6-2021

Needles in a Haystack: How Pooling Can Control Error Rates in Noisy Tests

Arockia David Roy Kulandai
Marquette University

J. Stella
Xavier Institute of Engineering Mahim

John Rose
Xavier Institute of Engineering Mahim

Thomas Schwarz
Marquette University, thomas.schwarz@marquette.edu

2021 IEEE International Conference on Communications Workshops (ICC Workshops), (June 2021). DOI.

Needles in a Haystack: How Pooling Can Control Error Rates in Noisy Tests

Arockia David Roy Kulandai
Marquette University, Milwaukee, WI, USA
J Stella
Xavier Institute of Engineering Mahim (West), Mumbai, India
John Rose
Xavier Institute of Engineering Mahim (West), Mumbai, India
Thomas Schwarz
Marquette University, Milwaukee, WI, USA

Abstract:

Testing many individuals for a reasonably rare condition using imperfect, time consuming, and expensive tests can be facilitated by pooling. Pooling groups samples from different individuals that are then tested for the existence of a pathogen. An individual is diagnosed as a carrier if a threshold of the tests to which the individual contributed samples is positive. Our assumptions dictate a testing strategy that is not adaptive, with the

exception of retesting positively diagnosed persons individually. Pooling is a standard proposal to stretch the supply of test kits. We show that it can also be used to control the false positive and false negative rate of tests, as long as errors are attributable to the lack of quality in the tests themselves and not to a lack of progression in the disease process where what is testable has still to develop. As the medical response to a new pandemic becomes more sophisticated, quality issues with tests will be less prevalent and our contribution will loose value. However, at the beginning of a new pandemic, wide-spread pooling with imperfect tests can prevent the disease from becoming a pandemic.

SECTION I. Introduction

When the first wave of the Corona Virus pandemy expanded world-wide, countries scrambled to implemement rapid and accurate testing at their borders. Even a year after the appearance of the virus, nationwide or even campus-wide testing exists only in rare circumstances, despite its obvious benefits in combatting the spread of the pandemic. Unfortunately, a new disease with pandemic potential is likely to appear in the near future. Early aggressive testing, even if the tests are imperfect, can control the spread to the extent that health authorities can trace the carriers who have not been caught in the dragnet of testing and isolate those they had contact with. This presupposes the early, wide-spread availability of tests that in the beginning of a pandemic will be quite imperfect.

Pooling or group testing is a long-established procedure to make better use of expensive tests [1]. Pooling is a testing strategy that collects several probes from an individual, combines samples from several individuals in a clever way into groups, and then tests each group for the presence of a pathogen or other indicator of a disease. An individual is diagnosed as positive if samples of the individual are in a threshold number of groups that tested positive for the disease. It was apparently first invented in 1943 in order to prevent waste in testing US army recruits for Syphilis [2], and has now been deployed against Covid as well [3], [4], [5], [6], [7]. Below, in Section III, we briefly review some pooling strategies.

The principal motivation for pooling is the better use of a scarce resource (the tests themselves) by allowing more individuals to be tested with the same number of tests. Here, we show that they can also be used to establish a trade-off between the false negative and the false positive rate of testing, in case the tests are unreliable, i.e. "noisy", which can be presumed at the beginning of the pandemic. To deal with false positives, we envision a one-step-and-a-half solution (similar to Mallapaty's Method 4 [8]) that retest only those diagnosed as positive. Our main goal is to limit the more dangerous false negative rate.

Our contribution shows that noisy tests can be made more reliable through pooling. The procedure is of value whenever there is a scramble to develop and produce tests. We envision using error-prone tests in an early stage of pandemic response, before more accurate tests become available, in order to test large swaths of the population and to eliminate as many non-symptomatic carriers as possible from circulation through quarantine. For Covid-19, this contribution comes too late, unless of course the virus mutates in a manner that make it undetectible by the current tests, or in situations of lack of resources. We envision the need for testing large populations, i.e. all people living in a certain town, all children at a school or day care facility, or all individuals crossing a border.

A test for an infectious disease can fail for two reasons. First, we might be testing for an indicator that is currently or not yet present in all infected individuals. This will create unavoidable false negatives. The second reason lies in defects in the test itself or in the application of the test, for example, because of an error by the health care worker administering the test. By not using tests from the same batch, we can assume that failures due to bad quality are almost independent of each other. By organizing testing, we can also achieve that failures

in test procedure are not only rare but also almost independent of each other, for instance, by producing several separated lines of test processing.

