

A Robust Pooled Testing Approach to Expand COVID-19 Screening Capacity

Douglas R Bish
University of Alabama
Information Systems, Statistics, and Management Science

Ebru K Bish
University of Alabama
Information Systems, Statistics, and Management Science

Hussein El-Hajj
Virginia Tech
Industrial and Systems Engineering

Hrayr Aprahamian
Texas A&M University
Industrial and Systems Engineering

Submitted: April 2020; Revised: July 2020; December 2020

Corresponding author:

Douglas R Bish; drbish@cba.ua.edu, Tuscaloosa, AL

Abstract

Limited testing capacity for COVID-19 has hampered the pandemic response. *Pooling* is a testing method wherein samples from specimens (e.g., swabs) from multiple subjects are combined into a pool and screened with a single test. If the pool tests positive, then new samples from the collected specimens are individually tested, while if the pool tests negative, the subjects are classified as negative for the disease. Pooling can substantially expand COVID-19 testing capacity and throughput, without requiring additional resources.

We develop a mathematical model to determine the best *pool size* for different *risk groups*, based on each group's estimated COVID-19 prevalence. Our approach takes into consideration the sensitivity and specificity of the test, and a dynamic and uncertain prevalence, and provides a *robust* pool size for each group. For practical relevance, we also develop a companion COVID-19 pooling design tool (through a spread sheet).

To demonstrate the potential value of pooling, we study COVID-19 screening using testing data from Iceland for the period, February-28-2020 to June-14-2020, for subjects stratified into high- and low-risk groups. We implement the robust pooling strategy within a *sequential* framework, which updates pool sizes each week, for each risk group, based on prior week's testing data. Robust pooling reduces the number of tests, over individual testing, by 88.5% to 90.2%, and 54.2% to 61.9%, respectively, for the low-risk and high-risk groups (based on test sensitivity values in the range [0.71, 0.98] as reported in the literature). This results in much shorter times, on average, to get the test results compared to individual testing (due to the higher testing throughput), and also allows for expanded screening to cover more individuals. Thus, robust pooling can potentially be a valuable strategy for COVID-19 screening.

INTRODUCTION

With around 46.8 million confirmed cases and 1.2M deaths in at least 188 countries (as of 11/2/2020) [1],

the COVID-19 pandemic continues to be the cause of considerable suffering and economic disruption.

Effective mitigation requires laboratory-based testing to identify COVID-19 positive subjects. As the

WHO Director-General Ghebreyesus puts it, "The most effective way to prevent infections and save lives

is breaking the chains of transmission, and to do that you must test and isolate. We cannot stop this

pandemic if we don't know who is infected. We have a simple message for all countries: test, test, test"

[2]. Testing is essential not only for symptomatic individuals, but also for asymptomatic individuals [3], because they are a major source of transmission [4].

Unfortunately, a major impediment to mitigation efforts in the United States (US), and in other parts of the world, has been the limited testing capacity for COVID-19, which, in the US, has constrained the number of tests that could be conducted, mostly restricting testing to those with symptoms for case identification, and less on mass screening and contact tracing efforts. Countries that were able to ramp up their COVID-19 testing capacity quickly and follow an aggressive testing strategy earlier during the epidemic were considerably more effective than others. For example, South Korea was able to curb the growth of the disease mainly through expanded testing, without a strong imposition of social distancing measures. By testing over 300,000 people out of its population of 52 million [5], South Korea was able to implement highly effective interventions early on, including contact tracing, followed by enforced quarantines and isolations. Another example is Iceland, which is able to offer free COVID-19 testing to the general population [6], and as of June 14, 2020, was able to test around 17% of its population [7].

The current mode of COVID-19 testing worldwide is *individual testing*, that is, each subject's specimen is tested with a single test. An alternative to individual testing, proposed by Dorfman in 1943 [8], is *pooled testing*, in which samples, sufficient for testing, are extracted from the specimens (e.g., nasopharyngeal swabs) from multiple subjects, and combined in a *pool*, and tested via a single test; if the pooled test's outcome is positive, then all subjects in the pool are individually tested (with the same type of test, using a new sample extracted from the previously collected specimen); and if the pooled test's outcome is negative, then all subjects in the pool are classified as disease-negative. Pooled testing does not require any additional resources beyond individual testing [9], and can substantially expand testing capacity over individual testing, especially when prevalence rates are low. Pooled testing, and in particular the Dorfman pooling method, is currently used in public health screening, including screening for sexually-transmitted diseases and screening donated blood for transfusion-transmittable infections [10-13]. An

important design decision in pooled testing is the *pool size* (i.e., number of specimens in each pool) so as to maximize the efficiency of testing (i.e., minimize the number of tests per subject). [11] provides a rigorous analysis on the selection of an optimal pool size for Dorfman pooling. Because pooling has obvious application to COVID-19 testing, there have been several recent pooling papers focused on COVID-19, for example, [14-16] offer various simplifications of the analysis in [11], with a focus on COVID-19. Unlike [14-16], this paper considers uncertainty in the prevalence rate, which is a defining feature of COVID-19; in other public health screening applications, the disease spread is less dynamic and prevalence changes at a much slower rate (e.g., sexually-transmitted diseases). For COVID-19, disease prevalence changes quickly, and there is high uncertainty around its prevalence at any point in time.

