scientific reports

Check for updates

OPEN

Identification-detection group testing protocols for COVID-19 at high prevalence

Marco Chiani[1]✉, Gianluigi Liva[2] & Enrico Paolini[1]

Group testing allows saving chemical reagents, analysis time, and costs, by testing pools of samples instead of individual samples. We introduce a class of group testing protocols with small dilution, suited to operate even at high prevalence (5–10%), and maximizing the fraction of samples classified positive/negative within the first round of tests. Precisely, if the tested group has exactly one positive sample then the protocols identify it without further individual tests. The protocols also detect the presence of two or more positives in the group, in which case a second round could be applied to identify the positive individuals. With a prevalence of 5% and maximum dilution 6, with 100 tests we classify 242 individuals, 92% of them in one round and 8% requiring a second individual test. In comparison, the Dorfman's scheme can test 229 individuals with 100 tests, with a second round for 18.5% of the individuals.

We consider those situations where it is necessary to check if some individuals are positive with respect to a given disease. With a direct approach, samples taken from the individuals can be tested one by one, with a number of tests equal to the number of individuals under test. In many cases, however, it is possible to pool samples taken from different individuals and test the pool: if the pool is negative then all the corresponding individuals are declared as negative, while if the pool is positive it means that at least one is positive. Several group testing (GT) techniques based on pooling to reduce the number of tests have been proposed, starting from the work by Dorfman[1]. When the disease prevalence is not too large, this brings considerable savings in terms of tests and therefore chemical reagents, analysis time, effort, and costs. Recently, due also to the cost of sophisticated tests like those based on polymerase chain reaction or transcription mediated amplification, the use of group testing has been advocated to enable mass screening in the context of the SARS-CoV-2 pandemic, with experimental campaigns implemented in a few countries[2]. In *adaptive* group testing, the tests are performed in sequence, with pools that are created based on the outcomes of the previous tests[3,4]. On the contrary, in non-adaptive group testing all pools are a-priori set, and tests are carried out in parallel. Both approaches have advantages and shortcomings: adaptive strategies can identify the status of individuals with fewer tests. Nevertheless, considering the time required to carry out each test, a pure adaptive strategy may require an excessive amount of time. Non-adaptive schemes require typically more tests to succeed, but they are faster as tests can be performed in parallel. To combine the advantages of both techniques, while mitigating their limitations, it is sometimes preferable to implement a hybrid approach, where a first screening is performed via a non-adaptive testing step, followed by an adaptive (or even individual) one for the population members that are identified as potentially infected. Approaches of this kind, which date back to the original work of Dorfman[1], enable remarkable savings in the number of tests. Several current investigations on the use of group testing for SARS-CoV-2 screening follow this line[5–12]. In particular, in the context of group testing for SARS-CoV-2, the simple Dorfman approach has been validated by verifying the sensibility of polymerase chain reaction tests with respect to the size n of the pools[6,7]. Non-adaptive protocols relying on Reed-Solomon error correcting codes to design the pools have been used to target low infection rate regime (e.g., prevalence below 1.3%)[8]. Bayesian approaches to identify the set of infected samples in a non-adaptive group testing approach have also been addressed[9], as well as schemes that exploit a quantitative knowledge on the viral load in the pools[10,11]. Other approaches to group testing in the low prevalence regime exploit a geometrical construction of the pools in a non-adaptive setting[12].

