Pooled testing to isolate infected individuals

Matthew Aldridge

School of Mathematics

University of Leeds

Leeds, U.K., LS2 9JT

Email: m.aldridge@leeds.ac.uk

Abstract—The usual problem for group testing is this: For a given number of individuals and a given prevalence, how many tests T^* are required to find every infected individual? In real life, however, the problem is usually different: For a given number of individuals, a given prevalence, and a limited number of tests T much smaller than T^*, how can these tests best be used?

In this conference paper, we outline some recent results on this problem for two models. First, the 'practical' model, which is relevant for screening for COVID-19 and has tests that are highly specific but imperfectly sensitive, shows that simple algorithms can be outperformed at low prevalence and high sensitivity. Second, the 'theoretical' model of very low prevalence with perfect tests gives interesting new mathematical results.

I. INTRODUCTION

A. The problem

When testing individuals for a disease such as COVID-19, we can take a sample from each individual and test them separately; for n individuals, this requires n tests. Alternatively, we can pool samples together and test the pooled sample; in an ideal model, a test is positive if at least one individual in the pool is infected, and is negative otherwise. When the prevalence p is low, the pooling method, known as 'group testing' or 'pooled testing', can identify all the infected individuals using fewer than n tests, thereby making better use of scarce tests. (For background on group testing, see the recent survey paper [1].)

Typically, the aim of group testing is to find every infected individual without erroneously declaring any uninfected individual to be infected. Thus the mathematical problem is typically this: Given the number of individuals n and the prevalence p, how many tests T^* are required to find the all the infected individuals, and what testing protocol achieves this. (There has also been some attention on the problem of finding a set that has a large overlap with the set of infected individuals [1], [2] but that may not classify every single individual correctly.) However, considering applications in modern settings, especially in the current coronavirus pandemic, we propose a different goal.

In a large workplace, healthcare facility, or university, for example, there may be a very limited number of tests T available each day – certainly $T < T^*$, so there are too few tests to accurately find every infected worker. Thus the aim should be to find as many infected individuals as possible with those limited tests, so that those individuals can be removed from the workforce to isolate at home (for example).

Under this criterion, one could take samples from just a small proportion of the workforce but try to find all or almost all of the infected individuals in that subset; alternatively, one could take samples from a much larger proportion of the workforce, but be satisfied with finding a smaller fraction of the infected individuals within the subset.

In order to compare different strategies, we propose using the *expected number of tests per infected individual found* (ETI) as the relevant figure of merit. That is, if a scheme uses an expected number $\mathbb{E}T$ of tests and finds an expected number $\mathbb{E}K$ of infected individuals, then the ETI is

$$\text{ETI} = \mathbb{E}T/\mathbb{E}K.$$

We would like this to be as large as possible.

B. The models

In this conference paper, we outline results for two different models, which we call the *theoretical* and *practical* models.

Both models have the following features in common:

- There are a large number n of individuals. This number is sufficiently large that looking at mathematical results in the limit as $n \to \infty$ is useful.
- The instantaneous prevalence rate $p \in (0, 1)$ is known. Throughout we write $q = 1 - p$. We use the *i.i.d. prior* where each individual is infected independently with probability p. It will be useful to write $k = pn$ for the expected number of infected individuals.
- Tests are very highly specific, in that a pool containing no infected samples is very highly likely to correctly give a negative result. We model this as the specificity being 1; thus under our model we can be certain that a positive test contains at least one infected sample.

In our first model, the *practical model*, we attempt to give a realistic model for coronavirus testing.

- The prevalence p is a fixed constant as $n \to \infty$. This is because the prevalence of COVID-19 is unlikely to change depending on the size of organisation being tested. For screening of COVID-19 among asymptomatic people, values of the prevalence p between 0.005 and 0.1 are likely to be of interest.
- Tests are only moderately sensitive, in that pools containing one or more infected samples may give an erroneous negative result. Our model for this is to say that each test containing at least one infected sample correctly gives a

positive result with probability $u \in (0, 1]$, independently between tests. Values of u in the range $u = 0.6$ to 0.9 are likely to be of interest for PCR tests.

