PROPOSAL FOR A METHOD TO OPTIMISE USE OF SARS-COV2 TESTS (COVID-19)

authors
Alberto Troisi (albmail@gmail.com)
Vincenzo Gervasi PhD (vincent.gervasi@gmail.com)

Version 5. Updates to this document will be uploaded to:
https://github.com/albmail/covid_pool_test_method

The proposed method makes it possible to reduce on average the number of molecular tests needed to diagnose the presence or absence of the virus in a group of individuals to be controlled.
The method is based on the fact that current epidemiological data show that tests carried out with negative results are prevalent over tests with positive results.
Therefore, by mixing the biological material of several individuals and testing on this mixed material, in case of a negative result, it is possible to exclude positivity on all these individuals by performing a single molecular test.

The main purpose of this paper is to show that if the tests carried out in a laboratory are positive for less than 30% of the cases analyzed then with the method described it is possible to obtain more diagnoses with the same number of molecular tests.

OPERATING PROCEDURE

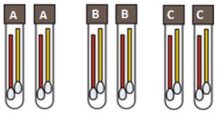
The procedure indicated serves only to illustrate the method for which it does not take account of any additional information required by national and international organisations.

The method:

1. Swab and collect biological material from 3 persons (individuals A,B,C)
   Two samples shall be taken for each person in order to have 2 separate samples for each individual under observation

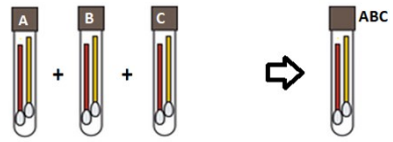
   

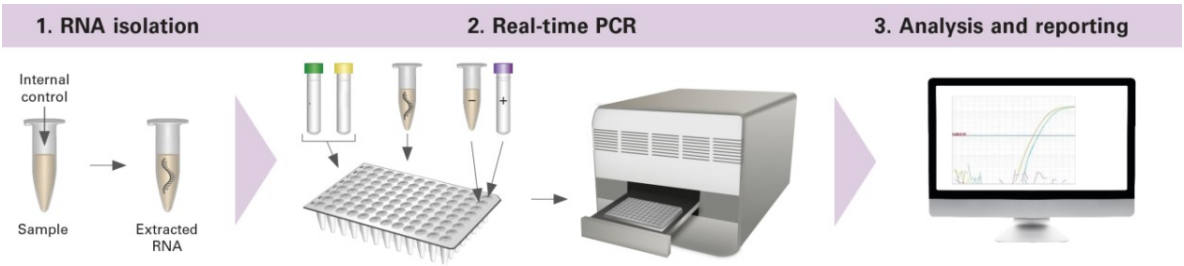
   We will then have collected for the group of 3 people 6 samples (2 for each individual)

2. **We prepare a mixed sample** by mixing the biological materials of the 3 individuals in the group (one of the 2 samples taken for each individual will be used for this purpose).

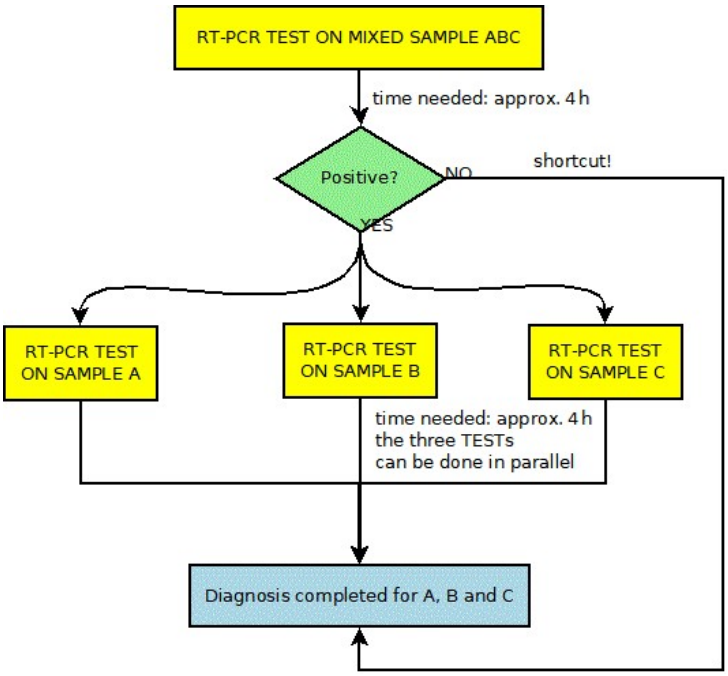
We then proceed to the molecular test on this mixed sample





If the test is negative we can conclude that there is no RNA of the virus in each of the 3 mixed samples and therefore all 3 individuals are not positive for the test. In this case, no further testing will be necessary.

