Economic Theory (2024) 77:413–444
https://doi.org/10.1007/s00199-023-01502-3

RESEARCH ARTICLE

Optimal group testing with heterogeneous risks

Nina Bobkova[1] · Ying Chen[2] · Hülya Eraslan[3,4]

Received: 9 August 2022 / Accepted: 8 May 2023 / Published online: 6 June 2023

Abstract

We consider optimal group testing of individuals with heterogeneous risks for an infectious disease. Our algorithm significantly reduces the number of tests needed compared to Dorfman (Ann Math Stat 14(4):436–440, 1943). When both low-risk and high-risk samples have sufficiently low infection probabilities, it is optimal to form heterogeneous groups with exactly one high-risk sample per group. Otherwise, it is not optimal to form heterogeneous groups, but homogeneous group testing may still be optimal. For a range of parameters including the U.S. Covid-19 positivity rate for many weeks during the pandemic, the optimal size of a group test is four. We discuss the implications of our results for team design and task assignment.

Keywords Group testing · Pooled testing · Positive assortative matching · Negative assortative matching · Heterogeneous risks

JEL Classification C610 · D0 · I18

We thank Bob Barbera, Olivier Gossner, Eric Schliesser, a referee, the Editor, and participants at HKBU, Kyoto KIER, NTU, Osaka ISER and Sinica joint seminar, Georgetown, Bonn, Kansas Workshop in Economic Theory, Glasgow and Rice theory group meeting for very helpful comments and suggestions. Eraslan gratefully acknowledges support from National Science Foundation under Grant SES-1730636. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

✉ Ying Chen
ying.chen@jhu.edu

Nina Bobkova
nina.bobkova@rice.edu

Hülya Eraslan
eraslan@rice.edu

1 Department of Economics, Rice University and CEPR, Houston, USA

2 Department of Economics, Johns Hopkins University, Baltimore, USA

3 Department of Economics, Rice University and NBER, Houston, USA

4 Institute of Social and Economics Research, Osaka University, Osaka, Japan

1 Introduction

The idea of group testing originated in the 1940s, when the U.S. government needed to save on the cost of screening World War II draftees for syphilis. Instead of testing each soldier individually, which was too costly, Dorfman (1943) proposed to pool samples of soldiers in groups and test them together to find out if at least one of the soldiers in the combined sample was infected. As Dorfman wrote, group testing could "yield significant savings in effort and expense when a complete elimination of defective units is required."

Allocation of tests has become an especially relevant economic problem during the global coronavirus pandemic because of the scarcity of testing resources in many countries. Group testing has been proposed as one important way to expand the testing capacity and use the available testing resources more efficiently and has been approved by the FDA in July 2020.[1] To improve group testing, our study takes into account that a population needing tests is heterogeneous in terms of risk: for example, some people are known to have been exposed to infected people and others are not; people may differ in terms of if and when they tested negative in the past, and people may have different levels of exposure depending on their living situations and workplace settings.[2] How does this heterogeneity affect the optimal testing of samples? Should we group samples with similar or different risks together?

In this paper, we formalize these questions and answer them by studying a model in which a decision maker, facing a population of high-risk and low-risk samples, aims at minimizing the cost of testing while detecting all infected samples. We propose an algorithm that modifies Dorfman (1943)'s algorithm (AD for short). In AD, a group of samples is tested together first. Since tests are assumed to be perfect, a negative group test clears all samples. If the group test is positive, each sample in the group is tested individually. We modify AD so that if the group test is positive, then each sample except the last one in the group is tested individually. If at least one of the individual tests is positive, then the last sample is tested as well. But if all the individual tests are negative, then the last sample is inferred to be positive and there is no need to test it. Under our group-testing algorithm (denoted by AG), for a group consisting of samples of heterogeneous risks, it is strictly better to test a high-risk rather than a low-risk sample last since it increases the probability that there is no need to test the last sample.

Given a heterogeneous population, what is the optimal testing schedule? To illustrate, suppose the group testing capacity is two, that is, up to two samples can be tested together. Group testing is better than individual testing if the probability that a group test is negative is high, so that no further testing is needed; otherwise individual testing is better. Hence, if the probability of infection for low-risk samples is above a certain threshold, then it is optimal to test all samples individually. If the probability

[1] For a list of countries implementing a group testing strategy against COVID-19, see https://en.wikipedia.org/wiki/List_of_countries_implementing_pool_testing_strategy_against_COVID-19.

[2] In campus surveillance testing that many universities conducted in the the academic year 2020–2021, it is recognized that people have different risks of exposure depending on residential setting or workplace environment. See, for example, https://uhs.berkeley.edu/coronavirus/testing-covid-19/campus-surveillance-testing.

of infection for low-risk samples is below the threshold, however, then the optimal testing schedule depends on the magnitude of the probability of infection for high-risk samples. If it is high, then it is optimal to test the high-risk samples individually and the low-risk samples using AG. But if it is low, then it is optimal to pool the high-risk and low-risk samples together and test them using AG. (We call this mixing of different risk types in group testing *negative assortative matching*, or NAM for short.)

It might seem somewhat surprising at first that NAM can be optimal. After all, as we show, testing homogeneous samples together achieves the lowest average probability of positive group tests. However, there is an advantage of NAM which comes in the second stage of AG after a positive group test. Recall that with AG, if the group test turns out to be positive, then it is strictly better to test the low-risk sample first when the two samples have different risks. When the two samples have the same risks, it is irrelevant which sample to test first. These imply that there is gain to be made in the second step of AG when the two samples are heterogeneous, making NAM optimal when both low-risk and high-risk are sufficiently small.

We generalize this characterization for any testing capacity. We show that if a mixed group is part of an optimal testing schedule, then it contains only one high-risk sample. Intuitively, the more high-risk samples are in a mixed group, the higher the number of expected tests required after a positive group test, so the gain from a mixed instead of a homogeneous group disappears if the group contains more than one high-risk sample. NAM is robust in the sense that it is optimal for any testing capacity when the probability of infection is small for all samples.

The main protocol we use has two variations compared to Dorfman (1943). First, it explicitly incorporates heterogeneity of risks. This aspect has been studied little in group testing. Our paper is the first one to provide a closed-form solution of the optimal testing schedule for small testing capacity and a partial characterization for arbitrary testing capacity under the assumption that each sample can be tested at most twice. Second, if a group of samples is tested in a pool, and if the result is positive, all samples are tested individually in Dorfman (1943). However, in our protocol, no individual test for the last sample is required if all previous samples in this group have tested negative. When there are no errors in testing, this second variation can save one test in a group that contains exactly one positive sample. If errors happen with a high probability in practice, this second variation might have limited applications. Below, we decompose the gains from our procedure compared to Dorfman (1943) and show that most of the gains can come from an explicit recognition of the heterogeneity of risks. For example, a simulation with the probability of infection being 0.05 for the low-risk individuals and 0.3 for the high-risk individuals and 80% of the population being low-risk shows that AG saves around 13.3% of tests compared with AD that ignores the heterogeneity of risk. The savings stem from both modifications of the AD algorithm: Of the 13.3% reduction of tests, about 9.3% is due to the incorporation of heterogeneity of risk and 4% is due possibly avoiding the last individual test in a group. Since information about a person's risk is often readily available either through observables (for example, presence/absence of symptoms, testing history, living/working situations) or self-reporting, efficiency of group testing can be significantly improved by the adoption of the algorithm that we propose and the optimal

sorting of samples. Overall, our analysis shows the importance of taking into account heterogeneity of risk in the population when conducting group testing.

Due to the assumption of an infinite population, our optimization problem can be written as a linear program. Even though we can solve the problem using established techniques in polynomial time, here we establish qualitative features of the solution which are robust to a range of parameters.[3] In addition, for particular cases (for example, sufficiently low infection probability), our results provide guidelines for policymakers and for labs, without requiring each of them to formalize the problem and to run a linear programming solver. Finally, even though we have framed our paper as characterizing optimal group testing with a population that has heterogeneous probabilities of infection, the results derived have implications for task assignment and team design in principal-agent relationships. We discuss various extensions in Sect. 5.

Related literature

As mentioned earlier, the idea of group testing dates back to Dorfman (1943). Dorfman (1943) derives the optimal group size for an infinite population of samples with homogeneous risk and perfect tests. In contrast to Dorfman (1943), our algorithm is adaptive to test outcomes and we allow for heterogeneous infection probabilities.

The algorithm in Dorfman (1943) is still widely in use in public health today, for example, for testing blood donations for HIV and hepatitis viruses (see, e.g., Aprahamian et al. 2016). Some papers have allowed for heterogeneous risk probabilities (e.g., Hwang 1975), proposed further improvements on the group testing algorithm (e.g., Du and Hwang 2000; Sterrett 1957) and characterized optimal group test design under various objectives when tests are imperfect (Aprahamian et al. 2019). Our model also connects to the group testing literature in computer science and industrial engineering (see, e.g., Black et al. (2012), Du and Hwang (2000) and the references therein), where the focus tilts towards comparing the (worst case) running time of different group testing algorithms.

Since the outbreak of the coronavirus pandemic in 2020, a lack of testing capacity has been a persistent problem in many countries and there has been renewed interest in group testing. Gollier and Gossner (2020) study optimal group testing, allowing for false positives and negatives; Augenblick et al. (2020) investigate how the optimal group testing strategy changes with repeated testing, correlation of infection probability between samples and uncertain prevalence rate, using a machine-learning approach. In contrast to our study, these studies do not consider heterogeneous risk probabilities in the population. Three recent papers allow individual heterogeneity, but do not consider group testing and address questions different from ours. Ely et al. (2021) analyze a (corona)-test allocation problem to individuals with heterogeneous risks. Deb et al. (2022) also incorporate heterogeneous agents, but in their model, the agents are different in terms of their beliefs about whether they are infected, and the paper combines testing and transfers to address the question of whether agents should be tested and how they should be incentivized. Makris (2021) considers a model of social distancing in which agents are heterogeneous with respect to infection-induced fatality risks. The closest paper to our framework is Lipnowski and Ravid (2021) which considers

[3] Many problems for which economists care about the qualitative properties, e.g, in mechanism design and information design, can be rewritten as a linear program. For example, see Vohra (2011).

testing a heterogeneous-risk population in groups. In their model, each agent can be tested at most once and not every individual needs to be tested.[4] In contrast to our model, they assume that each sample can be tested at most once (whether in a group or individually). Thus, they do not consider an adaptive group testing protocol after a positive group test. In Lipnowski and Ravid (2021), negative assortative matching between risk groups never arises. This is because the probability of an initial positive group result is lower when testing homogeneous samples rather than heterogeneous samples together in a group. This effect is also present in our model. However, after a positive group test in our paper, testing continues within the group, and there are potential gains to be made from including heterogeneous risk types in this later stage. Thus, in contrast to Lipnowski and Ravid (2021), negative assortative matching can be optimal in our model.