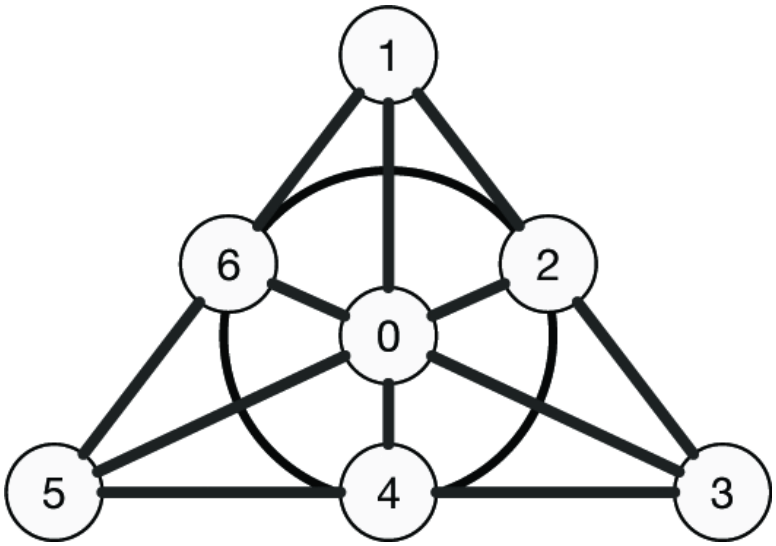


Fig. 1. Fano Plane.

As a more concrete example, assume a stream of travellers crossing a border. A few health care workers take independently repeated samples from the mouth of each traveller and one also from the nose. These samples receive an electronic tag. Within a few minutes, these samples are assigned to different batches based on a procedure that guarantees that samples from a pair of travellers do not end up in the same batch more than twice. When there are enough samples in a batch, the batch is tested, and the results are electronically stored and periodically evaluated. (Notice that the information processing needs are simple enough to be done on a smart phone). When the test results are available, individuals likely to be infected are identified and asked to quarantine immediately or confirm the presence or absence of an infection with a more accurate test. How this is done will depend of course on the time it takes to perform the tests, so the travellers might have already moved on. If instead of travellers we have children at a school or students at a university campus, the logistic challenges are much more limited.

In the remainder of this article, we first discuss grouping a batch of *N* individuals into fixed sized groups. Since testing takes time, we do not consider adaptive pooling schemes. We then calculate the false positive and false negative rates in dependence on the overall prevalence of the virus. Our schemes are successful, but can of course not work create an accuracy out of nothing, meaning that we cannot identify infected individuals whose infection cannot be picked up by the test, but can deal with shortcomings of individual tests.

SECTION II. Grouping

A. Fano Plane

To illustrate the power of our method, we start with a simple example. The Fano plane consists of seven points 0, 1, …, 6 grouped into seven lines, see Figure 1, which shows six lines as straight lines and one line as a circle. Each point lies in exactly three lines and each line has exactly three points. A line is uniquely determined by any two points. This makes the Fano Plane into a Steiner Triple System, named *S*(2, 3, 7), an instance of an Incomplete Block Design [9]. In our context, a point corresponds to an individual and the groups to the lines.

For decision making, we adopt as is common a threshold strategy. This choice is dictated to us because *a priori*, each individual is equally suspicious of being infected. First, we look at a threshold of $k = 3$, and accordingly, diagnose an individual as diseased if all three of the three samples containing a probe of that individual are positive.

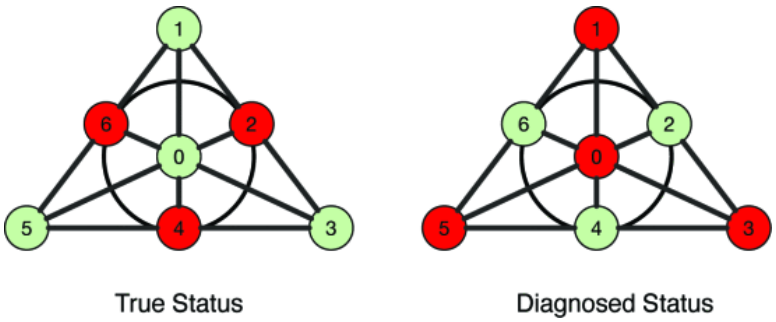


Fig. 2. Fano Plane: Example for a False Positive.

Even if the test is perfectly accurate, this scheme allows for false positives, but never for false negatives. Figure 2 gives an example, where individuals 0, 1, and 3 are infected. The probe of individual 5 is contained in groups {0, 2, 5}, {3, 4, 5}, and {1, 5, 6}, all of which contain a diseased individual's probe and are therefore positive. Thus, with any threshold (other than 0), individual 5 is diagnosed as likely to be infected. Presumably, this will result in an order to quarantine until a more accurate test or retesting will liberate it. Individuals 2, 4, and 6 escape this fate, since they form a group without infected individuals. Let ρ denote the incidence rate of infections, i.e. the probability that a randomly chosen individual is infected. Then the individual suffers a false positive diagnosis if (a) the individual is not infected (with probability $1 - \rho$), and (b) in each group, at least one of the other two individuals is infected. The probability that there is an infected person in a group of two is $\rho^2 + 2(1 - \rho)\rho$ and the probability for (b) is $2\rho^3 - \rho^6$. For an infection rate of 1%, this amounts to 7.8×10^{-6}. As we lower the threshold, the false positive probability moves from $8\rho^3 - 20\rho^4 + 18\rho^5 - 7\rho^6 + \rho^7$ for a threshold of three, to $12\rho^2 - 40\rho^3 + 55\rho 4 - 39\rho^5 + 14\rho^6 - 2\rho^7$ for threshold two, and to $6\rho - 21\rho^2 + 35\rho 3 - 35\rho^4 + 21\rho^5 - 7\rho^6 + \rho^7$ for threshold one.

