

## RESEARCH ARTICLE

# Evaluation of pool-based testing approaches to enable population-wide screening for COVID-19

Timo de Wolff<sup>1,2</sup> , Dirk Pflüger<sup>1,3</sup> , Michael Rehme<sup>2</sup>, Janin Heuer<sup>3</sup> , Martin-Immanuel Bittner<sup>1,4</sup> \*

**1** Young Academy (Junge Akademie) of the German National Academy of Sciences, Berlin, Germany, **2** Institute for Analysis and Algebra, Technical University of Braunschweig, Braunschweig, Germany, **3** Institute for Parallel and Distributed Systems, University of Stuttgart, Stuttgart, Germany, **4** Arctoris, Oxford, United Kingdom

 These authors contributed equally to this work.

\* [martin-immanuel.bittner@arctoris.com](mailto:martin-immanuel.bittner@arctoris.com)



## OPEN ACCESS

**Citation:** de Wolff T, Pflüger D, Rehme M, Heuer J, Bittner M-I (2020) Evaluation of pool-based testing approaches to enable population-wide screening for COVID-19. PLoS ONE 15(12): e0243692. <https://doi.org/10.1371/journal.pone.0243692>

**Editor:** Wenbin Tan, University of South Carolina, UNITED STATES

**Received:** July 3, 2020

**Accepted:** November 16, 2020

**Published:** December 21, 2020

**Copyright:** © 2020 de Wolff et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its [Supporting information](#) files; all code and calculations can be accessed via <https://github.com/SC-SGS/covid19-pooling>.

**Funding:** TdW, DP and MIB are supported by the Young Academy of the German National Academy of Sciences. TdW is funded by the German Research Foundation (DFG) grant WO 2206/1-1 under the Emmy Noether Programme. DP is funded by the Cluster of Excellence Data Integrated Simulation Science under Germany's Excellence

## Abstract

### Objective

Rapid testing is paramount during a pandemic to prevent continued viral spread and excess morbidity and mortality. This study investigates whether testing strategies based on sample pooling can increase the speed and throughput of screening for SARS-CoV-2, especially in resource-limited settings.

### Methods

In a mathematical modelling approach conducted in May 2020, six different testing strategies were simulated based on key input parameters such as infection rate, test characteristics, population size, and testing capacity. The situations in five countries were simulated, reflecting a broad variety of population sizes and testing capacities. The primary study outcome measurements were time and number of tests required, number of cases identified, and number of false positives.

### Findings

The performance of all tested methods depends on the input parameters, i.e. the specific circumstances of a screening campaign. To screen one tenth of each country's population at an infection rate of 1%, realistic optimised testing strategies enable such a campaign to be completed in ca. 29 days in the US, 71 in the UK, 25 in Singapore, 17 in Italy, and 10 in Germany. This is ca. eight times faster compared to individual testing. When infection rates are lower, or when employing an optimal, yet more complex pooling method, the gains are more pronounced. Pool-based approaches also reduce the number of false positive diagnoses by a factor of up to 100.

Strategy (EXC 2075 - DFG grant 390740016). MIB receives a salary from Arctoris Ltd. The funders provided support in the form of salaries for TdW, DP and MIB, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of the authors are articulated in the 'author contributions' section.

**Competing interests:** MIB is a shareholder and director of Arctoris Ltd. This does not alter our adherence to PLOS ONE policies on sharing data and materials. Apart from salary support, Arctoris Ltd had no additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Conclusions

The results of this study provide a rationale for adoption of pool-based testing strategies to increase speed and throughput of testing for SARS-CoV-2, hence saving time and resources compared with individual testing.

## Introduction

Pandemics such as COVID-19 pose a significant public health threat, leading to morbidity, mortality, and rapid and significant strain on the health system [1]. Faced with the global spread of a novel pathogen, the identification of cases and carriers and elucidation of transmission patterns is paramount [2, 3]. The ability to rapidly and reliably diagnose those infected is critical to 1) identify and control clusters of infection; 2) prepare the health system for the patient numbers to be expected; 3) deploy medical countermeasures in a targeted way; and 4) assess the effectiveness of any public health measures and adapt them accordingly [4].

The speed of testing is critical, but is limited by supply of and access to diagnostic tests, logistical challenges, and shortages in qualified personnel and/or laboratory facilities that could perform the necessary tests [5, 6].

In each of the above scenarios, maximising the number of people that can be tested in a given time is essential. Universal weekly testing has been proposed as the only viable exit strategy from lockdown [5, 7–11]. One potential approach of increasing testing efficiency is pooling of different samples in one test. This is a well-validated method used for example in transfusion medicine for HIV testing and has recently been experimentally deployed for SARS-CoV-2 in a small-scale study in California [12, 13].

Using a simulation approach, this study aims to identify the most effective testing strategy by comparing six mathematical procedures for mass testing a given population for infection with SARS-CoV-2. The primary objective thereby is to identify as many cases as possible as quickly as possible with a given limited testing capacity. In other words, we aim to deploy the available tests as effectively as possible, increasing the number of identified cases per test (ICPT).

## Methods

### Assumptions

We simulate a screening campaign aimed at an entire population assuming a range of different possible infection rates  $ir$  (including unreported cases). We assume a testing capacity of  $c$  (e.g. PCR-based) tests per day, and that each test takes 5 hours to process in a clinical laboratory, resulting in a capacity of  $c \frac{5}{24}$  tests in parallel. For reasons of consistency, sample logistics are not included as an input parameter.

In terms of test characteristics we assume a sensitivity of  $p = 0.99$  based on data reported by LabCorp to the FDA and an estimated false-positive rate of  $q = 0.01$  [14]. We assume the test characteristics to remain constant after pooling, based on a maximum pool size of  $k = 32$  which has been shown to provide reliable results for COVID-19 [15, 16].

To account for the case of a potential drop in sensitivity due to sample pooling via a dilution effect, we duplicated all analyses with a decreased sensitivity of  $p = 0.75$ . These data can be found in the S1 and S2 Figs.

In the following section, we provide an overview of the different pooling strategies. Critically, in this article we address the *specific* scenario of a mass screening campaign of an entire population, which during the course of a pandemic likely will have to be repeated several times. In this scenario—and especially in the case of an emerging pathogen—, it can be assumed that neither the number of test kits nor the logistics available would be sufficient for individual testing. These assumptions are also supported by our calculations presented in the Results section. Given the explicit goal of this study, focusing on sensitivity and specificity alone provides an incomplete picture as it ignores an urgent public health priority, which is mass testing of sufficiently large parts of the population. This study therefore employs “confirmed cases per test” as main parameter for optimisation. It follows the general idea that if we cannot find all cases in a screening campaign, then we should still aim to find as many cases as possible using all test kits available in a certain time period.

