

GROUP TESTING ENABLES ASYMPTOMATIC SCREENING FOR COVID-19 MITIGATION: FEASIBILITY AND OPTIMAL POOL SIZE SELECTION WITH DILUTION EFFECT

Yifan Lin
Yuxuan Ren
Jingyuan Wan
Enlu Zhou

Massey Cashore
Jiayue Wan
Yujia Zhang
Peter I. Frazier

Georgia Institute of Technology
755 Ferst Drive NW
Atlanta, GA 30332, USA

Cornell University
206 RHODES HALL
Ithaca, NY 14853, USA

ABSTRACT

Group testing pools multiple samples together and performs tests on these pooled samples to discern the infected samples. It greatly reduces the number of tests, however, with a sacrifice of increasing false negative rates due to the dilution of the viral load in the pooled samples. Therefore, it is important to balance the trade-off between number of tests and number of false negatives. We compare two popular group testing methods, namely linear array (a.k.a. Dorfman's procedure) and square array methods, and analyze the optimal pool size of a pooled sample that minimizes the number of false negatives per person under the constraint of testing capacity. We consider testing a closed community and determine the optimal testing cycle length that minimizes the final prevalence rate of infection at the end of the time period. Finally, we provide a testing protocol for practitioners to use these group testing methods in COVID-19 pandemic.

1 LINEAR ARRAY AND SQUARE ARRAY GROUP TESTING

Group testing provides a promising way to save the testing budget while detecting the infected samples out of a large population. However, pooling samples together dilutes the viral load in the pooled sample, and hence increases the number of false negatives. We take into account this dilution effect by fitting a logistic curve between (negative log) viral load and the false negative rate. We focus on two group testing methods that are easy to implement in practice and can be run in parallel, namely linear array and square array methods. To illustrate the two methods, we first introduce a set of notations. Denote by N the total number of people to test. Denote by p_0 the prevalence rate (i.e. the probability that an individual is infected). Denote by n the pool size in group testing (i.e., the number of samples pooled together). Denote by C_p the test capacity per person (i.e., the ratio of test capacity to the total number of people). We assume N is large so that all samples can be arranged into linear arrays and square arrays with pool size n . In addition, we assume the events that each individual gets infected are mutually independent. In the linear array group testing, we form $\lceil \frac{N}{n} \rceil$ linear array groups and conduct one test on each group. If a group tests positive, we will test each one in the group individually. In the square array group testing, we form $\lfloor \frac{N}{n^2} \rfloor$ square array groups of size $n \times n$ and pools all samples in each row and column, and thus we perform $2n$ pooled tests on each group. A sample is deemed suspicious if both its row and its column pooled samples test positive, and then these suspicious samples are tested individually to confirm their positiveness/negativeness. We assume N is large so that all samples can be arranged into linear/square arrays with pool size n . We evaluate the performance of each group testing method by minimizing the expected number of false negatives per person under the constraint that the expected number of tests per person not exceeding C_p . Figure 1 shows

the optimal objective values with respect to C_p and p_0 , of the linear array method, square array method, and the benchmark individual testing. In conclusion, if we are given a relatively large test capacity, we should consider using the linear array method. On the other hand, if we face the shortage of testing kits, and only have a relatively small test capacity, we prefer using the square array group test.

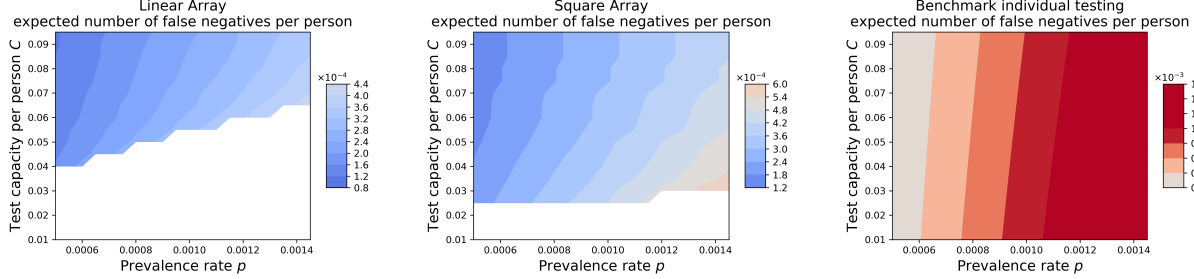


Figure 1: Comparison of the false negative detection under different settings of p_0 and C_p .

2 GROUP TESTING WITHIN A TESTING CYCLE

We consider testing a large population in a closed community such as college or nursing home with a mixed method of linear array and square array group testing. Due to limited daily testing capacity, we can only test the whole population in a testing cycle of multiple days. We propose a testing-quarantine-infection model, where group testing is conducted at the beginning of each day and people who test positive will be quarantined, while the infection keeps spreading. Because of the cost of dynamically adjusting pool size, we fix the pool size for each testing cycle. We optimize pool size to minimize the expected number of false negatives according to the prevalence rate at the beginning of a testing cycle. The number of false negatives will affect the number of people quarantined, which further impacts the prevalence rate over time. The influence will propagate to eventually impact the final prevalence rate. We aim to select the optimal testing cycle length t^* that minimizes the final prevalence rate at a given time horizon T . We set the time horizon $T = 14$, and simulate the testing-quarantine-infection model under different testing cycle lengths $t = 2, \dots, 7$ (note $t = 1$ is infeasible due to the limited daily test capacity). Figure 2 shows the prevalence rate on each day during the time period under different scenarios, compared against the benchmark of individual testing under the test capacity. Note that for the pessimistic scenario, $l = 1, 2, 3$ are not feasible at the beginning, and $l = i$ is not feasible at day $i, i = 4, 5, 6, 7$. Our simulation study finds that the final prevalence rate is sensitive to the growth rate of infection (currently we assume epidemic doubling time of three days) and the initial prevalence rate, while insensitive to the test capacity. We summarize the testing protocol as follows: given the total number of people, estimate of the initial prevalence rate, estimate of the infection growth rate and the test capacity, run our simulation model to output the optimal testing cycle length and the optimal pool size.

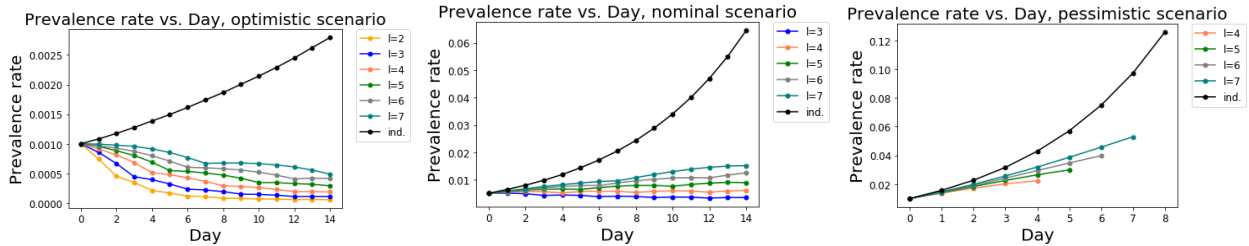


Figure 2: Prevalence rates of the mixed array method with different testing cycle length: (Left) optimistic scenario; (Middle) nominal scenario; (Right) pessimistic scenario.