Group Testing in a Pandemic: The Role of Frequent Testing, Correlated Risk, and Machine Learning*

Ned Augenblick, Jonathan Kolstad, Ziad Obermeyer, Ao Wang
UC Berkeley

July 17, 2020

Abstract

Group testing increases efficiency by pooling patient specimens, such that an entire group can be cleared with one negative test. Optimal grouping strategy is well studied for one-off testing scenarios, in populations with no correlations in risk and reasonably well-known prevalence rates. We discuss how the strategy changes in a pandemic environment with repeated testing, rapid local infection transmission, and highly uncertain risk. First, repeated testing mechanically lowers prevalence 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 prevalence and generates further efficiency. For this reason, increasing frequency can paradoxically *reduce* total testing cost. Second, we show that group size and efficiency increases with intra-group risk correlation, which is expected given spread within natural groupings (e.g., in workplaces, classrooms, etc). Third, because optimal groupings depend on disease prevalence and correlation, we show that better risk predictions from machine learning tools can drive large efficiency gains. We conclude that frequent group testing, aided by machine learning, is a promising and inexpensive surveillance strategy.

*Authors in alphabetical order. We are grateful to Katrina Abuabara, Sylvia Barmack, Kate Kolstad, Maya Petersen, Annika Todd, Johannes Spinnewijn, and Nicholas Swanson for helpful comments. All opinions and errors are our own.

1 Introduction

The current costs and supply constraints of testing make frequent, mass testing for SARS-CoV-2 infeasible. The old idea of group testing (Dorfman (1943)) has been proposed as a solution to this problem (Lakdawalla et al. (2020); Shental et al. (2020)): to increase testing efficiency, samples are combined and tested together, potentially clearing many people with one negative test. Given the complicated tradeoff between the benefits of increasing group size versus the cost of follow-up testing for a positive result, a large literature has emerged on optimal strategies (Dorfman (1943); Sobel and Groll (1959); Hwang (1975); Du et al. (2000); Saraniti (2006); Feng et al. (2010); Li et al. (2014); Aprahamian et al. (2018, 2019); Lipnowski and Ravid (2020)). This literature focuses 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 disease (Cahoon-Young et al. (1989); Behets et al. (1990); Quinn et al. (2000); Dodd et al. (2002); Gaydos (2005); Hourfar et al. (2007)). These environmental assumptions are violated when dealing with a novel pandemic with rapid spread. In this case, people may need to be tested multiple times, testing groups are formed from populations with correlated infection risk, and risk levels at any time are very uncertain. This paper notes how these different factors change the optimal testing strategy and open up ways to dramatically increase testing efficiency. We conclude that data-driven, frequent group testing – even daily – in workplaces and communities is a cost-effective way to contain infection spread.

We start with the well-known observation that group testing is more efficient when the population prevalence is lower, because the likelihood of a negative group test is increased. We then show how increased testing frequency mechanically lowers prevalence and therefore increases efficiency. For example, given reasonable levels of independent risk, testing twice as often cuts the prevalence 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 than requiring two times the numbers of tests, doubling 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 $(1 - 1/\sqrt{x})\%$.

The benefits to frequency are even greater when there is intra-group spread, as would be expected in a pandemic. In this case, testing more frequently has an additional benefit: by quickly removing infected individuals, infection spread is contained. This further lowers prevalence, and provides yet another driver of efficiency. We show that in this case – somewhat paradoxically – the quantity discount is so great that more frequent testing can actually *reduce* the total number of tests. Given that current testing for SARS-CoV-2 is done relatively infrequently, we therefore believe the optimal frequency is likely much higher.[1]

[1] As we show, these results do not rely on complex optimization of group size or sophisticated cross-pooling or multi-stage group testing. Furthermore, the results are qualitatively similar when using a range of reasonable group sizes.

Next, we show that grouping samples from people who are likely to spread the infection to each other – such as those that work or live in close proximity – increases the benefits of group testing. Intuitively, increased correlation of infection in a group with a fixed risk lowers the likelihood of a positive group test result, which increases efficiency. Consequently, we conclude that groups should be formed from people who are likely to infect each other, such as those in a work or living space. This has a key logistical advantage: it implies that simple collection strategies – such as collecting sequential samples by walking down the hallway of a nursing home – can encode physical proximity and therefore capture correlations without sophisticated modeling.

Finally, we note that, while there is a substantial literature noting that prevalence levels should be used to drive groupings (Hwang (1975); Bilder et al. (2010); Bilder and Tebbs (2012); Black et al. (2012); McMahan et al. (2012); Tebbs et al. (2013); Black et al. (2015); Aprahamian et al. (2019)), risk prediction methods are often coarse. This is appropriate in situations with stable risk rates, no correlation, and large amounts of previous outcome data, but less so in a quickly-changing pandemic with highly uncertain and correlated risk, such as SARS-CoV-2. We show this by quantifying the large efficiency losses from choosing groups based on incomplete or incorrect risk and correlation information. We then discuss how machine learning tools can produce highly accurate estimates of these parameters using observable data such as location, demographics, age, job type, living situation, along with past infection data.

