FEAT: A Flexible, Efficient and Accurate Test for COVID-19

Linjiajie Fang[1], Bing-Yi Jing[2], Shen Ling[3], and Qing Yang[4] *‡§

April 29, 2021

Abstract

Early detection of COVID-19 is critical in mitigating the spread of the virus. Commonly used tests include nucleic acid detection, antibodies detection via blood testing and CT imaging. Some tests are accurate but time-consuming, while others are cheaper but less accurate. Exactly which test to use is constrained by various considerations, such as availability, cost, accuracy and efficiency.

In this paper, we propose a Flexible, Efficient and Accurate Test (FEAT). FEAT is based on group testing with simple but careful design by incorporating ideas such as close contact cliques and repeated tests. FEAT could dramatically improve the efficiency and/or accuracy for any existing test. For example, for accurate but slow test such as RT-PCR, FEAT can improve efficiency by multiple times without compromising accuracy. On the other hand, for fast but inaccurate tests, FEAT can sharply lower the false negative rates (FNR) and greatly increase efficiency. Theoretical justifications are provided. We point out some scenarios where the FEAT can be effectively employed.

Keywords: Epidemiology Control, Blood Testing, COVID-19, Testing Efficiency

*[1]Mr. Linjiajie Fang is a PhD student in Data Science and Analytics, Hong Kong University of Science and Technology Email: lfangad@connect.ust.hk

†[2]Prof. Bing-Yi Jing is a Professor in Mathematics, Hong Kong University of Science and Technology Email: majing@ust.hk (Corresponding Author)

‡[3]Ms. Shen Ling is a PhD candidate in Mathematics and Statistics, Dalhousie University, NS, Canada Email: lisaling@dal.ca

§[4]Ms. Qing Yang is a PhD candidate in Mathematics, Hong Kong University of Science and Technology Email: qyangad@connect.ust.hk

1 Introduction

Population screening is essential for COVID-19 control. Commonly used tests for COVID-19 fall into three main categories: virus detection, antibodies detection via blood testing and CT image recognition. Reverse Transcription Polymerase Chain Reaction (RT-PCR) is one of the main testing methods for detecting coronavirus, due to its high accuracy and convenience. However, RT-PCR is a technical intensive testing method, which requires professional technicians to perform in some specialized laboratories [1]. As a result, RT-PCR is relatively expensive and time-consuming [2]. Its high cost and long testing time prevent it from widespread testing, particularly in less developed regions.

Efforts have been made around the world to have alternative diagnostic test kits in order to make the COVID-19 testing faster and more affordable. For example, in USA, Abbott developed a portable device "ID NOW", which can detect positive results within 5 minutes, and negative results within 13 minutes [3]; in Bangladesh, a research group developed another fast and cheap test, known as "Dot blot test", which reportedly costs only USD 3 and delivers results under 15 minutes, and the testing can be readily conducted in most labs in the country [4]. However, faster tests may not necessarily guarantee accuracy. In fact, there remains much concern about the high false negative rate (FNR) of these rapid tests. "*In Procop's analysis, Abbott's test had a false-negative rate of 14.8 percent*", according to *The Scientists* [5]. The nearly 15% FNR would result in a large number of individuals released by mistake, which would have serious adverse effects.

Clearly, each diagnostic test has its own limitations, see Table 1 for a brief summary. Given these limitations, the natural question is whether it is possible to design some viable procedures to overcome the shortcomings of each diagnostic test. In our view, an ideal testing procedure should possess some, if not all, of the following metrics: accuracy, efficiency, flexibility and feasibility. The accuracy is measured by FNR and false positive rate (FPR). The efficiency is assessed by the expected number of tests per person. The flexibility requires the testing to be easily adjusted for different scenarios. The feasibility concerns the practical implementation of the strategy, such as being affordable and timely. A good testing should be accurate enough for infection control, be efficient so that it is economical and time-saving, be flexible for different regions, and be feasible to be implemented in reality.

Our answer to the above-mentioned question is affirmative. In this paper, we propose some *Flexible, Efficient and Accurate Tests* (FEAT), which can achieve all the good metrics a testing strategy should possess. For instance, FEAT1 is designed for those tests with high specificity and sensitivity, for the purpose of improving the efficiency and maintaining accuracy; FEAT2 is proposed to improve the accuracy as the first priority, and increase the efficiency.

The rest of this paper is organized as follows. Section 2 includes a literature review of group testing and strategies for COVID-19 specifically. Section 3 introduces the methodology and basic techniques of FEAT strategies. Section 4 provides both theoretical and empirical results of FEAT. We compare the performances of FEAT1, FEAT2 and Dorfman's group testing as well as individual one-by-one testing in our empirical study. Finally, a conclusion is provided.

	Accuracy	Speed	Cost
RT-PCR	high	4 -6 hours	high(USD 125)
ID NOW	low	15 mins	medium
Dot blot test	low	15 mins	low(USD 3)

Table 1: Pros and cons of three main tests. RT_PCR is accurate and expensive [6]. ID NOW and Bangladesh 3 dollar fast and cheap tests, but less accurate [3] [4].

2 Related Work

The implementation of group testing strategy could be traced back to Dorfman [7], who introduced an approach to economically test syphilis using grouped blood samples. Wolf [8] gave an overview of the early history following Dorfman's original work. In formal terms of group testing problem, the items that cause a group to test positive are called *defective* items. And in our paper, we treat people who are infected with COVID-19 as *defective*.

In general, group testing could be classified by their testing design, the assumption of the diagnostic tests and the goal of detection.

- **Adaptive vs. Non-adaptive:** In the adaptive setting, the tests are processed sequentially, where the latter tests depend on the results from previous results. As for non-adaptive group testing, test pools are designed in advance (*design*) and the determination of defective items (*decoding* or *detection*) is given after all testing results are known. We consider adaptive group testing problem in our case, since the necessity of further tests is determined by the previous results.

- **Noiseless vs. Noisy testing:** In noiseless setting, diagnostic tests are assumed to work perfectly. However, we aim to tackle more realistic situations. And one of our contribution is exactly concerning the strategy of using unreliable diagnostic tests to give reliable results. Thus we only consider noisy testings.

- **Zero-error vs. Small-error:** Generally, the potential goal of group testing problems is either to recover all defective items (*zero-error*) or to recover the defective set with "high probability" (*small-error*). In reality, it is impossible to find out all infected people without any error by the unreliable tests. On the other hand, it is also hard to specify which "high probability" of recovering the set of infected people is acceptable. Thus we follow realistic demands: we use false negative rate and false positive rate as criterion. The goal of our group testing problem is to control the level of both false negative rates and false positive rates (such as being comparable to a single RT-PCR test), and at the same time to reduce the total number of tests.

Many algorithms have been developed to solve noiseless non-adaptive group testing problems, such as Combinatorial optimal matching pursuit (COMP), Definite defective algorithm (DD), Sequential COMP (SCOMP) and Smallest satisfying set (SSS) [9] [10]. And the aforementioned methods have their corresponding noisy versions, such as: noisy DD and noisy COMP [10]. Sterrett [11], Li [12] and Finucan [13] also proposed adaptive variants of the original Dorfman's two-stage group testing. Binary-splitting algorithms [14] [15] is very efficient, yet it required too many testing stages even for small population, which is troublesome in process. Finuan [13] provided the theoretical proof of the optimal group size to achieve the best efficiency. De Bonis [16] proposed the estimation of minimum number of tests required for different number of stages with group testing. Patsula [17] compared the multi-stage grouping with some other adaptive testing approaches. Mcmahan [18] and Aprahamian [19] considered the heterogeneity of infections, which proposed a group testing strategy to test individuals based on the risk level.

By the high efficiency of group testing, it is considered an essential procedure to speed up COVID-19 detection. The implementation of group testing to detect COVID-19 was first suggested by researchers at Technion and Rambam Health Care Campus [20] [21] using RT-PCR. Nguyen [22] proposed the optimal pooling size based on the prevalence, with the consideration of disease progression and dilution of samples. Eberhardt [23] introduced a multi-stage grouping testing method under noiseless assumption without a discussion of the selection of the optimal size. Pouwels [24] discussed the necessity and potential improvements for group testing, yet no actions suggested further.

In our work, we propose a flexible multi-stage noisy adaptive group testing strategy to handle the practical population screening for COVID-19 infections. FEAT provides a flexible bunch of testing

strategies catering for different prevalence levels, available virus detection methods and the demands for screening accuracy from place to place. Repeated tests are applied for the purpose of sensitivity control. Taking FEAT1 and FEAT2 as example, we explain how FEAT can balance the trade-off between accuracy and efficiency under different scenarios. Considering the heterogeneity nature of infections, close contact cliques (CCCs) are implemented. Additionally, FEAT considers at most three stages to guarantee feasibility. And operational delay in sequential setting can be shrunk to acceptable level when CCCs applied and stage number controlled. The detail and examples are discussed in Section 3.