Given the nature of the work, the study was exempted from ethics approval.

## Testing strategies

We summarise the methods and illustrate them in Fig 1.

**Individual testing.** The conventional approach of testing every person in a given population individually.

**2-level pooling.** Following a recent preprint by Hanel and Thurner we define a maximum pool size  $k$ . If the pooled test is positive, every sample in that pool is tested individually [17]. Specifically, if the first level test of the entire pool is negative, then the whole pool is cleared, without need for individual testing. This procedure was first introduced by Dorfman in 1943 and improved by Sterret in 1957 [18, 19].

**Binary splitting.** A well-known hierarchical multi-layer procedure: if a test of a pool size  $k$  is positive, the pool is split in two sets of size  $k/2$ , and a pooled test is performed on the two new sets [20]. If a pool is tested negative, it is cleared and no further tests are performed on the samples in this pool. This procedure is repeated recursively for those subsets with a positive (pooled) test until each individual case has been identified.

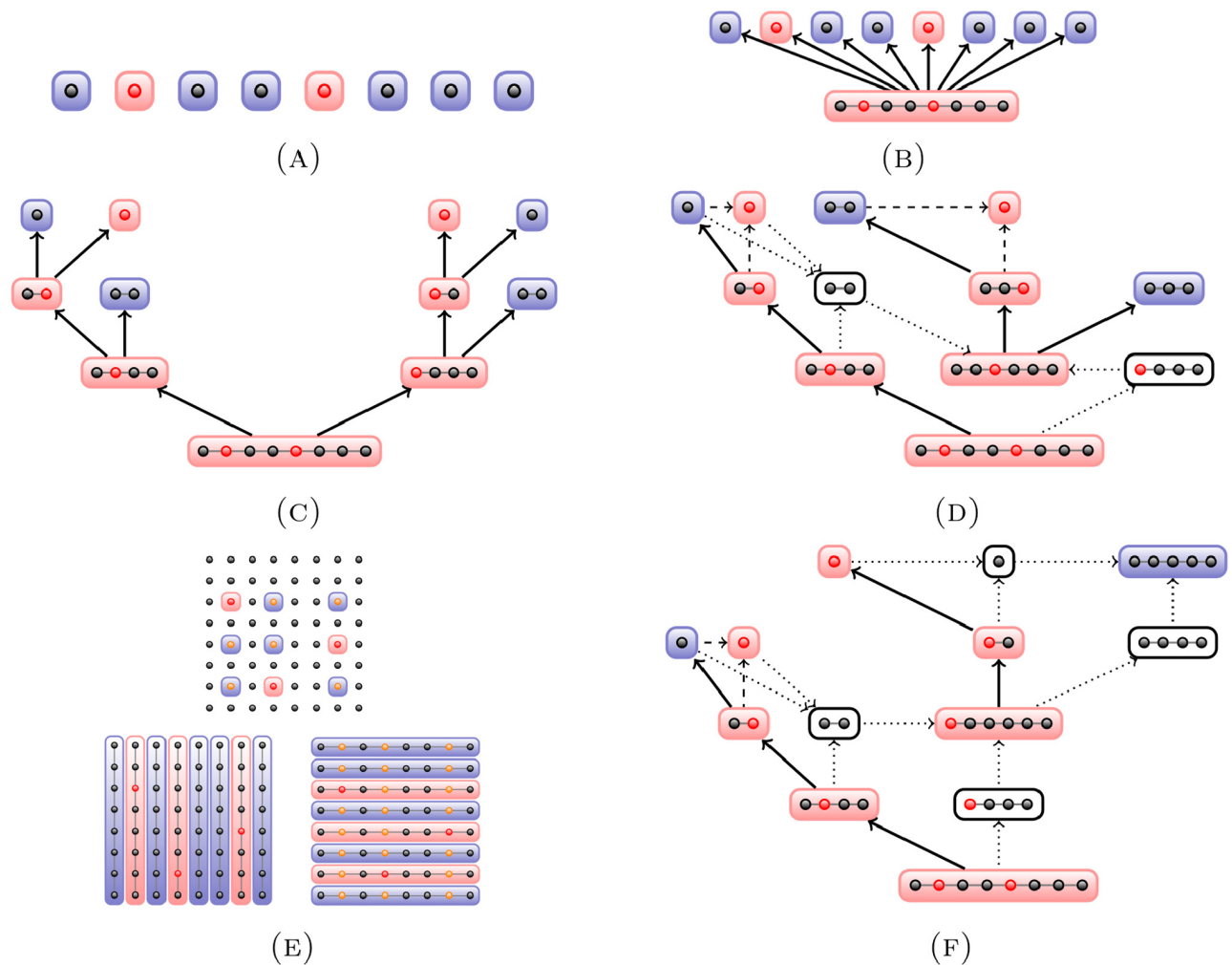
**Optimised recursive binary splitting.** A recent variation of binary splitting: if at a given level of the hierarchy only one pool tests positive, then the identification of a particular case continues via a binary search. Afterwards, all confirmed positive or confirmed negative samples are removed from the pool and the procedure continues with the remaining subjects in the unified pool [21]. We improved the method by choosing optimal initial pool sizes based on the infection rate.

Each level of the optimised recursive binary splitting (oRBS) procedure is executed as follows:

1. Starting with a pool  $P$  of (optimised) size  $k$  and testing the entire pool, unless it is known from a previous step of the algorithm that  $P$  must contain a positive sample.
2. If  $P$  is tested positive, then it is split into partitions  $P_1$  and  $P_2$  and tests are performed on these two new pools.
3. If both  $P_1$  and  $P_2$  are positive, the oRBS algorithm is run recursively on  $P_1$  and  $P_2$ . If only one partition, say  $P_1$ , tests positive, then all individuals in  $P_2$  are identified as healthy and a binary splitting subprocedure is run on  $P_1$ .

The purpose of this refinement step is to identify a single case in the pool  $P_1$  and to remove this case along with any individuals that have been identified as healthy.

