

# Increased SARS-CoV-2 Testing Capacity with Pooled Saliva Samples

Anne E. Watkins, Eli P. Fenichel, Daniel M. Weinberger, Chantal B.F. Vogels, Doug E. Brackney, Arnau Casanovas-Massana, Melissa Campbell, John Fournier, Santos Bermejo, Rupak Datta, Charles S. Dela Cruz, Shelli F. Farhadian, Akiko Iwasaki, Albert I. Ko, Nathan D. Grubaugh,<sup>1</sup> Anne L. Wyllie,<sup>1</sup> the Yale IMPACT Research Team<sup>2</sup>

We analyzed feasibility of pooling saliva samples for severe acute respiratory syndrome coronavirus 2 testing and found that sensitivity decreased according to pool size: 5 samples/pool, 7.4% reduction; 10 samples/pool, 11.1%; and 20 samples/pool, 14.8%. When virus prevalence is >2.6%, pools of 5 require fewer tests; when <0.6%, pools of 20 support screening strategies.

Limited laboratory capacity in the United States has hindered access to testing for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and has delayed results. To control outbreaks of coronavirus disease (COVID-19), testing capacity must be increased and maintained for the foreseeable future. One resource-saving, capacity-increasing approach is pooling samples, thereby testing multiple persons simultaneously. A negative result for the pool indicates that all samples were below the limit of detection, and a positive result for the pool requires individual re-testing of all samples. Pooled testing has been widely proposed as a way to expand capacity for large-scale screening (1,2; C.M. Verdun, unpub data, <https://doi.org/10.1101/2020.04.30.20085290>), a proactive strategy for early pathogen detection, primarily for persons who are not yet symptomatic.

Saliva is being used as a noninvasive source for SARS-CoV-2 testing (3,4) yet can be more difficult to process than traditional swab-based samples (5).

Author affiliations: Yale School of Public Health, New Haven, Connecticut, USA (A.E. Watkins, D.M. Weinberger, C.B.F. Vogels, A. Casanovas-Massana, A.I. Ko, N.D. Grubaugh, A.L. Wyllie); Yale School of the Environment, New Haven (E.P. Fenichel); Connecticut Agricultural Experiment Station, New Haven (D.E. Brackney); Yale University School of Medicine, New Haven (M. Campbell, J. Fournier, S. Bermejo, R. Datta, C.S. Dela Cruz, S.F. Farhadian, A. Iwasaki, A.I. Ko); Howard Hughes Medical Institute, New Haven (A. Iwasaki)

DOI: <https://doi.org/10.3201/eid2704.204200>

Given limited empirical evidence to properly inform projections of feasibility and cost-effectiveness of pooling, we explored the potential of pooling saliva to increase SARS-CoV-2 testing capacity.

## The Study

Using saliva collected from COVID-19 inpatients and at-risk healthcare workers (5), we combined 1 SARS-CoV-2-positive sample (<38 PCR cycle threshold [ $C_t$ ]) with SARS-CoV-2-negative saliva (Appendix, <https://wwwnc.cdc.gov/EID/article/27/4/20-4200-App1.pdf>) before RNA extraction in total pool sizes of 5 samples/pool ( $n = 23$  pools), 10 ( $n = 23$ ), and 20 ( $n = 31$ ). As pool size increased, detection sensitivity decreased independent of starting viral load (pool of 5, +2.2 cycle threshold [ $C_t$ ], 95% CI 1.4–3.0  $C_t$ ; 10, +3.1  $C_t$ , 95% CI 2.3–4.0  $C_t$ ; 20, +3.6  $C_t$ , 95% CI 2.7–4.4  $C_t$ ) (Figure 1; Appendix).

By applying the regression coefficients ( $C_t$  increase) to the  $C_t$  values from all SARS-CoV-2-positive saliva samples detected during our studies (6), we estimate that pool sizes will lead to detection sensitivities of 92.59% (95% CI 88.89%–95.56%) for pools of 5 samples, 88.89% (95% CI 80.00%–91.85%) for pools of 10, and 85.19% (95% CI 75.56%–91.11%) for pools of 20, relative to sensitivity of un-pooled samples (Appendix Figure 1). This loss in sensitivity could be minimized through protocol modifications: increasing the volume of pooled samples tested (400  $\mu$ L,  $n = 20$  pools of each size; Appendix Figure 2) and decreasing the elution volume.

On the basis of the calculated relative sensitivity loss resulting from pooling, we modeled the number of tests required (total of pooled and individual samples from positive pools tested) for a population of 10,000 with increasing SARS-CoV-2 prevalence

<sup>1</sup>These senior authors contributed equally to this article.

<sup>2</sup>Team members team are listed at the end of this article.

(Figure 2, panel A). We estimate that for populations with prevalence  $<0.6\%$ , pools of 20 require the fewest tests. However, for populations with prevalence  $>2.6\%$ , our analyses suggest that pooling of 5 samples leads to the fewest tests. For populations with prevalence  $>28.1\%$ , testing individual samples is more efficient than testing pools of any size. Thus, we suggest using an adaptive pooling strategy that accounts for SARS-CoV-2 prevalence for the population tested: as virus prevalence decreases, pool size can be increased, but as prevalence rises, pool size should be decreased.

