Statistical Science
2022, Vol. 37, No. 2, 229–250
https://doi.org/10.1214/22-STS857

Statistical Modeling for Practical Pooled Testing During the COVID-19 Pandemic

Saskia Comess, Hannah Wang, Susan Holmes and Claire Donnat

Abstract. Pooled testing offers an efficient solution to the unprecedented testing demands of the COVID-19 pandemic, despite their potentially lower sensitivity and increased costs to implementation in certain settings. Assessments of this trade-off typically assume the underlying infection statuses of pooled specimens to be independent and identically distributed. Yet, in the context of COVID-19, these assumptions are often violated: testing done on networks (housemates, spouses, co-workers) captures individuals with correlated infection statuses and risk, while infection risk varies substantially across time, place and individuals. Neglecting dependencies and heterogeneity may bias established optimality grids and induce a sub-optimal implementation of the procedure. As a lesson learned from this pandemic, this paper highlights the necessity of integrating field sampling information with statistical modeling to efficiently optimize pooled testing. Using real data, we show that (a) greater gains can be achieved at low logistical cost by exploiting natural correlations (nonindependence) between samples—allowing improvements in sensitivity and efficiency of up to 30% and 90%, respectively; and (b) these gains are robust despite substantial heterogeneity across pools (nonidentical). Our modeling results complement and extend the observations of Barak et al. (*Sci. Transl. Med.* **13** (2021) 1–8) who report an empirical sensitivity well beyond expectations. Finally, we provide an interactive tool for selecting an optimal pool size using contextual information.[1]

Key words and phrases: COVID-19, pooled testing, correlations, heterogeneity.

With an estimated 16% to 40% of asymptomatic SARS-CoV-2 cases and 50% of detection occurring prior to symptom onset (He et al., 2020, Oran and Topol, 2020, Pollock and Lancaster, 2020, Ma et al., 2021), widespread surveillance testing plays a crucial role in monitoring and controlling the spread of SARS-CoV-2 (Larremore et al., 2020, Gandhi, Yokoe and Havlir, 2020, Mina, Parker and Larremore, 2020, Dhillon et al., 2015, Nouvellet et al., 2015, Rannan-Eliya et al., 2021). Yet in practice, the inherent logistical costs of widespread testing have severely limited its deployment at scale. Throughout the pandemic, testing needs have outstripped availability: in November 2020, the United States fell short of its COVID-19 testing objective by 48%, performing a daily average of 1,193,000 tests out of the 2.3 million set as a minimum target (Collins, 2020). Testing shortages continue to persist globally, as reported during the Spring 2021 surge in India (Frayer, 2021), during the "third wave" throughout Africa in Summer 2021 (Mwai, 2021), and during the Winter 2021/22 Omicron-variant fueled surge in the U.S. (Heilweil, 2021, Pietsch, 2021). The unprecedented surge in testing demand has also strained the broader laboratory supply chain; from September 2020 through January 2021 in the U.S., shortages of testing materials (e.g., reagents, consumables, etc.) significantly impacted day-to-day testing for both COVID-19 and other infectious diseases (ASM, 2021). Consequently, despite rising vaccination rates, the threats of new variants, waning immu-

Saskia Comess is a PhD student, Emmett Interdisciplinary Program in Environment and Resources, Stanford University, Stanford, California (e-mail: saskiaco@stanford.edu). Hannah Wang is a resident physician, Department of Anatomic and Clinical Pathology, Stanford University School of Medicine, Stanford, California (e-mail: wangh1@stanford.edu). Susan Holmes is a Professor, Department of Statistics, Stanford University, Stanford, California (e-mail: susan@stat.stanford.edu). Claire Donnat is an Assistant Professor, Department of Statistics, The University of Chicago, Chicago, Illinois (e-mail: cdonnat@uchicago.edu).

[1]The corresponding shiny-app interface is available at: https://homecovidtests.shinyapps.io/Group-testing/.

nity, and localized outbreaks make the deployment of robust and continued large scale testing a priority.

In this context, pooled (or group) testing procedures have generated increasing interest during the pandemic (Abdalhamid et al., 2020). First proposed by Dorfman (1943) to screen soldiers for syphilis, the simplest form of pooled testing is a two-stage hierarchical procedure in which multiple laboratory specimens are first combined and tested. Only samples from positive pools are then subsequently individually retested. Since then, pooled testing has been successfully employed in a number of applications, ranging from the testing for low prevalence diseases (including HIV, chlamydia, and gonorrhea (McMahan, Tebbs and Bilder, 2012a, Wein and Zenios, 1996, Tu, Litvak and Pagano, 1995, Gaydos, 2005), to the detection of genetically modified organisms in crops (Yamamura and Hino, 2007).

Despite limited prior use of pooled testing for widespread epidemic management, the American Food and Drug Administration (FDA) approved its use for certain SARS-CoV-2 diagnostic tests in Summer 2020, with some restrictions. According to these guidelines, the sensitivity of the pooled procedure should be maintained above the threshold of 85% (FDA, 2020). This desideratum must also be weighed against logistical feasibility of implementing pooled testing—a tension recently described by the College of American Pathologists and which can be summarized along the following two axes (CAPS, 2020):

(a) Logistics. Prior to the pandemic, a substantial body of literature already considered the problem of optimizing pooled designs—either by grouping specimens according to a set of covariates (McMahan, Tebbs and Bilder, 2012a, Chen, Tebbs and Bilder, 2009, Bilder, Tebbs and Chen, 2010), or by placing the samples into an array matrix and pooling by combinations of rows and columns (McMahan, Tebbs and Bilder, 2012b) to allow the immediate identification of contaminated specimens and minimize individual retesting. A more nuanced approach, "informative array testing", combines these two methods and incorporates knowledge of population heterogeneity and covariates with an array scheme (McMahan, Tebbs and Bilder, 2012b). While these procedures can yield impressive efficiency gains from a purely statistical perspective, they simultaneously introduce more room for errors in specimen handling: if performed manually, specimen pooling can increase risk of specimen confusion and cross-contamination while increasing lab handling times. Automated robots become essential for aliquoting and attributing samples to pools following complex optimal designs. Thus, while mathematically optimal, these solutions are often difficult to implement without state-of-the-art (and often expensive) equipment.

(b) Context-dependent efficiency and sensitivity. The sensitivity and efficiency (number of tests per sample) of pooled testing are known to be functions of the pool size and disease prevalence. The latter determines the probability that a pool contains at least one positive individual, and therefore, that all individuals in the pool require retesting (Gastwirth, 2000). Larger pool sizes in low prevalence regimes ensure that fewer tests have to be carried out, while high prevalence levels imply higher risks for a majority of pools to require resampling and the pooling strategies to be rendered inefficient. Concurrently, pooling dilutes the amount of viral genetic material present in positive samples, thereby potentially reducing the sensitivity of the procedure. Previous studies have investigated the sensitivity of pooled testing under different prevalence levels in order to develop coarse recommendations for selecting an appropriate pool size at a given prevalence level (Kim et al., 2007, McMahan, Tebbs and Bilder, 2012a). Such studies typically assume independently and identically distributed (i.i.d) samples when estimating the appropriate pool size. However, such i.i.d assumptions may not be reasonable for SARS-CoV-2 given widespread community transmission and specimen collection procedures that capture highly connected networks—a phenomenon virtually unique to such a widespread epidemic scenario, and irrelevant to the low-prevalence-disease monitoring that pooled testing has historically been used for.

In fact, in many SARS-CoV-2 collection scenarios, infections and positive specimens are clustered, such as when testing students in dorms, coworkers in an office, individuals in households, or classrooms of children on a weekly basis (Eunjung Cha, 2021, Adam et al., 2020, Zhang et al., 2020). Transmission rates in these networks may be much higher than in less connected networks (Table 6 in the Appendix). By way of illustration, household transmission rates are estimated to vary from 4% to 55% (Koh et al., 2020)—highlighting transmission rates within close communities greater than the overall prevalence by orders of magnitude, but with significant uncertainty and heterogeneity. This leads us to a first lesson that we draw from our analysis.

LESSON 1. *The unprecedented data collection and sampling processes deployed during the pandemic have severely compromised the validity of classical statistical pipelines for the analysis of data—thereby leading to an inaccurate evaluation of pooled testing and potentially suboptimal deployment of this method. Statistical modeling is key to rapidly and efficiently re-adapting existing procedures to this novel setting, but its relevance is contingent on being able to bridge the gap between statistics-based optimality—which strives to make the greatest efficiency gains—and field-based optimality—which is informed by practical constraints and logistical considerations.*

To adapt to this novel situation, several studies have begun investigating the impact of deviations from the i.i.d. setting on the sensitivity and efficiency of pooled testing. To account for *nonindependence*, Rewley (2020) simulated correlations between consecutive persons in a testing queue, assuming an additive increased chance of a positive test given the previous person in line was positive. The simulations suggested that as the likelihood of clustered infections increased, pooling efficiency also increased, even with rising prevalence. Other simulation studies have similarly concluded that the Dorfman two-stage procedure is optimal when testing is performed on clusters of correlated individuals (Deckert, Bärnighausen and Kyei, 2020, Lin et al., 2020). So far, many of these studies have necessarily relied on simulating simplified settings, with arbitrary parameterisations and distributions and ignoring variability across pools. Few simulation studies have attempted to capture deviations from the *identically distributed* hypothesis, in part because there is minimal practical applicability for incorporating information on individual level covariates relevant to infection risk.

Taking an experimental data-driven approach, Barak et al. (2021) examined Dorfman-based pooled testing on over 130,000 SARS-CoV-2 samples. Pool size was adaptively chosen based on predicted prevalence levels in the community. They found that the rate of positives in pooled samples was best predicted by sorting samples into batches according to their source (such as by colleges, nursing homes or health care personnel) and also incorporating epidemiological information about the probability of infection in these different sources. Overall, they observed that pooled testing performance exceeded expectations both in terms of efficiency and sensitivity, which they attributed solely to the fact that there is a nonrandom distribution of positive samples in pools. Real-world data consequently supports the push to develop easily adaptable pooled testing strategies that exploit the non-i.i.d. nature of samples. Due to rapidly changing SARS-CoV-2 prevalence, laboratories require practical tools that allow them to adapt their procedures to the context and populations they treat.