To present transparent results, we consider a very stylized environment with a number of simplifications. 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. To pick one important example, we model a test with perfect sensitivity and specificity, but there is a natural concern that the sample dilution inherent in group testing leads to a loss of test sensitivity. However, the sensitivity loss of group testing given reasonable group sizes has been shown to be negligible in other domains (Shipitsyna et al. (2007); McMahan et al. (2012)) and more recently shown to be similarly low for SARS-CoV-2 in group sizes of 32 and 48 (Hogan et al. (2020); Yelin et al. (2020). Furthermore, even if there is a loss of sensitivity on a single test, this is counteracted by the large increase in overall sensitivity coming from running a larger number of tests given increased frequency.[2] Finally, if specificity is a concern, the past literature (Litvak et al. (1994); Aprahamian et al. (2019)) has clear methods to to optimize in the case of imperfect tests. There are multiple other issues, from grouping costs to regulatory barriers, but we believe the efficiency gains from frequent group testing are so high that addressing these issues is likely worth the cost.

The paper proceeds as follows: Section 2 reviews a main finding in the group testing literature that efficiency

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

rises as prevalence falls; Section 3 discusses the relationship between testing frequency and efficiency; Section 4 demonstrates how correlated infection leads to larger group sizes and greater efficiency; Section 5 discusses the usefulness of machine learning to estimate risk and correlation; and Section 6 concludes.

2 Group Testing: Benefits rise as prevalence falls

2.1 Background on Group Testing

To understand the basic benefits of group 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. To test each person individually – the conventional approach – requires 100 tests. Suppose instead that the individuals' samples are combined into five equally-sized groups of 20. Each of these combines samples are then tested with one test. If any one of the 20 individuals in a combined sample is positive then everyone in that group 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 group 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 grouping method is only around 23, a significant improvement over the 100 tests required in the non-grouped method.

The approach is well studied with a large literature focused on improving the efficiency of group testing. These include using optimal group size (e.g. in this example the optimal group size of 10 would lower the expected number of tests to around 20), placing people into multiple groups (Phatarfod and Sudbury (1994)), and allowing for multiple stages of group testing (Sterrett (1957); Sobel and Groll (1959); Litvak et al. (1994); Kim et al. (2007); Aprahamian et al. (2018)). There are also methods to deal with complications, such as incorporating continuous outcomes (Wang et al. (2018)). Any of these modifications can be incorporated in our group testing strategy.

For clarity of exposition, we present results for simple two-stage "Dorfman" testing – in which every person in a positive group is tested individually – to demonstrate that our conclusions are not driven by highly complex groupings and to make our calculations transparent.[3] As an example of this transparency, while the optimal group size and associated efficiency formulas under Dorfman testing are complicated, low-order Taylor-Series approximations are very simple and accurate at the low prevalences needed for group testing.[4,5] Specifically,

[3] In general, we advocate for these more sophisticated strategies when feasible as they further increase efficiency.

[4] The formula for the optimal group size (disregarding rounding) is $g^* = 2 \cdot W_0(-1/2\sqrt{-1/Ln(1-p)})/Ln(1-p)$ where $W_0(x)$ maps x to the principal solution for w in $x = we^w$. The expected number of tests is $(1-e^{2 \cdot W_0(-1/2\sqrt{-1/Ln(1-p)})} + Ln(1-p)/2 \cdot W_0(-1/2\sqrt{-1/Ln(1-p)})) \cdot n$

[5] For the prevalence rates we discuss in the paper, such as .1%, 1%, or 2%, the approximation of optimal group 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.

given a prevalence of p, the approximate optimal group size is

$$g^* \approx \frac{1}{2} + \frac{1}{\sqrt{p}} \tag{1}$$

and the resultant approximate expected number of tests given a population of n people is

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

2.2 Prevalence and Group Testing

For all of these different incarnations of group testing, the benefits of group testing rise as the prevalence rate falls in the population. Lower prevalence reduces the chance of a positive group test, thereby reducing the likelihood the entire pool must be retested individually. This is clear in Equation 2 as expected tests $2 \cdot \sqrt{p} \cdot n$ drop with prevalence. For example, if the prevalence drops from .1% to .01%, the optimal group size rises and the number of tests falls from around 20 to 6.3. There is still a large gain if the group size is fixed: expected tests drop from 23 to around 6.9 using a fixed group size of 20. Similarly, if the prevalence rises from .1% to 1%, the expected number of tests using the optimal group size rises to around 59 (or 93 given a fixed group 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 group sizes and visually highlights the results based on (i) individual testing – which always leads to n tests, (ii) using groups of 20, and (iii) using optimal grouping 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 groups sizes that are not divisible by n.[6] There are large gains from group testing at any prevalence level, though they are appreciably larger at low prevalence rates.