Considering pooled testing design, one concern is accuracy, in particular a potential increase in the number of false-negatives as a result of pooling. Pooling does not increase the number of false-positives over individual testing [11]. With imperfect tests, pooling will increase the false-negative rate (e.g., see [11]), because a disease-positive subject must be tested twice (first in a pool, then individually) to be classified as positive. Further, pooling may lead to the “dilution” of the infected specimen(s) in the pool, reducing the sensitivity of the pooled test, thus increasing the false-negative rate, especially for larger pools [13]. A common method for reducing dilution is to set a maximum pool size for, e.g., see [10, 12]. There are several lab-based studies that investigate the impact of pooling on the sensitivity of a PCR test for COVID-19, which is the most common type of test for COVID-19 screening. This preliminary research indicates that pooled testing leads only to a minor reduction in PCR sensitivity for larger pool sizes. For example, [9] shows that the PCR test was able to detect an infected specimen in pools of size of 32 in nine out of the ten pools tested. [17] shows a similar result for pools of size 30. In addition, several other studies find no loss of sensitivity for smaller pools, e.g., [18], [19], and [20] use pool sizes of 5, 8, and 10, respectively. Not surprisingly, [21, 22] show that the dilution effect is more pronounced when the infected specimens in the pool have low viral loads; the window period analysis in [21] indicates that the

low viral load typically occurs for specimens that are collected either too early or too late after infection. Related to the overall pooled testing design, [23] proposes a design where the pooled test is repeated to reduce false-negatives, and shows that such a design is still efficient compared to individual testing. On the other hand, [24] explores, via simulation, different pooling designs (e.g., adaptive and non-adaptive) and different metrics (e.g., efficiency and accuracy).

A PCR test run takes in the order of 2 hours [25, 26] to complete, thus, pooled testing, followed by individual testing as needed, is viable from a time perspective, and does not significantly increase the testing time for a particular subject. Further, testing machines have a capacity, which limits their throughput (e.g., number of completed tests per testing day), e.g., a limit of 96 samples per testing cycle (run) is common [25]. Because pooled testing increases the number of specimens tested per testing cycle, it also increases the throughput, shortening the average time to get the test results. The expanded testing capacity provided by pooling could enable more extensive testing for COVID-19.

This paper builds on previously published mathematical models by the authors; develops a robust approach for pooled testing design that can be customized for different testing populations and for a dynamic and uncertain disease prevalence; and demonstrates that the proposed pooled testing approach has the potential to substantially increase COVID-19 testing capacity. Further, we explore the impact of test sensitivity, and provide insight on how to manage the uncertainty in test sensitivity. This research is timely as limited COVID-19 testing capacity remains a serious problem, and more testing capacity is urgently needed, especially given the increasing number of infections, and the efforts to return to some form of “normalcy.”

METHODS

We study case identification via pooled testing; develop an easily implementable method for robust pooling design under prevalence uncertainty [27, 28], and for risk-stratified groups; and illustrate the

benefits through a case study on mass screening for COVID-19. Our model builds on our earlier work, in particular, we take the analytically complex models and results from [10], which uses risk-based pooling (to determine pool sizes and pool assignments for subjects given their individual disease risk), from [11], which uses robust optimization (to determine pool sizes under prevalence uncertainty), and from [29], which develops sequential pooling design for surveillance; and we develop a novel method for designing a simple, *robust pooling* strategy, which can be incorporated into a *sequential pooling design* framework to consider infection dynamics and prevalence uncertainty. For practical relevance, we develop a companion COVID-19 pooling design tool (through a spread sheet), which will be available online.

We consider a PCR test, which can be used for both individual and pooled testing. For pooled testing, we consider the Dorfman method [8] (hereafter, “pooling”). The test produces a binary outcome, with a positive (negative) outcome indicating the presence (absence) of the infection in the pool (for pooled testing) or in the individual specimen (for individual testing). We let Se and Sp respectively denote the test *sensitivity* (true positive probability) and *specificity* (true negative probability), and assume that pooling does not change the test’s efficacy up to a *maximum allowable pool size*. The terms “subject” and “specimen” respectively refer to the individual to be tested, and specimen collected from the individual, and are used interchangeably. We assume that each specimen has sufficient material (samples) for multiple tests, thus, if an individual follow-up test is needed for a subject, it is conducted on a new sample extracted from the subject’s previously collected specimen.

Model

We design a pooling strategy for a testing population divided into *risk groups* based on each group’s estimated disease prevalence, by determining, for each group, a *robust pool size*. The objective is to

maximize the efficiency of testing (minimize the expected number of tests per subject) under unknown prevalence. In the following, we detail the derivation of the robust pool size, discuss its extension to sequential pooling design, and derive the false-negative rate of pooling.

To simplify the subsequent notation, we omit the group index. Consider a given group, and let P denote the unknown prevalence for the group, with uncertainty set (range) $S(P)=[L, U]$, which consists of discrete *prevalence scenarios*, each with probability, $Pr(P=p)$, $p \in S(P)$; our method extends to any user-specified discrete or continuous distribution for P . The uncertainty set $S(P)$ is estimated by the tester, and does not necessarily correspond to the true support of random variable P , which is unknown in practice.

