bioRxiv preprint doi: https://doi.org/10.1101/2020.04.06.028431; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

RAPID, LARGE-SCALE, AND EFFECTIVE DETECTION OF COVID-19 VIA NON-ADAPTIVE TESTING

MATTHIAS TÄUFER

Abstract. Pooling of samples can increase lab capacity when using Polymerase chain reaction (PCR) to detect infections such as COVID-19. However, pool testing is typically performed via an *adaptive testing strategy* which requires a feedback loop in the lab and at least two PCR runs to confirm positive results. This can cost precious time. We discuss a non-adaptive testing method where each sample is distributed in a prescribed manner over several pools, and which yields reliable results after one round of testing. More precisely, assuming knowledge about the overall infection incidence rate, we calculate explicit error bounds on the number of false positives which scale very favourably with pool size and sample multiplicity. This allows for hugely streamlined PCR testing and cuts in detection times for a large-scale testing scenario. A viable consequence of this method could be real-time screening of entire communities, frontline healthcare workers and international flight passengers, for example, using the PCR machines currently in operation.

1. Introduction

One key to containing and mitigating the CoVid-19 pandemic is suggested to be rapid testing on a massive scale [HZW+20, SBY]. It would be beneficial to develop the ability to routinely, and in particular rapidly, test groups such as frontline healthcare workers, police officers, and international travellers. Testing for CoVid-19 is currently performed via the polymerase chain reaction (PCR) on nasopharyngeal swabs [TTY+20]. Typically, the population size significantly exceeds the capacity for testing, with the number of available PCR machines and reagents an important bottleneck in this process.

There are two basic approaches to PCR testing in populations: 1. individual tests, where every single sample is examined, and 2. pooled tests where larger sets of samples are mixed and tested en bloc. Pooled testing was pioneered by Dorfman in 1943 [Dor43] and led to a host of research activity, both on the lab side as well as the theoretical side [AJS19, DH99, DH06]. If the infection is rare in the population, then pooled testing may be advisable. In this case it can assist in optimizing precious testing capacity since most individual results would be negative. Pooling relies on the fact that the PCR is reasonably reliable under the combination of samples: the preprint [YAST+20] suggests that a detection of COVID-19 in pools of size 32 and possibly 64 is feasible.

Key words and phrases. Polymerase Chain Reaction, Pool testing, Non-adaptive testing, COVID-19.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.06.028431; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

While a classic pooling strategy has the advantage that less overall PCR tests are required, there are disadvantages in terms of lab organisation and – more crucially – time: pooling only indicates whether a pool contains at least one infected individual. If samples are tested in pools of size n and the incidence of the infection ρ is small (more precisely, if $\rho \cdot n$ is small) then in average

$$1-(1-\rho)^n = n\rho - \binom{n}{2}\rho^2 + \cdots \approx n\rho \quad \text{(remember that } \rho \cdot n \text{ is small)}$$

of pools will be positive and hence a portion $n \cdot \rho$ of individuals need to undergo a second test. In other words, pooled testing with individual verification of positive pools is an *adaptive testing strategy*, the lab organisation for which is a labour, management, and resource intensive process. It has several drawbacks, since it requires keeping multiple lab samples and re-running of the time-intensive PCR process. The lab feedback loop makes the entire workflow more susceptible to delays (see Figure 2). This may result in delays in individual results – a particular problem when the objective is to rapidly identify infected individuals, who may infect others while waiting for the test outcome. Furthermore, since the number of samples selected for a second round of testing is a random quantity, some reserve capacity is required to prevent further delays. This makes it more challenging for the lab to operate near its maximal capacity.

In the theoretical research on testing strategies the distinction is made between *adaptive testing*, for example when all samples in a positive pool undergo a second round of testing, and *non-adaptive* strategies, where all tests can be run simultaneously [DH99]. Testing every sample individually can be considered as a trivial non-adaptive strategy, but there exist non-adaptive strategies which combine the benefit of pooling with the advantages of non-adaptive testing.