Obviously, threshold one gives us an unacceptable high false positive rate, but the quadratic component of ρ for threshold two is quite acceptable for low infection rates. At an infection rate of 1%, the false positive rate is 0.116%.

For a completely accurate test, this added probability of false negatives does not buy us anything, but the situation is different if there is a non-negligible false negative rate. If individuum 1 is infected, then three instead of only one group can show this infection.

False negatives are much more dangerous than false positives, since they threaten whole communities instead of causing a severe inconvenience to an individual. Calculating rates exactly is however difficult and we use simulation that is easier to verify. We simulated the Fano plane testing plan using Python. We also calculated 99% confidence intervals, but do not display them in the graphs because we increased the number of simulations so much that the difference betweeen the upper and lower bound could not be visibly detected in our graphs. In our graphs, we give the rates in percents, not in absolute values. We assumed a false positive rate of 1% and false negative rates of 10%, 20%, and 30%. We give the results in Fig. 3, where we also give the rate of false positives and negatives if we just test individuals without pooling. As we can see, for Threshold 1 the false negative rates are very low, but the false positive rate very high, higher than the baseline of $1 \bullet (1 - \rho)\%$. For Threshold 3, the opposite happens, though the false negative rate is still improved over the baselines $10\rho\%$,

20ρ%, and 30ρ%, respectively. Threshold 2 balances false and negative rates successfully and makes our point that pooling is also useful to control error rates.

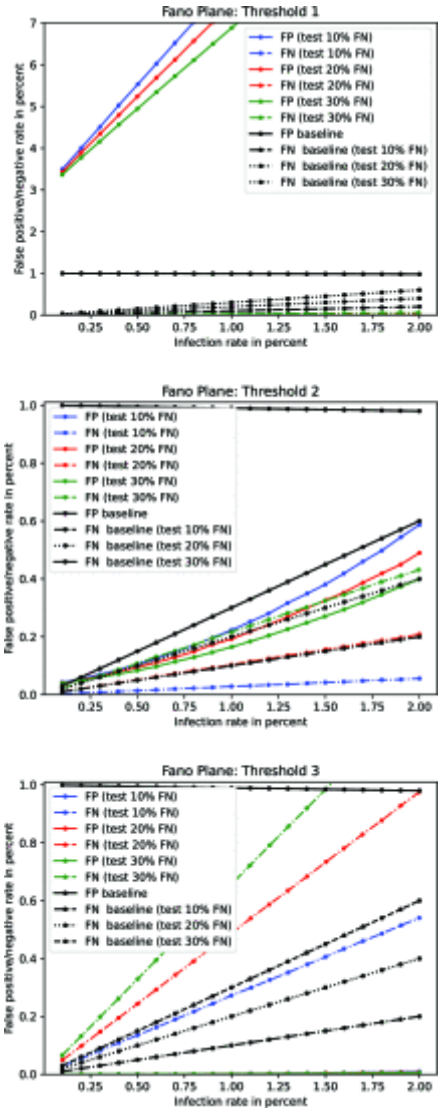


Fig. 3. Fano Plane with Treshold 2: False Positives and False Negatives for imperfect tests. The False Positives rate of the individual test is 1% and the false negative rates are 10%, 20%, and 30%.

B. Projective Plane of Order 3

An ideal layout assigns each individual to a fixed number of groups such that any two individuals share at most one group. A number of combinatorial designs exists where each pair of two individuals shares exactly one group. We now look at families of Incomplete Block Designs to show that this behavior is typical.

The projective plane of order 3 is such an arrangement. It is given by the following incidence matrix:

$$\Pi = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}$$

It consists of 13 points and 13 lines. The points correspond to the rows and the lines correspond to the columns of the matrix. A point is incident to a line exactly if the corresponding coefficient in Π is 1. As before, we identify individuals with points and testing groups with lines. Each point is in exactly four lines and each line contains exactly four points. The symmetry of the incidence matrix reflects the self-duality of the geometry.

Testing proceeds as before. The samples from four individuals corresponding to the points in the same line are mixed and evaluated together. We diagnose a diseased individual, if three of the four tests with samples from the individual are positive. We give the results in Figure 4, where in addition to the false positive and false negative rates we also give the false positive and false negative rates for the baseline, where we do not use grouping but just use one test per individual.

C. Singer Difference Set

A different combinatorial approach to find Incomplete Block Designs uses difference sets. We start with the (21, 5, 1) "Singer" difference set $D = \{0, 1, 6, 8, 18\}$ [9] and generate twenty-one groups by adding integers modulo 21 element-wise to D. This gives us twenty-one blocks such that each pair of numbers is in the same block exactly once. As before, we make these blocks into our testing groups. We use a threshold of t in order to diagnose an individual as infected. By varying the threshold between 2 and 3, we achieve moderate or extremely aggressive control of false negatives (Figure 5), where we also assume a native false positive rate of 1%.