3. If the test is positive we can conclude that in the group of 3 individuals there is at least one positive individual and therefore we must investigate further by testing all 3 separate samples kept aside for this purpose.



HOW MUCH DO YOU SAVE?

In order to answer this question, we need to analyse the probability that all three individuals tested are negative for the test. From now on I will use the following notation:

P(-) : the probability that a single test will result negative

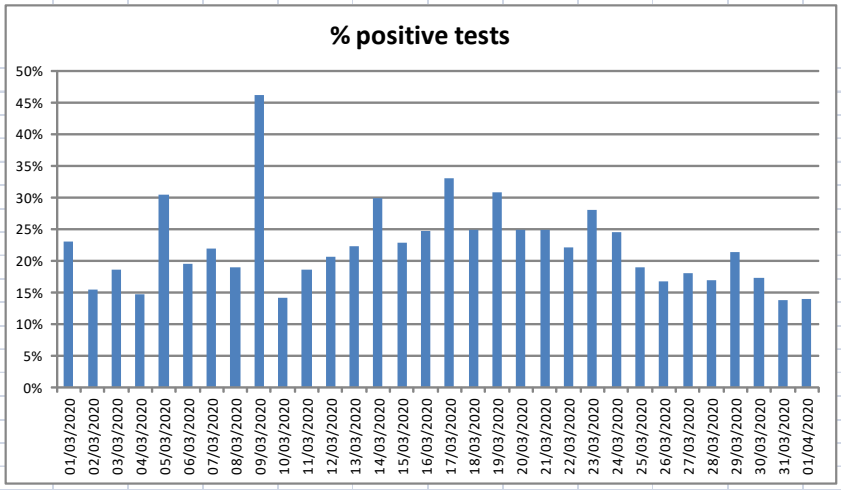
P(+) : the probability that a single test has a positive result

P(3-) : the probability that 3 tests will result negative, which is equivalent to the probability that a single test carried out on a mixture of 3 samples will result negative

Assuming that there is no correlation between the individuals analyzed we then have
$P(3-) = P(-)^3$

Let's take the Italian epidemiological data of the month of March

date	swabs	new positive cases	% positive tests
01/03/2020	2.466	566	23%
02/03/2020	2.218	342	15%
03/03/2020	2.511	466	19%
04/03/2020	3.981	587	15%
05/03/2020	2.525	769	30%
06/03/2020	3.997	778	19%
07/03/2020	5.703	1.247	22%
08/03/2020	7.875	1.492	19%
09/03/2020	3.889	1.797	46%
10/03/2020	6.935	977	14%
11/03/2020	12.393	2.313	19%
12/03/2020	12.857	2.651	21%
13/03/2020	11.477	2.547	22%
14/03/2020	11.682	3.497	30%
15/03/2020	15.729	3.590	23%
16/03/2020	13.063	3.233	25%
17/03/2020	10.695	3.526	33%
18/03/2020	16.884	4.207	25%
19/03/2020	17.236	5.322	31%
20/03/2020	24.109	5.986	25%
21/03/2020	26.336	6.557	25%
22/03/2020	25.180	5.560	22%
23/03/2020	17.066	4.789	28%
24/03/2020	21.496	5.249	24%
25/03/2020	27.481	5.210	19%
26/03/2020	36.615	6.153	17%
27/03/2020	33.019	5.959	18%
28/03/2020	35.447	5.974	17%
29/03/2020	24.504	5.217	21%
30/03/2020	23.329	4.050	17%
31/03/2020	29.609	4.053	14%
01/04/2020	34.455	4.782	14%
TOT	522.762	109.446	



As you can see the probability of a positive test result has varied daily from a minimum of 14% to a maximum of 46% with an average of 22%

We are going to use the average probability of 22% to do our analysis.
$P(+) = 0,22$ and so $P(-) = 1-P(+) = 0,78$
$P(3-) = P(-)^3 = 0,47$
So we'll have about 47% chance of finding the mixed sample negative for the test.