In this note, we propose a non-adaptive pooling strategy for rapid and large-scale screening for COVID-19 or other infections where detection time is critical. This allows for significant streamlining of the testing process and reductions in detection time. Firstly because only one round of PCR is required, and secondly because it eliminates actions in the lab workflow that require input from results determined in the lab, i.e. the testing infrastructure can be organized completely linearly, cf. Figure 2 for an illustration. The strategy will systematically overestimate the number of positives, but we can provide error bounds on the number of false positives which scale very favourably with large numbers and will be small in realistic scenarios.

2. Definition of the non-adaptive testing strategy: creation of multipools

Our testing strategy is as follows: every individual's sample is broken up into k samples and distributed over k different pools of size n such that no two individuals share more than one pool. An individual is considered as tested positive if all the pools in which its sample has been given are tested positive. Let us make this definition more formal:

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.06.028431; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

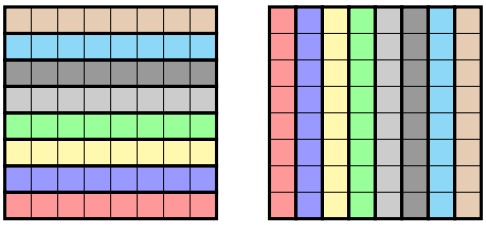

FIGURE 1. Pooling along rows and columns.

Definition 1 (Multipools)**.** *Let a* population *$(X_1, \dots, X_N)$ of size N, a* pool size *n, and a* multiplicity *k be given, and assume that Nk is a multiple of n. We call a collection of subsets/pools of $\{X_1, \dots, X_N\}$ an (N, n, k)-*multipool*, or briefly* multipool*, if all of the following three conditions hold:*

(M1) Every pool consists of exactly n elements.
(M2) Every sample X_i is contained in exactly k pools.
(M3) For any two different samples X_i, X_j there exists at most one pool which contains both X_i and X_j.

If $N = n^2$ and $k = 2$ then the construction of an (N, n, k)-multipool is quite straightforward, see also Figure 1: arrange the N samples in a rectangular grid and then pool along every row and column, cf. [SSW+16, FFLH, ZDF+14]. However, as we shall see below, $k = 2$ is in many realistic scenarios insufficient for the desired precision.

The recent preprint [FFLH] proposes to arrange samples in a (3 or higher dimensional) hypercube and to pool along all hyperplanes. This makes every individual sample appear in three or more pools, but it is *not* a multipool in the sense of Definition 1 above, since in dimension three and higher, any two hyperplanes will intersect in more than one point, in violation of Property (M3). This creates unnecessary correlations between different pools and impairs performance. One proper and flexible way to construct multipools is given by the Shifted Transversal Design [TM06] which we explain in Section 4.

3. Controlling the number of false positives

We always assume that the incidence ρ of the infection is small compared to the inverse pool size $1/n$. This is a reasonable requirement, also in classical pooling strategies (a ρn portion of samples will have to undergo second testing, thus a large ρn would attenuate the benefit of pooling).

Assuming perfect performance of the PCR, also under pooling (see Section 6 on how to deal with uncertainty here), multipooling will identify all infected individuals, since all their pools will be positive. However, someone might falsely be declared positive if all pools in which they are contained happen to contain an infected sample.