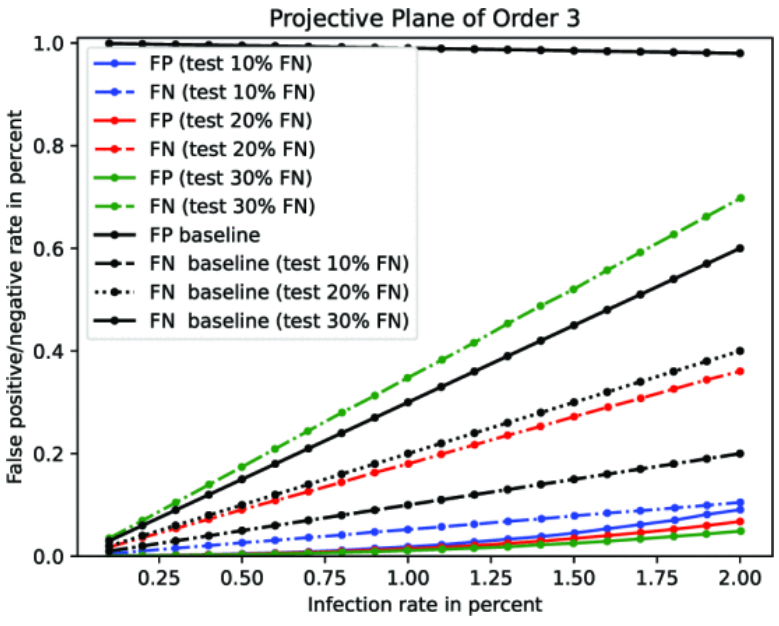
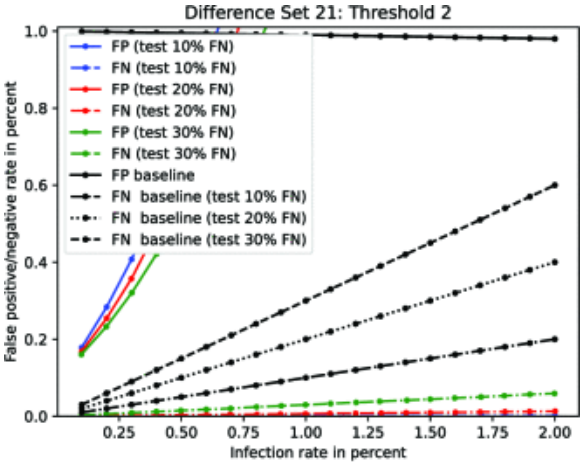


Fig. 4. Simulation results for the projective plane of dimension 3 with a threshold of 3.



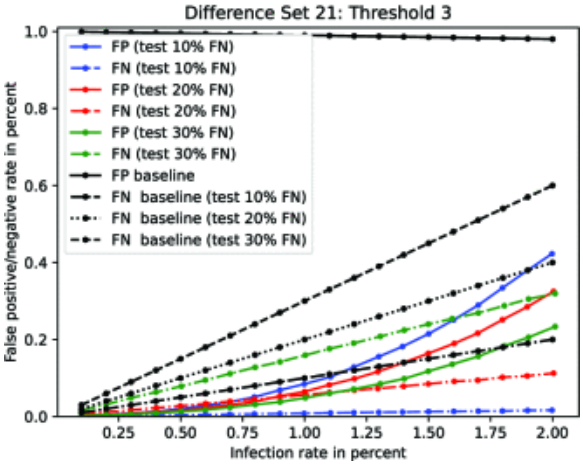


Fig. 5. Simulation results for the difference set with thresholds two and three.

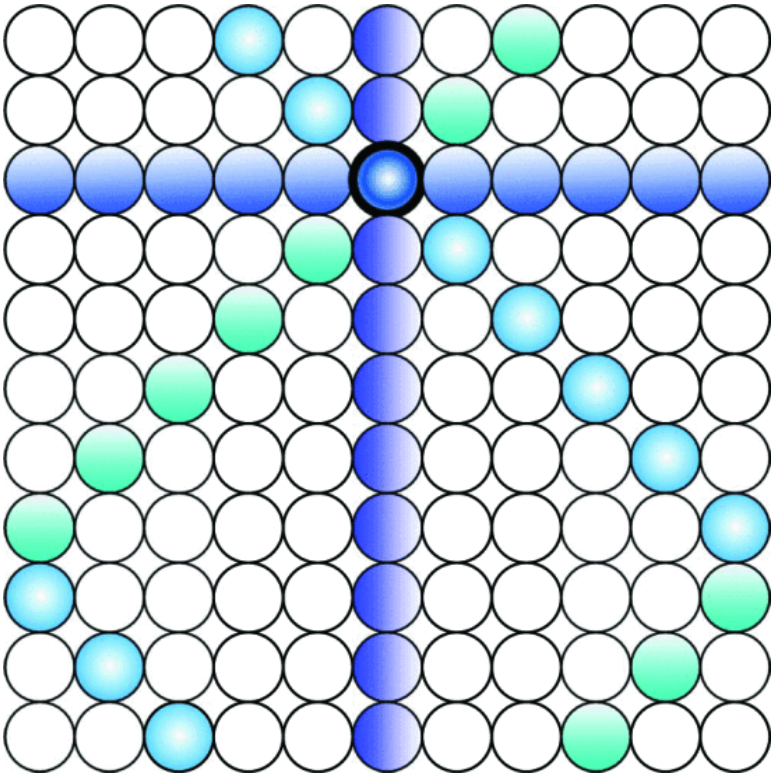


Fig. 6. Square layout.