At the time of writing, the study of optimal pooled testing under dependencies is made even more relevant by its increasing adoption across many real-world settings. Pooling has been proposed as a strategy for performing mass-testing at lower costs in K-12 schools (Simas et al., 2021). A cluster randomised trial examining pooled testing in German elementary schools found that pooled testing is an effective method for detecting positive cases, and is robust to different choices of sampling techniques (e.g., pooled saliva samples vs. oropharyngeal swabs) (Joachim et al., 2021). In Washington DC, pooled testing was performed during the 2020–2021 school year at an independent K-12 school on all staff and students using average pools of size 7.4 based on shared classes or work assignments. After the program was initiated, the proportion of students in remote learning decreased by between roughly 60% and 99%, depending on the grade/age of students, and the average cost of testing per student decreased (Berke et al., 2021). Pooled testing programs have also been instituted in California public schools on pods of five to twenty-five students (Jones, 2021) and in Massachusetts public schools on pools of maximum 10 samples using a Dorfman two-stage procedure (Massachussets, Department of Education, 2022, MASS.gov, 2021, Pollock et al., 2021). While these implementations implicitly take advantage of social relations and nonindependence (e.g., by grouping co-workers and children who share a classroom), the precise benefits of this method have not yet been formalized.

Objectives. We have two primary objectives:

1. *Show how a simple pooling method and accounting for correlated specimens in statistical modeling can yield unexpectedly efficient solutions*. To this end, we provide a straight-forward model that measures the efficiency of pooled testing under correlations, as well as formalizes and extends the lessons from Barak et al. (2021) in practical pooled sampling.
2. *Investigate and produce actionable recommendations that are ready for deployment during the COVID-19 pandemic*. For this reason, we focus on the Dorfman two-stage procedure rather than more mathematically optimal, but unscalable, pooling procedures.

We organize our discussion around three main lessons that our investigation of pooled testing for SARS-CoV-2 samples has taught us: (1) the importance of leveraging setting-specific information to optimize testing, (2) the necessity of evaluating efficiency through a set of practically meaningful measures, and (3) the importance of modeling the impact of uncertainty and/or heterogeneity. We show that pooled testing can efficiently identify infectious individuals despite natural deviations from i.i.d. hypotheses in the specimen collection process, with little detrimental effect on the accuracy of the procedure. Gains in sensitivity and efficiency can in fact be as much as 30% and 90% respectively compared to i.i.d. settings. In contrast to existing studies, our modeling of correlations (i) is focused on understanding pooled sampling's heightened sensitivity through its effect on the viral load of the sample, (ii) is informed by real data and practical constraints and above all, (iii) allows for the simultaneous consideration of nonindependence and population heterogeneity.

1. LEVERAGING CORRELATIONS AND THE HITCHHIKER EFFECT

This section focuses on a three-fold approach to understanding the impact of deviations from the i.i.d. hypotheses on the viral load of the sample (1) We introduce

a model that accounts for positive correlations between samples; (2) We provide a mathematical framework to quantify the observation by Barak et al. (2021) that pooled testing achieves higher than expected sensitivity due to the "hitchhiker effect"—a phenomenon whereby the detection of weakly positive tests is improved by borrowing strength from strongly positive tests; and (3) We use this framework to show how positive correlations between samples can improve the sensitivity of the pooling procedure beyond what is expected in an uncorrelated setting.

1.1 A Network Model for Pooled Specimens

Modeling Nonindependence (Network Effect). Consider that there are n samples in each pool. We assume that all the specimens are sampled from a network (co-workers, classroom, household, etc.) modeled by a fully connected graph on n nodes with edges indicating potential transmission between a given node and its neighbors (Figure 1). We denote by τ the network transmission probability between individuals in the pool (that is, the probability that an infected member infects another subject in the network), so that each edge weight $e_{ij} = \tau$ represents the probability that node i, if infected, transmits the disease to node j; in epidemiology, this parameter is commonly referred to as the secondary attack rate (SAR). The community transmission probability, or equivalently, the prevalence in the general population, is denoted by π. Let us denote as Y_i the indicator variable that specimen i is infected ($Y_i = 1$ if individual i infected, 0 otherwise). Since transmission can occur within the network or in the community, we decompose Y_i as follows: (1) Let $T_i^{(\text{cmty})}$ be the random variable indicating infection of individual i from outside the group (community transmission), and (2) Let $T_i^{(\text{ntw})}$ indicate infection from within the group (network transmission). The existence of a network effect is captured by writing the infectious status of individual i as the sum

$$Y_i = T_i^{(\text{cmty})} + (1 - T_i^{(\text{cmty})})T_i^{(\text{ntw})}$$
$$\text{with } T_i \in \{0, 1\},\ T_i^{(\text{cmty})} \perp\!\!\!\perp T_i^{(\text{ntw})} (M_1)$$
$$\text{so that } \mathbb{P}[Y_i = 1] = \mathbb{P}\big[\{T_i^{(\text{cmty})} = 1\} \cup \{T_i^{(\text{ntw})} = 1\}\big].$$

Network and community transmissions are themselves modeled as Bernoulli variables and tied to the SAR τ and prevalence π through the relations

$$\mathbb{P}\big[T_1^{(\text{cmty})} = 1\big] = \pi$$

and

$$\mathbb{P}\big[T_1^{(\text{ntw})} = 1 | T_2^{(\text{cmty})}, \ldots, T_n^{(\text{cmty})}\big] = 1 - (1 - \tau)^{\sum_{i=2}^n T_i^{(\text{cmty})}}.$$

In other words, network transmission is modeled as independent Bernoulli(τ) variables across edges, so that the probability that this transmission route fails is the product of the probability of failure across each edge: $\mathbb{P}[T_1^{(\text{ntw})} = 0 | T_2^{(\text{cmty})}, \ldots, T_n^{(\text{cmty})}] = (1 - \tau)^{\sum_{i=2}^n T_i^{(\text{cmty})}}$. A given pool can have a total of $K \leq n$ positive samples, k of which are infected from the community and $K - k$ of which are infected from within the network.

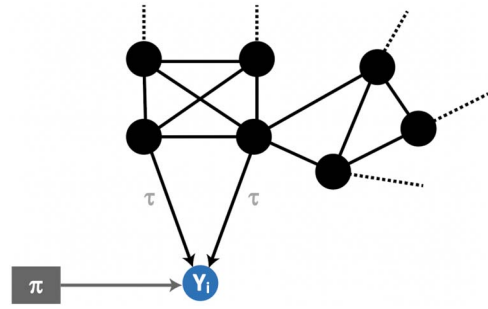


FIG. 1. *Representation of the network effect on the infectious status for any given node. This highlights the higher risk of sample i testing positive as soon as another sample in the network is also positive.*

Modeling Nonidentical Distributions (Heterogeneous Infection Probabilities). Individuals within a network are exposed to various levels of community transmission risk depending on a number of covariates, including age, profession, and lifestyle. This risk also varies considerably with time and epidemic kinetics (new variants, vaccination levels, etc.). At the granular level, this can be captured by introducing node covariates X_i's and allowing heterogeneous community infection rates,

$$\forall i, \quad \mathbb{P}\big[T_i^{(\text{cmty})} = 1\big] = \pi + f(X_i) + \epsilon_i,$$

where π is the general community prevalence level, $f(X_i)$ reflects deviations from this baseline level depending on a set of covariates (e.g., profession, lifestyle), and ϵ_i is a noise term capturing the stochasticity of the prevalence (e.g., temporal effects) and/or potential subject effects. Several studies have investigated using subject-level covariates to inform the risk function f, by classifying individuals either as high/low risk (McMahan, Tebbs and Bilder, 2012a, Bilder, Tebbs and McMahan, 2019, Donnat et al., 2020), or to inform retesting (Bilder and Tebbs, 2012). However, in the context of the COVID-19 pandemic, and due to the necessary volume and frequency of testing, introducing individual subject covariates yields impractical solutions: the collection of subject-level data and dispatching of samples in pools according to individual risk slows down the procedure and yields intricate pooling designs which are not feasible at scale. As such, we propose simply leveraging the specimen collection process to assume similar behaviors and covariates across pooled specimens ($f(X_i) \approx f(X_j)\ \forall i, j$), so that pools on any given day can be considered homogeneous. In this case, random effects—due for instance to temporal

variations, different variants, etc—are at the pool (instead of the individual subject) level:

$$\forall i, \quad \mathbb{P}\big[T_i^{(\text{cmty})} = 1\big] = \pi + \epsilon_{\text{pool}}.$$

Null and Alternative Models. This simple model allows us to capture a variety of situations. In particular, (i) the value of the SAR τ drives the balance between community and network transmission: the higher the value of τ, the more likely that any community transmission will yield more than one secondary infection in the pool. Conversely, for $\tau = 0$, there is no network transmissions and the samples are independent. (ii) Assuming this notation, the standard i.i.d. case studied in the literature is a homogeneous, fixed effect model and assumes $\epsilon = 0$ (so that $\forall i, \pi_i = \pi$) and $\tau = 0$. We refer to the uncorrelated, i.i.d. scenario with prevalence π and pool size n as our null model $H_0(\pi, n)$, and to the correlated, heterogeneous one as the alternative $H_a(\pi, n, \tau)$. (iii) This model allows us to study the law of the number of infected samples in the pool. As we will describe in subsequent subsections, the shift in probability distribution leads to improved performance of the pooling procedure. In Table 1, we summarize the properties of the distribution of number of positive samples per pool under both models.

Since (as in Table 1), the probability of the number of positive samples per pool is a complex polynomial function of τ, we propose an approximation when the prevalence is small. This allows us to gain greater insight into the intricate interplay between τ and π, without hindering the utility of the analysis since pooled sampling is predominantly deployed in low prevalence settings. More precisely, we make the following assumption.

