IDSA · Infectious Diseases Society of America · hivma hiv medicine association · OXFORD

Group Testing for Severe Acute Respiratory Syndrome–Coronavirus 2 to Enable Rapid Scale-up of Testing and Real-Time Surveillance of Incidence

Christopher D. Pilcher,[1,a] Daniel Westreich,[2,a] and Michael G. Hudgens[3]

[1]Department of Medicine, University of California San Francisco, San Francisco, California, USA, [2]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA, and [3]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

High-throughput molecular testing for severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2) may be enabled by group testing in which pools of specimens are screened, and individual specimens tested only after a pool tests positive. Several laboratories have recently published examples of pooling strategies applied to SARS-CoV-2 specimens, but overall guidance on efficient pooling strategies is lacking. Therefore we developed a model of the efficiency and accuracy of specimen pooling algorithms based on available data on SAR-CoV-2 viral dynamics. For a fixed number of tests, we estimate that programs using group testing could screen 2–20 times as many specimens compared with individual testing, increase the total number of true positive infections identified, and improve the positive predictive value of results. We compare outcomes that may be expected in different testing situations and provide general recommendations for group testing implementation. A free, publicly-available Web calculator is provided to help inform laboratory decisions on SARS-CoV-2 pooling algorithms.

Keywords. SARS-CoV-2; COVID-19; group testing; pooled testing; diagnostic testing; screening; surveillance.

Molecular tests of nasopharyngeal (NP) swab fluid for virus RNA remain the test of choice for early detection of severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2) infection, to identify new cases and to assess individual contagiousness. However, the high cost, limited throughput and imperfect specificity of molecular tests make them poorly suited to large-scale testing of populations with low expected rates of positivity.

Blood banks and human immunodeficiency virus (HIV) testing programs have addressed the problem of high-throughput molecular test screening for acute viral infections using group testing [1, 2]. In group testing, we first screen pools of specimens. When a pool is negative we declare the specimens in it negative; when a pool is positive, we retest subpools or individual specimens, depending on the strategy [2].

Several laboratories have recently published clinical validation studies in which SARS-CoV-2 RNA positive NP specimens from patients with coronavirus disease 2019 (COVID-19) have been tested in pools with RNA-negative clinical specimens

[3–7]. These studies have examined the use of unmodified assays and ad hoc pooling strategies comprising (variously) 5, 10, 32, or 64 total specimens. None of these studies have documented polymerase chain reaction (PCR) inhibition arising from pooling NP fluid samples.

Furthermore, 2 studies have confirmed that the analytic sensitivity of SARS-CoV-2 RNA PCR assays is lowered as expected when RNA in positive samples is diluted by negative samples in pools [3, 4]. For example, Abdalhamid et al [3] used an assay from the Centers for Disease Control and Prevention to test clinical specimens in pools of 5. Compared with individual testing, pooled testing resulted in cycle threshold (Ct) values that were on average 2.24 and 2.67 Ct higher (for targets N1 and N2), consistent with an increase of $\log_2(5) = 2.32$ Ct expected with 5-fold dilution (calculated as described elsewhere [3, table 1]). Based in part on these clinical validations, China used group testing to screen the population of Wuhan [8], and regional programs using expanded group testing are ongoing in Israel [7] and in Nebraska in the United States [3].

Unfortunately, NP swab group testing for SARS-CoV-2 has met with widespread skepticism. This is based in part on the perception that individual-specimen NP swab testing already has a "sensitivity problem"; diagnostic sensitivity in symptomatic patients has been found to be in the range of 60%–90% [9, 10]. To the extent that specimens have low virus loads, pooling dilution will reduce clinical detection even further [2].

In the current study, we seek first to describe the distribution of NP RNA viral loads that actually occur during the initial

"detection window" of acute SARS-CoV-2 infection. Second, we use these data to estimate how group testing will affect the outcomes of SARS-CoV-2 molecular testing efforts. Finally, we provide preliminary guidance for immediate implementation of efficient group testing algorithms for SARS-CoV-2.