Testing Efficiency

Let n_{max} denote the maximum allowable pool size (due to technological limitation, dilution effect, etc.). If a pool with n specimens tests negative, then only $1/n$ tests are needed per subject; and if the pool tests positive, then each subject requires an individual “follow-up” test, leading to $1+1/n$ total tests per subject (Table 1 displays the probability of each possible event). Then, for any pool size n and prevalence scenario p , the expected number of tests per subject tested (“expected tests”) under pooling, denoted by $E[T(n, p)]$, follows:

$$E[T(n, p)] = \frac{1 + n[Se - (Se + Sp - 1)(1 - p)^n]}{n}. \quad (1)$$

Table 1. Probabilities of All Possible Events in Pooled Testing (for a Pool of n Subjects).

Pooled Test Outcome True Subject Status in Pool	Pooled Test Negative (Pooled Test Only)	Pooled Test Positive (Pooled Test Plus n Individual Tests)
All Subjects Negative	$Sp(1 - p)^n$	$(1 - Sp)(1 - p)^n$

At Least One Subject Positive	$(1 - Se)(1 - (1 - p)^n)$	$Se(1 - (1 - p)^n)$
-------------------------------	---------------------------	---------------------

The probabilities for the intersections of all possible pooled test outcomes and subject status in pooled testing.

A key characteristic of COVID-19 is dynamic and uncertain prevalence rates, thus during testing the actual prevalence rate is unknown. To develop a pool size that is robust under prevalence uncertainty, we use the *Regret* measure (e.g., [11]), which can be computed for any pool size n and prevalence scenario p as follows:

$$Regret(n, p) = E[T(n, p)] - E[T(n(p), p)], \quad (2)$$

where $n(p)$ is the pool size that minimizes $E[T(n, p)]$ for prevalence scenario p (i.e., the minimizer of Eq. (1) under *perfect information* on p , the prevalence rate), and can be approximated as follows [11]:

$$n(p) \approx \operatorname{argmin}_{n \in (\lfloor \tilde{n}(p) \rfloor, \lceil \tilde{n}(p) \rceil, n_{max})} [E[T(n, p)]], \quad (3)$$

$$\text{where } \tilde{n}(p) = \sqrt{\frac{1}{p(Se + Sp - 1)}}.$$

In Eq. (3), we evaluate the expected tests, $E[T(n, p)]$, at the ceiling and floor of $\tilde{n}(p)$, and at the maximum allowable pool size, n_{max} , and select the value that yields a lower expected number of tests, in order to obtain an integer pool size, $n(p)$. A complex algorithm that derives the (exact) optimal pool size for scenario p is provided in our previous work [11].

The perfect information assumption used in Eq. (3), which implies a known prevalence rate, is obviously not realistic, and is utilized for the purpose of deriving a robust pool size when the prevalence rate is uncertain. Specifically, we determine a *robust pool size*, n^* , so as to minimize the expected *Regret* over the uncertainty set $S(P)$ of random variable P , that is:

$$n^* = \operatorname{argmin}_{n \in \mathbb{N}^+, n \leq n_{max}} [\sum_{p \in S(P)} Pr(P = p) \times Regret(n, p)]. \quad (4)$$

Utilizing the *Regret* objective in Eq. (4) requires the determination of a set of perfect information pool sizes *a priori* via Eq. (3). The *Regret* objective leads to a robust solution that is not overly conservative,

and a simple method for determining a robust pool size, compared to the analytically complex algorithm of [11].

In summary, to design a robust pooling strategy under prevalence uncertainty, we use Eqs. (3)-(4) with each risk group's respective parameters. Specifically, using Eq. (3), we first calculate a set of perfect information *pool sizes*, $n(p)$, for each scenario p in the range of P ; and then use Eq. (4) to determine the robust pool size for the group under prevalence uncertainty, n^* . This process is automated in our spread sheet, and the user needs to only input the problem parameters, including any user-specified probability distribution (discrete or continuous) for random variable P .

Sequential Pooling Design

Due to infection dynamics, we allow for updates to pooling design in each testing period (see, e.g., [29] for sequential pooling design for surveillance). To present our framework for sequential pooling design, we use index t to denote testing period $t \in Z^+$. At the beginning of each testing period t , we update the uncertainty set of random variable P based on the testing data obtained in periods $1, \dots, t-1$, compute each group's pool size in period t (Eqs. (3)-(4)), and use the updated pool sizes in testing period t . We repeat this process through the testing horizon.

False-negative Rate

We determine the expected false-negatives (FN s) per subject tested under pooling. When a pool contains infected specimen(s), FN (s) occurs if the pool tests negative, or the pool tests positive but the individual follow-up test for an infected specimen is negative, leading to:

$$E[FN^{Pool}] = (1 - Se^2)p. \quad (5)$$

On the other hand, for individual testing, when a subject is infected (with probability p), the test falsely provides a negative outcome (with probability $1-Se$), leading to:

$$E[FN^{Individual}] = (1 - Se)p. \quad (6)$$

Data

We demonstrate the efficiency of robust pooling for mass screening using published data for COVID-19, which is stratified into low- and high-risk groups based on symptom class. To this end, we construct each group's uncertainty set for the prevalence random variable to correspond to the 95% confidence interval (CI) from the dataset, with 100 equally-spaced prevalence scenarios (each with equal probability) within each uncertainty set; and develop a robust pooling design for each group (via Eqs. (3)-(4)).