The expected portion of false positives in a multipool strategy is

$$\begin{aligned} \rho_{\mathrm{fp}} &= \mathbf{P}(X_i \text{ negative but all its pools positive}) \\ &= (1-\rho) \cdot \mathbf{P}(\text{all pools containing } X_i \text{ positive} \mid X_i \text{ negative}) \\ &= (1-\rho)\left(1-(1-\rho)^{n-1}\right)^k \end{aligned} \tag{3.1}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.06.028431; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

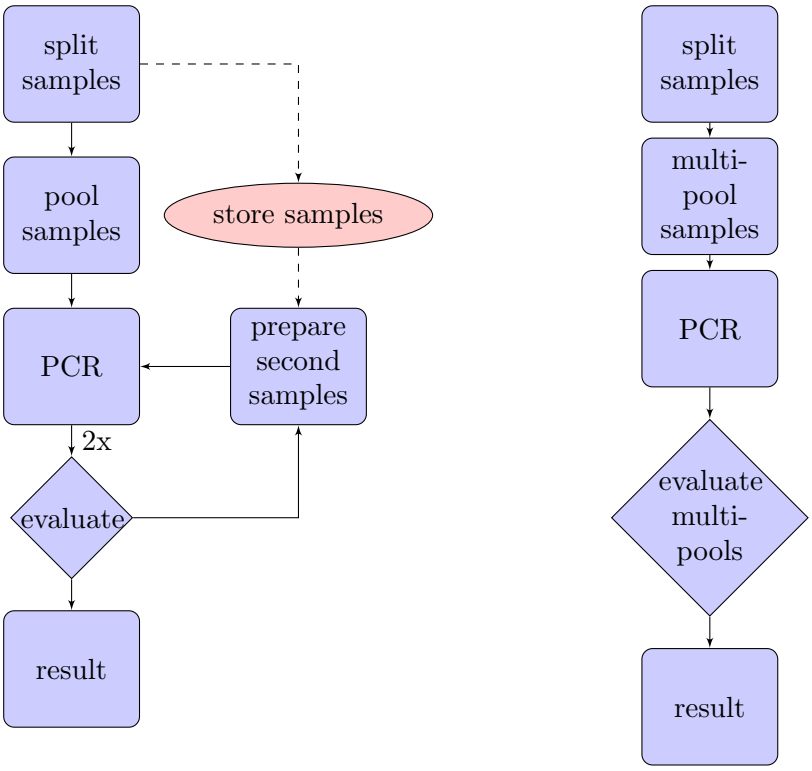

FIGURE 2. Illustration of the work flow for pooling with second test (left) and multipooling (right)

Here, the third identity crucially uses the property (M3) which guarantees independence between the poolmates in the different pools of a sample. By Bayes' rule, the probability to actually be negative when tested positive by the multipool (i.e. the portion of subjects falsely declared positive among all subjects declared positive) is

$$\begin{aligned}
&\mathbf{P}(X_i \text{ negative} \mid \text{all pools containing } X_i \text{ positive}) \\
=&\frac{\mathbf{P}(\text{all pools containing } X_i \text{ positive} \mid X_i \text{ negative})\mathbf{P}(\text{sample } X_i \text{ negative})}{\mathbf{P}(\text{all pools containing } X_i \text{ positive})} \\
=&\frac{(1-\rho)\left(1-(1-\rho)^{n-1}\right)^k}{\rho+(1-\rho)\left(1-(1-\rho)^{n-1}\right)^k}
\end{aligned}$$

Let us calculate for which k the probability of a positive test result being a false positive does not exceed $\epsilon_{\text{fp}} > 0$:

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.06.028431; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$
\begin{aligned}
&\frac{(1-\rho)\left(1-(1-\rho)^{n-1}\right)^k}{\rho+(1-\rho)\left(1-(1-\rho)^{n-1}\right)^k} \leq \epsilon_{\mathrm{fp}} \\
\Leftrightarrow\quad & (1-\rho)\left(1-(1-\rho)^{n-1}\right)^k \leq \epsilon_{\mathrm{fp}}(1-\rho)\left(1-(1-\rho)^{n-1}\right)^k + \epsilon_{\mathrm{fp}}\rho \\
\Leftrightarrow\quad & \left(1-(1-\rho)^{n-1}\right)^k \leq \frac{\epsilon_{\mathrm{fp}}\rho}{(1-\epsilon_{\mathrm{fp}})(1-\rho)} \\
\Leftrightarrow\quad & k \geq \frac{\ln\left(\frac{\epsilon_{\mathrm{fp}}}{1-\epsilon_{\mathrm{fp}}}\right)+\ln\left(\frac{\rho}{1-\rho}\right)}{\ln\left(1-(1-\rho)^{n-1}\right)}.
\end{aligned}
\tag{3.2}
$$