METHODS

Viral Dynamic Model Development

Estimating the effects of testing and pooling approaches on testing outcomes requires some knowledge of the distribution of biomarkers that will be found in the testing population. In testing for acute infections, the problem is complicated by the rapid flux of viral loads as well as antibody levels over time in infected individuals. If these dynamics are well described, however, and individuals arrive for testing uniformly during the detection window, one can estimate changes in clinical case detection compared to individual testing. Specifically, uniform presentation during the detection window allows the problem to be reduced to measuring how testing choices affect the length of time that an average individual case can be detected.

We therefore sought to describe NP viral loads that occur in infected individuals during the time window when RNA is detectable by standard RT-PCR assays. Assays results are reported using "cycle threshold (Ct)" units, indicating the number of PCR reaction doubling times (cycles) needed for the target RNA in the sample to be amplified to a detectable concentration. Higher initial RNA concentrations detected at lower Ct values. The RNA viral load can be expressed as ΔCt, the difference between sample Ct and the Ct taken as the assay's lower limit of detection. Each ΔCt represents an increase of 1.0 log2 relative to the viral load at assay's lower limit. We reviewed recent articles [9–15] containing data on SARS-CoV-2 viral dynamics. Most presented individual-level data in visual form only. Two articles displayed NP viral loads from multiple individuals who had been frequently sampled within days of first detection [14] or last detection [13]; these plots appeared to show a rapid rise and similarly rapid fall in viral load on either side of the detection window. In an analysis differentiating noncritical from critical cases of COVID-19, Tan and colleagues [10] confirmed this abrupt onset and equally abrupt ending of shedding among noncritical patients. They also showed that within a few days after symptoms NP RNA was already detectable at peak levels by PCR, an average of 14 cycles before cutoff (suggesting that average viral loads were $\geq 4.2 \log_{10}$ above cutoff.) These typical dynamics have been contrasted in several articles by the dynamics in patients with critical COVID-19, documenting delayed onset of NP shedding [14], very high levels of peak shedding [15], slower viral decay [10, 15], and a longer detection window [9, 10] among critically ill individuals.

Based on this information, we proposed a model of respiratory virus dynamics with the intent of conservatively representing SARS-CoV-2 dynamics in individuals during the detection window of typical (ie, noncritical) acute infection. The model is illustrated in Figure 1A. Parameters were as follows: detection window, 14 days [10, 12]; peak viral load, 4.2 $\log_{10}$ viral load [10]; rate of viral increase, +1.0 $\log_{10}$ viral load per day; and slope of viral decay, −1.0 $\log_{10}$ viral load per day. As a check on these parameters we used this model to estimate the distribution of viral loads that would be expected in a hypothetical testing population of individuals who all followed average dynamics and presented uniformly for testing. The distribution predicted by the model agreed closely with the distributions of SARS-CoV-2 viral load found in recent studies [3, 9] among individuals first testing positive for SARS-CoV-2. (Figure 1C).

Estimating Dilution Effects

We next used the above viral dynamic model to estimate the sensitivity of pooled testing for SARS-CoV-2 RNA, as follows: first, we calculated the average detection window that would be expected with or without dilution, as illustrated in Figure 1B. We calculated the estimate of pooled testing sensitivity (relative to individual testing sensitivity) by dividing the pooled detection window by the individual detection window.

These estimates assume that infected individuals are equally likely to present at all times during the detection window. These estimates also assume ordinary specimen pooling procedure, wherein specimens are processed individually before pooling, and the same volume of fluid is put into the assay from pools or from individual specimens. Finally, they assume that similar interpretation criteria (eg, detection cutoffs for positive status) are used for both individual specimens and pools.