We consider the PCR test for COVID-19 testing. In [9], ten pools of size 32 were constructed, where each pool contained one specimen infected with COVID-19 and 31 infection-free specimens. Using the PCR test, nine pools tested positive and one pool tested negative (due to dilution). Other studies suggest that the sensitivity of the PCR test for COVID-19 can be as low as 0.71 and as high as 0.98 [30]. As a result, we perform a one-way sensitivity analysis on the test sensitivity parameter, Se , over a wider range, of [0.70, 1.00], discuss the specific results for the published values (i.e., Se values of 0.71 and 0.98) in detail, and qualitatively discuss the results for other test sensitivity values. We also assume that the test has perfect specificity, i.e., $Sp=1$ [31], which is not subject to the dilution effect, and $n_{max} = 32$ [9].

Case Study Data

The case study focuses on mass screening of high- and low-risk subjects based on Iceland's COVID-19 testing dataset [7]. This dataset includes 63,134 subjects, and reports the number of subjects screened and the number of positive test outcomes for COVID-19 per day, based on testing data from two laboratories, which collectively conduct all COVID-19 screening in Iceland: 1) 21,576 high-risk subjects with "severe

symptoms and/or are at high risk of infection because of close contact with a diagnosed individual” [32], tested by the National University Hospital of Iceland (NUHI) between February-28-2020 and June-14-2020, leading to 1,628 positive-testing subjects; and 2) 41,558 low-risk subjects in the general population who have requested screening on a voluntary basis, tested by deCODE genetics between March-15-2020 and June-14-2020, leading to 182 positive-testing subjects.

We implement sequential pooling design, and compute the robust pool sizes for every week between March-05-2020 and June-14-2020 (high-risk group), and between March-19-2020 and June-14-2020 (low-risk group), i.e., after obtaining a number of days of testing data for each group. In particular, at the beginning of each week, we use the previous week’s testing data for each of the low- and high-risk groups, and the Wald’s method [33], to construct a 95% CI for each group’s prevalence:

$$L(\hat{p}) = \hat{p} - z_{0.975}\sqrt{\hat{p}(1 - \hat{p})/T}, \text{ and } U(\hat{p}) = \hat{p} + z_{0.975}\sqrt{\hat{p}(1 - \hat{p})/T},$$

where $\hat{p}$ is the point estimate, $z_{0.975}$ is the inverse of the CDF of the standard normal distribution at point 0.975, and T is the number of subjects tested in the previous week.

RESULTS

For illustrative purposes, Fig 1 reports the perfect information pool sizes (i.e., $(n(p))$ from Eq. (3) for prevalence scenario p) and expected number of tests per subject (simply, expected tests), for various prevalence scenarios between 0.01 and 0.292, and test sensitivity values ($Se=0.71, 0.98, 1$), considering that the test has perfect specificity. The expected tests under perfect information on prevalence rate provides a *lower bound* (LB) on the number of tests achievable via pooling, but is not realistic due to the inherent uncertainty in the prevalence rate (i.e., the lack of perfect information), which is typically the case for a highly contagious disease like COVID-19. Pooling is not more efficient than individual testing for prevalence scenarios above 0.292, due to the large number of individual follow-up tests. Relatedly, Fig 2 shows the expected tests ($E[T(n(p),p)]$) for various prevalence scenarios between 0.025 and 0.30, and test

sensitivity values of 0.98 (Fig 2 (a)) and 0.71 (Fig 2 (b)) over a range of pool sizes, and thus shows the impact of pool size on efficiency for each prevalence scenario and test sensitivity.

Fig 1. Perfect Information Pool Size and Expected Number of Tests per Subject versus Prevalence Rate for Various Test Sensitivity Values.

The perfect information pool size, $n(p)$, and expected number of tests per subject, $E[T(n(p), p)]$, for each prevalence scenario p between 0.01-0.292 for test sensitivity values, $Se=0.71, 0.98, 1$.

Based on the Iceland dataset [7], we develop a robust pooling design within a sequential framework for the high-risk (those tested by NUHI) and low-risk (those tested by deCODE) groups, updated every week during the study period. We consider that the weekly robust pool sizes are used each day of that week, and examine the daily testing results.

Fig 2. Expected Number of Tests per Subject versus Pool Size for Various Prevalence Rates for Test Sensitivity Values (a) $Se = 0.98$ and (b) $Se = 0.71$.

The expected number of tests per subject, $E[T(n(p), p)]$, for various prevalence scenarios p between 0.01–0.25 for test sensitivity values (a) $Se=0.98$ and (b) $Se = 0.71$ for pool sizes from 2-32.

Fig 3 depicts the weekly robust pool sizes (n^*) for two test sensitivity values ($Se=0.71, 0.98$) as well as the optimal pool size for each prevalence scenario (i.e., the perfect information pool size for each p , $n(p)$, from Eq. (3)) within the 95% CI on the weekly prevalence forecast (with each CI discretized into 100 equally-spaced prevalence scenarios), along with the actual weekly prevalence rate. Note that $n(p)$ decreases as the prevalence rate p increases (Eq. (3)). Thus, the robust pool size (n^*) is bounded from above by the optimal pool size for the lower limit of the CI on prevalence, and from below by the optimal pool size for the upper limit of the CI. We see that the robust pool size tends to be closer to the optimal pool size for the upper limit of the CI, that is, closer to the lower bound pool size. This figure also shows that as the test sensitivity increases, the robust pool size decreases.

