Pooled Testing Efficiency Increases with Test Frequency

Ned Augenblick[a], **Jonathan Kolstad**[b], **Ziad Obermeyer**[c], **and Ao Wang**[d]

[a]Haas School of Business, University of California, Berkeley, 2220 Piedmont Ave, Berkeley, CA 94720; [b]Haas School of Business and Department of Economics, University of California, Berkeley, 2220 Piedmont Ave, Berkeley, CA 94720; [c]School of Public Health, University of California, Berkeley, 2121 Berkeley Way, Berkeley, CA 94704; [d]Department of Economics, University of California, Berkeley, 530 Evans Hall #3880, Berkeley, California 94720

This manuscript was compiled on November 9, 2021

Pooled testing increases efficiency by grouping individual samples and testing the combined sample, such that many individuals can be cleared with one negative test. This short paper demonstrates that pooled testing is particularly advantageous in the setting of pandemics, given repeated testing, rapid spread, and uncertain risk. Repeated testing mechanically lowers the infection probability at the time of the next test by removing positives from the population. This effect alone means that increasing frequency by x times only increases expected tests by around $\sqrt{x}$. However, this calculation omits a further benefit of frequent testing: removing infections from the population lowers intra-group transmission, which lowers infection probability and generates further efficiency. For this reason, increasing testing frequency can paradoxically *reduce* total testing cost. Our calculations are based on the assumption that infection rates are known, but predicting these rates is challenging in a fast-moving pandemic. However, given that frequent testing naturally suppresses the mean and variance of infection rates, we show that our results are very robust to uncertainty and misprediction. Finally, we note that efficiency further increases given natural sampling pools (e.g., workplaces, classrooms, etc.) that induce correlated risk via local transmission. We conclude that frequent pooled testing using natural groupings is a cost-effective way to provide consistent testing of a population to suppress infection risk in a pandemic.

Pooled Testing | COVID-19 | Surveillance Testing

The COVID-19 pandemic has generated a health and economic crisis not seen in more than a century. Opening businesses and schools is necessary to regain economic activity, but the potential public health costs are dramatic. One policy to circumvent this stark trade-off is to open the economy, while implementing *surveillance testing* that can quickly identify infected individuals – particularly those without symptoms – and prevent them from spreading the disease. Unfortunately, testing at this scale appears infeasible given the cost and capacity constraints. This paper makes a simple but essential point about these costs: when using pooling testing, frequent testing of correlated samples makes testing dramatically more efficient (and therefore less costly) than understood both by existing research and policymakers.

In pooled testing (1), multiple samples are combined, tested together using one test, and the entire pool is cleared given a negative test result. Pooling is an old concept and a large literature has emerged on optimal strategies (1–10) and, more recently, others have discussed how it might be used to increase COVID-19 test efficiency (11, 12). However, all of these papers focus on one-time testing of a set of samples with known and independent infection risk, which matches common use cases such as screening donated blood for infectious diseases (13–18). These environmental assumptions are violated when dealing with a novel pandemic with rapid spread. In this case, people need to be tested multiple times, testing pools are likely formed from populations with correlated infection risk, and risk levels at any time are very uncertain. How do these changes impact testing strategy?

We start with the well-known observation that pooled testing is more efficient when the infection probability is lower, because the likelihood of a negative pooled test is increased. This observation has been used to conclude that pooled testing is not cost-effective for "high-risk" populations, such as health care workers or for people in areas experiencing an outbreak. While this statement is true for one-off testing, it does not hold when the population is tested repeatedly. As an extreme example, if a person in a high-risk area was just tested and determined to be negative, their probability of infection when tested an hour later is extremely low, simply because there is not much time to be infected between the tests. In other words, the infection probability at the time of testing depends both on the flow rate of infection as well as the timing of testing.

We quantify the impact of testing frequency on infection probability and its consequent impact on pooled-testing efficiency. For example, we show that given reasonable levels of independent risk, testing twice as often cuts the infection probability at the time of testing by (about) half, which lowers the expected number of tests at each testing round to about 70% of the original number. The savings are akin to a "quantity discount" of 30% in the cost of testing. Therefore, rather

Significance Statement

Frequent mass testing can slow a rapidly-spreading infectious disease by quickly identifying and isolating infected individuals from the population. One proposed method to reduce the extremely high costs of this testing strategy is to employ pooled testing, in which samples are combined, tested together using one test, and the entire pool is cleared given a negative test result. This paper demonstrates that frequent pooled testing of individuals with correlated risk – even given large uncertainty about infection rates — is particularly efficient. We conclude that frequent pooled testing using natural groupings is a cost-effective way to suppress infection risk in a pandemic.

Please provide details of author contributions here.

The authors have ownership stake in Berkeley Data Ventures, which has provided advice on COVID-19 testing strategies.

[1]Authors contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: augenblick@berkeley.edu

than requiring two times the number of tests, doubling the frequency only increases costs by a factor of 1.4. More generally, we demonstrate that testing more frequently requires fewer tests than might be naively expected: increasing frequency by x times only uses about $\sqrt{x}$ as many tests, implying a quantity discount of $(1 - \frac{1}{\sqrt{x}})$.

The benefits to frequency are even greater when the disease spreads within the testing population. In this case, testing more frequently has an additional benefit: by quickly removing infected individuals, infection spread is contained, future infection probabilities are lowered, and testing efficiency rises further. We analytically quantify this additional benefit as a function of the exponential-like growth path of the disease. We show that in this case – somewhat paradoxically – the quantity discount can be so great that more frequent testing can actually *reduce* the total number of tests. For example, if the disease dynamics are such that doubling the testing frequency reduces the infection probability at the time of testing by more than fourfold, then doubling the frequency will require fewer tests in expectation.

In our simple model, we assume that infection probabilities are known when constructing optimal pool sizes and efficiency statistics. However, the prediction of infections in a fast-changing pandemic is an extremely difficult inference problem (see, for example, (19)). Given this issue, it is appropriate to worry that uncertainty and potential misprediction will make pool size choices challenging, reduce pooled testing efficiency, and render our conclusions void. For example, testing data from Massachusetts in the fall of 2020 shows high average testing positivity rates (7%) that vary widely across time and space (s.d. of 6%) in potentially unpredictable ways.[*] Using *one-off* pooled testing given this population — which has a extremely high positivity rate partially due to *self-selection* of people who desire a test — will be very inefficient given the high rates and the potential for mis-optimization. However, as discussed above, *frequent* testing of a *consistent* population reduces the mean and variance of infection probabilities at the time of testing because there is little time between testing for mean- and variance-inducing spread to occur, and the selection issue is removed. For example, as noted in (20), the town of Wellesley, Massachusetts, employed weekly testing of consistent subpopulations in the fall of 2020 and the average positivity rates stayed low (0.3%) and didn't vary considerably (s.d. of 0.3%). When positivity rates have low mean and variance, we show that the efficiency of pooled testing is strongly robust to reasonably-miscalibrated estimations and constant pool sizes, such that pooled testing remains very attractive. Finally, we note that better estimation of the positivity distribution is also helped by frequent testing, which naturally produces a constant stream of recent test-result data from the relevant population.

