Statistics and Applications {ISSN 2452-7395 (online)}
Volume 18, No. 1, 2020 (New Series), pp 281–294

Reducing the Number of Tests for COVID-19 Infection via Group Testing Methodologies

Jyotirmoy Sarkar
Department of Mathematical Sciences
Indiana University–Purdue University Indianapolis, Indiana, USA

Received: 25 June 2020; Revised: 09 July 2020; Accepted: 12 July 2020

Abstract

Total economic shutdown being detrimental to a nation's prosperity, most governments are reopening businesses and schools with the requirement of frequent and mass-scale testing to determine each person's status of COVID-19 infection. Obviously, the costs add up quickly and impose a heavy economic toll. As a way out of this dilemma, employers and administrators should consider seriously the application of group testing methodologies.

Group testing methods check samples in batches, rather than individually, for the presence of a disease. If the group tests positive, then the group is investigated further to identify who all are positive. On the other hand, if the group tests negative, not just once but also a second or a third time, then everyone within the group is cleared for activity. With a carefully chosen protocol, group testing costs can be 30-80% lower than those of individual testing, with the savings being higher when prevalence of the disease is lower.

Key words: Design of Experiments; Sensitivity; Specificity; False positive; False negative.

AMS Subject Classifications: 62K05

1. Introduction

Originating in Wuhan Province of China in November 2019, the novel corona virus has inflicted the SARS COVID-19 pandemic across the globe. In an immediate attempt to curtail the spread of the virus in the absence of a vaccine, many governments imposed lock-downs on their respective jurisdictions. However, the economies of many states, provinces and countries have been severely damaged because of lock-downs and stay-at-home orders. To prevent a total collapse of the economy, many governments are forced to reopen businesses and schools, notwithstanding the risk of spreading the disease. Therefore, it has become imperative to isolate people who have the disease and quarantine people who test positive for the virus. Consequently, frequent and mass-scale testing for COVID-19 infection has become a necessary precondition for restarting the economy. See, for example, The White House, *et al.* (2020) report.

The cost of testing has become burdensome on the payers (individuals, employers, administrators, insurance companies and governments). Also, testing capacity is often limited. In the midst of this dire situation, the celebrated group testing methodologies can offer a valuable relief without compromising safety.

Corresponding Author: Jyotirmoy Sarkar
Email: jsarkar@iupui.edu

Even though little in this paper is theoretically a new finding, except perhaps the extension to imperfect tests, the importance and necessity of the day prompt us to review this methodology anew. A recent paper by Aprahamian *et al.* (2017) mentions that it is precisely a lack of understanding of how an optimal pooling scheme should be designed to maximize classification accuracy under a budget constraint that hampers screening efforts. Perhaps a wide-spread familiarity with the techniques will not only reduce the cost of administering the tests, but also put suspicious minds at rest knowing that safety is not compromised in an attempt to reduce cost. In fact, Conger (2020) reports how pooling patient samples for COVID-19 testing helped Stanford researchers track the early spread of the virus in the Bay Area prior to the last week of February, 2020.

Section 2 gives a brief history of the group testing methodologies (GTMs). Section 3 determines the optimal group size for a perfect test and computes percentage savings for various choices of group sizes. Section 4 studies the more realistic case of imperfect tests, evaluating sensitivity and specificity of a group's sample as functions of those same quantities for an individual's sample and prevalence of disease, and determines the optimal size and cost savings. Section 5 gives some advancements in Group Testing Designs (GTDs) and GTMs that shed light on identifying a few defective items intermixed with many good items. Section 6 draws some practical implications of GTMs in the COVID-19 context.

All figures are drawn using the freeware R; and the codes are given in the Annexure.

2. A Brief History of Group Testing Methodology Used in Medicine

We borrow the history of group testing from Ding-Zhu and Hwang (1993). GTMs have had a humble start in Dorfman (1943). During the World War II, the United States Public Health Service and the Selective Service carried out a large-scale project to isolate all syphilitic men called up for induction. Testing an individual for syphilis involved drawing a blood sample and then analyzing the sample to determine the presence or absence of syphilis. At the time, performing this test was expensive, and testing every soldier individually would have been very expensive and inefficient.

Here is how the methodology works: Suppose that there are N soldiers. Testing each individual separately requires N tests, which is a reasonable approach if a large proportion of the people are infected. However, if a small proportion of men are infected, there is a much more efficient testing scheme: Split the soldiers into groups, and in each group combine the blood samples together. If one or more of the soldiers in this group has syphilis, then the test will be positive; and each member of the group has to be tested individually to find which soldier(s) are syphilitic. On the other hand, if the test is negative all members of the group are declared free of syphilis using only one test.

