Title: Fighting COVID-19 with Flexible Testing: Models and Insights

Authors: Jiali Yu[1], Zuo-Jun (Max) Shen[2]*.

Affiliations:

[1]Tsinghua-Berkeley Shenzhen Institute (TBSI), Tsinghua University

[2]Department of Industrial Engineering and Operations Research, Department of Civil and Environmental Engineering, TBSI, University of California, Berkeley

*Correspondence to: maxshen@berkeley.edu.

Abstract:

Increasing testing capacities plays a substantial role in safely reopening the economy and avoiding a new wave of COVID-19. Pooled testing can expand testing capabilities by pooling multiple individual samples, but it also raises accuracy concerns. In this study, we propose a flexible testing strategy that adapts pool sizes to epidemic dynamics. We present an analytical method to calculate the optimal pool size and the prevalence threshold between individual and pooled testing. Incorporating an epidemic model, we show pooled testing is more effective in containing epidemic outbreaks and can generate more reliable test results than individual testing because the reliability of test results is relevant to both testing methods and prevalence. Our study is the first to evaluate the interplay between pooled testing and a rapidly evolving outbreak to the best of our knowledge. Our results allay accuracy concerns about pooled testing and provide theoretical supports to empirical studies.

Introduction

Testing plays a critical role in containing the COVID-19 outbreak and lifting containment measures, which have caused enormous economic losses and posed health risks, particularly among disadvantaged populations. However, increasing the testing capability to meet the demand and improving the reliability of test results remains to be challenging tasks.

Pooled testing is to test samples drawn from multiple individuals as a pooled sample, also called a pool[1]. If a pool tests negative, everyone in this pool is considered negative. If the result is positive, which indicates that at least one individual in this pool is infected, then all the samples need to be retested individually. Prevalence, test accuracy, and pool size all have great impacts on pooled testing. When a large proportion of pools test negative, pooled testing can clear groups of patients using fewer assays. Conversely, if most pools yield positive results, individual testing outperforms pooled testing.

Although pooled testing provides an alternative to expand the testing capability, policymakers have remained conservative about pooled testing, such as mandating a small pool size when the positive result rate is low. For example, the Indian Council of Medical Research suggested pooling five specimens for COVID-19 testing in communities where the positive rate was less than 5%[2]. The U.S. Food and Drug Administration (FDA) has approved only one COVID-19 diagnostic test for pooling up to four individual samples[3]. However, recent empirical studies have shown that the pool size could be expanded to 32 nasopharyngeal swabs while maintaining acceptable test accuracy[4] and pools of 20 saliva samples help control the epidemic when the prevalence is less than 1%[5].

Recent studies mainly focused on investigating efficiency gains and the accuracy losses of different pooled algorithms[4-17], such as multi-stages and multi-dimensions procedures. In this study, we focus on the Dorfman's classic two-stage procedure which is the most utilized algorithm in practice because of its simplicity and high efficiency. Studies on pooled testing typically assume a given prevalence, approximate the theoretical pool size without considering testing accuracy, or fail to include testing capacity constraints. However, during a pandemic, the interplay between testing and fast-changing prevalence remains unclear; newly developed tests have uncertain accuracy; testing capacity becomes insufficient to meet surging demand. No studies to date, however, have evaluated the overall performance of pooled testing in a rapidly evolving outbreak nor given the analytical solution for the optimal pool size to maximize the effectiveness of pooled testing.

We aim to provide a broad picture of testing in the pandemic from a public health policy perspective. We translate scientific evidence into a mathematical decision-making tool, a flexible testing strategy that adapts the pool size to the epidemic dynamics. We present an analytical

method to calculate the optimal pool size and the prevalence threshold to determine when testing strategies must be changed. We use an epidemic model to compare flexible and individual testing in population screening. Our results contradict the common belief that pooled testing is less accurate than individual testing and that only highly accurate tests facilitate the safe reopening of the economy. We show that flexible testing is more cost-effective for containing the outbreaks and can generate more reliable results than individual testing. The higher capacity and frequency achieved by appropriately adopting a pooled testing strategy or newly developed tests ease the containment measures and prevent infections, even if tests have a low ability to detect the virus.

Results

Flexible testing strategy. The objective of testing during a pandemic is to identify and remove infected patients from the transmission chain as soon as possible. We measure the overall efficiency and effectiveness of testing by the expected number of tests required to identify an infected patient, defined as the cost-performance ratio (CPR). CPR depends on the prevalence p, pool size n, test sensitivity S_e (true positive rate), and specificity S_p (true negative rate) as follows (see methods for details on the derivation of the CPR):

$$CPR(n) = \begin{cases} \dfrac{1}{S_e p}, & n = 1, \\ \dfrac{1}{S_e^2 p}\left[\dfrac{1}{n} + S_e - (S_e + S_p - 1)(1-p)^n\right], & \text{otherwise.} \end{cases} \tag{1}$$

During the pandemic, the flexible testing strategy with an optimal pool size n^* that minimizes the CPR follows a threshold policy, as follows:

$$n^* = \begin{cases} 1, & p \geq \underline{p}, \\ n_0, & p < \underline{p}, \end{cases} \tag{1}$$

where $\underline{p} \equiv 1 - e^{-(S_e+S_p-1)e^{-1}}$, $n_0 \equiv \frac{2}{\ln(1-p)} W_0\left(-\frac{1}{2}\left(\frac{\ln\left(\frac{1}{1-p}\right)}{S_e+S_p-1}\right)^{\frac{1}{2}}\right)$. $W_0(\cdot)$ is the principal branch of Lambert W function[18,19] (see supplementary material for more details).