Another paper that allows for heterogeneous risks among the samples using Dorfman's algorithm is Hwang (1975). Notably, in contrast to our model, Hwang (1975) considers a finite number of samples, each of which can have a different risk probability. This raises the problem of how to group samples of heterogeneous risks, for which Hwang develops a dynamic programming algorithm. Note that if the sample population in Hwang (1975) were countably infinite as in our model, then the optimal group testing using AD would have a very simple solution: only test homogeneous risk types in a group. The argument is the same as in Lipnowski and Ravid (2021) — testing samples in homogeneous groups minimizes the probability of a positive group test, and hence, the number of subsequently required individual tests in the non-adaptive AD. Hence, Hwang (1975) tackles the integer problem of the finite sample population assumption for AD.[5] In contrast, our adaptive algorithm AG does not require testing all samples after a positive group result. Therefore, the optimal testing schedule may require testing heterogeneous risk types in a group even for a countably infinite sample population, and not just as an integer problem effect of a finite sample population assumption as in Hwang (1975).

2 Model

A decision maker (henceforth DM) faces a countably infinite set of samples (test swabs).[6] Each sample s_i ($i \in \mathbb{N}$) is either infected or healthy. The samples consist of two heterogeneous risk groups categorized by their probabilities of infection. A type $\theta \in \{h, \ell\}$ sample has a probability of infection p_θ where $1 > p_h \geq p_\ell > 0$. The

[4] In contrast to our paper and most of the literature on group testing, Lipnowski and Ravid (2021) focus on quarantine decisions: the decision-maker's goal of testing is to maximize the number of uninfected agents who are not quarantined.

[5] While Hwang (1975) allows for several arbitrary risk types, he does not provide a closed-form expression for how to form groups beyond the algorithm of his Theorem 1.

[6] We make this assumption for expositional convenience. Alternatively, we could assume that the DM faces a finite set of samples, but it would involve considering many special cases without generating additional insights.

fraction of low-risk samples in the overall population is α. The probability of being infected is independent across samples.[7]

The decision maker has access to an infinite number of tests. She can test several samples together (group testing) up to a capacity of n, or test each sample individually (individual testing). For simplicity, we assume that the testing technology is perfect for both group and individual testing. When multiple samples are tested together via group testing, the test outcome is positive if and only if at least one of the samples in the tested group is infected.[8] Like Dorfman (1943) and many papers following it, we impose the following feasibility assumption on the number of admissible tests per sample:

Assumption 1 The decision maker can test each sample at most twice.

Repeated (that is, more than twice) testing of the same sample might be difficult due to dilution of the initial sample so that detection becomes harder, due to complexity considerations on procedures within a lab, or due to regulatory restrictions such as CDC guidelines.[9] In addition, solving for the optimal testing algorithm without Assumption 1 makes the problem computationally complex.[10]

Tests are costly and we assume that the objective of the DM is to minimize the expected number of tests per sample while identifying all infected samples. Before we formalize this problem, we introduce some notation. Recall that the test capacity is n, that is, up to n samples can be tested together in a group test. Let $\mathcal{S}_n$ be the set of all groups up to size n. We describe each sample by its type and each group S by an enumeration $\{s_i\}_{i=1}^k$ of the samples in it, where $k = 1, \ldots, n$ and $s_i \in \{\ell, h\}$ for each i. For example, for $n = 2$, the set of possible groups are $\{\ell, h, \ell\ell, hh, \ell h, h\ell\}$. Let $|S|$ be the cardinality of a group $S \in \mathcal{S}_n$, for example, $|\ell\ell h| = 3$. Similarly, let $|S|_\ell$ be the cardinality of the ℓ-types and $|S|_h$ be the cardinality of the h-types in S, for example, $|\ell\ell h|_\ell = 2$ and $|\ell\ell h|_h = 1$. It will be convenient to partition the set $\mathcal{S}_n$ into (i) $\mathcal{S}_n^\ell$, a set containing only homogeneous groups consisting of low-risk types, (ii) $\mathcal{S}_n^h$, a set containing only homogeneous groups consisting of high-risk types, and (iii) $\mathcal{S}_n^m$, a set containing every mixed group:

$$\mathcal{S}_n = \mathcal{S}_n^\ell \cup \mathcal{S}_n^h \cup \mathcal{S}_n^m.$$

2.1 Two algorithms

When there is only one sample to test in a group, that is, when the group S is singleton, the only algorithm to use is the individual testing algorithm which we denote by AI. To use individual testing protocol on a group S with $|S| \geq 2$ is equivalent to first

[7] We discuss the possibility of correlation in Sect. 5.3.

[8] We discuss what happens if there are type I and type II errors in Sect. 5.2.

[9] See https://www.cdc.gov/coronavirus/2019-ncov/lab/pooling-procedures.html.

[10] We relax Assumption 1 in Sect. 5.4 by allowing each sample to be tested three times which substantially increases the difficulty of the optimization problem. In fact, without an assumption on maximal tests per sample, the computational complexity class of this problem even with homogeneous samples has not yet been determined but conjectured to be hard in some complexity class (see Du and Hwang, 2000).

partitioning S into the singleton groups containing its elements, and then using the individual testing protocol for each $\{s_i\}$ for $s_i \in S$.

We define the group testing algorithm AG below only for non-singleton groups $S = \{s_i\}_{i=1}^{k}$ where $k = 2, \ldots, n$:

Algorithm (AG: Group Testing) *Test samples in* $\{s_i\}_{i=1}^{k}$ *where* $s_i \in \{\ell, h\}$ *together.*

1. *If negative, stop. Each sample in* $\{s_i\}_{i=1}^{k}$ *is healthy.*
2. *If positive, test each sample in* $\{s_i\}_{i=1}^{k-1}$ *individually.*
   (a) *If every sample in* $\{s_i\}_{i=1}^{k-1}$ *is negative, stop. Only* s_k *is infected.*
   (b) *If at least one sample in* $\{s_i\}_{i=1}^{k-1}$ *is positive, test* s_k *and stop. If* s_k *negative, only the samples in* $\{s_i\}_{i=1}^{k-1}$ *who tested positive are infected. Otherwise, the samples in* $\{s_i\}_{i=1}^{k-1}$ *who tested positive and* s_k *are infected.*

Our group testing protocol AG differs from Dorfman's (1943) protocol AD in that under AD, if the group test turns out to be positive, then all samples are tested individually whereas under our protocol, if the group test turns out to be positive, we test all but one sample individually. If the individual tests are all negative, then we do not need to test the last sample and infer that it is infected.[11]

To find the optimal testing protocol under Assumption 1, we do not need to consider other testing algorithms. After an individual test, the DM knows whether that individual is infected or not. After a group test, the DM can test each sample in the group only one further time (due to Assumption 1). If she tests them next in a group instead of separately, she might not learn whether the agent is positive or negative. This is ruled out by assumption as the DM needs to identify all positive samples.

Let ϕ_S denote the average number of tests required per sample within group S with the assumption that for $|S| \geq 2$, samples in S are tested with algorithm AG. Trivially, ϕ_S is 1 when $|S| = 1$. Next, consider $|S| \geq 2$.

If the group test is negative (this happens with probability $(1-p_\ell)^{|S|_\ell}(1-p_h)^{|S|_h}$), then the DM only requires one test. If the group test is positive, then the DM has to start testing the samples individually and there are two possibilities: (i) there is at least one positive sample among the first $|S|-1$ individual tests, or (ii) the first $|S|-1$ individual tests are negative. In case (i), the DM requires $|S|+1$ test in total—the initial group test and an individual test for every single sample in $|S|$. In case (ii), the DM requires $|S|$ tests in total as the last sample does not need to be tested in AG.

Thus, when testing a homogeneous group S consisting only of type θ samples with $p_\theta = p$, the average number of tests required per sample is

$$\begin{aligned}\phi_S &= \frac{1}{|S|}\left[(1-p)^{|S|} + (1-p)^{|S|-1}p|S| + (1-(1-p)^{|S|}-(1-p)^{|S|-1}p)(|S|+1)\right] \\ &= 1-(1-p)^{|S|} + \frac{1}{|S|}(1-p(1-p)^{|S|-1}). \qquad (1)\end{aligned}$$

[11] Finucan (1964) mentions a refinement of the AD protocol for a group of size two, of which our AG protocol is a generalization. Moreover, Finucan (1964) considers only a homogeneous population where the order of the different risk types is irrelevant.

Next, consider the average number of tests per sample in a heterogeneous group with AG. After a positive group test, the DM can only avoid testing all $|S|$ individual samples if the first $|S|-1$ individual tests are negative. To maximize the probability of this happening, a high risk sample should be tested last after a positive group test:

Lemma 1 *In AG, for any mixed group, it is strictly better to test a high-risk, rather than a low-risk sample, last following a positive group test result.*

Using Lemma 1, the average tests per sample in a heterogeneous group S is

$$\begin{aligned}\phi_S &= \frac{1}{|S|}[(1-p_\ell)^{|S|_\ell}(1-p_h)^{|S|_h} + (1-p_\ell)^{|S|_\ell}(1-p_h)^{|S|_h-1}p_h|S| \\ &\quad + \left(1-(1-p_\ell)^{|S|_\ell}(1-p_h)^{|S|_h} - (1-p_\ell)^{|S|_\ell}(1-p_h)^{|S|_h-1}p_h\right)(|S|+1)] \\ &= 1-(1-p_\ell)^{|S|_\ell}(1-p_h)^{|S|_h} + \frac{1}{|S|}(1-p_h(1-p_\ell)^{|S|_\ell}(1-p_h)^{|S|_h-1}). \quad (2)\end{aligned}$$

2.2 Optimization problem of the DM

Let f_S denote the fraction of the overall sample population that the DM tests in a group $S \in \mathcal{S}_n$. For example, if $f_{\ell\ell h} = 0.6$, then 60% of the overall sample population is tested in the group $\{\ell\ell h\}$. As $|S|_\ell/|S| = 2/3$ of those are low risk (1/3 are high risk), this requires that at least 40% of the sample population is of low risk ($\alpha \geq 0.4$), and at least 20% is of high risk for $f_{\ell\ell h} = 0.6$ to be part of a feasible testing schedule. Since the DM seeks to minimize the average number of tests per sample, her optimization problem can be expressed as the following linear minimization problem subject to two binding feasibility constraints,

$$\begin{aligned}&\min_{\{f_S\}_{S\in\mathcal{S}_n}} \sum_{S\in\mathcal{S}_n} f_S\phi_S \\ &\text{subject to } \sum_{S\in\mathcal{S}_n^\ell} f_S + \sum_{S\in\mathcal{S}_n^m} \frac{|S|_\ell}{|S|} f_S = \alpha \quad (3) \\ &\qquad \sum_{S\in\mathcal{S}_n^h} f_S + \sum_{S\in\mathcal{S}_n^m} \frac{|S|_h}{|S|} f_S = 1-\alpha \\ &\qquad f_S \in [0,1] \text{ for all } S \in \mathcal{S}_n. \quad (4)\end{aligned}$$

Note that in our optimization problem, we assume that for any group S with $|S| \geq 2$, algorithm AG is used. This is without loss of generality since if AI were better, then the DM would have shifted the weight to f_S such that $|S| = 1$. As this is a linear program with a bounded objective function and a non-empty feasible constraint set, a solution exists.