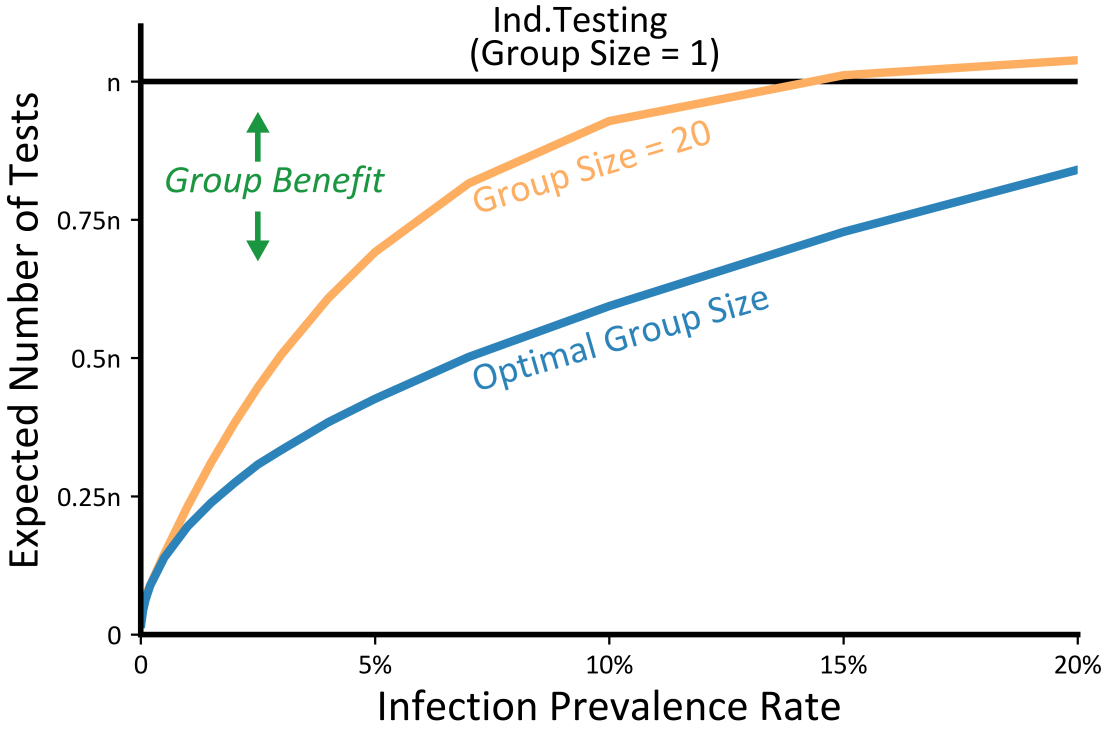
3 Increasing Test Frequency

3.1 Interaction Between Frequent Testing and Group Testing

Our first insight is the important complementarity between group testing and testing frequency. Intuitively, the benefits of group testing rise as prevalence falls and frequent testing keeps the prevalence at each testing period low. Continuing with our example, suppose that 100 people have a 1% independent chance of being positive over the course a given time period. As discussed above, one could either sample everyone (requiring 100 tests), use group testing with a group size of 20 (requiring ≈23 expected tests), or use group testing with an optimal group size (requiring ≈20 expected tests).

[6] We note that this figure replicates many similar figures already in the literature going back to Dorfman (1943).

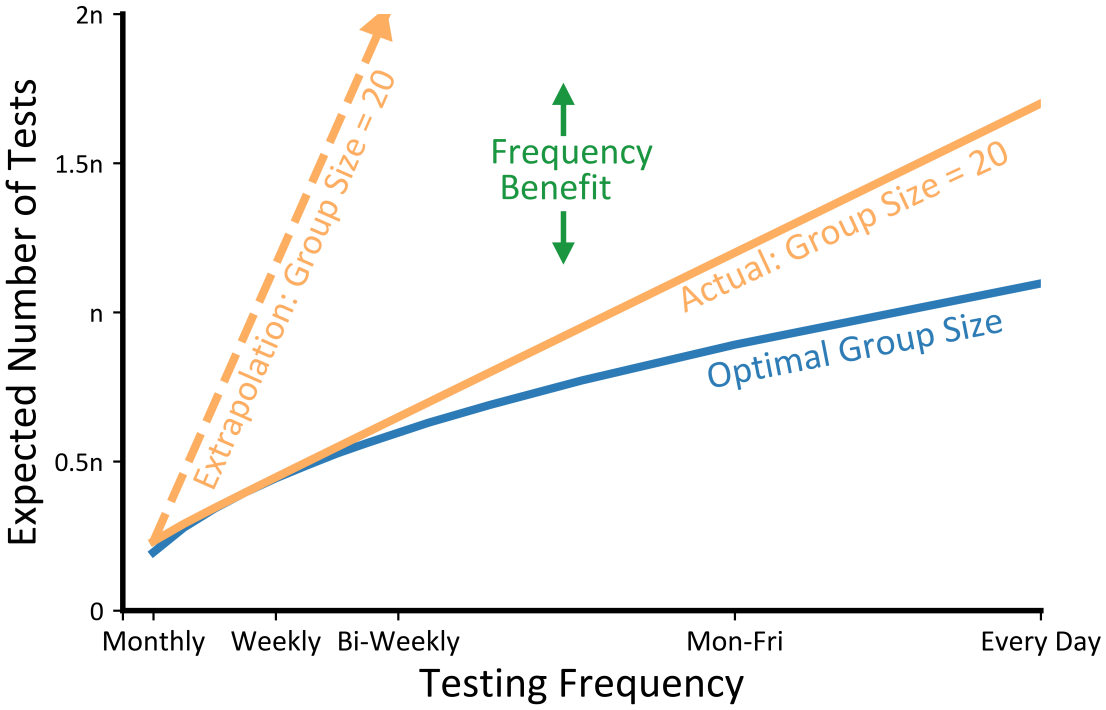
Figure 1: Efficiency of group testing rises with prevalence