Sterrett (1957) improved Dorfman's procedure: Perform individual testing on a positive group sequentially; stop as soon as a defective is identified; then test the remaining items in the group, as a smaller group, since it is likely that none of them are defective. The first thorough treatment of group testing was given by Sobel and Groll (1959) who described five new procedures, including when prevalence is unknown. For the most optimal procedure, they provided an explicit formula for the expected number of tests it would use. Ungar (1960) proved that the range of prevalence of disease (or proportion of defective items), for which there is a group testing plan with expected number of perfect tests less than the

numer of items, is $[0, (3-\sqrt{5})/2) \approx [0, .382)$. Hwang (1972) detects all defective members in a population by group testing. Sobel and Elashoff (1975) used group testing for estimation.

The above methods work under the simplistic assumption that testing is error-free. However, in reality, most often a diagnostic test is "imperfect" in the sense that there is some chance that the result of a test is erroneous—exhibiting either false positive (that is, the test comes out positive when the sample contains no defectives) or false negative (that is, the test is negative even though the sample contains defectives).

As with most diagnostic tests, COVID-19 tests are imperfect. In particular, Yang *et al.* (2020) suggests that the sensitivity of polymerase chain reaction (PCR) tests on samples collected by nasal swab is around 70%, implying that about 30% of infected patients will return a false negative test result. Much of this error is caused by factors related to sample collection (for example, the patient does not have high enough viral-load levels at the time of collection or the swab did not reach the right place) rather than failures of the PCR test itself. But the danger of a false negative in this situation is unacceptably high: If someone believes he or she is free of disease contrary to truth, they risk spreading the disease to other employees, customers or students—precisely what the testing is trying to prevent!

GTMs have seen a lot of advancements during the last 40 years. Applications abound in both industrial product testing and in medical diagnostic tests. We completely skip the industrial applications (referring readers to Wikipedia, n. d.). In Section 5 of this paper, we mention some important advancements on GTDs and GTMs. Subsection 5.1 illustrates use of one such advancement. Here, we highlight a few references on medical diagnostics. Keeler E. and Berwick D. (1976) presents models of how test performance is degraded by pooling, and of the financial savings that pooling allows. They demonstrate the method of computing optimal pool size on a screening test and on a test for gastrin. Schisterman and Vexler (2008) examines the effect of different sampling strategies of biospecimens for exposure assessment that cannot be detected below a detection threshold. They compare use of pooled samples to a randomly selected sample from a cohort to evaluate the efficiency of parameter estimates.

To apply these GTMs, one must ensure that the following assumptions hold: (i) individual samples can be combined into groups, (ii) group tests have comparable accuracy to individual tests, and (iii) results of group tests can be correctly interpreted.

3. Optimal Group Size and Savings: Perfect Test

Consider first the simpler case of group testing with a perfect test. We explain the relationship among group sizes, number of tests and infection rates. Suppose that an employer splits up N workers into groups of equal size g and tests each group (with one or more group tests) in the first stage. In the second stage, all workers who belong to groups that tested positive (at least once) are subjected to individual testing, as in Dorfman (1943).

Extensions of the above simple protocol is not too hard: One can apply this framework to strategies with more than two stages. For example, lower-prevalence regions can afford to test larger pools and carry out multiple pooled testing rounds before beginning individualized testing. Moreover, employers can learn the actual prevalence over time based on the prevalence level revealed in earlier rounds of testing through statistical predictions. If

prevalence is higher or lower than anticipated, pool size can be adjusted in the next round.

Here, we also assume that workers are homogeneous in their risk of testing positive. But the method extends to heterogeneous population: The employer can simply stratify its workforce into subgroups by risk, and solve the test-minimization problem on each subpopulation separately using a smaller group size for a high risk stratum and a larger group size for a low risk stratum. We leave these extensions to the interested readers, and focus on a homogeneous population of employees tested in two stages—group and individual.

Suppose that the prevalence of active infection in the workforce is θ, and it is *known* to the employer. (Even if θ is unknown, the savings might decrease, but still there will be considerable savings as long as θ is estimated reasonably well.) Suppose also that N employees are split into N/g groups of size g each. Then the probability that a pool of g workers contains at least one infected member is $\theta_g = 1 - (1-\theta)^g$. The expected number of groups testing positive is given by $\theta_g\, N/g$. All g members of these COVID-19 positive pools have to be tested individually. Therefore, the expected number of workers receiving follow-up tests will be $\theta_g\, N$. Hence, the total number of tests will be

$$\frac{N}{g} + N\theta_g = N\left\{\frac{1}{g} + 1 - (1-\theta)^g\right\} \tag{1}$$

On the other hand, an individual testing protocol requires exactly N tests.