We provide the analytical solutions of the optimal pool size n^* and the prevalence threshold between pooled and individual testing $\underline{p}$. When the prevalance is higher than the threshold, individual testing is the optimal strategy. Otherwise, pooled testing with pool size n_0 is preferred. In supplementary material, we prove the optimal pool size and present a more realistic formula for calculating the optimal pool size that includes an integer constraint and an upper limit on the pool size to account for technological limitations and dilution effects.

Moreover, we find that n^*, $CPR(n^*)$, and $\underline{p}$ have the simple monotonicity that offers clear directions for decision-makers to modify testing strategies to identify an infected patient with minimum tests. The results are as follows (see proof in supplementary material):

1) The prevalence threshold increases in sensitivity and specificity, as shown in Figure 1A. If the testing sensitivity and specificity are 70% and 95%, the prevalence threshold between individual and pooled testing is 21.2%. If tests were perfect, the prevalence threshold is up to 30% and we can apply pooled testing in areas with higher prevalence.
2) The optimal pool size monotonically decreases with the increasing prevalence, sensitivity, and specificity, as shown in Figure 1B, D. The optimal pool size varies more with prevalence than with test sensitivity and specificity, especially when prevalence is lower than 1%. When the prevalence is higher than 1%, the optimal pool size changes no more than one sample for 10% fluctuation of sensitivity or specificity.
3) The CPR of optimal pool size decreases with increasing prevalence, sensitivity, and specificity, as shown in Figure 1C, E. As more people get infected, or we have more accurate testing technology, we can identify infected patients with fewer tests.

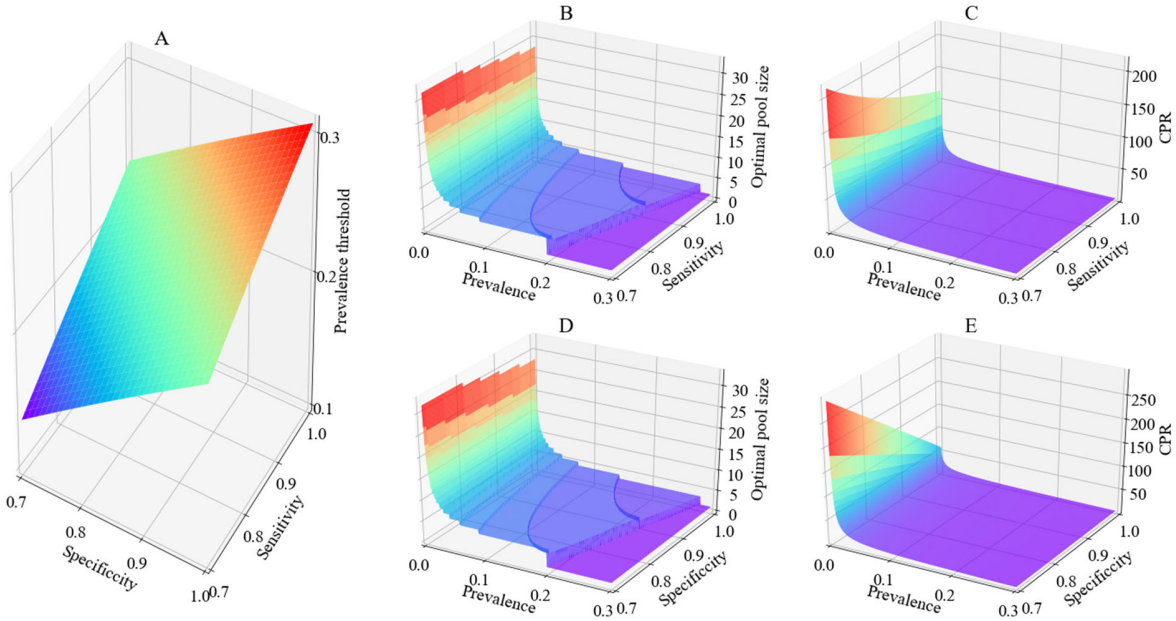

Figure 1. Monotonicity of the prevalence threshold, optimal pool size, and CPR.

(A): The prevalence threshold is monotonically increasing in sensitivity and specificity. (B) and (C): the optimal pool size and the corresponding CPR is monotonically decreasing in prevalence and sensitivity, assuming the specificity is 90%. (D) and (E): the optimal pool size and the corresponding CPR is monotonically decreasing in prevalence and specificity, assuming the sensitivity is 90%. The prevalence ranges from 0.15%–30%.

Application of flexible testing strategy. We illustrate the application of flexible testing using testing data from the United States. Figure 2 shows the seven-day average positive rate, the optimal pool sizes, and the CPRs of different testing strategies for 50 U.S. states and Washington, D.C. Based on testing data from the COVID Tracking Project https://covidtracking.com/, between October 28–November 2, 2020, five states reported a seven-day average positive rate higher than 20%, namely, South Dakota (50.6%), Iowa (37.4%), Kansas(36.0%), Idaho (33.4%), and Wyoming(30.4) (figure 2A). Note that positive rates are higher than prevalence rates where testing capacity is limited because hospitals prioritize testing individuals with symptoms or higher risk of infection. For simplicity, we use test positive rates to represent the prevalance in the tested population.