We note one final efficiency benefit associated with the most natural implementation of frequent testing. When frequently testing a consistent subpopulation (such as those living or working together), it is likely that the infection spreads within the sub-population. This correlation increases the benefits of pooled testing even in a static testing environment (a finding concurrently noted in (21)). Intuitively, an increased correlation in a pool with fixed individual risk lowers the likelihood of a positive pooled test result, which increases efficiency.

Throughout the paper, we consider a very stylized environment with a number of simplifications to present transparent results. While removing these constraints further complicates the problem and raises a number of important logistical questions, we do not believe that their inclusion changes our main insights. For example, our simple model assumes that a person who becomes infected will test positive indefinitely, whereas in reality they will potentially recover at some point. This does not impact our results when the time between tests is less than the recovery period, but lowers the relative cost of pooled testing when frequency is low because the prevalence is lower due to recoveries. However, our main qualitative conclusion — testing more frequently leads to fewer tests for each testing period — still holds in this case.

Another important simplification is that we model a test with perfect sensitivity.[†] There are multiple ways in which pooled testing interacts with test sensitivity. First, there is a natural negative impact: combining samples can potentially dilute the viral load below the limit-of-detection of the test. However, this implies that the false negatives will occur when the viral load is very low and the person is less likely to be infectious[‡]. Second, this dilution concern is counteracted when testing frequently by the large increase in overall sensitivity coming from running a larger number of tests.[§] Third, as noted in (22), false negatives may result from poor-quality samples. However, frequency again has benefits: by testing the same population repeatedly, subjects become better experienced with proper sampling protocols and those who provide poor samples can be identified and corrected.

Finally, we largely abstract away various practical implementation costs and constraints. First, we assume that every test, whether individual or pooled, has the same cost. However, pooled testing necessitates a more complicated setup in the lab, requiring more space and trained personnel (or a robotic setup) to correctly mix the samples together. While these costs are relatively moderate if spread over a long period of time, a laboratory might be reluctant to change their operations when the duration of the pandemic is very unclear. Second, we assume that there is no time delay between testing and receiving the test result. In reality, it takes time to transport samples to the lab and test them, and pooled testing takes more time than individual testing because it potentially requires an additional re-testing step. Fortunately, the difference in these delays can be minimized when using the common "hold-out" method: only a portion of each individual sample is used to construct the pooled sample, such that remaining portions of the individual samples can be immediately individually tested if the pooled tests positive. However, even if the difference is minimized, any delay still impacts our analysis. In particular, by assuming no delay, increasing the testing frequency minimizes the likelihood of undiscovered new infections in the time between tests, such that the infection probability at the

[*] This data is publicly available at: https://www.mass.gov/info-details/covid-19-response-reporting

[†] As noted in (22), test specificity of standard protocols such as PCR appears to be very close to one. However, if specificity is a concern, the past literature (9, 23) has clear methods to optimize in the case of imperfect tests.

[‡] Furthermore, empirically, the sensitivity loss of pooled testing given reasonable pool sizes has been shown to be negligible in other domains (24, 25) and more recently shown to be similarly low for SARS-CoV-2 in pool sizes of 32 and 48 (26, 27), although the results of (28) shows lower specificity (81%) for pools of 50.

[§] For example, if pooled testing leads the sensitivity to drop from 99% to 90% on a single test, sampling x times as frequently will increase overall sensitivity to $1-(0.10)^x$ if errors are independent. Even with extreme correlation in the error – suppose the false negative rate for a pool given a previous false negative for that pool is 50% – pooled testing 4-5 times as frequently will recover the same false positive rate as individual testing.

time of testing can be kept arbitrarily low. But, when there is a delay in receiving test results, it is not possible to stop infection and spread during the delay period even if testing is continuous. Therefore, it might be simply impossible to lower the infection probability below the ∼5% threshold at which the cost benefit of pooled testing is considered clear. In this extreme case, we do not recommend pooled testing. However, if the risk and spread are so extreme that 5% of a group is expected to be newly infected every few days even with very frequent testing, an alternative policy relying on isolation seems far more likely.

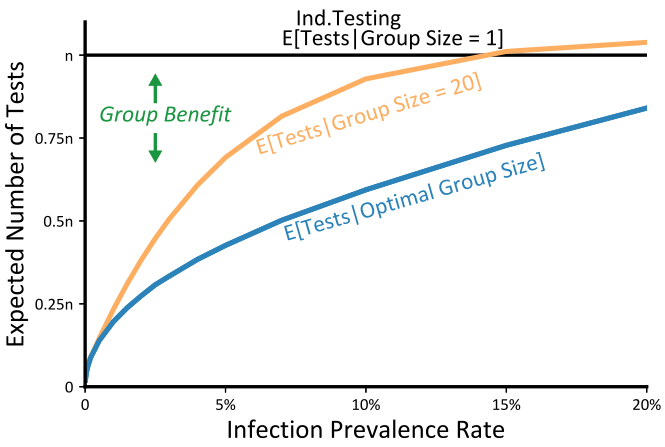
Although we see this paper as noting a general insight of the relationship between pooled testing and testing frequency, it is useful to discuss the particular historical context in which the paper was written. The first paper draft of the paper was completed in June 2020, during the first wave of the COVID-19 pandemic. At that point, testing supply was low and prices were high because laboratories were building up testing capacity in a relatively strict regulatory environment. By early 2021, multiple organizations – such as Mirimus, Ginkgo, and the Broad – were offering frequent pooled testing at much cheaper prices than individual testing and multiple organizations with correlated risk – such as employers, cities, and school districts – were employing these tests. For example, in February 2021, Massachusetts implemented a policy of providing universal weekly pooled testing for all K-12 students and faculty and staff.[¶] and, nationally, the Rockefeller Foundation called for use of frequent pooled testing as an essential aspect of school reopening (30).[‖] The authors, based on the main insights of this paper, supported many of these policy initiatives and recommendations. Interestingly, the cost of pooled testing in Massachusetts (between $3 and $10 per student per test) is almost precisely the predicted amount using pooling in the first draft of the paper, providing a useful empirical validation of the model.