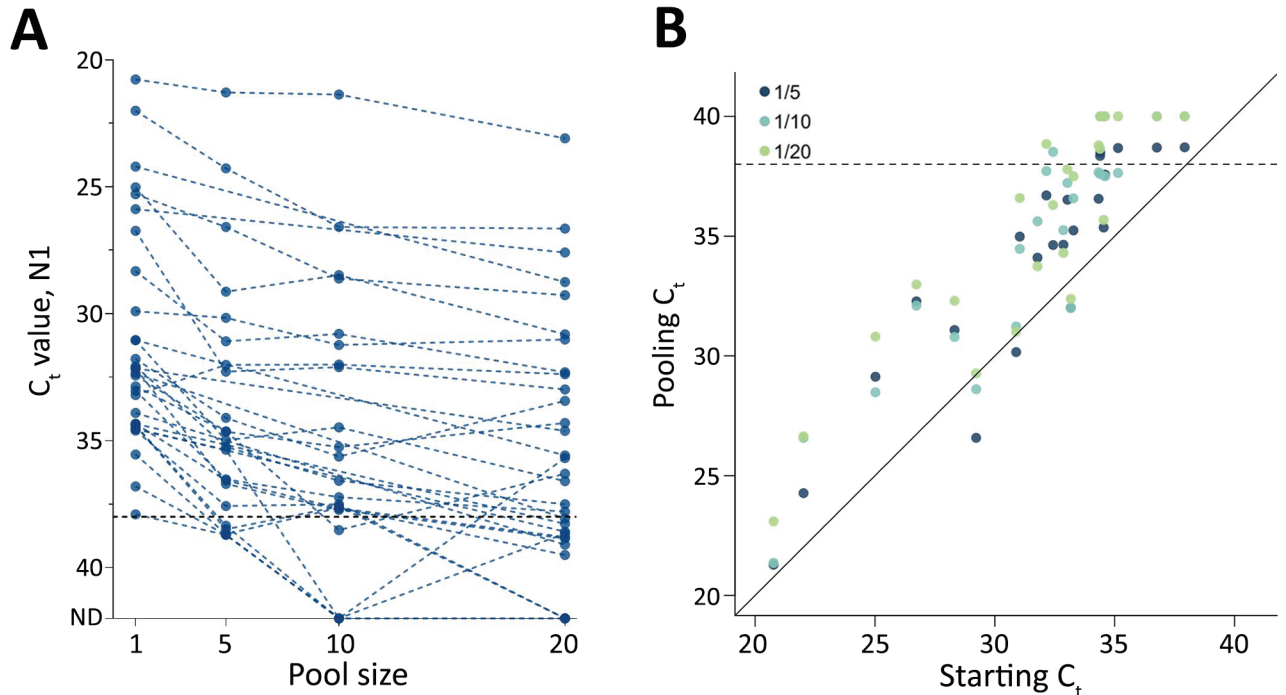
Because sensitivity varies by pooling design (Figure 1), a different number of positive results will be detected for a given population with a given SARS-CoV-2 prevalence. As virus prevalence decreases, we estimate that cost savings of pooled testing will increase (Figure 2, panel B). For example, if SARS-CoV-2 prevalence for a 10,000-person population was  $0.5\%$ , then pooling by 20 would require only 1,318 tests, including retesting of all persons from test-positive pools. If tests cost US\$30 each, the savings would be \$260,453 relative to individual testing while still identifying  $\approx 43$  of 50 infected persons. The savings will vary on a scale relative to test prices. Ultimately, the net benefits of pooled testing can continue to increase

even as virus prevalence decreases with increased pool sizes, which is essential for ongoing screening.