Fig 3. Weekly Robust Pool Size and 95% Confidence Interval Pool Sizes for (a) High-risk and (b) Low-risk Groups for Two Test Sensitivity Values, and Actual Weekly Prevalence Rates for (c) High-risk and (d) Low-risk Groups.

The robust pool size, n^* , for each week along with pool sizes corresponding to the 95% confidence interval of the prevalence forecast for (a) high-risk and (b) low-risk groups for test sensitivity values, $Se=0.71$ and 0.98, and the actual weekly prevalence rates for (c) high-risk and (d) low-risk groups.

Fig 4 reports the daily number of tests for individual testing (i.e., the actual number of tests conducted per day in the dataset), and the expected number of tests for robust pooling (i.e., using n^*) and the perfect information lower bound (LB , i.e., using $n(p)$) for (a) high-risk and (b) low-risk subjects for a test sensitivity of 0.98, and (c) high-risk and (d) low-risk subjects for a test sensitivity of 0.71. Recall that the perfect information lower bound is unattainable, because the prevalence rate is uncertain, and a prevalence rate forecast is required due to changing disease dynamics. Considering test sensitivity values of 0.71 and 0.98, the total reduction in the number of tests, over individual testing, is 61.9% and 54.2% respectively for robust pooling (62.7% and 55.6% respectively under perfect information) for the high-risk group during the period of March-05-June-14; and 90.2% and 88.5% respectively for robust pooling (91.0% and 89.5% respectively under perfect information) for the low-risk group during the period of March-19-June-14.

Fig 4. Daily Expected Number of Tests for the High-risk Group for Test Sensitivity Values of (a) 0.98 and (c) 0.71 and for the Low-risk Group for Test Sensitivity Values of (b) 0.98 and (d) 0.71.

The daily expected number of tests required for the (a) high-risk group with $Se=0.98$, (b) low-risk group with $Se=0.98$, (c) high-risk group with $Se=0.71$, and (d) low-risk group with $Se=0.71$ for the perfect information lower bound, robust pooling, and individual testing.

In the above analysis, we use a maximum allowable pool size, n_{max} , of 32, to limit the dilution effect, which is set based on preliminary studies (see Introduction), hence we next study how the value of n_{max} affects the results. Table 2 reports the efficiency of robust pooling, that is, the percent reduction in the number of tests that can be achieved via robust pooling compared to individual testing, for lower values of n_{max} , and for test sensitivity values of $Se=0.71$ and 0.98, and the high- and low-risk groups. As expected, very low n_{max} values impact the efficiency of testing more for the low-risk group compared to the high-risk group, because the latter group already uses small pool sizes due to their high prevalence.

Table 2. The Percent Reduction in the Expected Number of Tests via Robust Pooling Compared to Individual Testing for High-risk and Low-risk Groups and Two Test Sensitivity Values for Various Maximum Allowable Pool Sizes, n_{max} .

	High-risk Group	Low-risk Group
--	-----------------	----------------

Maximum Allowable Pool Size (n_{max})	$Se=0.98$	$Se=0.71$	$Se=0.98$	$Se=0.71$
32	54.2%	61.9%	88.5%	90.2%
30	54.1%	61.8 %	88.5%	90.1%
20	54.0%	61.7%	88.2%	89.8%
10	53.4%	61%	86.2%	87.4%
5	51.3%	58.8%	79.8%	80.8%

By requiring, on average, fewer tests per subject, robust pooling can substantially expand testing capacity. To illustrate this concept, we select a week during the study period, i.e., the week spanning April-2 to April-8 of 2020, during which 7,967 low-risk subjects were tested via 7,967 tests, of which 54 subjects tested positive. We report the expected number of subjects that could be tested using the 7,967 tests under the different strategies. For this measure, the perfect information lower bound on the expected tests per subject provides a *perfect information upper bound (UB)* on the expected number of subjects tested. Fig 5 (a) displays the expected number of low-risk subjects that could be tested via 7,967 tests for the individual testing and robust pooling (n^*) strategies, and the perfect information upper bound (UB), for a test sensitivity range $[0.70, 1.00]$. For each pooling strategy, the number of low-risk subjects screened reduces as test sensitivity increases. For example, as test sensitivity increases from 0.70 to 1.00, the number of low-risk subjects screened reduces from 66,010 to 54,898 under perfect information, and from 60,140 to 49,822 under robust pooling. This follows because a higher test sensitivity implies a higher likelihood that a pool containing an infected specimen will test positive, necessitating further individual testing for all subjects in the pool. As a result, optimal pool sizes are non-increasing in test sensitivity (Fig 1).

Fig 5. Expected Number of Low-risk Subjects Screened with 7,967 Tests (a) under Different Strategies and (b) under Different Assumed Test Sensitivities, versus True Test Sensitivity, and the Expected Number of (c) False-negative cases and (d) Missed and False-negative Cases versus Test Sensitivity.