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

Suppose instead that people are tested ten times as frequently. Testing individually at this frequency requires ten times the number of tests, for 1000 total tests. It is therefore natural think that group 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 previous test) from 1% to only around .1%. This drop in prevalence reduces the number of expected tests – given groups 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 using optimal group sizes: the one-time group test 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: Efficiency of group testing rises with frequency



Notes: This graph presents the effect of testing frequency (x-axis) on the expected number of tests (y-axis), given a prevalence in the population at 1% over a month. When the frequency is once a month, the points correspond to those in Figure 1 given prevalence of 1%: n for individual testing, $.23 \cdot n$ when using a group size of 20 and $.20 \cdot n$ tests when using the optimal group size. The dotted orange line represents the (incorrect) extrapolation that if a group 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 group size.

Figure 2 represents this effect more generally for different levels of test frequency given a prevalence of 1% over the course of a month. Note that, at a frequency of a once a month, the numbers match those in Figure 1, which was based on one test at a prevalence of 1%. Unlike in Figure 1, we do include the results for individual testing in this graph as testing individually everyday requires 20-30 times more tests than group testing, which renders the graph unreadable. The dotted orange line represents the naive (and incorrect) calculation for group 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 group 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. Group testing is in fact much cheaper due to the reduction in prevalence — the central contribution of this section. We therefore denote the savings between the extrapolation line and the actual requirements of group testing as the "frequency benefit."

The exact level of savings of the frequency benefit changes in a complicated way depending on the prevalence p given one test and the frequency x. However, the Taylor-Series approximation given the optimal group size is

again very accurate for reasonable prevalence rates[7] and makes the relationship clear:

$$E[tests^*|x] \approx 2 \cdot \sqrt{p} \cdot \sqrt{x} \cdot n. \tag{3}$$

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

3.2 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.[8] Beyond the obvious benefits, removing people from the testing population earlier stops them from infecting others, which reduces the prevalence, and therefore increases the benefit of group 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 instead be correlated.

Precisely modeling spread in a given population is challenging. The infection path is a complicated random variable that depends on exactly how and when different members of each particular testing population interact. Our goal is therefore not to make precise quantitative predictions but rather make a set of qualitative points based on the exponential-like infection growth common across virtually all models.[9] Consequently, we focus on the results of the simplest simulation that captures correlated exponential growth.

Specifically, we suppose that the 100 people continue to face a 1% chance of being infected over the course of a time period due to contact with someone outside of the test population. However, once a person is infected, this person can spread the infection to other people. We then chose an infection transmission rate such that if a person was infected at the start of the time period and was not removed from the population, the infected person

[7] For example, even given $p = 5\%$, the approximation is within 1.3%, 0.6%, and 0.008% of the true number for x of 5, 10, and 100, respectively.

[8] Barak et al. (2020) 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.

[9] We consistently use the term "exponential-like" spread as we are considering smaller testing populations in which the infection rate slows as infected people are removed or become immune, such that the spread is actually logistic. However, we focus on situations in which the spread is caught early and therefore still close to exponential.

would have a total independent probability s% over the time period of personally spreading the infection to each other person.[10]

Now, consider the spread over the entire time period. With 37% probability, no one in the population is infected. However, with a 63% chance, a person is infected from the outside of the group, setting off exponential-like growth within the rest of the test population (which ends only after testing and infected people are removed).Given this growth, for example, if s=1%, 5%, 10%, approximately 2, 8, and 26 people will be infected at the end of the time period. However, by testing ten times in the time period, infected people are removed more quickly and there is less time for a person to infect others, such that only around 1, 2, or 4 people are infected by the end of the time period.