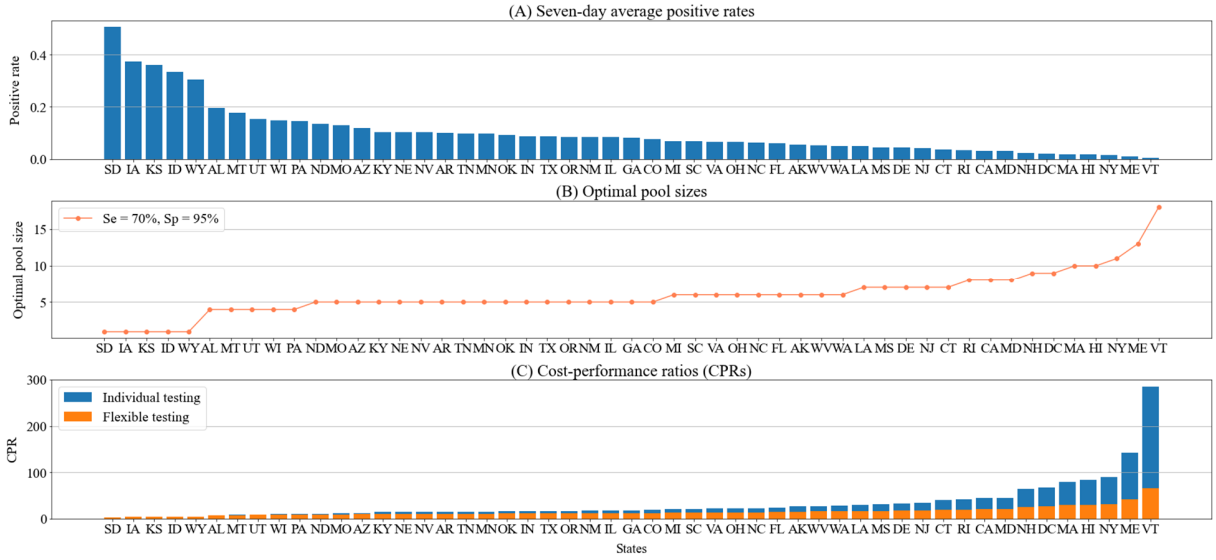

Figure 2: COVID-19 testing in the United States.

(A) The seven-day average positive rate of 50 U.S. states and Washington, D.C., between October 28–November 2, 2020. (B) The optimal pool size of the flexible testing strategy. (C) The CPRs of the flexible and individual testing strategy. We assume that sensitivity and specificity are 70% and 95%, respectively.

If the test sensitivity is 70% and specificity is 95%, which are reasonable estimates of main diagnostic test accuracy (involving a nasopharyngeal swab)[20], 45 states and Washington, D.C., where positive rates are lower than 21.2%, can apply pooled testing to increase testing capabilities. The maximum pool size approved by the U.S. FDA is four individual samples, which is the optimal pool size for five states where positive rates are 12.5%–21.2%. 40 states and Washington, D.C., with lower than 12.5% positive rates, can implement larger pool sizes than four to increase testing capabilities further. The current limitation on the maximum pool size is too conservative and restraints the effectiveness of pooled testing. When the prevalence is less than 1%, flexible testing saves more than 70% tests to identify an infected patient than individual testing. Taking Vermont as an example, the positive rate is 0.5%. Flexible testing with the pool size of 18 takes 66 tests to identify an infected patient, while individual testing takes 286 tests. Flexible testing can alleviate the current shortage of coronavirus tests.

Comparison of individual and flexible testing strategies. A flexible testing strategy should consider the prevalence rate, which changes over time among different groups. In practice, we can implement targeted testing tactics by categorizing people into different risk groups based on their local prevalence rates, symptoms, and exposure histories. In this study, we focus on the change of the prevalence rate over time. We use a susceptible–infected–removed (SIR) model to compare individual and flexible testing strategies in population screening considering different containment measures, test sensitivities, specificities, and capacities.

In baseline scenarios, the initial prevalence is 1%. The testing sensitivity and specificity are 73.3% and 98.6%, respectively. Limited testing capacity allows individually test all population within 100 days. Containment measures suppress the reproductive number to 1.5. As shown in Figures 3A and 3B, the total infection rate increases exponentially without testing and more drastic containment measures. Individual testing only flattens the curve. Flexible testing contains the outbreak more effectively and causes fewer infections than individual testing.

Because the prevalence is lower than the threshold, flexible testing adopts pooled testing to increase testing capabilities. As the prevalence in pooled testing decreases (Figure 2B), the optimal pool size increases until it reaches the upper limit (Figure 2C). Pooled testing increases testing capabilities make it possible to test people more frequently. In baseline scenarios, pooled testing tests all population five times within the initial 100 days, while individual testing only tests all population once. Although pooled testing provides more results than individual testing (Figure 2D), it generates fewer false positive results (Figure 2E), avoiding unnecessary treatments and isolations.

We measure the reliability of the test results by positive predictive values (PPVs) and negative predictive values (NPVs). PPVs and NPVs incorporate information about the prevalence, pool size, test sensitivity, and specificity (see the supplementary material for details of the PPV, NPV, sensitivity, and specificity). High PPVs and NPVs correspond to reliable positive and negative results, respectively.

Contrary to the common belief that pooled testing is less accurate than individual testing, we show that pooled testing can generate more reliable results because the reliability of test results is relevant to testing methods and prevalence. More precisely, pooled testing can be associated with higher PPVs during a pandemic (Figure 2F) because each positive sample is assayed twice,

include a pooled test and an individual retest. On the other hand, the NPVs of pooled testing are initially lower than those of individual testing, but they quickly increase and become higher than those of individual testing (Figure 2G) when the prevalence rate under pooled testing is significantly lower than that under individual testing (Figure 2B). However, research that assumes a fixed prevalence shows that pooled testing always leads to lower NPVs than those of individual testing [21].