The expected number of low-risk subjects screened with 7,967 tests (a) for the perfect information upper bound, robust pooling, n^* , and individual testing versus test sensitivity, (b) assuming a prevalence rate, p , of 0.0065 and a test sensitivity, Se , of 0.70, 0.85, and 1 versus true test sensitivity, (c) FN s (out of the 7,967 low-risk subjects tested under both strategies), and (d) missed cases plus FN s for robust pooling and individual testing (out of 52,173 subjects, which corresponds to the expected number of low-risk subjects tested via robust pooling) versus test sensitivity.

The sensitivity of the test may not be known with certainty. To examine the effect of the test sensitivity estimate, we calculate the robust pool sizes for the low-risk group for the week spanning April-2 to April-8, using three assumed test sensitivity values, of 0.70, 0.85, and 1. Fig 5 (b) depicts the expected number of low-risk subjects that would be screened with 7,967 tests under a range of true test sensitivity values, Se , in $[0.70, 1]$. At the extremes, when the true sensitivity is 0.70 and the pooling strategy is based on a sensitivity of 1, 837.2 fewer subjects can be tested in expectation compared to using the true sensitivity; and when the true sensitivity is 1 and the pooling strategy is based on a sensitivity of 0.70, 540.4 fewer subjects can be tested in expectation.

Next, we study the effect of pooling on the expected number of FN s for various test sensitivity values. While pooling increases the number of FN s over individual testing for the same number of subjects (Eqs. (5)-(6)), it also allows for expanded testing, thus reducing potential *missed cases* over individual testing (i.e., infected subjects not tested by individual testing, who could have been tested under pooling). To illustrate this point, we consider again the low-risk subjects tested during the week spanning April-2 to April-8 in the Iceland dataset. Fig 5 (c) displays the number of FN s for both robust pooling and individual testing (for the 7,967 subjects tested under both strategies), as a function of test sensitivity, while Fig 5 (d) displays the FN s for robust pooling, and the sum of FN s and missed cases for individual testing, for 52,173 subjects that would have been screened under robust pooling (with 7,967 tests).

DISCUSSION

In this paper, we provide a robust pooled testing approach to screen for COVID-19 and demonstrate its value for overcoming difficulties associated with COVID-19 testing. These difficulties include dynamic disease prevalence, which leads to high uncertainty in current prevalence, a wide range of possible test sensitivity values, different risk groups, and limited testing resources. For illustrative purposes, we use

the Iceland dataset [7], in which the risk groups are based on whether the subject was tested by NUHI (high-risk due to symptoms and/or potential contacts) or by deCODE (low-risk, voluntary testing). The proposed robust pooling strategy significantly reduces the expected number of tests required to accomplish the screening conducted in Iceland for COVID-19 compared to individual testing, and the differences are more pronounced for the low-risk group, see Figs 3. Overall, the reductions in the number of tests are 54.2% to 61.9% for the high-risk group, for test sensitivity of 0.98 and 0.71, respectively, and 88.5% to 90.2% for the low-risk group, for test sensitivity of 0.98 and 0.71, respectively. These reductions are based on weekly forecasted point estimates and confidence intervals for the prevalence. The robust pooling strategy based on these forecasts does nearly as well as having perfect prevalence information (for which the respective reductions are 55.6% to 62.7% and 89.5% to 91.0%, for the high-risk and low-risk groups, for test sensitivity of 0.98 and 0.71), which, of course, we only have in hindsight. Thus, the proposed robust pooling strategy can be used to substantially expand COVID-19 testing capacity. This expansion is more pronounced at lower prevalence rates (due to fewer follow-up tests and larger pool sizes).

Dilution is an important issue when considering pooling, especially for large pools. In our models, we use a maximum allowable pool size, n_{max} , to limit the dilution effect, which is a viable practice in pooled testing, especially when lab-based data on the magnitude of the dilution effect for different pool sizes is scarce (as is the case with COVID-19). Informed by the research that describes pool sizes for which pooling does not significantly affect the test sensitivity for the PCR test for COVID-19 (see Introduction), we consider that n_{max} is 32. Our analysis (Table 2) indicates that reducing n_{max} , and thus further reducing the potential for dilution, does decrease the efficiency of robust pooling, but even at low values of n_{max} , e.g., $n_{max}=5$, robust pooling is still much more efficient than individual testing. As an alternative, if comprehensive data on the dilution effect for COVID-19 become available, one can derive a pooled sensitivity function, as a function of pool size, and use this function in the analytical expressions (Eqs. (1),

(3), (5)) to model dilution (e.g., see [21, 34]). This will be an important future research direction once such data become available.