Differently from previous works, we here are not limited to low prevalence. Specifically, in this paper we describe a new class of protocols for group testing where the main objective is to maximize the probability that classification of samples is completed within the first round of tests. If the tested group has exactly one positive sample, then the protocols detect that there is only one positive and identify it without the need for a second

[1]DEI, University of Bologna, Bologna, Italy. [2]Institute of Communications and Navigation, German Aerospace Center (DLR), Wessling, Germany. ✉email: marco.chiani@unibo.it

round of individual testing. The protocols also detect the presence (without identification) of two or more positives in the group, in which case a second round must be applied to identify the positive individuals. These protocols are thus analogous to error control codes able to correct one error and detect two or more errors occurring in a group[13]. Due to this capability to directly identify one positive in the first round, this work is specially suited for high prevalence (5–10%) scenarios, differently from other methods which address the low prevalence case[2,6–8,12]. Also, due to the problem of dilution which can lead to false negatives, this work is particularly focused on those pool sizes which can be realistically used in a diagnostic laboratory[2,5,7,14–16]. We present next the main results of the investigation, based on the probabilistic analysis detailed at the end of the paper.

Results

In the following we will refer to polymerase chain reaction (PCR) for the test, but the procedure is general for any possible test. With "prevalence" we will indicate the probability that an individual is positive. The direct approach to testing consists of performing individual tests, with one PCR for each individual sample, to determine if it is positive or negative. The number of tests in this case equals the number of samples to classify.

In (GT), individual samples are grouped (pooling), and the pools are tested: if a pool is negative it is assumed that all individuals participating to that pool are negative. Thus, the number of PCR tests can be reduced with respect to individual testing, if the prevalence is not too high. The saving is more marked for low prevalence. In this paper we will refer to GT with a first round of pooled tests, possibly followed by a second round of some (hopefully few) individual tests to complete the classification. While the advantage is clear, it must be considered that implementing GT imposes a reorganization of the testing process, whose impact should not be underestimated. In fact, with GT a phase of preparation of the pools is necessary. This phase should be automated to avoid errors in the processing: this is already possible, as machines currently available in many diagnostic laboratories can be suitably reprogrammed for pooling. Also, while individual testing ends in a single round of PCR, in the case of GT it is sometimes necessary to carry out a second round of PCRs for some individuals (thus requiring additional time). If the number of samples to retest is large, managing the second round, where individual samples needing an individual PCR must be reexamined, should be automated to avoid errors and contamination. When full automation of the process is not available, it would be preferred to adopt GT schemes with a low fraction of samples needing an individual retest. Also, it must be remarked that large pool sizes can lead to a dilution of the viral load affecting the sensitivity of the test, therefore causing false negatives. For this reason, we will concentrate on schemes with limited pool sizes, which justifies our assumption that the false negative rate is negligible.

The baseline protocol is that originally proposed by Dorfman in 1943, where: individuals are grouped into groups of n; one pool is used to analyze all n individuals (dilution n); the mother tubes of the n individuals are set aside; the single pool is tested. If the pool is negative, all n individuals are declared negative, and no other tests are needed. If, on the contrary, the pool is positive, it is necessary to carry out a second round of individual tests on all n individuals[1].

Identification-detection for group testing: the *Pnp* protocols. We propose a new class of pooling schemes with small dilution, for high prevalence testing scenarios. Assume a group test employing p pools $P_1, P_2, \ldots, P_p$ to test a group of $n > p$ individuals $I_1, I_2, \ldots, I_n$. The pooling can be described by a test matrix, where each row is a pool and each column is an individual. The matrix elements are 0 or 1, where a 1 in row i and column j indicates that individual I_j participates in pool P_i.

For identification-detection we propose to use test matrices composed by columns all with a fixed number c of 1s, so that each individual sample is copied into exactly c pools. With this choice, there would be c positive pools if and only if the group has exactly one positive individual. A number of positive pools larger than c indicates that there are two or more positive individuals in the group. The largest group size n for a given number of copies c and pools p is

$$n = \binom{p}{c}. \tag{1}$$

We will assume always the largest n, as for GT the objective is to test the largest possible number of individuals for a given p. The pooling matrix columns are thus all possible vectors with $p - c$ elements to 0 and c elements to 1. The number of individuals per pool (dilution) is indicated as d. It can be checked that the dilution for p pools and n individuals, each participating in c pools, is

$$d = \binom{p-1}{c-1}. \tag{2}$$

Hence, by testing the p pools (first round), the scheme allows to classify immediately the cases of zero positives per group or one positive per group. Therefore, for up to one positive per group there is no need for individual tests. The scheme also detects the presence of two or more positives per group, in which case a second round of individual tests is required.