Estimating Testing Program Outcomes

The above model and estimation procedures were used to adapt a previously described software package designed to optimize group testing in acute HIV infection [2]. All calculations assumed that a representative assay would be used, with analytic sensitivity of 0.95 and specificity of 0.99, both of which we judged to be reasonable for molecular testing. To determine a possible upper limit on pool size, we estimated the maximum fold-dilution of specimen that would reduce the analytic sensitivity of pooled testing by <20% compared with individual testing. Here we distinguish between analytic sensitivity and diagnostic sensitivity, similarly to Saah and Hoover [16]. Specifically, diagnostic sensitivity is the probability that a testing protocol correctly identifies an individual with COVID-19 as infected (and as noted above, may be as low as 60% in some clinical circumstances [9, 10]); analytic (or test) sensitivity is the probability that an assay correctly classifies as positive a sample with viral load above the molecular level of detection. Reductions in analytic sensitivity result in proportional, relative reductions in diagnostic sensitivity. For example, suppose a testing protocol that does not involve pooling has a diagnostic sensitivity of 70%. If adding specimen pooling reduces the
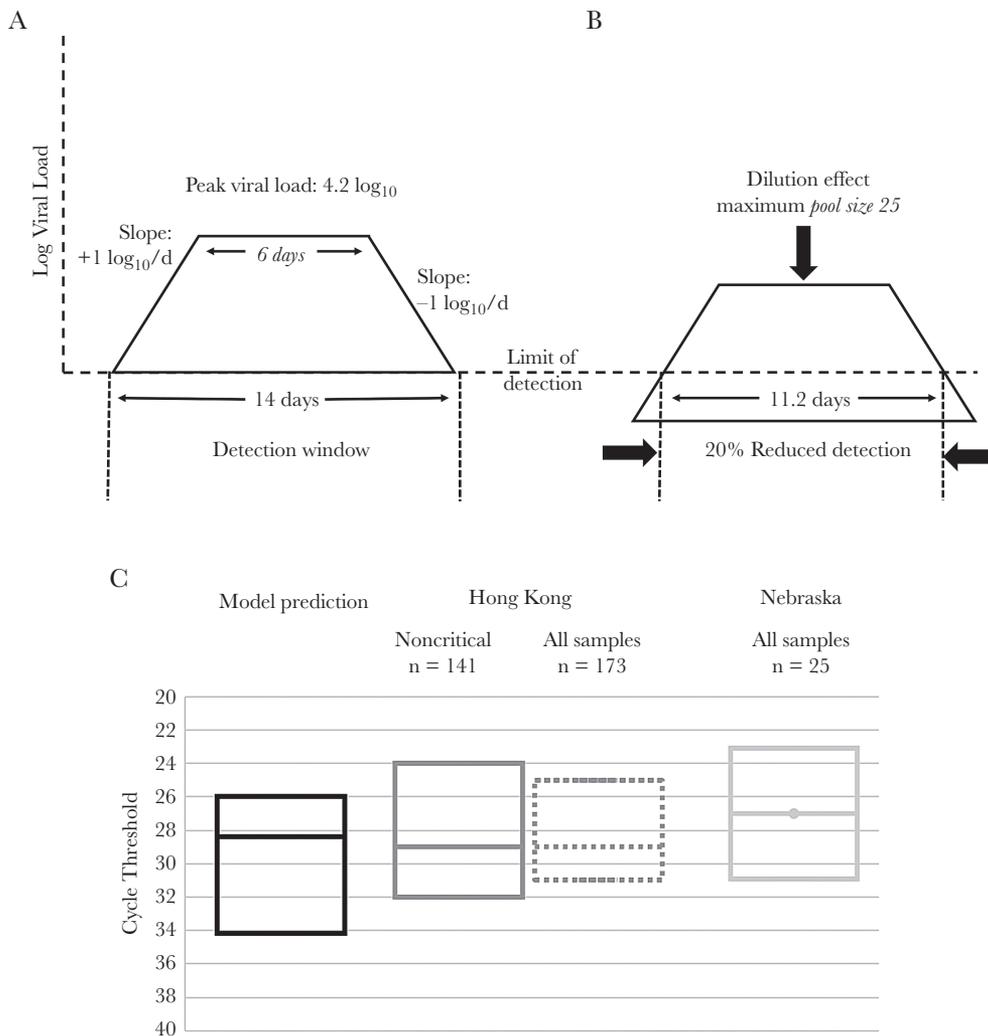
Figure 1. Model for nasopharyngeal (NP) RNA in acute severe acute respiratory syndrome–coronavirus 2 (SARS CoV-2) infection. The viral dynamic model shown here reflects viral load dynamics for an average acutely infected individual from the first to the last day with RNA levels above the limit of detection (of a standard assay, used on an individual specimen). Model parameters were selected to reflect typical (noncritical) infection. All viral load information is shown on a log scale. *A,* Model parameters, with a 14-day window of detection, 1 $\log_{10}$/day up and down slope on either side of a 6-day viral load plateau. *B,* Effect of specimen pooling on the window of detection and how sensitivity is estimated based on changes in the detection window. The calculation used to determine the maximum allowable pool size of 25 (see Results) is shown. *C,* Distribution of viral loads predicted by the model compared with those reported for testing populations in Hong Kong [7] and Nebraska [3]. The model-predicted distribution was estimated assuming that individuals with noncritical SARS CoV-2 infection would arrive for testing at uniform times during the detection window. For the Hong Kong study, results are those reported for individuals who were NP RNA positive at their first test. The Nebraska group reported the distribution for all positive testers. Boxes illustrate medians and interquartile ranges (IQRs); to allow comparison across studies, we estimated the IQR for the Nebraska study based on mean (standard deviation) of 27 (5.8) and an assumption of normality.