For any $\mathbf{f} = \{f_S\}_{S\in\mathcal{S}_n}$, if $f_S = 0$ for all S with $|S| \geq 2$, we say $\mathbf{f}$ is an *individual (IND) testing schedule*; if $f_S > 0$ for some $S \in \mathcal{S}_n^m$, we say $\mathbf{f}$ is a *negative assortative matching (NAM) testing schedule*; otherwise we say $\mathbf{f}$ is a *positive assortative matching (PAM) testing schedule*.

3 Optimal testing with capacity two

In this section, we focus on a testing capacity of $n = 2$. This case is interesting in its own right and serves as a building block for arbitrary testing capacity. For this capacity, is it ever optimal to test two samples together with AG? And if yes, is it ever optimal to choose a NAM testing schedule? To answer these, it is instructive to first consider the optimal algorithm for samples with a homogeneous risk probability.

In what follows, let $\hat{p} := \frac{1}{2}(3 - \sqrt{5}) \approx 0.38$.

Lemma 2 (Homogeneous Population) *Fix testing capacity $n = 2$. Consider an infinite number of samples with homogeneous infection probability $p \in (0, 1)$.*

(i) *AI (individual testing) is optimal if and only if $p \geq \hat{p}$.*
(ii) *AG (group testing) is optimal if and only if $p \leq \hat{p}$.*

Proof of Lemma 2 For a singleton group S, the average number of tests per sample with AI is $\phi_S = 1$. According to (1), the average tests per sample in a homogenous group S of size $|S| = 2$ with infection probability p is

$$\phi_S = \frac{1}{2}(1 + 3p - p^2).$$

Thus, a group S requires strictly less than one test on average per sample if and only if $1 > \frac{1}{2}(1 + 3p - p^2)$, or equivalently, $p < \hat{p}$. □

Intuitively, for group testing to be optimal, the probability of a positive group test has to be sufficiently low, which happens for sufficiently low probability of infection. If a group test is positive, the DM requires at least 2 tests for the group, which would have made AI a better choice. This is consistent with the CDC guideline that "pooling should be used only in areas or situations where the number of positive test results is expected to be low – for example in areas with a low prevalence of SARS-CoV-2 infections."

Lemma 2 is useful for our following main result in this section which characterizes the optimal testing schedule for heterogenous population when the testing capacity is two.

Proposition 1 (Optimal testing schedule with capacity two) *Fix testing capacity $n = 2$.*

(i) *(IND) If $\hat{p} \leq p_\ell$, then it is optimal to test every sample individually.*
(ii) *(PAM) If $p_\ell \leq \hat{p}$ and $p_h \geq \frac{1-p_\ell^2-p_\ell}{2(1-p_\ell)}$, then it is optimal to test all ℓ-types with AG and all h-types with AI.*
(iii) *(NAM) If $p_\ell \leq \hat{p}$ and $p_h \leq \frac{1-p_\ell^2-p_\ell}{2(1-p_\ell)}$, then it is optimal to maximize the fraction of samples in a mixed group $S = \{\ell h\}$ and to test the remaining homogeneous samples according to Lemma 2.*

Figure 1 shows the three parameter regions divided according to the above result: (i) all samples are tested individually (IND region); (ii) all low-risk samples are tested

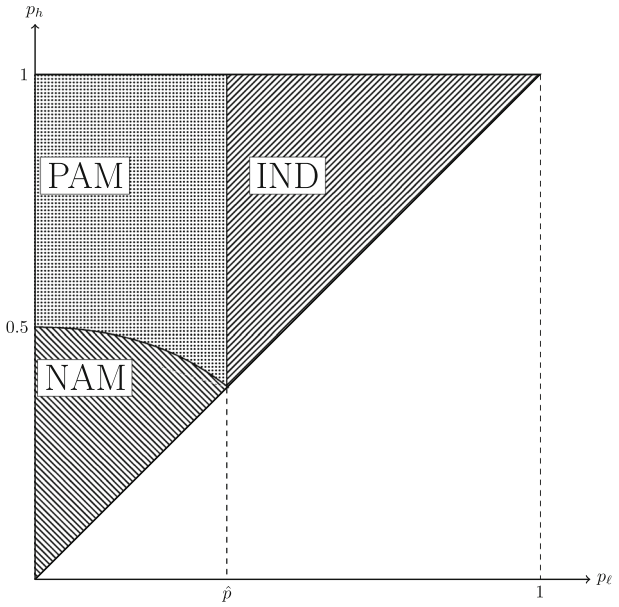

Fig. 1 Optimal testing schedule with capactiy $n = 2$

in groups of 2, while the high-risk samples are tested individually (PAM region); (iii) the DM maximizes the number of heterogeneous groups which she tests in groups of two (NAM region).

In region IND, the probability of infection for a low-risk sample is already so high that testing two ℓ-samples together is not profitable (see Lemma 2). Thus, the optimal testing schedule consists only of individual tests: $f_\ell = \alpha$ and $f_h = 1 - \alpha$.

Under which conditions does the optimal testing schedule involve a positive fraction of mixed groups $S = \{\ell h\}$ tested with AG? For this to be optimal, splitting up any mixed groups into either homogeneous groups (with AG) or individual samples tested with AI should not decrease the average number of expected tests per sample. With testing capacity of 2, this boils down to whether the DM prefers to test four samples $\{\ell\ell hh\}$ in two heterogeneous groups or separately for each risk-type (according to AI or AG based on Lemma 2).

The difference between p_h and p_ℓ determines whether PAM or NAM testing is optimal. For p_h sufficiently large, the DM prefers to test all low-risk samples with AG, and all high-risk samples with AI. This is depicted in the PAM region, and corresponds to a testing schedule $f_{\ell\ell} = \alpha$ and $f_h = 1 - \alpha$.

For sufficiently low p_h, NAM becomes optimal. It may seem somewhat surprising at first that NAM can be optimal. We give some intuition here. Recall from Lemma 1 that with the AG protocol, if the group test turns out to be positive, then it is strictly better to test the lower risk sample first if the two samples have different risks. If the two samples have the same risks, it is irrelevant which sample to test first. This implies that there are gains to be made in the second step of AG when the two samples are heterogeneous, making NAM optimal when both p_ℓ and p_h are sufficiently low, the

Table 1 Optimal testing schedule in the NAM region

	$p_h \geq \hat{p}$	$p_h \leq \hat{p}$
$\alpha < 1/2$	$f_{\ell h} = 2\alpha$, $f_h = 1 - 2\alpha$	$f_{\ell h} = 2\alpha$, $f_{hh} = 1 - 2\alpha$
$\alpha > 1/2$	$f_{\ell h} = 2 - 2\alpha$, $f_{\ell\ell} = 2\alpha - 1$	$f_{\ell h} = 2 - 2\alpha$, $f_{\ell\ell} = 2\alpha - 1$

conditions for AG to be better than AI. Strikingly, this region where NAM is optimal extends beyond the case when $p_\ell \leq p_h \leq \hat{p}$: Fig. 1 shows that the NAM region partially encompasses the region where $p_h > \hat{p}$. For such parameter constellations, high-risk samples would be tested individually (AI) when tested among themselves, but the DM is strictly better off testing them in groups with NAM in a population with heterogeneous risks.

In the NAM region, if $\alpha = 1/2$, then the DM can form exclusively mixed groups and $f_{\ell h} = 1$ is the optimal testing schedule. Otherwise, for $\alpha \neq 1/2$, the optimal testing schedule in the NAM region depends on (i) which risk-types are in the majority and need to be tested separately after the DM maximizes the fraction of NAM groups $f_{\ell h} = \min\{2\alpha, 2(1-\alpha)\}$, and (ii) the infection probability for these left-over samples which determines whether it is optimal to test them with AI or AG according to Lemma 2.

Table 1 summarizes all optimal test schedules that can arise in the NAM region.

To summarize, for any parameters p_ℓ and p_h, the DM never requires more than two different groups in the optimal testing schedule. The parameter α is only relevant for determining the fraction of samples which can be matched optimally via heterogeneous groups, but not for determining if the optimal testing schedule is PAM or NAM.

4 Optimal testing with general capacity

Next, we consider the general case in which the DM can choose any group size below some arbitrary testing capacity $n \geq 2$. Theorem 1 summarizes the main results.

Theorem 1 *Fix testing capacity n. In an optimal testing schedule, any mixed group contains exactly one h-type sample. Moreover, for p_h sufficiently low, any optimal testing schedule*

(i) *consists of at most two different group types,*
(ii) *contains exactly one mixed group type,*
(iii) *contains groups of size n only.*

The result that any mixed group contains exactly one h-type sample does not depend on p_h and follows immediately from Lemma 4 which we establish in Sect. 4.2. Of course when p_h is high, it is optimal to test high-risk samples individually. Therefore, NAM arises only when p_h is sufficiently low. In this case, all samples are tested at capacity n. Moreover, the optimal schedule contains a mixed group with exactly one h-type, and the remaining samples are tested in homogeneous groups. These follow from Lemma 6 which is also established in Sect. 4.2. The proof of Lemma 6 provides a precise bound on p_h for these results to be satisfied.

Table 2 Optimal group size for a homogeneous population depending on infection probability p

Optimal group size	p range in AG	p range in AD
1	0.382–1	0.307–1
2	0.203–0.382	–
3	0.103–0.203	0.124–0.307
4	0.060–0.103	0.066–0.124
5	0.039–0.060	0.041–0.066
6	0.027–0.039	0.028–0.041
7	0.020–0.027	0.021–0.028
8	0.016–0.020	0.016–0.021

The remainder of this section provides intuition for the theorem and establishes additional results.

4.1 Testing homogeneous samples

We start by considering the optimal testing algorithm if the DM has to test homogeneous samples with an infection probability of p. This problem arises for every risk group separately under PAM, or if there are samples with the same risk probability left over after forming as many mixed groups as possible.[12]

First, consider the unconstrained problem of finding the optimal homogeneous group size absent a capacity constraint. An optimal group size solves

$$\arg\min_{n\in\mathbb{N}} 1-(1-p)^n+\frac{1}{n}(1-p(1-p)^{n-1}). \tag{5}$$

Table 2 shows the optimal group size as it varies in probability p, up to a group size of 8. For comparison, it includes a column for the optimal group size with AD. Under group testing, the number of tests is strictly lower than individual testing only if the group result is negative. Otherwise, the number of tests is higher than that of under individual testing. The larger the probability of being infected, the more likely that a group test will turn out positive, and therefore, the number of tests under group testing will exceed the number of tests under individual testing. To compensate for this effect, optimal group size decreases in the probability of infection.

In mid-July 2020 (when the FDA approved the pooled testing for the coronavirus), the positivity rate for the coronavirus tests in the U.S. averaged around 7%.[13] If no data about high or low risk is available, then exactly 4 samples should be grouped together, both in our proposed algorithm AG as well as in AD. Notably, this exactly corresponds to the maximal group size which the FDA approved.

The following result establishes that for a sufficiently small infection probability, the testing capacity is a binding constraint.

[12] Similarly, this is the DM's problem if she does not know the risk group of each sample.

[13] See https://coronavirus.jhu.edu/testing/individual-states.