To keep the dilution as small as possible, we investigate in particular the case $c = 2$, where each sample is copied in two pools. With this choice, from (2) the dilution is $d = p - 1$. We now explicit the test matrices for dilution up to $d = 6$.

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

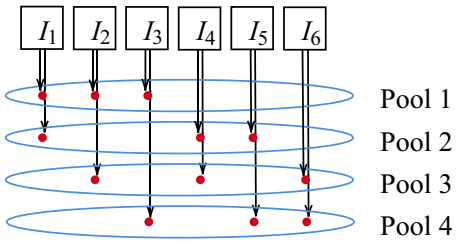

Figure 1. Pooling matrix and its interpretation for the *P*64 protocol. $I_1, ..., I_6$: individuals.

Pools result	Positive individual	Further test
0 0 0 0	None	No
1 1 0 0	I_1	No
1 0 1 0	I_2	No
1 0 0 1	I_3	No
0 1 1 0	I_4	No
0 1 0 1	I_5	No
0 0 1 1	I_6	No
0 1 1 1		I_4, I_5, I_6
1 0 1 1		I_2, I_3, I_6
1 1 0 1		I_1, I_3, I_5
1 1 1 0		I_1, I_2, I_4
1 1 1 1		$I_1, I_2, I_3, I_4, I_5, I_6$

Table 1. Decision rule for *P*64 (1 = positive pool, 0 = negative pool).

Protocol *P*64. From Eq. (1), the smallest group size for which $n > p$ is obtained with $p = 4$ pools, each individual participating in $c = 2$ pools, and therefore with a number of individuals per group $n = 6$. The test matrix for *P*64 is reported in Fig. 1.

Each pool contains the samples from exactly three individuals, so the dilution is $d = 3$, as given by Eq. (2). The protocol is described as follows: individuals are arranged into groups of $n = 6$ (indicated in the figure as $I_1, ..., I_6$); $p = 4$ pools are used to analyze the 6 individuals; each individual participates in $c = 2$ pools according to the scheme in the figure, with exactly 3 individuals in each pool; the mother tubes of the 6 individuals are set aside; the four pools are tested (e.g., by PCR).

Based on the results of the 4 tests, the following cases may arise (see Table 1):

- All 4 pools are negative: in this case all six individuals are declared as negative. No other tests are needed.
- Exactly 2 out of the 4 pools are negative: in this case only one individual is positive, uniquely identified according to the decoding table. No other tests are needed.
- One pool is negative and the other 3 are positive: a second round of individual tests is required for three individuals according to the scheme of Table 1 (or, to simplify, individual test on all six individuals).
- All 4 pools are positive: a second round of individual tests is required for all six individuals.

Protocol *P*105. With $p = 5$ pools and $c = 2$ we have groups of $n = 10$ individuals, pooled according to the test matrix

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

In this case the dilution is $d = 4$. This matrix identifies one positive and detects two or more positives per group of $n = 10$ individuals.

Protocol *P*156. With $p = 6$ pools and $c = 2$ we have groups of $n = 15$ individuals, pooled according to the test matrix

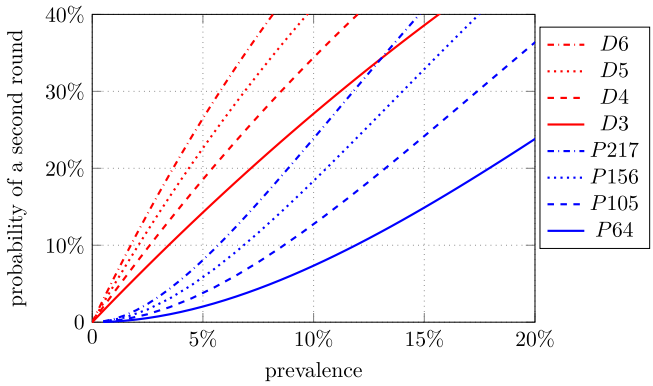

Figure 2. Probability of a second testing round for an individual, as a function of the prevalence. *Dn* = Dorfman with *n* individuals and one pool; *Pnp* = protocol P with *n* individuals and *p* pools.