analytic sensitivity by 10%, then the new diagnostic sensitivity would be 70% × 90%, or 63% [16].

We addressed outcomes of 2 kinds of pooling strategies: 2 stage (*N*:1, that is 1 pool of *N* specimens, followed if necessary by retesting of individual specimens) and 3 stage (*kN*:*N*:1, where typically *k* = *N* and so N^2:*N*:1; that is, 1 pool of *kN* specimens, followed if necessary by retesting of *k* pools each containing *N* specimens, followed if necessary by retesting of individual specimens in individual positive subpools). We identified group testing algorithms for either strategy that would increase specimen throughput, increase actual case identification, and

increase the positive predictive value (PPV) of results, at levels of prevalence ranging from 1 per 1000 to 10% positive tests.

We reported on the following outcomes: average time to results (measured as mean number of rounds of testing, assuming that individual testing requires 1 round); efficiency, defined as expected number of specimens screened (or alternatively, individual results obtained) per molecular assay used, where individual testing allows screening of 1 specimen per assay; reduction in sensitivity compared with individual testing (given the above assumptions); and PPV. The PPV calculation assumes uncorrelated errors between rounds of testing, under

which assumption group testing usually leads to substantial increases in PPV [2]. A free, publicly-available Web calculator of the model is available to help inform laboratory decisions on SARS-CoV-2 pooling algorithms (http://www.bios.unc.edu/~mhudgens/SARS-CoV-2.pooling.home.html).

RESULTS

The proposed viral dynamic model is summarized in Figure 1. In Figure 1C, the distribution of viral load values predicted by the model is shown to be similar to those published in clinical studies of SARS-CoV-2 testing. For example, our model predicted that 50% of viral loads would have ΔCt values >11.6, similar to the 50% of viral loads >11 ΔCt found by Zhao and colleagues [9] among newly positive testers with noncritical COVID-19 in Hong Kong.

The model indicated that pool sizes >25 are expected to reduce analytic sensitivity by >20%, calculated as follows. First, as illustrated in Figure 1B), our assumption of uniform presentation during the detection window means that a 20% loss of analytic sensitivity is due to a loss of 20% of 14 days (2.8 days; or 1.4 days on the rise and 1.4 days on the fall). The dilution that causes us to lose 1.4 days on each side of the curve is calculated as $10^{1/d \times 1.4 \text{ days}} = 25.12$; we

therefore set 25 as a limit on upper pool size for subsequent analyses.

These SARS-CoV-2-specific estimates were then incorporated into a software package described elsewhere [2]. Predicted outcomes are shown for selected algorithms in Table 1 and in Figures 2 and 3. Gains in efficiency appeared to be large, allowing 2–20 times the number of specimens to be processed with the same number of tests. When the prevalence was >1%, simple pooling schemes and smaller pools (eg, 6:1 "minipools" for prevalence of .05) were more comparable in efficiency to larger and/or more complex pooling schemes. Below 1% prevalence, larger minipools could be several-fold more efficient (in terms of results per test used) than 5:1 minipools (Figure 3). When prevalence was 1 in 1000, larger pools and particularly 3-stage pools were substantially more efficient (Table 1 and Figure 3). Below 1% prevalence, adding the intermediate pool stage generally resulted in much higher testing efficiency.