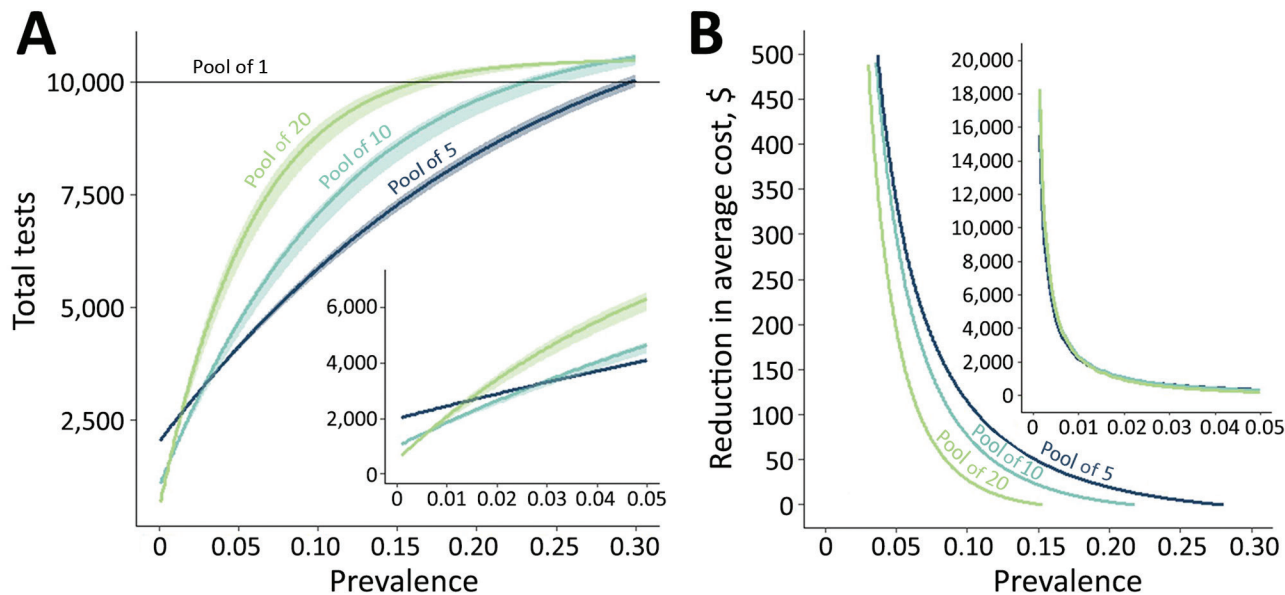
## Conclusions

The cost of SARS-CoV-2 testing can be prohibitive when positive samples are rarely found, presenting a major barrier to prolonged screening strategies. Pooling of samples can help overcome this barrier. Our model demonstrates that as local outbreaks fluctuate, adapting pool sizes will have resource-savings benefits.

The benefits of pooled testing will always be accompanied by decreased detection sensitivity. However, the lower overall number of tests required and the lower associated costs expands testing capacity, permitting more frequent testing, and testing persons more often mitigates the loss of sensitivity (7). By enabling broader testing, pooling has the potential to identify more infected persons than more limited (or no) individual testing. Infected persons can then be isolated from the population, thus reducing the probability of contact between a susceptible and an infectious person, ultimately reducing transmission. Given our findings, we urge the US Food and Drug Administration to develop new guidelines for pooled-testing approaches. Although the first Emergency Use



**Figure 1.** Effect of pooling on detection of severe acute respiratory syndrome coronavirus 2, by pool size and between samples tested. A) As the pool size increased, so did the  $C_t$  value (dotted lines connect pools comprising the same positive sample).  $C_t$  for positivity is set to 38. Samples falling on the x-axis indicated samples from which signal was not detected by reverse transcription quantitative PCR. B) As the pool size increased, so did the  $C_t$ . We equated this change by using linear regression (pool of 5 samples, dark blue,  $+2.2 C_t$ , 95% CI 1.4–3.0  $C_t$ ; pool of 10, light blue,  $+3.1 C_t$ , 95% CI 2.3–4.0  $C_t$ ; pool of 20, green,  $+3.6 C_t$ , 95% CI 2.7–4.4  $C_t$ ). Dashed lines indicate  $C_t$  38 (cutoff for sample positivity). 1/5, pool of 5; 1/10, pool of 10; 1/20, pool of 20.  $C_t$ , cycle threshold.



**Figure 2.** The resource-saving benefit of sample pooling for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) testing, based on size of the pool and expected prevalence of SARS-CoV-2 within the population. A) We modeled the number of tests required to test 10,000 persons (results qualitatively scale with population) when pools contain 5, 10, or 20 samples (and individually retesting samples within positive pools) compared with testing samples individually (pool = 1 sample). As prevalence increases, so does the number of pools positive for SARS-CoV-2, thereby increasing the required number of confirmatory tests of individual samples. Therefore, over a prevalence of 2.6%, pooled samples of 5 result in fewer overall tests required than do larger pool sizes. B) At lower prevalence rates, such as when outbreaks have been controlled but ongoing screening is required, pools of 10 or 20 samples yield substantial cost savings for the same expected level of positive detections, after accounting for sensitivity differences. Values are shown in US\$. Insets show the region with <5% prevalence.

Authorization for SARS-CoV-2 pooled testing ( $\leq 4$  swab samples in 1 test) (8) will be most useful in high-prevalence settings, the  $\approx 12\%$ – $15\%$  losses in sensitivity when pooling 10–20 samples would probably not pass current authorization criteria ( $>95\%$  sensitivity).

Going forward, screening strategies need to be reviewed separately from traditional diagnostic testing, taking into consideration the repeated testing of individuals performed during screening. For strategies considering twice-weekly sampling, such as in the reopening plans for many US colleges, even if larger pools have lower sensitivity per test, the probability of 2 repeated false-negative results for any person will often be less than the probability of a false-negative result for a single test from a small pool. For example, a small pool (or individual test) may have the probability of a false-negative result of 2% but cost may limit testing to once per week. Conversely, the lower per-person cost of a large pool with a per-test probability of a false-negative result of 14% is more likely to allow for testing twice per week. Therefore, persons tested twice in larger pools have a per-week false-negative probability of only 1.96%. In the context of prolonged community screening, sensitivity should be thought of as per unit time, and the testing regimen should be taken into account.