Lemma 3 (Binding test capacity for sufficiently low risk) *Fix capacity n and a countably infinite number of homogeneous samples with infection probability p. There exists $\tilde{p} > 0$, such that for all $p \leq \tilde{p}$, testing n samples in a group S' with $|S'| = n$ requires strictly fewer expected tests than any other group S containing fewer samples:*

$$\phi_{S'} < \phi_S \quad \text{for all } S \text{ such that } \quad |S| < n.$$

4.2 Testing heterogeneous samples

We now consider testing heterogeneous samples. When more than two samples can be tested together, then the DM can form different mixed groups. For example, with test capacity of $n = 4$, there are six candidates for a mixed group: $\{\ell h\}$, $\{\ell\ell h\}$, $\{\ell hh\}$, $\{\ell\ell\ell h\}$, $\{\ell\ell hh\}$, and $\{\ell hhh\}$.[14] Which of these can be part of an optimal testing schedule? Strikingly, the next result shows that the answer is independent of p_ℓ and p_h, and is the same for every n: no mixed group is optimal if it contains two or more high-risk samples.

Lemma 4 (No mixed groups with more than one h-type) *For all $S \in \mathcal{S}_n^m$ with $|S|_h > 1$, $f_S = 0$ in any optimal testing schedule.*

Lemma 4 drastically reduces the types of mixed groups which can arise in the optimal testing schedule. For example, if the DM ever tests a mixed group containing 4 samples, it will be group $\{\ell\ell\ell h\}$; testing $\{\ell\ell hh\}$ or $\{\ell hhh\}$ is never optimal.

The composition of a group S that is tested with AG determines (i) the probability of a positive group test result, and (ii) the probability of requiring $|S| - 1$ individual tests, which happens if the group tests positive and all but the last sample test negative. To get an intuition for Lemma 4, we unravel the effects of (i) and (ii) under various mixed groups.

Let $\lambda(S)$ denote the probability that group S requires one test with AG, i.e., the probability that group test result is negative. Similarly, let $\mu(S)$ denote the probability that group S requires $|S|$ tests with AG, i.e., the joint probability that the group test result is positive and the first $|S| - 1$ samples tested individually are negative. These probabilities are given by

$$\lambda(S) = (1 - p_\ell)^{|S|_\ell}(1 - p_h)^{|S|_h}, \tag{6}$$

and

$$\mu(S) = \begin{cases} p_\ell(1 - p_\ell)^{|S|-1} & \text{if } |S|_\ell = |S|, \\ p_h(1 - p_\ell)^{|S|_\ell}(1 - p_h)^{|S|_h - 1} & \text{if } |S|_\ell < |S|. \end{cases} \tag{7}$$

[14] Recall that we identify a group with the enumeration of the samples in it. By Lemma 1, the only relevant groups are those that have a high risk sample last in the enumeration. It is straightforward to see that any two groups that are permutations of each other except for the last sample are equivalent to each other as long as their last samples have the same risk types. Therefore, it is without loss to restrict attention to groups which test all ℓ-types first before testing the remaining h-types.

Then, the average tests per sample (recall (2)) is given by

$$\phi_S = \frac{1}{|S|}\left[\lambda(S) + \mu(S)|S| + (1 - \lambda(S) - \mu(S))\,(|S| + 1)\right]. \tag{8}$$

An increase in $\lambda(S)$ or $\mu(S)$ decreases ϕ_S. How do $\lambda(S)$ and $\mu(S)$ vary with the composition of S? As expected, $\lambda(S)$ is increasing in $|S|_h$: a higher number of high-risk samples translates into a higher probability of the initial group test being positive. This is illustrated in Fig. 2.

The probability of avoiding the last test is decreasing in $|S|_h$, except at $|S|_h = 1$, a non-monotonicity that we will explain in detail later. This is intuitive. The last sample is tested only when at least one of the non-last samples tests positive. When the probability of any non-last sample being infected is higher, it is more likely that this sample will test positive. This is illustrated in Fig. 3.

Suppose now contrary to the statement of Lemma 4 that there exists some mixed group S with $f_S > 0$ and $|S| = j$ in the optimal testing schedule that contains $k > 1$ high-risk samples. We show that there exists a strictly profitable restructuring of the S-groups into two groups S' and S'' of equal cardinality, one containing only one high-risk sample and the other containing only high-risk samples. Choose S' and S'' with $|S'| = |S''| = j$, $|S'|_h = 1$ and $|S''|_h = j$. To preserve the total fraction of high- and low-risk samples who were initially assigned to group S in the restructuring, we can split $j - 1$ groups of type S into $j - k$ groups of type S' and $k - 1$ groups of type S''. This restructuring achieves two things:

(i) it increases the probability of a negative initial group test result since $\lambda(S) < \frac{j-k}{j-1}\lambda(S') + \frac{k-1}{j-1}\lambda(S'')$, and

(ii) it increases the probability of requiring exactly $|S|$ tests since $\mu(S) < \frac{j-k}{j-1}\mu(S') + \frac{k-1}{j-1}\mu(S'')$.

Intuitively, (i) follows from the fact that λ is convex as a function of the number of high risk samples its argument S contains, and (ii) follows from the fact that the same is true for μ when S has at least one high-risk sample.

Together, both these effects lower the average tests per sample, as they both imply that the probability of requiring $|S| + 1$ tests decreases. This can be seen in (8) as $\phi(S)$ is decreasing in $\lambda(S)$ and $\mu(S)$. Essentially, whenever there is more than one high-risk sample in a group, there exists a strictly profitable restructuring that leads to (i) a lower probability of the initial group test being positive, and (ii) fewer tests conditional on a positive group test. This contradiction establishes Lemma 4.

The only mixed groups which are not ruled out by Lemma 4 in an optimal test schedule are the groups containing exactly one high-risk sample. When does a mixed group arise in an optimal testing schedule? The answer depends on the parameters because of two countervailing forces.

A necessary condition for the optimal testing schedule to contain a mixed group S with exactly one high-risk sample is that the S-groups cannot be profitably restructured into other groups. The next result shows that restructuring into homogeneous groups cannot be profitable when p_ℓ and p_h are sufficiently close.

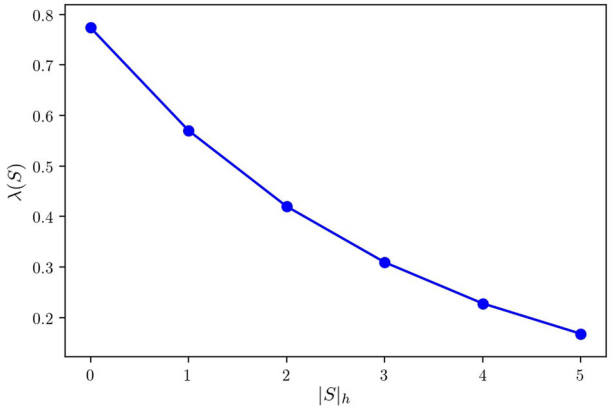

Fig. 2 Probability λ of a negative group test for $|S| = 5$, $p_\ell = 0.05$ and $p_h = 0.3$

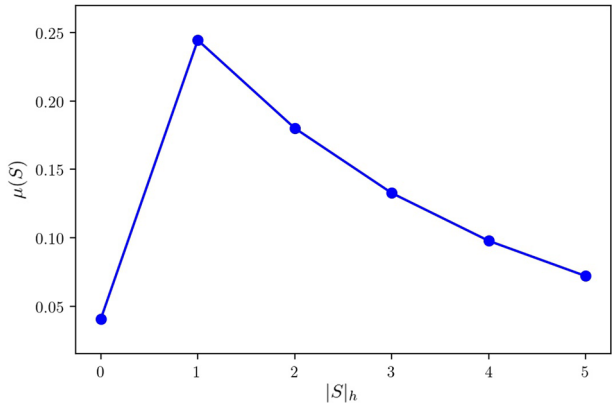

Fig. 3 Probability μ of requiring $|S|$ tests for $|S| = 5$, $p_\ell = 0.05$ and $p_h = 0.3$

Lemma 5 *Fix group size j. Then, for p_h sufficiently close to p_ℓ, testing a mixed group S^m with $|S^m| = j$ and $|S^m|_h = 1$ requires weakly fewer tests than restructuring the samples into homogeneous groups of size j.*

The restructuring of the mixed group of size j into homogeneous groups of size j each leads to two countervailing effects:

(i) it increases the average probability of a negative group test result,
(ii) it decreases the average probability of requiring exactly j tests and increases the average probability of requiring $j + 1$ tests.

The first effect is advantageous for the restructuring since a negative group test result requires only one test in total. The second effect results in a higher probability that the group requires $j + 1$ tests. This is disadvantageous for the restructuring. Lemma 5 says the second effect dominates when the risk types have sufficiently similar infection probabilities.

The first effect has been previously observed in the literature, e.g., by Lipnowski and Ravid (2021). Consider the extreme case of degenerate probabilities $p_\ell = 0$ and $p_h = 1$ and testing samples in pairs. In each mixed group $\{\ell h\}$, the probability of a positive group test is 1. Restructuring two mixed groups into two homogeneous groups $\{\ell\ell\}$ and $\{hh\}$ yields probabilities 0 and 1 of a positive group test respectively, and reduces by half the average probability of a positive group test.

Figure 3 illustrates the second effect. A mixed group containing only one high-risk sample has the highest probability of requiring exactly j tests among all groups of size j. Strikingly, the expected number of tests conditional on a positive group test is *higher* in a homogeneous group with low-risk samples than in a mixed group containing exactly one high-risk sample. Recall that μ is convex as a function of the number of high risk samples its argument S contains when S has at least one high-risk sample. It follows that the average probability of requiring $j+1$ tests per group after restructuring into homogeneous groups is higher than that of the mixed group (of size j with exactly one high-risk sample), since the number of high-risk samples in the mixed group is a convex combination of 0 and j which are the number of high-risk samples in the homogeneous groups after the restructuring. Hence, any restructuring (in our case into two homogeneous groups) leads to a higher probability of requiring $j+1$ tests.

To see the benefit of mixed groups conditional on a positive group test, consider the extreme example with almost degenerate risk probabilities $p_h \approx 1$ and $p_\ell \approx 0$. If the mixed group $\{\ell\ell h\}$ tests positive, then in expectation the DM requires only approximately two additional tests. This is because by Lemma 1, the high-risk sample is tested last, and the probability that the two low-risk samples test negative (which avoids the final test) is almost one. If instead a homogeneous group $\{\ell\ell\ell\}$ tests positive, then the probability of avoiding the last test is approximately $1/3$ and the expected number of tests is higher than in the mixed group.[15] In a mixed group with one high-risk sample, a positive group result is disproportionately attributed to the high-risk rather than the low-risk samples, and hence, the probability of avoiding the final test is higher.

The following result establishes sufficient conditions for NAM to be optimal.

Lemma 6 (NAM optimal for small probabilities of infection) *Fix testing capacity n. There exists $\underline{p} > 0$ such that for all $p_h < \underline{p}$, the optimal testing schedule contains at most two types of groups, and*

(i) *maximizes the fraction f_S for the mixed group S with $|S| = n$ and $|S|_h = 1$,*
(ii) *tests the remaining homogeneous samples in groups at capacity n.*

This establishes that for any possible testing capacity, the optimal testing schedule is NAM when infection probabilities are sufficiently small. Thus, the fact that the optimal testing schedule in Sect. 3 contains mixed groups is not an artefact of a small test capacity, but a relevant for *any* testing capacity n.