Figure 3 shows expected differences in efficiency for different pooling strategies with master pool sizes of 5, 15 or 25, at various levels of prevalence. A 5:1 minipool format was selected by Abdalhamid and colleagues [3] for their initial Nebraska demonstration in a population with 5% expected prevalence: our analysis predicts 2.5 results per test in this scenario, similar

Table 1. Expected Outcomes of Group Testing in Samples with Different Prevalences of Detectable Severe Acute Respiratory Syndrome–Coronavirus 2[a]

| | Scenario | | | | Performance Relative to Individual Testing | | |
| | | | | | | Positive Predictive Value[f] | |
Pooling Strategy	Prevalence	Recommended Algorithm[b]	Time to Results[c]	Results Obtained per Test Used[d]	Reduction in Sensitivity vs Individual Testing[e]	Group	Individual
2-Stage testing	0.001	25:1	1.03	14.5	0.20	0.73	0.09
	0.005	17:1	1.07	7.6	0.18	0.85	0.32
	0.01	12:1	1.10	5.4	0.15	0.90	0.49
	0.05	6:1	1.23	2.5	0.11	0.96	0.83
	0.1	4:1	1.31	1.8	0.09	0.98	0.91
3-Stage testing	0.001	25:5:1	1.03	20.0	0.20	0.96	0.09
	0.005	25:5:1	1.12	12.7	0.20	0.96	0.32
	0.01	25:5:1	1.22	8.8	0.20	0.96	0.49
	0.05	16:4:1	1.59	3.1	0.17	0.97	0.83
	0.1	9:3:1	1.73	2.0	0.14	0.98	0.91

[a]Pool sizes suggested are those predicted to give the highest number of specimens screened per test used, while not reducing analytic sensitivity by >20% in a specimen pool. All estimates reflect assumed viral dynamics, dilution effects, and baseline assay performance in individual specimens (see text); they also assume that assay results are interpreted similarly (eg, using the same cycle threshold for a quantitative polymerase chain reaction assay) when testing pools or individual specimens.

[b]Two- and 3-stage algorithms are described in Estimating Testing Program Outcomes, within Methods.

[c]The time to results is estimated as the mean number of testing *rounds* required to obtain all results, because the time to completing a run will vary according to the assay and platform used by a laboratory. Here individual testing is assumed to require 1 round; in group testing most negative results require 1 round but some 2 or 3; and all positive results require 2 or 3 rounds (for 2- or 3-stage testing, respectively). These estimates depend on sensitivity and specificity.

[d]The number of test results generated in each group testing scenario was divided by the number of assays used in the process; this can be implicitly compared with individual testing, where this ratio is always 1. The ratio of results to tests indicates the increase in testing capacity that a laboratory can expect with an algorithm where test kits and supplies are the limiting factors. Greater efficiencies can be achieved if increased pool sizes (and increased dilution, and therefore lower sensitivity) are allowed.

[e]This is the degree to which group testing (given the master pool size in this row) reduces analytic sensitivity compared with individual testing. For example, in the first row of results, the 95% analytic sensitivity is reduced by 20% to [95% × (1 – 20%)] = 76%. This reflects losses in detection due to dilution only, based on assumptions about viral dynamics (see text).