3.1. Minimize the number of tests

To compare the pooled testing protocol versus the individual testing, it suffices to minimize the number of test *per worker* less one, or to solve the minimization problem:

$$\min_{1\le g\le N}\left\{\frac{1}{g} - (1-\theta)^g\right\}$$

Using calculus, the argmin may be obtained by the solving

$$g^2(1-\theta)^g = [-\ln(1-\theta)]^{-1}$$

and then taking the largest integer no more than the solution, since g is a whole number (because people are indivisible). Alternatively, we can compare a contemplated $g \ge 3$ against the previous value $(g-1)$; and we prefer g over $(g-1)$ if and only if $g(g-1)\theta(1-\theta)^{g-1} < 1$. If this condition holds, we increase g by one and check this condition again; otherwise, we stop and declare the previous g as optimal. Hence, for a given prevalence θ, the optimal group size is given by

$$g^*(\theta) = \max\{g, 1 \le g \le N : i(i-1)\theta(1-\theta)^{i-1} < 1, \text{for all } i = 3, 4, \ldots, g\} \tag{2}$$

See the Annexure for the R codes to compute g^*. Indeed, the optimal g^* is a non-increasing step function of θ. For example, $g^*(0.05) = 5$, $g^*(0.02) = 8$, $g^*(0.01) = 11$, $g^*(0.005) = 16$, $g^*(0.002) = 24$, $g^*(0.001) = 32$. See more details in Figure 1. Of course, aligning with our intuition, we have $g^*(0+) = \lim_{\theta\to 0^+} g(\theta) = N$ and $g^*(1-) = \lim_{\theta\to 1^-} g(\theta) = 1$.

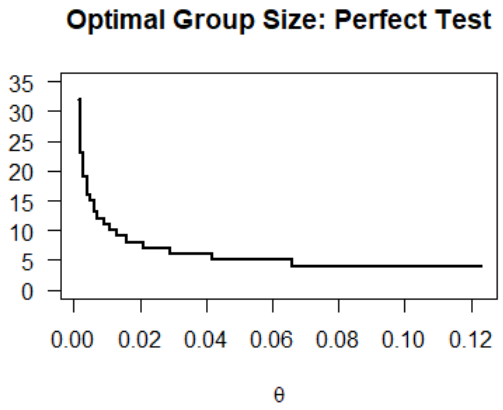

Figure 1: For a perfect test, the optimal group size $g^*(\theta)$

Alternatively, solving the dual problem, one can determine the range of values of θ for which a given group size g is optimal. For instance, $g = 4$ is optimal for $\theta \in (.0655, .1235)$, $g = 5$ is optimal for $\theta \in (.0415, 0.0655)$, etc. Again, more details are in Figure 1. Thus, oftentimes an imprecise knowledge of θ does not significantly alter the optimal choice of g.

3.2. Maximize the percentage savings

For a perfect test, instead of N individual tests, a group testing protocol with a fixed group size g requires fewer number of tests given by Eq. (1). Hence, a group testing protocol achieves a percentage saving over the individual testing protocol given by

$$\text{Savings (Perfect Test)} = \left((1-\theta)^g - \frac{1}{g}\right) \times 100\% \tag{3}$$

Clearly, savings are maximum when the optimal group size $g^*(\theta)$ is chosen; otherwise, savings are reduced. See Figure 2 (and the R codes in the Annexure).

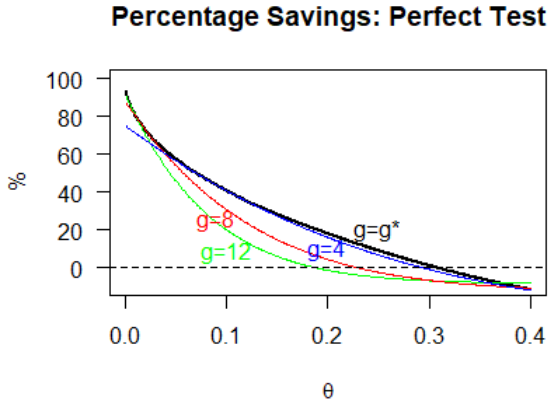

Figure 2: For a perfect test, the percentage savings for various group sizes: optimum and fixed

Caution: The group size must be chosen with careful consideration of the prevalence. For otherwise, savings may be *negative* when group size is chosen far from the optimum. For example, if prevalence is 12% then using the optimum group size $g^* = 4$, given by Eq. (2), the savings would be about 35%, using Eq. (3). However, if group size is (incorrectly) set at 32, then the savings would be -1.45%.

To illustrate the use of Figure 2, employers can reduce the number of tests by using groups of 4 workers, so long as prevalence is under 25 percent. Employers who know the prevalence with higher precision can further reduce the number of tests by choosing larger group sizes. But if employers underestimate the true prevalence rate and use larger group sizes than warranted, then the number of tests will increase.

4. Determining Group Size: Imperfect Test

The best COVID-19 test available to date is still imperfect. Let the sensitivity of the test be $\tau = \Pr\{+|D\}$, the probability that a person with the disease (D) will test positive. Then $(1 - \tau)$ is the probability of a false negative. Likewise, let the specificity of the test be $\eta = \Pr\{-|N\}$, the probability that a person with no disease (N) will test negative. Then $(1 - \eta)$ is the probability of a false positive. Assume that the sensitivity and the specificity remain the same whether we are testing a nasal swab of an individual or the combined swabs of the group (of any size). As mentioned earlier, PCR test for COVID-19 infection has sensitivity 0.70 and specificity 0.99.