3 Methodology

3.1 Motivations and Key Components of FEAT

The successful implementation of FEAT hinges on the four key components: group testing structure, repeated tests, close contact cliques and flexibility of performance parameters. The usage of three-stage group testing structure guarantees high efficiency; repeated testing hedges the loss of sensitivity of group testing methods and can further enhance accuracy; applying close contact cliques improves the testing efficiency by considering the heterogeneity of infections; the flexibility of performance parameters makes it possible to easily apply FEAT in different situations.

Group Testing

The improvement in efficiency of Dorfman's group testing procedure is substantial. For example, if Dorfman's procedure was applied to RT-PCR tests in USA, the overall efficiency would be boosted seven more times than the current practice (Fig.8). Similar performance is also observed for other diagnostic tests.

However, one very undesirable drawback with naive use of Dorfman's procedure is that the FNR would be doubled: when Dorfman's two-stage group testing is adopted, the probability that a positive sample been falsely diagnosed as negative is

$$\gamma_- + (1-\gamma_-)\cdot\gamma_- \sim 2\gamma_-, \text{where } \gamma_- \text{ is the FNR of each diagnostic test.}$$

This problem is made even more pronounced when it is applied to tests with low sensitivity. For example, if the test kit has sensitivity level at 85%, or equivalently FNR at 15%, then Dorfman's procedure would lead to overall 30% FNR. Intuitively, regarding to each single positive individual to be tested by Dorfman's procedure, infections have twice chance to escape from being detected than individual tests. Similarly, despite the high efficiency, the direct use of N-stage group testing structure would amplify FNR to the order of $N \cdot \gamma_-$, which should not be ignored under the noisy setting.

Repeated Tests

To solve the increased FNR caused by group testing, we propose to use *repeated testing* procedure to either individuals or groups/subgroups level. When repeating the test once, we provide a positive result when there is either one test showing positive result. As a result, the false positives would be reduced significantly. For instance, if FNR = 15%, testing an individual twice would decrease FNR to 2.25%. A further theoretical discussion of FNR and FPR of FEAT is given in the *Appendix*.

Naturally, repeated tests would increase the total number of tests. The ensuing question is how to keep a balance between accuracy and efficiency, which is one of the focal points in this paper. We also provide flexible choices of FEAT1 and FEAT2 to encounter two situations where the accuracy of diagnostic tests are different.

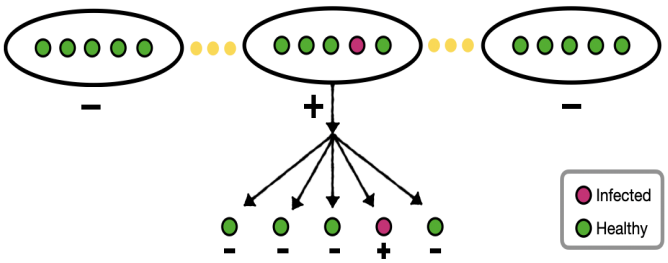


Figure 1: Dorfman's two-stage group testing. In this graph, the pooling size is five (m =5). The groups of samples are tested and positive groups are later tested individually.

Close Contact Cliques (CCCs)

COVID-19 is a disease related to respiratory system. Contact with carriers is one of the necessary conditions for the spread of the virus [25]. It is observed that many infections occur within *close contact cliques* (CCCs), e.g., within family members, friends, colleagues, passengers from the same flights. Under these circumstances, when applying group testing, it would make sense to test CCCs together. Our results show the efficiency can be improved significantly. It is worth mentioning that FEAT performs very well even without using information of CCCs (see Fig.7).

Feasibility and Flexibility

Feasibility and flexibility must be factored in any test strategy. In FEAT, we try to keep it as simple as possible so that it can be easily implemented in practice. Many existing designs are simply not feasible at all. For instance, the binary splitting method has been toyed around in the literature to minimize total number of tests [26], where individuals or groups, when tested positive, are further divided into equal halves, and so on. As a result, some individuals might need to undergo many rounds of tests, which we regard as unacceptable in practice. Along similar lines, when we employ repeated tests, we only need to require individuals to provide two independent samples at the same time. Such practice is very much standard in, say, providing urine samples for athletes in sporting competitions. Candidates are only required to gather A and B samples, then the repetition of gathering procedure is at most three times.

Another advantage of FEAT is its flexibility. Besides FEAT1 and FEAT2, users can choose any number of repetitions in each step by three parameters FEAT(a,b,c) with each number corresponding to the number of test repetition at each stage. In our case, FEAT1 and FEAT2 are the same as FEAT(1,2,2) and FEAT(2,2,2) respectively. The flexible choice of group size, i.e. the number of CCCs pooled in a group, can be adjusted according to the specific prevalence rate and the average size of such CCCs to achieve the best efficiency.

3.2 Methods

3.2.1 FEAT1 for tests with high accuracy and low efficiency

For diagnostic tests with high accuracy but low efficiency, such as RT-PCR, we propose FEAT1 to largely improve the efficiency, see Fig.2 for illustration. We first form small subgroups, which are chosen as CCCs if available (see details in Section 3.2.3). Examples of CCCs include close family members, close neighbors, passengers from the same flight, etc. These subgroups are then pooled together to form

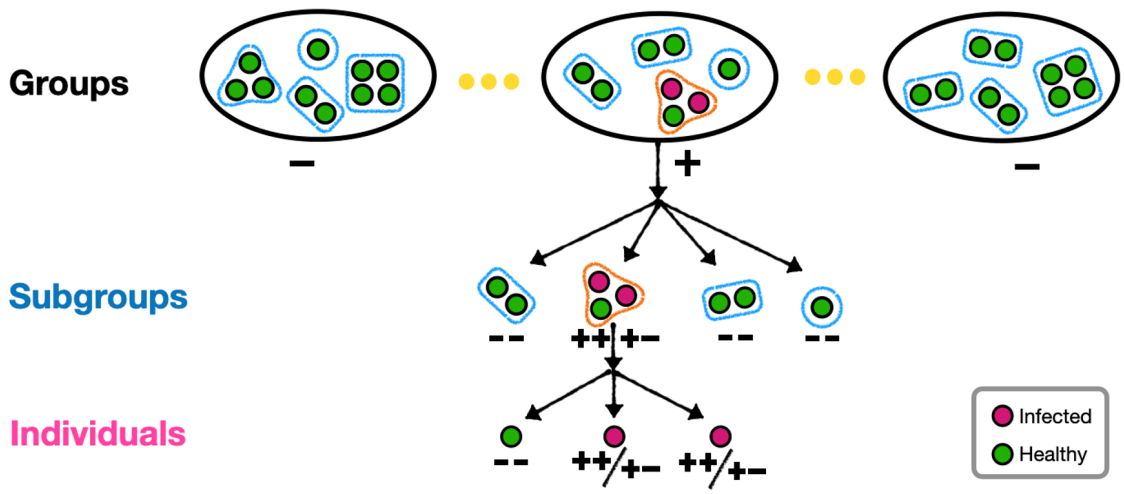


Figure 2: An illustration of FEAT1. Each group contains four CCCs ($k^* = 4$, where the subgroup sizes $m \in \{1, 2, 3, 4\}$). For group-level tests, FEAT1 performs one diagnostic test and stops the screening for negative groups. Subsequently, FEAT1 performs two diagnostic tests for each test simultaneously, and either one of the two diagnostic tests shows positive, such subgroup or individual would be labeled positive.

larger groups. Tests will be performed for each group. Testing stops at this first stage only when the group is tested negative; otherwise it will move to the second stage, where each subgroup will be tested twice. Testing stops at the second stage only when the subgroup is tested negative in both occasions (- -); otherwise testing continues into the third stage where each individual will be tested twice. Again, testing stops only when the individual is tested negative in both tests (- -). For simplicity, we will refer to this test strategy as FEAT1.

3.2.2 FEAT2 for tests with high efficiency and low accuracy

When the sensitivity of available diagnostic tests is relatively low (e.g. 85%), one can consider some variants of FEAT1. For instance, we could test the groups twice in the first stage (Fig.3), and testing stops only when both testing results are negative; otherwise testing continues into the second stage, where the procedures are the same as in FEAT1. For simplicity, we will refer to this test strategy as FEAT2. As we will see later, FEAT2 can work extremely well for low sensitivity test kits.

4 Theoretical Results

This section provides theoretical results of FEAT. Proofs of theorems are given in the *Appendix*.