[f]Positive predictive value (PPV) is the probability that, given a final positive result, the specimen is truly positive. Substantial increases in PPV comparing group testing with individual testing result from the effect of retesting positive samples in the group testing procedure, assuming uncorrelated errors between testing rounds. The model emphasizes that molecular tests with imperfect specificity (0.99 in our models) have inherently limited utility in low-prevalence situations such as SARS CoV-2 surveillance [5] where false-positive individual results could swamp true-positive results.
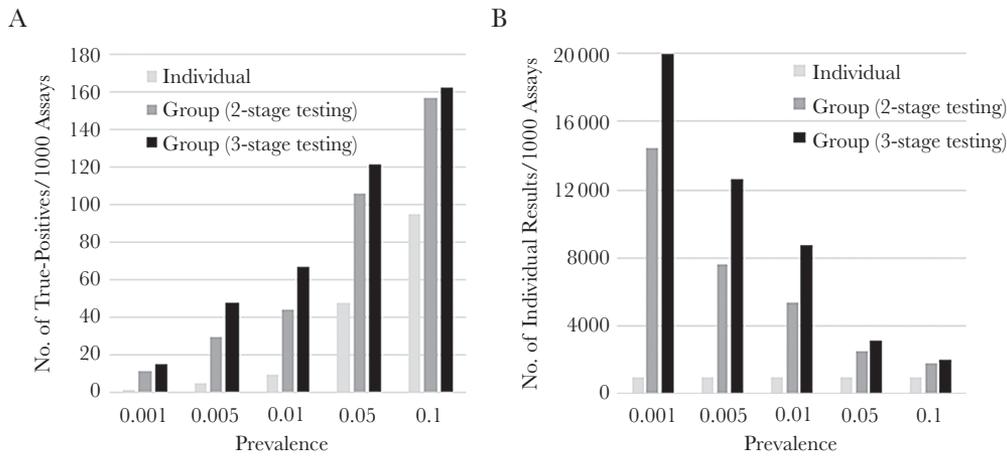
Figure 2. Expected number of patients tested and coronavirus disease 2019 (COVID-19) cases identified by individual versus group testing using a fixed number of molecular tests. Expected results are estimated for laboratories according to the prevalence of detectable severe acute respiratory syndrome–coronavirus 2 (SARS CoV-2) RNA in tested samples. Scenarios span situations with low prevalence (0.001; eg, surveillance [5]) and high prevalence (0.10; eg, clinical testing), and expected results are compared for individual testing, 2-stage group testing, and 3-stage group testing. *A,* COVID-19 cases identified, shown as the expected number of COVID-19 cases (ie, true-positive samples) found per 1000 assays used. *B,* Individuals tested, shown as the estimated number of samples with finalized results per 1000 assays, for individual and group testing.

to the 2.3 results per test seen in that study. As expected, the predicted efficiency of 5:1 minipooling was lower at higher prevalence values (10%, 25%, and 50%) (Figure 3). At lower prevalence values, all strategies are increasingly efficient; at a prevalence of 1 per 1000, the efficiency estimates for 5:1, 25:1,

and 25:5:1 strategies were 4.7, 14.6, and 20.1 results per test, respectively.

DISCUSSION

The results of this analysis suggest group testing schemes should be effective in expanding the capacity and throughput of molecular testing for SARS CoV-2. Simple-to-implement algorithms can allow between 2 and 20 results to be generated for every molecular test used, depending on the testing scenario. The highest gains in efficiency and testing performance were predicted for testing situations where the expected prevalence of disease is low; because testing is at present generally limited to those with high likelihood of having the disease, this means that the potential for greater efficiency with group testing in such low-prevalence settings has been underestimated. Situations with low expected prevalence of disease include screening of low-risk, asymptomatic healthcare workers, performing universal testing in healthcare facilities or in the general population, or large surveillance studies. Indeed, it is difficult to see how molecular testing can be rapidly scaled up in such settings without group testing. It is equally difficult to see how alternatives such as antibody tests, which identify past infection, will be able to identify new cases in an ongoing epidemic.

Importantly, our results indicate that the 5-specimen minipool protocol recently demonstrated by Abdalhamid and colleagues [3] can increase efficiency even when 10%–25% of samples are positive and could thus be an effective standard protocol. Having a standard minipool protocol allows laboratories to tailor testing for specific sample sets. For example, a laboratory processing samples from a general population survey would anticipate much greater efficiency at larger pool sizes of 15 or 25 and perhaps by using 3-stage testing (Figure 3);
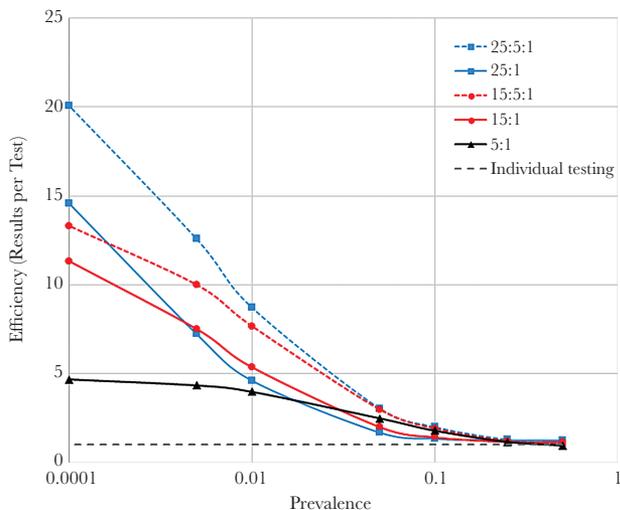