Our estimates are conservative; the number of tests required is most likely lower than predicted, especially if behavioral or geographic information can be used to stratify the population so that the adaptive pooling strategy can be applied differentially to different sampled subpopulations. However, this approach needs to be balanced with feasibility in the laboratory because pooled testing adds additional steps and complexity to the system, all of which must be reliably implemented. Furthermore, pooled approaches could incorporate retesting individual samples from pools generating any SARS-CoV-2-specific signal in quantitative reverse transcription PCR regardless of  $C_t$  (in place of those pools with the  $<38 C_t$  cutoff applied here) (9). Although pooling has traditionally focused on extracted nucleic acid before quantitative reverse transcription PCR (10–12), because of the expense of RNA extraction and a comparable effect on detection sensitivity (Appendix), we recommend pooling before RNA extraction. Validation of our work in additional settings and on a larger scale will help better inform our models.

The cost-savings benefits of adaptive pooling *saliva* for community screening for SARS-CoV-2 provides a mechanism to maintain testing as virus spread

is brought under control and to avoid resurgence. Even if prevalence is very low, it is probably desirable to increase pool sizes before stopping testing altogether. Together with the ease of saliva collection, pooling samples should be considered as an effective testing strategy for expanding the breadth of testing and continued screening during the ongoing COVID-19 pandemic.

Yale IMPACT Research Team authors: Kelly Anastasio Michael H. Askenase, Maria Batsu, Sean Bickerton, Kristina Brower, Molly L. Bucklin, Staci Cahill, Yiyun Cao, Edward Courchaine, Giuseppe Delulisi, Rebecca Earnest, Bertie Geng, Ryan Handoko, Christina A. Harden, Chaney C. Kalinich, William Khoury-Hanold, Daniel Kim, Lynda Knaggs, Maxine Kuang, Eriko Kudo, Melissa Linehan, Peiwen Lu, Alice Lu-Culligan, Anjelica Martin, Irene Matos, David McDonald, Maksym Minasyan, Adam J. Moore, M. Catherine Muenker, Nida Naushad, Allison Nelson, Jessica Nouws, Abeer Obaid, Camila Odio, Ji Eun Oh, Saad Omer, Isabel M. Ott, Annsea Park, Hong-Jai Park, Xiaohua Peng, Mary Petrone, Sarah Prophet, Tyler Rice, Kadi-Ann Rose, Lorenzo Sewanan, Lokesh Sharma, Denise Shepard, Mikhail Smolgovsky, Nicole Sonnert, Yvette Strong, Codruta Todeasa, Maria Tokuyama, Jordan Valdez, Sofia Velazquez, Arvind Venkataraman, Pavithra Vijayakumar, Elizabeth B. White, and Yexin Yang.

### Acknowledgments

We gratefully acknowledge the study participants for their time and commitment to the study. We thank all members of the clinical team at Yale-New Haven Hospital for their dedication and work, which made this study possible. We also thank S. Taylor and P. Jack for technical discussions.

This study was supported by the Huffman Family Donor Advised Fund (N.D.G.), Fast Grant funding support from the Emergent Ventures at the Mercatus Center, George Mason University (N.D.G.), the Yale Institute for Global Health (N.D.G.), and the Beatrice Kleinberg Neuwirth Fund (A.I.K.). C.B.F.V. is supported by NWO Rubicon 019.181EN.004.

### About the Author

Ms. Watkins is a master's degree student and research assistant in Epidemiology of Microbial Diseases, Yale School of Public Health, in New Haven, CT. Her primary research interests are the intersection of infectious and chronic diseases with a focus on respiratory and neurologic pathogens.

### References

- Cherif A, Grobe N, Wang X, Kotanko P. Simulation of pool testing to identify patients with coronavirus disease 2019 under conditions of limited test availability. *JAMA Netw Open*. 2020;3:e2013075. <https://doi.org/10.1001/jamanetworkopen.2020.13075>
- Abdalthamid 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. <https://doi.org/10.1093/ajcp/aqaa064>
- Yokota I, Shane PY, Okada K, Unoki Y, Yang Y, Inao T, et al. Mass screening of asymptomatic persons for SARS-CoV-2 using saliva. *Clin Infect Dis*. 2020 Sep 25 [Epub ahead of print]. <https://doi.org/10.1093/cid/ciaa1388>
- Vogels CBF, Brackney D, Wang J, Kalinich CC, Ott I, Kudo E, et al. SalivaDirect: simple and sensitive molecular diagnostic test for SARS-CoV-2 surveillance [cited 2021 Jan 27]. [https://www.cell.com/med/pdf/S2666-6340\(20\)30076-3.pdf](https://www.cell.com/med/pdf/S2666-6340(20)30076-3.pdf)
- Landry ML, Criscuolo J, Peaper DR. Challenges in use of saliva for detection of SARS CoV-2 RNA in symptomatic outpatients. *J Clin Virol*. 2020;130:104567. <https://doi.org/10.1016/j.jcv.2020.104567>
- Wyllie AL, Fournier J, Casanovas-Massana A, Campbell M, Tokuyama M, Vijayakumar P, et al. Saliva or nasopharyngeal swab specimens for detection of SARS-CoV-2. *N Engl J Med*. 2020;383:1283–6. <https://doi.org/10.1056/NEJMc2016359>
- Larremore DB, Wilder B, Lester E, Shehata S, Burke JM, Hay JA, et al. Test sensitivity is secondary to frequency and turnaround time for COVID-19 surveillance. *Sci Adv*. 2021;7:eabd5393. <https://doi.org/10.1126/sciadv.abd5393>
- Quest Diagnostics. Quest SARS-CoV-2 rRT-PCR package insert [cited 2020 Aug 20]. <https://www.fda.gov/media/136231>
- Pasomsab E, Watcharananan SP, Watthanachockchai T, Rakmanee K, Tassaneeritthep B, Kiertiburanakul S, et al. Saliva sample pooling for the detection of SARS-CoV-2. *J Med Virol*. 2020 Aug 25 [Epub ahead of print]. <https://doi.org/10.1002/jmv.26460>
- Eis-Hübinger AM, Hönemann M, Wenzel JJ, Berger A, Widera M, Schmidt B, 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. <https://doi.org/10.1016/j.jcv.2020.104381>
- Wyllie AL, Rümke LW, Arp K, Bosch AATM, Bruin JP, Rots NY, et al. Molecular surveillance on *Streptococcus pneumoniae* carriage in non-elderly adults; little evidence for pneumococcal circulation independent from the reservoir in children. *Sci Rep*. 2016;6:34888. <https://doi.org/10.1038/srep34888>
- Yelin I, Aharony N, Tamar ES, Argoetti A, Messer E, Berenbaum D, et al. Evaluation of COVID-19 RT-qPCR test in multi-sample pools. *Clin Infect Dis*. 2020;71:2073–8. <https://doi.org/10.1101/2020.03.26.20039438>