This provides a lower bound on the necessary multiplicity k in terms of the sample size n, the knowledge on the incidence ρ of the infection, and the acceptable portion ϵ_{fp} of falsely positive tests among all positives. Assuming $\epsilon_{\mathrm{fp}} < 1$ and $\rho \leq \frac{1}{2}$ (which are both reasonable assumptions, recall that ρn is small), the lower bound in (3.2) is monotone increasing in ρ. Hence, if the exact incidence is unknown but we have an upper bound on it, we can work with the largest/worst case ρ. Let us summarize our findings in the following

Theorem 1. *Let the incidence of the infection be at most $\rho \leq \frac{1}{2}$, let n be such that $n\rho \leq 1$ and let $0 < \epsilon_{\mathrm{fp}} < 1$. Then if*

$$
k \geq \frac{\ln\left(\frac{\epsilon_{\mathrm{fp}}}{1-\epsilon_{\mathrm{fp}}}\right)+\ln\left(\frac{\rho}{1-\rho}\right)}{\ln\left(1-(1-\rho)^{n-1}\right)} \tag{3.3}
$$

then in any multipooling strategy with pool size n and multiplicity k, the probability of a positive test being a false positive does not exceed ϵ_{fp}.

The number of tests required in a multipool strategy is Nk/n, an improvement compared to individual testing by a factor n/k. A key observation is that the lower bound on k in Ineq. (3.3) scales very favourably with large multiplicities n. Indeed, recall that in an adaptive pooling strategy one wants on the one hand large pool sizes n, but on the other hand $n\rho$ should be small. It is therefore reasonable to have n proportional to the inverse of ρ, i.e. $n\rho \approx C$. Using that $1-\rho \approx 1$ and $1-(1-\rho)^{n-1} \approx (n-1)\rho \approx n\rho$, the lower bound in (3.3) behaves approximately as

$$
k \gtrsim \frac{\ln\left(\frac{\epsilon_{\mathrm{fp}}}{1-\epsilon_{\mathrm{fp}}}\right)+\ln\rho}{\ln(n\rho)} \approx \frac{\ln\left(\frac{\epsilon_{\mathrm{fp}}}{1-\epsilon_{\mathrm{fp}}}\right)+\ln(n/C)}{\ln C}.
$$

that is k grows only logarithmically with the pool size n. An analogous analysis shows that k also grows logarithmically with the inverse of ϵ_{fp} when the error probability ϵ_{fp} is sent to zero.

We have sketched the interplay between pool size n, multiplicity k, false positive probability ϵ_{fp} and the minimal k required for some choices of n, ρ, and ϵ_{fp} in Table 1.

4. Generating multipools

The question for which combiniations (N, n, k) a multipool exists seems to be in general a non-trivial combinatorial problem. We focus here on the

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.06.028431; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

n	ρ	ϵ_{fp}	k
16	1.0%	0.50	2.34
16	1.0%	0.25	2.90
16	1.0%	0.10	3.45
16	1.0%	0.05	3.83
16	1.0%	0.01	4.67
16	0.5%	0.50	2.02
16	0.5%	0.25	2.43
16	0.5%	0.10	2.85
16	0.5%	0.05	3.14
16	0.5%	0.01	3.77
16	0.2%	0.50	1.76
16	0.2%	0.25	2.08
16	0.2%	0.10	2.39
16	0.2%	0.05	2.60
16	0.2%	0.01	3.07
16	0.1%	0.50	1.64
16	0.1%	0.25	1.90
16	0.1%	0.10	2.16
16	0.1%	0.05	2.34
16	0.1%	0.01	2.73