ASSUMPTION 1 (Low Prevalence). Throughout our analysis, whenever we make use of approximations to derive greater insights into pooled sampling, we assume that the prevalence is such that $\pi n \leq 0.10$. To put this number into context, this scenario is aligned with situations observed in Summer 2020 or Spring 2021 in Europe and the United States: in early June 2021 for instance, the reported prevalence of COVID-19 in the United Kingdom was estimated around 0.70%, which would allow us to look at pool sizes of up to 15, or to sizes of up to 50 for low prevalence levels under 0.2% observed in some parts of the world (such as for instance Israel in late June 2021 (Ritchie et al., 2020) [1]). This is a convenient threshold that allow us to simplify the analysis while still providing insight into the interplay between community and network transmission, as highlighted by the following observations that are a consequence of Assumption 1.

[1]Chart of percentage of positive samples per test is provided in Figure 1 in the Supplementary Material (Comess et al., 2022).

OBSERVATION 1 (Number of positive samples per pool in the low-prevalence regime under H_0). Under Assumption 1 ($\pi n \leq 0.10$), the probability that under H_0 (i.e., no correlation), there are two or more infected samples in the pool is less than 0.01. This follows from the following simplification:

$$\begin{aligned}
&\mathbb{P}_{H_0}\left[\sum_{i=1}^{n} Y_i > 1\right] \\
&\quad = 1 - (1-\pi)^n - n\pi(1-\pi)^{n-1} \\
&\quad = 1 - (1-\pi)^{n-1}\big(1 + (n-1)\pi\big) \\
&\quad \leq 1 - \left(1 - (n-1)\pi + \frac{(n-1)(n-2)}{2}\pi^2 \right. \\
&\qquad \left. - \frac{(n-1)(n-2)(n-3)}{6}\pi^3\right) \times \big(1 + (n-1)\pi\big) \\
&\quad = \frac{n(n-1)}{2}\pi^2 - n\frac{(n-1)(n-2)}{3}\pi^3 \\
&\qquad + \frac{(n-1)^2(n-2)(n-3)}{6}\pi^4 \\
&\quad \leq \frac{(n\pi)^2}{2} + \frac{n^4\pi^4}{6} \leq 0.006.
\end{aligned}$$

Similarly, in a pool with at least one infected sample, we can show that

$$\mathbb{P}[S \geq 2 | S \geq 1] \leq \frac{n}{2}\pi \leq 0.05$$

(see Supplementary Material (Comess et al., 2022), equation (4)). This means that with 95% confidence, an infected pool contains only a single infected sample. This fact is useful for simplifying our computations in subsequent paragraphs.

OBSERVATION 2 (Number of positive samples per pool in the low-prevalence regime under H_a). Concurrently, in this scenario, the number of positive samples per pool *with correlation* (under H_a) can be approximated by:

$$\mathbb{E}_{H_a}\left[\sum_{i=1}^{n} Y_i\right] = n\pi + n(1-\pi)\big(1 - (1-\tau\pi)^{n-1}\big)$$

while the probability that there are more than one infected samples in the pool is

$$\begin{aligned}
&\mathbb{P}_{H_a}\left[\sum_{i=1}^{n} Y_i > 1\right] \\
&\quad = 1 - (1-\pi)^n - n\pi(1-\pi)^{n-1}(1-\tau)^{n-1} \\
&\quad = \mathbb{P}_{H_0}\left[\sum_{i=1}^{n} Y_i > 1\right] + \big(1 - (1-\tau)^{n-1}\big)n\pi(1-\pi)^{n-1}
\end{aligned}$$

This illustrates the striking difference in behavior between the distribution of positive samples with and without correlation: the probability of having more than two samples in a correlated pool increases rapidly as τ increases.

TABLE 1

Comparison of the properties of number of infected samples per pool (size $= n$)*, with and without correlation* (*derivations provided in the Supplementary Material* (Comess et al., 2022))*. Note the shift in expected number of samples as a function of the network transmission rate* (*or SAR*), τ

Quantity	$H_0(\pi, n)$	$H_a(\pi, n, \tau)$
Distribution of # of contaminated samples $\sum_{i=1}^n Y_i$	$\sum_{i=1}^n Y_i \sim \text{Binomial}(n, \pi)$	$\sum_{i=1}^n T_i^{(\text{cmty})} + \text{Binomial}(n - \sum_{i=1}^n T_i^{(\text{cmty})}, \tau)$ with $\sum_{i=1}^n T_i^{(\text{cmty})} \sim \text{Binomial}(n, \pi)$
$\mathbb{P}[Y_i = 1]$	π	$1 - (1-\pi)(1-\pi\tau)^{n-1}$
Correlation $c(Y_i, Y_j)$	0	$\frac{(1-\pi)^2[(1-\pi+(1-\tau)^2\pi)^{n-2} - (1-\pi\tau)^{2n-2}]}{(1-(1-\pi)(1-\pi\tau)^{n-1})((1-\pi)(1-\pi\tau)^{n-1})}$ $= (1 + \frac{1-\tau}{1+(n-1)\tau})\tau + o(\tau)$
Average number of positive samples per pool	$n\pi$	$n\pi + \sum_{k=1}^{n-1} \binom{n}{k} \pi^k (1-\pi)^{n-k}(1-(1-\tau)^k)(n-k)$ $= n\pi(1 + (n-1)\tau) + o(n\pi)$
Average number of positive samples per contaminated pool	$\frac{n\pi}{1-(1-\pi)^n}$ ≈ 1	$1 + \sum_{k=1}^{n-1} \binom{n}{k} \frac{\pi^k (1-\pi)^{n-k}}{1-(1-\pi)^n}(1-(1-\tau)^k)(n-k)$ $\approx (1 + (n-1)\tau)$
$\mathbb{P}[\sum_{i=1}^n Y_i = 0]$	$(1-\pi)^n$	$(1-\pi)^n$
$\mathbb{P}[\sum_{i=1}^n Y_i = k], k > 0$	$\binom{n}{k} \pi^k (1-\pi)^{n-k}$	$\sum_{j=1}^k \binom{n}{j} \pi^j (1-\pi)^{n-j} \binom{n-j}{k-j} (1-(1-\tau)^j)^{k-j} (1-\tau)^{j(n-k)}$

To illustrate the relative effect of the correlation τ on this probability, let us consider the Taylor expansion of the previous equation around $\pi = 0$ in the low-prevalence regime of Assumption 1. In this case, we can show that

$$\mathbb{P}_{H_a}\left[\sum_{i=1}^n Y_i > 1\right] = \mathbb{P}_{H_0}\left[\sum_{i=1}^n Y_i > 1\right]\left(1 + \frac{2(1-(1-\tau)^{n-1})}{(n-1)\pi} - \frac{2((n+1)((1-\tau)^{n-1}-1))}{3(n-1)} + o(n\pi)\right)$$

Lower bounding $(1-\tau)^{n-1}$ by $1-(n-1)\tau$, this means that a lower bound for the probability of having strictly more than one positive sample in a pool under H_a has leading term $(1 + \frac{2\tau}{\pi})\mathbb{P}_{H_0}[\sum_{i=1}^n Y_i > 1]$, which is linear with τ. This illustrates the nonnegligible effect of the pool correlation τ on the likelihood of getting several contaminated samples.

OBSERVATION 3 (Expectation of the number of positive samples in infected pools in the low prevalence regime). Conditional on the pool being positive, the expectation of number of positive samples per pool *with correlation* can be well approximated by: $\mathbb{E}_{H_a}[\sum_{i=1}^n Y_i] = 1 + (n-1)\tau + O(n\pi)$ (proof in the Supplementary Material (Comess et al., 2022))—we note that this simplified approximation makes sense in that we expect τ to be much greater than π in low-prevalence regimes. Consequently, the mode of the number of infected samples (conditional on $\sum_{i=1}^n Y_i > 1$) is shifted from 1 under H_0 (with 95% confidence) to (roughly) $1 + (n-1)\tau$ with correlations. This further supports the observation by Barak et al. (2021) regarding the existence of positive correlations between samples: our relationship quantifies the effect and the strength of the interactions on the distribution of the number of positive samples per pool.

Our simple model thus highlights the fact that with moderate correlations between samples, observing more than one sample becomes in fact highly probable—whereas it is unlikely with high probability (0.95) under the i.i.d. null.

1.2 Modeling the Hitchhiker Effect

Armed with the previous set of observations, we now turn to the quantification of the "hitchhiker effect" observed by Barak et al. (2021)—a phenomenon whereby the detection of weakly positive samples borrows strength from strongly positive samples in the pool, thus increasing their chances of detection. This requires us to model the influence of the number of contaminated samples in the pool on the viral load present in the sample, as detailed in the generative model presented in Figure 2.

For the purpose of our study, we consider a reverse transcription polymerase chain reaction (RT-PCR) test, as it has been an FDA-approved gold standard for SARS-

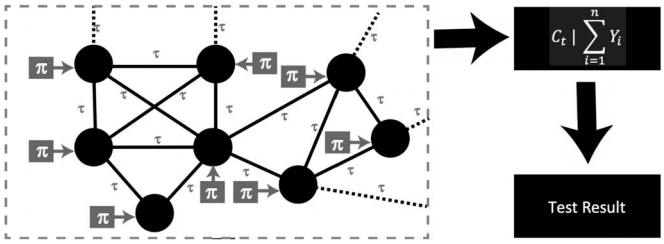


FIG. 2. *Graphical model for the procedure: the probability of a positive pooled test is a function of the number of infected samples (through community and network transmission) through the pooled C_t value.*