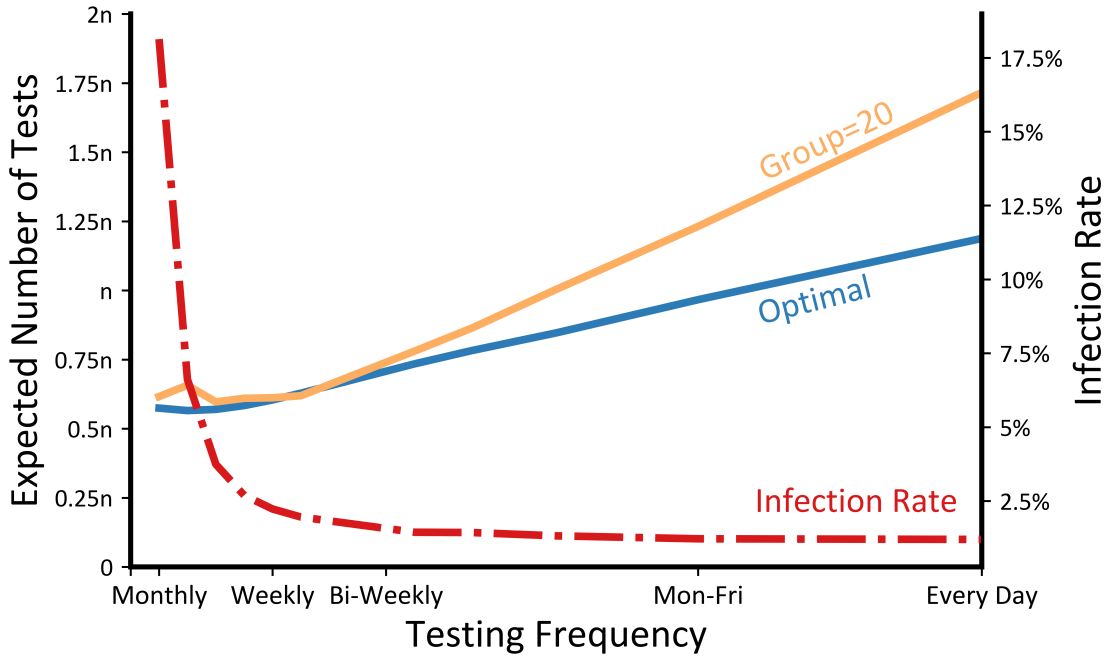
Not only are infections rates reduced, but the reduction in prevalence reduces the relative cost of more frequent group testing. For example, using a group size of 20 and testing once over the time period leads to 23 tests in expectation given no spread. However, with spread given s=1%, 5%, 10%, this number rises to 30, 50, and 58, respectively. On the other hand, the expected number of tests when testing ten times as frequently does not grow at the same rate due to controlling the spread and therefore prevalence. Specifically, the number of tests needed rises from 69 (with no spread) to 71, 84, and 115, respectively. That is, for example, at s=5%, the number of needed tests rises by 150% when testing once due to spread, but only rises by 22% when testing ten times as frequently.

These effects are shown in Figure 3. We plot the expected number of tests (left y-axis) and infection rate (right y-axis) for different testing frequencies assuming s=5%.[11] The infection rate 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 group sizes of 20 (orange) and 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. The difference is due to the fact that, with intra-group infections, testing more frequently has the additional benefit of lowering the prevalence rate by containing infections. For example, the quantity discount from a frequency of 2 is higher than 50% in the case of optimal group sizes, such that the total cost from doubling the frequency actually falls.

[10] For example, if s=5% and the time period is a month, then infected person i who has not been removed has a independent $1-(1-.05)^{(1/30)} \approx .17\%$ daily chance of infecting non-infected person j, $\approx .17\%$ daily chance of infecting non-infected person k, etc.

[11] There is one minor difference between this figure and the example in the text. In the figure, we use a population of 120 rather than 100, as it allows for easier division into many group sizes. This has a very slight impact as it slightly increases the time for the exponential spread to be curbed by increasing population immunity.

Figure 3: Increased frequency lowers infections with few additional tests given intra-group spread



Notes: This graph presents the effect of testing frequency (x-axis) on the expected number of tests (y-axis 1) and infection rate (y-axis 2) given a model with intra-group spread over the course of a month. As shown in red dot-dashed line of infection rates, increased frequency reduces the exponential-like spread because infected people are removed from the population. The number of expected tests required is shown for group size 20 in orange and in blue for the optimal group size. There is not much increase (and even an early decrease) in the number of tests required as frequency increases because the increased frequency reduces the spread and therefore prevalence.

3.3 Optimal Testing Frequency

The main benefit of increased 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.[12] Consequently, at very higher 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. However, our strong suspicion given

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

our results and the economic costs of testing versus the economic (and human) costs of higher infections is that the optimal frequency is higher than the seemingly-common policy of testing on the order of monthly or only when an employee is symptomatic.[13]

4 Correlated Infection

In this Section, we discuss and isolate an additional factor implicitly included in the above example, which includes correlation between people whose samples are grouped together. We use our simple example to separate the insight that this correlation alone (i) is complementary with group testing in that it reduces the number of expected tests, (ii) leads to larger optimal group sizes, and (iii) has greater reduction if the structure is known and used to determine the composition of the groups.

To first understand the benefit of correlation given group testing, it is useful to broadly outline the forces that determine the expected number of tests with simple two stage testing with a group size of g and a large testing population n. In the first stage, a test will be run for every n/g group, while in the second stage, every n/g group faces a probability q that at least one sample will be positive, such that all g people in the group will need to be individually tested. Combining and simplifying these factors leads to a simple formula of the expected number of tests given a group 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 group size $g > 1$. For example, with two people in a group whose infections have 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 group has a positive result if either person 1 *or* person 2 is infected, which – holding p constant – is less likely when infections are correlated and therefore more likely to occur simultaneously.