The paper proceeds as follows: Section 2 reviews one important finding in the pooled testing literature that efficiency rises as infection probability falls; Section 3 discusses the relationship between testing frequency and efficiency; Section 4 demonstrates how correlated infection leads to larger pool sizes and greater efficiency; and Section 5 concludes.

Pooled Testing: Benefits Rise as Infection Probability Falls

Background on Pooled Testing. To understand the basic benefits of pooled testing, consider a simple example: 100 people, each with an independent likelihood of being positive of 1% and a test that (perfectly) determines if a sample is positive. The conventional approach of testing each person individually requires 100 tests. Suppose instead that the individuals' samples are combined into five equally-sized pools of 20, and then each of these combined samples is tested using one test. If any one of the 20 individuals in a combined sample is positive, everyone in that pool is individually tested, requiring 20 more tests (21 in total). The probability that this occurs is $1-(1-.01)^{20} \approx 18\%$. However, if no one in the pool is positive – which occurs with probability $\approx 82\%$ – no more testing is required. Because the majority of tests require no testing in the second case, the expected number of tests for this simple pooling method is only around 23, a significant improvement over the 100 tests required in the non-pooled method.

Fig. 1. Efficiency of pooled testing rises with infection probability



Notes: This figure plots the expected number of tests (y-axis) from pooled testing given a population of n people as the population infection probability (x-axis) changes. The black flat line shows the number of tests from individual testing (equivalent to a pool size of 1), which always requires n tests regardless of infection probability. The results from using a pool size of 20 is orange, while the blue line represents the number of tests given the optimal pool size for a given probability. Finally, the green text notes that benefit from pooled testing is the distance between the black individual-testing line and those from pooled testing. For example, as noted in the text, using a pool size of 20 for a probability of 1% leads to $.23 \cdot n$ tests rather than n tests, while the optimal pool size (10) leads to $.20 \cdot n$ tests.

The approach is well studied, with a large literature focused on improving the efficiency of pooled testing. These include using the optimal pool size (e.g. in this example, the optimal pool size of 10 would lower the expected number of tests to around 20), placing people into multiple pools (31), and allowing for multiple stages of pooled testing (2, 8, 23, 32, 33). There are also methods to deal with complications, such as incorporating continuous outcomes (34). Any of these modifications can be incorporated in our pooled testing strategy.

For clarity of exposition, we present results for simple two-stage "Dorfman" testing – in which every person in a positive pool is tested individually – to demonstrate that our conclusions are not driven by highly complex poolings and to make our calculations transparent, although we advocate for more sophisticated strategies when feasible. As an example of this transparency, while the optimal pool size and associated efficiency formulas under Dorfman testing are complicated, approximations around infection probability $p = 0$ are very simple and accurate at the low probabilities needed for pooled testing. Specifically, given a relatively-low infection probability

[¶] See e.g. https://covidedtesting.com which provides a detailed description of the approach to pooled testing in schools and the experience in Massachusetts. (29) covers the role the program can play in Massachusetts and as a model for the U.S.

[‖] This report notes that, even with vaccination, testing will continue to be a key policy lever. In wealthy countries, the potential for low vaccination take-up and much slower vaccination approval for children suggests that schools and possibly many employers will continue to need surveillance testing. Furthermore, many low income countries will not achieve large-scale vaccination for multiple years.

p, the approximate optimal pool size is:

$$g^* \approx \frac{1}{2} + \frac{1}{\sqrt{p}} \quad [1]$$

and a good approximation of the expected number of tests given a population of n people is:

$$E[tests^*] \approx 2 \cdot \sqrt{p} \cdot n. \quad [2]$$

In this paper, we create simple statements about the impact of increasing frequency that are approximately correct for low p. Note that this does not imply that our results are only appropriate for low-risk populations. The amount of infection at the time of testing depends on the testing frequency. Therefore, for the same population (even a high-risk population), p will be high when considering testing every month, but much lower when testing every day. By focusing on situations in which p is relatively low, we are not focusing on low-risk populations, but rather focusing on frequencies in which p remains relatively low for the given population. That is, for a high-risk population, our approximation formulas will be reasonably accurate when comparing the benefits of testing once vs. twice a week, but less accurate when comparing the benefits of testing once vs. twice a month. In other words, our focus on low levels of p is not an assumption about the population risk level, but an assumption that we are only comparing frequencies in which infection does not spread out of control in a given population.

How good is the approximation? For the magnitude of infection probabilities we discuss in the paper, such as 2%, 1%, or .1%, the approximation of the optimal pool size is within 0.3%, 0.1%, 0.01%, of the true optimal, respectively, and the approximation of the number of tests is within 3.1%, 2.3%, and 0.7% of the true number. However, given that there are multiple possible formulas in the literature, it is also useful to discuss the origin of our formulas. The formula we use for the pool size is from (35), who uses Taylor expansion to create an approximation around $p = 0$ given that pool size is continuous. However, he also notes that $ceiling(\sqrt{\frac{1}{p}})$ is a better approximation if pool size is constrained to be an integer. Meanwhile, (36) notes that the *exact* solution is either 1 or $1 + floor(\sqrt{\frac{1}{p}})$ or $2 + floor(\sqrt{\frac{1}{p}})$ depending on p. Similarly, for expected tests, the low-order Taylor approximation (also from (35)) is $2 \cdot \sqrt{p} \cdot n - \frac{p}{2} \cdot n$. Here, we only use the term $2 \cdot \sqrt{p} \cdot n$ as it creates simpler formulas with little loss of accuracy in the approximation.

Infection Probability and Pooled Testing. As noted by many previous authors, for all the different incarnations of pooled testing, the benefits of pooled testing rise as the infection probability falls in the population. Lower probabilities reduce the chance of a positive pooled test, thereby reducing the likelihood that the entire pool must be retested individually. This is clear in Equation 2 as expected tests $2 \cdot \sqrt{p} \cdot n$ drop with infection probabilities. For example, if the probability drops from 1% to .1%, the optimal pool size rises and the number of tests falls from around 20 to 6.3. There is still a large gain if the pool size is fixed: expected tests drop from 23 to around 6.9 using a fixed pool size of 20. Similarly, if the probability rises from 1% to 10%, the expected number of tests using the optimal pool size rises to around 59 (or 93 given a fixed pool size of 20).