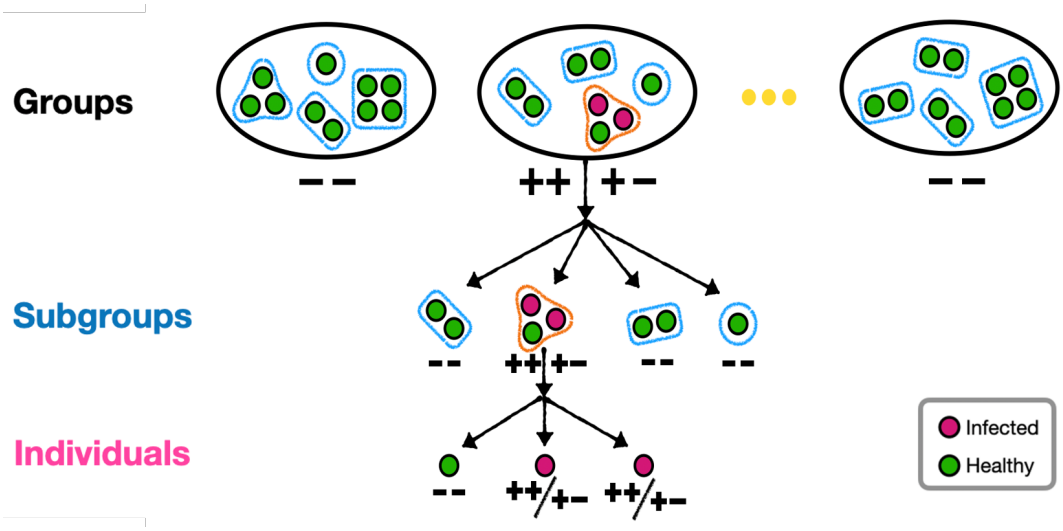


Figure 3: An illustration of FEAT2. Each group containing four CCCs ($k^* = 4$, where the subgroup sizes $m \in \{1, 2, 3, 4\}$). For group-level tests, FEAT2 performs two diagnostic tests and stops the screening for two negative results. Subsequently, FEAT2 performs two diagnostic tests for each subgroup/individual simultaneously, and either one of the two tests shows positive, such subgroup or individual would be labeled positive.

4.1 Notations

The notations used in the following sections are summarized as below:

ρ : prevalence of the whole population.

β : probability that an individual is infected by the infected source in the same subgroup.

k^* : number of subgroups in a group in three-stage group testing. See examples in Fig.2, Fig.3.

m : size of a subgroup.

$\bar{m}$: average size of subgroups.

k : average size of groups. It satisfies $k = m \cdot k^*$.

ρ^* : fraction of infected subgroups. Also, $\rho^* = \dfrac{\bar{m}}{1 + (\bar{m} - 1)\beta}\rho$, (see *Appendix)* .

γ_- : FNR of each diagnostic test.

γ_+ : FPR of each diagnostic test.

γ_-^* : FNR of the overall group testing.

γ_+^* : FPR of the overall group testing.

4.2 Optimal Pooling Sizes

The optimal number of subgroups in each group (see Fig.2, Fig.3) is closely related to the testing efficiency (see Fig.4, Fig.5). We extend the relationship between the group size and the prevalence level of two-stage methods in Finucan (1964) [13] to FEAT1 and FEAT2 respectively. The subgroups are naturally determined by clusters of people. For example, to screen a community, the subgroups are

simply families; to screen travelers in the airports, the subgroups are people having close seat positions. We aim to find the number of subgroups per group to maximize the expected number of persons per test, i.e., to maximize the testing efficiency.

We refer the number of subgroups k^* that maximizes the efficiency as "optimal", and denote that number as k^*_{opt}.

Theorem 4.1 *For FEAT1, given the prevalence of subgroups ρ^*, the optimal number of subgroups per group k^*_{opt} is*

$$k^*_{opt} \approx 1/\sqrt{2\rho^*}.$$

In our simulation (Fig.4), we assume $\bar{m} = 4.76$ [27], $\beta = 0.7$, $\rho = 0.005$, thus $\rho^* \approx 0.006$. We can quickly compute $k^*_{opt} = 1/\sqrt{2 * 0.006} \approx 9$, equivalently to pool $9 * 4.76 \approx 43$ people in a group, which is the same as our simulation result.

Theorem 4.2 *For FEAT2, the optimal number of subgroups per group k^*_{opt} is*

$$k^*_{opt} \approx 1/\sqrt{\rho^*}.$$

Under the same setting in FEAT1, k^*_{opt} of FEAT2 is estimated by $1/\sqrt{0.006} \approx 13$, or equivalently to pool 63 people in a group, which is a little different from the simulation results in Fig.5. Fig.5 shows that it is robust for FEAT2 to apply group size that differs one or two subgroups from the optimal while maintain almost the best efficiency.

4.3 Prevalence Condition for Efficiency

For Dorfman's two-stage group testing, it has been shown that it is better than individual testing if and only if $\rho < 1 - 1/3^{1/3} \approx 0.306$ [28]. It is natural to ask under which conditions the FEAT strategy is more efficient than testing individually. Here we give the results:

Theorem 4.3 *Suppose the diagnostic tests are perfect and the population is large enough. Then,*

- *FEAT1 is more efficient than individual testing if and only if:*

$$\rho < \frac{1 + (\bar{m} - 1)\beta}{\bar{m}} \left(\frac{1}{2} - \frac{1}{1 + \sqrt{1 + \bar{m}^2}} \right).$$

- *FEAT2 is more efficient than individual testing if and only if:*

$$\rho < \frac{1 + (\bar{m} - 1)\beta}{\bar{m}} \left(\frac{1}{2} - \frac{1}{1 + \sqrt{1 + \bar{m}^2/2}} \right).$$

The upper bound for the prevalence is determined by the size of subgroups $\bar{m}$ as well as the infection rate β. As illustration, if $\bar{m} = 4$ and $\beta = 0.7$, FEAT1 requires $\rho < 0.24$ and FEAT2 requires $\rho < 0.19$ to be efficient. If $\bar{m}$ is changed to $\bar{m} = 3$, such requirements is changed to $\rho < 0.21$ and $\rho < 0.16$ respectively.

4.4 Accuracy of FEAT

We analyse the accuracy of FEAT by the overall FNR γ_-^* and FPR γ_+^*.

Theorem 4.4 *Suppose ρ is small, the order of the accuracy of FEAT is given by:*

- *For FEAT1:*

$$\begin{cases} \gamma_-^* \sim \gamma_- \ , \\ \gamma_+^* \sim 2\rho(1-\beta)(\bar{m}-1)\gamma_+ \ . \end{cases}$$

- *For FEAT2:*

$$\begin{cases} \gamma_-^* \sim 3\gamma_-^2 \ , \\ \gamma_+^* \sim 2\rho(1-\beta)(\bar{m}-1)\gamma_+ \ . \end{cases}$$

FEAT1 maintains the same order of FNR as individual testing while FEAT2 lowers the FNR to the second order. Although FEAT1 does not reduce FNR as FEAT2, FEAT1 is much more efficient without performing repeated testing in the first stage. The choice of FEAT1, FEAT2 or even other variants depends on the accuracy of diagnostic tests.

5 Empirical Results

To evaluate the performance of FEAT, we first perform empirical study of FEAT1 and FEAT2 for data of Bangladesh. In our study, we try to use real data as much as possible, such as: the prevalence, household distribution and accuracy of diagnostic tests. For other parameters. we make reasonable assumptions followed by sensitivity analysis. Finally, FEAT is implemented of FEAT for five countries, and compared with Dorfman's group testing.

Numerical Setup

In Fig.4 and Fig.5, we assume prevalence $\rho = 0.005$ and household distribution of Bangladesh is based on data [29]. That is, we choose the number of family members from $\{1, 3, 5, 7\}$ with probabilities equal to $\{0.02, 0.30, 0.45, 0.23\}$ respectively, and the average family size is 4.76. We also assume the infection rate within close contact cliques to be $\beta = 0.7$.

We first generate 20,000 families (as subgroups) with the fraction of infection equal to 0.006. Then we generate the number of members and infections of each family by the household distribution and the infection rate to make sure the overall individual prevalence is 0.005. We use pooling size, referring to the number of individuals in each group, as the changing performance parameter in comparing performances of different methods (Fig.4, Fig.5). For each pooling size, we simulate 50 replicates and evaluate the average performance.

In Fig.7, we follow the same configuration as in Fig.4 except for changing the infection rate β from 0 to 1 with a gap 0.1. For each β, we evaluate performances of each method by their optimal pooling sizes. Similarly, in Fig.6 we fix $\beta = 0.7$ and change prevalence from 10^{-4} to 10^{-2} with a gap of log-scale equal to 0.2 in order to evaluate the best performance on different prevalence levels. To avoid the variance when searching for the best efficiency, we simulate 30 replicates of simulation set containing 20,000 families and the average performance is compared.