CoV-2 detection throughout the pandemic, due to its generally high sensitivity (probability of detecting true positive cases) and specificity (probability of detecting true negative cases) (Giri et al., 2021, FDA, 2020). RT-PCR identifies the presence of even small quantities of SARS-CoV-2 genetic material by increasing the amount of viral nucleic acid present through repeated "amplification cycles" (Public Health England, 2020). The amount of viral material in a sample (or pool of samples) after RT-PCR is quantified using the *cycle threshold* (C_t) of the sample, which measures the number of cycles in an RT-PCR assay needed to amplify viral RNA to detectable levels (Public Health England, 2020). Specifically, the C_t value refers to the number of cycles needed to amplify viral RNA to reach a fixed background level T of fluorescence at which the diagnostic result of the real-time PCR changes from negative (not detectable) to positive (detectable) (Tom and Mina, 2020). Since at every cycle the amount of viral RNA is (roughly) doubled, the C_t value is thus an indirect measure of the viral load $n^{(0)}$ in the sample

$$T = n^{(0)} 2^{C_t} \implies C_t = \log_2(T/n^{(0)}).$$

Effect of Dilution on C_t *Values.* The amount of viral material determines the sensitivity of the test: lowering the initial amount of viral material in the sample $n^{(0)}$ increases the C_t value, potentially below the Limit Of Detection (LOD) and consequently decreases the sensitivity of the test. Let us begin to formalize and quantify the extent of this phenomenon. Denoting the individual samples' C_t values as $\mathrm{C}_t^{(i)}$ (corresponding to respective initial viral loads $n_i^{(0)}$), we can write the overall C_t value for the pool (denoted $\mathrm{C}_t^{(\mathrm{dilution})}$) as

$$\begin{aligned} C_t^{(\mathrm{dilution})} &= \log_2(T) - \log_2\left(\frac{\sum_{i=1}^n n_i^{(0)}}{n}\right) \\ &= \log_2(T) - \log_2\left(\frac{\sum_{i=1}^n T Y_i 2^{-C_t^{(i)}}}{n}\right) \\ &= -\log_2\left(\sum_{i=1}^n Y_i 2^{-C_t^{(i)}}\right) + \log_2(n). \end{aligned} \tag{1}$$

Consequently, as shown in equation (1), the sensitivity of the pooled test depends on the C_t values of the individual samples. Note that while the samples' infection status Y are correlated, given the Y_i's, the C_t values can themselves be considered as independent—that is, there is no evidence (at least, at the time of writing) of the value of the C_t depending on context (e.g., a sample's corresponding age, gender or genetics). To study the efficiency of this procedure, we consider the $C_t^{(\mathrm{dilution})}$ value in the following two sufficient and mutually exclusive scenarios:

1. *For* $\sum_{i=1}^n Y_i = 1$ *infected sample in the pool.* If we assume (without loss of generality) the first sample is positive, then the C_t of the dilution is distributed as $C_t^{\mathrm{dilution}} \overset{D}{=} C_t^{(1)} + \log_2(n)$: that is, the dilution translates the distribution of the C_t value by $\log_2(n)$, and the impact on the sensitivity can be computed directly by translating the sensitivity curve by $\log_2(n)$. For example, in a pool of size $n = 8$, an individual sample with with $\mathrm{C}_t = 40$ has an effective pooled C_t value of 43. Equivalently, we can consider how much greater the viral load would need to be in order to detect a positive sample in a diluted pool: if the limit of detection for an individual sample is $\mathrm{C}_t = 40$, then a pool of size n with only one positive sample is detected only if that specimen has $\mathrm{C}_t = 40 - \log_2(n)$; with $n = 8$, the positive sample must have $\mathrm{C}_t \le 37$, or for a pool size of $n = 20$, $\mathrm{C}_t \le 35$.

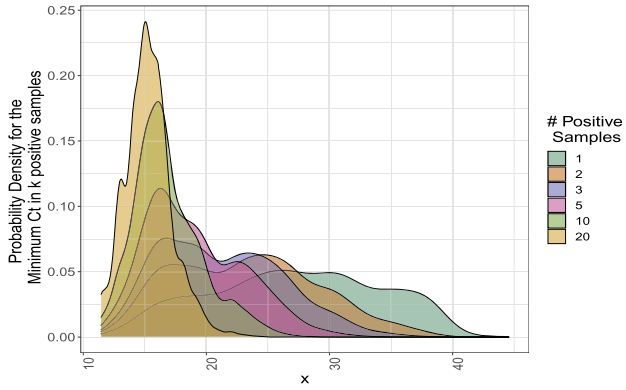
2. *For* $\sum_{i=1}^n Y_i = k \ge 2$ *infected samples in the pool.* In this case, the analysis of the C_t value of the pooled sample becomes more difficult to formalize in closed form. In this case, we resort to the following approximation to gain further insights into the hitchhiker mechanism. Assuming without loss of generality that the k first samples are positive, a Taylor expansion around the minimum C_t yields an approximation of the pooled C_t as

$$\begin{aligned} C_t^{\mathrm{pooled}} \overset{D}{\approx} &\min_{j\in[1,k]} C_t^{(j)} + \log_2(n) \\ &- \frac{1}{\log(2)} \sum_{i \neq \operatorname{argmin} C_t^{(j)}} 2^{\min_{j\in[1,k]}\{C_t^{(j)}\} - C_t^{(i)}}. \end{aligned}$$

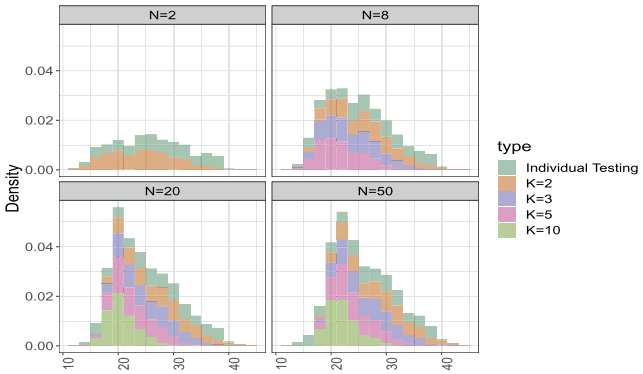
(See the Supplementary Material (Comess et al., 2022) for details.) Empirical data has shown that the distribution of the C_t values tends to be heavily skewed right, and $2^{\min_{j\in[1,k]} C_t^{(j)} - C_t^{(i)}} \ll 1$ with high probability. It then follows that the dilution C_t roughly behaves as: $C_t^{\mathrm{dilution}} \overset{D}{=} \min_{j\in[1,k]} C_t^{(j)} + \log_2(n)$. This fact is simply a formalization of the observations by Barak et al. (2021) according to which the C_t value of the pooled sample is dominated by the value of the minimum. Thus, for a weakly positive sample (high C_t) to be detected, it is sufficient for it to be combined with a strongly positive sample. We further add to this argument that as the number of positive samples k in the pool increases, we expect the distribution of the dilution to be increasingly small and eventually counterbalance the $\log_2(n)$ offset.

Applications to Reference Population. We now apply our previous findings regarding the effect of dilution on C_t value to a reference population, using data provided by Wang et al. (2021). These data consist of C_t values from naso- and oro-pharyngeal swab samples collected from both symptomatic and asymptomatic inpatients and outpatients presenting to Stanford-affiliated medical facilities in the San Francisco Bay Area, California, in the summer of 2020 (Wang et al., 2021) (details in the Appendix).

To visualize the relationship between quantity of viral material, C_t value, and probability of detection, we fit a probit curve to the dataset of Stanford samples Wang et al. (2021). We create a probabilistic model of this relationship (see the Appendix for visualization plots), where specimens with C_t greater than the limit of detection are assigned a decreasing probability of detection based on the probit regression curve. We further observe in this population that in a random sample of $k\mathrm{C}_t$ values, the distribution of the minimum C_t has a a decreasing mode and variance as k increases (Figure 3(a)). While this figure



(a) Distribution of the minimum C_t value in the Stanford data collected by Wang et al. (2021) for different numbers of positive samples. As the number of positive samples increases, the distribution of the minimum observed C_t value shifts lower.



(b) Distribution of the C_t values in our reference population under varying pool sizes and differing numbers of positive samples per pool. For as few as $k = 2$ samples, the distribution of the C_t is comparable to the distribution of individual C_t. Increasing the number of positive samples in the pool yields a C_t distribution for the dilution with a smaller mode.

FIG. 3. *Distribution of minimum C_t and $C_t^{(\mathrm{dilution})}$ in the Stanford reference population data.*

uses data from a specific population, we expect this phenomenon to generalize across other populations, since C_t values are known to exhibit important spread (Tso et al., 2021) (see the Supplementary Material (Comess et al., 2022)). This leads to the following observation for the hitchhiker effect when the independence assumption is violated (under Assumption 1 ($\pi n < 0.10$)).

OBSERVATION 4(A) (Distribution of the pooled C_t value in the i.i.d. case). For a pool size n, by the total law of probability formula, the C_t value of the dilution behaves as a mixture of n distributions:

$$C_t^{(\mathrm{pooled})} = \log_2(n) + \sum_{k=1}^{n} 1_{\sum_{i=1}^{n} Y_i = k} C_t^{(\mathrm{pooled})}[k, n],$$

where $C_t^{(\mathrm{pooled})}[k, n]$ is the C_t of the pool with k infected samples. Since in this setting, given that the pool is positive, the probability of having more than two samples is below 0.01, we have

$$C_t \overset{D}{\approx} \log_2(n) + C_t^{(1)}$$

so that the $\mathrm{C}_t^{(\mathrm{pooled})}$ value behaves roughly like a translated C_t curve. In other words, in low prevalence settings, the limit of detection for pooled samples would be $T - \log_2(n)$, and the pooled sample procedure is thus likely to miss contaminated samples with high initial individual C_t.

To quantify the hitchhiker effect under correlation, we use Monte Carlo simulations to model the behavior of the dilution's C_t. We display the results in Figure 3(b), where we show the distribution of the C_t values in our reference population under varying levels of dilution (pool sizes) and different number of positive samples per pool. Note that for as few as $k = 2$ samples, the distribution of the C_t is comparable to the distribution of individual C_t, and, for these pool sizes, having three positive samples in the pool yields a C_t distribution for the dilution with a smaller mode. This again leads to the following observation for the hitchhiker effect when the independence assumption is violated (under Assumption 1 ($\pi n < 0.10$)).

OBSERVATION 4(B) (Distribution of the pooled C_t value with network effects). Assuming positive correlations between samples, by Observation 3, we know that we expect $(n-1)\tau$ more positive samples than under the null scenario H_0. This means that the distribution of the minimum C_t is more heavily skewed towards strongly positive samples, thus ensuring a better probability of detection in the pool. In fact, we show in the Supplementary Material (Comess et al., 2022, equation (7)) how to compute the probability that a pooled C_t with K positive samples exceeds the average individual C_t.

