The Right Kind of Pooled Testing for the Novel Coronavirus: First, Do No Harm

As with so much else in the COVID-19 pandemic, the United States is so far losing the SARS-CoV-2 testing race. In a growing number of areas of the country, as well as at educational and work sites, testing and obtaining timely results cannot keep pace with demand. Because the effort and cost of so much testing is considerable, and because some testing supplies are still scarce, it's important to conduct tests as efficiently and cost-effectively as possible.

If the nation is to overcome its testing challenges, it may have to turn to more pooled testing, in which a number of samples are combined and processed as if they were a single specimen. Some pooling of tests is already under way in other countries; the US Department of Health and Human Services (HHS) is reportedly considering it, and in mid-July 2020, the US Food and Drug Administration (FDA) gave the commercial laboratory company Quest Diagnostics the green light to pool samples of its diagnostic test for the virus.[1] But pooling tests in any fashion will not be a solution; in fact, there are right ways and wrong ways to undertake pooling.

The protocol currently being evaluated by the HHS, and to be used by Quest, is the so-called Dorfman protocol,[2] a highly touted procedure initially designed to test the urine of Army recruits for syphilis during World War II. Under this approach, for example, 10 samples may be pooled together; if the pool tests negative, all samples within the pool will be declared negative. If the pool tests positive, that will suggest that at least one sample in the pool is positive, so each sample will be tested individually to identify the positive samples.

Pooling tests this way may produce some greater testing efficiency and some savings as a result. But employing the Dorfman approach would be extremely dangerous, because it would produce high rates of false negative results—which means that potentially thousands of people who were in fact infected with the virus would test negative instead. With that false assurance, they might continue to behave as though they were uninfected, probably contributing to the even greater community spread of COVID-19. False negative test results are thus far more dangerous than false positives, which may simply make people distraught until they are retested and learn that they are actually negative (as occurred recently in Manchester, Vermont).[3]

There is a far better way to conduct pooled testing: use an approach known as "split pooling,"[4] which is almost the obverse of the Dorfman protocol. Under this approach, if a pool of samples tests negative, a new pool consisting of the same individual samples is assembled and tested to confirm the initial result. (For extra assurance, a different testing kit, with different levels of sensitivity and specificity, could also be used to perform this second test.) If the pool tests negative this second time, all the tests in the pooled sample are declared negative and none are tested again. However, if the pool tests positive, it is then split up into two pools of half the size and retested. This process of splitting all pools that test positive, and retesting them, is repeated as many times as necessary, even to the point where a single sample is retested if needed.

Now consider a scenario in which 5 million individuals are tested daily for SARS-CoV-2 virus in the United States, as some experts have recommended.[5] We will pose two different circumstances—one in which the prevalence of the virus is very low, at 0.04% of the population to be tested, and another in which it is much higher, at 2.4%. This range is broad, but realistic and allows us to contrast the effects of testing protocols. Because no test is perfect, and all have varying degrees of sensitivity and specificity,[6] greater prevalence will, in general, produce more false negatives.

Next, predicate the testing scenarios outlined on two assumptions, which are the same whether the Dorfman protocol or split pooling is employed. The first assumption is that pooling does not result in any diminution of specificity or sensitivity for pools of 16 samples or fewer, as Yelin et al. have verified.[7] The second assumption is that the results of the second round of testing under the split pooling protocol are independent from the first round (and here again, for further assurance, using a different testing kit from the one used in the first round would reinforce the independence assumption). Results are

ABOUT THE AUTHORS

Eugene Litvak is President and CEO of the Institute for Healthcare Optimization, Newton, MA, and is with the Harvard T. H. Chan School of Public Health, Boston, MA. Susan Dentzer is Senior Policy Fellow with the Robert J. Margolis Center for Health Policy, Duke University, Washington, DC. Marcello Pagano is Professor of Biostatistics, Harvard T. H. Chan School of Public Health.

Correspondence should be sent to Susan Dentzer, Senior Policy Fellow, Duke University, Robert J. Margolis Center for Health Policy, 1201 Pennsylvania Avenue, Suite 500, Washington, DC 20004 (e-mail: susan.dentzer@duke.edu). Reprints can be ordered at http://www.ajph.org by clicking the "Reprints" link.