D. Square Layout n^2

Our examples so far have not saved in the number of tests. For example, the Fano plane yields a test battery of seven tests for seven individuals and the projective plane a battery of thirteen tests for thirteen individuals. Drawing on experience in designing failure resilient disk arrays, we generalize now a square layout for disk arrays, [10]. Other designs from this field can also be applied to pooling.

We arrange a group of n^2 individuals in a square. Each individual is grouped into four groups of n individuals, consisting of a row, a column, a main diagonal, and a secondary diagonal. If the size of the square is even, $n^2/2$ of the diagonals will intersect a secondary diagonal in two fields of the square. This is not a big detriment to our schemes as our simulations will show, but it is an incentive to find a cleaner design.

Unlike the previous schemes discussed above, a square layout with $n > 4$ saves in the total number of tests administered, but at the cost of introducing false positives, even if each individual test does not have false positives. As before, we denote the probability of an infection with ρ. A healthy individual will see one of their tests pooled with an infected individual with probability

$$p_1 = 1 - (1 - \rho)^{n-1}.$$

If m tests are performed, a false positive happens with probability $p_m = p_1^m$.

The rate at which we observe false positives is this probability multiplied with the rate of infected individuals, i.e. multiplied with $1 - \rho$. Figure 7 shows the result. Clearly, the incidence rate of false positives for the critical infection rate range is quite low and can be justified just by the expected savings in tests.

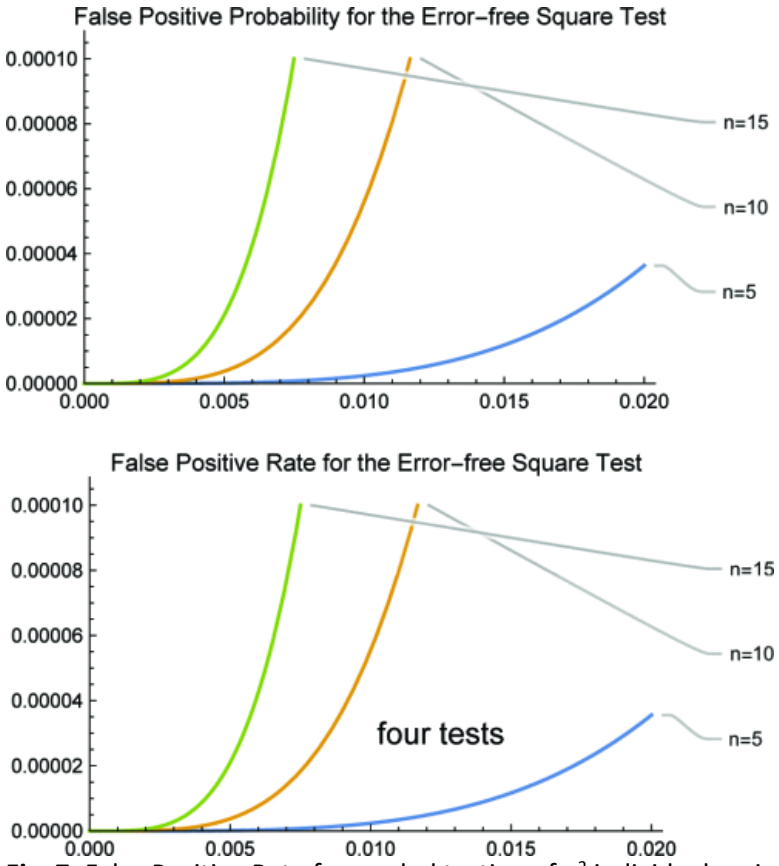


Fig. 7. False Positive Rate for pooled testing of n^2 individuals using three (top) and four (bottom) tests.

If there are no infected people, an n^2 scheme tests n^2 individuals with 3 or 4 tests per individual and uses $3n$ or $4n$ tests, respectively, or $3/n$ or $4/n$ tests per individual. We can change to a semi-adaptive scheme where individuals diagnosed as positive through pooling are tested again. In today's situation, the re-testing could be done with more accurate tests, but for argument's sake, we assume that we use the same type of tests. The infection rate determines the number of additional tests, as we now elaborate. From the point of view of a single person, this person is either infected and tests positive and therefore needs an additional test, or the person is healthy, but is diagnosed positive falsely. The probability that a person will need to be tested once in addition is therefore $\rho+(1-\rho)p_m$. For high infection rates, almost every person will have to be tested again, so that no savings accrue.

For low infection rates this is not the case. Figure 8 gives the result of our calculations for infection rates of up to 5% and shows that the savings of pooling are mostly maintained even if we retest. Notice that even with threshold testing, an infected individual tests positive in all groups and is therefore correctly diagnosed. With other words, there are no false negatives.

What happens if the tests are not error-free? The threshold number of positive tests used to diagnose an individual as infected is then important. First, we assume that the false positive rate is negligible but a non-negligible false negative rate *gamma* for each test. If we set a threshold of four out of four tests, then the false negative rate will be almost four times as high. A threshold of three therefore seems more prudent. We now calculate the probability of false positives and negatives.