In this population, the probability that the C_t value of the mixture is greater than the expected C_t of the individual samples is less than 0.67 for $K = 2$, and 0.45 for

$K = 3$ (assuming worse case $n = 100$). Thus, after $K = 2$ to 3 positive samples, the hitchhiker effect ensures that the C_t value of the mixture will be favorable. While this result uses data from our reference study, we expect the behaviour exhibited here to generalize across populations, as similar phenomena were empirically observed by Barak et al. (2021).

To summarize, in this section, we have shown through simple derivations, backed by simulations, that the dilution only induces a significant drop in sensitivity if a single sample is positive, but that this effect shrinks as the number of positive samples in the pool increases—thus confirming and quantifying empirical observations made by Barak et al. (2021). This allows us to conclude:

LESSON 2. *Simple models tailored to the relevant statistical assumptions are useful. The realization that correlated/clustered specimen sampling could yield such considerable gains in C_t value is only possible through a careful consideration of practical field constraints and observed measurements.*

2. EVALUATING PRACTICAL EFFICIENCY

2.1 Metrics of Interest and Benchmarks

Having studied the impact of the network transmission on the amount of viral material present in the sample, we now assess the impact of these departures from the traditional framework on our metrics of interest. To measure the efficiency of pooled testing under correlations, we introduce a set of four performance metrics, consistent with the group testing literature: the sensitivity (s), relative sensitivity (s_r) (which compares individual to pooled testing procedures), expected number of tests per sample (η), and proportion of missed cases per sample.

Sensitivity (s) and Relative Sensitivity (s_r). Sensitivity (s) is the proportion of true-positives that are detected by the test, that is, the probability of a positive test result given that there is at least one positive sample in the pool. *Sensitivity:*

$$\begin{aligned} s &= \mathbb{P}\left[\text{test is positive} \mid \sum_{i=1}^{n} Y_i \geq 1\right] \\ &= \frac{\sum_{k=1}^{n} \mathbb{P}[\text{test is positive} \cap \{\sum_{i=1}^{n} Y_i = k\}]}{\mathbb{P}[\sum_{i=1}^{n} Y_i \geq 1]} \\ &= \frac{\sum_{k=1}^{n} \mathbb{P}[\text{test is positive} \mid \sum_{i=1}^{n} Y_i = k] p_k}{\mathbb{P}[\sum_{i=1}^{n} Y_i \geq 1]}, \end{aligned}$$

with $p_k = \mathbb{P}[\sum_{i=1}^{n} Y_i = k]$. We assume that specificity (probability of a false positive) is zero.

However, tests are inherently imperfect, and such sensitivity might not be realistically achievable. As such, a more informative metric is the sensitivity of the pooled procedure, compared to individual testing (which we consider to be our "gold-standard"). Relative sensitivity serves as a comparison of the probability of a positive test result in a pooled testing scenario to an individual testing scenario which might differ as a product of the dilution effect.

Relative Sensitivity:

$$s_r = \frac{\mathbb{P}[\text{test is positive} \mid \sum_{i=1}^{n} Y_i \geq 1]}{\mathbb{P}[\text{Individual test is positive} \mid Y_1 = 1]}.$$

Contrary to the sensitivity, this measure can take values in $\mathbb{R}^+$: a relative sensitivity lower than one indicates a lower sensitivity of the pooled procedure relative to individual testing, whereas a value of s_r greater than one would indicate a better sensitivity in the case of pooled testing with respect to individual testing.

Both sensitivities can be decomposed as a function of number of positive samples per pool and C_t value as follows: $\forall x$,

$$\begin{aligned} &\mathbb{P}[s > x] \\ &\quad = \int \mathbb{P}[s > x \mid C_t] p(C_t)\, dCt \\ &\quad = \int \sum_{k=1}^{n} \mathbb{P}[s > x \mid C_t] p\left(C_t, \sum_{i=1}^{n} Y_i = k\right) dCt \\ &\quad = \sum_{k=1}^{n} \mathbb{P}\left[\sum_{i=1}^{n} Y_i = k\right] \\ &\qquad \times \underbrace{\int \mathbb{P}[s > x \mid C_t] p\left(C_t \mid \sum_{i=1}^{n} Y_i = k\right) dCt}_{= s_k, \text{ by definition}} \\ &\quad = \sum_{k=1}^{n} p_k s_k, \end{aligned}$$

where we have considered here the sensitivity of the test s_k to be a function of the number of positives (marginalized over C_t values). In light of our discussion of the effect of the network effects on the C_t value of the dilution, we conclude that the existence of correlations has a positive effect on the sensitivities.

Expected Number of Tests per Sample (η). This measures the efficiency of the pooled testing procedure. Since in the two-step Dorfman procedure every sample has to be retested if the pool is tested positive, the efficiency (assuming 100% sensitivity of the test) is

$$\begin{aligned} \eta &= \frac{1}{n}\mathbb{E}\left[1 + n\mathbb{P}[\text{test is positive}]\right] \\ &= \frac{1}{n} + \mathbb{P}[\text{test is positive}] \\ &= \frac{1}{n} + \sum_{k=1}^{n} \mathbb{P}\left[\text{test is positive} \mid \sum_{i} Y_i = k\right] p_k. \end{aligned}$$

This measure has to be compared against the benchmark value of $\eta_0 = 1$, which is the efficiency of the individual testing procedure (pool size $n = 1$). However, it is important to note that the expected number of tests per sample must be considered in conjunction with other metrics, as it is only a partial indicator of the validity of the procedure. Indeed, a faulty test which is always negative will achieve the best efficacy $\eta = \frac{1}{n}$, but with zero sensitivity.

Proportion of Missed Cases per Sample. We also consider the proportion of cases that the grouped testing procedure fails to detect per test:

$$= \frac{\sum_{k=1}^{n} k\mathbb{P}[\text{test is positive} \mid \sum_{i=1}^{n} Y_i = k](1 - p_k)}{\frac{1}{n} + \mathbb{P}[\text{test is positive}]}.$$

Simulations. To illustrate our analysis, we perform Monte Carlo simulations and calculate our metrics of interest at varying pool sizes and values of π and τ. We simulate two-stage pooled testing setting where individuals' infection statuses (positive vs. negative) are correlated, and compare this to our null model of assuming uncorrelated individuals. To make our analysis more realistic, the values of π and τ are informed by fitting Beta distributions to published literature and data. For τ, we fit Beta distributions to SAR values reported for a range of settings, including households with symptomatic index cases, households with asymptomatic index cases, transmission between spouses, healthcare settings, and from a child index case. For π, we fit Beta distributions to prevalence data at differing times over the course of the pandemic and geographic locations in the United States. The chosen time points and geographic locations are intended to be representative of varying prevalence levels and stages of the pandemic (e.g., rising cases, falling cases, etc.). Details of the methodology for fitting the distributions, as well as information on the settings and distribution parameters are described in the Appendix. Pool sizes range from n of 1 to n of 30.

To compute the sensitivity of the PCR test given a pool of size n containing $1, \ldots, n$ positive samples, we use empirically collected data on the distribution of C_t values from Wang et al. (2021), which is best represented by a Weibull distribution with shape parameter $s = 4.5$ and scale $\eta = 30$. The distribution of the C_t values depends on a number of factors, including the population tested (i.e., hospital admissions vs. general population, COVID variant, etc.). To create a realistic distribution of C_t values with the appropriate amount of spread, we sample and shift the Weibull distribution of Wang et al. (2021): we sample from their fitted distribution to create a mock distribution of individual C_t values, and shift it to model a population in which 30% of samples are above $C_t = 35$.

For each combination of π, τ, and pool size, we calculate the metrics of interest (sensitivity, relative sensitivity, expected number of tests per sample, missed cases per sample), weighted by the probability of observing k positives in that particular pool. We simulate the situation of testing a population of correlated individuals, where we either ignore correlation and erroneously treat the individuals as independent (H_0), or correctly consider networks of correlated individuals (H_a). In the null model, we make i.i.d. assumptions and the expected number of positives per pool has a binomial distribution (probability of observing K positives in n trials (pool size) with success probability π and $\tau = 0$. For the alternative model H_a, the probability of K positives is computed exactly using the Poisson Binomial distribution (Appendix).

Discussion of Simulation Results. Across all prevalence settings and levels of network transmission, the model that accounts for network transmission (correlations in pools) performs better in terms of higher sensitivity and fewer missed cases per sample than the null model (which ignores correlations) (Figure 4). Accounting for correlations between individuals can result in large percentage increases in sensitivity over the null model; for example, for spousal network transmission in both low prevalence (e.g., Maine October 2020) and high prevalence (e.g., Alabama January 2021) settings, we observe 31.25% and 19.14% increases in sensitivity, respectively, compared to the null model (Table 2).

Comparing the sensitivity of the pooled procedure to individual testing, high levels of network transmission (such as observed between spouses and in a household with a symptomatic index case) results in sensitivity greater than the individual test, and far exceeding the minimum FDA threshold (0.85). In low prevalence settings or weak network transmission (such as healthcare settings or households with an asymptomatic index case), the sensitivity of the pooled testing procedure may fall below the FDA threshold (0.85) at large pool sizes. At sufficiently high prevalence levels (such as observed in Alabama in January 2021), pools of all sizes (including as large as 30) exceed the FDA threshold for pooled testing sensitivity.

The pooled procedure also results in large decreases in the number of tests needed per sample, when compared to individual testing. Implementing pooled testing in Maine during October 2020 (low prevalence) among households with asymptomatic index cases, households with symptomatic index cases, and spouses could reduce the number of tests per sample by over 92% in all three network transmission settings (Table 2). In higher prevalence settings, reductions in testing associated with pooled testing are more modest, but still upwards of 20% (Table 2). From these results, we draw the following conclusion.