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 group 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 group 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 group size given correlation (and the savings are very significant). Second, note that while the group size with the lowest number of expected tests given these potential group sizes is 10 when there is no correlation, larger group sizes are better given correlation. This statement is more general: higher correlation raises the optimal group size. The intuition is that the marginal

[13] We also note a an important comparative static: it is more valuable to more frequently test a person who is more likely to catch and spread the infection (such as a person who meets with many other people versus a person who works alone in the back room).

benefit of higher group size (reducing the $1/g$ 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 group size given $p = 1\%$ is 10 given no correlation, the optimal group sizes given r of 0, 0.2, 0,4, 0.6, 0.8 are 11, 22, 44, 107, and 385, respectively. Finally, note that when $r = 1$, the optimal group size is unbounded, because adding an extra person to a group adds benefit by reducing first-stage tests, but has no cost because the probability of a positive in the group remains constant at p. Obviously, extremely large group sizes are potentially technologically infeasible, but the intuition remains that correlation within testing groups should raise group size.

5 Machine Learning Predictions of Correlated Risk

Individual infection risk and correlation in a pandemic is very uncertain and likely constantly changing. Unfortunately, without accurate assessments of these parameters, it is challenging to optimize group composition. In this section, we first demonstrate the value of accurate prediction in terms of testing efficiency. Then, we discuss how machine learning can be used to estimate key parameters using information on the testing population and indirect information about network structure. While we outline a general strategy, we stop short of developing a specific algorithm as the prediction methods are likely to be specific to each testing setting.

5.1 The Role of Accurate Prediction

Our goal is to understand the value of accurate predictions of individual risk and correlation structure in terms of expected tests. Consequently, we introduce heterogeneity and correlated spread into our simple static simulation by considering a large population of n people in which half of the population is high risk (5% chance of being infected), half is low risk (.1% chance of infection), and the risk groups consist of clusters of 40 people whose infections are pairwise correlated with $r = 0.25$.[14] Next, we quantify the efficiency loss when group composition is decided based on various incorrect or incomplete information about those parameters, with the results shown in Table 1.

The first line of the table considers the baseline of individual testing, which requires n tests for n people. Next, we consider the efficiency of group testing given accurate estimation of heterogeneity and correlation. Accurate estimates prompt the designer to place correlated low risk individuals into large groups of 40 and correlated high risk individuals into smaller groups of 10. This leads to large efficiency gains: expected tests drops by 82% to $.18n$. That is, even when half of the individuals are high risk, group testing still provides large efficiency gains

[14] These clusters could represent floors of an office building or apartment building, in which people can easily spread the infection to each other in a cluster but don't spread across clusters.

Table 1: Efficiency Gains From Information

Belief	**Implied group size**	**Exp Tests**	**Issue**
Baseline			
No group testing	1	n	Individual Testing Inefficient
Correct belief			
Half .1%, Half 5% (corr)	.1%→40,5%→10	$.18n$	Optimal
Incorrect beliefs			
Everyone .1%	34	$.58n$	Believe all low risk $\Longrightarrow$ groups too big
Everyone 5%	5	$.32n$	Believe all high risk $\Longrightarrow$ groups too small
Everyone ˜2.5%	8	$.30n$	Believe homogenous $\Longrightarrow$ don't split
Half .1%, Half 5% (no corr)	.1%→34,5%→5	$.24n$	Believe $\rho = 0 \Longrightarrow$ groups too small

Notes: This table outlines the effect of different beliefs on grouping strategy, in a situation where half of the n people have .1% correlated risk with $r = 0.25$ and the other half have 5% correlated risk with $r = 0.25$. Row 2 shows the optimal strategy, where the designer correctly estimates risk and correlation. Rows 3-6 show scenarios in which the group designer holds false beliefs. For example, in Row 3, she believes everyone is a low risk type with .1% uncorrelated risk, causing her to form (believed-to-be-optimal) groups of 34. This leads the expected number of tests to be $.58n$, which is sub-optimal because the groups for the high-risk types are too big.

given accurate estimations.

However, without sophisticated estimation tools, the designer is likely to have an incorrect or incomplete assessment of these parameters. For example, we next consider the effect of the mistaken belief that everyone is low risk. This belief leads the designer to mix both low and high risk people into large groups of 34. These groupings are inefficient because large groups containing high risk individuals are very likely to have at least one infected individual and therefore require a costly second stage of testing. Consequently, the expected number of tests rises by over 300% to $.58n$, largely mitigating the efficiency gains of group testing. That is, group testing is only as good as the risk and correlation estimates driving the groupings. Conversely, the third line considers the effect of falsely believing that everyone is high risk, which leads to small groups of 5. While expected tests drop to $.32n$, these grouping are still inefficient because small groups necessarily leads to more first-stage tests, which are not necessary for the low-risk individuals. Next, the fourth line considers the belief that everyone shares the same population average risk ($\approx$2.5%), which would occur if the overall risk estimate was accurate, but there was no way to specifically identify high- and low-risk individuals. Here, tests again drop to $.30n$, but groups are inefficient because the group sizes for the different risk groups does not vary. Finally, the fifth line notes that expected tests drop by 20% to $.24n$ when individuals can be accurately separated into risk types. In this case, low risk individuals are placed in groups of 34 and high risk into groups of 5, which are optimal given the respective risk levels if there was no risk correlation. Still, the lack of correlation estimation leads to 33% more tests in comparison to the optimal groupings.