$$\begin{pmatrix} 1&1&1&1&1&0&0&0&0&0&0&0&0&0&0 \\ 1&0&0&0&0&1&1&1&1&0&0&0&0&0&0 \\ 0&1&0&0&0&1&0&0&0&1&1&1&0&0&0 \\ 0&0&1&0&0&0&1&0&0&1&0&0&1&1&0 \\ 0&0&0&1&0&0&0&1&0&0&1&0&1&0&1 \\ 0&0&0&0&1&0&0&0&1&0&0&1&0&1&1 \end{pmatrix}.$$

In this case the dilution is $d = 5$. This matrix identifies one positive and detects two or more positives per group of $n = 15$ individuals.

Protocol *P217*. With $p = 7$ pools and $c = 2$ we have groups of $n = 21$ individuals, pooled according to the test matrix

$$\begin{pmatrix} 1&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 1&0&0&0&0&0&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&1&0&0&0&0&1&1&1&1&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&1&0&0&0&1&0&0&0&1&1&1&0&0&0 \\ 0&0&0&1&0&0&0&0&1&0&0&0&1&0&0&1&0&0&1&1&0 \\ 0&0&0&0&1&0&0&0&0&1&0&0&0&1&0&0&1&0&1&0&1 \\ 0&0&0&0&0&1&0&0&0&0&1&0&0&0&1&0&0&1&0&1&1 \end{pmatrix}.$$

In this case the dilution is $d = 6$. This matrix identifies one positive and detects two or more positives per group of $n = 21$ individuals.

Other protocols, even for different values of c, can be similarly designed. For all protocols, the decoding rule can be reformulated succinctly as follows: in the first round, classify as negative all individuals participating in a negative pool, and test individually the others.

Performance. We present the performance of our protocol compared with the Dorfman's protocol, for dilutions ranging from $d = 3$ up to $d = 6$. Considering both the first round of test on p pools and the occasional second round on some or all individuals, for a generic protocol we define the performance in terms of:

- efficiency, quantified by the average number of individuals tested with 100 PCR tests;
- probability that an individual is tested in a second round, indicated as P_{SR}.

We assume a prevalence ϵ and independent positivity from individual to individual. To calculate efficiency, let us denote with T_g the number of tests needed to identify all positives in the group. By indicating with $\mathbb{E}\{\}$ the statistical expectation, the average number of tests per individual is $\mathbb{E}\{T_g\}/n$. Therefore, on the average, with 100 PCRs we classify a number of individuals equal to:

$$100\,\frac{n}{\mathbb{E}\{T_g\}} \quad \text{[individuals classified with 100 tests]}. \tag{3}$$

The statistical characterization of T_g is provided in the Methods section, and leads to Eq. (9). About P_{SR}, we observe that a second round is needed if the number of positive pools, indicated as Y, is greater than 2. The statistic of Y, provided in the Methods section, leads to Eq. (8).

Results as functions of the prevalence are shown in Figs. 2 and 3, where the probability of a second testing for an individual, given by Eqs. (6) and (8), and the efficiency, given by Eqs. (3) and (9), are reported. For example, with *P*64 we find that, with a prevalence $\epsilon = 5\%$, about 146 individuals are classified with 100 PCR tests. Of all individuals, 98% are classified in the first round, and only 2% need a second round.