LESSON 3. *The utility of pooled testing is context dependent, but statistical models informed by observed data in a range of prevalence and network settings demonstrate that accounting for non-i.i.d. settings uniformly improves the expected performance of the procedure.*

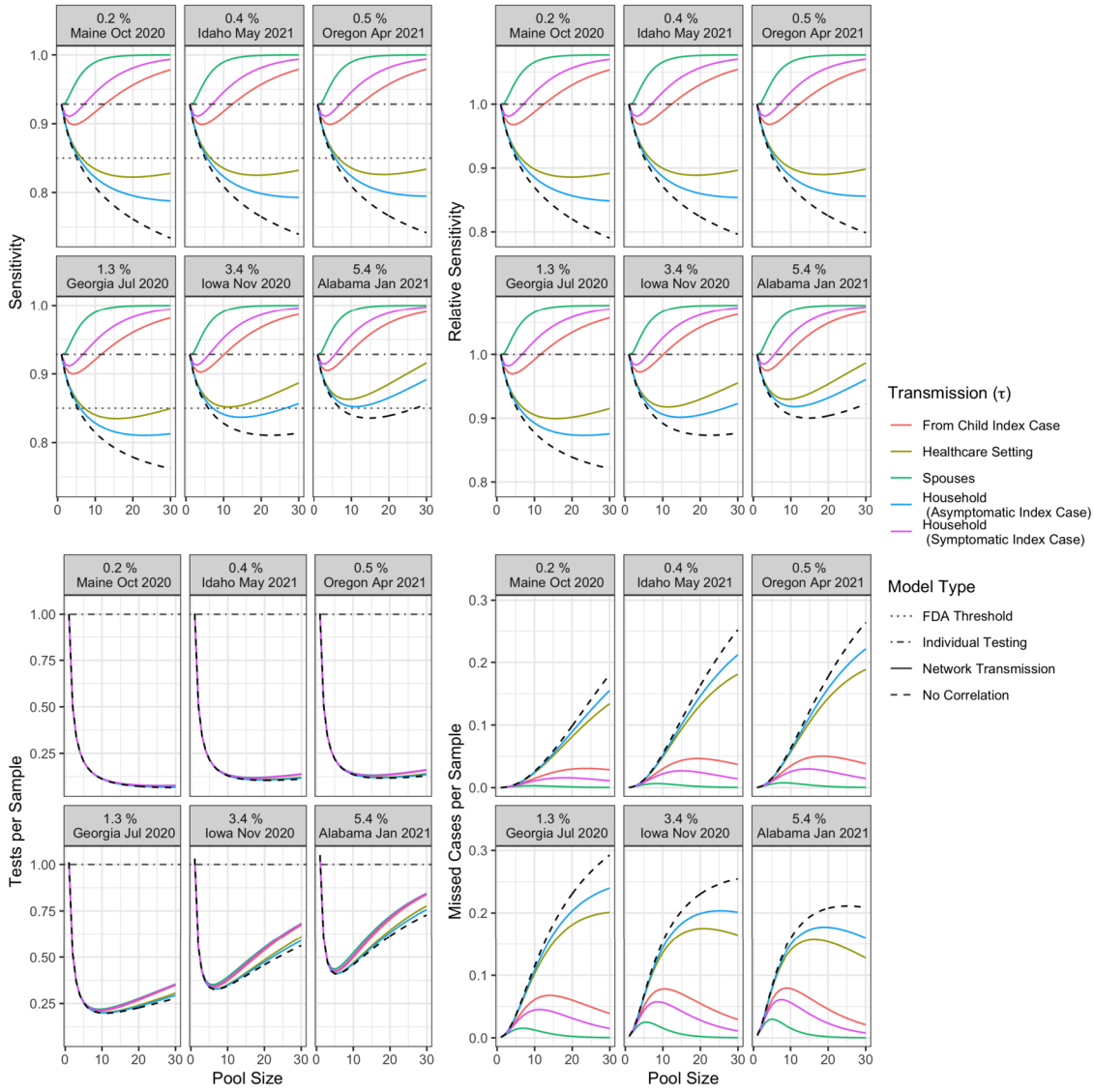


FIG. 4. Fixed Model: *Model-estimated parameters* (*sensitivity*, *relative sensitivity*, *expected tests per sample*, *missed cases per sample*) *by pool size*, *prevalence* (π), *and network transmission probability* (τ). *Null* (*no correlation*) *model*, *individual testing*, *and FDA sensitivity threshold are also indicated*, *where relevant*.

3. TESTING UNDER UNCERTAINTY

So far, our discussion has focused on situations where the prevalence and the transmission levels are fixed, known quantities. As such, all the quantities that we have computed are conditional expectations given τ, π. However, in practice, these are estimates with associated levels of uncertainty and thus can themselves be modeled as random variables, whose variability has to be taken into account as we compute metrics of interest. Given the nonlinear nature of the model and the wide uncertainty around the value of the network transmission rate (or SAR) τ, it is important to evaluate how much this added variability affects our estimates of the performance of pooled sampling. Uncertainty and heterogeneity are crucial aspects of COVID-19 kinetics that need to be accounted for to ensure accurate epidemiological predictions (Cirillo and Taleb, 2020, Cave, 2020, Gómez-Carballa et al., 2020, Zhang et al., 2020): most COVID-19 forecasting models—whether geared towards the prediction of the incidence rate, underascertainment bias, or towards the performance of pooled testing, such as the one considered in this paper—are indeed nonlinear functions of many unknown and/or highly variable quantities. When solely considering the average rather than accounting for the distributional nature of these variables, the error can rapidly amplify, and thus needs to be appropriately characterized and controlled (Donnat and Holmes, 2021). In this section, we (a) show that in the prevalence/SAR regimes that we are considering, the main driver of the heterogeneity lies in the uncertainty around network transmission and is a function of the behavior of the distribution at the tail, rather than of its variability, (b) highlight ranges of parameters (n, π, τ) which are robust to this heterogeneity,

TABLE 2
Expected performance of pooled testing for the fixed model under select prevalence, network transmission, and pool size scenarios

	Network Transmission (τ)	Prevalence (π)	*Fixed Model*	
			% Increase Sensitivity Relative to H_0	% Decrease Tests/Sample Relative to Individual Testing
Pool Size $n = 5$	Household Asympt. (0.012)	Maine Oct. 2020 (0.2%)	0.62	79.35
		Alabama Jan. 2021 (5.4%)	0.57	58.92
	Household Symptomatic (0.18)	Maine Oct. 2020 (0.2%)	7.77	79.31
		Alabama Jan. 2021 (5.4%)	7.11	57.54
	Spouses (0.38)	Maine Oct. 2020 (0.2%)	13.02	79.28
		Alabama Jan. 2021 (5.4%)	11.85	56.55
Pool Size $n = 20$	Household Asympt. (0.012)	Maine Oct. 2020 (0.2%)	4.46	92.63
		Alabama Jan. 2021 (5.4%)	3.12	36.88
	Household Symptomatic (0.18)	Maine Oct. 2020 (0.2%)	28.85	92.07
		Alabama Jan. 2021 (5.4%)	17.87	28.57
	Spouses (0.38)	Maine Oct. 2020 (0.2%)	31.25	92.02
		Alabama Jan. 2021 (5.4%)	19.14	27.86

Percent increase in sensitivity relative to $H_0 = 100 * (\text{Sens}_{H_a} - \text{Sens}_{H_0})/(\text{Sens}_{H_0})$; Percent decrease in tests per sample relative to individual testing $= 100 * (\text{Tests}_{H_a} - n)/(n)$, ($n$ = pool size).

and (c) show via experiments how to construct (and interpret) prediction intervals for the performance of pooled sampling under uncertainty.

OBSERVATION 5 (The sensitivity and efficiency (tests per sample) of the alternative model can never be worse than the null model). This is due to the fact that the (true) effective number of tests per samples can be written as

$$\eta = \frac{1 + \mathbb{P}[\sum_{i=1}^{n} Y_i > 0]}{n} = \frac{1}{n} + \frac{1 - (1-\pi)^n}{n} = \frac{1}{n} + \pi + o(\pi)$$

in both scenarios. From this formulation, we observe that the efficiency is not a function of the network transmission, and depends only on community transmission rates. As such, results on the efficiency are robust for all parameterizations and levels of uncertainty in the value of τ, but will solely depend on the uncertainty for π. This fact also highlights the necessity of having accurate estimates of the prevalence, tailored to the population at hand in order to correctly optimize pooled testing. In this context, estimates of the prevalence for the sampling population, using hyper-local data and/or additional covariates such as vaccination rates can be crucial in further reducing this uncertainty (Stevens et al., 2021, Cramer et al., 2021, Donnat et al., 2021, Zhou et al., 2020).

For the sensitivity, since correlations can only increase the number of positive samples per pool and sensitivity is an increasing function of the C_t, the sensitivity can only be improved by taking into account correlations between individuals in the alternative model.

OBSERVATION 6 (We can identify settings in which pooled sampling will have worse sensitivity than individual testing). Since under correlations, the number of positive samples per infected pool is well approximated by $1 + \text{Binomial}(n-1, \tau)$ (since with 95% probability, community transmission yields only one infected sample), the number of infected samples can vary quite substantially:

$$\begin{aligned}
&\mathbb{V}_{H_a}\left[\sum_{i=1}^{n} Y_i \mid \sum_{i=1}^{n} Y_i > 0\right] \\
&= \mathbb{E}\left[\mathbb{V}_{H_a}\left[\sum_{i=1}^{n} Y_i \mid \tau, \sum_{i=1}^{n} Y_i > 0\right]\right] \\
&\quad + \mathbb{V}\left[\mathbb{E}_{H_a}\left[\sum_{i=1}^{n} Y_i \mid \tau, \sum_{i=1}^{n} Y_i > 0\right]\right] \\
&\approx \mathbb{E}\big[(n-1)\tau(1-\tau)\big] + \mathbb{V}\big[1 + \tau(n-1)\big] \\
&\approx (n-1)(n-2)\sigma_\tau^2 + (n-1)\mu_\tau(1-\mu_\tau),
\end{aligned}$$

where μ_τ and σ_{τ^2} are respectively the mean and variance of τ. We can now examine the effect of this variance on the sensitivity. As seen in the first section, the sensitivity is impacted if the number of positive samples in infected pools falls below 2. This would mean that network transmission only accounts for a single additional infected sample in the pool. This event, which has (by property of the binomial) probability equal to $(n-1)\tau(1-\tau)^{n-2}$ can be deemed highly unlikely as long as it happens with probability less than 0.05. This allows us to solve for regions Ω of the parameter space for τ where we expect robust performances of pooled testing, which improve upon