4. Finally, we run the oRBS algorithm recursively on all remaining samples in  $P_1$ .



**Fig 1. An illustration of finding cases in a pool.** Demonstrated for a pool of eight samples for individual testing (A), 2-level pooling (B), binary splitting (C), and Sobel-R1 (F). For recursive binary splitting (D), we show part of the search tree highlighting the difference to binary splitting. For Purim (E), we show the search for three cases in an overall set of  $8^2 = 64$ , starting from 16 pools of size eight (i.e. 8 horizontal and 8 vertical in a matrix arrangement) before testing the cross-sections.

<https://doi.org/10.1371/journal.pone.0243692.g001>

Consider the example step in Fig 1(D). The starting point is a positive sample pool  $P_1$  of size  $k = 8$  from a previous oRBS step. Executing the binary subprocedure to identify a first case in  $P_1$ , the pool is split into two subpools  $S_1$  and  $S_2$  of size 4. To identify this case, it is sufficient to perform a test on one of the subpools. In our example, the tested subpool  $S_1$  contains one case, leading to another binary split. Only one of the resulting pools  $S_{1,1}$  and  $S_{1,2}$  of size 2 is tested.  $S_{1,1}$  is identified as positive and split into two single samples. One of these two samples is tested negative, implying that the second sample has to be positive.

At this point we have identified one case and one healthy individual in the pool  $P_1$ . Removing these two individuals, the oRBS algorithm is performed analogously on the remaining six samples in  $P_1$  until all cases in  $P_1$  have been identified.

**Purim.** A matrix-based pooling approach where one-dimensional overlapping pools are arranged in a matrix, and only cross-sections of pools that have tested positive are tested individually [22].

In the example illustrated in Fig 1(E), we consider a set of 64 individuals to be tested and choose a fixed pool size of  $k = 8$ . The samples are arranged into a  $8 \times 8$  matrix and the eight pools formed by the columns of this matrix are tested. As a next step, the eight pools formed by the rows of the matrix are tested, followed by individual tests on samples where both column and row have tested positive.

We only consider Purim’s 2D variant and neglect the 3D variant; the latter becomes impractical for low infection rates, requiring handling of up to  $32^3 = 32,768$  samples simultaneously.

**Sobel-R1.** A decision tree approach based on the assumption of a binomial distribution of the test results. Pool sizes are adapted according to the minimisation of the expected number of remaining tests [20].

This approach assumes that all  $n$  individuals to be tested are partitioned into (at most) two sets; a set of size  $m \geq 0$  containing at least one person that is known to be infected, and a binomial set of size  $n - m \geq 0$ . Denote by  $q$  the probability of being tested negative and by  $k$  the next pool size we need to determine. The expected number of pooled tests remaining to be performed for a set of size  $m$  and a binomial set of size  $n - m$  is given recursively by

$$G(0, n) = 1 + \min_{1 \leq k \leq n} \{q^k G(0, n - k) + (1 - q^k)G(k, n)\},$$

for  $m = 0$ , and

$$G(m, n) = 1 + \min_{1 \leq k \leq n} \left\{ \left( \frac{q^k - q^m}{1 - q^m} \right) G(m - k, n - k) + \left( \frac{1 - q^k}{1 - q^m} \right) G(k, n) \right\},$$

for  $n \geq m \geq 2$ , with boundary conditions

$$\begin{aligned} G(0, 0) &= 0, \\ G(1, n) &= G(0, n - 1) \quad \text{for } n = 1, 2, \dots \end{aligned}$$

Thus, we can determine the size of the next pool by computing the expected number of remaining tests.

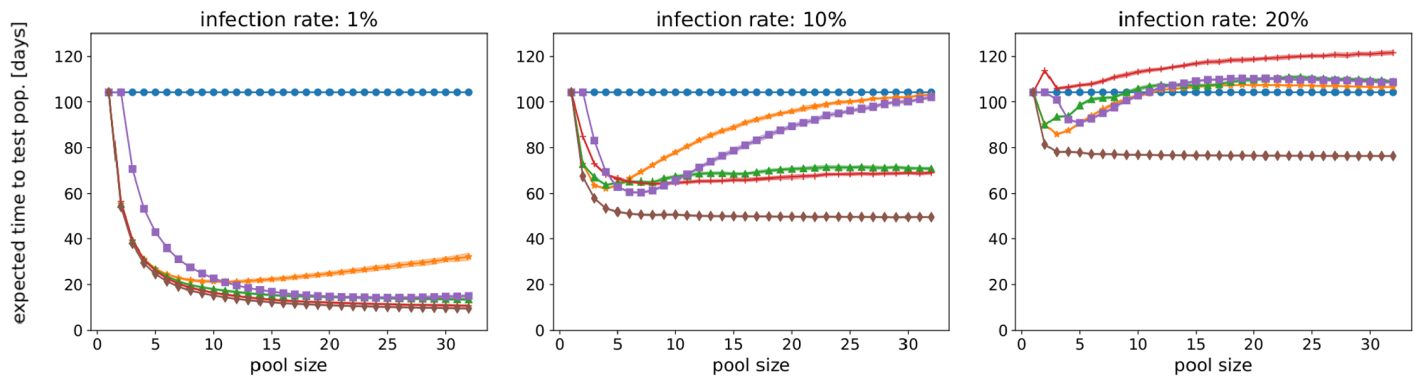
If the infection rate is known, it can be shown that this approach is a stochastically optimal search variant and therefore serves as an upper bound for the number of cases correctly identified per test.

In the example depicted in Fig 1(F), we start with a pool size 8 with an infection rate  $p = 0.25 = 1 - q$ . Computing  $G(0, 8)$  yields  $k = 8$ , and as a first step the entire pool is tested. Since the test is positive, we continue by computing  $G(8, 8)$  and inferring that we need to choose  $k = 4$  as the size of the next pool to be tested. This procedure is repeated until the end of the branch, i.e. until a single sample needs to be tested. All samples that have not yet been classified are now collected in a new pool, which in our example has size 6. Since computing  $G(0, 6)$  yields  $k = 6$ , the entire pool is tested. The result is positive, so we compute  $G(6, 6)$  and continue analogously to the procedure as described for the previous branch.

The algorithm ends once all samples in the initial pool have been classified as positive or have been found to be negative.

### Implementation and statistics

Samples within the virtual populations are designated as positive randomly using Python’s numpy random number generator in accordance with the specified infection rate. To account for randomness and to obtain robust results, we simulate in repetitions of ten and then calculate the expectation value and standard deviation.