Address for correspondence: Anne L. Wyllie, Yale School of Public Health, 60 College St, New Haven, CT 06511, USA; email: [anne.wyllie@yale.edu](mailto:anne.wyllie@yale.edu); Eli P. Fenichel, Kroon Hall, 195 Prospect St, New Haven, CT 06511, USA; email: [eli.fenichel@yale.edu](mailto:eli.fenichel@yale.edu)



# Increasing SARS-CoV-2 Testing Capacity with Pooled Saliva Samples

## Appendix

### Methods

#### Sample pooling

Saliva was collected as a part of the Yale IMPACT Biorepository (1) from COVID-19 inpatients and healthcare workers at the Yale-New Haven Hospital (Yale Human Research Protection Program Institutional Review Boards FWA00002571, Protocol ID. 2000027690) (1). RNA was extracted and tested by RT-qPCR for SARS-CoV-2 RNA (N1) (2).

Saliva samples were combined into pools of 5 (n=23), 10 (n=23), and 20 (n=31). Each pool contained equal amounts of one SARS-CoV-2 positive sample (as determined by RT-qPCR; 31 positive samples total) and the respective number of individual SARS-CoV-2 negative samples required to complete the target pool size. RNA extraction from pooled samples and RT-qPCR for SARS-CoV-2 detection were performed according to the biorepository's standard operating procedures (1–3) with either 300  $\mu$ l (equating to 60  $\mu$ l, 30  $\mu$ l, and 15  $\mu$ l of the original sample) or 400  $\mu$ l (equating to 80  $\mu$ l, 40  $\mu$ l, and 20  $\mu$ l of the original sample; n=20 for each of 5, 10 and 20 pools) total extraction input volume with RNA eluted into a total volume of 75  $\mu$ l. Later, RNA extracted from saliva (n=10) was tested individually or together in pool sizes of 5 or 10 and tested in RT-qPCR for SARS-CoV-2 detection (2). The cut-off for all RT-qPCR assays in this study was set at a cycle threshold (Ct) of 38 (2).

#### Statistical analyses

##### Sensitivity analyses

We fit a linear regression to the experimental pooling data to model the change in Ct values of positive samples following pooling. Let ' $\Delta$ Ct' be the change in Ct value of pooled samples and let 'ratio' be the categorical ratio of pool size (i.e. 1/5, 1/10, 1/20). Analyses were done separately by input volume in order to determine the effect of pool size under both 300  $\mu$ l

and 400 µl extraction conditions. This equation was used, separately, for both pre-extraction saliva and post-extraction RNA pooling. Ratio in this model can be interchanged with “condition” for the model of the 1/20 PBS and water dilution data.

We found that the change in Ct value post-pooling was independent of the Ct value of the undiluted sample (Pearson’s,  $r=-0.004$ ; 95% CI: -0.240, 0.233), thus it was not included in the model. Confidence intervals were generated by simulating from the covariance matrix of the parameters from the fitted model using the `mvrnorm` function in the R package “MASS” (4), and quantile functions.

#### Modeling the resource-saving benefit of sample pooling for SARS-CoV-2 testing

The problem of pooling can be approached modularly. We model pooling based on the expected prevalence in a test population of known size at a given time. By approaching the problem this way, we abstract from the problem of estimating prevalence in the sampled population at a given time. Nevertheless, our approach can be plugged into broader population level models with epidemiological dynamics.