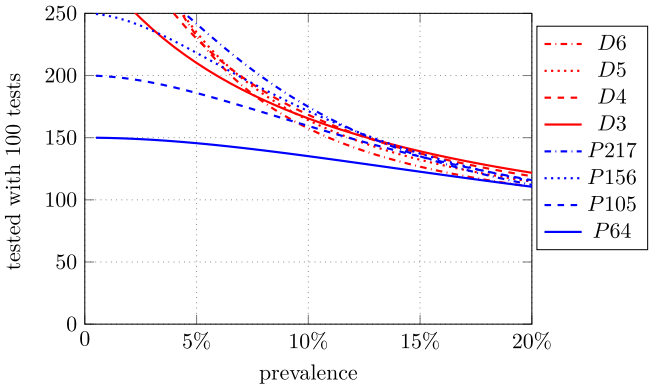

Figure 3. Average number of individuals tested with 100 tests as a function of the prevalence. *Dn* = Dorfman with *n* individuals and one pool; *Pnp* = protocol P with *n* individuals and *p* pools.

	Prevalence 5%		Prevalence 10%	
	tested with 100 tests	retested rate	tested with 100 tests	retested rate
*P*64	146	2%	135	7.3%
*P*105	186	3.7%	159	12.7%
*P*156	218	5.8%	171	18.4%
*P*217	242	8%	175	23.9%
*D*3	210	14.2%	165	27.1%
*D*4	229	18.5%	168	34.4%
*D*5	235	22.6%	164	40.9%
*D*6	232	26.5%	157	46.8%

Table 2. Performance of the analyzed protocols, dilutions $d = 3, \ldots, 6$.

For the Dorfman's scheme the results, shown in Figs. 2 and 3, have been derived by using Eqs. (10) and (11). The exact numbers are shown for prevalence of 5% and 10% in Table 2. For example, with a pool of $n = 4$ individuals and a prevalence $\epsilon = 5\%$, an average of 229 individuals are tested with 100 PCR tests with Dorfman's scheme. However, about 18.5% of all individuals need a second round of individual tests.

Discussion

We have investigated group testing consisting of a first round of pooled tests, followed by individual testing, applied to a population with high prevalence, to save resources.

Two main issues must be discussed for a practical usage of group testing. First, one should consider that the maximum pool size is limited due to the dilution of the sample viral load and the consequent problem of false negatives. For COVID-19, a conservative current estimation suggests that dilutions in the order of 5–8 would still allow a negligible false negative rate, although higher dilutions have been investigated, with some conflicting reports[2,5,7,14–16]. Second, for group testing the diagnostic laboratory must be organized to handle the whole process (pooling, first round of tests, reopening and second round of tests). Automation systems and robots, currently available in many diagnostic laboratories, can be suitably reprogrammed for pooling. The main issue is related to the management of the second rounds of tests. If the fraction of samples to retest is large, picking back the original samples of some individual to be reexamined should be automated, to avoid errors and contamination. When the process is not fully automated it could be necessary to use protocols able to complete the positive identification mostly within the first round of tests, with a small rate of individuals to be retested. In fact, with low rates of second rounds it may be possible to handle the retesting process even manually, thus simplifying the organization of a diagnostic laboratory. Reducing the second round tests will also have the advantage of giving a faster classification.

Limiting the discussion to dilutions up to 6, we have found that in the range of prevalence 5–10% the best choice is represented by the identification-detection scheme *P*217, which outperforms all the others in terms of efficiency while still having a low rate of second round tests. For example, at $\epsilon = 5\%$ it allows to classify 242 individuals with 100 tests, with a rate of second round individual tests of about 8%. The scheme also performs better than the Dorfman's scheme both in terms of efficiency and rate of second round individual tests. Compared with the proposal, the Dorfman's schemes are in fact less efficient and have much larger rates of individuals tested twice (one time in group, then individually). They are therefore not suitable at high prevalence. Identification-detection schemes with dilutions $3 - 5$ give less advantages in terms of number of tests, but offer a smaller rate of second

$x \diagdown y$	3	4
2	12	3
3	4	16
4	0	15
5	0	6
6	0	1

Table 3. $a(x, y)$ for $P64$.

round tests. The scheme $P64$, with dilution 3, is the one with the smallest rate of retested individuals, having a rate of individual retest of 0.088% at $\epsilon = 1\%$, of 2% at $\epsilon = 5\%$, and of 7% at $\epsilon = 10\%$. The choice of the specific protocol imposes therefore, for a given maximum dilution, a trade-off among efficiency and second round rates.