We demonstrate the benefits of pooling in more detail using the week of April-2 to April-8 of 2020 for the low-risk group (with 7,967 low-risk subjects individually tested) considering a test sensitivity of 0.85. In contrast to the 7,967 tests required for individual testing, robust pooling uses a pool size of 14, thus requiring $\left\lceil \frac{7,967}{14} \right\rceil = 570$ pools. Given the true prevalence of 0.68% (calculated from the dataset), the probability that any pool will test positive (thus requiring individual follow-up tests for the subjects in the pool) is given by $Se(1 - (1 - p)^n) = 0.07722$ (see Table 1). Therefore, for the 570 pools, we have, on average, $0.07722 \times 570 \times 14 = 616.2$ individual follow-up tests. Thus, with pooling, the 7,967 subjects would require 1,186.2 ($=570+616.2$) tests in expectation, an 85.1% reduction over individual testing. This reduction is closely related to a cumulative reduction in testing time, which we illustrate using the same week assuming a PCR testing machine that has a capacity of 96 tests, that is, 96 tests can run at the same time [25], and assuming that a test run takes 2 hours. Under individual testing, the number of machine runs is $\left\lceil \frac{7,967}{96} \right\rceil = 83$, thus requiring 166 hours to complete testing. For the pooled strategy, $\left\lceil \frac{570}{96} \right\rceil = 6$ runs are required for the pools, plus $\left\lceil \frac{616}{96} \right\rceil = 7$ runs for the expected individual follow-up tests, thus requiring 26 hours of testing in expectation, compared to 166 hours for individual testing.

Next we discuss the effect of test sensitivity on pooling by examining Fig 5, which again considers the week of April-2 to April-8 of 2020 for the low-risk group of 7,967 subjects. If the 7,967 tests were used in the pooled strategy, between 49,822 (at a sensitivity of 1.00) and 60,140 (at a sensitivity of 0.70) subjects could have been tested (in expectation), see Fig 5 (a). This is a considerable difference, and as Fig 5 (b) depicts, even if the pooling strategy is derived under the wrong test sensitivity, pooling still does well, e.g., at a true sensitivity of 0.70, a strategy derived based on perfect sensitivity (of 1.00) tests 59,122 subjects, equivalently, 89.6% of those that could be tested if the true test sensitivity were known (i.e., derived using

a test sensitivity of 0.70). This reduction is due to using pools that are too small (13 versus 15 under the true sensitivity of 0.70), which results in loss of efficiency, but this is somewhat mitigated by a lower probability that subjects in a pool will need individual follow-up testing. Fig 5 (b) suggests that a good strategy to handle uncertainty in test sensitivity is to pick the mid-point of the potential sensitivity range.

Next, we compare the *FNs* under the different strategies. *FN* rate is larger under pooling than individual testing (Eqs (5)-(6)). For instance, Fig 5 (c) shows that for a test sensitivity of 0.70, individual testing has 16 expected *FNs*, while pooling has 28 (for the 7,976 subjects). Of course, pooling uses many fewer tests for these 7,976 subjects, thus allowing for expanded testing. Considering this expanded testing population, Fig 5 (d) shows that at a test sensitivity of 0.70, there are 171 *FNs* from pooling, and 16 *FNs* plus 335 missed cases from individual testing. As the test sensitivity increases, the *FNs* decrease fairly fast compared to the reduction in missed cases.

As we demonstrate in this study, robust pooling can substantially expand the testing capacity, and allow the testing of many more subjects for COVID-19. However, to achieve the maximum benefits of pooled testing, it is important to use pool sizes that are customized for the different risk groups in the population, and pool sizes that are designed to hedge against prevalence that is dynamic and uncertain; these are the key features of the robust pooling approach developed in this paper.

Limitation

One limitation is the complex nature of the test sensitivity function for the PCR test for COVID-19. PCR tests tend to have high sensitivity, but for COVID-19, the clinical sensitivity may be lower [35] (due, for

example, to low quality specimens), and the reasons for this lower clinical sensitivity, and ways for improving the clinical sensitivity, have implications for pooling.

Acknowledgments

We would like to thank the Academic Editor, Dr. Baochuan Lin, along with Dr. Robert Rubin, and three anonymous reviewers for providing excellent comments that improved the analysis in this manuscript.

This material is based upon work supported in part by the National Science Foundation under Grant #1761842. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

1. Global cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University 2020 [cited 2020]. Available from: <https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>.
2. WHO Director-General's opening remarks at the media briefing on COVID-19-16 March 2020: World Health Organization; 2020 [cited 2020]. Available from: <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---16-march-2020>.
3. Adalja AA, Toner E, Inglesby TV. Priorities for the US Health Community responding to COVID-19. JAMA. 2020. doi: 10.1001/jama.2020.3413.
4. Bai Y, Yao L, Wei T, Tian F, Jin D-Y, Chen L, et al. Presumed asymptomatic carrier transmission of COVID-19. JAMA. 2020. doi: 10.1001/jama.2020.2565.
5. Zastrow M. South Korea is reporting intimate details of COVID-19 cases: has it helped? Nature. 2020. doi: 10.1038/d41586-020-00740-y.
6. Bamforth E. Iceland tests anyone for coronavirus, shows significant amount are asymptomatic 2020 [cited 2020]. Available from: <https://www.cleveland.com/news/2020/04/iceland-tests-anyone-for-coronavirus-shows-significant-amount-are-asymptomatic.html>.
7. COVID-19 in Iceland – Statistics: The Directorate of Health Iceland; 2020 [cited 2020]. Available from: <https://www.covid.is/data>.
8. Dorfman R. The detection of defective members of large populations. The Annals of Mathematical Statistics. 1943;14(4):436-40.
9. Yelin I, Aharony N, Shaer Tamar E, Argoetti A, Messer E, Berenbaum D, et al. Evaluation of COVID-19 RT-qPCR test in multi-sample pools. Clinical Infectious Diseases. 2020. doi: 10.1093/cid/ciaa531.