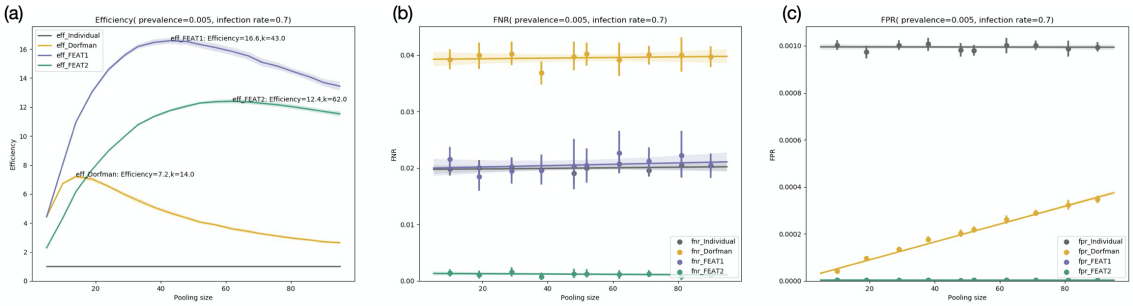


Figure 4: Comparisons of FEAT1, FEAT2, Dorfman's method and the individual test on 50 simulation sets. Each simulation set contains 20,000 families with the distribution of household sizes following data in Bangladesh [29]. The FNR and the FPR of each single test are supposed to be 0.02 and 0.001 respectively. We assume the individual prevalence $\rho = 0.005$ and the infection rate within close contact clique $\beta = 0.7$. We evaluate the efficiency and accuracy along different pooling sizes. (a): The efficiencies versus pooling sizes. (b): The FNR versus pooling sizes. (c): The FPR versus pooling sizes. The FPR of both FEAT1 and FEAT2 overlaps each other.

Results for Testing with High Specificity and Sensitivity

Fig.4 shows the results when accurate tests, such as RT-PCR, are used. We assume the FNR = 0.02 and FPR=0.001. We simulate our data following the household distribution of Bangladesh under prevalence rate 0.005, the results in other places are similar and the optimal performances are summarized in Fig.8, will be discussed later. Thus we make the following observations.

- FEAT1 and FEAT2 are over 16 and 12 times efficient as the individual test with optimal pooling sizes to be 43 and 62 respectively, as shown in Fig.4(a). The maximum efficiency of Dorfman's method is 7.2, which is far less than FEAT. Here efficiency is measured by the average number of persons implemented for each single test.
- In terms of the accuracy, Dorfman's method doubles the FNR from 0.02 to 0.04. FEAT1 nicely controls the FNR to be the same level of the individual test. Although FEAT2 is less effective than FEAT1, the FNR of FEAT2 is extremely small, see Fig.4(b).
- The FPRs of FEAT1 and FEAT2 are nearly zero and overlap each other, and the FPR of Dorfman's method is almost linear in pooling size. Fig.4(c) shows that the FPR of all the group testing methods are far less than the individual test.
- Overall, FEAT1 could improve efficiency remarkably and at the mean time maintain the high accuracy.

Results for Testing with Lower Specificity and Sensitivity

If RT-PCR is not available or not affordable in some nations or regions, we would consider testing strategies applying only cheap but inaccurate tests. Fig.5 shows the performances of group testing strategies applying inaccurate tests. The configuration of the simulation is the same as above except that we assume FNR=0.15 and FPR=0.05. We make following remarks.

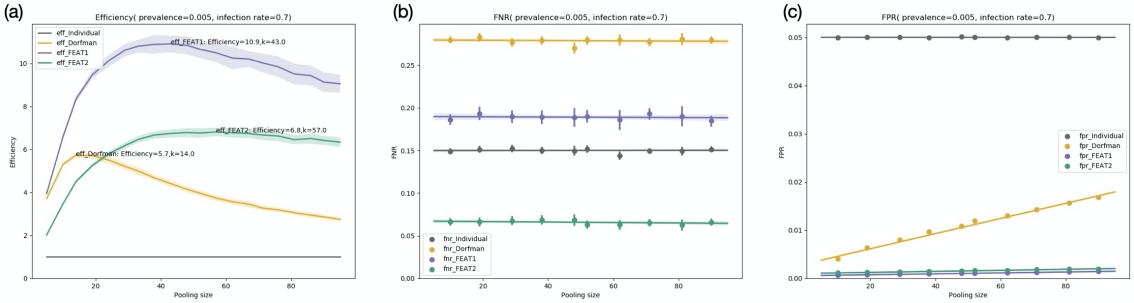


Figure 5: Comparisons of FEAT1, FEAT2, Dorfman's method and the individual test on 50 simulation sets. The configuration is the same as Fig.4 except that the FPR and FNR are assumed to be 0.15 and 0.05 respectively. (a): The efficiency versus pool size. (b): The FNR versus pool size. (c): The FPR versus pool size.

- FEAT1, FEAT2 and Dorfman's method reach maximum efficiency 10.9, 6.8 and 5.7 respectively, refering to Fig.5(a). It is noticed that efficiencies under inaccurate tests are less than the scenario when accurate tests are applied, as higher FPR (0.05 v.s. 0.001) could lead to redundant tests on groups with no infection.
- From Fig.5(b). Dorfman's method even doubles the FNR from 0.15 to 0.30, which is unacceptable. FEAT1 could no longer control the FNR to be the same as the individual test in such scenario, which is increased to be around 0.2.
- FEAT2 reduces the FNR substantially from 0.15 to 0.0675. In terms of the FPR, Fig.5(c) shows that FEAT have consistently better sensitivity than Dorfman's method.
- FEAT2 can leverage cheap inaccurate tests and largely reduces the overall FNR. For example, if FNR=0.10 for the diagnostic test, then FEAT2 can reduce the FNR to 0.03; and if FNR=0.05, FEAT2 can reduce it to 0.0075. Details of the calculation are given in *Appendix*.

Effects of the Prevalence

Fig.6 compares the performance of FEAT1, FEAT2 and Dorfman's method under different prevalence levels. To control the number of parameters, we fix the infection rate within close contact clique $\beta = 0.7$ and change the prevalence from 0.01% to 1%. Fig.6(a) shows that optimal pooling sizes of all group testing methods decreases when the prevalence increases. Fig.6(b) shows that the efficiencies of three methods also decrease with prevalence. The smaller the prevalence, the more efficient the group testings should be.

Effects of the Infection Rate within Subgroups

Fig.7 compares the performance of FEAT1, FEAT2 and Dorfman's method with optimal pooling sizes under different infection rates $\beta \in [0, 1]$. Here we fix the prevalence $\rho = 0.005$. Fig.7(a) shows that FEAT methods always pool more samples than Dorfman's method and the pooling size is monotonously increasing with respect to β. Fig.7(b) shows that when $\beta \geq 0.15$, FEAT2 would be more efficient than Dorfman's method, and FEAT1 always performs better than Dorfman's regardless of β. Since SARS-CoV-2 spreads very easily and sustainable between people [30], it is reasonable to speculate that FEAT

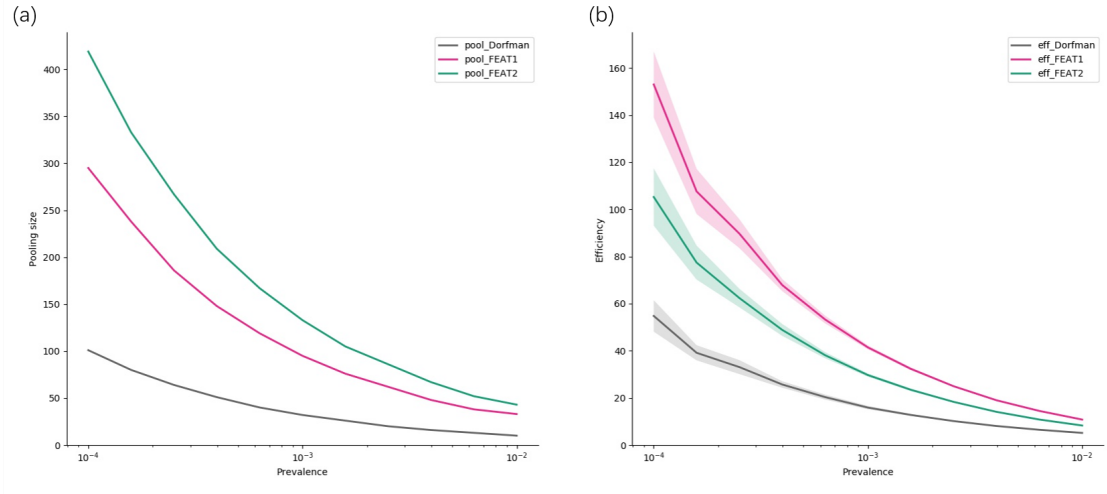


Figure 6: The best performance of FEAT1, FEAT2 and Dorfman's method under different prevalence. The infection rate is fixed to be the same as Fig.4.(a): The optimal pooling sizes of three methods versus different prevalence. (b): The efficiencies of three methods versus different prevalence.

would be more efficient in reality.