If samples are independent of each other, pulled from the same well-mixed population (identically distributed), and that anyone in a test-positive pool needs to be re-tested individually, then binomial sampling theory provides the tool to compute the number of tests needed, which has been used for over half a century (5,6). The number of positive groups is  $P = [1 - (1 - \sigma(g)m)^g](N/g)$ , given a total test population of size  $N$  that is divided into groups of size  $g$  yield  $(N/g)$  groups, with a prevalence of infection in the sampled population equal to  $m$  and a test sensitivity  $\sigma(g)$ , where sensitivity can be a function of group size. The total number of tests need is  $T = (N/g) + Pg$ . The R script to implement these calculations are available at <https://github.com/efenichel/pooled-saliva-testing>.

To calculate the total number of tests and the number of test positive groups, we assume the expected prevalence is computed with error or that any error is orthogonal to the sampling error associated with the estimates of sensitivity. Therefore, to propagate the uncertainty associated with sensitivity sampling error, we make the calculations for the number of positive groups and total tests using a single predefined, conservative cut-off value. This mimics the existence of an established protocol. Variation in the cycle thresholds used would increase the sampling uncertainty for sensitivity, and would expect the point estimate of the sensitivity,

conditional on pool size, to be a convex combination of the estimates using individual cycle thresholds.

In practice, those coordinating testing need to consider what constitutes a single, well-mixed sampled population. For example, demographic or socio-economic information may be used to group samples into distinct subpopulations prior to pooling and testing. This would be called stratifying the population. If these subpopulations have different expected prevalences, then different sized pools may be optimal for the different subpopulations. However, stratification requires population specific data that is invariant to the test itself or stronger assumptions. The possibility of embedding an adaptive pooling approach into a model of a system that brings population level data to bear is a strength of the approach.

Another consideration that the model does not directly address is selection into the sampled population. If there is selection, then the well-mixed assumption is violated. There are two reasons to be concerned about this in practice. First, if people who are more likely to test positive are also more likely to get tested when there is a binding test capacity constraint, then as the constraint is relaxed with pooling, the expected prevalence in the population is likely to fall. This is a reason why stratifying the sample based on observable features, e.g., self-assessed probability of infection prior to pooling might be important. Conversely, consider a segment of the population that tries to avoid testing and engages in high risk behaviors (i.e., people who believe COVID-19 is a hoax). These people select out of testing. If it is easier to include these people in a testing regime with greater capacity due to pooling, then expected prevalence may actually rise. This can also be addressed with stratification of the population.

Further statistical analyses were conducted in GraphPad Prism 8.0.0 as described in the text and figure legends.

## **Results**

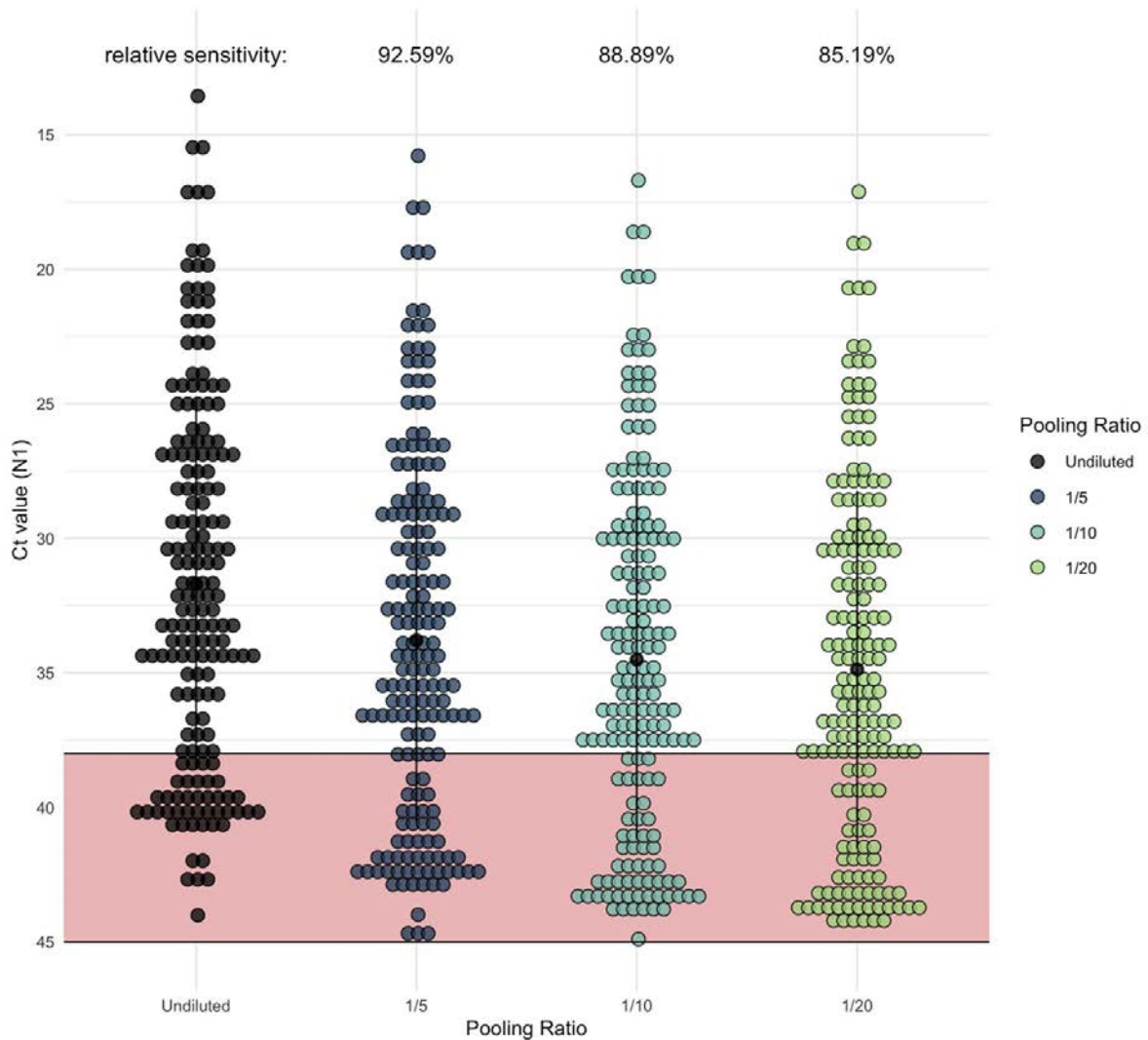
When pooling saliva samples, the effect on the sensitivity of detection was independent of the Ct value of the undiluted sample (Pearson's,  $r=-0.004$ ; 95% CI: -0.240, 0.233), i.e. the sensitivity loss in a sample with a higher Ct value (lower viral load) was not more than that of a sample with a lower Ct value (higher viral load).