The full relationship is shown in Figure 1, which plots the expected number of tests in a population of n people given different pool sizes and visually highlights the results based on (i) individual testing – which always leads to n tests, (ii) using pools of 20, and (iii) using optimal pooling given two stages. For simplicity, we construct these figures by assuming that n is large to remove rounding issues that arise from breaking n people into pools sizes that are not divisible by n. There are large gains from pooled testing at many infection probabilities, though they are appreciably larger at lower probabilities. We note that this figure replicates many similar figures already in the literature going back to (1).

Increasing Test Frequency

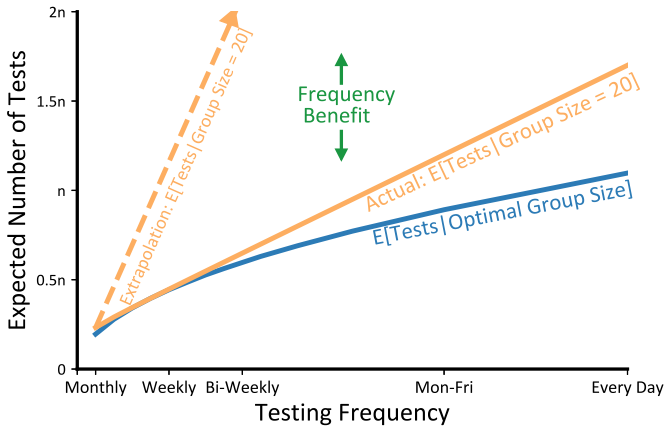
Interaction Between Frequent Testing and Pooled Testing. Our main insight is the important complementarity between pooled testing and testing frequency. Intuitively, the benefits of pooled testing rise as the infection probability falls, and frequent testing keeps this probability low at each testing period. Continuing with our example, suppose that 100 people have a p=1% independent chance of being positive over the course of an arbitrary *baseline length of time*, such as a month.

The baseline length represents the longest length of time between tests that we consider in our analysis. The variable p is then determined by both the fundamental disease characteristics and the baseline length. For example, p will be lower for a less-infectious disease and lower if the baseline length is shorter (as there is less time for infection). In our analysis below, we fix the disease and baseline length under consideration, and therefore fix p. Note then that time enters our model through the baseline length, via the variable p. For our later approximation formulas to be accurate, the baseline length must be short enough that p remains relatively low (<5%) for the given population. In other words, our approximations are appropriate when comparing different frequencies as long as all of the frequencies under consideration keep the expected p under some control.

Returning to our example, suppose that people are instead tested ten times a month. Testing individually at this frequency requires ten times the number of tests, for 1000 total tests. It is therefore natural to think that pooled testing also requires ten times the number of tests, for more than 200 total tests. However, this estimation ignores the fact that testing ten times as frequently reduces the probability of infection at the point of each test (conditional on not being positive at the previous test) from 1% to only around .1%. [**] This drop in probability reduces the number of expected tests – given pools of 20 – to 6.9 at each of the ten testing points, such that the total number is only 69. That is, testing people 10 times as frequently only requires slightly more than three times the number of tests. Or, put in a different way, there is a quantity discount of around 65% by increasing frequency. The same conclusion holds for optimal pool sizes: the one-time pool test

[**] Our analysis assumes that background risk is spread uniformly across time. Instead, one might consider a model in which the risk is changing. In this case, a main question of interest becomes *when* to test. That is, when the background risk is constant, it is appropriate to test at equal intervals because this places an equal amount of infection risk at each test. However, if the risk is changing, is is appropriate to test at unequal intervals. For example, if the background risk is exponentially rising, testing should occur more frequently over time, such that the interval between tests is exponentially decreasing.

Fig. 2. Efficiency of pooled testing rises with frequency



Notes: This graph presents the effect of testing frequency (x-axis) on the expected number of tests (y-axis), given infection probability for each individual in the population of 1% over a month. When the frequency is once a month, the points correspond to those in Figure 1 given probability of 1%: n for individual testing, $.23 \cdot n$ when using a pool size of 20 and $.20 \cdot n$ tests when using the optimal pool size. The dotted orange line represents the (incorrect) extrapolation that if a pool size of 20 leads to $.23 \cdot n$ tests when frequency is once a month, it should equal $x \cdot .23 \cdot n$ if frequency is x times a month. In reality, the expected tests are much lower, due to a quantity discount or "frequency benefit," highlighted by the green text. Finally, the blue line highlights tests given the optimally-chosen pool size.

would require 20 expected tests, while testing ten times as frequently requires 6.3 tests at each testing point, for a total of 63. The savings relative to the 1000 tests using individual testing are dramatic, with only approximately 6% of the total tests required.

Figure 2 represents this effect more generally for different levels of frequency given an infection probability of 1% over the course of a month. Note that, at a frequency of once a month, the numbers match those in Figure 1, which was based on one test given a probability of 1%. Unlike in Figure 1, we do not include the results for individual testing in this graph as testing individually every day requires 20-30 times more tests than pooled testing, which renders the graph unreadable. The dotted orange line represents the naive (and incorrect) calculation for pooled testing by extrapolating the cost of testing multiple times by using the number of tests required for one test. That is, as above, one might naively think that testing x times using a pool size of 20 in a population of n would require $x \cdot .23 \cdot n$ tests given that testing once requires $.23 \cdot n$ tests. Pooled testing is in fact much cheaper due to the reduction in the probability of infection at the time of each testing – the central contribution of this section. We therefore denote the savings between the extrapolation line and the actual requirements of pooled testing as the "frequency benefit."

The exact level of savings of the frequency benefit changes in a complicated way depending on the infection probability p given one test and the frequency x. However, approximations again provide a useful guide: the probability of a person being positive at testing $P(x) = 1 - \sqrt[x]{1-p}$ is well approximated using the first-order Taylor Series around $p = 0$ by: [††]

$$P(x) \approx \frac{p}{x}. \quad [3]$$

Plugging this into our previous approximation, the expected number of tests given the optimal pool size is then well approximated by:

$$E[tests^*|x] \approx 2 \cdot \sqrt{p} \cdot \sqrt{x} \cdot n. \quad [4]$$

Again, for our purposes, the most important fact is that these approximations are accurate for low p. For example, given p of 2%, 1%, or .1%, the approximation is within 1%, 0.5%, 0.02%, of the true number for all $x < 100$. The approximation for $E[tests^*|x]$ even given $p = 5\%$ is within 1.3%, 0.6%, and 0.008% of the true number for x of 5, 10, and 100, respectively.