Given a population, let t_{AG} be the average number of tests per sample in the optimal testing schedule under AG (that is, the value function of the optimization problem in

[15] The probability of avoiding the final test in this extreme case approaches $\lim_{p_\ell \to 0} \frac{p_\ell(1-p_\ell)^2}{1-(1-p_\ell)^3} = 1/3$. See the proof of Lemma A2 in the Appendix for further details.

Sect. 2.2), t_{AD} be the average number of tests per sample in the optimal testing schedule under AD when heterogeneity of risk is ignored and t'_{AD} be the average number of tests per sample in the optimal testing schedule under AD when heterogeneity of risk is incorporated. When heterogeneity is ignored, we find the optimal size of a group test under AD using the average probability of infection in the population and calculate the average number of tests per sample. When heterogeneity is incorporated (we denote this by AD′), we find the optimal sizes of group test for both the low-risk samples and the high-risk samples, calculate the corresponding average number of tests per sample, and then find the population average. Table 3 provides an example that compares the optimal testing schedule under AG, AD′ and AD as we vary p_h, with p_ℓ and α fixed.

It is worth noting that our protocol can lead to significant efficiency gains. For example, as Table 3 illustrates, when $p_\ell = 0.05$, $p_h = 0.3$ and $\alpha = 0.8$, the average number of tests required per sample is 0.515 under AG and 0.594 under AD that ignores heterogeneity of risk. Thus, our testing algorithm saves about 13.3% of tests. Note that under AD that incorporates heterogeneity of risk, the average number of tests required per sample is 0.539. Hence, of the 13.3% reduction of tests, about 9.3% is due to the incorporation of heterogeneity or risk and 4% is due to the modification in the algorithm.

5 Extensions and Discussion

5.1 More than two risk types

In this section, we consider an extension in which the samples consist of more than two heterogeneous risk types. Let $\{1, \ldots, m\}$ denote the set of risk types and let p_θ denote the infection probability for type θ. We assume without loss of generality that the set of types is enumerated so that $0 < p_1 < \ldots < p_m < 1$. Let q_θ denote the fraction of type θ samples in the population.

We start by considering the case in which testing capacity is 2. We first show that if $p_\theta > \min\{\frac{1-2p_1}{1-p_1}, \frac{1-p_1^2-p_1}{2(1-p_1)}\}$, then it is optimal to test all θ types individually in any optimal testing schedule.

Proposition 2 (Optimality of individual testing for high risk types) *If* $p_\theta > \min\{\frac{1-2p_1}{1-p_1}, \frac{1-p_1^2-p_1}{2(1-p_1)}\}$, *then* $f_\theta = q_\theta$.

Intuitively, if it is not optimal to group type θ with the lowest risk type, then it is not optimal to group it with any other type either.

In what follows, let κ denote the highest type with $p_\kappa < \hat{p}$ where $\hat{p}$ is as defined in Sect. 3. Notice that $\frac{1-2p_1}{1-p_1} = \frac{1-p_1^2-p_1}{2(1-p_1)}$ when $p_1 = \hat{p}$, i.e., the cutoff probability of infection for group testing in a homogeneous population (recall Lemma 2). Thus, p_θ satisfying the hypothesis of Proposition 2 must necessarily exceed $\hat{p}$.

When $\kappa = 2$, Proposition 2 implies that all types $\theta \in \{3, \ldots, m\}$ are tested individually and DM's problem reduces to optimal testing when attention is restricted to types 1 and 2. In particular, an analogous version of Proposition 1 continues to hold with ℓ-types and h-types in Proposition 1 replaced with type 1 and type 2, respectively.

Table 3 Optimal testing schedule and average number of tests per sample under AG and AD with $p_\ell = 0.05, \alpha = 0.8$

	AG	AD′	AD
$p_h = 0.1$	$f_{\ell\ell\ell h} = 0.8,\ f_{\ell\ell\ell\ell\ell} = 0.2, t_{AG} = 0.449$	$f_{\ell\ell\ell\ell\ell} = 0.8,\ f_{hhhh} = 0.2, t'_{AD} = 0.460$	Optimal group size= 5, $t_{AD} = 0.466$
$p_h = 0.2$	$f_{\ell\ell\ell\ell\ell} = 0.8,\ f_{hhh} = 0.2, t_{AG} = 0.49$	$f_{\ell\ell\ell\ell\ell} = 0.8,\ f_{hhh} = 0.2, t'_{AD} = 0.505$	Optimal group size= 4, $t_{AD} = 0.534$
$p_h = 0.3$	$f_{\ell\ell\ell\ell\ell} = 0.8,\ f_{hh} = 0.2, t_{AG} = 0.515$	$f_{\ell\ell\ell\ell\ell} = 0.8,\ f_{hhh} = 0.2, t'_{AD} = 0.539$	Optimal group size= 4, $t_{AD} = 0.594$

By Lemma 2, AG is the optimal algorithm for any homogeneous population $\theta \in \{1, \ldots, \kappa\}$ since $p_\theta \leq p_\kappa < \hat{p}$. Consequently, $f_\theta = 0$ for all $\theta \in \{1, \ldots, \kappa\}$ in any optimal testing schedule, since otherwise we can profitably regroup individuals into homogeneous groups. If $f_{\theta\theta} > 0$ and $f_{\theta'\theta'} > 0$ for some $\theta, \theta' \in \{1, \ldots, \kappa\}$, then regrouping two homogeneous groups $\{\theta'\theta\}$ and $\{\theta'\theta'\}$ into two heterogeneous groups $\{\theta\theta'\}$ decreases the number of required tests, i.e., $2\phi_{\theta\theta'} \leq \phi_{\theta\theta} + \phi_{\theta'\theta'}$. To see this, note that

$$2(1 + p + 2p' - p'p) \leq (1 + 3p - p^2) + (1 + 3p' - p'^2)$$

for any p and p' such that $1 \geq p \geq p' \geq 0$. By taking $p = \max\{p_\theta, p_{\theta'}\}$ and $p' = \min\{p_\theta, p_{\theta'}\}$, the left hand side of the above expression is the required number of tests after regrouping, and the right hand side is the required number of tests before regrouping. It follows that there can be at most one homogeneous group in an optimal schedule when the probability of infection is sufficiently small for all types:

Proposition 3 (No individual tests and at most one homogeneous group for low infection probabilities) *If $p_\kappa < \hat{p}$, then $f_\theta = 0$ for all $\theta \in \{1, \ldots, \kappa\}$. Furthermore, if $f_{\theta\theta} > 0$ for some $\theta \in \{1, \ldots, \kappa\}$, then $f_{\theta'\theta'} = 0$ for all $\theta' \neq \theta$.*

Since it is optimal to test all samples with types $\{1, \ldots, \kappa\}$ in groups, the feasibility constraints for the DM's problem in this case are given by $f_{\theta\theta'} \in [0, 1]$ for all θ, θ' and

$$\frac{1}{2} \sum_{\theta' \neq \theta} f_{\theta\theta'} + f_{\theta\theta} = q_\theta, \tag{9}$$

for all $\theta \in \{1 \ldots, \kappa\}$. If $\kappa \geq 2$ and $q_\theta > \frac{1}{2}$ for some $\theta \in \{1, \ldots, \kappa\}$, i.e., if there are at least two low-risk types and if there is a majority risk type among the low-risk types, then it is not possible to find a feasible solution to the system of equations given by (9) with $f_{\theta\theta} = 0$. In that case, the optimal testing schedule necessarily includes a homogeneous group. Otherwise, i.e., if $q_\theta \leq \frac{1}{2}$ for all θ, one can find a feasible solution with $f_{\theta\theta} = 0$ for all θ. In this case, it might be optimal to have no homogeneous groups.

What happens when the testing capacity is higher than 2? The next result shows that it is never optimal to test a mixed group of size three with three types present in it.

Proposition 4 (No mixed group with three risk levels) *If the testing capacity is $n > 2$, then $f_{ijk} = 0$ for any three distinct types i, j, k.*

5.2 Imperfect tests

So far we assumed that the tests are perfect. That is, if a sample is not infected, then the test result is negative with probability 1 and if a sample is infected, then the test is positive with probability 1. Realistically, tests are imperfect and have some errors. In

what follows, we compare the probability of errors under individual testing protocol AI and group testing protocol AG.

Suppose $k \geq 1$ is the number of samples in a test. Denote by $x(k)$ the false positive rate of the test (the conditional probability that the test result is positive given that all samples in the test are uninfected) and $y(k)$ the false negative rate (the conditional probability that the test result is negative given that at least one sample in the test is infected). Note that we assume here that $y(k)$ depends only on k and does not depend on how many samples are infected.[16]

Under the individual testing algorithm AI, for each sample s_i, the probability of a false positive is $x(1)$ and the probability of false negative is $y(1)$.

Now consider the group testing algorithm AG. Under AG, a sample is found to be false negative if it is infected but either the group test is negative or the group test is positive but the subsequent individual test on this sample is negative. A sample is found to be false positive if it is uninfected but the group test is positive and either (i) the subsequent individual test on this sample is also positive or (ii) no individual test on this sample was conducted but all other individual samples were tested negative and therefore this sample is inferred to be positive.

Proposition 5 *Suppose* $k \geq 2$ *and* $x(k), y(k), y(1) \in (0, 1)$. *The false positive rate is strictly lower for any sample, except for the last one, under group testing protocol AG than under individual testing protocol AI. The false negative rate is strictly higher for any sample under AG than AI if* $y(k) \geq y(1)$.

Under group testing, if the initial group test result is negative, then no more tests are conducted, implying a higher false negative rate for a sample under AG than AI if the false negative rate is increasing in the number of samples in a test. But at the same time, AG lowers the false positive rate for any sample other than the last one since to be found false positive, both the group test and the individual test have to be (false) positive. For the last sample, however, the false positive rate could be higher under AG since it can be found false positive when it is uninfected, but the group test is positive and all the individual tests on other samples are negative and therefore it is (mistakenly) inferred to be positive.

5.3 Correlated samples

Our paper has focused on heterogeneity in the probability of infection and for simplicity we have assumed independence of infection probability across samples. Another interesting dimension of heterogeneity may be correlation: For example, the infection probability of people from the same household or same workplace may be correlated but it is independent across households and workplaces. When conducting group testing, should we pool independent or correlated samples together? In what follows, we provide a simple example to illustrate that positive assortative matching (that is, pooling positively correlated samples together) is optimal. Suppose there are two samples

[16] Biology suggests that $y(k)$ may be increasing in k because of the dilution of positive samples in a group test, though this effect may be small if the number of samples in a group test is small. See, for example, Cherif et al. (2020).

with the joint distribution illustrated in Table 4 below (1 means infected and 0 means healthy).