4.1. Specificity of group testing

Suppose that a person has no COVID-19 infection. The group's test will be negative because of two disjoint and exhaustive cases: (1) all other $(g - 1)$ group members are also negative and the group tests negative (which is a true negative); (2) not all group members are negative, but the group's test comes out negative (which is a false negative). Writing $a = (1 - \theta)^{g-1}$, the probability that the group's test is negative, obtained by adding the probabilities of true negative and false negative, is

$$a\,\eta + (1 - a)(1 - \tau) = 1 - \tau + a(\eta + \tau - 1) \tag{4}$$

Similarly, the probability that the group's test will be positive, obtained by adding the probabilities of false positive and the true positive, is

$$a\,(1 - \eta) + (1 - a)\tau = \tau - a(\eta + \tau - 1) \tag{5}$$

which is one minus the probability that the group's test is negative, as it should be.

In this case of imperfect test, whenever the group tests negative, no individual tests are run: All individuals are declared negative. However, if the group tests positive, the individual without disease must get a negative result during the second-stage individual testing in order to be declared disease-free. Note that the specificity of the group testing is the same as the probability that the person without the disease will be declared negative after passing through both stages of the group testing protocol. As such, using Eqs. (4) and (5), the specificity of the group testing is given by

$$1 - \tau + a(\eta + \tau - 1) + [\tau - a(\eta + \tau - 1)]\eta = 1 - (1 - \eta)[(1 - a)\tau + a(1 - \tau)] > \eta \tag{6}$$

since the factor within square brackets on the right hand side of Eq. (6) is less than one. Thus, a group testing protocol increases specificity compared to individual testing. This reduces the probability of false positive results—a happy achievement. However, as we shall see in the next Subsection, it comes at the cost of increasing the probability of false negative results, which is highly risky in the COVID-19 application.

4.2. Sensitivity of group testing

A person with disease exhibits a positive result if and only if the group tests positive and so does the individual. Hence, The probability that a person with disease exhibits a positive result is τ^2, which is less than τ. This is a cause for concern, since the probability of false negative, $1-\tau^2$, is quite high.

A possible remedy is to run a follow-up group testing when the first group test is negative. Then the sensitivity of the group testing protocol, with two group-tests performed sequentially and if either group-test is positive then testing all group members individually, is the probability that a patient with disease shows a positive test result, and is given by

$$\tau^2 + (1-\tau)\tau^2 = \tau^2[2-\tau] > \tau^2 \tag{7}$$

For COVID-19 test, with $\tau = .7$, a single group test has sensitivity $\tau^2 = .49$; but a second follow-up group testing protocol has an overall group-test sensitivity $\tau^2(2-\tau) = .637$, which is tolerable. If a third follow-up group testing is used, then the sensitivity further rises to $\tau^2(3-3\tau+\tau^2) = .681$. Of course, while multiple follow-up group testing will increase sensitivity (though it will never exceed τ), it will also reduce savings. Therefore, let us settle on the protocol of *at most two group testings* before the group is found to be either positive (at least once) or declared to be negative. For this protocol, the sensitivity is given by the left hand side of Eq. (7). We leave it to the reader to study the group testing protocol that allows at most three group testings.

Suppose that a group has at least one COVID-19 patient, which happens with probability $\theta_g = 1-(1-\theta)^g$. Following the "two-group-testing protocol," the group's test will be positive with probability

$$1-(1-\tau)^2 = \tau(2-\tau) \tag{8}$$

Similarly, if a group has no COVID-19 patient, which happens with probability $(1-\theta)^g$, the probability that this group's test will be positive (falsely) is

$$1-\eta^2 \tag{9}$$

4.3. Minimize the number of tests

For the two-group-testing protocol, using Eqs. (8) and (9), the number of tests *per worker* is $2/g + [\theta_g\tau(2-\tau) + (1-\theta)^g(1-\eta^2)]$, which simplifies to

$$\frac{2}{g} + \tau(2-\tau) - (1-\theta)^g[\tau(2-\tau) + \eta^2 - 1] \tag{10}$$

Consequently, to determine the optimum group size we may drop the constant $\tau(2-\tau)$ from Eq. (10) and solve the minimization problem:

$$\min_{1\le g\le N}\left\{\frac{2}{g} - (1-\theta)^g[\eta^2 + \tau(2-\tau) - 1]\right\}$$

Analogously to the method of solution in case of a perfect test, here in the imperfect test case with two-group-testing protocol, given θ, the optimal group size $g^{\#}(\theta)$ is the largest g such that

$$i(i-1)\theta(1-\theta)^{i-1} < \frac{2}{\eta^2 + \tau(2-\tau) - 1}, \text{ for all } i = 3, 4, \ldots, g \tag{11}$$

The R codes to compute $g^{\#}$ are given in the Annexure. Note that the optimal $g^{\#}$ is a non-increasing step function of θ. For example, $g^{\#}(0.05) = 10$, $g^{\#}(0.02) = 16$, $g^{\#}(0.01) = 23$, $g^{\#}(0.005) = 33$. See more details in Figure 3, where the thick (black) curve shows $g^{\#}$ and the thinner (green) curve shows g^{*} already depicted in Figure 1 corresponding to the perfect test and repeated here for easy comparison. The optimal group size in the imperfect test case, with the two-group-testing protocol, is about 40–80% larger than the optimal group size in the perfect test case.