Intuitively, testing at a frequency of x cuts the probability to around $\frac{p}{x}$ by Equation 3, such that the expected tests at each testing time is around $2 \cdot \sqrt{\frac{p}{x}} \cdot n$, such that testing x times requires $2 \cdot \sqrt{\frac{p}{x}} \cdot x \cdot n$ total tests, which simplifies to Equation 4. Therefore, the expected cost of pooled testing x times as frequently is around $\sqrt{x}$ when using optimal-pool-sized two-stage pooled testing, and asymptotes to this exact amount as p falls to zero. In other words, the quantity discount of increased frequency is close to $(1 - \frac{1}{\sqrt{x}})$. So, for example, pooled testing using optimally-sized pools every week (about 4 times a month) costs around $\sqrt{4} \approx 2$ times the number of tests from pooled testing every month, implying a quantity discount of 50%. Or, testing every day (around 30 times a month) costs about $\sqrt{30} = 5.5$ times the number of tests, implying a quantity discount of 82%.

Avoiding Exponential Spread Through Frequent Testing. The logic above ignores a major benefit of frequent testing: identifying infected people earlier and removing them from the population.[‡‡] Beyond the obvious health benefits, removing people from the testing population earlier stops them from infecting others, which reduces infection probability, and therefore increases the benefit of pooled testing. In the previous section, we shut down this channel by assuming that every person in the testing population had an independent probability of becoming infected. If the testing population includes people that interact, such as people who work or live in the same space, infections will spread at a higher rate within the testing population once someone is infected from the outside.

Precisely modeling spread in a given population is challenging and situation-dependent. Our goal is not to make specific statements about a particular disease in a particular situation, but provide more general and portable statements about the efficiency of frequent pooled testing in many situations. To that end, we consider a very stylized model in which we affix an exponential multiplier function $\exp(\frac{\lambda}{x})$ on $P(x)$ to capture

[††] In general, the first-order Taylor Series approximation of $f(x,p)$ around $p = 0$ is $f(x,0) + \frac{\partial f(x,p)}{\partial p}|_{p=0} \cdot (p - 0)$. In our case, then: $1 - \sqrt[x]{1-p} \approx 1 - \sqrt[x]{1-0} + \frac{\partial(1-\sqrt[x]{1-p})}{\partial p}|_{p=0} \cdot (p-0) = 0 + \frac{1}{x} \cdot (1-0)^{\frac{1}{x}-1} \cdot (p-0) = \frac{p}{x}$.

[‡‡] (37) notes a similar effect given a fixed budget of individual tests: it is more efficient to spread testing out over time because infected people are discovered earlier and removed.

the exponential-like growth associated with untamed spread within the population prior to saturation,[§§, ¶¶] such that:

$$P_{spread}(x) \equiv P(x) \cdot \exp(\frac{\lambda}{x}) \approx \frac{p}{x} \cdot \exp(\frac{\lambda}{x}). \quad [5]$$

Intuitively, given $\lambda \geq 0$, the multiplier $\exp(\frac{\lambda}{x})$ causes the probability of infection to rise above $\frac{p}{x}$, with a stronger impact as frequency drops and spread continues unchecked. Given no intra-group spread ($\lambda = 0$), $P_{spread}(x)$ reverts to $P(x)$. Just as the parameter p was chosen above to represent the probability of outside infection during the chosen baseline length of time given no testing, λ is calibrated such that $p \cdot \exp(\lambda)$ equals the probability of infection over that period when including unchecked intra-group spread. Therefore, for example, when considering a time period of a month, if outside infection alone is expected to lead to a 1% infection rate at the end of the month given no testing, but the inclusion of intra-group spread causes this rate to rise to 4%, then $\lambda = \text{Ln}[4]$. Given this addition, Equation 4 then changes to:

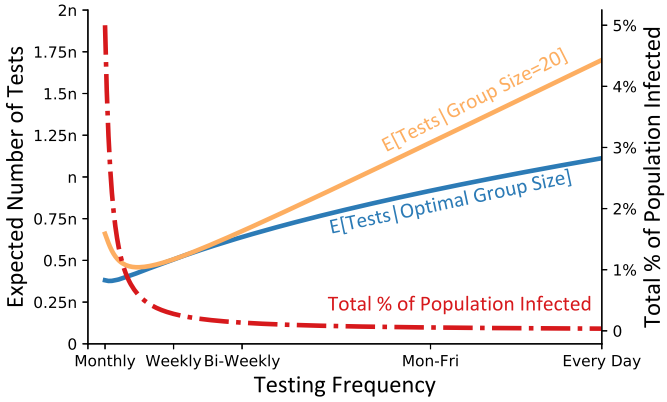
$$E[tests^*_{spread}|x] \approx 2 \cdot \sqrt{p} \cdot \sqrt{x} \cdot \sqrt{\exp(\frac{\lambda}{x})} \cdot n. \quad [6]$$

Recall that, without spread, Equation 4 implied that increasing frequency from 1 to $x > 1$ doesn't cost x times as many tests, but rather $\sqrt{x}$ times, for a quantity discount of $(1 - \frac{1}{\sqrt{x}})$. Equation 6 implies that, given spread, this benefit is increased: testing x times now lowers to $\sqrt{x} \cdot \sqrt{\exp(\lambda \cdot \frac{1-x}{x})}$ as many tests for an increased discount of $(1 - \frac{1}{\sqrt{x}} \cdot \sqrt{\exp(\lambda \cdot \frac{1-x}{x})})$. This discount is rising in x (as before) and is also rising in λ.

Somewhat counter-intuitively, the quantity discount can be so great that testing more frequently requires fewer total tests in expectation. For example, it is (approximately) cheaper to test twice as frequently if $\lambda \geq \text{Ln}[4]$, which implies the probability of infection at the time of testing is reduced by more than 4 times by testing twice as often. Intuitively, if spread is very aggressive, the efficiency gains from reduced infection probabilities arising from increasing frequency are so great that they overwhelm the increasing number of testing times. Note, however, that this "free lunch" does not necessarily continue if x doubles again from 2 to 4. For example, when $\lambda = \text{Ln}[4]$, testing four times as often costs 1.2 as many times as testing once or twice.