Implementation of FEAT

In Fig.8, we evaluate the optimal performances of FEAT in different countries. The number of confirmed cases, death and recovered (data collected from [31] before June, 2020) and the population [32] are used to calculate prevalence in each country. The distribution of households is gathered from United Nations reports in 2017 [29]. For simplicity, here we compare FEAT1 equipped with accurate diagnostic tests (FNR =0.02, FPR=0.001), FEAT2 equipped with inaccurate diagnostic tests (FNR=0.15, FPR=0.05), Dorfman's with accurate tests (FNR =0.02, FPR=0.001, denoted by DT-ac) and Dorfman's with inaccurate tests (FNR =0.15, FPR=0.05, denoted by DT-inac). Group testing methods with cheap but inaccurate tests are shadowed.

Fig.8 provides FEAT solutions for five countries whose number of confirmed cases is large. Comparing FEAT1 with DT-ac, The accuracy of all five countries are well controlled, where FNR is around 0.02 and FPR is less than 0.0001. In terms of efficiency, FEAT1 surpasses DT-ac, and averagely doubles the efficiency. Additionally, in some countries with larger average family size such as Bangladesh, the efficiency of FEAT2 is as good as DT-inac. However, its accuracy can be largely improved.

6 DISCUSSION

How Accurate Can Group Testing Achieve?

If repetition is not adopted, three-stage group testing, which applies to the groups of subgroups, is much more efficient than two-stage grouping. However, compared with individual tests, two-stage grouping

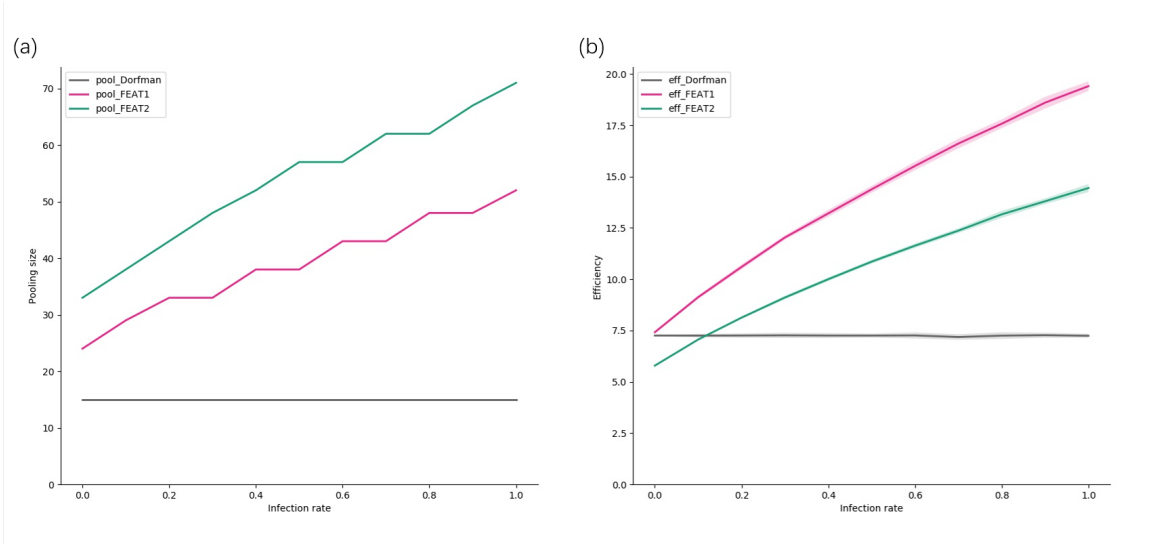


Figure 7: The best performance of FEAT1, FEAT2 and Dorfman's strategy under different infection rate within close contact clique. The prevalence is fixed to be the same as Fig.4. (a): The optimal pooling sizes versus infection rate. (b): The best efficiencies versus infection rate.

(like Dorfman's method) doubles FNR, and its FPR is nearly linear to the group size. Three-stage grouping triples FNR, and their FPR is nearly linear to the subgroup size. These facts direct us to avoid overmuch number of stages and prevent pooling too many individual samples in the last stage of group testing strategy.

To alleviate the problem of multiplying FNR without loss of much efficiency, FEAT adopts a three-stage grouping structure and performs repeated tests at proposed levels, whose FNR is controlled to be at the level of second order and FPR is stable along different pooling sizes. Although repetition intuitively increases the total number of tests, by virtue of three-stage structure and infection information, FEAT maintains comparable efficiency as Dorfman's method.

Necessity of Group Testing

When the prevalence becomes very high, both efficiency and accuracy of group testing will be greatly reduced. Thus we have to consider if it is necessary to implement group testing when the level of prevalence is severe. Fig.9 provides theoretical curves of naive two-stage grouping testing with repeated tests, which suggests that when prevalence is over 10%, the number of tests of grouping patients would outnumber the individual tests. It could be a loss of the original intention for designing group testing. For areas with very severe epidemics, small pooling size will be appreciated or just remain as simple as individual testing. However, to the best of our knowledge, prevalence in most areas for COVID-19 is about 1% [33] or even lower, so the necessity of group testing is still evident for COVID-19 pandemic.

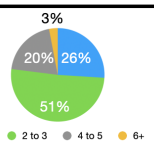
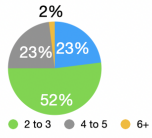
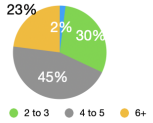
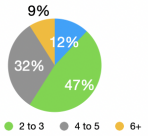
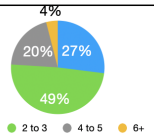
Countries	Prevalence	Distribution of Household	Methods	Optimal Pooling Size	Efficiency	Average FNR	Average FPR
Russia	0.001621		FEAT1	48	21.7(0.3)	0.026(0.019)	<0.0001(0.0001)
			FEAT2	69	6.7(0.2)	0.073(0.022)	0.0013(0.0001)
			DT-ac	25	12.6(0.2)	0.039(0.012)	<0.0001(0.0001)
			DT-inac	25	8.2(0.2)	0.28(0.03)	0.0040(0.0002)
Spain	0.001438		FEAT1	52	23.5(0.2)	0.020(0.012)	<0.0001(0.0001)
			FEAT2	74	7.0(0.3)	0.064(0.029)	<0.0001(0.0001)
			DT-ac	27	13.2(0.3)	0.046(0.017)	<0.0001(0.0001)
			DT-inac	27	8.5(0.3)	0.27(0.03)	0.0040(0.0003)
Bangladesh	0.000372		FEAT1	157	69.5(1.8)	0.013(0.022)	<0.0001(0.0001)
			FEAT2	219	12.33(0.86)	0.081(0.049)	0.0011(0.0001)
			DT-ac	52	25.8(1.0)	0.041(0.018)	<0.0001(0.0001)
			DT-inac	52	11.9(0.4)	0.28(0.05)	0.0033(0.0002)
Brazil	0.002034		FEAT1	53	23.0(0.4)	0.021(0.018)	<0.0001(0.0001)
			FEAT2	75	7.4(0.4)	0.074(0.033)	0.0014(0.0002)
			DT-ac	23	11.2(0.3)	0.041(0.012)	<0.0001(0.0001)
			DT-inac	23	7.7(0.3)	0.27(0.03)	0.0043(0.0003)
USA	0.004538		FEAT1	30	12.4(0.1)	0.022(0.011)	<0.0001(0.0001)
			FEAT2	42	5.0(0.1)	0.064(0.013)	0.0016(0.0001)
			DT-ac	15	7.6(0.1)	0.039(0.006)	<0.0001(0.0001)
			DT-inac	15	5.9(0.1)	0.28(0.02)	0.0049(0.0003)

Figure 8: This table provides the FEAT and Dorfman's two-stage group testing (DT-ac/DT-inac) results for five countries: Russia, Spain, Bangladesh, Brazil and United States of America (USA). The quantities are given by mean values with standard deviation in brackets. FEAT1 and DT-ac use accurate diagnostic tests(FNR =0.02, FPR=0.001). The FEAT2 and DT-inac use cheap but inaccurate diagnostic tests (FNR =0.15, FPR=0.05), whose rows are shadowed to differ from methods with accurate diagnostic tests. The infection rate within close contact cliques is supposed to be 0.7.