- As many stages of testing will be impractical and slow, we limit adaptive strategies to two stages of testing.
- There are high costs for a false positive declaration (that is, wrongly declaring a noninfected individual to be infected) – for example, a healthcare worker may have to self-isolate for at least seven days for no reason. For this reason, false positive declarations will not be permitted, and we may only count individuals who we are certain are infected (under the above model assumptions).
- So that individuals can be sure they have the virus before isolating, we require suspicion that an individual is infected to be definitively confirmed with an individual non-pooled test. Thus we have a first stage of pooled testing, then a second stage of limited individual testing. This is known as 'trivial two-stage testing'.

(For more detailed information on the practicality of pooled testing for COVID-19, with detailed consideration of real-life issues and accurate modelling, see the forthcoming book chapter [3].)

We give here some further justification for our consideration on trivial two-stage testing. When the sensitivity u is less than 1, it is impossible to definitively rule out individuals as definitely noninfected, since any negative tests might have been false negatives. Thus the only way to definitively confirm an individual is infected is by them receiving an individual test and that test being positive. Thus the second stage in any of our algorithms must be individual tests, as pooled tests in the final stage will be worthless under the modelling assumptions and success criteria we have set out.

In our second model, the *theoretical model*, we follow the most common set-up for theoretical results on group testing – see, for example [1], [4]–[7].

- As $n \to \infty$, the prevalence p scales like $p \sim n^{-(1-\alpha)}$ for $\alpha \in [0, 1)$. Thus the true number of infected individuals is strongly concentrated around $k = pn \sim n^\alpha$.
- Schemes must be fully nonadaptive, meaning all the tests are decided on in advance and are conducted in a single stage.
- The criterion for success is that *every* individual declared to be infected is indeed infected. However, we do allow failure to meet this criterion, provided that the probability of such failure tends to 0 as $n \to \infty$.
- The testing procedure is perfect. Any pool containing no infected individuals always gives a negative result, and any pool containing one or more infected individuals always gives a positive result.

II. RESULTS FOR THE PRACTICAL MODEL

We consider trivial two-stage algorithms with two parameters, r and s, as studied in [8], [9] for perfect noiseless tests. In the first stage, any individual that is sampled is sampled in r pools, and each pool samples s individuals. Typically r is

very small (1 and 2 are the most common values, but 3 and 4 are sometimes used).

For individual non-pooled testing, we take $r = 1$, $s = 1$, and don't require a second stage. For all other sensible parameters we have $s > 1$, and in the second stage we retest any individuals that were positive in all r pooled tests. (Although we don't consider it here, it may be worthwhile to retest individuals whose r pooled tests were mostly – but not entirely – positive. We intend to study this in future work.)

Setting $r = 1$, $s > 1$, gives the simplest pooled algorithm, named *Dorfman's algorithm*, after Robert Dorfman's original group testing procedure [10]. Here, the sampled individuals are split into pools of size s. If a pooled test is negative, it is assumed that all those individuals are noninfected. (This is not certain to be correct, but is strong evidence that retesting them is a poor use of resources compared with testing a new untested pool.) If the pooled test is positive, then those individuals are individually tested in the second stage.

For general r, s, the most mathematically convenient method is to fix the number of individuals $m \leq n$ to be sampled, then to choose a testing strategy uniformly at random, subject to each test having weight s and each individual having weight r. This requires m to be divisible by r and s, but since these are typically small, this is not much of a restriction.

In practice, it can sometimes be more convenient to use a *hypercube design*. We explain this first by considering the case $r = 2$, where we also use the term *grid design*. Here, we take m individuals where m is a multiple of s^2. We imagine the individuals placed on square grids of size $s \times s$. Then each grid corresponds to $2s$ tests: s tests each pooling the individuals in one row, and s tests each pooling the individuals in one column. The grid design can be better than the random design for small n (although the asymptotic performance is the same). However, s^2 can be quite large, so the divisibility issue can be awkward. The grid design was studied recently by Broder and Kumar [9].

For $r > 2$ the hypercube design requires $s = a^{r-1}$ for some integer $a > 1$. We place each a^r individuals in an $a \times a \times \cdots \times a$ r-dimensional hypercube. Each test corresponds to one of the rs $(r - 1)$-dimensional 'slices' of the hypercube. The case $r = 3$, $s = 9$, with a 27 individuals in a $3 \times 3 \times 3$ hypercube was prominently studied in [11] – this has nine 2-dimensional 3×3 slices: three front to back, three left to right, and three top to bottom. With a hypercube design, the divisibility restrictions are much stronger, but the extra structure can be more convenient for a laboratory to carry out.