Having to handle few second round tests, all new protocols seems suitable even for non automated laboratories at prevalence up to 5%, with few percent of the individuals needing a second round. At prevalence 10% the only protocols with small probability of second round are $P105$ and $P64$, with respectively 12.7% and 7% of the individuals which need retesting.

We derived also analytical expressions for the performance of the new protocols, allowing the design of identification-detection group testing pooling schemes for arbitrary dilutions and for the targeted prevalence rates.

Methods

Performance of the *Pnp* protocols. In this section we derive expressions for the performance of the proposed protocols assuming $c = 2$, which is the most effective value of c to limit dilution. The analysis can however be generalized to other values of c.

Let us denote as X the number of positive individuals in a group of n individuals, and as Y the number of positive pools out of the p pools. For an identification-detection protocol able to identify one single positive in a group of n and detect two or more positives, the probability that the group must be reopened for a second round is

$$P_{GR} = \Pr\{X \geq 2\} = \sum_{x=2}^{n} \binom{n}{x} \epsilon^x (1-\epsilon)^{n-x}. \tag{4}$$

The average number of tests per group can be bounded by assuming that the second round is taken on all n individuals

$$\mathbb{E}\{T_g\} \leq p + n \sum_{x=2}^{n} \binom{n}{x} \epsilon^x (1-\epsilon)^{n-x}. \tag{5}$$

To derive a precise analysis we must consider that the second round occurs on subsets of the group, depending on the number of positive pools. To this aim, we observe that the probability that y pools are positive is

$$\Pr\{Y = y\} = \sum_{x=0}^{n} \Pr\{Y = y | X = x\} \Pr\{X = x\} = \sum_{x=0}^{n} a(x,y) \epsilon^x (1-\epsilon)^{n-x} \tag{6}$$

where $a(x, y)$ is the number of group configurations with x positive individuals and y positive pools, so that, for example, it is $a(1, 2) = n$. The values of $a(x, y)$ for arbitrary x, y can be derived by combinatorial analysis. Specifically, we prove at the end of the paper that $a(x, y)$ is given by the recursion

$$a(x,y) = \binom{p}{y} \binom{(y-1)y/2}{x} - \sum_{\ell=1}^{y-1} a(x,\ell) \binom{p-\ell}{p-y}. \tag{7}$$

Values of $a(x, y)$ needed to evaluate $\Pr\{Y = y\}$ are those for $y > 2$, which are reported for some protocols of interest in Tables 3, 4, 5 and 6.

Then, we observe that if there are $Y = y$ positive pools the number of individuals to retest in a second round is $\binom{y}{2}$. Therefore, the probability that an individual needs a second round is

$$P_{SR} = \frac{1}{n} \sum_{y=3}^{p} \binom{y}{2} \Pr\{Y = y\}. \tag{8}$$

The exact average number of tests needed to classify all n individuals in a group is then

$$\mathbb{E}\{T_g\} = p + \sum_{y=3}^{p} \binom{y}{2} \Pr\{Y = y\} = p + nP_{SR} \tag{9}$$

and the number of individuals tested with 100 tests, defined by (3), is rewritten as $100/(p/n + P_{SR})$.

$x \diagdown y$	3	4	5
2	30	15	0
3	10	80	30
4	0	75	135
5	0	30	222
6	0	5	205
7	0	0	120
8	0	0	45
9	0	0	10
10	0	0	1

Table 4. $a(x, y)$ for $P105$.