Figure 3. Expected differences in efficiency for different pooling strategies with master pool sizes of 5, 15, or 25, at various levels of prevalence. Testing efficiency as a function of group testing strategy and severe acute respiratory syndrome–coronavirus 2 (SARS CoV-2) prevalence. Model estimates for testing efficiency are shown for 5 testing strategies at levels of prevalence from 0.001 to 0.50. The strategies shown are individual testing, 5:1 minipools, 15:1 minipools, 15:5:1 (3-stage) pools, 25:1 minipools, and 25:5:1 (3-stage) pools. Prevalence was defined as the proportion of specimens with RNA levels above the assay cutoff. Efficiency was defined as the number of all results obtained (positive and negative) divided by the number of tests performed and expressed as results per test; individual testing has an efficiency of 1.

the larger pools needed could be simply created by pooling the standard minipools at the end of a standard processing procedure, given sufficient fluid volume.

It is essential to reiterate that pooling can sacrifice the analytic sensitivity of molecular tests for some low viral load specimens [2–6]. However, early results suggest that SARS-CoV-2 infection has fast-on/fast-off viral dynamics in NP fluid [13, 14], making it an ideal candidate for group testing. In particular, the rate of viral increase in acute infection in SARS Cov-2 seems to be substantially faster than that in HIV-1 [17, 18], and group testing is successful and indeed standard for acute HIV-1 [1]. As more information on the window of detection and the speed of viral load increase and/or decrease becomes available, our viral dynamics model should be reassessed, and the effects of pool size on dilution and analytic sensitivity should be reevaluated. In particular if the rise and fall of viral load is more gradual than we assumed, smaller maximum pool sizes may be desirable. However, some loss of sensitivity due to pooling may be acceptable, because at present there are insufficient molecular tests available for all the individuals who need them in many settings. In such a situation, the comparison is not between pooled sensitivity and sensitivity of individual testing (because individual testing is an impossibility), but rather between pooled sensitivity and the sensitivity of not testing at all. (Not testing, of course, has a sensitivity of 0% [95% confidence interval, 0%–0%]).

Adding serologic tests will help address the problem of diagnostic sensitivity of SARS CoV-2 RNA testing. For instance, Zhao and colleagues [9] showed that even in the first week of illness, when antibody tests had lower diagnostic sensitivity (38%) than NP RNA testing (67%), combining antibody and RNA results increased diagnostic sensitivity to 79%. In serial antibody/viral load testing algorithms (for instance, where only NP specimens from antibody-negative individuals are tested), group testing for viral RNA has been shown to be especially sensitive and efficient [1, 2]. Removing specimens from antibody-positive individuals may reduce the number of RNA positives in a sample set and thus may reduce the proportion that contain low viral loads. Estimating algorithms for this situation would require a modification of our present model, one taking antibody test dynamics into account.

The model of viral dynamics for this study was based on rapidly emerging clinical data. As more groups report results of group testing and viral dynamic studies, the assumptions of our model may change and our group testing Web tool will be updated accordingly and transparently, including a linked change log.

In summary, the need for group testing to make widespread high-throughput molecular testing feasible is clear. Although caution around group testing has been reasonable, numerous types of data now suggest that SARS CoV-2 is an ideal candidate for group testing. There are

logistical issues involved [2] that can be especially challenging for smaller laboratories. However, the Wuhan example of testing 6.5 million residents over a period of days in June 2020 shows that group testing can be efficiently implemented at scale [8]. The specimen pooling protocols needed for such have been published [3–5]. These can be cleanly implemented as long as laboratories are assured they will have appropriate regulatory clearances, reimbursement, and technical support. Regardless of location, the first step is for authorities to task large laboratories (both public health and commercial) with expanding testing, and to encourage that group testing be used if test availability is a limiting factor.