n	ρ	ϵ_{fp}	k
32	1.0%	0.50	3.49
32	1.0%	0.25	4.32
32	1.0%	0.10	5.15
32	1.0%	0.05	5.72
32	1.0%	0.01	6.97
32	0.5%	0.50	2.73
32	0.5%	0.25	3.30
32	0.5%	0.10	3.86
32	0.5%	0.05	4.25
32	0.5%	0.01	5.10
32	0.2%	0.50	2.21
32	0.2%	0.25	2.60
32	0.2%	0.10	2.99
32	0.2%	0.05	3.26
32	0.2%	0.01	3.85
32	0.1%	0.50	1.98
32	0.1%	0.25	2.29
32	0.1%	0.10	2.61
32	0.1%	0.05	2.82
32	0.1%	0.01	3.30

n	ρ	ϵ_{fp}	k
64	0.5%	0.50	4.05
64	0.5%	0.25	4.89
64	0.5%	0.10	5.73
64	0.5%	0.05	6.31
64	0.5%	0.01	7.57
64	0.2%	0.50	2.91
64	0.2%	0.25	3.43
64	0.2%	0.10	3.94
64	0.2%	0.05	4.29
64	0.2%	0.01	5.07
64	0.1%	0.50	2.47
64	0.1%	0.25	2.86
64	0.1%	0.10	3.26
64	0.1%	0.05	3.52
64	0.1%	0.01	4.11
128	0.1%	0.50	3.25
128	0.1%	0.25	3.77
128	0.1%	0.10	4.28
128	0.1%	0.05	4.63
128	0.1%	0.01	5.41

TABLE 1. Lower bounds on the number k required to make the probability of a positive result being a false positive smaller than ϵ_{fp} for different choices of pool size n and bound on the incidence ρ.

case when $N = n^2$ and on constructions based on the Shifted Transversal Design [TM06]. As mentioned in [FFLH], assuming N to be a square number is not really a restriction since numbers N which are not a square can be either divided into smaller squares or extended to a larger square number by adding negative "dummy samples".

It is useful to imagine all N samples arranged in an $n \times n$-square and denote the sample in the i-th row and the j-th column by X_{ij}. For $k = 2$, a (N, n, k)-multipool can be constructed by pooling along rows and columns, as for instance in Figure 3.

We have seen in Table 1 that for many relevant parameters a multiplicity $k = 2$ is not sufficient for the required precision. To deal with the case $k = 3$, one can sample along diagonals, where the diagonals are continued periodically, see Figure 3.

Theorem 2. *Let $N = n^2$ and $n \geq 2$. Then there exists an $(N, n, 3)$-multipool, obtained by sampling along rows, columns, and all periodically continued north-west-to-south-east diagonals.*

In order to pass to $k = 4$, one might be tempted to now pool along the other (north-east-to-south-west) diagonals, but this is not going to yield a

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.06.028431; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

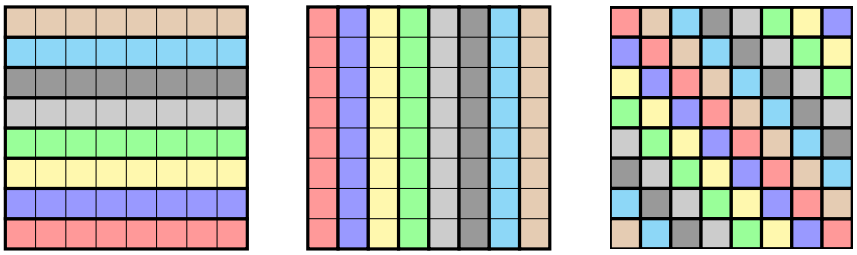

FIGURE 3. Pooling along rows, columns, and periodically continued diagonals.