This editorial was accepted August 22, 2020.

https://doi.org/10.2105/AJPH.2020.305945

TABLE 1—SARS-CoV-2 Testing Protocol Results at Two Testing Prevalence Rates: United States

SARS-CoV-2 Prevalence	No. Tests	False Positives, No.	False Negatives, No.
4 per 10 000			
Individual	5 000 000	99 960	40
Dorfman	700 000	54 978	236
Split pool	500 000	5 998	24
240 per 10 000			
Individual	5 000 000	97 600	2 400
Dorfman	2 800 000	53 680	14 160
Split pool	2 050 000	5 856	1 440

Note. Results are summarized for 5 million tests and initial pools of size 32.

summarized in Table 1 for 5 million tests and initial pools of size 32.

As Litvak and Pagano have previously demonstrated, split pooling produces significantly lower numbers of both false positives and false negatives.[4] The sensitivity and specificity used in our calculations for the example is 0.98 for all pools of size 16 or less, including for an individual test. For pool sizes of 32, the calculations decrease the sensitivity and specificity to 0.90. Both these values for pools of 32 and fewer are conservative when it comes to the tests used today in screening for COVID-19. Indeed, according to the FDA, sensitivity and specificity of several testing kits currently in use are much higher.

Under a prevalence rate of 0.04% or 2.40%, then, split pooling requires fewer tests than individual testing—just 10% of the individual test number at the lower prevalence and 41% at the higher prevalence. Still, even with this many fewer tests at the lower prevalence rate, fewer than 6000 false positives will occur by contrast to individual testing, which produces 99 960 false positives. The contrast at the higher prevalence is similar. What's more, the split pool generates 60% of the false negatives produced by individual testing at both prevalence rates.

The Dorfman protocol falls between the split pooling protocol and the individual test protocol at the low and high prevalence rates in the number of tests required and the number of false positives. But when it comes to false negatives, the Dorfman protocol produces far worse results than split pooling—almost 10 times (9.8) as many as the split pooling method delivers in both prevalence situations.

With modern automated laboratory equipment, it is easy to carry out split pool testing. Furthermore, to lower the number of tests that must be carried out, the size of the pools can also be adjusted depending on how widespread the virus is in the community. Given the rapidly rising number of infections now emerging in a number of states, it is surely time to try new strategies such as pooled testing. But by no means should pooled testing follow the Dorfman protocol. We can't afford to adopt strategies, such as individual testing and Dorfman pooling, that could run the risk of giving false assurances to large numbers of infected people, potentially making the toll of this terrible pandemic worse than it already is. AJPH

Eugene Litvak, PhD
Susan Dentzer, BA
Marcello Pagano, PhD

CONTRIBUTORS

The authors contributed equally to this editorial.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

REFERENCES

1. US Food and Drug Administration. Coronavirus (COVID-19) update: FDA issues first emergency authorization for sample pooling in diagnostic testing. 2020. Available at: https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-issues-first-emergency-authorization-sample-pooling-diagnostic. Accessed August 28, 2020.

2. Dorfman R. The detection of defective members of large populations. *Ann Math Stat*. 1943;14(4):436–440. https://doi.org/10.1214/aoms/1177731363

3. Jickling K. No COVID outbreak in Manchester, health commissioner says. *VTDigger*, July 21, 2020. Available at: https://vtdigger.org/2020/07/21/no-covid-outbreak-in-manchester-health-commissioner-says. Accessed August 28, 2020.

4. Litvak E, Tu XM, Pagano M. Screening for the presence of a disease by pooling sera samples. *J Am Stat Assoc*. 1994;89(426):424–434. https://doi.org/10.1080/01621459.1994.10476764

5. Pearlstein S. With 5 million tests a day, the rest of us can get back to work. *Washington Post*, May 21, 2020. Available at: https://www.washingtonpost.com/business/2020/05/21/with-5-million-tests-day-rest-us-can-get-back-work. Accessed August 28, 2020.

6. Pearce N, Vandenbroucke JP, VanderWeele TJ, Greenland S. Accurate statistics on COVID-19 are essential for policy guidance and decisions. *Am J Public Health*. 2020;110(7):949–951. doi: https://doi.org/10.2105/AJPH.2020.305708

7. Yelin I, Aharony N, Tamar ES, et al. Evaluation of COVID-19 RT-qPCR test in multi-sample pools. *Clin Infect Dis*. 2020; Epub ahead of print. https://doi.org/10.1093/cid/ciaa531