In case the mixed sample is negative, with a single test we can determine the negativity of all three subjects, otherwise to reach a conclusion will require 3 more tests (one for each sample) and then 4 tests in total.
The probability of completing the analysis with a single test coincides with the probability that the mixed sample is negative.
P(1 Test needed) = P(3-) = 0,47
The probability that 4 tests are necessary is complementary and therefore
P(4 Tests needed) = 1- P(1 Test needed) = 0,53
The average number of tests required to determine whether or not an individual is positive is therefore
[P(1 Test needed) + P(4 Tests needed)*4]/3 = 0,86

What does that mean?
It means that using this method you can know whether or not 100 individuals are positive by performing an average of 86 molecular tests instead of the 100 that would have been necessary normally. Considering that in the last week more than 20,000 individuals a day are being tested with this method you could save over 2,800 molecular tests every day.
And what would happen when the probability of detecting a positive case has dropped to 5% for example?
P(+) = 0,05 so P(-) = 1-P(+) = 0,95
$P(3-) = P(-)^3 = 0,86$
P(1 Test needed) = P(3-) = 0,86
P(4 Tests needed) = 1- P(1 Test needed) = 0,14
The average number of tests required to determine whether or not an individual is positive is therefore
[P(1 Test needed) + P(4 Tests needed)*4]/3 = 0,47
Assuming we're screening 20,000 individuals a day we'd have in this case:
a saving of about 10,600 molecular tests per day,
more than half the tests saved!
This method could contribute to the search for asymptomatic patients when the number of new symptomatic cases detected daily will have dropped significantly. When testing non symptomatic groups of people the probability P(+) would be very low and the lower the probability the more effective this method is.

HOW WOULD THE TIMING OF THE DIAGNOSIS CHANGE?

Considering that currently the duration of a molecular test (RT-PCR) is about 4 hours
According to the methodology described, the case in which only one test is sufficient would give an answer in 4 hours.
In the case of a positive result on the mixed sample we then proceed with the other 3 tests on the individual samples (and assuming that these can be done simultaneously) and then are needed 4 hours more: then 8 hours in total.

However, for a correct statistical analysis of the average time required for each diagnosis, it is necessary to take into account the fact that on average more diagnosis can be made for the same number of tests. This could then be further investigated and adapted to the specific operating conditions of the analysis laboratories.

CRITICALITY AND LIMITS

The most relevant aspect to consider is that in the mixed sample the density of the viral RNA to be detected could be lowered by a third. However, this should not significantly affect the RT-PCR analysis as this test has a high sensitivity and requires limited amounts of the starting sample.
Obviously, this should be further investigated by those who know the technical details of the test.
It should also be considered that in different areas and at different times of the epidemic the probability P(+) may be different and if the probability is too high this method may be counterproductive. In general the method has a convenience in terms of the number of tests when the average number of tests needed for an individual is less than 1, i.e.
(P(1 Test needed) + P(4 Tests needed)*4)/3 < 1
so
[P(3-)+(1-P(3-))*4]/3 < 1
$4/3-P(-)^3 < 1$