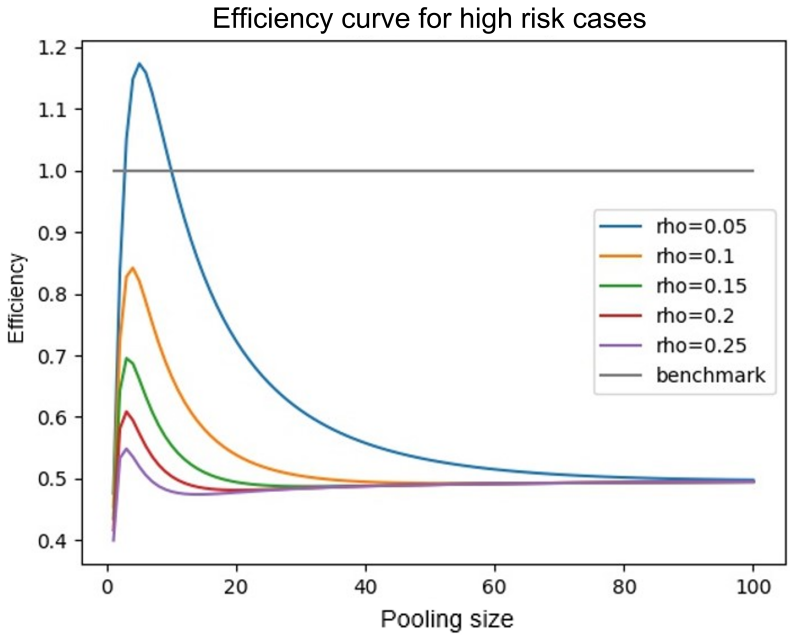


Figure 9: The efficiency of two-stage group testing method with repeated tests when the prevalence level is high.

Determination of Close Contact Cliques

The concept of close contact cliques can be freely generalized to any clique of individuals with close contacts, not only families. For example, the residential relations, roommates staying in the same dormitory unit. Furthermore, when applying group testing strategies in airport to diagnose the virus status of commuters, passengers sitting in front and back rows on the same plane can be pooled in the same close contact clique.

Unlike Dorfman's method, which pools a fixed number of individuals, FEAT pools samples by a fixed number of close contact cliques. Since the size of close contact cliques is flexible, FEAT pools different number of samples from group to group, which concerns more about the contact information than restrictions on specific pooling size. In this paper, the contact between groups or subgroups is ignored. However, in reality, the contact between people is much more complicated. The contact information of patients plays an important role to distinguish high risk susceptible patients from low risk ones. A flexible grouping strategy based on contact network might help cluster the infected patients to certain pools and largely reduce the number of tests.

Potential Operational Delay

Group testing is a sequential testing strategy, where operational delay could be a natural concern. In this part, we would explain there should be few individuals delayed by three-stage structure. And under current quarantine, delay will not lead to any inconvenience in some situations.

First, compared to some multi-stage group testing methods, FEAT allows at most two times delay

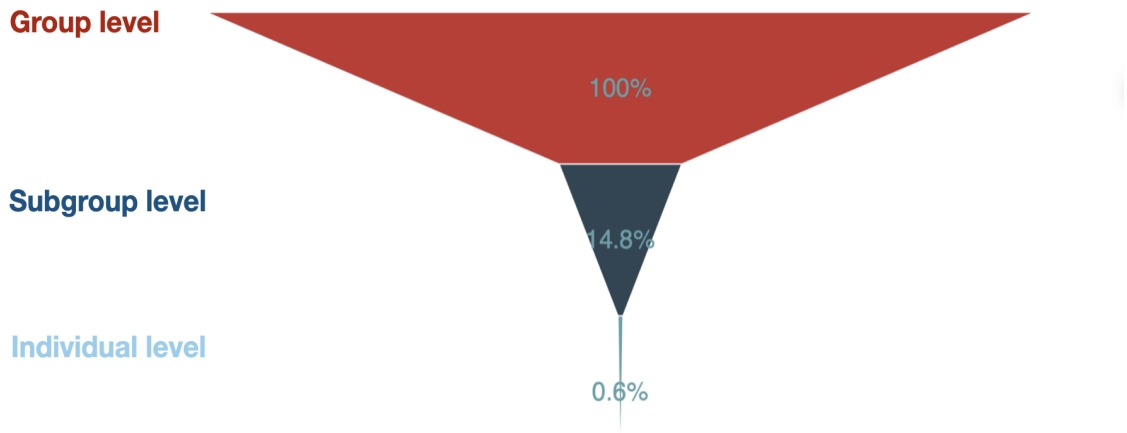


Figure 10: The operational delay of FEAT. We assume prevalence $\rho = 0.005$, group size $k = 32$ and the diagnostic tests are perfect. Only a very small fraction of people ($< 1\%$) will take all three stage tests.

from individual tests. Fig.10 shows the operational delay of FEAT when $\rho = 0.005$ and $k = 32$. Only less than 1 percentage of the whole population would be required to take the third-stage test. However, binary splitting for group tests delays four times when the population size is only 30. On the other hand, FEAT is a set of methods introduced for testing strategies, where users also can implement two-stage structure to reduce the total operational delay.

Second, under current quarantine policy, international travellers are mandatory to take at least 2-week quarantine in the hotel when access to China. In this situation, three-stage group testing does no harm, as infected cases have no chance to contact with others before confirmed.

Next, FEAT as well as other group testing based strategies is design for population screening, instead of using with high-risk cohorts or in clinical settings to conclude the viral status of potential patients who already have symptoms. For that reason, the overall prevalence of population should not be high. With the low prevalence, the majority of population should be screened out as negative at the first stage of group testing, where there would be no operational delay.

"Predictive" Group Testing

In some areas, government would propose universal testing among all the population, where millions of individuals could be involved. In order to perform a more efficient testing, it is prioritized to test people who have high probability of getting infected. Before testing, we can make prediction of the risk level for someone based on their contact information, symptoms and residential areas. Such risk prediction is usually non-uniform among different groups of people. In the future, we hope to develop methods to predict the risk of people and build testing strategies based on different predictions.

7 Conclusion

We proposed the FEAT procedures by incorporating several ideas such as group testing, close contact cliques (CCCs) and repeated tests. Different versions of FEAT can be designed to cater for different sce-

narios. We designed two such versions, FEAT1 and FEAT2, corresponding to several different diagnostic testing kits. FEAT can increase the overall testing efficiency and accuracy. The improvement in efficiency can reduce the number of kits required for a specific size of population. In other words, the financial burden for testing will be reduced significantly. Additionally, the contribution of increasing accuracy is also meaningful. Highly accurate test methods can be applied in some areas when those accurate tests are not available. In terms of testing flexibility, different versions of FEAT (e.g. FEAT1 and FEAT2) can cater for different situations to balance both accuracy and efficiency. Finally, the mechanisms and application in countries are described in detail in this paper.

8 Appendix

8.1 Proofs

Proof of Theorem 4.1. For simplicity, we assume each *subgroup* has the same number of individuals. Once the subgroup is tested to be positive, the number of tests implemented for each subgroup (in the third stage) is a constant. We also assume the population size is large enough. In a group with k^* subgroups, the average number of tests performed for FEAT1 in the first and the second stage is given by:

$$\begin{aligned} T &= 1 + \mathbb{P}(\text{containing infections})(1-\gamma_-)\cdot 2k^* + \mathbb{P}(\text{no infections})\cdot\gamma_+\cdot 2k^* \\ &= 1 + (1-(1-\rho^*)^{k^*})(1-\gamma_-)\cdot 2k^* + (1-\rho^*)^{k^*}\cdot\gamma_+\cdot 2k^* \\ &= (1-(1-\rho^*)^{k^*})(1-\gamma_- - \gamma_+)\cdot 2k^* + \gamma_+\cdot 2k^* + 1. \end{aligned}$$

An optimal k^*_{opt} is selected to minimize the average number of tests per subgroup T/k^*, i.e.

$$k^*_{opt} = \text{argmin}_{k^*} 2(1-(1-\rho)^{k^*})(1-\gamma_- - \gamma_+) + \frac{1}{k^*} + const.$$

Since $(1-(1-\rho^*)^{k^*}) \approx \rho^* k^*$ for small ρ^* by Taylor expansion, we aim to find

$$k^*_{opt} = \text{argmin}_{k^*} 2\rho k^*((1-\gamma_- - \gamma_+) + \frac{1}{k^*},$$

which gives $k^*_{opt} = 1/\sqrt{2\rho^*(1-\gamma_- - \gamma_+)} \approx 1/\sqrt{2\rho^*}$ when γ_-, γ_+ are small. □

Proof of Theorem 4.2. Similar to the proof of theorem 4.1, the average number of tests performed for FEAT2 in the first and the second stage of a group is given by:

$$T = 2 + \mathbb{P}(\text{containing infections})(1-\gamma_-^2)\cdot 2k^* + \mathbb{P}(\text{no infections})\cdot(2\gamma_+ - \gamma_+^2)\cdot 2k^*.$$