We also evaluated the effect of pooling post-RNA extraction and pooled RNA templates extracted from undiluted saliva samples by 5 and by 10 (n=10). While we observed a similar decrease in sensitivity (pool: of 5, +2.2 Ct, 95% CI: 1.7-2.6; pool of 10, +3.1 Ct, 95% CI: 2.6-3.6) as to when pooled prior to RNA extraction, the degree to which each sample varied was less with less overall variation as compared to pre-extraction pooling (F test, pools: of 5,  $p = 0.061$ ; pools of 10,  $p = 0.009$ , Appendix Figure 3).

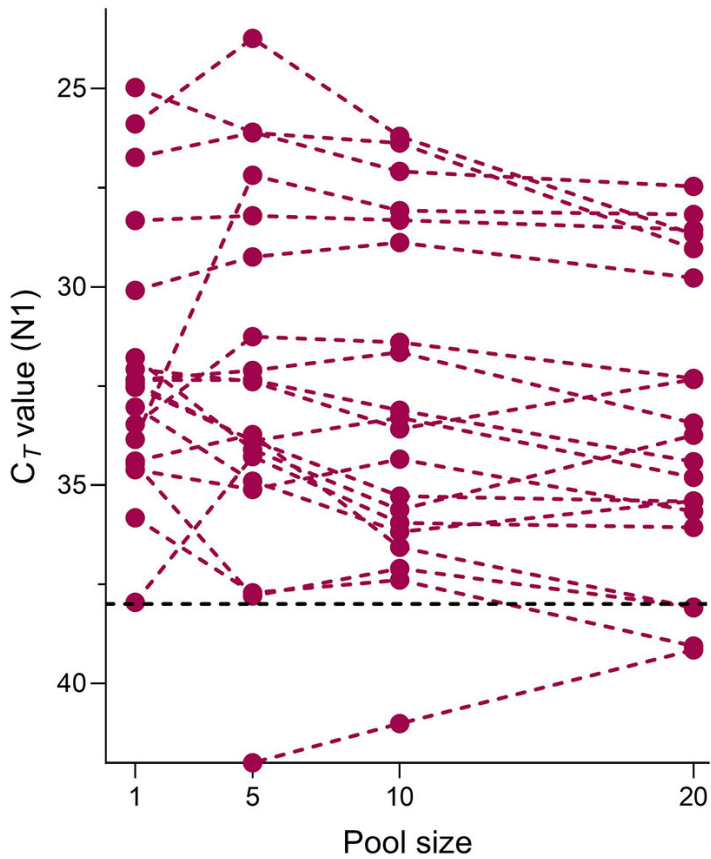
## References

1. Wyllie AL, Fournier J, Casanovas-Massana A, Campbell M, Tokuyama M, Vijayakumar P, et al. Saliva or nasopharyngeal swab specimens for detection of SARS-CoV-2. *N Engl J Med*. 2020;383:1283–6. [PubMed https://doi.org/10.1056/NEJMc2016359](https://doi.org/10.1056/NEJMc2016359)
2. Vogels CBF, Brito AF, Wyllie AL, Fauver JR, Ott IM, Kalinich CC, et al. Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT-qPCR primer-probe sets. [Internet]. *Nat Microbiol*. 2020;5:1299–305. [PubMed https://doi.org/10.1038/s41564-020-0761-6](https://doi.org/10.1038/s41564-020-0761-6)
3. Ott I, Vogels C, Grubaugh N, Wyllie A. Saliva collection and RNA extraction for SARS-CoV-2 detection v1 [cited 2020 Dec 12]. <https://doi.org/10.17504/protocols.io.bg3pjymn>
4. Ripley B, Venables B, Bates DM, Hornik K, Gebhardt A, Firth D, et al. Package “MASS.” [cited 2021 Jan 28]. <https://cran.r-project.org/web/packages/MASS/MASS.pdf>
5. Dorfman R. The detection of defective members of large populations. *Ann Math Statist*. 1943;14:436–40. <https://doi.org/10.1214/aoms/1177731363>
6. Gollier C, Gossner O. Group testing against Covid-19 [cited 2021 Jan 28]. *Covid Economics*. 2020;1:32–42.

