$0 \leq 1 - \frac{2}{\hat{CV}^2 + 1} < w < \frac{2}{\hat{CV}^2 + 1} \leq 1. \tag{29}$$

If w does not satisfy Eq. (29), then β_1 or β_2 will not be positive. For all w that satisfies Eq. (29),

Figure 5: Kolmogorov-Smirnov test statistic as a function of w



the two solutions are given by:

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

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

In this case, the system of equations given in Eq.s (26)-(27) has a unique solution for all w such that:

$$w < \frac{2}{\hat{CV}^2 + 1} \quad \text{or} \quad w > 1 - \frac{2}{\hat{CV}^2 + 1}. \tag{30}$$

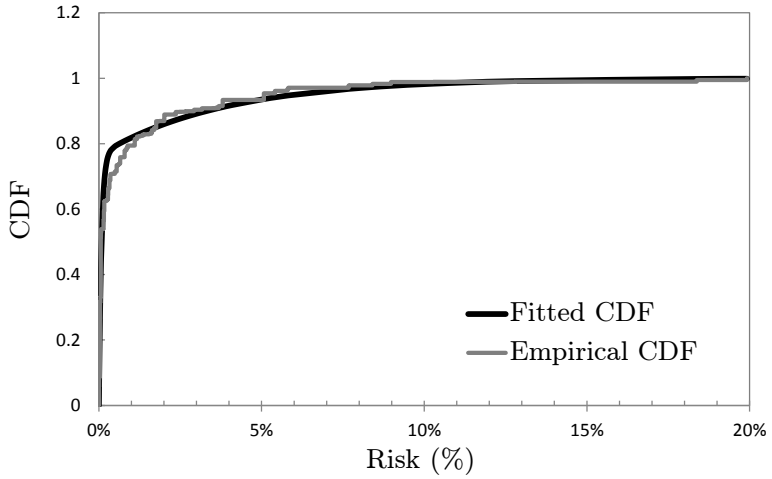
Note that due to the condition imposed in Case II, i.e., $\hat{CV} \geq \sqrt{3}$, w can satisfy at most one of the two inequalities in Eq. (30). If w does not satisfy any of the inequalities in Eq. (30), then β_1 or β_2 will not be positive. For all w that satisfies Eq. (30), the unique solution is given by:

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

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 β_1 and β_2. As such, by limiting the case to:

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

Figure 6: Empirical and fitted cumulative distribution functions (CDF), when $w = 0.235$, $\beta_1 = 25.708$, and $\beta_2 = 1,291.832$



the unique solution reduces to:

$$\frac{1}{\beta_1} = \hat{\mu}_p \left(1 + \sqrt{\frac{(\hat{CV}^2 - 1)(1-w)}{2w}}\right), \tag{31}$$

$$\frac{1}{\beta_2} = \hat{\mu}_p \left(1 - \sqrt{\frac{(\hat{CV}^2 - 1)w}{2(1-w)}}\right). \tag{32}$$

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 Eq.s (31) and (32), respectively.

As mentioned above, w is chosen so as to minimize the Kolmogorov-Smirnov test statistic (KS) [28], i.e., the maximum distance between the empirical and fitted cumulative distribution functions. That is, w is the optimal solution to the following optimization problem:

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

Figure 5 plots $KS(w)$ as a function of w for the data set used in the case study, and shows that function $KS(w)$ is unimodal in w, with a global minimizer, $w = 0.235$, leading to $KS(0.235) = 0.125$. Then, from Eq.s (31) and (32), 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 [28], and hence, $KS(0.235) = 0.125 < 0.163$, and 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.

D The Effect of Dilution

Figure 7: Fitted linear sensitivity function versus empirical results [39] as a function of group size