$x \diagdown y$	3	4	5	6
2	60	45	0	0
3	20	240	180	15
4	0	225	810	330
5	0	90	1332	1581
6	0	15	1230	3760
7	0	0	720	5715
8	0	0	270	6165
9	0	0	60	4945
10	0	0	6	2997
11	0	0	0	1365
12	0	0	0	455
13	0	0	0	105
14	0	0	0	15
15	0	0	0	1

Table 5. $a(x, y)$ for $P156$.

$x \diagdown y$	3	4	5	6	7
2	105	105	0	0	0
3	35	560	630	105	0
4	0	525	2835	2310	315
5	0	210	4662	11067	4410
6	0	35	4305	26320	23604
7	0	0	2520	40005	73755
8	0	0	945	43155	159390
9	0	0	210	34615	259105
10	0	0	21	20979	331716
11	0	0	0	9555	343161
12	0	0	0	3185	290745
13	0	0	0	735	202755
14	0	0	0	105	116175
15	0	0	0	7	54257
16	0	0	0	0	20349
17	0	0	0	0	5985
18	0	0	0	0	1330
19	0	0	0	0	210
20	0	0	0	0	21
21	0	0	0	0	1

Table 6. $a(x, y)$ for $P217$.

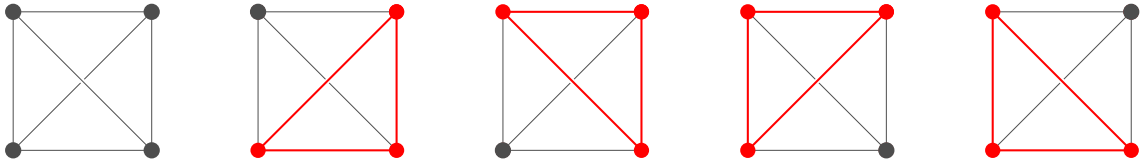

Figure 4. Left graph: the *P*64 matrix with four pools (vertices) and six individuals (edges). The four graphs on the right represent the cases with three positive individuals (red edges) producing three positive pools (red vertices).

Performance of the Dorfman's scheme. For completeness, we review also the performance for the Dorfman's protocol[1]. In the same hypothesis above, the probability that a second round is needed for the Dorfman's scheme (in this case all individuals have to be retested) is

$$P_{SR} = \sum_{x=1}^{n} \binom{n}{x} \epsilon^x (1-\epsilon)^{n-x} = 1 - (1-\epsilon)^n \tag{10}$$

and the average number of tests per group is

$$\mathbb{E}\{T_g\} = 1 + n\, P_{SR}. \tag{11}$$

The number of individuals tested with 100 tests is then $100/(1/n + P_{SR})$.

Recursive computation of $a(x, y)$. When $c = 2$, the pooling matrix is amenable of a simple graphical description. In particular, a pooling matrix with p rows and n columns can be represented as a graph G with p vertices, each one associated with a matrix row (equivalent, with a pool), and n edges, each one associated with a matrix column (equivalently, with an individual). An edge connects two vertices if and only if the individual corresponding to the edge participates in the two pools corresponding to the vertices. An example is provided in Fig. 4 for the *P*64 pooling matrix.

The value of $a(x, y)$ equals the number of sub-graphs of G having having y vertices (with nonzero degree) and x edges. Since the pooling matrix columns are *all* length-p binary vectors with two 1s, we can focus on a specific subset S of vertices with cardinality y and search for the number of sub-graphs of G with y vertices (with nonzero degree) and x edges, all vertices belonging to S. This number is denoted by $\tilde{a}(x, y)$ and is related to $a(x, y)$ by