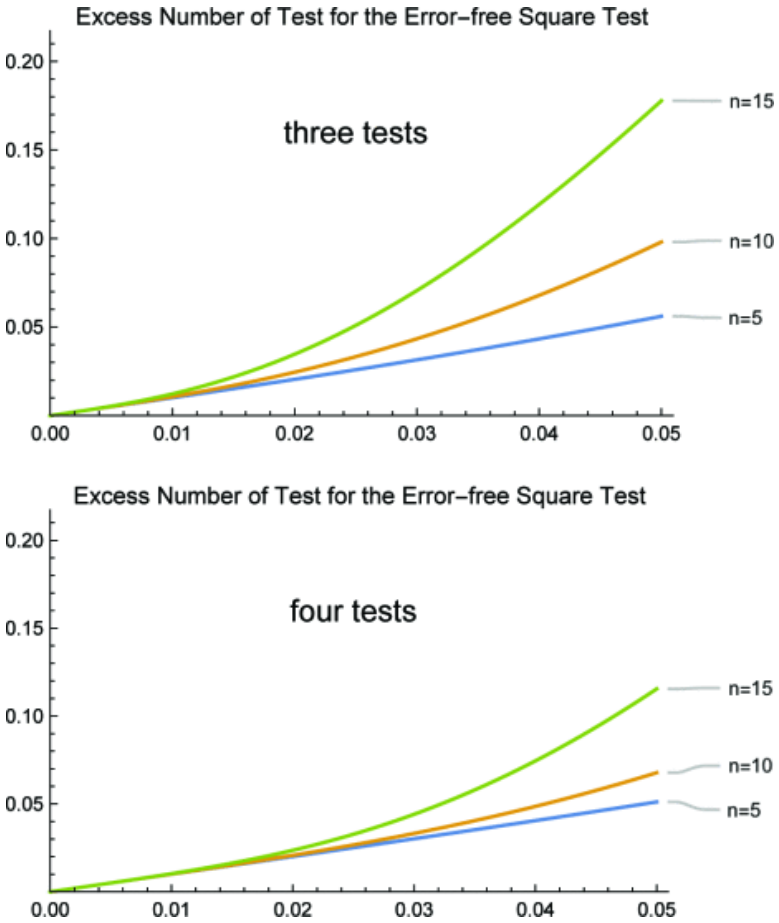


Fig. 8. Retests per individual necessary after pooled testing of n^2 individuals using three (top) and four (bottom) tests.

A healthy person will be diagnosed infected if three or four of the groups test positive. A group tests positive if at least one of its $n - 1$ other members is infected and the test is indeed positive and not a false negative. The first condition occurs with probability $1 - (1 - \rho)^{n-1}$, the second with probability $1 - \gamma$. Therefore, a non-infected individual will have one of its four group tests come back positive with probability $p_1 = (1-\gamma)(1-(1-\rho)^{n-1})$ and the battery will declare this healthy individual infected with probability

$$p_{fp} = \binom{4}{3}(1 - p_1)p_1^3 + p_1^4.$$

Since a portion of $1-\rho$ of the population are not infected, the rate at which we observe false positives is

$$(1 - \rho)p_{fp}.$$

A false negative can only happen to an infected individual. A single group test fails with probability $1 - \gamma$. Two or more of the four tests need to fail for the individual to not be diagnosed, which happens with probability

$$\binom{4}{2}\gamma^2(1 - \gamma)^2 + \binom{4}{3}\gamma^3(1 - \gamma) + \gamma^4.$$

Multiplying this number with ρ gives the rate at which we observe false negatives.

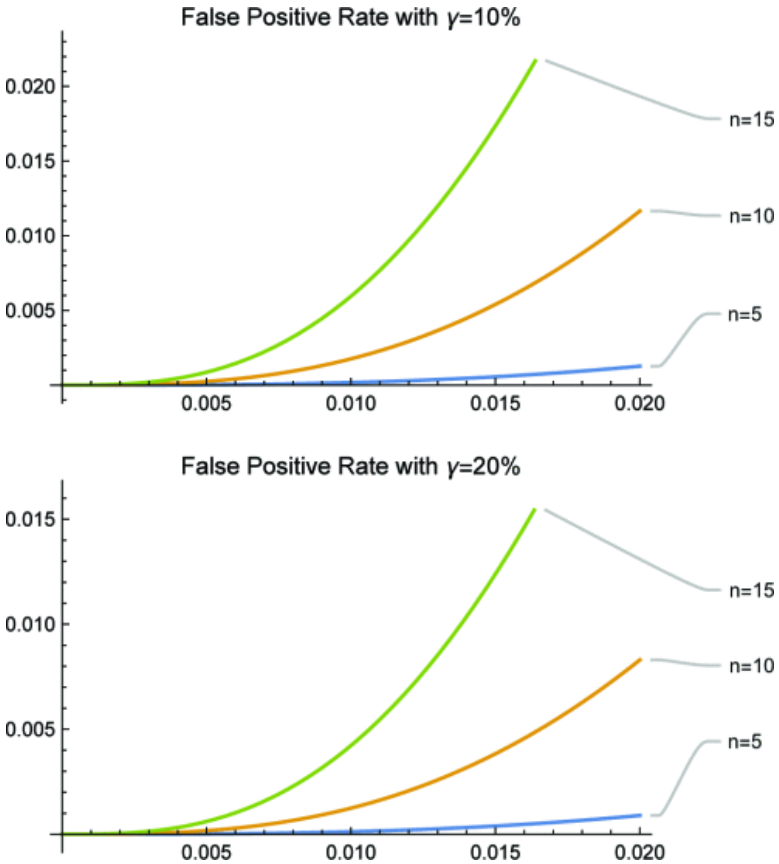


Fig. 9. False Positive Rate for pooled testing of n^2 individuals assuming 10% and 20% false negative rates in the individual group tests.