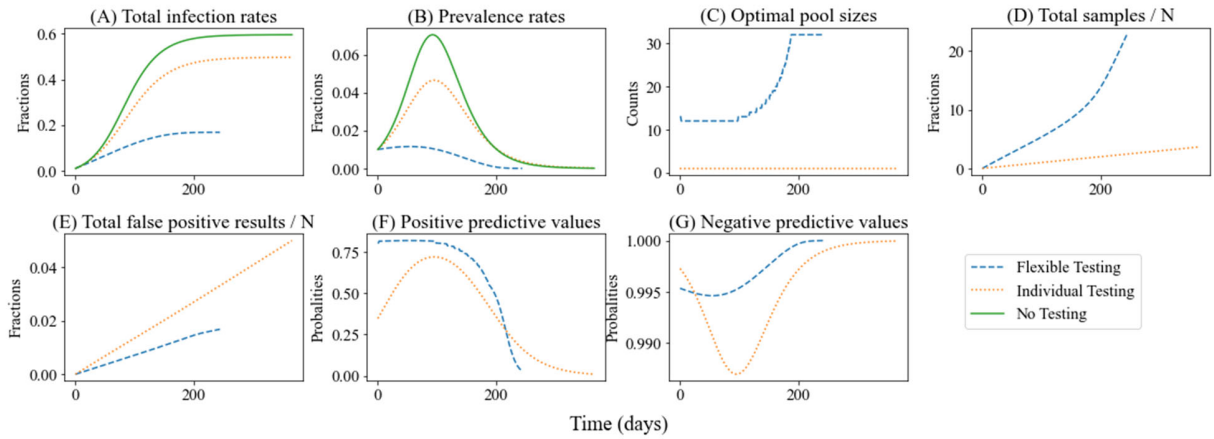

Figure 3: Epidemic dynamics and test performances of individual and flexible testing.

During the initial stage of the outbreak, we faced great challenges in designing, developing, and expanding the diagnostic tests. Drastic containment measures, such as a complete lockdown, effectively suppressed the total infection rate but halted most economic activities. Flexible testing offers an alternative to ramp up testing capabilities within a short time and accelerate the suppression of outbreaks (Figure 4A, B). With the lifting of containment measures (Figure 4C, D), the effect of flexible testing approaches that of individual testing, only mitigating the outbreak. If we let the virus spread with few containment measures, the total infection rate will increase exponentially and surpass 90%, and limited tests have little effect on containing the outbreak (Figure 4D). Containment measures and testing should be coordinated to reduce both infections and economic losses. To keep total infection rates lower than the same level, flexible testing requires less stringent containment measures than individual testing.

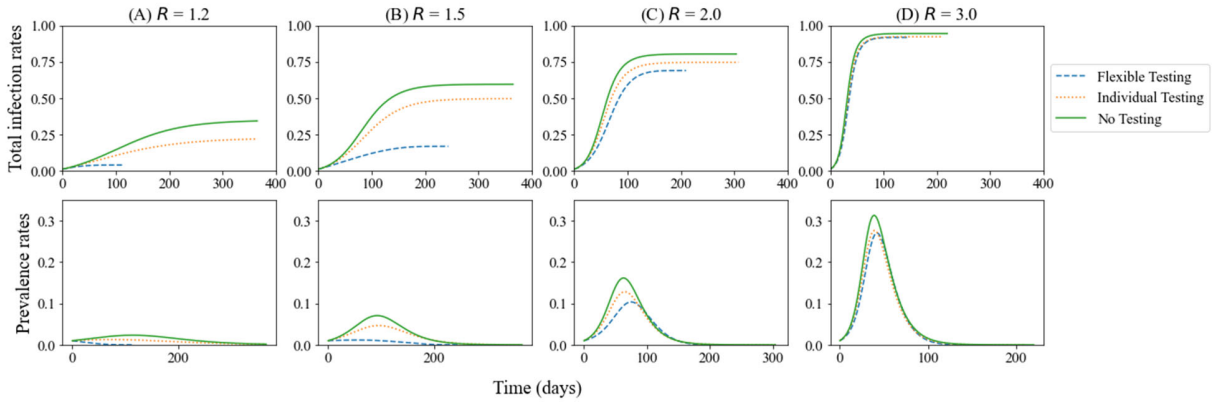

Figure 4: Effects of testing under different containment measures.

Smaller R implies more stringent containment measures. The top and bottom show total infection rates and prevalence rates, respectively.

Test sensitivity and specificity affect epidemic dynamics differently under different testing strategies, as shown in Figure 5. Test sensitivity impacts epidemic dynamics and test performances for both individual and pooled testing. Higher test sensitivity reduces total infection rates (Figure 5A) and increases NPVs (Figure 5B), which imply more reliable negative results.

Specificity has a negligible influence on epidemic dynamics under individual testing (Figure 5A top). However, we show that specificity is also important for pooled testing. Lower specificity causes more false positive results of pooled samples in the first stage and then leads to more unnecessary individual retests. Increasing testing sensitivity or specificity can dramatically strengthen the effectiveness of pooled testing on containing outbreaks (Figure 5A bottom). Moreover, higher test specificity reduces false positive results (Figure 5D) and increases PPVs (Figure 5C), which imply more reliable positive results.