These effects are shown in Figure 3 for $p = .01$ and $\lambda = \text{Ln}[5]$. We plot the expected number of tests (left y-axis) and final portion of the population infected (right y-axis) for different testing frequencies. The number of infections rises in an exponential-like manner as frequency decreases and the infection is allowed to spread. The expected number of tests given different frequencies uses the same colors to represent a pool size of 20 (orange) and the optimal size (blue). Comparing Figures 2 and 3 is instructive. In Figure 2, we see a consistent increase in the tests required as the frequency of testing is increased. In Figure 3, though, the tests required are relatively flat and even decrease for early frequency increases.

Fig. 3. Increased frequency lowers infections with few additional tests given intra-pool spread



Notes: This graph presents the effect of testing frequency (x-axis) on the expected number of tests (y-axis 1) and final portion of the population that are infected (y-axis 2) given the model with intra-population spread outlined in Equation 5 with $p = .01$ and $\lambda = \text{Ln}[5]$. As shown in red dot-dashed line of final infection proportions, increased frequency reduces the exponential-like spread because infected people are removed from the population. The number of expected tests required is shown for pool size 20 in orange and in blue for the optimal pool size. There is not much increase (and even an early decrease) in the total number of tests required because frequency increases leads to increased testing efficiency.

Optimal Testing Frequency. The main benefit of increasing frequency is reducing the exponential rise in infections. As shown in Figure 3, the marginal benefit from reduced infections due to increasing frequency is high at low frequencies and drops as frequency rises, eventually to zero. Interestingly, as shown in Figure 3, the number of tests can actually fall as frequency rises when starting at low frequencies. Therefore, for low frequencies, there is, in fact, no trade-off of raising frequency: it both reduces infections and reduces tests.

As testing frequency rises, the number of expected tests will inevitably rise, leading to a trade-off between marginal benefit and cost.[***] Consequently, at very high frequencies, there is an increased cost without a large benefit. The optimal frequency lies between these extremes, but depends on the value of reducing infections versus the cost of tests, which is an issue beyond the scope of this paper.

[§§] This formula captures exponential growth in the most simplistic way possible. However, we believe it is also a good approximation for more complicated models. For example, the simple model does not apparently capture the fact that people who are infected later in a time period will cause less spread that those who are infected earlier. Solving for that more complete model leads to $P_{spread}(x) = p/\gamma \cdot (e^{\gamma/x} - 1)$ for a different spread parameter γ where $\lambda = Log((e^{\gamma} - 1)/\gamma)$. However, for reasonable parameters, our simple formula is very good approximation, particular for a low γ. For $\gamma = Log[2]$ (every person is expected to infect 1 other person over the baseline length), our formula is within 0%, 0.31%, 0.18%, and 0.002% of the true formula for x equal to 0, 5, 10, and 100 respectfully for every p. Even when $\gamma = Log[6]$, the percentages are 0%, 2.1%, 1.2%, and 0.1%. Therefore, as in the rest of the paper, we choose the simpler approximation as it leads to more transparent and intuitive conclusions.

[¶¶] If the population is saturated with infections, growth will not continue to be exponential. Therefore, the baseline testing length under consideration needs to be short enough to not allow the disease to saturate the population in expectation between tests.

[***] As an extreme example, if testing is so frequent that the infection probability at each test is effectively zero, then increasing the frequency by 1 will lead to an additional test for each pool without meaningfully reducing this probability at each testing period. This can be seen in Figure 3 for pool size of 20 where, at a frequency of around bi-weekly, the number of expected tests rises close to linearly with a slope of $\frac{1}{20} = .05 \cdot n$.

Robustness to Uncertainty and Misprediction of Infection Rates

The above analysis is predicated on the assumption that the probability of infection at the time of testing is known. However, infection rates vary wildly over time and space in a pandemic and are extremely challenging to estimate correctly (see, for example, (22)). Given this, one concern is that our main results and bullish conclusions about the efficiency of frequent pooled testing will fall apart given this uncertainty and the potential misprediction of infection rates. However, as we show in the section, our results are strongly robust to this concern.

An important benefit of frequent testing — as repeatedly noted above — is that it suppresses the infection rate at the time of testing and therefore reduces uncertainty. Using an extreme example, if people are tested every hour, the likelihood of observing any new infections at the time of testing is necessarily very small and, if any new infections are observed, they will likely be very few. That is, by testing frequently, the mean and variance of the infection rate at the time of testing is kept very small.

A less-extreme empirical example is mentioned in the introduction: the town of Wellesley, Massachusetts tested a consistent group of school staff and students weekly in the Fall of 2020. During this time, the second wave of COVID-19 was ravaging the United States. Data from Massachusetts as a whole in this period shows extremely high average weekly testing positivity rates (7%) with large variation across time (s.d. of 6%). However, the average weekly positivity rates in Wellesley stayed low (0.3%) and didn't vary considerably (s.d. of 0.3%).[†††] While the Wellesley testing population was consistently and frequently tested, the Massachusetts testing population likely consisted of many one-off tests from people who sought out a test, presumably because they were exposed or experienced symptoms. As we have noted throughout the paper, it is therefore not correct to observe high self-selected positivity rates in the general population and conclude that frequent pooled testing on a consistent subpopulation would be inefficient.

Even though frequent testing reduces the mean and variance of positivity rates, there is still uncertainty. However, we now show that this uncertainty has little impact on our conclusions. In particular, suppose that rather than being known, the infection rate is uncertain and drawn from a gamma distribution with mean μ and standard deviation σ (we chose the gamma distribution as it has the ability to reasonably match the empirical distribution of positivity rates for both Wellesley and Massachusetts as a whole). In our previous analysis, we effectively assumed that the pool designer is aware of the exact realization of the infection rate and can optimize pool size accordingly. What if, instead, the designer knows the distribution but not the specific draw? And, what happens if the designer (mistakenly) believes that the distribution is actually characterized by $\hat{\mu} = \alpha \cdot \mu$ and $\hat{\sigma} = \alpha \cdot \sigma$? Does this uncertainty or misprediction destroy the efficiency from pooled testing?