Table 4 Joint distribution of infection

	1	0
1	$p^2 + \delta p(1-p)$	$(1-\delta)p(1-p)$
0	$(1-\delta)p(1-p)$	$(1-p)^2 + \delta p(1-p)$

Under AG, if both samples are healthy, one test is conducted; if only one sample is healthy, then with probability $\frac{1}{2}$, two tests are conducted, and with probability $\frac{1}{2}$, three tests are conducted; if both samples are infected, then three tests are conducted. Hence, the expected number of tests is given by

$$\begin{aligned} &3[p^2 + \delta p(1-p)] + \frac{5}{2}[(2(1-\delta)p(1-p)] + (1-p)^2 + \delta p(1-p) \\ &\quad = 3p^2 + (1-p)^2 + p(1-p)(5-\delta). \end{aligned}$$

Since the expected number of tests is strictly decreasing in δ, the correlation coefficient between the two samples, it is better to pool together positively-correlated samples than independent ones.[17]

5.4 Testing more than two times

The preceding analysis relied on testing each samples no more than twice. This, together with the assumption that each infected sample ought to be identified, reduces the feasible testing algorithms to IND and AG. Next, we generalize the previous analysis by allowing each sample to be tested at most three times.

Assumption 2 Each sample can be tested at most three times.

Just one additional test per sample substantially increases the admissible testing protocols, complicating the analysis. To keep the analysis tractable, we also make the following assumption:

Assumption 3 Retesting across groups is inadmissible.

This assumption says that if a sample has been group-tested, it cannot be retested in a group test with another sample from a different group.[18] This is practical to implement in laboratories where different groups of samples might be split up to be tested in different locations and reduces complexity by ruling out many retesting configurations.

[17] Augenblick et al. (2020) make a similar observation about how the expected number of tests decreases in the correlation among samples in a group test, but do not discuss the implications for how to sort samples when there is heterogeneity in correlation among samples in the population.

[18] For example, if group $\{\ell\ell h\}$ and another group $\{\ell\ell h\}$ both test positive in a group test, one may not test the two h-samples from the two different groups together in one follow-up group test. Instead, aside from individual testing, only group tests $\{\ell h\}$ or $\{\ell\ell\}$ within each group should be considered.

Table 5 Tests per group in ADG and AGG

Sample composition	Probability	ADG	AGG
s_1, s_2, s_3 healthy	$(1-p_1)(1-p_2)(1-p_3)$	1	1
s_1, s_2 healthy, s_3 infected	$(1-p_1)(1-p_2)p_3$	3	2
s_1, s_3 healthy, s_2 infected	$(1-p_1)p_2(1-p_3)$	4	4
s_2, s_3 healthy, s_1 infected	$p_1(1-p_2)(1-p_3)$	3	5
s_1, s_2 infected, s_3 healthy	$p_1p_2(1-p_3)$	5	5
s_1, s_3 infected, s_2 healthy	$p_1(1-p_2)p_3$	4	5
s_2, s_3 infected, s_1 healthy	$(1-p_1)p_2p_3$	4	4
s_1, s_2, s_3 infected	$p_1p_2p_3$	5	5

In what follows, we discuss the optimal testing algorithms when testing capacity is $n = 3$. In this environment with $n = 3$, in addition to IND and AG, there are two other algorithms to consider. Both of them start with the group test of three samples $\{s_1s_2s_3\}$ and stop if the group test is negative (all three samples are healthy). If the group test is positive, there are two protocols:

1. Dynamic Group-Testing Algorithm (ADG): Test s_1 individually. If s_1 is positive, test $\{s_2s_3\}$ with AG. If s_1 is negative, test s_2 individually. If s_2 is negative, stop (only s_3 infected). If s_2 is positive, test s_3 individually.
2. Twice Group-Testing Algorithm (AGG): Test $\{s_1s_2\}$ via AG. Test s_3 individually only if group $\{s_1s_2\}$ is positive (if $\{s_1s_2\}$ is negative, only s_3 is infected).

Table 5 shows the number of tests for three ordered samples with infection probabilities $\{p_1, p_2, p_3\}$ with ADG and AGG. The following expected number of tests required per sample in ADG and AGG follows from the table:

$$\phi^{ADG}(p_1, p_2, p_3) = \frac{1}{3}(1 + 2p_1 + 3p_2 - p_1p_2 + (2 - p_1)(1 - p_2)p_3), \tag{10}$$

$$\phi^{AGG}(p_1, p_2, p_3) = \frac{1}{3}(1 + 4p_1 + 3p_2 - 3p_1p_2 + (1 - p_1)(1 - p_2)p_3). \tag{11}$$

Homogeneous samples. With homogeneous samples and $n = 3$, AGG is not optimal for any infection probability $p \in [0, 1]$. To see this, suppose a homogeneous group of size three is tested positive. Then, group-testing two of these samples cannot be optimal: each sample is infected with probability at least $1/3$, so a group with two samples will be positive with probability above $5/9$. Individual testing of the two samples requires two tests, while grouping them together requires strictly more tests in expectation.

In a homogeneous population, for a sufficiently low infection probability p, individual testing is not optimal. Following a positive group test of three samples, suppose that the individual test of sample s_1 is positive. This, in turn, substantially reduces the probability that s_2 or s_3 are positive. Hence, for sufficiently low p, group-testing $\{s_2s_3\}$ with ADG will dominate testing them individually following a positive test of s_1. On the other hand, unless p is sufficiently small, grouping only two samples and

using AG might be the optimal testing schedule. The following result summarizes the optimal testing algorithms for different infection probabilities.

Lemma 7 *Let $n = 3$ be the testing capacity. Under Assumptions 2 and 3, there exist $\underline{p} < \overline{p}$ such that the optimal testing algorithm is*

1. *ADG if $p \leq \underline{p}$,*
2. *AG with group size 2 if $p \in [\underline{p}, \overline{p}]$,*
3. *IND if $p \geq \overline{p}$.*

The equations pinning down $\underline{p} \approx 0.26$ and $\overline{p} \approx 0.35$ can be found in the appendix. Some of the insights in Lemma 7 can be generalized beyond capacity $n = 3$. First, testing a sample in a group test more than once is optimal only if the belief about this sample being infected following previous tests does *not* increase too much (as discussed in the case of AGG). Second, the adaptive nature of the retesting configuration can be important, even under Assumption 3: for a certain parameter range, the optimal testing algorithm requires that the retesting configuration depends on the outcome of the previous tests (for example, in ADG, whether s_2 and s_3 are tested in a group or individually depends on the test outcome of s_1).

Heterogeneous samples. With a heterogeneous population, which mixed groups of size three can arise in an optimal testing schedule? For AG, in Lemmas 1 and 4 we have established that only $\{\ell\ell h\}$ is a viable candidate. For ADG and AGG, there are six candidates to consider for each algorithm, three with one h-type in a group, and three with two h-types.

For ADG, with the expected tests per sample given in (10), we have $\phi^{ADG}(p_\ell, p_\ell, p_h) < \phi^{ADG}(p_h, p_\ell, p_\ell) < \phi^{ADG}(p_\ell, p_h, p_\ell)$ for $0 < p_\ell < p_h < 1$. Hence, only $\{\ell\ell h\}$ can arise. Similarly, we have $\phi^{ADG}(p_h, p_\ell, p_h) < \phi^{ADG}(p_\ell, p_h, p_h) < \phi^{ADG}(p_h, p_h, p_\ell)$. Hence, $\{h\ell h\}$ is the unique candidate for ADG among the groups containing two h-types and one ℓ-type. For AGG, among all six mixed groups, only $\{\ell\ell h\}$ is a candidate. This is because $\phi^{AGG}(p_\ell, p_\ell, p_h) < \phi^{AGG}(p_\ell, p_h, p_\ell) < \phi^{AGG}(p_h, p_\ell, p_\ell)$ and $\phi^{AGG}(p_\ell, p_h, p_h) < \phi^{AGG}(p_h, p_\ell, p_h) < \phi^{AGG}(p_h, p_h, p_\ell)$, and $\phi^{AGG}(p_\ell, p_h, p_h) > \phi^{ADG}(p_h, p_\ell, p_h)$.

With four different mixed groups to consider, the set of possible initial groupings and retesting configurations to consider increases substantially relative to our main analysis which requires consideration of only one mixed group for any group size. A numerical exercise shows that none of the four mixed groups can be ruled out since each is optimal for some parameter range: For example, if $\alpha = 0.5$, $p_\ell = 0.03$ and $p_h = 0.1$, the optimal testing schedule consists of two mixed groups, each for one of the two new algorithms: $f^{AGG}_{\ell\ell h} = 0.5$ and $f^{ADG}_{h\ell h} = 0.5$. On the other hand, if $\alpha = 0.5$, $p_\ell = 0.08$ and $p_h = 0.1$, the optimal testing schedule only utilizes ADG: $f^{ADG}_{\ell\ell h} = 0.5$ and $f^{ADG}_{h\ell h} = 0.5$. Moreover, as shown in Lemma 7, AG can still be optimal.

Characterizing the optimal testing algorithm when samples can be tested more than three times is beyond the scope of our paper. In what follows, we briefly discuss the challenges and what insight we can still apply from our analysis.

Suppose the samples can be tested more than three times and consider a K-stage testing algorithm where at each stage, any retesting configuration (that is, the number

of subgroups, their sizes and their members) is allowed. The problem becomes very complicated very quickly as K increases because the number of possible retesting configurations grows rapidly. Black et al. (2015) investigate this problem computationally and they show that if the initial group test contains N samples and the samples are ordered by the probability of being positive in the retesting configurations, then a K-stage algorithm has approximately $(K-1)^{N-1}$ possible retesting configurations. For a large N, the computational time to find the optimal testing protocol is impractical even for a small K. When there are two risk types, the number of possible retesting configurations is smaller than $(K-1)^{N-1}$, but still very large for a large N.

Partly because of the large number of possible retesting configurations, relatively simple algorithms have been proposed and analyzed by researchers and used in practice. A well-known example is the halving algorithm, which involves successively splitting positive groups into two equal-sized halves. Positive groups are halved until all groups test negatively or until individual testing occurs. Black et al. (2012) show that ordering the samples according to risk probabilities when halving them into subgroups instead of treating them as homogeneous and halving them randomly reduces the number of expected tests. Our analysis shows that further reduction is possible from simple modifications of the algorithm. Following a positive group test and after halving the group into two subgroups, we should first test the subgroup that has the highest probability of a negative test. If it is negative, we can conclude that the other subgroup will test positive and we either halve it further or start individual testing, depending on the updated infection probabilities. When individual test of a mixed group occurs, we should test those with the low risk first and leave the high-risk sample last to minimize the expected number of tests.

5.5 Task assignment and team design

Our analysis of group testing with heterogeneous risks can be applied to other situations, for example, to task assignment and team design in principal-agent relationships. Specifically, suppose a principal, having heterogeneous priors about a set of agents of uncertain abilities, can assign tasks and form teams to learn about their abilities. We assume that the principal can learn each agent's ability by assigning an individual task, which is perfectly informative about the individual's ability, or a team task. The principal's goal is to learn agents' abilities using a minimal number of tasks.