Theorem 1: In the practical regime, the above algorithm has ETI

$$\text{ETI} = \frac{1}{pu}$$

for individual testing $r = s = 1$, and ETI

$$\text{ETI} = \frac{\frac{r}{s} + u^r \left(p + q(1 - q^{s-1})^r\right)}{pu^{r+1}}$$

for $s > 1$.

TABLE I
ETI FOR INDIVIDUAL NON-POOLED TESTING WITH EXAMPLE PARAMETERS OF THE SENSITIVITY u AND PREVALENCE p IN THE PRACTICAL MODEL

p	$u = 0.6$	$u = 0.7$	$u = 0.8$	$u = 0.9$
0.1	16.7	14.3	12.5	11.1
0.05	33.3	28.6	25.0	22.2
0.02	83.3	71.4	62.5	55.6
0.01	167	143	125	111
0.005	333	286	250	222

TABLE II
ETI FOR POOLED TESTING WITH EXAMPLE PARAMETERS OF THE SENSITIVITY u AND PREVALENCE p IN THE PRACTICAL MODEL. VALUES OF THE FIRST-STAGE TESTS-PER-INDIVIDUAL r AND INDIVIDUALS-PER-TEST s ARE GIVEN IN BRACKETS AS (r, s).

p	$u = 0.6$		$u = 0.7$		$u = 0.8$		$u = 0.9$	
0.1	12.4	$(1, 5)$	9.93	$(1, 5)$	8.21	$(1, 4)$	6.91	$(1, 4)$
0.05	18.0	$(1, 7)$	14.4	$(1, 6)$	11.8	$(1, 6)$	9.48	$(2, 10)$
0.02	29.1	$(1, 10)$	23.0	$(2, 21)$	17.2	$(2, 19)$	13.3	$(3, 27)$
0.01	41.7	$(1, 14)$	29.8	$(2, 32)$	22.2	$(2, 29)$	16.2	$(3, 42)$
0.005	54.0	$(2, 55)$	38.2	$(2, 49)$	28.2	$(3, 76)$	19.7	$(3, 68)$

Proof: We give a brief justification of this result. The result for an individual test is clear: the test is positive if the individual is infected (with probability p) and the test correctly gives a positive result (with probability u) for an ETI of $1/(pu)$.

Now consider a trivial two-stage algorithm with parameters r and s. The number of first-stage tests-per-individual is r/s.

An individual could potentially be retested in the second stage if it is either infected, with probability p, or it is not infected but all r of its tests has one of the other $s-1$ individuals infected, with probability asymptotically $q(1 - q^{s-1})^r$. (This would be exactly true if the test results were independent, but if an individual shares more than one pool with our given individual, that would not be the case. However, we prove in our full paper that the equation is asymptotically accurate in spite on the dependences.) If one of these criteria is fulfilled, the individual will be tested again if all r tests correctly give a positive result, with probability u^r. Hence the expected number of second-stage tests per individual is $u^r (p + q(1 - q^{s-1})^r)$.

An infected individual is found if it is indeed infected, with probability p, all r pooled tests are correctly positive, with probability u^r, and the individual test is correctly positive, with probability u. All together, this is pu^{r+1}.

Finally, the ETI is the ratio of these two terms. ∎

Tables I and II shows the expected tests per infected individual found (ETI) for various plausible values of the sensitivity u and the prevalence p. The numbers in brackets are optimal values of (r, s), determined numerically. Note that

$(1, s)$ denotes Dorfman's simple pooling algorithm.

Note that the pooled schemes are better than individual testing for all values of the parameters. The gain is biggest when the prevalence is low and the sensitivity is high. Dorfman's algorithm ($r = 1$) is best for moderately high prevalence or moderately low sensitivity.

For a similar analysis of Dorfman's algorithm that also allows for imperfect specificity, see [3].

III. RESULTS FOR THE THEORETICAL MODEL

Recall that in the theoretical model all tests are perfectly accurate, and the number of infected individuals is very close to $k = pn$, which scales like $k \sim n^{\alpha}$ with $\alpha \in [0, 1)$.