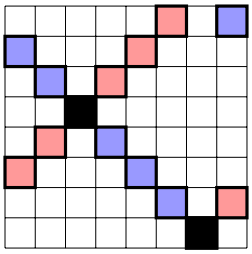

FIGURE 4. The two diagonals (red and blue) intersect in two points (black). They cannot both be used as pools in a multipool.

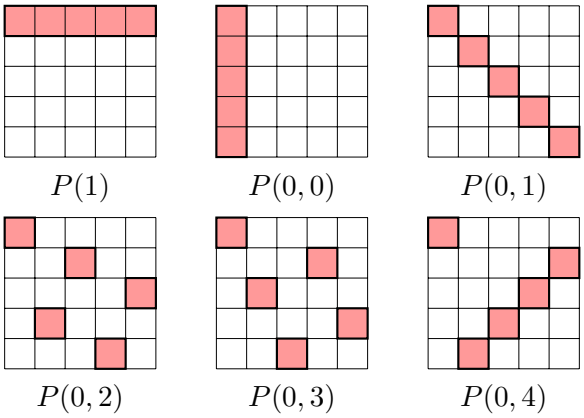

FIGURE 5. Multipools of different slopes as in Theorem 3 for $n = 5$.

multipool in general, see for instance Figure 4 for the case of $n = 8$ where two diagonals intersect in more than one point.

Another option is to sample along lines of different slopes, that is to construct a multipool using pools of the form

$$P(l,m) := \left\{X_{j,l+jm(\bmod n)} : j = 0,\dots,n-1\right\}, \quad l,m \in \{0,\dots,n-1\},$$

where $\bmod n$ means that we calculate modulo n, that is as soon as we exceed $n-1$, we start counting from 0 again. We sketch such sets in Figure 5.

It follows from standard algebraic properties of the integers modulo n (more precisely from the fact that $\mathbb{Z}(\bmod n)$ is a field having multiplicative

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.06.028431; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

inverses if and only if n is prime) that any two such pools do not intersect in more than one point if and only if n is a *prime number*. This leads to:

Theorem 3. *Let n be a prime number and let $N = n^2$. Then, there exists a (N, n, k)-multipool for $k = (n + 1)$, and consequently also for every smaller k. This multipool is given by pooling along all rows, that is:*

$$P(l) := \{X_{l,j} : j = 0, \dots, n-1\}, \quad l \in \{0, \dots, n-1\}.$$

and pooling along all slopes, that is:

$$P(l, m) := \left\{X_{i,l+mj(\mathrm{mod}\, n)} : j = 0, \dots, n-1\right\}, \quad l, m \in \{0, \dots, n-1\}.$$

We show in Figure 5 an illustration in the case $k = 5$. This construction is also referred to as the *Shifted Transversal Design* in [TM06].

If N is a square, but not a square of a prime, an adapted strategy (with smaller pool size n) can be devised by exploiting the fact that $\sqrt{N}$ can be written as a product of a prime and another number.

From a practical perspective it seems to be reasonable to generate multipools by a sequence of unions of two equally diluted pools. This leads to pool sizes which are a power of 2, certainly not a prime number (except for 2 itself). One approach to accomodate for that would be population sizes $N = n^2$ where n is a prime just less than a power of 2, e.g. $n = 31$, which is just less than 32 or $n = 61$ which is just less than 64. Then pools of size n can be mixed by adding a small number of negative dummy samples and proceeding as if n was a power of 2.