Interestingly, reasonable uncertainty and misprediction have very little impact on efficiency. For example, in the case resembling Wellesley where $\mu = 0.003$ $\sigma = 0.003$, the expected number of tests given full knowledge of the infection realization is $.093 \cdot n$. When the designer knows the distribution but is unaware of the realization, the optimal pool size is 19 and the resultant expected number of tests only rises to $.103 \cdot n$. That is, the lack of knowledge only costs $.01 \cdot n$ tests in expectation. Finally, given mistaken beliefs where α equals .5, .75, 1.5, and 2, tests only rise to $.109 \cdot n$, $.105 \cdot n$, $.106 \cdot n$, and $.109 \cdot n$, respectively. That is, mistaken beliefs have little impact on efficiency. This lack of impact is a result of the robustness of efficiency to wrongly-chosen pool sizes. For example, whereas the correct beliefs about the distribution lead to an optimal pool size of 19, mistakenly using pool sizes of 10, 15, 30, and 40 only increases tests to $.126 \cdot n$, $.109 \cdot n$, $.113 \cdot n$, and $.130 \cdot n$, respectively. The main driver of efficiency is not perfect optimization of pool sizes, but rather the mean infection rate, which is suppressed by frequent testing.[‡‡‡]

Finally, we note one additional benefit of frequent testing with respect to uncertainty. When performing a one-off test on a random new population, it is very challenging to create an accurate estimate of the risk distribution. However, performing frequent testing on a consistent population naturally generates a byproduct of past positivity realizations for the same population, which can be used to create a more accurate estimate.[§§§] For example, a surprisingly high positivity rate one week might shift beliefs about the next week's distribution upwards, leading to smaller pool sizes or more frequent testing.

Correlated Infection Further Increases Efficiency

When subpopulations are frequently tested, it is natural to pool individuals with correlated risk, such as people who live or work together. We very briefly note a result (concurrently noted by (21)), that this correlation can even further increase efficiency.

To understand the benefit of correlation given pooled testing, it is useful to broadly outline the forces that determine the expected number of tests with simple two stage testing with a pool size of g and a large testing population n. In the first stage, a test will be run for every n/g pool, while in the second stage, every n/g pool faces a probability q that at least one sample will be positive, such that all g people in the pool will need to be individually tested. Combining and simplifying these factors leads to a simple formula of the expected number of tests given a pool size: $n \cdot (1/g + q)$. As noted above, in the case of infections with independent probability p, $q = 1-(1-p)^g$. However, as infections become more positively correlated, q falls for every pool size $g > 1$. For example, with two people in a pool whose infections have a correlation r, q can be shown to be $1 - (1-p)^2 - r \cdot p \cdot (1-p)$. That is, when $r = 0$, we recover the original formula $1-(1-p)^2$, while raising r linearly drops the probability until it is p when $r = 1$. Intuitively, the pool has a positive result if either person 1 *or* person 2 is infected, which – holding p constant – is less

[†††] The use of frequent testing was not randomly assigned to Wellesley. Therefore, one fear is that Wellesley's rates are fundamentally low due to specific population characteristics. While only (non-existent) random assignment can solve this identification problem, we do note that Norfolk County – the home of Wellesley – had similar infection rates to other counties in December 2020, when Massachusetts began publishing county-by-county statistics.

[‡‡‡] While efficiency losses rise with mean and variance, reasonable robustness to uncertainty holds more generally. For example, for μ and σ of [0.01,.01],[0.03,0.03], and [0.06,0.06], the efficiency costs of lacking knowledge about the realization of distribution are $.018 \cdot n$, $.026 \cdot n$, and $.032 \cdot n$, respectively, and the efficiency loses from a mistaken beliefs of $\alpha = .5$ are $.026 \cdot n$, $.036 \cdot n$, and $.045 \cdot n$ respectively.

[§§§] In a previous version of the paper located at https://www.nber.org/papers/w27457, we discuss how employing machine learning techniques on known data can create more accurate estimates and increase testing efficiency.

likely when infections are correlated and therefore more likely to occur simultaneously.

To understand larger pools, we repeatedly simulate each individual drawing a $N(0,1)$ random variable and assigning them to be infected if their draw is above a critical value (i.e. 2.326 for a 1% infection rate). To simulate correlation, an individual's draw is a convex combination of a shared and individual normally-distributed variable, where the weights are calibrated such that the pairwise correlation between any two people is r. As an example of how q falls with more people and consequently reduces the number of tests, suppose that $p = 1\%$: when infections are uncorrelated, q is around 9.6%, 18.2%, 26.0%, and 33.1% given respective pool sizes 10, 20, 30, and 40, while q respectively drops to around 3.1%, 3.9%, 4.4%, and 4.8% when every person is pairwise-correlated with $r = 0.5$. Therefore, the respective expected number of tests given these pool sizes falls from $.196 \cdot n$, $.232 \cdot n$, $.294 \cdot n$, and $.356 \cdot n$ when uncorrelated to $.131 \cdot n$, $.089 \cdot n$, $.077 \cdot n$, and $.073 \cdot n$ when $r = 0.5$. First, note that the number of expected tests is universally lower at every pool size given correlation (and the savings are very significant). Second, note that while the pool size with the lowest number of expected tests given these potential pool sizes is 10 when there is no correlation, larger pool sizes are better given correlation. This statement is more general: a higher correlation raises the optimal pool size. The intuition is that the marginal benefit of higher pool size (reducing the $\frac{1}{g}$ first-stage tests) is the same with or without correlation, but the marginal cost (increasing the probability of second-stage testing) is reduced with higher correlation, thus leading to a higher optimum. As an example, while the optimal pool size given $p = 1\%$ is 10 given no correlation, the optimal pool sizes given r of 0, 0.2, 0,4, 0.6, 0.8 are 11, 22, 44, 107, and 385, respectively.

Conclusions

This paper shows that pooled testing is particularly efficient when frequently performed on pools with correlated risk (e.g., in workplaces or schools). Our key insight is that repeated testing reduces the infection probability at the time of each test and – since pooled testing is more efficient given lower probabilities – increases the efficiency of pooled testing. Therefore, contrary to a commonly stated rule, pooled testing is appropriate and cost-effective even for high-risk populations, as long as the frequency of testing rises in relation to this risk.[¶¶¶]. In fact, we are starting to see frequent pooled testing offered at prices of $3-10 per person per test. Consequently, frequent pooled testing is increasingly being adopted at scale, such as the state of Massachusetts where all K-12 students are offered weekly pooled testing. The cost of testing in those programs validates these estimates and demonstrates the feasibility of these strategies.

ACKNOWLEDGMENTS. Authors in alphabetical order. This paper was previously circulate under a different title: "Group Testing in a Pandemic: The Role of Frequent Testing, Correlated Risk, and Machine Learning." We are grateful to Katrina Abuabara, Sylvia Barmack, Kate Kolstad, Maya Petersen, Annika Todd, Lam Keng Siak, Johannes Spinnewijn, and Nicholas Swanson for helpful comments. All opinions and errors are our own.