Notes

References

1. Pilcher CD, Fiscus SA, Nguyen TQ, et al. Detection of acute infections during HIV testing in North Carolina. N Engl J Med **2005**; 352:1873–83.
2. Westreich DJ, Hudgens MG, Fiscus SA, Pilcher CD. Optimizing screening for acute human immunodeficiency virus infection with pooled nucleic acid amplification tests. J Clin Microbiol **2008**; 46:1785–92.
3. Abdalhamid B, Bilder CR, McCutchen EL, Hinrichs SH, Koepsell SA, Iwen PC. Assessment of specimen pooling to conserve SARS CoV-2 testing resources. Am J Clin Pathol **2020**; 153:715–8.
4. Yelin I, Aharony N, Shaer Tamar E, et al. Evaluation of COVID-19 RT-qPCR test in multi-sample pools [published online ahead of print May 2, 2020]. Clin Infect Dis doi:10.1093/cid/ciaa531
5. Hogan CA, Sahoo MK, Pinsky BA. Sample pooling as a strategy to detect community transmission of SARS-CoV-2. JAMA **2020**; 323:1967–9.
6. Eis-Hübinger AM, Hönemann M, Wenzel JJ, et al. Ad hoc laboratory-based surveillance of SARS-CoV-2 by

real-time RT-PCR using minipools of RNA prepared from routine respiratory samples. J Clin Virol **2020**; 127:104381.

7. Ben-Ami R, Klochendler A, Seidel M, et al; The Hebrew University-Hadassah COVID-19 diagnosis team. Large-scale implementation of pooled RNA-extraction and RT-PCR for SARS-CoV-2 detection [published online ahead of print June 22, 2020]. Clin Microbiol Infect S1198-743X(20)30349-9. doi:10.1016/j.cmi.2020.06.009

8. Wee SL, Wang V. Here's how Wuhan tested 6.5 million for coronavirus in days. New York Times. Published 26 May **2020** (updated 3 June 2020). https://www.nytimes.com/2020/05/26/world/asia/coronavirus-wuhan-tests.html. Accessed 12 June 2020.

9. Zhao J, Yuan Q, Wang H, et al. Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019 [published online ahead of print March 28, 2020]. Clin Infect Dis doi:10.1093/cid/ciaa344

10. Tan W, Lu Y, Zhang J, et al. Viral kinetics and antibody responses in patients with COVID-19. medRxiv [Preprint]. 26 March 2020. Available from: https://doi.org/10.1101/2020.03.24.20042382.

11. Pan Y, Zhang D, Yang P, Poon LLM, Wang Q. Viral load of SARS-CoV-2 in clinical samples. Lancet Infect Dis **2020**; 20:411–2.

12. Wölfel R, Corman VM, Guggemos W, et al. Virological assessment of hospitalized patients with COVID-2019. Nature **2020**; 581:465–9.

13. Xu T, Chen C, Zhu Z, et al. Clinical features and dynamics of viral load in imported and non-imported patients with COVID-19. Int J Infect Dis **2020**; 94:68–71.

14. Liu Y, Yan LM, Wan L, et al. Viral dynamics in mild and severe cases of COVID-19. Lancet Infect Dis **2020**; 20:656–7.

15. To KK, Tsang OT, Leung WS, et al. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. Lancet Infect Dis **2020**; 20:565–74.

16. Saah AJ, Hoover DR. "Sensitivity" and "specificity" reconsidered: the meaning of these terms in analytical and diagnostic settings. Ann Intern Med **1997**; 126:91–4.

17. Fiebig EW, Wright DJ, Rawal BD, et al. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. AIDS **2003**; 17:1871–9.

18. Pilcher CD, Porco TC, Facente SN, et al; Consortium for the Evaluation and Performance of HIV Incidence Assays (CEPHIA). A generalizable method for estimating duration of HIV infections using clinical testing history and HIV test results. AIDS **2019**; 33:1231–40.