**Fig 2. The best pool size (lowest total time) depends on the infection rate, here for  $ir = 1\%$ ,  $10\%$ ,  $20\%$ .** For low infection rates, all methods but 2-level pooling prefer pool sizes that are as large as possible. With increasing infection rate, the optimal pool size decreases—with the exception of the Sobel-R1 method—until they approach pool size 1. Parameters: sensitivity  $p = 0.99$ , false positive rate  $q = 0.01$ , population 50,000, test duration 5h, averaged over 10 runs. Blue: individual testing; orange: 2-level pooling; green: binary splitting; red: recursive binary splitting; purple: Purim; brown: Sobel-R1.

<https://doi.org/10.1371/journal.pone.0243692.g002>

For all pooling approaches we limit the maximum pool size to 32 which was shown to be a viable upper limit enabling pool-based testing without a significant loss of sensitivity [15, 16]. However, for increasing infection rates, smaller pool sizes result in more cases identified per test, which means the initial pool size can be optimised, see e.g. Hanel and Thurner and Xiong et al. [17, 23]. We optimise the pool size for each testing strategy and infection rate using a test population of 50,000 individuals using every pool size in  $\{1, \dots, 32\}$ . The optimal pool size is the one leading to the fastest screening of the test population. It is calculated once, stored, and then reused for the appropriate combination of testing method and infection rate.

As shown in Fig 2, the optimal pool size for each method in terms of lowest expected time to test an entire population depends on the infection rate. For an infection rate  $ir = 1\%$ , the expected total time decreases with increasing pool size for all pooling methods except 2-level pooling. For increasing infection rates, the optimal pool size decreases for all pooling methods but Sobel-R1, and eventually approaches pool size 1 and thus individual testing.

The code and all data are available at

<https://github.com/SC-SGS/covid19-pooling>.

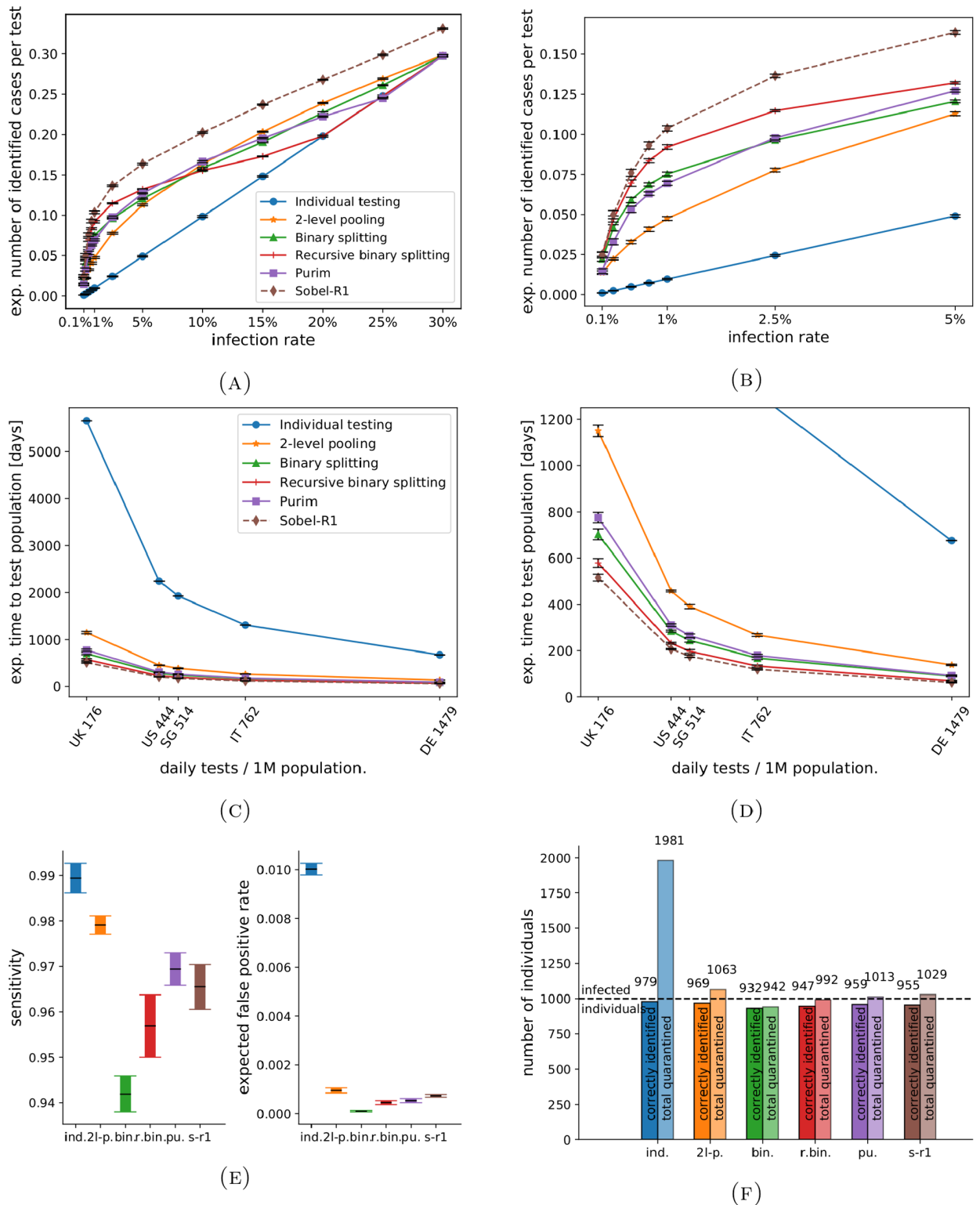
## Results

This study employs six different methods to model the most effective strategy to screen a given population within the shortest time and with the smallest number of tests possible. Since the choice of ideal testing strategy depends on the input parameters (such as size and infection rate of a population, test sensitivity, etc.), the models we developed can be fully customised and run via

<https://covid19.enfunction.com>,

enabling the reader to find the best possible strategy for their specific needs.