[¶¶¶] However, the initial round that begins the frequent testing is an exception: when testing starts in a high-risk population, infection probabilities at the time of testing are likely to be high because there have been no actions taken to contain spread. It is therefore potentially cost-effective to use individual testing for this one initial round before switching to pooled testing for all following rounds. For these later rounds, infection probability is kept low because frequent testing is allowing for the constant removal of infected individuals.

1. R Dorfman, The detection of defective members of large populations. *The Annals Math. Stat.* **14**, 436–440 (1943).
2. M Sobel, PA Groll, Group testing to eliminate efficiently all defectives in a binomial sample. *Bell Syst. Tech. J.* **38**, 1179–1252 (1959).
3. F Hwang, A generalized binomial group testing problem. *J. Am. Stat. Assoc.* **70**, 923–926 (1975).
4. D Du, FK Hwang, F Hwang, *Combinatorial group testing and its applications*. (World Scientific) Vol. 12, (2000).
5. BA Saraniti, Optimal pooled testing. *Heal. care management science* **9**, 143–149 (2006).
6. J Feng, L Liu, M Parlar, An efficient dynamic optimization method for sequential identification of group-testable items. *IIE Transactions* **43**, 69–83 (2010).
7. T Li, CL Chan, W Huang, T Kaced, S Jaggi, Group testing with prior statistics in *2014 IEEE International Symposium on Information Theory*. (IEEE), pp. 2346–2350 (2014).
8. H Aprahamian, EK Bish, DR Bish, Adaptive risk-based pooling in public health screening. *IISE Transactions* **50**, 753–766 (2018).
9. H Aprahamian, DR Bish, EK Bish, Optimal risk-based group testing. *Manag. Sci.* **65**, 4365–4384 (2019).
10. E Lipnowski, D Ravid, Pooled testing for quarantine decisions, (Becker Friedman Institute), Working Paper 2020-85 (2020).
11. D Lakdawalla, E Keeler, D Goldman, E Trish, Getting americans back to work (and school) with pooled testing (2020).
12. N Shental, et al., Efficient high throughput sars-cov-2 testing to detect asymptomatic carriers (2020).
13. B Cahoon-Young, A Chandler, T Livermore, J Gaudino, R Benjamin, Sensitivity and specificity of pooled versus individual sera in a human immunodeficiency virus antibody prevalence study. *J. Clin. Microbiol.* **27**, 1893–1895 (1989).
14. F Behets, et al., Successful use of pooled sera to determine hiv-1 seroprevalence in zaire with development of cost-efficiency models. *AIDS (London, England)* **4**, 737–741 (1990).
15. TC Quinn, et al., Feasibility of pooling sera for hiv-1 viral rna to diagnose acute primary hiv-1 infection and estimate hiv incidence. *Aids* **14**, 2751–2757 (2000).
16. R Dodd, E Notari IV, S Stramer, Current prevalence and incidence of infectious disease markers and estimated window-period risk in the american red cross blood donor population. *Transfusion* **42**, 975–979 (2002).
17. CA Gaydos, Nucleic acid amplification tests for gonorrhea and chlamydia: practice and applications. *Infect. Dis. Clin.* **19**, 367–386 (2005).
18. MK Hourfar, et al., Blood screening for influenza. *Emerg. infectious diseases* **13**, 1081 (2007).
19. CF Manski, Bounding the accuracy of diagnostic tests, with application to covid-19 antibody tests. *Epidemiology* **32**, 162–167 (2020).
20. S Doron, et al., Weekly sars-cov-2 screening of asymptomatic students and staff to guide and evaluate strategies for safer in-person learning (2021).
21. J Rewley, Specimen pooling to conserve additional testing resources when persons' infection status is correlated: A simulation study. *Epidemiology* **31**, 832–835 (2020).
22. CF Manski, F Molinari, Estimating the covid-19 infection rate: Anatomy of an inference problem. *J. Econom.* **220**, 181–192 (2021).
23. E Litvak, XM Tu, M Pagano, Screening for the presence of a disease by pooling sera samples. *J. Am. Stat. Assoc.* **89**, 424–434 (1994).
24. E Shipitsyna, K Shalepo, A Savicheva, M Unemo, M Domeika, Pooling samples: the key to sensitive, specific and cost-effective genetic diagnosis of chlamydia trachomatis in low-resource countries. *Acta dermato-venereologica* **87**, 140–143 (2007).
25. CS McMahan, JM Tebbs, CR Bilder, Informative dorfman screening. *Biometrics* **68**, 287–296 (2012).
26. CA Hogan, MK Sahoo, BA Pinsky, Sample pooling as a strategy to detect community transmission of sars-cov-2. *Jama* **323**, 1967–1969 (2020).
27. I Yelin, et al., Evaluation of covid-19 rt-qpcr test in multi-sample pools (2020).
28. AC Bateman, S Mueller, K Guenther, P Shult, Assessing the dilution effect of specimen pooling on the sensitivity of sars-cov-2 pcr tests. *J. Med. Virol.* **93**, 1568–1572 (2021).
29. K Wu, Massachusetts actually might have a way to keep schools open (2021).
30. ea Aspinall, How did covid-19 and stabilization policies affect spending and employment? a new real-time economic tracker based on private sector data, (Rockerfeller Foundation), Technical report (2020).
31. R Phatarfod, A Sudbury, The use of a square array scheme in blood testing. *Stat. Medicine* **13**, 2337–2343 (1994).
32. A Sterrett, On the detection of defective members of large populations. *The Annals Math. Stat.* **28**, 1033–1036 (1957).
33. HY Kim, MG Hudgens, JM Dreyfuss, DJ Westreich, CD Pilcher, Comparison of group testing algorithms for case identification in the presence of test error. *Biometrics* **63**, 1152–1163 (2007).
34. D Wang, CS McMahan, JM Tebbs, CR Bilder, Group testing case identification with biomarker information. *Comput. statistics & data analysis* **122**, 156–166 (2018).
35. H Finucan, The blood testing problem. *J. Royal Stat. Soc. Ser. C (Applied Stat.* **13**, 43–50 (1964).
36. S Samuels, The exact solution to the two-stage group-testing problem. *Technometrics* **20**, 497–500 (1978).
37. B Barak, M Nitzan, NR Tannenbaum, J Yuval, Optimizing testing policies for detecting covid-19 outbreaks (2020) https://www.boazbarak.org/Papers/COVID19_arxiv.pdf.