The optimal number of subgroups for each group k^*_{opt} is given by:

$$k^*_{opt} = \text{argmin}_{k^*} 2\rho^* k^*(1 - 2\gamma_+ - \gamma_-^2 + \gamma_+^2) + \frac{2}{k^*},$$

which gives $k^*_{opt} = 1/\sqrt{\rho^*(1-2\gamma_+ - \gamma_-^2 + \gamma_+^2)} \approx 1/\sqrt{\rho^*}$ when γ_-, γ_+ are small. □

Proof of Theorem 4.3. For FEAT1, the expected number of tests taken in a group is given by:

$$\begin{aligned}
T &= 1+\sum_{i=1}^{k^*}\binom{k^*}{i}(\rho^*)^i(1-\rho^*)^{k^*-i}(2k^*+2i\bar{m}) \\
&= 1+2\sum_{i=0}^{k^*}\binom{k^*}{i}(\rho^*)^i(1-\rho^*)^{k^*-i}(k^*+i\bar{m})-2(1-\rho^*)^{k^*}k^* \\
&= 1+2k^*+2\bar{m}k^*\rho^*-2(1-\rho^*)^{k^*}k^* \\
&\approx 1+2k^*+2\bar{m}k^*\rho^*-2(1-k^*\rho^*)k^* \\
&= 1+2\bar{m}k^*\rho^*+2(k^*)^2\rho^*.
\end{aligned}$$

Since a group has $\bar{m}k^*$ individuals, the expected number of tests per person is given by:

$$T/\bar{m}k^* = \frac{1}{\bar{m}k^*}+2\rho^*+\frac{2k^*\rho^*}{\bar{m}}.$$

Replace k^* by $k^*_{opt}=1/\sqrt{2\rho^*}$,

$$T/\bar{m}k^* = \frac{2\sqrt{2\rho^*}}{\bar{m}}+2\rho^*.$$

FEAT1 is more efficient than individual testing if and only if $T/\bar{m}k^*<1$, which gives

$$\rho^* < \frac{1}{2}-\frac{1}{1+\sqrt{1+\bar{m}^2}}.$$

Recall that $\rho^*=\frac{\bar{m}}{1+(\bar{m}-1)\beta}\rho$, we have

$$\rho < \frac{1+(\bar{m}-1)\beta}{\bar{m}}\left(\frac{1}{2}-\frac{1}{1+\sqrt{1+\bar{m}^2}}\right).$$

Similarly, FEAT2 is better than individual test if and only if

$$\rho < \frac{1+(\bar{m}-1)\beta}{\bar{m}}\left(\frac{1}{2}-\frac{1}{1+\sqrt{1+\bar{m}^2/2}}\right). \qquad \square$$

8.2 Analysis of group testing accuracy

In this part we provide detailed analysis of the accuracy of group testing strategies to further illustrate the performance in our simulations. When repeated tests are performed, we always apply "once-positive" rule, i.e., a group/subgroup/individual is labeled as positive if either one of the two diagnostic tests is positive.

- **The individual testing**

$$\begin{cases}\gamma^*_- = \gamma_- \ , \\ \gamma^*_+ = \gamma_+ \ .\end{cases}$$

  Since the false positive rate and false negative rate are directly determined by the accuracy of a single RT-PCR test.

- **Two-stage strategy with single testing**:

$$\begin{cases} \gamma_-^* \sim 2\gamma_- \quad , \\ \gamma_+^* \sim \rho(k-1)\gamma_+ \quad . \end{cases}$$

For two-stage strategy like Dorfman's procedure, each false negative could result from a false negative group testing or, successfully detected in group testing while failed in its individual testing. So $\gamma_-^* = \gamma_- + (1-\gamma_-)\gamma_- \sim 2\gamma_-$. A false positive could come from an infected group or uninfected group. $\gamma_+^* = \mathbb{P}(\text{Healthy, in an infected group})(1-\gamma_-)\gamma_+ + \mathbb{P}(\text{Healthy, in an uninfected group})\gamma_+^2 = (1-(1-\rho)^{k-1})\gamma_+(1-\gamma_-) + (1-\rho)^k\gamma_+^2 \sim \rho(k-1)\gamma_+$ for small ρ by Taylor expansion. This explains the simulation results of two-stage methods that γ_+^* is almost linear in k when ρ is small.

- **Three-stage strategy with single testings**:

$$\begin{cases} \gamma_-^* \sim 3\gamma_- \quad , \\ \gamma_+^* \sim \rho(m-1)\gamma_+ \quad . \end{cases}$$

Three-stage strategy with single testings models the case if FEAT does not adopt repeated tests nor infection information in subgroups. For three-stage methods, $\gamma_-^* = \gamma_- + (1-\gamma_-)\gamma_- + (1-\gamma_-)^2\gamma_- \sim 3\gamma_-$. As for false positives, if he comes from a non-infected group, the false positive rate is at order γ_+^3. If he comes from a non-infected subgroup and at the same time in an infected group, the false positive rate is at order γ_+^2. So the decisive order is given by a false positive when he is in an infected subgroup.$\gamma_+^* \sim \mathbb{P}(\text{An healthy person in an infected subgroup})(1-\gamma_-)^2\gamma_+ \sim (1-(1-\rho)^{\bar{m}-1})\gamma_+ \sim \rho(\bar{m}-1)\gamma_+$. This explained why three-stage procedures have low and stable false positive rates.

The false negative rate of group testing strategies without repetition would be doubled or tripled depending on the number of stages, which confirms our simulation results.

- **Two-stage strategy with double testings**:

$$\begin{cases} \gamma_-^* \sim 2\gamma_-^2 \quad , \\ \gamma_+^* \sim 2\rho(k-1)\gamma_+ \quad . \end{cases}$$

Applying repetition with "once-positive" rule amplifies false positive rate from γ_+ to $\gamma_+^2 + 2\gamma_+(1-\gamma_+) \sim 2\gamma_+$, while decrease false negative rate from γ_- to γ_-^2 compared with a single dianostic test.

For two-stage methods, $\gamma_-^* = \gamma_-^2 + (1-\gamma_-^2)\gamma_-^2 \sim 2\gamma_-^2$, which is controlled far below the false negative rate of widely used individual test at a higher order. $\gamma_+^* \sim \mathbb{P}(\text{An healthy person in an infected group})(1-\gamma_-^2) * 2\gamma_+ + \mathbb{P}(\text{An healthy person in an uninfected group}) * (2\gamma_+)^2 \sim 2\rho(k-1)\gamma_+$ which doubles the false positive rate than before. However, since prevalence ρ is small in most cases, false positive rate is well controlled.

- **Three-stage strategy with double testings**:

$$\begin{cases} \gamma_-^* \sim 3\gamma_-^2 \quad , \\ \gamma_+^* \sim 2\rho(m-1)\gamma_+ \quad . \end{cases}$$

For three-stage strategies with uniform distributed infections, $\gamma_-^* = \gamma_-^2 + (1-\gamma_-^2)\gamma_-^2 + (1-\gamma_-^2)^2\gamma_-^2 \sim 3\gamma_-^2$. False negative rate of three-stage strategies shrinks to order two. For $\gamma_- = 0.02$ in our cases, three-stage strategies have 1/16 false negative rate compared to the individual test. As for false positives, when false positives appears in an uninfected group or subgroup, false positive rates

changed to $(2\gamma_+)^3 \sim 8\gamma_+^3$ or $(2\gamma_+)^2 \sim 4\gamma_+^2$ respectively. Even though they are higher than the scenarios when repetitions are not performed, such increasing in false positives is still negligible by their orders. The decisive order of γ_+^* is: $\gamma_+^* \sim \mathbb{P}(\text{An healthy person in an infected subgroup})(1-\gamma_-^2)^2 \cdot 2\gamma_+ \sim (1-(1-\rho)^{m-1}) \cdot 2\gamma_+ \sim 2\rho(m-1)\gamma_+$.

- **FEAT (proof of theorem 4.4)**:

  For FEAT1:

$$\begin{cases} \gamma_-^* \sim \gamma_- \ , \\ \gamma_+^* \sim 2\rho(1-\beta)(\bar{m}-1)\gamma_+ \ . \end{cases}$$

  For FEAT2:

$$\begin{cases} \gamma_-^* \sim 3\gamma_-^2 \ , \\ \gamma_+^* \sim 2\rho(1-\beta)(\bar{m}-1)\gamma_+ \ . \end{cases}$$

  The discussion of the false negative rate of FEAT2 is the same as other three-stage strategies. For FEAT1, $\gamma_-^* = \gamma_- + (1-\gamma_-)\gamma_-^2 + (1-\gamma_-)(1-\gamma_-^2)\gamma_-^2 \sim \gamma_-$. As for the false positive rate, we only need to figure out the probability of an healthy person in a positive subgroup with non-zero infection rate. Once the probability is determined, $\gamma_+^* \sim \mathbb{P}(\text{An healthy person in an infected subgroup}) \cdot 2\gamma_+$ for both FEAT1 and FEAT2. Here we assume the spread of COVID-19 in the same subgroup are independent conditional on the "source" positive individual.

$$\begin{aligned} &\mathbb{P}(\text{An healthy person in an infected subgroup}) \\ &= 1 - \mathbb{P}(\text{Other } (\bar{m}\text{-1}) \text{ people are healthy} \mid \text{One healthy}) \\ &= 1 - \mathbb{P}^{\bar{m}-1}(\text{Another person is healthy} \mid \text{One healthy}) \\ &= 1 - (1 - \mathbb{P}(\text{Another person is sick} \mid \text{One healthy}))^{\bar{m}-1} \\ &= 1 - (1 - \frac{\mathbb{P}(\text{Another is healthy} \mid \text{One is sick})\mathbb{P}(\text{sick})}{1-\mathbb{P}(\text{sick})})^{\bar{m}-1}, \end{aligned}$$

  by Bayes theorem.