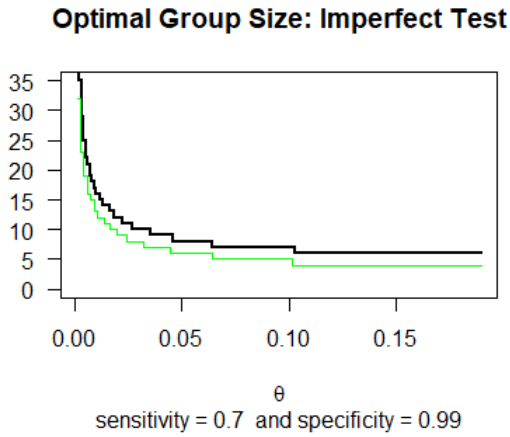

Figure 3: For an imperfect test, the optimal group size $g^{\#}(\theta)$ is shown by a thick, black curve. The thin, green curve shows g^{*} for the perfect test as in Figure 1

4.4. Maximize the percentage savings

For an imperfect test, with the two-group-testing protocol and fixed group size g, the percentage savings is given by

$$\text{Savings (ImperfectTest)} = \left([\eta^2 + \tau(2-\tau) - 1](1-\theta)^g - \frac{2}{g} \right) \times 100\% \tag{12}$$

Once again, the percentage savings is maximum when the optimal group size $g^{\#}(\theta)$ is chosen; otherwise, savings is reduced as the group size deviates from the optimum size. See details in Figure 4 (and the R codes in the Annexure).

As it happened in the perfect case, so also in the imperfect case, employers can reduce the number of tests by using groups of 8 workers, so long as prevalence is under 20 percent. Employers who know the prevalence more precisely can further reduce the number of tests by choosing larger group sizes. But if employers underestimate the true prevalence rate and choose larger group sizes, then the number of tests will increase.

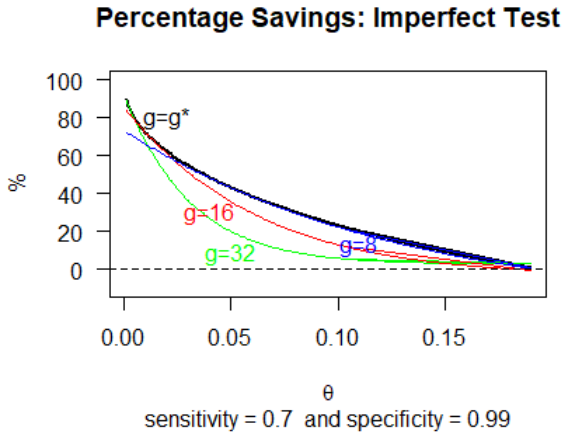

Figure 4: For an imperfect test, the percentage savings for various group sizes: optimum and fixed

5. Further Improvements in Group Testing Methodologies

As mentioned in Section 2, GTMs have undergone tremendous advancement during the last 40 years. We mention some of these results that focus on identifying a few defective items from among many good items, since that is the situation we find ourselves in when a group tests positive in the first-stage. We hope this short review of the GTD research landscape will inspire other researchers pursue this fascinating field of study.

Bush *et al.* (1984) introduced a new class of combinatorial designs with completeness property on t symbols, and used them in group testing to separate defective items from good ones using fewer number of tests than items. In their language: "if a large population of v items has exactly 1 bad item, it can be detected in b tests, where b is only a very tiny fraction of v." Weideman and Raghavarao (1987 a) carried out a systematic study of non-adaptive hypergeometric GTDs for identifying two defectives items from among n, obtaining bounds for n, given the number of tests. Weideman and Raghavarao (1987 b) extended the work to identifying *at most* two defective items.

Das and Roy Choudhury (1987) provides simple methods of forming a small number of groups out of a large number of individuals so that the group test results uniquely (and easily) determine all defective individuals. The methods consist of first encoding suitably all individuals and then forming the groups by using certain properties of the encoders. Hwang and Sos (1987) mentions that even though adaptive or sequential designs (which keep constructing new groups based on the results of previous groups) typically outperform non-adaptive combinatorial designs (which declare the groups at the outset), with the advent of parallel processing, the time-saving feature of non-adaptive designs remains attractive.