Fig 3(A) compares the effectiveness in terms of ICPT for all methods for different infection rates under the assumptions given in the Methods section. Sobel-R1 is the optimal method and provides a theoretical upper bound for the achievable ICPT. However, this method is



**Fig 3. Screening the whole population.** Parameters: sensitivity  $p = 0.99$ , false positive rate  $q = 0.01$ , test duration 5h, averaged over 10 runs. Optimal (max.) pool size each (c.f. Fig 2); for  $ir = 1\%$  as in (B)–(D) we obtain individual testing: 1; 2-level pooling: 12; binary splitting: 32; recursive binary splitting: 32; Purim: 32, Sobel-R1: 31. (A): Expected number and standard deviation of identified cases per test for different infection rates with optimised pool sizes; right: zoom-in for small infection rates, (B): Expected time in days and standard deviation to test the whole population depending on daily test capacity per 1m population, infection rate 1%; right: zoom-in excl. individual testing, (C): Expected true positives and false positives for the six methods (pop. 100,000), (D): Numbers of identified cases and quarantined individuals (pop. 100,000).

<https://doi.org/10.1371/journal.pone.0243692.g003>

Table 1. Data used in the simulations based on country.

| Country | Population $n$ | Daily testing capacity $c$ | Tests per 1m population |
|---------|----------------|----------------------------|-------------------------|
| US      | 328.24 m       | 146,000                    | 444                     |
| DE      | 83.15 m        | 123,000                    | 1,479                   |
| UK      | 67.89 m        | 12,000                     | 176                     |
| IT      | 60.31 m        | 46,000                     | 762                     |
| SG      | 5.64 m         | 2,900                      | 514                     |

14 April 2020, various sources (top to bottom: The Atlantic; Robert-Koch-Institute; gov.uk; ourworldindata.org; The Straits Times)

<https://doi.org/10.1371/journal.pone.0243692.t001>

restricted in its practical applicability (see below). For an infection rate of 1%, the hierarchical approaches such as binary and recursive binary splitting show almost an eight and ten fold ICPT increase compared to the current status quo of individual testing respectively, enabling the testing of large groups with a single test. Disregarding Sobel-R1, we obtain the following results: for infection rates up to 6%, recursive binary splitting is the optimal method. For infection rates up to 2.5%, the next best option is binary splitting, and for infection rates between 2.5% and 6%, the next best method is the matrix-based Purim method. For infection rates between 6% and 12%, Purim is the optimal method. For infection rates of 12% and higher, 2-level pooling with optimised pool sizes yields the highest ICPT. Of note, the current standard approach of individual testing is never found to be the best choice in the scenarios we modelled.

To better examine the potential of hierarchical approaches, we deployed our models to simulate the situation for five countries: the US, the UK, Germany, Italy, and Singapore; see Table 1. We assume that as of May 2020  $ir = 1\%$  of the population is infected. Fig 3(B) shows the overall time required to test the whole population in each of the five countries. Even in Germany, which has the highest relative testing capacity per day of these five countries, it would take 675 days with the current screening approach (individual testing) to test every individual; the US on the other hand would require 2,244 days. For the US, binary splitting reduces the time required to 285 days (about 9.5 months), and the optimised recursive binary splitting to 232 days. If only one tenth of the population needs to undergo screening—e.g. when prioritising frontline medical staff and public workers—then this would mean such a campaign could be completed with binary splitting in about 29 days in the US, 71 in the UK, 25 in Singapore, 17 in Italy, and 10 in Germany. When the infection rates are considerably lower, or when employing an optimal, yet logistically more challenging, pooling method, the gains are more pronounced. With Sobel-R1, the same screening campaign could be completed in 21 days in the US, 52 in the UK, 18 in Singapore, 12 in Italy, and 7 in Germany. As a case study, we conducted a comparison using the six approaches for three different infection rates for the US, see Table 2.

It is important to note that a screening campaign based on a hierarchical approach on average identifies fewer cases than individual testing. Assuming a test sensitivity of 0.99 for an individual test, employing hierarchical testing has a probability of 0.01 (or 1%) to miss a certain case on each of several test stages. Fig 3(C) shows that the (compounded) expected rate of identified cases (true positives) therefore drops by between 1% and 5% in total. Hence, the *sensitivity* of a hierarchical approach is lower compared to individual testing. In contrast, the likelihood of incorrectly classifying a subject as infected is reduced from 1% to almost 0%, see Fig 3(D); i.e. when carrying out a hierarchical approach the *specificity* is improved compared to individual testing.



**Table 2. Expected time in days to test 10% of the US population for three different infection rates.**

| Method                     | <i>ir</i> = 0.1% |         | <i>ir</i> = 1% |         | <i>ir</i> = 2.5% |         |
|----------------------------|------------------|---------|----------------|---------|------------------|---------|
|                            | days             | speedup | days           | speedup | days             | speedup |
| Individual testing         | 224.4            |         | 224.4          |         | 224.4            |         |
| 2-level pooling            | 16.4             | 13.7x   | 45.8           | 4.9x    | 70.1             | 3.2x    |
| Binary splitting           | 10.3             | 21.8x   | 28.5           | 7.9x    | 55.4             | 4.1x    |
| Recursive binary splitting | 8.8              | 25.5x   | 23.2           | 9.7x    | 46.4             | 4.8x    |
| Purim                      | 14.6             | 15.4x   | 31.1           | 7.2x    | 55.2             | 4.1x    |
| Sobel-R1                   | 8.7              | 25.8x   | 20.7           | 10.8x   | 39.4             | 5.7x    |

Test capacities and simulation parameters as of April 14, 2020, see [Table 1](#)

<https://doi.org/10.1371/journal.pone.0243692.t002>

To illustrate why the number of identified cases per test (ICPT) is a suitable measure to compare the effectiveness of different pooling methods, we simulate the application of 100,000 tests on a population of 1 million individuals, see [Table 3](#). We demonstrate that while the number of false negatives increases when employing pooling methods, two other measures change to a much greater degree: the number of untested (infected) individuals decreases *and* the number of true positives increases. As before, we assume an infection rate of 1%. The effectiveness of each approach can directly be observed via the number of individuals tested in a given population. While for individual testing the number of tested individuals equals the number of tests available, 2-level pooling examines 490,792 individuals on average with 100,000 tests, and Sobel-R1 can test the whole population more than once (1.08 times). Binary splitting is best suited to illustrate the occurrence of false negatives: the overall sensitivity is reduced from 99% to about 94.2% (see [Fig 3\(C\)](#)), and the false negative rate increases correspondingly from 1% to 6.8%. In our simulation, the expected number of cases that are not correctly classified as infected increases from 10 to 471. However, binary splitting at the same time correctly identifies 8,363 more infected individuals which individual testing was not able to test. Our results clearly show that all pooling methods significantly outperform individual testing, with the additional benefit of fewer people being sent to quarantine despite being healthy (fp).