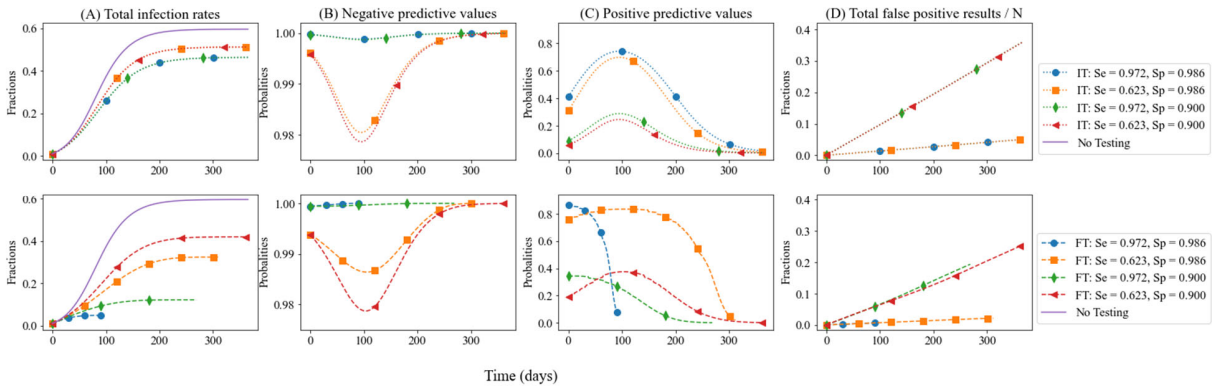

Figure 5: Comparison of different testing strategies and technologies.

The top and bottom rows show individual testing (IT) and flexible testing (FT), respectively. Testing technologies vary in sensitivity S_e and specificity S_p, as indicated.

The testing capacities achieved with different technologies vary significantly. In-home rapid tests enable the possibility of screening a population frequently. When the initial prevalence is 5% and there are few containment measures (R=3), individual testing twice a week can reverse the outbreak and keep the total infection rate no more than 20%, even using an inaccurate testing technology (sensitivity: 62.3% and specificity: 90%) (Figure 6A, B). This result contradicts the common belief that highly accurate tests are prerequisites to safely reopening the economy. As the testing frequency increases, inaccurate technologies reduce infections (Figure 6A, B); however, FPs rapidly increase because of the low specificity (Figure 6C).

Frequent testing, such as the twice-weekly testing, could enable the economy to reopen while maintaining the total infection rate much lower than the herd immunity level. Compared with the devastating economic cost of containment measures, increasing testing capacity is more cost-effective in containing the outbreak. Our results are consistent with the findings of the previous study that increasing testing frequency contributes to a safer reopening, even if tests are inaccurate[22].

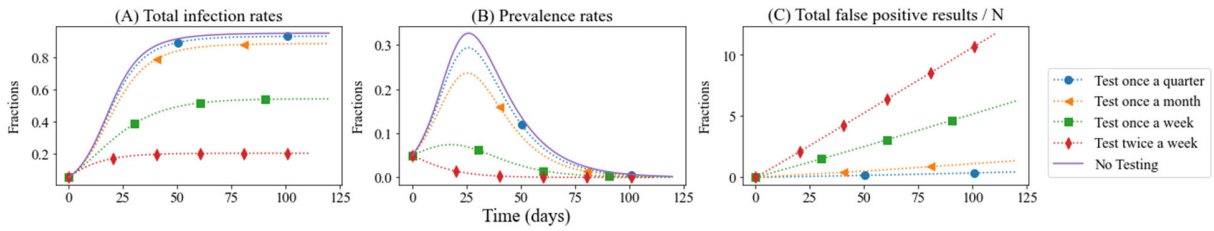

Figure 6: Comparison of different testing frequencies of individual testing.

These scenarios show total infection rates (A), prevalence rates (B), and false positives (C) of individual testing with an initial prevalence of 5% using an inaccurate technology (62.3% sensitivity and 90% specificity). The reproductive number is 3. We compare the quarterly, monthly, weekly, and twice-weekly testing frequency.

Discussion and conclusion

Flexible testing is more cost-effective in containing the outbreaks and can generate more reliable results than individual testing during the pandemic. We present an analytical method to calculate the optimal pool size that maximizes the effectiveness of limited resources. Simulation results show that increasing testing capability is a more feasible strategy to expedite identifying infected patients than improving testing sensitivity and specificity in short term. Moreover, pooled testing

can increase the reliability of negative results by lowering the prevalence and improve the reliability of positive results by testing positive samples twice.

Our analysis also subjects to several limitations. First, we derive the analytical solutions by assuming the prevalence is known. During the pandemic, developing methods to estimate the prevalence of COVID-19 is crucial for adjusting containment policies. Moreover, we use a well-mixed SIR model to capture the difference between individual and flexible testing. More detailed transmission models that consider the heterogeneous population and social contact patterns can provide a more specific testing performance evaluation. In addition, assessing the diagnostic accuracy of tests administered to pooled samples and asymptomatic individuals is also an urgent priority[5,23].

Flexible testing and the newly developed tests, which are low cost and noninvasive, make widespread and frequent testing accessible and help navigate uncertainty after the reopening. Compared with the painful containment measures, increasing testing frequency enables people to return to workplaces and students to go back to schools more safely. For example, the U.S. National Football League is individually testing all players and team personnel every day or every other day during the regular season except game day because false positives may cause the delay or cancellation of a game. The seven-day positive rate is only 0.017% between August 30–September 5, 2020[24]. Pooled testing can save more than 95% of tests for the National Football League. Moreover, the University of Illinois Urbana-Champaign (UIUC) is testing students twice-weekly using rapid saliva tests[25]. Between July 6–November 14, 2020, the daily positive rate is no more than 2.86% (August 30, 2020). As these positive rates are much lower than the threshold, compared with individual testing, pooled testing can save substantial tests and reduce numerous false positives by testing each positive sample twice. For example, the University of Cambridge and the University of Nottingham are implementing pooled testing to keep campuses open. Meanwhile, healthcare providers should offer more information on interpreting test results and organize timely follow-up testing to rule out false positive and false negative results.