Whereas Ding-Zhu and Hwang (1993) develops the conventional disjunct search model, D'yachkov *et al.* (2001) discusses two non-standard models of nonadaptive combinatorial search for a small number of defective elements contained in a finite population in the presence of inhibitors. Hung and Swallow (1999) discusses the robustness of group testing for estimating proportions when the underlying assumptions of no testing errors and independent individuals are violated. Adhikari, Ghosh and Sinha (2001) considered a multi-component

system which can be tested even though the components cannot be tested separately. For a five-component system, they utilized Taguchi's L16 orthogonal array design to identify defective components and to estimate the proportion of defectives. They also designed a sequential experiment which reduced the number of tests from 16 to 12.

Du and Hwang (2006) surveys both sequential pooling designs and nonadaptive group testing. Rao, Rao and Sinha (2006) obtained improved bounds on the number of group tests necessary for both adaptive and nonadaptive GTDs. They established that when a group of 2^n items have *at most two* defective items, any nonadaptive GTD needs at least $2n$ group tests for identifying all defective items, and the optimal nonadaptive GTDs can be constructed using generalized Petersen graphs. In the same context, they presented an adaptive multistage GTD with a maximum of $2n$ group tests, and constructed a family of two-stage adaptive GTDs. Rao, Sinha and Rao (2013) gives a comprehensive review of both adaptive and non-adaptive GTDs, exposing hidden combinatorial and probabilistic challenges and offering a storehouse of unsolved problems.

We leave to the reader to apply these advanced GTDs to the COVID-19 detection problem. We describe below a situation where a group that tested positive in the first-stage is subjected to another group testing before conducting individual tests.

5.1. A second-stage group testing

According to our proposed two-group-testing protocol, if two successive group tests are both negative, we declare the entire group negative. But when the first or the second group-test turns out positive, then we test every member of the group individually. As mentioned in Section 3, if the group size is big (say, bigger than 10), it is more efficient to apply GTMs on this group, before moving on to individual testing, and harvest additional savings.

Suppose that a university estimated the prevalence of COVID-19 among its 25,000 students, faculty and staff members to be very low, say 0.1% or 0.001, and accordingly chose $g = 49$ (using Eq. (11)). The number of infected individuals in each group of 49 has an approximate Poisson(0.049) distribution. Hence, about 25 groups are likely to test positive; and within each group that tests positive, with a very high probability (about .9757) exactly one member is positive, with a low probability (about .0239) two members are positive, and with a negligible probability (about .0004) more than two members are positive.

Therefore, instead of testing individually every member of a positively tested group, as our proposed first-stage two-group-test protocol recommends, the employer can update the prevalence (for each positively tested group) to be $1/25 = .04$, or more conservatively to be $2/25 = .08$, and adopt a second-stage group testing with $g = 7$ (again using Eq. (11) or Figure 3). Then with a high probability (.9757) it will take 13 group tests to identify one positive subgroup, or with a low probability (.0239) it will take 12 group tests to identify two positive subgroups, or with a negligible probability (.0004) it will take at most 11 group tests to identify more than two positive subgroups.

It may not be worth applying the same logic to the one (or two or three) subgroups of size 7 that tested positive during the second-stage group testing because an individual's nasal swab specimen can be divided into about four or five sub-specimens while still retaining tractability of COVID-19 infection. It will be best to simply apply an individual test on every

member of the latest positively tested subgroup using the fifth sub-specimen, if any, or collect fresh new specimens from them. Thus, overall the 49 individual tests are reduced *on average* to about $.9757 * (13 + 7) + .0239 * (12 + 2 * 7) + .0004 * (11 + 3 * 7) = 20.15$ tests, achieving a 59% reduction on costs after first-stage testing.

To summarize, the first stage two-group-test protocol requires about $2 * \lceil 25000/49 \rceil - 25 = 997$ tests. Thereafter, (i) a second stage individual testing would involve $25 * 49 = 1225$ tests, for a total of 2222 tests. But (ii) a second stage two-group-test followed by a third stage individual testing would involve about $25 * 20.15 = 504$ tests. Thus, strategy (i) saves 91% compared to testing all 25000 people individually, and strategy (ii) additionally saves 721 tests, or 2.9%. Finally, strategy (ii), when compared to strategy (i), harvests a $721/2222 = 32.4\%$ saving across all tests in all stages combined.

6. Implications of Group Testing for COVID-19 Detection

Economic shutdown causes a tremendous loss in GDP (gross domestic product). On the other hand, reopening the economy necessitates frequent and mass-scale testing of all employees. Although the latter cost is relatively smaller than the former cost, if employers have to bear all the cost, then a cost effective way to carry out these tests is imperative. Group testing, with a follow-up test for the group if the first group test is negative (or when two group tests are negative), is a promising methodology to mitigate the challenge.

We have shown significant savings when prevalence is known with a high degree of certainty. We have also demonstrated that the quality of diagnostics made using the group testing protocol with two-group-tests is reasonably close to that of the individual testing protocol. The quality can be increased slightly (at the cost of reducing savings) by using a three-group-tests protocol before declaring the group negative. Also, there can be a second-stage group testing followed by individual testing to reap additional savings. Until testing sensitivity improves, employers should implement group testing together with other strategies, such as symptom monitoring and contact tracing.