10. Aprahamian H, Bish DR, Bish EK. Optimal risk-based group testing. *Management Science*. 2019;65(9):4365-84.
11. Aprahamian H, Bish DR, Bish EK. Optimal group testing: Structural properties and robust solutions, with application to public health screening. *INFORMS Journal on Computing*. 2020; 32(4):895-911.
12. McMahan CS, Tebbs JM, Bilder CR. Informative Dorfman screening. *Biometrics*. 2012;68(1):287-96.
13. Wein LM, Zenios SA. Pooled testing for HIV screening: capturing the dilution effect. *Operations Research*. 1996;44(4):543-69.
14. Aragón-Caqueo D, Fernández-Salinas J, Laroze D. Optimization of group size in pool testing strategy for SARS-CoV-2: A simple mathematical model. *Journal of Medical Virology*. 2020.
15. Bilder CR, Iwen PC, Abdalhamid B. Pool size selection when testing for severe acute respiratory syndrome coronavirus 2. *Clinical Infectious Diseases*. 2020.
16. Cherif A, Grobe N, Wang X, Kotanko P. Simulation of pool testing to identify patients with coronavirus disease 2019 under conditions of limited test availability. *JAMA Network Open*. 2020;3(6):e2013075-e.
17. Lohse S, Pfuhl T, Berkó-Göttel B, Rissland J, Geißler T, Gärtner B, et al. Pooling of samples for testing for SARS-CoV-2 in asymptomatic people. *The Lancet Infectious Diseases*. 2020.
18. Singh AK, Nema RK, Joshi A, Shankar P, Nema S, Raghuwanshi A, et al. Evaluation of pooled sample analysis strategy in expediting case detection in areas with emerging outbreaks of COVID-19: A pilot study. *PLOS ONE*. 2020;15(9):e0239492.
19. Becker MG, Taylor T, Kiazzyk S, Cabiles DR, Meyers AFA, Sandstrom PA. Recommendations for sample pooling on the Cepheid GeneXpert® system using the Cepheid Xpert® Xpress SARS-CoV-2 assay. *PLOS ONE*. 2020;15(11):e0241959.
20. Lim KL, Johari NA, Wong ST, Khaw LT, Tan BK, Chan KK, et al. A novel strategy for community screening of SARS-CoV-2 (COVID-19): Sample pooling method. *PLOS ONE*. 2020;15(8):e0238417.
21. Pilcher CD, Westreich D, Hudgens MG. Group testing for Sars-Cov-2 to enable rapid scale-up of testing and real-time surveillance of incidence. *The Journal of Infectious Diseases*. 2020.
22. Wacharapluesadee S, Kaewpom T, Ampoot W, Ghai S, Khamhang W, Worachotsueptrakun K, et al. Evaluating the efficiency of specimen pooling for PCR-based detection of COVID-19. *Journal of Medical Virology*. 2020.
23. Hanel R, Thurner S. Boosting test-efficiency by pooled testing for SARS-CoV-2—Formula for optimal pool size. *PLOS ONE*. 2020;15(11):e0240652.
24. Furstenau TN, Cocking JH, Hepp CM, Fofanov VY. Sample pooling methods for efficient pathogen screening: Practical implications. *PLOS ONE*. 2020;15(11):e0236849.
25. Carter LJ, Garner LV, Smoot JW, Li Y, Zhou Q, Saveson CJ, et al. Assay techniques and test development for COVID-19 diagnosis. *ACS Publications*; 2020.
26. Won J, Lee S, Park M, Kim TY, Park MG, Choi BY, et al. Development of a laboratory-safe and low-cost detection protocol for SARS-CoV-2 of the Coronavirus disease 2019 (COVID-19). *Experimental Neurobiology*. 2020;29(2):107.
27. Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLOS ONE*. 2020;15(3):e0230405.

28. Petropoulos F, Makridakis S. Forecasting the novel coronavirus COVID-19. PLOS ONE. 2020;15(3):e0231236.
29. Nguyen NT, Bish EK, Aprahamian H. Sequential prevalence estimation with pooling and continuous test outcomes. Statistics in Medicine. 2018;37(15):2391-426.
30. Watson J, Whiting PF, Brush JE. Interpreting a COVID-19 test result. BMJ. 2020;369. doi: 10.1136/bmj.m1808.
31. He J-L, Luo L, Luo Z-D, Lyu J-X, Ng M-Y, Shen X-P, et al. Diagnostic performance between CT and initial real-time RT-PCR for clinically suspected 2019 coronavirus disease (COVID-19) patients outside Wuhan, China. Respiratory Medicine. 2020:105980.
32. Stock JH, Aspelund KM, Droste M, Walker CD. Estimates of the undetected rate among the SARS-CoV-2 infected using testing data from Iceland: Chicago Booth; 2020 [cited 2020]. Available from: http://www.igmchicago.org/wp-content/uploads/2020/04/Covid_Iceland_v10.pdf.
33. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. Statistics in Medicine. 1998;17(8):857-72.
34. Nguyen NT, Aprahamian H, Bish EK, Bish DR. A methodology for deriving the sensitivity of pooled testing, based on viral load progression and pooling dilution. Journal of Translational Medicine. 2019;17(1):252. doi: 10.1186/s12967-019-1992-2.
35. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology. 2020:200642.