Countries are striving to increase their testing capabilities to control the COVID-19 outbreak and safely reopen their economies. We suggest governments and agencies increase the testing capability and frequency by the approval of larger pool sizes and more fast test technologies. As

we learn more about the pandemic by testing more people and conducting more research, we will gain confidence and win the fight against COVID-19.

Methods

The Notation and Assumption. Let S_e denote test sensitivity, which is the proportion of diseased subjects who test positive. Let S_p denote test specificity, which is the proportion of subjects without the disease who test negative. $p \in (0,1)$ represents the prevalence of the disease. Because healthy people have a higher probability of receiving negative test results than infected people, without the loss of generality, $S_e + S_p - 1 \in (0,1)$. In this research, we limit the dilution effect by considering an upper limit M on pool size.

Derivations of the Performance Indicators. The cost-performance ratio $CPR(n)$ evaluates the overall performance of the efficiency and effectiveness of the testing strategy. Thus, we first derive the indicators of efficiency and effectiveness as functions of the prevalence p, pool size n, sensitivity S_e, and specificity S_p [19].

The expected number of tests for each subject measures the efficiency of different test strategies. For individual testing, each subject only takes one test. For pooled testing of size n $(n \geq 2)$, the pooled sample is tested in the first stage. If the pooled sample tests negative, then all subjects in this pool are classified as negative. If pooled sample tests positive, then all the subjects in this pool need to be retested individually. Note that we should consider the false positive and false negative results in the first stage.

If $n \geq 2$, the per-subject expected number of tests $\mathbb{E}\left[T(n)\right]$ is:

$$\mathbb{E}\left[T(n)\right] = (1-p)^n \left(\frac{1}{n} S_p + \frac{n+1}{n}(1-S_p)\right) + \left(1-(1-p)^n\right)\left(\frac{n+1}{n} S_e + \frac{1}{n}(1-S_e)\right).$$

Thus, for all n, the per-subject expected number of tests $\mathbb{E}\left[T(n)\right]$ can be written as:

$$\mathbb{E}\left[T(n)\right] = \begin{cases} 1, & n = 1, \\ \frac{1}{n} + S_e - \left(S_e + S_p - 1\right)(1-p)^n, & \text{otherwise.} \end{cases}$$

To measure the effectiveness, we first calculate the per-subject expected number of four outcomes, including true positives $\mathbb{E}\left[TP(n)\right]$, false negatives $\mathbb{E}\left[FN(n)\right]$, false positives $\mathbb{E}\left[FP(n)\right]$, true negatives $\mathbb{E}\left[TN(n)\right]$. Let $N^{+}(n) \sim Binomial(n, p)$ denotes the random number of true positives in a pool of size n with prevalence p.

If $n \geq 2$, the per-subject expected number of true positives $\mathbb{E}\left[TP(n)\right]$ is:

$$\begin{aligned}\mathbb{E}\left[TP(n)\right] &= \sum_{k=1}^{n} \mathbb{E}\left[TP(n) \mid N^{+}(n)=k\right] \mathbb{P}\left(N^{+}(n)=k\right) \\ &= \sum_{k=1}^{n} \frac{kS_e^2}{n}(1-p)^{n-k} p^k \\ &= \frac{S_e^2}{n} \sum_{k=1}^{n} k(1-p)^{n-k} p^k = S_e^2 p.\end{aligned}$$

If $n \geq 2$, the per-subject expected number of false positives $\mathbb{E}\left[FP(n)\right]$ is:

$$\begin{aligned}\mathbb{E}\left[FP(n)\right] &= \sum_{k=0}^{n-1} \mathbb{E}\left[FP(n) \mid N^{+}(n)=k\right] \mathbb{P}\left(N^{+}(n)=k\right) \\ &= \left(1-S_p\right)^2 \mathbb{P}\left(N^{+}(n)=0\right) + \sum_{k=1}^{n-1} \frac{(n-k)\left(1-S_p\right)S_e}{n} \mathbb{P}\left(N^{+}(n)=k\right) \\ &= S_e\left(1-S_p\right)(1-p) - \left(1-S_p\right)\left(S_e+S_p-1\right)(1-p)^n.\end{aligned}$$

As $\mathbb{E}\left[TP(n)\right] + \mathbb{E}\left[FN(n)\right] = p$ and $\mathbb{E}\left[TN(n)\right] + \mathbb{E}\left[FP(n)\right] = 1-p$, we can calculate $\mathbb{E}\left[FN(n)\right]$ and $\mathbb{E}\left[TN(n)\right]$. Thus, we have the following results:

$$\mathbb{E}\left[TP(n)\right] = \begin{cases} S_e p, & n=1, \\ S_e^2 p, & \text{otherwise}, \end{cases}$$

$$\mathbb{E}\left[FN(n)\right] = \begin{cases} \left(1-S_e\right) p, & n=1, \\ \left(1-S_e^2\right) p, & \text{otherwise}, \end{cases}$$

$$\mathbb{E}\left[FP(n)\right]=\begin{cases}(1-S_p)(1-p), & n=1,\\ S_e(1-S_p)(1-p)-(1-S_p)(S_e+S_p-1)(1-p)^n, & \text{otherwise,}\end{cases}$$

$$\mathbb{E}\left[TN(n)\right]=\begin{cases}S_p(1-p), & n=1,\\ (1-p)-(1-S_p)\left[S_e(1-p)-(S_e+S_p-1)(1-p)^n\right]. & \text{otherwise.}\end{cases}$$