Our result (stated slightly informally in this conference paper, with formalities to follow in a later paper full paper later) is the following.

Theorem 2: In the theoretical model, we can achieve an ETI of

$$\text{ETI} = \min\{\text{ETI}_{\text{full}}, \text{ETI}_{\text{saff}}\},$$

where

$$\text{ETI}_{\text{full}} = \max\left\{\log_2 \frac{n}{k}, \frac{1}{\ln 2} \ln k\right\},$$

$$\text{ETI}_{\text{saff}} = 2\text{e} \log_2 \frac{n}{k}.$$

Proof: We give a brief justification for this result, again with a full formal proof to appear in a later paper. We use

similar techniques to those used in [2] for the 'find one defective item' and 'approximate recovery' problems.

First, ETI$_{\text{full}}$ is the ETI achieved when using

$$T^* = \max\left\{k\log_2\frac{n}{k}, \frac{1}{\ln 2}k\ln k\right\}$$

tests to find all k infected individuals [6], [7]. Suppose we instead have $T = cT^*$ tests, where $c < 1$. We 'cut our losses' by immediately discarding all but cn individuals. This subset will have very close to ck infected individuals, and we can find all of them in

$$\max\left\{ck\log_2\frac{cn}{ck}, \frac{1}{\ln 2}ck\ln ck\right\}$$
$$= \max\left\{ck\log_2\frac{n}{k}, c\frac{1}{\ln 2}k\ln k + O(k)\right\}$$
$$\sim c\left\{k\log_2\frac{n}{k}, \frac{1}{\ln 2}k\ln k\right\},$$
$$= cT^*$$

tests, as required.

Second, ETI$_{\text{saff}}$ is achieved by an idea inspired by the SAFFRON scheme of [12] (see also [1]). Suppose for the moment that we have $T = 2\log_2 n/k$ tests. Take a subset of $n/k = 1/p$ items, and note that it contains exactly one infected individual with probability

$$\frac{n}{k}p(1-p)^{n/k-1} = (1-p)^{1/p-1} \to e^{-1}.$$

Number the individuals from 1 to n/k, and let $\mathbf{v}_i \in \{0,1\}^l$ be the number i written in binary, where the length is $l = \log_2 n/k = T/2$ (up to rounding). Further let $\bar{\mathbf{v}}_i\{0,1\}^l$ be $\mathbf{v}_i$ with the 0s and 1s flipped. Then individual i is placed in the tests that correspond to the positions of the 1s in the vector $(\mathbf{v}_i\bar{\mathbf{v}}_i) \in \{0,1\}^{2l} = \{0,1\}$. Note that this means each item is in exactly $l = T/2$ of the $2l = T$ tests.

If the set contains zero infected individuals, then all T tests will be negative, and we can rule all the individuals to be noninfected. If exactly one of the individuals is infected, then exactly $l = T/2$ of the T tests will be positive, and the first $T/2$ test outcomes will 'spell out' in binary the number of the infected individual. If there are two or more infected individuals, then strictly more than $l = T/2$ of the tests are positive, and we find none of the expected items. Thus with our T tests we find one infected individual if and only if the set contains exactly one infected individual, with is an expected number of e^{-1}. Hence the ETI is

$$\text{ETI}_{\text{saff}} = \frac{T}{e^{-1}} = \frac{2\log_2\frac{n}{k}}{e^{-1}} = 2e\log_2\frac{n}{k}.$$

Given the actual value of T we have, we split it into segments of size $2\log_2 n/k$, and run the SAFFRON-inspired algorithm separately on each, obtaining the same ETI.

For a given parameter α, we choose whichever out of the 'full' and 'SAFFRON' methods gives the best ETI. This proves the theorem. ∎
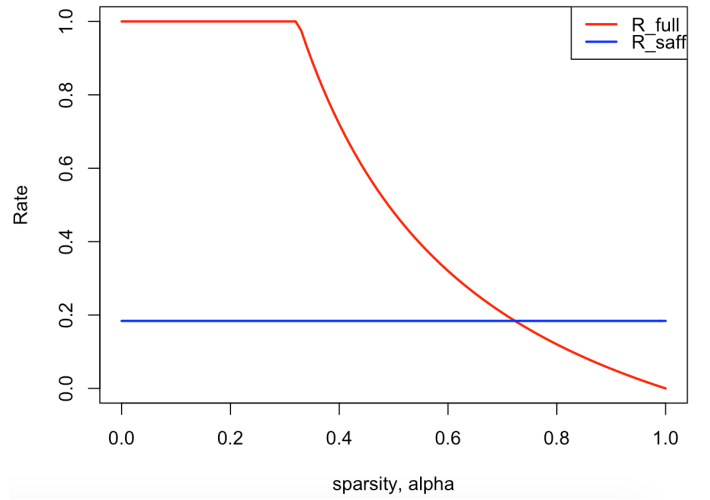