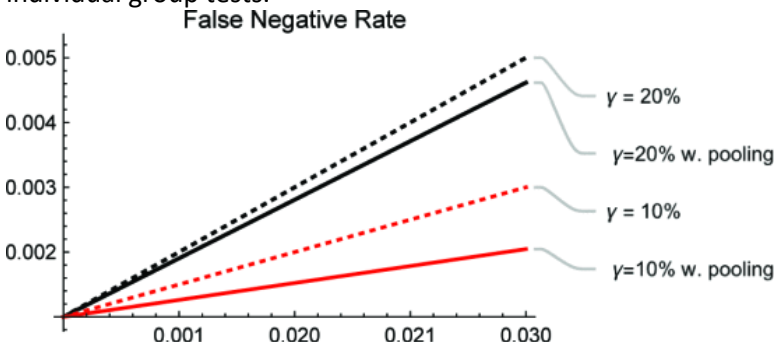


Fig. 10. False Negative Rate for pooled testing of n^2 individuals assuming 10% and 20% false negative rates in the individual group tests.

Figures 9 and 10 show the results. Again, we succeeded in controlling the false negative rate and overall lower the inaccuracy of the test. For lack of space, we do not present our simulation results that confirm the theoretical results.

SECTION III. Related Work

The idea of pooling statistical tests appears to be made first in an informal disccussion on lowering the number of Wasserman blood tests for Syphilis among US army inductees by David Rosenblatt at the Research Division of Price Administration in Washington in 1943 [2]. The idea was taken up by Dorfman who wrote the seminal paper about it [11]. Besides a short note by Sterret [12], it did not receive much attention in the post-war period. Sobel and Groll wrote a 74-page paper on group testing motivated by an industrial problem, namely whether a gas

was leaking from a device [13]. Since then, group testing or pooling has been the subject of a plethora of articles and yielded various avenues of research.

A first division of group testing is the distinction between adaptive and non-adaptive method. Adaptive testing lets the administration of testing depend on previous tests. This necessarily leads to longer times to decision and is not appropriate in our context. The goal of testing can be exact recovery vs. partial recovery, where the former is guaranteed to yield a correct diagonosis for each individual tested. Similarly, but differently, we can distinguish between zero error probability regimens and small error probability depending on whether the set of infected individuals is always recognized or only with high probability. A noiseless testing regimen assumes no false positives or false negatives for each individual test.

Some algorithms require to know the number of infected individuals and sometimes assume that they are uniform randomly distributed among the test set (*a combinatorial prior*), others assume that each individual is infected with a fixed probability (a *i.i.d. prior*).

Here, we assumed a i.i.d. prior, and use non-adaptive threshold methods for noisy tests.

During the Covid pandemic, pooling / grouping has been proposed to help stretch rare tests. The work by Mutesa and colleagues and Beunardeau and colleagues are outstanding examples [3], [5]. Their focus is in the original line of Rosenbblatt and Dorfman, namely of making better use of limited or expensive tests. Here, we deal with a quality problem in early tests.

SECTION IV. Conclusions

Pooling is a strategy used to make better use of scarce tests. Here, we showed that pooling is also a strategy to deal with somewhat defective tests. Each test in practice has false positives and negatives, though the rates might be negligible. The false negative rate reflects two different phenomena: First and by far more common in mature tests, an infected individual might not exhibit the marker or set of markers for which we test. For example, the "Fast Covid Tests" used in Germany are said to be not capable of detecting a Covid infected individual with 30% probability five days after infection, simply because the infection process has not progressed sufficiently. In contrast, the false negative probability for several fast tests are far less than the 20% set as a minimum standard by the World Health Organization for symptomatic patients [14], [15]. The second and by now presumably very rare cause for existing tests for Covid 19 is a defect in the test itself, reflecting the spectacular success of the medical community in addressing the current pandemic. However, this is likely to change for the next pandemic or a mutation of the virus strong enough to escape all current testing strategies. While the second of these scenarios is unlikely, the next new infection disease will come soon, in which case fast, wide-spread testing will be crucial in preventing it to become another pandemic.

We have shown that pooling not only has the potential to spread an insufficient number of tests, but also to deal with variance in the quality of the tests. To this end, different groups need to be tested with different batches of the same test, with different versions of the same test, or maybe even with completely different tests. By adjusting the thresholds for diagnosis, we can not only control the probability and rate of false negatives, but also balance between the rate of false positives and negatives.