5. Examples and scenarios

Let us sketch some concrete examples where the pool sizes are a prime number and where the multipooling strategy might be useful:

N = 961, $\rho \leq 1\%$, $n = 31$. Let the population size be $N = 31^2 = 961$. This could for instance be the number of employees in a company or passengers which depart from an international airport within a certain time window. Let the incidence rate ρ be no more than 1.0% and let us work with a pool size $n = 31$. Theorem 1 then allows to bound the probability of a positive test being erroneous for different multiplicities k, see Table 2. Accepting for

k	ϵ_{fp}	k/n
4	0.32	0.129
5	0.11	0.161
6	0.03	0.194
7	0.008	0.226

Table 2. Probability of a positive result being a false positive and the portion k/n of tests required with respect to individal testing for $n = 31$, $\rho \leq 0.01$ and different k.

instance a false positive probability of 3% requires $6N/n = 186$ PCR tests, 19.4% of what would be required in individual testing. Let us emphasize again here that this means that 3% among the *results flagged as positive* will be false positives, not 3% of the overall population.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.06.028431; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

N = 3721, $\rho \leq 0.1\%$**,** $n = 61$**.** The multipool method scales well with larger numbers. Let the population size be $N = 61^2 = 3721$ and the pool size 61, which is of the order of pools being used for the PCR today [YAST+20]. Let furthermore be the incidence rate be no larger than 0.1%. The error bounds in Theorem 1 then lead to Table 3. If we choose $k = 4$ and accept

k	ϵ_{fp}	k/n
3	0.17	0.049
4	0.012	0.066
5	0.0007	0.082

TABLE 3. Probability of a positive result being a false positive and the portion k/n of tests required with respect to individal testing for $n = 61$, $\rho \leq 0.001$ and different k.

$\epsilon_{\text{fp}} = 1.2\%$ as the probability for positive results being false positives, we need $Nk/n = 244$ tests in order to fast and efficiently test 3721 individuals, that is 6.6% of what would be needed for individual testing.

N = 16129, $\rho \leq 0.1\%$**,** $n = 127$**.** Let us consider a hypothetical scenario where a dilution of samples by a factor 127 is acceptable, that is $N = 127^2 = 16129$. The error bounds given by Theorem 1 are shown in Table 4. If one

k	ϵ_{fp}	k/n
4	0.165	0.031
5	0.023	0.039
6	0.0028	0.047

TABLE 4. Probability of a positive result being a false positive and the portion k/n of tests required with respect to individal testing for $n = 127$, $\rho \leq 0.001$ and different k.

works with $k = 5$ and accepts a false positive rate of $\epsilon_{\text{fp}} = 2.3\%$, then only $Nk/n = 635$ PCR tests would be required, that is 3.9% of what would be needed in the case of individual testing.

6. Discussion and possible extensions

The non-adaptive multi-pooling strategy provides a streamlined and efficient organisation of the testing process and cuts in detection time. This significant benefit comes with potential reductions in accuracy compared with adaptive testing, but this false positive rate can be tightly controlled and tailored to suit the circumstance. The false positive probability ϵ_{fp} deemed an acceptable cost for the increased testing efficiencies may depend on, for example, the infection characteristics, the government policy and resource levels.

A small modification of our strategy might furthermore allow for an improvement of the false negative rate – even compared to usual adaptive pool testing strategies: even though commonly used, pooling samples can potentially dilute samples close to the identification threshold of the PCR and increase the probaility of false negatives. The recent preprint [YAST+20]

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.06.028431; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

estimates a false negative rate of 10% when detecting COVID-19 in pools of size 32. One can reduce this type of false negative in our strategy by declaring all samples which are *in at least* $k-1$ *positive pools* as tested positive. This will on the one hand lower the probability of false negatives, but more importantly it will only mildly affect the false positive rate. This could be seen by adding a next-order term in the error analysis performed leading up to Theorem 1. Furtherore, knowledge on the false positive rate gained through experiments would be required, but the general message that the necessary multiplicity k will grow slowly with large n and small ϵ_{fp} remains.

Let us finally note that the basic idea is close to compressed sensing and sparse recovery [CT06, FH13]. While in our situation the output space consists of $\{0,1\}$-vectors, which make the mathematics we use rather elementary, there also seem to be applications of the PCR where quantitative measurements are taken and where compressive sensing techniques might be applied.