Uncertainty about prevalence reduces savings only marginally if prevalence can be estimated well. As more information becomes available to estimate prevalence better, higher savings are anticipated. Also, as the sensitivity of the test improves—when researchers discover better tests—the optimal group sizes will rise yielding additional savings. Moreover, such better tests will reduce complexity of testing and discomfort to the employees. For instance, saliva-based tests may become a suitable alternative to nasal swabs. Let us look forward to those better days until a vaccine or a cure becomes available and affordable.

Acknowledgements

I thank Dr. Honglang Wang for suggesting some improvements on a first draft. From an anonymous reviewer, I received a treasure trove of references, which give a bird's eye view of the entire landscape of advancement in GTMs. I express my heart-felt thanks to this generous reviewer. Thanks are due to Professor Bikas K. Sinha for providing a hard-to-find reference on short notice. I am grateful to the Chair Editor for his patience, guidance, encouragement and instructions at multiple stages.

References

Adhikari, S., Ghosh, J. K. and Sinha, Bikas K. (2001). Group testing for identification of defective components: An application of orthogonal arrays. *Statistics and Applications*, **3(1-2)**, 89–96.

Aprahamian, H., Bish, E. K. and Bish, D. R. (2017). Adaptive risk-based pooling in public health screening. *Institute of Industrial and Systems Engineers (IISE) Transactions*, **50(9)**, 753–766. https://doi.org/10.1080/24725854.2018.1434333

Bush, K. A., Federer, W. T., Pesotan, H. and Raghavarao, D. (1984). New combinatorial designs and their applications to group testing. *Journal of Statistical Planning and Inference*, **10**, 335–343.

Conger, K. (2020). Testing pooled samples for COVID-19 helps Stanford researchers track early viral spread in Bay Area. *Stanford Medicine News Center*. https://med.stanford.edu/news/all-news/2020/04/testing-pooled-samples-to-track-early-spread-of-virus

Das, M. N. and Roy Choudhury, D. (1987). On problems of search using group testing. *Sankhya B*, **49**, 137–147.

Ding-Zhu, D. and Hwang, F. K. (1993). *Combinatorial Group Testing and Its Applications*. Singapore: World Scientific.

Dorfman, R. (1943). The detection of defective members of large populations. *The Annals of Mathematical Statistics*, **14(4)**, 436–440.

Du, D. and Hwang, F. (2006). *Pooling Designs and Nonadaptive Group Testing*. Boston: Twayne Publishers.

D'yachkov, A. G., Macula, A. J., Torney, D. C. and Vilenkin, P. A. (2001). Two models of nonadaptive group testing for designing screening experiments. In: Atkinson A. C., Hackl P., Müller W. G. (eds) *mODa 6 – Advances in Model-Oriented Design and Analysis*, pp 63–75. Contributions to Statistics. Physica, Heidelberg. https://doi.org/10.1007/978-3-642-57576-1_8

Hung, M. and Swallow, W. H. (1999). Robustness of group testing in the estimation of proportions. *Biometrics*, **55(1)**, 231–237. doi:10.1111/j.0006-341X.1999.00231.x

Hwang, F. K. (1972). A method for detecting all defective members in a population by group testing. *Journal of the American Statistical Association*, **67(339)**, 605–608. doi:10.2307/2284447.

Hwang, F. K. and Sos, V. T. (1987). Non-adaptive hypergeometric group testing. *Studia Scientiarum Mathematicarum Hungarica*, **22**, 257–263.

Keeler E. and Berwick D. (1976). Effects of pooled samples. *Health Laboratory Science*, **13(2)**, 121–128. https://www.ncbi.nlm.nih.gov/pubmed/1270261

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rao, P. S. S. N. V. P., Rao, S. B. and Sinha, Bikas K. (2006). Improved bounds for group testing designs. *Journal of Statistical Planning and Inference*, **136(1)**, 260–269.

Rao, S. B., Sinha, Bikas K. and Rao, Prasad. (2013). Group testing designs: A combinatorial marvel. *Statistics and Applications*, **11(1-2)**, 111–126.

Schisterman, E. F., and Vexler, A. (2008). To pool or not to pool, from whether to when: applications of pooling to biospecimens subject to a limit of detection. *Paediatric and Perinatal Epidemiology*. https://doi.org/10.1111/j.1365-3016.2008.00956.x

Sobel, M. and Groll, P. A. (1959). Group testing to eliminate efficiently all defectives in a binomial sample. *Bell System Technical Journal*, **38(5)**, 1179–1252.

doi:10.1002/j.1538-7305.1959.tb03914.x

Sobel, M. and Elashoff, R. M. (1975). Group testing with a new goal, estimation. *Biometrika*, **62(1)**, 181–193. doi:10.1093/biomet/62.1.181

Sterrett, A. (1957). On the detection of defective members of large populations. *The Annals of Mathematical Statistics*, **28 (4)**, 1033–1036. doi:10.1214/aoms/1177706807

The White House, Centers for Disease Control and Prevention, and Food and Drug Administration. (2020). *Testing Blueprint: Opening Up America Again.* https://www.whitehouse.gov/wp-content/uploads/2020/04/Testing-Blueprint.pdf

Ungar, P. (1960). The cutoff point for group testing. *Communications on Pure and Applied Mathematics*, **13**, 49–54.