  Here,

$$\begin{aligned} &\mathbb{P}(\text{Another is healthy} \mid \text{One is sick}) \\ &= \mathbb{P}(\text{Not being infected} \mid \text{One is sick}) \\ &= 1 - \beta. \end{aligned}$$

  Since $\rho/(1-\rho) \sim \rho$,

$$\begin{aligned} &\mathbb{P}(\text{An healthy person in an infected subgroup}) \\ &= 1 - (1 - (1-\beta)\rho/(1-\rho))^{\bar{m}-1} \sim \rho(1-\beta)(\bar{m}-1). \end{aligned}$$

  Thus,

$$\gamma_+^* \sim 2\rho(1-\beta)(\bar{m}-1)\gamma_+. \qquad \square$$

8.3 Relationship between ρ and ρ^*

The relationship between the prevalence $\rho = \mathbb{P}(\text{A person is infected})$ and $\rho^* = \mathbb{P}(\text{A subgroup contains infections})$ is as follows:

$$\rho^* = \frac{\sum_i m \cdot p_m}{1 + (\sum_i m \cdot p_m - 1)\beta}\rho,$$

where the size of a subgroup is distributed as $\mathbb{P}(n_i = m) = p_m$, and $\beta = \mathbb{P}(\text{within-subgroup infection rate})$. The proof of (1) follows from:

$$\begin{aligned}
\rho &= \mathbb{P}(\text{Infected}) \\
&= \sum_m \mathbb{P}(\text{Infected}|\text{From infected subgroup } m) \times \mathbb{P}(\text{From infected subgroup m}) \\
&= \sum_m (\frac{1}{m} + \frac{m-1}{m}\beta)\frac{m \cdot p_m}{\sum_i i \cdot p_i}\rho^* \\
&= \frac{1 + (\sum_m m \cdot p_m - 1)\beta}{\sum_m m \cdot p_m}\rho^*. \qquad \square
\end{aligned}$$

9 CODE AVAILABILITY

Our code is available on https://github.com/Fang-Lin93/FAST_group_testing

References

[1] Tania Nolan, Rebecca E Hands, and Stephen A Bustin. Quantification of mrna using real-time rt-pcr. *Nature protocols*, 1(3):1559, 2006.

[2] Kamalini Ramdas, Ara Darzi, and Sanjay Jain. 'test, re-test, re-test': using inaccurate tests to greatly increase the accuracy of covid-19 testing. *Nature Medicine*, pages 1–2, 2020.

[3] An update on abbott's work on covid-19 testing, url https://www.abbott.com/corpnewsroom/product- and -innovation/an-update-on-abbotts-work-on-covid-19 -testing.html, 2020.

[4] https://www.aljazeera.com/news/2020/03/ bangladesh-scientists-create-3-kit-detect-covid-19 -200323035631025.html, 2020.

[5] Lisa Winter. False negatives in quick covid-19 test near 15 percent: Study. *The Scientist URL: https://www.the-scientist.com/news-opinion/false- negatives-in-quick-covid-19-test-near-15-percent-study -67451*, 2020.

[6] Rudolf Hanel and Stefan Thurner. Boosting test-efficiency by pooled testing strategies for sars-cov-2. *arXiv preprint arXiv:2003.09944*, 2020.

[7] Robert Dorfman. The detection of defective members of large populations. *The Annals of Mathematical Statistics*, 14(4):436–440, 1943.

[8] Jack Wolf. Born again group testing: Multiaccess communications. *IEEE Transactions on Information Theory*, 31(2):185–191, 1985.

[9] Matthew Aldridge, Leonardo Baldassini, and Oliver Johnson. Group testing algorithms: Bounds and simulations. *IEEE Transactions on Information Theory*, 60(6):3671–3687, 2014.

[10] Matthew Aldridge, Oliver Johnson, and Jonathan Scarlett. Group testing: an information theory perspective. *arXiv preprint arXiv:1902.06002*, 2019.

[11] A. Sterrett. On the detection of defective members of large populations. *The Annals of Mathematical Statistics*, page 1033–1036, 1957.

[12] C. H. Li. A sequential method for screening experimental variables. *Jour- nal of the American Statistical Association*, page 455–477, 1962.

[13] HM Finucan. The blood testing problem. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 13(1):43–50, 1964.

[14] Dingzhu Du, Frank K Hwang, and Frank Hwang. *Combinatorial group testing and its applications*, volume 12. World Scientific, 2000.

[15] Cassidy Mentus, Martin Romeo, and Christian DiPaola. Analysis and applications of adaptive group testing methods for covid-19. *medRxiv*, 2020.

[16] Annalisa De Bonis, Leszek Gasieniec, and Ugo Vaccaro. Optimal two-stage algorithms for group testing problems. *SIAM Journal on Computing*, 34(5):1253–1270, 2005.

[17] Liane N Patsula. A comparison of computerized adaptive testing and multi-stage testing. 1999.

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

[19] Hrayer Aprahamian, Douglas R Bish, and Ebru K Bish. Optimal risk-based group testing. *Management Science*, 65(9):4365–4384, 2019.

[20] Nathan Jeffay. To ease global virus test bottleneck israeli scientists suggest pooling samples pooling method for accelerated testing of covid-19, url https://www.timesofisrael.com/to-ease-global-virus-test- bottleneck-israeli-scientists-suggest-pooling-samples/, 2020.

[21] Pooling method for accelerated testing of covid-19, url https://www.technion.ac.il/en/2020/03/pooling-method- for-accelerated- testing-of-covid-19/, 2020.

[22] Ngoc T Nguyen, Hrayer Aprahamian, Ebru K Bish, and Douglas R Bish. A methodology for deriving the sensitivity of pooled testing, based on viral load progression and pooling dilution. *Journal of translational medicine*, 17(1):252, 2019.

[23] Jens Niklas Eberhardt, Nikolas Peter Breuckmann, and Christiane Sigrid Eberhardt. Multi-stage group testing improves efficiency of large-scale covid-19 screening. *Journal of Clinical Virology*, page 104382, 2020.

[24] Koen B Pouwels, Laurence SJ Roope, Adrian Barnett, David J Hunter, Terry M Nolan, and Philip M Clarke. Group testing for sars-cov-2: Forward to the past? *PharmacoEconomics Open*, 4(2):207, 2020.

[25] Public health guidance for community-related exposure, url https://www.cdc.gov/coronavirus/2019-ncov/php/public-health-recommendations.html, 2020.

[26] Keqin Li. Pooling strategy optimization for accelerating asymptomatic covid-19 screening. 2020.

[27] Erin Duffin. Distribution of households in the united states from 1970 to 2019, by household size, url https://www.statista.com/statistics/242189/distribution-of -households-in-the-us -by-household-size/, November 2019.

[28] S. M. Samuels. The exact solution to the two-stage group-testing problem. *Technometrics*, 20(4):497–500, 1978.

[29] Economic Social Affairs. Household size and composition around the world 2017. *United Nations' Data Booklet. URL: https://www.un.org/en/development/desa/population/ publications/pdf/ageing/household_size_and_ composition_around_the_world_2017_data_booklet.pdf. Title from the screen*, 2017.

[30] Coronavirus disease 2019, url cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how -covid-spreads.html, 2020.

[31] Covid-19 dashboard by the center for systems science and engineering (csse) at johns hopkins university (jhu), url https://coronavirus.jhu.edu/map.html, 2020.

[32] United Nations Department of Economic, Population Estimates Social Affairs, Population Division, and Projections Section. World population prospects, the 2019 revision. *World Population Prospects, the 2019 Revision.*, 2019.

[33] Stefan Lohse, Thorsten Pfuhl, Barbara Berkó-Göttel, Jürgen Rissland, Tobias Geißler, Barbara Gärtner, Sören L Becker, Sophie Schneitler, and Sigrun Smola. Pooling of samples for testing for sars-cov-2 in asymptomatic people. *The Lancet Infectious Diseases*, 2020.