Finally, we can calculate the cost-performance ratio $CPR(n)$, which is the average number of tests per confirmed case, as follows:

$$CPR(n)=\frac{\mathbb{E}\left[T(n)\right]}{\mathbb{E}\left[TP(n)\right]}=\begin{cases}\dfrac{1}{S_e p}, & n=1,\\ \dfrac{1}{S_e^2 p}\left[\dfrac{1}{n}+S_e-(S_e+S_p-1)(1-p)^n\right], & \text{otherwise.}\end{cases}$$

In the supplementary material, we prove the optimal pool size follows a threshold policy by solving $\dfrac{\partial CPR(n)}{\partial n}=0$. Then, we further discuss the impacts of the integer constraint and dilution impact on the optimal pool size.

Epidemic Model: The SIR model captures the prevalence variation over time. Susceptible individuals S_t are infected through contact with unidentified infected individuals I_t. The removed people R_t consist of recovered people and confirmed cases who are isolated until they pose no risk. Although it seems more natural to utilize population counts, we use fractions of the total population N to simplify the discussion and calculation, so that $S_t+I_t+R_t=1$. Therefore, the prevalence is I_t. The epidemic dynamics are described by the following system of ordinary differential equations:

$$\frac{dS_t}{dt}=-\beta S_t I_t,$$

$$\frac{dI_t}{dt}=\beta S_t I_t-\frac{C_0}{CPR(n_t)}-\sigma I_t,$$

$$\frac{dR_t}{dt} = \frac{C_0}{CPR(n_t)} + \sigma I_t, \tag{2}$$

where β denotes the transmission rate, and σ is the recovery rate. To prevent infections as many as possible, we assume that decision-makers make full use of daily testing capacity C_0. The number of infected individuals who are isolated after testing positive is $C_0 / CPR(n_t)$, where n_t is the pool size during time interval dt. For flexible testing strategy, the optimal pool size n_t^* is chosen dynamically according to the prevalence I_t to minimize the $CPR(n_t)$. Individual testing test one sample per assay all the time (i.e., $n_t = 1$).

Simulation Settings: We compare the epidemic dynamics and performances of individual and flexible testing strategies in a pandemic by simulating different containment measures, testing sensitivity, specificity, and capacity. The baseline values and the value ranges used for sensitivity analyses are listed in Table 1. We end the population screening when the prevalence is lower than 10^{-5}.

Table 1. Model parameter values at baseline and the ranges used for sensitivity analyses.

Parameter	Baseline Value	Range	Details and references
Maximum pool size M	32		A recent study revealed that the COVID-19 real-time quantitative polymerase chain reaction (RT-qPCR) test sensitivity retained 90% for a batch of up to 32 nasopharyngeal swabs[4].
Recovery time σ^{-1}	14		The median recovery time for mild cases is approximately two weeks[26].
Reproductive number R with containment measures	1.5	1.2, 1.5, 2.0, 3.0.	The reproductive number can be thought of the expected number of new infections that are generated directly from one infected individual. We calculate β as $\beta = R \cdot \sigma$. We use R to capture the overall impact of containment measures. A smaller R corresponds to more stringent measures, which cause more economic losses.

Sensitivity S_e	0.733	0.623, 0.972.	The test accuracy varies because of human error and variability in specimens. We use the nasopharyngeal swab PCR test as the baseline. We analyze the main range of sensitivity and specificity of PCR tests (involving the nasopharyngeal swab, saliva, and sputum samples) based on a systematic review of COVID-19 test accuracy[27].
Specificity S_p	0.986	0.900, 0.986.	
Testing capacity C_0 / N	0.010	0.011, 0.033, 0.143, 0.286.	Testing capacity $C_0 = 0 \cdot 01N$ represents that we can individually test N people within 100 days. Sensitivity analysis compare the quarterly, monthly, weekly, and twice-weekly testing frequency under individual testing.
Initial prevalence I_0	1%	5%	The initial prevalence is 1% of people are infected with COVID-19 but not removed from the transmission chain.

Contributors

Jiali Yu developed the model and prepared the first draft of the paper. Zuo-Jun (Max) Shen supervised the study, contributed to data interpretation. All authors contributed to writing the manuscript, verified the underlying data, and approved the final version..

Declaration of interests

We declare no competing interests.

Data sharing

We use the testing data provided by the COVID Tracking Project who collects, cross-checks, and publishes COVID-19 testing data of the United States. To view more descriptions about data, please visit https://covidtracking.com/about-data. Johns Hopkins Coronavirus Resource Center provides a visualization tool for testing outcomes https://coronavirus.jhu.edu/testing/individual-states. The on-campus COVID-19 testing data dashboard of UIUC is available at https://go.illinois.edu/COVIDTestingData. All code is available for free use at https://github.com/JL-YU/Flexible-Testing.

References:

1. Dorfman R. The Detection of Defective Members of Large Populations. *The Annals of Mathematical Statistics* 1943; **14**(4): 436-40. http://www.jstor.org/stable/2235930.

2. Indian Council of Medical Research. Advisory on feasibility of using pooled samples for molecular testing of COVID-19, 2020. https://www.icmr.gov.in/pdf/covid/strategy/Advisory_on_feasibility_of_sample_pooling.pdf.