Weideman, C. A. and Raghavarao, D. (1987 a). Some optimum nonadaptive hypergeometric group testing designs for identifying two defectives. *Journal of statistical Planning and Inference*, **16**, 55–61.

Weideman, C. A. and Raghavarao D. (1987 b). Nonadaptive hypergeometric group testing designs for identifying at most two defectives. *Communications in Statistics*, **16 (10)**, 2991–3006.

Wikipedia Contributors (no date). *Group Testing. Wikipedia, The Free Encyclopedia.* Retrieved 13 June 2020. https://en.wikipedia.org/wiki/Group_testing

Yang Y., Yang, M., Shen, C., Wang, F., Yuan, J., Li, J., Zhang, M., Wang, Z., Xing, L., Wei, J., Peng, L., Wong, G., Zheng, H., Liao, M., Feng, K., Li, J., Yang, Q., Zhao, J., Zhang, Z., Liu, L. and Liu, Y. (2020). Evaluating the accuracy of different respiratory specimens in the laboratory diagnosis and monitoring the viral shedding of 2019-nCoV infections. *MedRxiv*. https://doi.org/10.1101/20 20.02.11.20021493

ANNEXURE

Here we document the R codes used to draw Figures 1–4.

```
### Fig 1. Optimal group size (Perfect test)
gs=function(the){b=3
  for (i in 4:40){
    if (i*(i-1)*(1-the)^(i-1)*the<1){b=i}
    else{break} }
b}

theta=seq(0.001, .123, .001)
gsp=rep(3,123)
for (j in 1:123){gsp[j]=gs(j/1000)}
plot(theta, gsp, type='s', las=1, lwd=2, ylim=c(0,35),
  ylab='', xlab=expression(theta),
  main='Optimal Group Size: Perfect Test')

### Fig 2. Savings (Perfect test)
thet=c(theta,seq(.124,.400,.001))
save=100*c((1-theta)^gsp-1/gsp, (1-seq(.124,.400,.001))^3-1/3)
plot(thet, save, type='l', las=1, lwd=2, ylim=c(-10,100),
```

```
   ylab='%', xlab=expression(theta),
   main='Percentage Savings: Perfect Test')

lines(thet, 100*((1-thet)^12-1/12), col='green')
lines(thet, 100*((1-thet)^8-1/8), col='red')
lines(thet, 100*((1-thet)^4-1/4), col='blue')
abline(c(0, .400),c(0,0), lty=2)
text(0.1, 8,'g=12', col='green'); text(0.09, 25,'g=8', col='red')
text(0.20, 10,'g=4', col='blue'); text(0.25, 20,'g=g*', col='black')

### Fig 3. Optimal group size (Imperfect test)
tau=.70; eta=.99;
(mult=eta^2+tau*(2-tau)-1)
gs=function(the){b=3
  for (i in 6:80){
    if (i*(i-1)*(1-the)^(i-1)*the<2/mult){b=i}
    else{break} }
b}

theta=seq(0.001, .190, .001)
gst=rep(6,190)
for (j in 1:190){gst[j]=gs(j/1000)}
plot(theta, gst, type='s', las=1, lwd=2, ylim=c(0,35),
  ylab='', xlab=expression(theta),
  main='Optimal Group Size: Imperfect Test',
  sub=paste('sensitivity =', tau, ' and specificity =', eta) )
par(new=TRUE)
plot(theta[1:123], gsp, type='s', las=1, ylim=c(0,35),
   xlab='', ylab='', xaxt='n', yaxt='n', lty=1, col='green')


### Fig 4. Savings (Imperfect test)
thet=c(theta,seq(.104,.190,.001))
save=100*c(1-tau*(2-tau)+(1-theta)^gst*mult-2/gst,
   1-tau*(2-tau)+(1-seq(.104,.190,.001))^6*mult-1/3)
plot(thet, save, type='l', las=1, lwd=2, ylim=c(-10,100),
   ylab='%', xlab=expression(theta),
   main='Percentage Savings: Imperfect Test',
   sub=paste('sensitivity =', tau, ' and specificity =', eta) )

lines(thet, 100*(1-tau*(2-tau)+(1-thet)^32*mult-1/16), col='green')
lines(thet, 100*(1-tau*(2-tau)+(1-thet)^16*mult-1/8), col='red')
lines(thet, 100*(1-tau*(2-tau)+(1-thet)^8*mult-1/4), col='blue')
abline(c(0, .200),c(0,0), lty=2)
text(0.05, 8,'g=32', col='green'); text(0.04, 30,'g=16', col='red')
text(0.11, 13,'g=8', col='blue');  text(0.02, 80,'g=g*', col='black')
```