Consider first the case in which all team tasks are disjunctive tasks, that is, the team output is a success if at least one team member is of high ability and a failure if all team members are of low ability. Thus, following a failure the principal can infer that all team members are of low ability but a success requires further learning to determine the team members' abilities. To cast this situation in our group-testing model, high ability corresponds to an individual being infected and p_i corresponds to the prior of agent i having high ability. We call i a member of the underdog subgroup and j a member of the favored subgroup if $p_i < p_j$. Applying the results of our paper, the optimal team assignment consists of at most two different team compositions, and at most one mixed team with exactly one member from the favored subgroup.

Next consider the case in which all team tasks are conjunctive tasks, that is, the team output is a failure if at least one team member is of low ability and a success if all team members are of high ability. Thus, following a success the principal can infer that all team members are of high ability but a failure requires further learning. As such, low ability corresponds to an individual being infected and p_i corresponds to the prior of agent i having low ability. In this case, we call i a member of the underdog subgroup and j a member of the favored subgroup if $p_i > p_j$. Again, the optimal team assignment consists of at most two different team compositions, and at most one mixed team with exactly one member now from the underdog subgroup.

In our analysis, we showed that group testing reduces the expected number of tests only when the probability of infection is sufficiently low. Hence, if members of the underdog subgroup have a high probability of having high ability, then only conjunctive tasks are suitable as team tasks. If members of the favored group has a low probability of having high ability, then only disjunctive tasks are suitable as team tasks. In each of these environments, if the principal can choose the type of team task, she would use conjunctive and disjunctive tasks, respectively.

A Omitted proofs

Proof of Lemma 1 Fix a mixed group S. The probability of a negative group test, $(1-p_\ell)^{|S|_\ell}(1-p_h)^{|S|_h}$, is independent of the order of individual tests following a positive group test result. Since

$$(1-p_\ell)^{|S|_\ell}(1-p_h)^{|S|_h-1}p_h > (1-p_\ell)^{|S|_\ell-1}(1-p_h)^{|S|_h}p_\ell \quad \Leftrightarrow \quad p_\ell < p_h,$$

the joint probability of a positive group test result and the first $|S|-1$ individual tests being negative in the second step is higher if a high risk sample is tested last following a positive group test result. Hence, the probability of requiring a total of $|S|$ tests instead of $|S|+1$ tests is strictly higher if a high-risk sample is tested last. □

Proof of Proposition 1 The following result summarizes when NAM is optimal in a testing schedule. Note that the highest fraction of samples that can be part of a $S=\{\ell h\}$ mixed group is twice the fraction of the minority risk type, $\min\{2\alpha, 2(1-\alpha)\}$.

Recall $\hat{p} = \frac{1}{2}(3-\sqrt{5})$.

Lemma A1 (i) *If $\hat{p} \le p_\ell < p_h$, then $f_{\ell h} = 0$ in any optimal testing schedule.*

(ii) *If $p_\ell < \hat{p} < p_h$, then $f_{\ell h} = \min\{2\alpha, 2(1-\alpha)\}$ in any optimal testing schedule when $p_h \le \frac{p_\ell^2+p_\ell-1}{2(p_\ell-1)}$, and $f_{\ell h} = 0$ when $p_h > \frac{p_\ell^2+p_\ell-1}{2(p_\ell-1)}$ in any optimal testing schedule.*

(iii) *If $p_\ell < p_h \le \hat{p}$, then $f_{\ell h} = \max\{2\alpha, 2(1-\alpha)\}$ in any optimal testing schedule.*

Proof of Lemma A1 By equation (2), the average number of tests per sample in a mixed group of size 2 is

$$\phi_{\{\ell h\}} = \frac{1}{2}(1+p_h+2p_\ell - p_h p_\ell).$$

(i): If $p_h > p_\ell \geq \hat{p}$, then $1 = \phi_{\{h\}} = \phi_{\{\ell\}} < \phi_{\{\ell h\}}$ implying that testing a mixed group is worse than individually testing the samples in it. Hence, $f_{\ell h} = 0$.

(ii): If $p_\ell < \hat{p} \leq p_h$, then by Lemma 2, it is optimal to test high-risk samples with AI and low-risk samples with AG. A mixed group cannot be optimal if we can find a profitable restructuring. Consider restructuring two mixed groups $\{\ell h\}$ into of one homogeneous group $\{\ell\ell\}$ and two individual high-risk groups $\{h\}$. This restructuring reduces the average number of tests per sample if and only if

$$2(1 + p_h + 2p_\ell - p_h p_\ell) > 1 + 1 + (1 + 3p_\ell - p_\ell^2),$$

which holds if and only if $p_h > \frac{p_\ell^2 + p_\ell - 1}{2(p_\ell - 1)}$.

(iii): If $p_\ell < p_h < \hat{p}$, then by Lemma 2, AG is optimal for any homogeneous population. Then, NAM is optimal if and only if restructuring two homogeneous groups $\{\ell\ell\}$ and $\{hh\}$ into two mixed groups $\{\ell h\}$ reduces the average number of tests per sample. This holds if and only if

$$2(1 + p_h + 2p_\ell - p_h p_\ell) \leq (1 + 3p_\ell - p_\ell^2) + (1 + 3p_h - p_h^2),$$

which is always satisfied.

□

Lemma A1 pins down the fraction of samples that belong to mixed groups $\{\ell h\}$ in any optimal testing schedule. Lemma 2 determines the remainder of the optimal testing schedule, as all samples that are not tested in $S = \{\ell h\}$ are tested in a homogeneous group or individually. □

Proof of Lemma 3 For $n = 1$, there is nothing to prove. Let $n \geq 2$. For simplicity of notation, in the rest of the proof let S_k be a homogeneous group containing k samples with risk probability p.

First, note that by (1), under AG, increasing the group size from k to $k + 1 \leq n$ strictly decreases the expected number of tests per sample iff

$$p(1 - p)^{k-1}(1 + k + k^2 - pk^2) < 1. \tag{A1}$$

Since $\lim_{p \to 0} p(1-p)^{k-1}(1+k+k^2 - pk^2) = 0$ for every k, it follows that for every $2 \leq k \leq n - 1$, there exists $\tilde{p}(k) > 0$ such that (A1) is satisfied for every $p < \tilde{p}(k)$.

Second, note that by (1) testing group S_n with AG requires a fewer expected number of tests per sample than testing each sample in it with AI whenever

$$(1 - p)^{n-1}(n - np + p) > 1. \tag{A2}$$

Since $\lim_{p \to 0}(1 - p)^{n-1}(n - np + p) = n$ and $n \geq 2$, there exists $\tilde{p}(1) > 0$ such that for all $p < \tilde{p}(1)$ the inequality (A2) always holds.

Finally, let $\tilde{p} = \min\{\tilde{p}(1), \tilde{p}(2), \ldots, \tilde{p}(n-1)\}$. Then, for every $p < \tilde{p}$, increasing the group size by one up until the capacity n is reached strictly decreases the expected number of tests per sample. □

Proof of Lemma 4 Suppose to the contrary that the optimal testing schedule contains a positive fraction of mixed groups S such that S contains $k > 1$ high-risk samples. Let $j = |S|$. We show that the restructuring $j - 1$ of these S groups into $j - k$ groups containing only one high-risk sample, and $k - 1$ homogeneous groups containing only high-risk samples strictly reduces the number of tests.

Let S' and S'' denote the size-j groups with $|S'|_h = 1$ and $|S''|_h = j$. It suffices to show that

$$\phi(S) > \frac{j-k}{j-1}\phi(S') + \frac{k-1}{j-1}\phi(S'').$$

By (8), this is satisfied whenever

$$\lambda(S) < \frac{j-k}{j-1}\lambda(S') + \frac{k-1}{j-1}\lambda(S'') \tag{A3}$$

and

$$\mu(S) < \frac{j-k}{j-1}\mu(S') + \frac{k-1}{j-1}\mu(S''). \tag{A4}$$

are satisfied. Note that (A3) is equivalent to

$$(1-p_\ell)^{j-k}(1-p_h)^{k-1} < \frac{j-k}{j-1}(1-p_\ell)^{j-1} + \frac{k-1}{j-1}(1-p_h)^{j-1}.$$

This holds by the weighted AM-GM equation. Likewise, (A4) is always satisfied. □

Proof of Lemma 5 Consider three different groups S^m, S^ℓ, S^h, each containing j samples, with $|S^m|_h = 1$, $|S^\ell|_h = 0$ and $|S^h|_h = j$. Note that j of the S^m groups can be restructured into $(j-1)$ of the S^ℓ groups and one S^h group. The original testing schedule with mixed groups yields a lower expected number of tests if

$$j\phi_{S^m} \le (j-1)\phi_{S^\ell} + \phi_{S^h}.$$

Plugging in equations (1) and (2) for the expected number of test per sample, the inequality simplifies to

$$(1-p_h)^{j-1}(p_h + j - p_h j) \le (1-p_\ell)^{j-1}(p_\ell + j - p_\ell j + j(p_\ell - p_h)(j-1)). \tag{A5}$$

This is satisfied with equality when $p_h = p_\ell$. Furthermore, the right hand side is increasing in p_ℓ when $p_h - p_\ell$ is sufficiently small. To see this, note that

$$\frac{\partial(1-p_\ell)^{j-1}(p_\ell + j - p_\ell j + j(p_\ell - p_h)(j-1))}{\partial p_\ell} = -(1-p_\ell)^{j-2}(j-1)(1-(p_h-p_\ell)(j-1)j).$$

This is negative whenever $(p_h - p_\ell) < \frac{1}{(j-1)j}$. Thus, the inequality (A5) is satisfied for p_h sufficiently close to p_ℓ. □

Lemma A2 *Conditional on a positive test result, the expected number of tests is lower in a mixed group with one high-risk sample than in that of a homogeneous group of same size.*

Proof Fix any group size $k \geq 2$. Conditional on a positive group test, the probability of avoiding the last test of a group $S = \{s_i\}_{i=1}^k$ with respective risk-types $\{p_i\}_{i=1}^k$ is

$$\begin{aligned}\Pr(\{s_i\}_{i=1}^{k-1} \text{ healthy}|\text{positive group test}) &= \frac{\Pr(\{s_i\}_{i=1}^{k-1} \text{ healthy and } s_k \text{ infected})}{\Pr(\text{positive group test})} \\ &= \frac{p_k \Pi_{i=1}^{k-1}(1-p_i)}{1-\Pi_{i=1}^{k}(1-p_i)}.\end{aligned}$$

Applying this equation, the probability of avoiding the last test in mixed group of size k with one high-risk sample is

$$\frac{p_h(1-p_\ell)^{k-1}}{1-(1-p_\ell)^{k-1}(1-p_h)},$$

and the probability of avoiding the last test in a homogeneous low-risk group of size k is

$$\frac{p_\ell(1-p_\ell)^{k-1}}{1-(1-p_\ell)^{k}}.$$

Since $p_\ell < p_h$, we have

$$\frac{p_h(1-p_\ell)^{k-1}}{1-(1-p_\ell)^{k-1}(1-p_h)} > \frac{p_\ell(1-p_\ell)^{k-1}}{1-(1-p_\ell)^{k}}.$$

Hence, a mixed group which tested positive is more likely to require $k-1$ instead of k additional tests than the homogeneous low-risk group. A similar argument shows that a mixed group which tested positive is more likely to require $k-1$ instead of k additional tests than the homogeneous high-risk group as well. □

Proof of Lemma 6 First, we establish that for p_h sufficiently small, a mixed group that is below capacity is never optimal.