Fig. 1. Rate of finding isolated individuals in the theoretical model, from Theorem 3. For given α, the better method is the one with higher rate.

Note that the 'full' algorithm tests fewer individuals, but seeks to find every infected individual who was tested. In contrast, the SAFFRON-inspired algorithm tests more individuals, but only expects to find $e^{-1}/(1-e^{-1}) = 58\%$ of those infected.

One convenient way of interpreting Theorem 2 is by considering the *rate* of group testing [1], [4]. Here, we define the rate to be

$$\text{rate} = \frac{H(p)}{p\text{ETI}},$$

where $H(p)$ is the binary entropy. Standard information theoretic conditions tell us that the rate is bounded above by 1; we want the rate to be as close as possible to 1, the higher the better.

We can rewrite Theorem 2 as follows.

Theorem 3: When $p \sim n^{-(1-\alpha)}$, the theoretical model can achieve rates up to

$$R = \max\{R_{\text{full}}, R_{\text{saff}}\},$$

where

$$R_{\text{full}} = \min\left\{1, (\ln 2)^2\frac{1-\alpha}{\alpha}\right\}, \qquad R_{\text{saff}} = \frac{1}{2e}.$$

This is illustrated in Fig. 1. Note that, for large α, the SAFFRON method gives a higher rate than R_{full}, the usual rate for finding all defectives in the usual full group testing problem.

REFERENCES

[1] M. Aldridge, O. Johnson, and J. Scarlett, "Group testing: an information theory perspective," *Foundations and Trends in Communications and Information Theory*, vol. 15, no. 3-4, pp. 196–392, 2019.

[2] L. V. Truong, M. Aldridge, and J. Scarlett, "On the all-or-nothing behavior of bernoulli group testing," *IEEE Journal on Selected Areas in Information Theory*, 2020, https://arxiv.org/abs/2001.10137.

[3] M. Aldridge and D. Ellis, "Pooled testing in the COVID-19 pandemic," 2021, book chapter, forthcoming.

[4] M. Aldridge, L. Baldassini, and O. Johnson, "Group testing algorithms: bounds and simulations," *IEEE Transactions on Information Theory*, vol. 60, no. 6, pp. 3671–3687, 2014.

[5] O. Johnson, M. Aldridge, and J. Scarlett, "Performance of group testing algorithms with near-constant tests per item," *IEEE Transactions on Information Theory*, vol. 65, no. 2, pp. 707–723, 2019.

[6] A. Coja-Oghlan, O. Gebhard, M. Hahn-Klimroth, and P. Loick, "Information-theoretic and algorithmic thresholds for group testing," *arXiv*, 2019, https://arxiv.org/abs/1902.02202.

[7] ——, "Optimal group testing," *arXiv*, 2020, https://arxiv.org/abs/1911.02287.

[8] M. Aldridge, "Conservative two-stage group testing," *arXiv*, 2020, https://arxiv.org/abs/2005.06617.

[9] A. Z. Broder and R. Kumar, "A note on double pooling tests," *arXiv*, 2020, https://arxiv.org/abs/2004.01684.

[10] R. Dorfman, "The detection of defective members of large populations," *The Annals of Mathematical Statistics*, vol. 14, no. 4, pp. 436–440, 1943.

[11] L. Mutesa et al, "A pooled testing strategy for identifying SARS-CoV-2 at low prevalence," *Nature*, 2020.

[12] K. Lee, K. Chandrasekher, R. Pedarsani, and K. Ramchandran, "Saffron: A fast, efficient, and robust framework for group testing based on sparse-graph codes," *IEEE Transactions on Signal Processing*, vol. 67, no. 17, pp. 4649–4664, 2019.