5.2 Machine Learning to Predict SARS-CoV-2 Risk

Risk prediction is a specific case of a "prediction policy problem" (Kleinberg et al. (2015)), in the sense that the likelihood of a negative (positive) group test is an important driver of the value of group testing. Optimizing group sizes does not require causal estimates of *why* someone has a specific risk or has correlated risk with another person, but rather just accurate *predictions* of these parameters. Therefore, our goal is to sequentially outline the basic methods to estimate individual risk and the underlying transmission network as a function of observable characteristics.

A crucial point about individual risk in a pandemic – which complicated estimation – is that it is likely changing constantly. As a simple example, our discussions of the benefits of frequency are based on the notion that testing directly changes risk assessment: even a potentially "high-risk" worker has a low risk of being positive today if they tested negative yesterday. Similarly, a "low-risk" worker living in area with a temporary outbreak has a temporary high risk of being infected. Fortunately, it is possible to estimate these effects using machine learning given data on testing outcomes, population demographics, home locations, or even geo-coded travel data from a mobile device or known contacts where technology enabled contact tracing tools are implemented. A crucial benefit of machine learning is the ability to identify non-linearities and interaction effects, which are likely important in this complicated prediction problem. As a simple example, going out frequently in a low-prevalence area leads to low risk, staying home in a high-prevalence area leads to low risk, but going out in a high-prevalence area leads to high risk.

Predicting the infection network is also challenging. The basic strategy is to use observable "connections" data about people – such as shared physical work space, shared living space, shared demographic data, social network data, etc. – to predict the unobservable infection network. One natural approach would be to infer this mapping using the relationship between the connection variables and observable infections. Unfortunately, we believe that the sparsity of infections will cause this method to fail. However, given the natural assumptions that infection networks are based on physical interactions, we can instead shift to using connection data to predict physical proximity. Broadly, given z ways in which people might interact, the goal is to reflect proximity by estimating a z dimensional distance between them.

Finally, we point to the potential to use easily-obtainable sample-collection data as an important indirect measure of connection to provide significant insight into people's physical interaction (and therefore infection) network. The simple idea is that the order and location of sample collection can shed light on proximity. For example, if a nurse simply walked down a hall of a nursing home and sequentially collected samples, the ordering of sampling would largely encode physical distance. Then, given that the efficient testing strategy is to combine correlated samples, the samples taken closest to each other could be combined. That is, a fortunate coincidence

is that some of the most natural collection methods also provide a large amount of indirect information about connection with little additional cost. We therefore advocate taking this effect into account when designing testing collection strategies and deciding how to combine samples.

6 Conclusions

The combination of high-frequency testing with machine learning predictions and knowledge of risk correlation (e.g., in workplaces or facilities) is more plausible and cost-efficient than previously thought. While a formal cost-effectiveness analysis is beyond the scope of this paper, we estimate that, given marginal test cost of $50, daily group testing could be offered for between $3-5 per person per day. This makes high-frequency, intelligent group testing a powerful new tool in the fight against SARS-CoV-2, and potentially other infectious diseases.

References

Aprahamian, Hrayer, Douglas R Bish, and Ebru K Bish, "Optimal risk-based group testing," *Management Science*, 2019, *65* (9), 4365–4384. 1, 2

_ , Ebru K Bish, and Douglas R Bish, "Adaptive risk-based pooling in public health screening," *IISE Transactions*, 2018, *50* (9), 753–766. 1, 3

Barak, Boaz, Mor Nitzan, Neta Ravid Tannenbaum, and Janni Yuval, "Optimizing testing policies for detecting COVID-19 outbreaks," 2020. https://www.boazbarak.org/Papers/COVID19_arxiv.pdf. 7

Behets, Frieda, Stefano Bertozzi, Mwamba Kasali, Mwandagalirwa Kashamuka, L Atikala, Christopher Brown, Robert W Ryder, and Thomas C Quinn, "Successful use of pooled sera to determine HIV-1 seroprevalence in Zaire with development of cost-efficiency models.," *AIDS (London, England)*, 1990, *4* (8), 737–741. 1

Bilder, Christopher R and Joshua M Tebbs, "Pooled-testing procedures for screening high volume clinical specimens in heterogeneous populations," *Statistics in medicine*, 2012, *31* (27), 3261–3268. 2

_ , _ , and Peng Chen, "Informative retesting," *Journal of the American Statistical Association*, 2010, *105* (491), 942–955. 2

Black, Michael S, Christopher R Bilder, and Joshua M Tebbs, "Group testing in heterogeneous populations by using halving algorithms," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 2012, *61* (2), 277–290. 2

_ , _ , and _ , "Optimal retesting configurations for hierarchical group testing," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 2015, *64* (4), 693–710. 2

Cahoon-Young, B, A Chandler, T Livermore, J Gaudino, and R Benjamin, "Sensitivity and specificity of pooled versus individual sera in a human immunodeficiency virus antibody prevalence study.," *Journal of Clinical Microbiology*, 1989, *27* (8), 1893–1895. 1

Dodd, RY, EP Notari IV, and SL Stramer, "Current prevalence and incidence of infectious disease markers and estimated window-period risk in the American Red Cross blood donor population," *Transfusion*, 2002, *42* (8), 975–979. 1