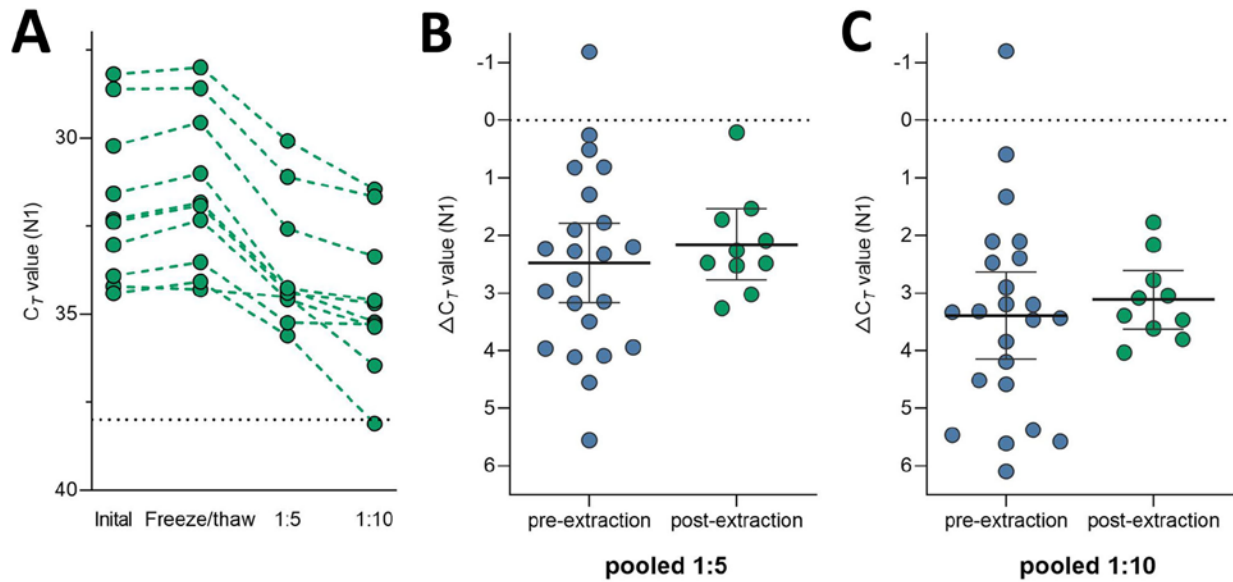




**Appendix Figure 1.** As pool size increases, more samples would be classified as negative in comparison to samples tested individually (unpooled). Each dot represents one of the 180 Yale IMPACT saliva samples which generated signal when tested by RT-qPCR for SARS-CoV-2 N1. Of these, 135 fell below the cycle threshold (Ct) of 38 (solid line) and were classified as positive for virus. The regression coefficient (representing expected Ct increase) for each of the pool sizes was added to the Ct value generated from the undiluted sample (shown in black) to determine the relative level of sensitivity for each pool size. The area shaded in red indicates the Ct range in which N1-signal is considered to be below the limit of detection.



**Appendix Figure 2.** Cycle threshold (Ct) values of saliva samples (n=20) tested individually (pool size = 1) at a total volume of 300  $\mu$ L, or when diluted with an increasing number of negative samples (total pool sizes of 5, 10 and 20) and a total extraction volume of 400  $\mu$ L. When extracting from 400  $\mu$ L volumes of pooled samples, we observed improved detection (pool of 5, -0.1 Ct, 95% CI -1.2, 1.1; pool of 10, 0.3 Ct, 95% CI -0.8, 1.5; pool of 20, 1.1 Ct, 95% CI -0.1, 2.2; linear regression). Dotted lines connect pools comprised of the same positive sample. Ct threshold for positivity is set to 38. Samples falling below the x-axis indicated samples from which signal was not detected in RT-qPCR.



**Appendix Figure 3.** Less variation in cycle threshold (Ct) values when pooling RNA templates. (A) Ct values of SARS-CoV-2 positive RNA (n=10) extracted from saliva samples when tested individually (pool size = 1) on day of sample collection (initial) and following storage of RNA at -80°C (freeze/thaw), or when diluted with 4 or 9 SARS-CoV-2 negative RNA templates (total pool sizes of 5 and 10). Dotted lines connect pools comprised of the same positive sample. While the median change in Ct value was comparable whether pooling samples or RNA templates by (B) 5 (Mann-Whitney,  $p = 0.499$ ) or (C) 10 (Mann-Whitney,  $p = 0.556$ ), pooling of samples resulted in more varied Ct changes (F test,  $p = 0.061$  and  $p = 0.009$ , respectively).