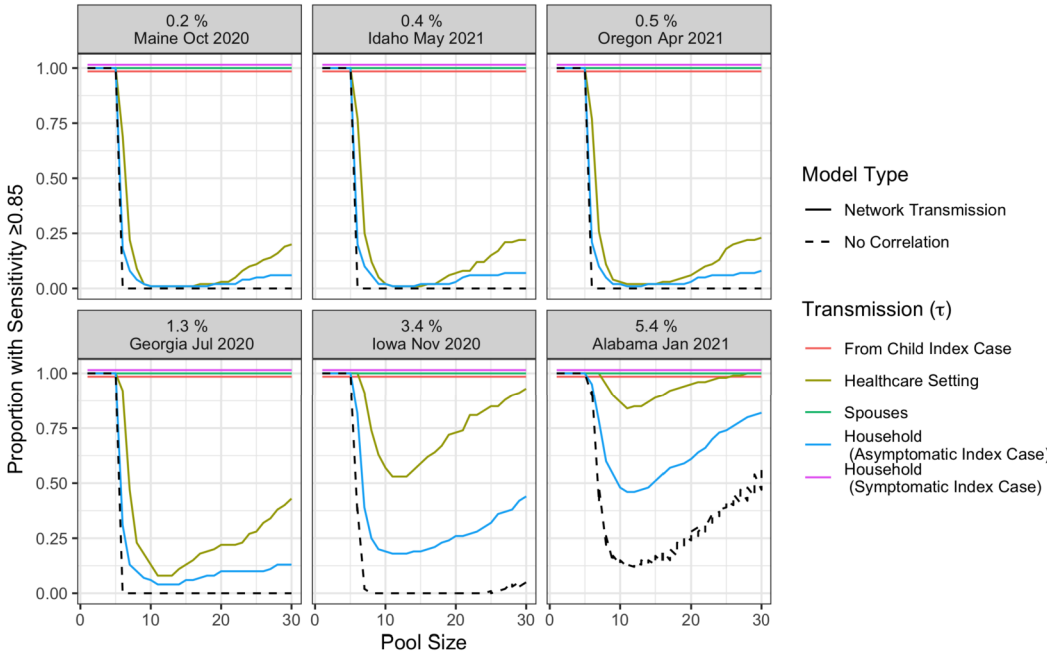


FIG. 5. *Proportion of simulated results with sensitivity equal to or greater than the FDA threshold for pooled testing* (0.85), *using the Random Prevalence/Random Network Effect Model.*

individual testing:

$$\Omega = \{\tau \in [0,1] : (n-1)\tau(1-\tau)^{n-2} \leq 0.05\}.$$

This allows the user to select a pool size such that, with probability $1 - \mathbb{P}[\tau \in \Omega]$, the procedure is better than individual testing. Choosing the value of η is setting a tolerance threshold, and should be defined by the user.

Experiments

We adapt our simulations in the previous section to account for heterogeneity and uncertainty in π and τ by performing the simulations under three further settings:

Setting 1—*Fixed* π, *Random* τ (Fixed Prevalence, Random Network Effect): Sample τ from a Beta prior distribution and calculate the corresponding probability of k positives due to network transmission. π assumed fixed and equal to the point estimate.

Setting 2—*Random* π, *Fixed* τ (Random Prevalence, Fixed Network Effect). Sample π from a Beta prior distribution. τ assumed fixed and equal to the point estimate.

Setting 3—*Random* π, *Random* τ (Random Prevalence, Random Network Effect): Sample both π and τ, as described above.

When sampling π and τ, for each point estimate and group size, we perform $B = 100$ simulations of sampling from the prior distribution.

The results of the Monte Carlo simulations for Random τ and π (Random Prevalence/Random Network Effect) are shown in Figure 6; results for random τ (Fixed Prevalence/Random Network Effect) and random π (Random Prevalence/Fixed Network Effect) are presented in the Appendix, Figures 10(a) and 10(b). These simulations inform the following observation.

LESSON 4. *These findings highlight the importance of developing* adaptive *pooled testing procedures. With reasonably specified models for prevalence and network transmission, the optimal pool size can be chosen to maximize sensitivity and ensure minimum FDA thresholds are met. However, even with significant uncertainty and heterogeneity in pools, results are robust.*

In settings with high network transmission (e.g., child index cases, spouses, and household symptomatic index cases), all of the simulation results have sensitivity greater than the FDA threshold for sensitivity (Figure 5). For low network transmission settings (e.g., healthcare or household asymptomatic index cases), pooled testing with small pools ($n \leq 5$) may still meet minimum FDA standards, but larger pools may not be appropriate (Figure 5). Supporting our previous observations, heterogeneity in results is primarily driven by uncertainty in τ, not in π.

Additionally, adaptability of the overall model is critical in a pandemic setting, where transmissibility and susceptibility vary over time and space as a function of the particular viral variants circulating and the prevalence of vaccination (and efficacy of vaccines against variants). The simulations presented in this paper can easily be adapted to settings where more transmissible variants are widespread (by increasing the value of the transmission parameter) or as vaccination rates increase (by decreasing the value of the prevalence parameter). The robustness of

results to significant uncertainty in parameter values is especially critical in this setting.

Considering the efficiency of the pooled testing procedure, there is very little variation in expected tests per sample across different specifications of the model and parameters (Figures 6; Table 3). In all cases, the pooled sampling procedure results in significant reductions in tests per sample relative to the individual sampling procedure.

4. CONCLUSION

Violations of i.i.d. assumptions and high prevalence settings have typically been considered impediments to using pooled testing; however, we demonstrate, mathematically and via simulation, that significant gains can be made in terms of testing efficiency and sensitivity by taking advantage of correlated samples and high probabilities of positive samples. Clustering multiple positive samples in a single pool improves both efficiency (by focusing retesting in a directed fashion) and sensitivity (by increasing the concentration of viral genetic material). We also investigate the effects of heterogeneity in the input parameters—specifically prevalence and network transmission—on the outcome metrics of interest. We find that our overall findings are robust to heterogeneity. This suggests that these methods can be useful even in real-world settings where precise values of prevalence and network transmission may not be known.

For example, at the peak of the pandemic in Alabama in January 2021, pooled samples of size 5 collected from households with symptomatic index cases could have reduced testing needs by 58%; at that time, the positive test rate in Alabama exceeded 20%, and greater testing capacity was urgently needed. Pooling in this setting and with this pool size is predicted to have sensitivity far exceeding individual tests, even when accounting for uncertainty in both prevalence and network transmission rates. Even greater savings can be observed in other settings, such as Maine in October 2020, a period when prevalence was low: pooled testing in pools of size 20 would have reduced testing requirements by over 90% relative to individual testing in both high and low network transmission settings. For moderately high values of network transmission (household transmission with a symptomatic index case and spouses), the sensitivity of this procedure again exceeds individual testing.

In conclusion, as a lesson learned from the pandemic, we highlight the importance of "field-aware" statistical modeling and the importance of adaptive models. To develop an actionable response to the pandemic and its unprecedented conditions, it is important to develop statistical model that accurately optimize procedures to the population at hand. This development has to be done in close collaboration with clinicians, to ensure the feasibility and scalability of the proposed solution. Here, we demonstrate that leveraging correlations in specimen collection procedures and incorporating knowledge about local prevalence and network transmission parameters can lead to better informed, logistically-feasible, and adaptive pooled testing.

APPENDIX: SIMULATIONS

A.1 Mathematical Formulations of Probability Laws

In computing the metrics of interest, we must also compute $p_k = P[\sum_{i=1}^{n} Y_i = K]$, the probability of K positives in a pool of size n. The probability law for computing this depends on whether we invoke the null or alternative (network transmission) models, that is, whether we account for correlation between members of a pool. In the null model, the probability of observing K positives in a pool of size n is given by the binomial distribution: $p_k \sim \text{Binom}(K, n)$. When independence assumptions are violated (correlated individuals), we calculate the probability of having K total positives in a pool of size n, k of which are infected in the community (where probability of infection equals prevalence, π) and $K - k$ of which are infected via network transmission (where probability of infection equals a homogeneous network transmission probability between all individuals, τ). We calculate the total probability over all possible values of k (Equation (1)).

If we allow τ to be heterogeneous between groups, but homogeneous within a group (i.e., each groups has its own τ_i) and then consider the probability of seeing K positives average over the total number of groups (n_{group}), we obtain Equation (2), where $P(\tau = \tau_i) = \frac{1}{n_{\text{groups}}}$.

Similarly, we can further account for heterogeneity in risk among individuals by allowing π to vary, reflecting the fact that individuals may have greater or less risk of being infected from the community depending on their behaviors, profession, etc. (Equation (3)), where $P(\pi = \pi_i) = \frac{1}{n_{\text{groups}}}$.

Combining equations (2) and (3) represents the model in which we account for heterogeneity in both τ and π.

A.2 C_t Value Data

We use empirically collected data on the distribution of C_t values from Wang et al. (2021). Briefly, nasopharyngeal or oropharyngeal swab specimens obtained for SARS-CoV-2 testing were obtained by the Stanford Clinical Virology Laboratory from tertiary-care academic hospitals and affiliated outpatient facilities in the San Francisco Bay Area, California, from June 10–June 19, 2020 and July 6–July 23, 2020. Samples were collected both from symptomatic and asymptomatic inpatients and outpatients, either for clinical care or via COVID-related surveillance studies and drug trials.