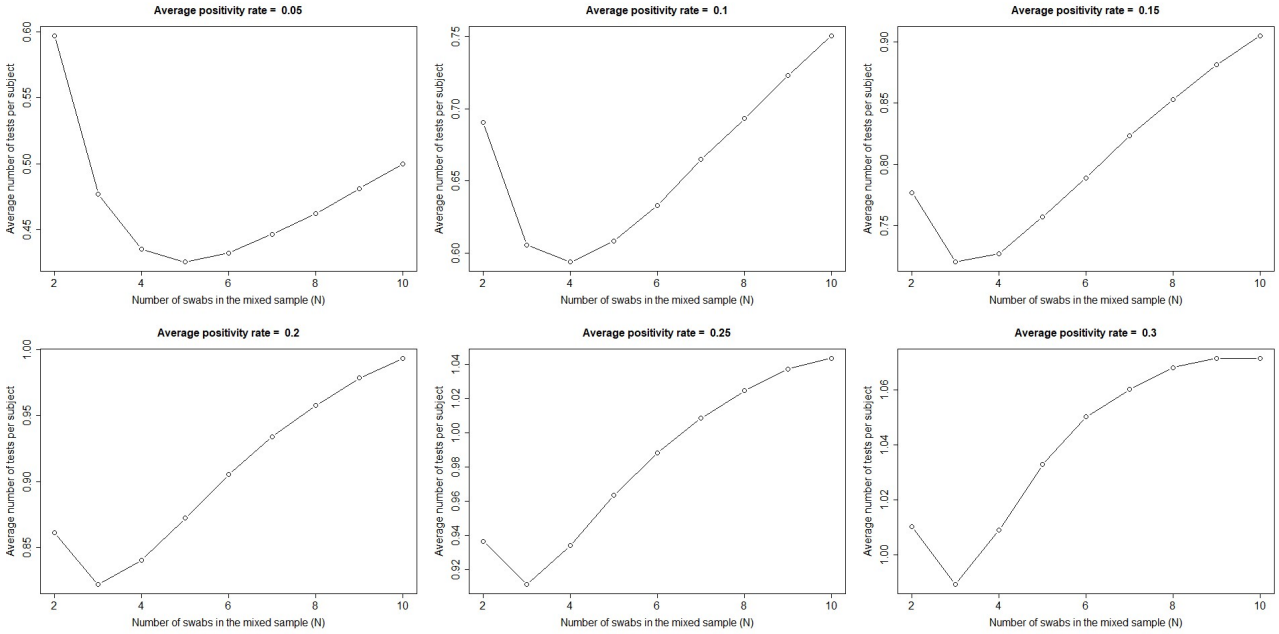
$P(-) > \sqrt[3]{\frac{1}{3}}$

the saving is therefore obtained when P(-) is greater than about 0.7 and therefore P(+) is less than 0.3. If the probability of having a positive test is therefore greater than 30%, it is no longer convenient to use the method.

WHAT IF WE MIX THE BIOLOGICAL MATERIAL OF MORE THAN THREE INDIVIDUALS FOR THE MIXED SAMPLE?

By creating a computational model we made simulations that show how the efficiency of the described method varies with the expected value of positivity P(+) but also with the number of individuals from which the mixed sample (N) is formed.

The results are synthesized in the following graphs:



There is a saving especially when the positivity rate of the test is low, and the optimal number of samples to combine changes in different cases

Positivity rate	Number of samples to mix	Saving %
0.05	5	60%
0.1	4	40%
0.15	3	30%
0.2	3	20%
0.25	3	10%
0.3	3	5%

It is intuitive that with even lower rates of positivity, even more samples could be pooled and the choice of protocol could be made according to epidemiological conditions. However, the study should be deepened also taking into account the sensitivity variation of the molecular test.

In the current situation, adapting the method to the different areas of origin of the samples (to which corresponds a different rate of positivity), it is possible to obtain a very significant increase in diagnosis, but to assess exactly how much this saving would correspond to, a more in-depth analysis is necessary.

An analytical study of the proposed method by varying the number of samples to be pooled can be developed from the equation [P(3-)+(1-P(3-))*4]/3 (indicating the average number of tests required per individual with the 3-sample method) and generalising it to a generic N number of samples.
The following formula is obtained:

$$\frac{P(-)^N + (1 - P(-)^N)(N+1)}{N} = 1 + \frac{1}{N} - P(-)^N$$

Where **P(-)** is the rate of negativity and **N** is the number of individuals from which the mixed sample is formed.
So the limit that determines whether the method is convenient or not is determined by the formula:

$$1 + \frac{1}{N} - P(-)^N < 1$$

$$\boldsymbol{P(-)^N > \frac{1}{N}}$$

OTHER STUDIES

Similar procedures to the one described are being investigated in several countries,
we have not carried out an in-depth search but there are articles on the internet that talk about Israeli, German and US researches:

We report the links to the identified articles (some of these links may not work anymore in the future):

- https://aktuelles.uni-frankfurt.de/englisch/pool-testing-of-sars-cov-02-samples-increases-worldwide-test-capacities-many-times-over/
- https://www.hindustantimes.com/india-news/govt-mulls-new-testing-approach/story-eD4ZiGlrdaynBRPSebpJoO.html
- https://www.medrxiv.org/content/10.1101/2020.04.03.20051995v1.full.pdf
- https://www.moneycontrol.com/news/coronavirus/coronavirus-pandemic-pool-testing-for-covid-19-may-benefit-india-as-us-study-highlights-test-feasibility-efficacy-5111811.html
- https://www.washingtonpost.com/outlook/2020/03/31/coronavirus-testing-groups/
- https://www.timesofisrael.com/to-ease-global-virus-test-bottleneck-israeli-scientists-suggest-pooling-samples/
- https://www.jpost.com/HEALTH-SCIENCE/Acceleration-in-multiple-coronavirus-tests-at-once-by-Israel-research-team-621533

REFERS:

Italian national data covid-19
https://github.com/pcm-dpc/COVID-19/tree/master/dati-andamento-nazionale

WHO - Coronavirus disease (COVID-19) technical guidance: Laboratory testing for 2019-nCoV in humans
https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/laboratory-guidance

WHO - Real-time RT-PCR assays for the detection of SARS-CoV-2
https://www.who.int/docs/default-source/coronaviruse/real-time-rt-pcr-assays-for-the-detection-of-sars-cov-2-institut-pasteur-paris.pdf?sfvrsn=3662fcb6_2

PANDORA-ID-NET Consortium - method for using RT-qPCR to diagnose the COVID-19
https://www.youtube.com/watch?v=5f_wieig4iQ
https://drive.google.com/drive/folders/1z_VvGEvenQ66aaIR7PLFPu2-nPKugMg4