Our method is limited by the effects of diluting a diseased individual's probe with others' probes, though Perchetti *et al.* cite a community consensus that up to 30 probes are safe [6]. Our examples are below this limit, but this might not be true for the next pandemic. A more important restriction is to approximate our assumptions that test failures are independent of each other. Failure of a test can be caused by a failure in the taking of the probe such as a too shallow or too deep nasal swab or contanimation with another probe. We can assume that these happen rarely enough that we can treat these incidences as occurring independently of each

other. Another cause would be individual variations in the testing ingrediences. However, most noxious to our assumption is the *bad batch*, where all tests from the same production run are partially or completely unreliable. To counteract the bad-batch problem, testing will have to be organized such that the probes taken from an individual use different batches, which can be done by using tests in large numbers in short periods of times, because then the batches can be easily mixed. Ideally of course would be tests that use different methodologies. We have not considered the implications of this because of the difficulty of modeling sets of different false positive and false negative rates.

The design of a test battery based on pooling relies ultimately on Mathematical Combinatorics, which provides many classes of block designs, especially if a small number of tests is desired. We could generalize our results by replacing block designs that work for a fixed sized-set of individuals (e.g. 13) with open designs capable of placing a continuous stream of individuals into a set of small groups such that no two individuals are placed into the same group twice or more.

Finally, wide-spread testing is just one of many ways to successfully counteract the spread of an infectious disease. Contact tracing, hot spot identification, preventive measures as well as fast vaccine development all combine with testing in a successful strategy.

In summary, in addition to help stretch scarce testing resources, pooling can be used to control noisiness in tests. This allows for more effective mass screening at the beginning of a pandemic when diagnostics have not yet matured. This aspect will loose importance as the medical response to a pandemic becomes more effective.

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. 192-392, 2019.
2. D. Du, F. K. Hwang and F. Hwang, "Combinatorial group testing and its applications", *World Scientific*, vol. 12, 2000.
3. M. Beunardeau, É. Brier, N. Cartier, A. Connolly, N. Courant, R. Géraud-Stewart, et al., "Optimal Covid-19 pool testing with a priori information", 2020.
4. S. Lohse, T. Pfuhl, B. Berkó-Göttel, J. Rissland, T. Geißler, B. Gärtner, et al., "Pooling of samples for testing for SARS-CoV-2 in asymptomatic people", *The Lancet Infectious Diseases*, vol. 20, no. 11, pp. 1231-1232, 2020.
5. L. Mutesa, P. Ndishimye, Y. Butera, J. Souopgui, A. Uwineza, R. Rutayisire, E. L. Ndoricimpaye, E. Musoni, N. Rujeni, T. Nyatanyi et al., "A pooled testing strategy for identifying SARS-CoV-2 at low prevalence", *Nature*, pp. 1-8, 2020.
6. G. A. Perchetti, K.-W. Sullivan, G. Pepper, M.-L. Huang, N. Breit, P. Mathias, et al., "Pooling of SARS-CoV-2 samples to increase molecular testing throughput", *Journal of Clinical Virology*, vol. 131, pp. 104570, 2020.
7. I. Torres, E. Albert and D. Navarro, "Pooling of nasopharyngeal swab specimens for SARS-CoV-2 detection by RT-PCR", *Journal of medical virology*, vol. 92, no. 11, pp. 2306-2307, 2020.
8. S. Mallapaty, "The mathematical strategy that could transform coronavirus testing", *Nature*, vol. 583, no. 7817, pp. 504-505, 2020.
9. C. J. Colbourn, CRC handbook of combinatorial designs, Chapman & Hall, CRC press, 2010.
10. T. Schwarz, D. D. Long and J.-F. Pâris, "Reliability of disk arrays with double parity", *2013 IEEE 19th Pacific Rim International Symposium on Dependable Computing*, pp. 108-117, 2013.
11. R. Dorfman, "The detection of defective members of large populations", *The Annals of Mathematical Statistics*, vol. 14, no. 4, pp. 436-440, 1943.

12. A. Sterrett, "On the detection of defective members of large populations", *The Annals of Mathematical Statistics*, vol. 28, no. 4, pp. 1033-1036, 1957.
13. M. Sobel and P. A. Groll, "Group testing to eliminate efficiently all defectives in a binomial sample", *Bell System Technical Journal*, vol. 38, no. 5, pp. 1179-1252, 1959.
14. D. O. Andrey, P. Cohen, B. Meyer, G. Torriani, S. Yerly, L. Mazza, A. Calame, I. Arm-Vernez, I. Guessous, S. Stringhini et al., "Head-to-head accuracy comparison of three commercial COVID-19 IgM/IgG serology rapid tests", *Journal of clinical medicine*, vol. 9, no. 8, pp. 2369, 2020.
15. A. Berger, M. T. Ngo Nsoga, F. J. Perez-Rodriguez, Y. A. Aad, P. Sattonnet-Roche, A. Gayet-Ageron, et al., "Diagnostic accuracy of two commercial SARS-CoV-2 antigen-detecting rapid tests at the point of care in community-based testing centers", *medRxiv*, 2020, [online] Available: https://www.medrxiv.org/content/early/2020/11/23/2020.11.20.20235341.