3. FDA. Coronavirus (COVID-19) Update: FDA Issues First Emergency Authorization for Sample Pooling in Diagnostic Testing, 2020. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-issues-first-emergency-authorization-sample-pooling-diagnostic.

4. Yelin I, Aharony N, Tamar ES, et al. Evaluation of COVID-19 RT-qPCR Test in Multi sample Pools. *Clinical Infectious Diseases* 2020. DOI: 10.1093/cid/ciaa531.

5. Watkins AE, Fenichel EP, Weinberger DM, et al. Pooling saliva to increase SARS-CoV-2 testing capacity. *medRxiv* 2020: 2020.09.02.20183830. DOI: 10.1101/2020.09.02.20183830.

6. Augenblick N, Kolstad JT, Obermeyer Z, Wang A. Group Testing in a Pandemic: The Role of Frequent Testing, Correlated Risk, and Machine Learning. *National Bureau of Economic Research Working Paper Series* 2020; **No. 27457**. DOI: 10.3386/w27457.

7. Majid F, Omer SB, Khwaja AI. Optimising SARS-CoV-2 pooled testing for low-resource settings. *The Lancet Microbe* 2020; **1**(3): e101-e2. DOI: 10.1016/S2666-5247(20)30056-2.

8. Libin PJK, Willem L, Verstraeten T, Torneri A, Vanderlocht J, Hens N. Assessing the feasibility and effectiveness of household-pooled universal testing to control COVID-19 epidemics. *medRxiv* 2020: 2020.10.03.20205765. DOI: 10.1101/2020.10.03.20205765.

9. Deckert A, Bärnighausen T, Kyei N. Pooled-sample analysis strategies for COVID-19 mass testing: a simulation study. *Bulletin of the World Health Organization* 2020; **98**: 590-8. DOI: 10.2471/BLT.20.257188.

10. Praharaj I, Jain A, Singh M, et al. Pooled testing for COVID-19 diagnosis by real-time RT-PCR: A multi-site comparative evaluation of 5- & 10-sample pooling. *Indian J Med Res* 2020; **152**(1 & 2): 88-94. DOI: 10.4103/ijmr.IJMR_2304_20.

11. Ben-Ami R, Klochendler A, Seidel M, et al. Large-scale implementation of pooled RNA extraction and RT-PCR for SARS-CoV-2 detection. *Clinical Microbiology and Infection* 2020; **26**(9): 1248-53. DOI: 10.1016/j.cmi.2020.06.009.

12. 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. DOI: 10.1001/jamanetworkopen.2020.13075.

13. Bukhari SUK, Khalid SS, Syed A, Shah SSH. Smart Pooled sample Testing for COVID-19: A Possible Solution for Sparsity of Test Kits. *medRxiv* 2020: 2020.04.21.20044594. DOI: 10.1101/2020.04.21.20044594.

14. Žilinskas J, Lančinskas A, Guarracino MR. Pooled testing with replication: a mass testing strategy for the COVID-19 pandemics. *medRxiv* 2020: 2020.04.27.20076422. DOI: 10.1101/2020.04.27.20076422.

15. Regen F, Eren N, Heuser I, Hellmann-Regen J. A simple approach to optimum pool size for pooled SARS-CoV-2 testing. *International Journal of Infectious Diseases* 2020; **100**: 324-6. DOI: 10.1016/j.ijid.2020.08.063.

16. Mutesa L, Ndishimye P, Butera Y, et al. A pooled testing strategy for identifying SARS-CoV-2 at low prevalence. *Nature* 2020. DOI: 10.1038/s41586-020-2885-5.

17. 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. DOI: 10.1371/journal.pone.0240652.

18. Corless RM, Gonnet GH, Hare DEG, Jeffrey DJ, Knuth DE. On the LambertW function. *Advances in Computational Mathematics* 1996; **5**(1): 329-59. DOI: 10.1007/BF02124750.

19. 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. DOI: 10.1287/ijoc.2019.0942.

20. Woloshin S, Patel N, Kesselheim AS. False Negative Tests for SARS-CoV-2 Infection — Challenges and Implications. *New England Journal of Medicine* 2020. DOI: 10.1056/NEJMp2015897.

21. Singh AK, Nema RK, Joshi 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. DOI: 10.1371/journal.pone.0239492.

22. Larremore DB, Wilder B, Lester E, et al. Test sensitivity is secondary to frequency and turnaround time for COVID-19 surveillance. *medRxiv* 2020: 2020.06.22.20136309. DOI: 10.1101/2020.06.22.20136309.

23. Woloshin S, Patel N, Kesselheim AS. False Negative Tests for SARS-CoV-2 Infection — Challenges and Implications. *New England Journal of Medicine* 2020; **383**(6): e38. DOI: 10.1056/NEJMp2015897.

24. Shook N. Eight positive cases in latest COVID-19 testing data from NFL, NFLPA. September 13, 2020 2020. https://www.nfl.com/news/eight-positive-cases-in-latest-covid-19-testing-data-from-nfl-nflpa2020 (accessed October 19, 2020).

25. UIUC. Past seven-day case positivity rate. https://go.illinois.edu/COVIDTestingData (accessed November 16, 2020).

26. World Health Organization. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19), 2020. https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf.

27. Böger B, Fachi MM, Vilhena RO, Cobre AF, Tonin FS, Pontarolo R. Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19. *American Journal of Infection Control* 2020. DOI: 10.1016/j.ajic.2020.07.011.