**Table 3. Effectiveness of conducting 100,000 tests on a population of 1 million; full statistics over 10 simulations each.**

|   | Method             |                 |                  |                          |           |                              |
|---|--------------------|-----------------|------------------|--------------------------|-----------|------------------------------|
|   | Individual testing | 2-level pooling | Binary splitting | Recursive bin. splitting | Purim     | Sobel-R1                     |
| No. individuals tested  | 100,000.0          | 490,791.9       | 804,985.4        | 976,797.2                | 720,729.5 | 1,083,181.0 <sup>(*)</sup>   |
| Cases found (tp)  | 990.0              | 4,810.7         | 7,578.5          | 9,353.0                  | 6,993.6   | 10,442.2 <sup>(*)</sup>      |
| False-positives (fp)  | 990.3              | 460.5           | 76.6             | 459.5                    | 393.6     | 850.4 <sup>(*)</sup>         |
| Cases missed (fn)   | 10.0               | 97.2            | 471.3            | 415.0                    | 213.7     | 389.6 <sup>(*)</sup>         |
| Cleared (tn)  | 98,009.7           | 485,423.4       | 796,858.9        | 966,569.8                | 713,128.6 | 1,071,498.7 <sup>(*)</sup>   |
| Infected individuals not even tested due to limited test capacity | 9,000.0            | 5,092.1         | 1,950.1          | 232.0                    | 2,792.7   | 0.0 (-831.8 <sup>(*)</sup> ) |

Simulating 100,000 tests on a total population of 1 million with an infection rate of 1%. We assume a sensitivity of 99% and a false negative rate of 1% for a single PCR, and no dilution due to pooling. We report how many individuals are tested with 100,000 (pooled) tests as well as true positives (tp), false positives (fp), false negatives (fn) and true negatives (tn). The reported numbers are averaged over 10 runs testing the whole population, measuring a relative standard deviation of <0.3% each.

(\*) Note that Sobel-R1 is able to test the whole population 1.08 times; this methods identifies 10, 442.2 > 10, 000 cases, identifying 832 cases in the second run leading to a negative value in the last entry.

<https://doi.org/10.1371/journal.pone.0243692.t003>

Note that with binary splitting a single sample can undergo up to six sequential testing steps. Purim and 2-level testing can be carried out in two sequential testing steps. For recursive binary splitting and Sobel-R1, re-pooling the batch sizes can lead to large numbers of sequential stages (up to 17 and 23 hierarchical steps in the worst case scenario for multiple cases in a pool, respectively). For Sobel-R1 as the (theoretical) optimal method, at an infection rate of 1%, only 5% of tests could be carried out with at most 5 hierarchical steps, whereas 95% could be carried out with at most 13 hierarchical steps.

## Discussion

To the best of our knowledge, this study presents the first comprehensive, comparative assessment of optimised testing strategies for effective large-scale screening for infection with SARS-CoV-2. Our simulations indicate that population-level diagnostics in a pandemic the scale of COVID-19 will only be possible by making use of pool-based strategies—at the time of our study in May 2020, testing even 10% of the US population with the available testing capacity would have taken more than seven months.

Our study enables the following key conclusions for mass testing for COVID-19:

1. Screening the entire population can be carried out several times faster via hierarchical or matrix-based approaches compared with individual testing in every considered scenario.
2. Among the methods that can be considered immediately applicable in practice, we see that with respect to an optimal ICPT:
  - (a). Binary splitting is the best method for infection rates between 0% and 2.5%.
  - (b). Purim is the best method for infection rates between 2.5% and 12%.
  - (c). 2-level pooling is the best method for infection rates beyond 12%.
3. Recursive binary splitting and the Sobel-R1 method would allow to significantly improve the ICPT. However, to be practically applicable, first, either the time required to process each individual test would have to be reduced or the possible storage time of samples would have to be increased by a factor of about two (given the average number of hierarchical steps expected). Second, sufficient sample material would have to be available for a large number of repeat tests. Third, software support for lab technicians would need to be provided to guide lab operations and ensure that the method can be carried out swiftly and correctly.

With a maximum pool size of 32, binary splitting requires up to six sequential testing steps for a single sample. With an assumed duration of 5 hours per test, individual results are available after 30h and six subdivisions of the sample for a pool size of 32, which we consider as a realistic scenario. Purim and 2-level testing can be carried out in two sequential testing steps. These methods conclude testing after 10h and can realistically be applied in practice as of now.

Of note, all hierarchical models require adequate scheduling of the processes within the involved laboratories, as the performance of certain tests is conditional upon the outcome of previous tests. However, these scheduling problems are well studied, with standard-type optimisations available that can easily be implemented.

For the purpose of this study, algorithms were implemented as described in the literature. Therefore, certain adaptations to account for medical and legal requirements might be needed (e.g. Sobel-R1 can identify cases via exclusion, requiring an additional confirmation step to be included for full clinical applicability).

In recent publications and preprints published in the course of the COVID-19 pandemic, 2-level pooling with optimised pool size, matrix-based approaches for pooled screening and binary splitting were presented [17, 22, 24–28]. Our simulations include models of all of these central categories of approaches, benchmarking and comparing the various testing strategies, and demonstrating for which scenarios each method can become the optimal choice. Our model calculations prioritise identifying infected individuals as soon as possible to prevent further spread, i.e. optimising for speed of completion of a screening campaign. This is especially important since regular re-testing will be necessary given the continuous risk of infection and potentially even recurrence as highlighted by recent case reports.

Of course, differences in test characteristics—sensitivity, rate of false positives, and processing time—have a significant impact on the choice of testing strategy. In particular, test sensitivities were recently reported to be in the range of 0.75 for sputum or nasal swabs [29]. However, based on the same study, the decreased sensitivity can mostly be attributed to the quality of the samples and variations in distribution of viral load in patients—in other words, the samples taken will or will not contain viral material that could be amplified during the PCR tests (with their sensitivity remaining constant at 0.99). Thus, if a sample contains no viral material, then any method will fail, whether individual or pooled. If a sample, however, contains viral material, pooling—which happens after sample taking—can be conducted and is advisable as discussed in this paper. For illustration, a 0.75 test sensitivity scenario is given in the S1 and S2 Figs, showing that even under these assumptions, pooling is still advisable.

In addition, all variations can be modelled via our online tool <https://covid19.enfunction.com>.