Lemma A3 *Fix capacity n. There exists $p' > 0$ such that for all $p_h < p'$, $f_S = 0$ for all $S \in \mathcal{S}_{n-1}^m$ in any optimal testing schedule.*

Proof of Lemma A3 For simplicity of notation, let $S_{a,b}$ be a group containing a low-risk and b high-risk samples. We show that for p_h sufficiently small, it is strictly better for

the DM to restructure $k(k+1)$ number of $S_{k-1,1}$ groups into $(k-1)(k+1)$ number of $S_{k,1}$ groups and one $S_{0,k+1}$ groups, i.e., we show that

$$k(k+1)\phi_{S_{k-1,1}} > (k-1)(k+1)\phi_{S_{k,1}} + \phi_{S_{0,k+1}}.$$

From (2),

$$\phi_{S_{k,1}} = 1 - (1-p_\ell)^k(1-p_h) + \frac{1}{k+1}(1 - p_h(1-p_\ell)^k)$$

is decreasing in k when p_h is sufficiently low. Hence, $\phi_{S_{k-1,1}} > \phi_{S_{k,1}}$ for p_h sufficiently low. From (1) and (2), we have $\phi_{S_{k,1}} = \phi_{S_{0,k+1}}$ when $p_h = p_\ell$. Hence, when p_h is sufficiently low (which implies that it is sufficiently close to p_ℓ since $p_h > p_\ell$), we have $\phi_{S_{k-1,1}} > \phi_{S_{0,k+1}}$. Hence,

$$k(k+1)\phi_{S_{k-1,1}} > (k-1)(k+1)\phi_{S_{k,1}} + \phi_{S_{0,k+1}}$$

when p_h is sufficiently low.

Hence, for every $k \leq n-1$, there exists $p(k)$ such that for all $p_h < p(k)$, the mixed group $S_{k-1,1}$ can be restructured into two groups, each containing one more sample and each requiring fewer expected number of tests per sample. Let $p' := \min\{p(2), p(3), \ldots, p(n-1)\}$. Then, for all $p_h < p'$, a mixed group of capacity lower than n requires more tests than restructuring it into a mixed group and a homogeneous group at capacity n. □

Next, we combine Lemmas 3, 5 and A3 to establish the proof of Lemma 6. Fix capacity n. Let

$$p'' = \frac{1}{(n-1)n}. \tag{A6}$$

If $p_h < p''$, then using the proof of Lemma 5, $p_h - p_\ell$ is sufficiently close to p_ℓ for any $p_\ell < p_h$ such that forming mixed groups containing one high-risk sample is better than homogeneous groups of the same size for any group size $k \leq n$.

Define a lower bound $\underline{p} := \min\{\tilde{p}, p', p''\}$ where the $\tilde{p}$ and p' are given by Lemmas 3 and A3 respectively, and p'' is given by (A6). Then, for all $p_h < \underline{p}$, in any optimal testing schedule,

(i) homogeneous samples are tested at testing capacity n, i.e., $f_S = 0$ for all $S \in \mathcal{S}^\ell_{n-1} \cup \mathcal{S}^h_{n-1}$ (Lemma 3),
(ii) mixed groups are formed only at capacity n, i.e., $f_S = 0$ for all $S \in \mathcal{S}^m_{n-1}$ (Lemma A3),
(iii) the fraction f_S for the mixed group S at capacity n with $|S|_\ell = n-1$ and $|S|_h = 1$ is maximized. (Lemma 5).

Hence, for p_h sufficiently low, the optimal testing schedule is NAM and contains at most two groups: a mixed group at capacity n with exactly one high-risk sample, and a homogeneous group at capacity of the type that cannot be feasibly tested in mixed groups. □

Proof of Proposition 2 The expected number of tests in a group with type 1 and type θ is given by

$$-p_1 p_\theta + 2p_1 + p_\theta + 1.$$

Restructuring a $\{1\,\theta\}$ group into two individual tests results in strictly lower expected number of tests when

$$-p_1 p_\theta + 2p_1 + p_\theta + 1 > 2$$

i.e., when $p_\theta > \frac{1-2p_1}{1-p_1}$. Since the expected number of tests in any group $\{\theta'\,\theta\}$ for $\theta' \in \{2, \ldots, m\}$ is higher than in $\{1\,\theta\}$, it is strictly profitable to restructure these groups into two individual tests as well if $p_\theta > \frac{1-2p_1}{1-p_1}$.

If $p_1 < \hat{p}$, then $\frac{1-p_1^2-p_1}{2(1-p_1)} < \frac{1-2p_1}{1-p_1}$. In that case, it is more profitable to restructure two $\{1\,\theta\}$ group tests into a $\{1\,1\}$ group test and two individual θ tests. Again, the expected number of tests is lower with this restructuring. A similar argument shows that it is also never optimal to test a θ type in a group with any other type θ' such that $\theta' < \hat{p}$. □

Proof of Proposition 4 Without loss of generality assume $i < j < k$. It suffices to show that the regrouping of two $\{ijk\}$ groups into an $\{iik\}$ group and an $\{jjk\}$ group reduces the expected number of tests. A straightforward extension of the AG algorithm and Lemma 1 shows that given any three samples $\theta, \theta', \theta''$, it is optimal to test the lowest risk sample first and highest risk sample last in any group test under AG. Thus, the expected number of tests for an $\{ijk\}$ group is given by

$$\begin{aligned} 3\phi_{ijk} &= (1-p_i)(1-p_j)(1-p_k) + 3(1-p_i)(1-p_j)p_k \\ &\quad + 4[1-(1-p_i)(1-p_j)(1-p_k) - (1-p_i)(1-p_j)p_k] \\ &= 1 + 3p_i + 3p_j - 3p_i p_j + 2(1-p_i)(1-p_j)p_k. \end{aligned} \tag{A7}$$

In the expression above, $(1-p_i)(1-p_i)(1-p_k)$ is the probability that no sample is infected, in which case only one test is needed, $(1-p_i)(1-p_j)p_k$ is the probability that only sample k is infected, in which case three tests are needed, and

$$\big(1-(1-p_i)(1-p_j)(1-p_k) - (1-p_i)(1-p_j)p_k\big)$$

is the probability that sample i or sample j are infected, in which case four test are needed. The required number of tests per sample in each of the $\{\theta\theta k\}$ groups for $\theta = i, j$ is given in (2). Tedious but straightforward algebra shows that $2\phi_{ijk} > \phi_{iik} + \phi_{jjk}$, which suffices to establish the proposition. □

Proof of Proposition 5 For sample s_i where $i \in \{1, \ldots, k-1\}$, the probability of a false positive is

$$\left[\prod_{j=1, j\neq i}^{k} (1-p_j)x(k)x(1) + (1 - \prod_{j=1, j\neq i}^{k} (1-p_j))(1-y(k))x(1)\right],$$

Since $x(k) < 1$ and $1 - y(k) < 1$, this is strictly lower than $x(1)$, the false positive rate under AI.

For sample s_i where $i \in \{1, \ldots, k-1\}$, the probability of a false negative is $y(k) + (1 - y(k))y(1) > y(k) \geq y(1)$. For sample s_k, the probability of a false negative is $y(k) + (1 - y(k))zy(1)$ where $z > 0$ is the probability that all the $(k-1)$ samples other than s_k test negative on their individual tests following the positive group test. Since $y(k) + (1 - y(k))zy(1) > y(k) \geq y(1)$, it follows that any sample s_i where $i \in \{1, \ldots, k\}$ has a strictly higher false negative rate under AG than AI. □

Proof of Lemma 7 Simplifying equations (11) and (10) for $(p_1, p_2, p_3) = (p, p, p)$ yields $\phi^{AGG}(p, p, p) = 1/3(1+p(8+(-5+p)p))$ and $\phi^{ADG} = 1/3(1+p(7+(-4+p)p))$. Comparing this to the number of tests per sample in IND and AG for capacity 2 and 3 yields the result. In particular, $\overline{p}$ is pinned down by $1/3(2-p(7+(-4+p)p)) = 0$ (when $\phi^{ADG}(p, p, p) = 1 = \phi^{IND}$) and $\underline{p}$ by $1/6(1 + p(-5 + (5 - 2p)p)) = 0$ (when $\phi^{ADG}(p, p, p) = \phi^{AG}(p, p)$). □

References

Aprahamian, H., Bish, D.R., Bish, E.K.: Residual risk and waste in donated blood with pooled nucleic acid testing. Stat. Med. **35**(28), 5283–5301 (2016)

Aprahamian, H., Bish, D.R., Bish, E.K.: Optimal risk-based group testing. Manag. Sci. **65**(9), 4365–4384 (2019)

Augenblick, N., Kolstad, J.T., Obermeyer, Z., Wang, A.: Group testing in a pandemic: The role of frequent testing, correlated risk, and machine learning. Technical report, National Bureau of Economic Research (2020)

Black, M.S., Bilder, C.R., Tebbs, J.M.: Group testing in heterogeneous populations by using halving algorithms. J. R. Stat. Soc. Ser. C (Appl. Stat.) **61**(2), 277–290 (2012)

Black, M.S., Bilder, C.R., Tebbs, J.M.: Optimal retesting configurations for hierarchical group testing. J. R. Stat. Soc. Ser. C (Appl. Stat.) **64**(4), 693–710 (2015)

Cherif, A., Grobe, N., Wang, X., Kotanko, P.: Simulation of pool testing to identify patients with coronavirus disease 2019 under conditions of limited test availability. JAMA Netw. Open **3**(6), e2013075–e2013075 (2020)

Deb, R., Pai, M., Vohra, A., Vohra, R.: Testing alone is insufficient. Rev. Econ. Design **26**(1), 1–21 (2022)

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

Du, D., Hwang, F.K.: Combinatorial Group Testing and Its Applications, vol. 12. World Scientific, Singapore (2000)

Ely, J., Galeotti, A., Jann, O., Steiner, J.: Optimal test allocation. J. Econ. Theory **193**, 105236 (2021). https://doi.org/10.1016/j.jet.2021.105236

Finucan, H.: The blood testing problem. J. R. Stat. Soc. Ser. C (Appl. Stat.) **13**(1), 43–50 (1964)

Gollier, C., Gossner, O.: Group testing against covid-19. Covid Econ. **2** (2020)

Hwang, F.: A generalized binomial group testing problem. J. Am. Stat. Assoc. **70**(352), 923–926 (1975)

Lipnowski, E., Ravid, D.: Pooled testing for quarantine decisions. J. Econ. Theory **198**, 105372 (2021). https://doi.org/10.1016/j.jet.2021.105372

Makris, M.: Covid and social distancing with a heterogenous population. Econ. Theory (2021). https://doi.org/10.1007/s00199-021-01377-2

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

Vohra, R.V.: Mechanism Design: A Linear Programming Approach, vol. 47. Cambridge University Press, Cambridge (2011)