Acknowledgements. The author thanks Christoph Schumacher for helpful discussions. Comments by Emma Lawrance, Albrecht Seelmann and Sasha Sodin are gratefully acknowledged.

References

[AJS19] M. Aldridge, O. Johnson, and J. Scarlett. Group testing: An information theory perspective. *Foundations and Trends® in Communications and Information Theory*, 15(3-4):196–392, 2019.

[CT06] E. J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, December 2006.

[DH99] D.-Z. Du and F. K. Hwang. *Combinatorial Group Testing and Its Applications*. WORLD SCIENTIFIC, December 1999.

[DH06] D.-Z. Du and F. K. Hwang. *Pooling Designs and Nonadaptive Group Testing*. WORLD SCIENTIFIC, June 2006.

[Dor43] R. Dorfman. The detection of defective members of large populations. *The Annals of Mathematical Statistics*, 14(4):436–440, December 1943.

[FFLH] B. I. Fargion, D. Fargion, P. G. De Sanctis Lucentini, and E. Habib. Purim: a rapid method with reduced cost for massive detection of covid-19. arXiv:2003.11975 [q-bio.PE].

[FH13] S. Foucart and Rauhut H. *A Mathematical Introduction to Compressive Sensing*. Springer New York, 2013.

[HZW+20] G. Huang, W. Zeng, W. Wang, Y. Song, X. Mo, J. Li, P. Wu, R. Wang, F. Zhou, J. Wu, B. Yi, Z. Xiong, L. Zhou, F. Wang, Y. Tian, W. Hu, X. Xu, R. Zhai, K. Yuan, X. Li, X. Qiu, J. Qiu, and A. Wang. Triaging patients in the outbreak of the 2019 novel coronavirus. March 2020.

[SBY] A. F. Siegenfeld and Y. Bar-Yam. Eliminating covid-19: A community-based analysis. arXiv:2003.10086[q-bio.PE].

[SSW+16] D. Sint, M. Sporleder, C. Wallinger, O. Zegarra, J. Oehm, N. Dangi, Y. P. Giri, J. Kroschel, and M. Traugott. A two-dimensional pooling approach towards efficient detection of parasitoid and pathogen DNA at low infestation rates. *Methods in Ecology and Evolution*, 7(12):1548–1557, August 2016.

[TM06] N. Thierry-Mieg. *BMC Bioinformatics*, 7(1):28, 2006.

[TTY+20] K. Kai-Wang To, O. Tak-Yin Tsang, C. Chik-Yan Yip, K.H. Chan, T.V. Wu, J. Man-Chun Chan, W.S. Leung, T. Shiu-Hong Chik, C. Yau-Chung Choi, D.H. Kandamby, D.C. Lung, A.R. Tam, R. Wing-Shan Poon, A. Yim-Fong Fung, I. Fan-Ngai Hung, V. Chi-Chung Cheng, J. Fuk-Woo Chan, and

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.06.028431; this version posted April 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

K.Y. Yuen. Consistent detection of 2019 novel coronavirus in saliva. *Clinical Infectious Diseases*, February 2020.

[YAST+20] I. Yelin, N. Aharony, E. Shaer-Tamar, A. Argoetti, E. Messer, D. Berenbaum, E. Shafran, A. Kuzli, N. Gandali, T. Hashimshony, Y. Mandel-Gutfreund, M. Halberthal, Y. Geffen, M. Szwarcwort-Cohen, and R. Kishony. Evaluation of COVID-19 RT-qPCR test in multi-sample pools. March 2020.

[ZDF+14] P.C. Zuzarte, R.E. Denroche, G. Fehringer, H. Katzov-Eckert, R.J. Hung, and J.D. McPherson. A two-dimensional pooling strategy for rare variant detection on next-generation sequencing platforms. *PLoS ONE*, 9(4):e93455, April 2014.

M. Täufer, School of Mathematical Sciences, Queen Mary University of London

Email address: m.taeufer@qmul.ac.uk