$$a(x, y) = \binom{p}{y} \tilde{a}(x, y). \tag{12}$$

The value of $\tilde{a}(x, y)$ equals the number of ways in which, given the set S of y vertices, we can place x edges in such a way that each vertex is connected to at least one edge. This is equal to the total number of ways in which the x edges can be placed, $y(y-1)/2$, minus the number of edge configurations in which only ℓ nodes are "touched", for $\ell \in \{1, 2, \ldots, y-1\}$. This yields

$$\tilde{a}(x, y) = \binom{y(y-1)/2}{x} - \sum_{\ell=1}^{y-1} \tilde{a}(x, \ell) \binom{y}{\ell} \tag{13}$$

which in particular gives $\tilde{a}(x, 2) = 1$ if $x = 1$ and $\tilde{a}(x, 2) = 0$ otherwise. Incorporating Eq. (13) into Eq. (12) gives, after some simplifications, Eq. (7).

Received: 10 April 2021; Accepted: 2 February 2022
Published online: 28 February 2022

References

1. Dorfman, R. The detection of defective members of large populations. *Ann. Math. Stat.* **14**, 436–440 (1943).
2. Mallapaty, S. The mathematical strategy that could transform coronavirus testing. *Nature* **583**, 504–505 (2020).
3. Aldridge, M., Johnson, O. & Scarlett, J. Group testing: An information theory perspective. in *Foundations and Trends in Communications and Information Theory*. (2019).
4. Sobel, M. & Groll, P. A. Group testing to eliminate efficiently all defectives in a binomial sample. *Bell Labs Tech. J.* **38**, 1179–1252 (1959).
5. Vandenberg, O., Martiny, D., Rochas, O., van Belkum, A. & Kozlakidis, Z. Considerations for diagnostic COVID-19 tests. *Nat. Rev. Microbiol.* **19**, 171–183 (2021).
6. Yelin, I. *et al.* Evaluation of COVID-19 RT-qPCR test in multi sample pools. *Clin. Infect. Dis.* **71**, 2073–2078 (2020).
7. Hirotsu, Y. *et al.* Pooling RT-qPCR testing for SARS-CoV-2 in 1000 individuals of healthy and infection-suspected patients. *Sci. Rep.* **10**, 18899 (2020).
8. Shental, N. *et al.* Efficient high-throughput SARS-CoV-2 testing to detect asymptomatic carriers. *Sci. Adv.* **6** (2020).
9. Liva, G., Paolini, E. & Chiani, M. Optimum detection of defective elements in non-adaptive group testing. in *Proceedings of the 55th Annual Conference on Information Science and Systems (CISS)*. (2021).
10. Ghosh, S. *et al.* Tapestry: A single-round smart pooling technique for COVID-19 testing. *medRxiv* (2020).

11. Heidarzadeh, A. & Narayanan, K. Two-stage adaptive pooling with RT-QPCR for Covid-19 screening. in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2021).
12. Mutesa, L. *et al.* A pooled testing strategy for identifying SARS-CoV-2 at low prevalence. *Nature* **589**, 276–280 (2021).
13. Ryan, W. & Lin, S. *Channel Codes - Classical and Modern* (Cambridge University Press, New York, 2009).
14. Ben-Ami, R. *et al.* Large-scale implementation of pooled RNA extraction and RT-PCR for SARS-CoV-2 detection. *Clin. Microbiol. Infect.* **26**, 1248–1253 (2020).
15. Abid, S. *et al.* Assessment of sample pooling for SARS-CoV-2 molecular testing for screening of asymptomatic persons in Tunisia. *Diagn. Microbiol. Infect. Dis.* **98**, 113 (2020).
16. Barak, N. *et al.* Lessons from applied large-scale pooling of 133,816 SARS-CoV-2 RT-PCR tests. *Sci. Transl. Med.* (2021).

Acknowledgements
The authors would like to thank Prof. Vittorio Sambri for discussions and comments about COVID-19 diagnostic laboratory activities. This work was supported in part by Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) under the program "Departments of Excellence (2018-2022) - Precise-CPS".

Author contributions
The authors contributed equally to this work.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to M.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.