We show that a hierarchical testing approach increases the specificity of a screening campaign. The reduced number of false positives is important, as it equates to ca. 1,000 healthy individuals who would otherwise have been erroneously quarantined in a population of 100,000. At the same time, binary and recursive binary splitting correctly identify up to ten times more cases per test compared to the currently employed individual testing. Furthermore, they enable multiple screening campaigns within the same time normally required for a single screening campaign where each sample is tested individually.

One of the biggest caveats of any modelling approach is the need to show that theoretical simulations can be successfully translated into public health measures. Hogan et al. and Yelin et al. recently presented their findings on the practical applicability of sample pooling in California and Israel, respectively, thereby providing important experimental validation for our modelling approach [12, 15]. In fact, Yelin et al. even showed that a pool size of up to 64 still provides an acceptable sensitivity, potentially enabling larger pool sizes that would increase the benefits of hierarchical pooling approaches. The exact maximum number of samples that can be pooled and analysed while maintaining adequate sensitivity will likely depend on the characteristics of each test and is an important question for follow-on work. Of note, on 13 April 2020, the Indian Council of Medical Research published guidance recommending limited 2-level pooling to increase screening capacity—to our best knowledge the first pooling approach formally adopted as of now, but certainly not the last [30].

## Conclusion

Rapid identification of patients, asymptomatic carriers, and the modes of transmission of a given pathogen are key goals of pandemic response, that can then be embedded into a larger set of medical countermeasures [31]. This study provides a theoretical framework and practical guidance to frontline medical staff, public health authorities, and governments on how to best

deploy limited testing resources to maximise the number of people tested in the shortest amount of time possible—in the case of COVID-19 as well as future pandemic outbreaks.

## Supporting information

**S1 Fig. Decreased test sensitivity: The best pool size (lowest total time) depends on the infection rate, here for  $ir = 1\%$ ,  $10\%$ ,  $20\%$ .** Parameters: sensitivity  $p = 0.75$ , false positive rate  $q = 0.01$ , population 50,000, test duration 5h, averaged over 10 runs. Blue: individual testing; orange: 2-level pooling; green: binary splitting; red: recursive binary splitting; purple: Purim; brown: Sobel-R1.  
(EPS)

**S2 Fig. Decreased test sensitivity: Screening the whole population.** Parameters: sensitivity  $p = 0.75$ , false positive rate  $q = 0.01$ , test duration 5h, averaged over 10 runs. Optimal (max.) pool size each (c.f. Fig 2); for  $ir = 1\%$  as in (B)–(D) we obtain individual testing: 1; 2-level pooling: 12; binary splitting: 32; recursive binary splitting: 32; Purim: 31, Sobel-R1: 32. (A): Expected number and standard deviation of identified cases per test for different infection rates with optimised pool sizes; right: zoom-in for small infection rates, (B): Expected time in days and standard deviation to test the whole population depending on daily test capacity per 1m population, infection rate 1%; right: zoom-in excl. individual testing, (C): Expected true positives and false positives for the six methods (pop. 100,000), (D): Numbers of identified cases and quarantined individuals (pop. 100,000).  
(EPS)

**S1 URL. Code and underlying data.** The code and all underlying data are available at <https://github.com/SC-SGS/covid19-pooling>.  
(TXT)

## Acknowledgments

The authors would like to thank Professor Roy Kishony (Technion—Israel Institute of Technology) for valuable contributions on PCR test characteristics; Dr Shmona Simpson (Bill and Melinda Gates Foundation, USA) for valuable discussions regarding planning and implementation of pandemic response measures; Dr Stefan Zimmer (University of Stuttgart, Germany) for assistance in code review and discussions regarding optimised pooling approaches; Dr Nikita Kaushal (Nanyang Technological University, Singapore) for manuscript review; and Dr Semen Trygubenko (Arctoris, UK) for assistance in implementing the online tool that allows for user-directed modelling for different testing scenarios.