Dorfman, Robert, "The detection of defective members of large populations," *The Annals of Mathematical Statistics*, 1943, *14* (4), 436–440. 1, 4

Du, Dingzhu, Frank K Hwang, and Frank Hwang, *Combinatorial group testing and its applications*, Vol. 12, World Scientific, 2000. 1

Feng, Jiejian, Liming Liu, and Mahmut Parlar, "An efficient dynamic optimization method for sequential identification of group-testable items," *IIE Transactions*, 2010, *43* (2), 69–83. 1

Gaydos, Charlotte A, "Nucleic acid amplification tests for gonorrhea and chlamydia: practice and applications," *Infectious Disease Clinics*, 2005, *19* (2), 367–386. 1

Hogan, Catherine A, Malaya K Sahoo, and Benjamin A Pinsky, "Sample pooling as a strategy to detect community transmission of SARS-CoV-2," *Jama*, 2020, *323* (19), 1967–1969. 2

Hourfar, Michael Kai, Anna Themann, Markus Eickmann, Pilaipan Puthavathana, Thomas Laue, Erhard Seifried, and Michael Schmidt, "Blood screening for influenza," *Emerging infectious diseases*, 2007, *13* (7), 1081. 1

Hwang, FK, "A generalized binomial group testing problem," *Journal of the American Statistical Association*, 1975, *70* (352), 923–926. 1, 2

Kim, Hae-Young, Michael G Hudgens, Jonathan M Dreyfuss, Daniel J Westreich, and Christopher D Pilcher, "Comparison of group testing algorithms for case identification in the presence of test error," *Biometrics*, 2007, *63* (4), 1152–1163. 3

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer, "Prediction policy problems," *American Economic Review*, 2015, *105* (5), 491–95. 13

Lakdawalla, Darius, Emmet Keeler, Dana Goldman, and Erin Trish, "Getting Americans back to work (and school) with pooled testing," *USC Schaeffer Center White Paper*, 2020. 1

Li, Tongxin, Chun Lam Chan, Wenhao Huang, Tarik Kaced, and Sidharth Jaggi, "Group testing with prior statistics," in "2014 IEEE International Symposium on Information Theory" IEEE 2014, pp. 2346–2350. 1

Lipnowski, Elliot and Doron Ravid, "Pooled Testing for Quarantine Decisions," Working Paper 2020-85, Becker Friedman Institute 2020. 1

Litvak, Eugene, Xin M Tu, and Marcello Pagano, "Screening for the presence of a disease by pooling sera samples," *Journal of the American Statistical Association*, 1994, *89* (426), 424–434. 2, 3

McMahan, Christopher S, Joshua M Tebbs, and Christopher R Bilder, "Informative dorfman screening," *Biometrics*, 2012, *68* (1), 287–296. 2

Phatarfod, RM and Aidan Sudbury, "The use of a square array scheme in blood testing," *Statistics in Medicine*, 1994, *13* (22), 2337–2343. 3

Quinn, Thomas C, Ron Brookmeyer, Richard Kline, Mary Shepherd, Ramesh Paranjape, Sanjay Mehendale, Deepak A Gadkari, and Robert Bollinger, "Feasibility of pooling sera for HIV-1 viral RNA to diagnose acute primary HIV-1 infection and estimate HIV incidence," *Aids*, 2000, *14* (17), 2751–2757. 1

Saraniti, Brett A, "Optimal pooled testing," *Health care management science*, 2006, *9* (2), 143–149. 1

Shental, Noam, Shlomia Levy, Shosh Skorniakov, Vered Wuvshet, Yonat Shemer-Avni, Angel Porgador, and Tomer Hertz, "Efficient high throughput SARS-CoV-2 testing to detect asymptomatic carriers," *medRxiv*, 2020. 1

Shipitsyna, Elena, Kira Shalepo, Alevtina Savicheva, Magnus Unemo, and Marius Domeika, "Pooling samples: the key to sensitive, specific and cost-effective genetic diagnosis of Chlamydia trachomatis in low-resource countries," *Acta dermato-venereologica*, 2007, *87* (2), 140–143. 2

Sobel, Milton and Phyllis A Groll, "Group testing to eliminate efficiently all defectives in a binomial sample," *Bell System Technical Journal*, 1959, *38* (5), 1179–1252. 1, 3

Sterrett, Andrew, "On the detection of defective members of large populations," *The Annals of Mathematical Statistics*, 1957, *28* (4), 1033–1036. 3

Tebbs, Joshua M, Christopher S McMahan, and Christopher R Bilder, "Two-stage hierarchical group testing for multiple infections with application to the infertility prevention project," *Biometrics*, 2013, *69* (4), 1064–1073. 2

Wang, Dewei, Christopher S McMahan, Joshua M Tebbs, and Christopher R Bilder, "Group testing case identification with biomarker information," *Computational statistics & data analysis*, 2018, *122*, 156–166. 3

Yelin, Idan, Noga Aharony, Einat Shaer-Tamar, Amir Argoetti, Esther Messer, Dina Berenbaum, Einat Shafran, Areen Kuzli, Nagam Gandali, Tamar Hashimshony et al., "Evaluation of COVID-19 RT-qPCR test in multi-sample pools," *MedRxiv*, 2020. 2