## Author Contributions

**Conceptualization:** Timo de Wolff, Dirk Pflüger, Michael Rehme, Martin-Immanuel Bittner.

**Data curation:** Michael Rehme, Janin Heuer, Martin-Immanuel Bittner.

**Formal analysis:** Martin-Immanuel Bittner.

**Funding acquisition:** Timo de Wolff, Dirk Pflüger, Martin-Immanuel Bittner.

**Investigation:** Timo de Wolff, Dirk Pflüger, Michael Rehme, Janin Heuer, Martin-Immanuel Bittner.

**Methodology:** Timo de Wolff, Dirk Pflüger, Michael Rehme, Janin Heuer, Martin-Immanuel Bittner.

**Project administration:** Timo de Wolff, Dirk Pflüger, Martin-Immanuel Bittner.

**Resources:** Timo de Wolff, Dirk Pflüger, Janin Heuer, Martin-Immanuel Bittner.

**Software:** Timo de Wolff, Dirk Pflüger, Michael Rehme, Janin Heuer.

**Supervision:** Timo de Wolff, Dirk Pflüger, Martin-Immanuel Bittner.

**Validation:** Timo de Wolff, Dirk Pflüger, Michael Rehme, Janin Heuer, Martin-Immanuel Bittner.

**Visualization:** Timo de Wolff, Dirk Pflüger, Michael Rehme, Martin-Immanuel Bittner.

**Writing – original draft:** Timo de Wolff, Dirk Pflüger, Michael Rehme, Janin Heuer, Martin-Immanuel Bittner.

**Writing – review & editing:** Timo de Wolff, Dirk Pflüger, Michael Rehme, Janin Heuer, Martin-Immanuel Bittner.

## References

1. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al. Fair Allocation of Scarce Medical Resources in the Time of Covid-19. *New England Journal of Medicine*. 2020; 382(21):2049–2055. <https://doi.org/10.1056/NEJMsb2005114> PMID: 32202722
2. Kelly-Cirino CD, Nkengasong J, Kettler H, Tongio I, Gay-Andrieu F, Escadafal C, et al. Importance of diagnostics in epidemic and pandemic preparedness. *BMJ Global Health*. 2019; 4(Suppl 2). <https://doi.org/10.1136/bmjgh-2018-001179> PMID: 30815287
3. Lippi G, Plebani M. The critical role of laboratory medicine during coronavirus disease 2019 (COVID-19) and other viral outbreaks. *Clinical Chemistry and Laboratory Medicine (CCLM)*. 2020; 58(7):1063–1069. <https://doi.org/10.1515/cclm-2020-0240> PMID: 32191623
4. Pulia MS, O'Brien TP, Hou PC, Schuman A, Sambursky R. Multi-tiered screening and diagnosis strategy for COVID-19: a model for sustainable testing capacity in response to pandemic. *Annals of Medicine*. 2020; 52(5):207–214. <https://doi.org/10.1080/07853890.2020.1763449> PMID: 32370561
5. Peto J. Covid-19 mass testing facilities could end the epidemic rapidly. *BMJ*. 2020; 368:m1163. <https://doi.org/10.1136/bmj.m1163> PMID: 32201376
6. Burki TK. Testing for COVID-19. *The Lancet Respiratory Medicine*. 2020; 8(7):e63–e64. [https://doi.org/10.1016/S2213-2600\(20\)30247-2](https://doi.org/10.1016/S2213-2600(20)30247-2) PMID: 32479792
7. Shah P. Testing Is Our Way Out; 2020.
8. McClellan M, Gottlieb S, Mostashari F, Rivers C, Silvis L. A National COVID-19 Surveillance System: Achieving Containment. Duke-Margolis Center for Health Policy; Durham, NC, USA; 2020.
9. Peto J, Alwan NA, Godfrey KM, et al. Universal weekly testing as the UK COVID-19 lockdown exit strategy. *The Lancet*. 2020. [https://doi.org/10.1016/S0140-6736\(20\)30936-3](https://doi.org/10.1016/S0140-6736(20)30936-3) PMID: 32325027
10. Edmond J Safra Center for Ethics at Harvard University. Roadmap to Pandemic Resilience. Harvard University; Boston, MA, USA; 2020.
11. Rubin EJ, Baden LR, Morrissey S. Audio Interview: Loosening Covid-19 Restrictions. *New England Journal of Medicine*. 2020; 382(18):e67. PMID: 32348666
12. Hogan CA, Sahoo MK, Pinsky BA. Sample Pooling as a Strategy to Detect Community Transmission of SARS-CoV-2. *JAMA*. 2020. <https://doi.org/10.1001/jama.2020.5445> PMID: 32250394
13. Pilcher CD, Fiscus SA, Nguyen TQ, et al. Detection of Acute Infections during HIV Testing in North Carolina. *New England Journal of Medicine*. 2005; 352(18):1873–1883. <https://doi.org/10.1056/NEJMoa042291> PMID: 15872202
14. LabCorp. Accelerated Emergency Use Authorization (EUA) Summary COVID-19 RT-PCR Test. Laboratory Corporation of America; 2020.
15. Yelin I, Aharony N, Tamar ES, Argoetti A, Messer E, Berenbaum D, et al. Evaluation of COVID-19 RT-qPCR Test in Multi sample Pools. *Clinical Infectious Diseases*. 2020. <https://doi.org/10.1093/cid/ciaa531> PMID: 32358960
16. 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. [https://doi.org/10.1016/S1473-3099\(20\)30362-5](https://doi.org/10.1016/S1473-3099(20)30362-5) PMID: 32530425

17. Hanel R, Thurner S. Boosting test-efficiency by pooled testing strategies for SARS-CoV-2. Preprint, ArXiv 200309944. 2020.
18. Sterrett A. On the detection of defective members of large populations. *Ann Math Stat.* 1957; 28:1033–1036. <https://doi.org/10.1214/aoms/1177706807>
19. Dorfman R. The Detection of Defective Members of Large Populations. *Ann Math Statist.* 1943; 14(4):436–440. <https://doi.org/10.1214/aoms/1177731363>
20. Sobel M, Groll PA. Group Testing To Eliminate Efficiently All Defectives in a Binomial Sample. *Bell System Technical Journal.* 1959; 38(5):1179–1252. <https://doi.org/10.1002/j.1538-7305.1959.tb03914.x>
21. Cheng Y, Du DZ, Zheng F. A new strongly competitive group testing algorithm with small sequentiality.
22. Fargion BI, Fargion D, De Sanctis Lucentini PG, Habib E. Purim: a rapid method with reduced cost for massive detection of CoVid-19. Preprint, ArXiv 200311975. 2020.
23. Xiong W, Lu H, Ding J. Determination of Varying Group Sizes for Pooling Procedure. *Computational and mathematical methods in medicine.* 2019; 2019:4381084. <https://doi.org/10.1155/2019/4381084> PMID: 31065292
24. Täufer M. Rapid, large-scale, and effective detection of COVID-19 via non-adaptive testing. *Journal of Theoretical Biology.* 2020; 506:110450. <https://doi.org/10.1016/j.jtbi.2020.110450> PMID: 32814072
25. Sinnott-Armstrong N, Klein D, Hickey B. Evaluation of Group Testing for SARS-CoV-2 RNA. medRxiv. 2020. <https://doi.org/10.1101/2020.03.27.20043968>
26. Shani-Narkiss H, Gilday OD, Yayon N, Landau ID. Efficient and Practical Sample Pooling for High-Throughput PCR Diagnosis of COVID-19. medRxiv. 2020. <https://doi.org/10.1101/2020.04.06.20052159>
27. Narayanan K, Frost I, Heidarzadeh A, et al. Pooling RT-PCR or NGS samples has the potential to cost-effectively generate estimates of COVID-19 prevalence in resource limited environments. medRxiv. 2020. <https://doi.org/10.1101/2020.04.03.20051995>
28. Noriega R, Samore MH. Increasing testing throughput and case detection with a pooled-sample Bayesian approach in the context of COVID-19. bioRxiv. 2020. <https://doi.org/10.1101/2020.04.03.024216>
29. Yang Y, Yang M, Shen C, Wang F, Yuan J, Li J, et al. Evaluating the accuracy of different respiratory specimens in the laboratory diagnosis and monitoring the viral shedding of 2019-nCoV infections. medRxiv. 2020. <https://doi.org/10.1101/2020.02.11.20021493>
30. Prakash S, Jain A, Rade K. Advisory on feasibility of using pooled samples for molecular testing of COVID-19. Indian Council of Medical Research; 2020.
31. Simpson S, Kaufmann MC, Glozman V, Chakrabarti A. Disease X: accelerating the development of medical countermeasures for the next pandemic. *The Lancet Infectious Diseases.* 2020. [https://doi.org/10.1016/S1473-3099\(20\)30123-7](https://doi.org/10.1016/S1473-3099(20)30123-7) PMID: 32197097