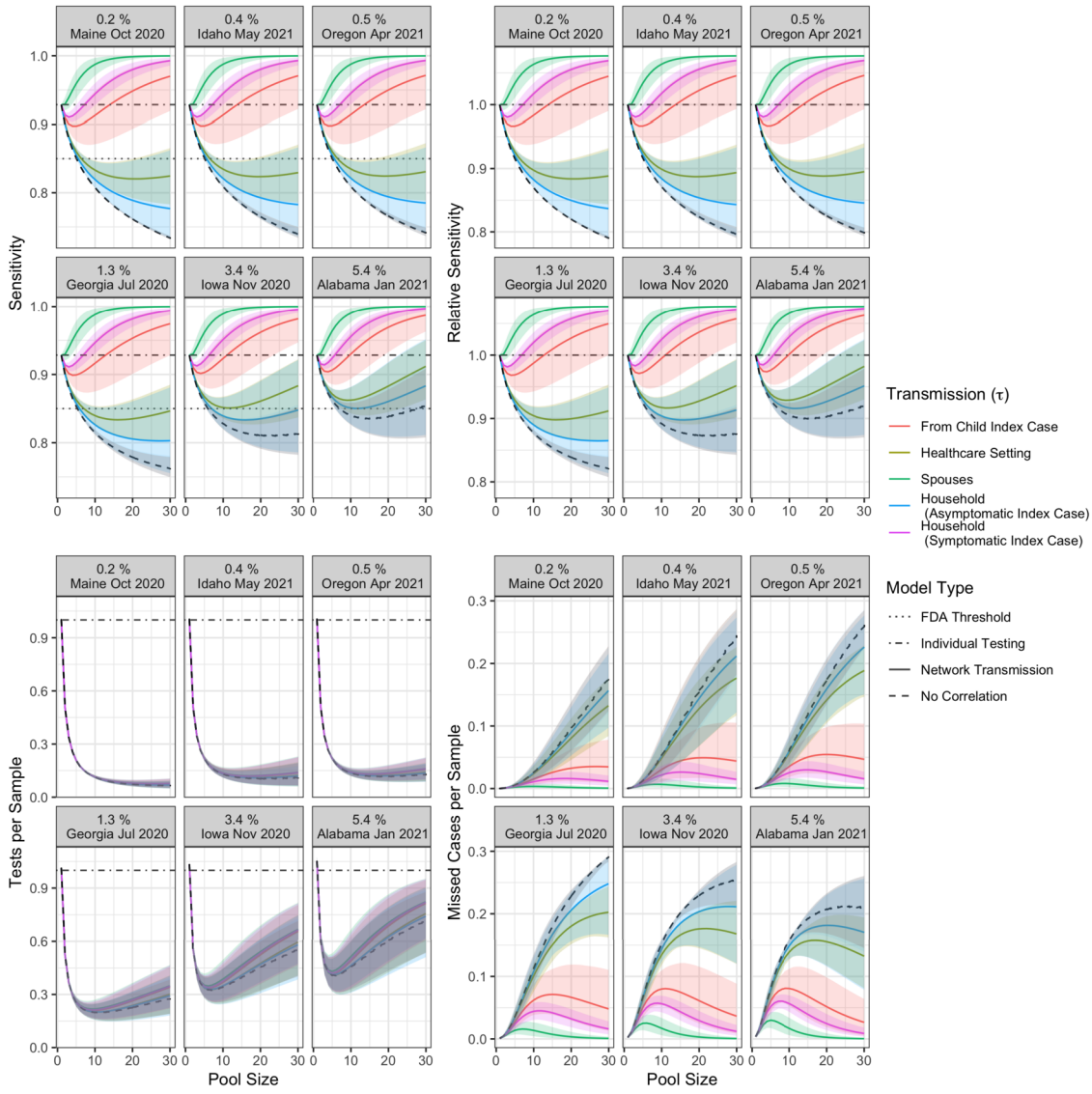


FIG. 6. Random Prevalence/Random Network Effect Model: *Model-estimated parameters* (*sensitivity*, *relative sensitivity*, *expected tests per sample*, *missed cases per sample*) *by pool size*, *prevalence* (π), *and network transmission* (τ) *when* π *and* τ *are sampled from Beta distributions corresponding to the specified scenario*. *Shaded area indicates the empirical* 95 *percent credible prediction interval*. *Proportion of samples with* C_t *above the* 95% *LoD* (*held constant at LoD* $= 35$) *is constant*, *equal to* 25%. *The null* (*no correlation*) *model*, *individual testing*, *and FDA sensitivity threshold are also indicated*, *where relevant*.

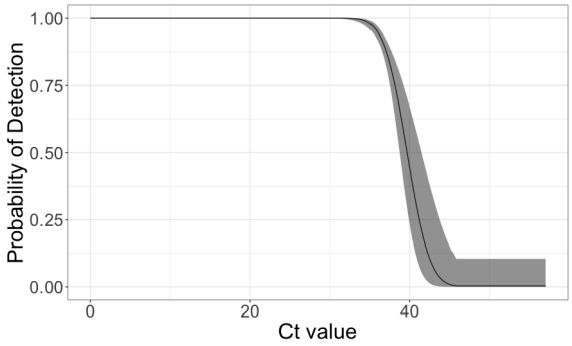


FIG. 7. *Probabilistic model for sensitivity of the test as a function of the* C_t *value in our reference population*.

The distribution of C_t values is best represented by a Weibull distribution with shape parameter $s = 4.5$ and scale $\eta = 30$. The distribution of the C_t values depends on a number of factors, including the population tested (i.e., hospital admissions vs. general population, COVID variant, etc.). To create a realistic distribution of C_t values with the appropriate amount of spread, we sample and shift the Weibull distribution of Wang et al.: we sample from their fitted distribution to create a mock distribution of individual C_t values, and shift it so that 25.0% of samples are above $C_t = 35$.

A.3 Prior Distributions on π and τ

To sample π and τ from informed priors, we fit Beta distributions to published data on prevalence and SAR,

TABLE 3
Expected performance of pooled testing under select prevalence, network transmission, model type, and pool size scenarios

			Model			
			Random Prevalence/Fixed Network Transmission		Random Prevalence/Random Network Transmission	
	Network Transmission (τ)	Prevalence (π)	% Increase Sensitivity Relative to H_0	% Decrease Tests/Sample Relative to Individual Testing	% Increase Sensitivity Relative to H_0	% Decrease Tests/Sample Relative to Individual Testing
Pool Size $n = 5$	Household Asympt. (0.012)	Maine Oct. 2020 (0.2%)	0.52	79.34	0.58	79.36
		Alabama Jan. 2021 (5.4%)	0.50	58.77	0.53	58.92
	Household Symptomatic (0.18)	Maine Oct. 2020 (0.2%)	7.71	79.31	7.86	79.31
		Alabama Jan. 2021 (5.4%)	7.04	58.00	7.19	57.52
	Spouses (0.38)	Maine Oct. 2020 (0.2%)	12.8	79.28	13.09	79.27
		Alabama Jan. 2021 (5.4%)	11.67	56.94	11.92	56.54
Pool Size $n = 20$	Household Asympt. (0.012)	Maine Oct. 2020 (0.2%)	3.62	92.60	3.94	92.64
		Alabama Jan. 2021 (5.4%)	2.55	37.74	2.71	37.12
	Household Symptomatic (0.18)	Maine Oct. 2020 (0.2%)	28.69	92.07	28.86	92.07
		Alabama Jan. 2021 (5.4%)	17.76	30.51	17.88	28.57
	Spouses (0.38)	Maine Oct. 2020 (0.2%)	31.17	92.03	31.20	92.02
		Alabama Jan. 2021 (5.4%)	19.09	29.70	19.11	27.87

Percent increase in sensitivity relative to $H_0 = 100 \times (\text{Sens}_{H_a} - \text{Sens}_{H_0})/(\text{Sens}_{H_0})$; Percent decrease in tests per sample relative to individual testing $= 100 \times (\text{Tests}_{H_a} - n)/(n)$, where $n =$ pool size.

TABLE 4
Probability laws for number of positive samples in a pool

Scenario	Equation	Reference
Homogeneous τ, π	$P(\sum Y_i = K) = \sum_{k=1}^{K}\left(\binom{n}{k}\pi^k(1-\pi)^{n-k}\binom{n-k}{K-k}(1-(1-\tau)^k)^{K-k}((1-\tau)^k)^{(n-K)}\right)$	Eq. (1)
Heterogeneous τ	$P(\sum Y_i = K) = \sum_{k=1}^{K}\left(\binom{n}{k}\pi^k(1-\pi)^{n-k}\sum_{\tau_i=1}^{n_{\text{groups}}} P(\tau=\tau_i)\binom{n-k}{K-k}(1-(1-\tau_i)^k)^{K-k}((1-\tau_i)^k)^{(n-K)}\right)$	Eq. (2)
Heterogeneous π	$P(\sum Y_i = K) = \sum_{k=1}^{K}\left(\sum_{pi_i=1}^{n_{\text{groups}}} P(\pi=\pi_i)\binom{n}{k}\pi_i^k(1-\pi_i)^{n-k}\binom{n-k}{K-k}(1-(1-\tau_i)^k)^{K-k}((1-\tau_i)^k)^{(n-K)}\right)$	Eq. (3)

TABLE 5
Parameters for select Beta prior distributions for community prevalence π in different settings. The Beta parameters (α, β) are estimated from the 95% confidence intervals of the mean estimated active cases in a given state and month, obtained from covidestim.org

Setting	Mean Monthly Prevalence (95% CI)	Beta(α, β)
Georgia, July 2020	1.3% (0.7, 2.0)	Beta(16.67, 1282.88)
Maine, October 2020	0.2% (0.07, 0.3)	Beta(9.94, 6561.33)
Iowa, November 2020	3.4% (2.0, 5.2)	Beta(16.99, 477.12)
Alabama, January 2021	5.4% (3.0, 8.4)	Beta(14.38, 251.01)
Oregon, April 2021	0.5% (0.2, 0.7)	Beta(13.06, 2836.41)
Idaho, May 2021	0.4% (0.1, 0.7)	Beta(5.77, 1543.33)

respectively. Beta distributions to the 95% confidence intervals of reported metrics of interest (prevalence, SAR) using the publicly available `beta.params.from.quantiles` function (Joseph and Belisle, 2017).

To estimate distributions for the community prevalence, we fit a Beta distribution to reported 95% confidence intervals on the estimated rates of COVID-19 infections over time in every U.S. state. Specifically, we use the estimate of true number of infections, which is adjusted for reporting delays and potential under-counting, provided by https://covidestim.org/. The methodology for adjusting case counts is described in a preprint paper by the authors (Chitwood et al., 2021). Data extend from the first reported case (January 13, 2020) through present (data downloaded 28 May 2021). We only consider states for which 95% CI data are available, and then further sub-

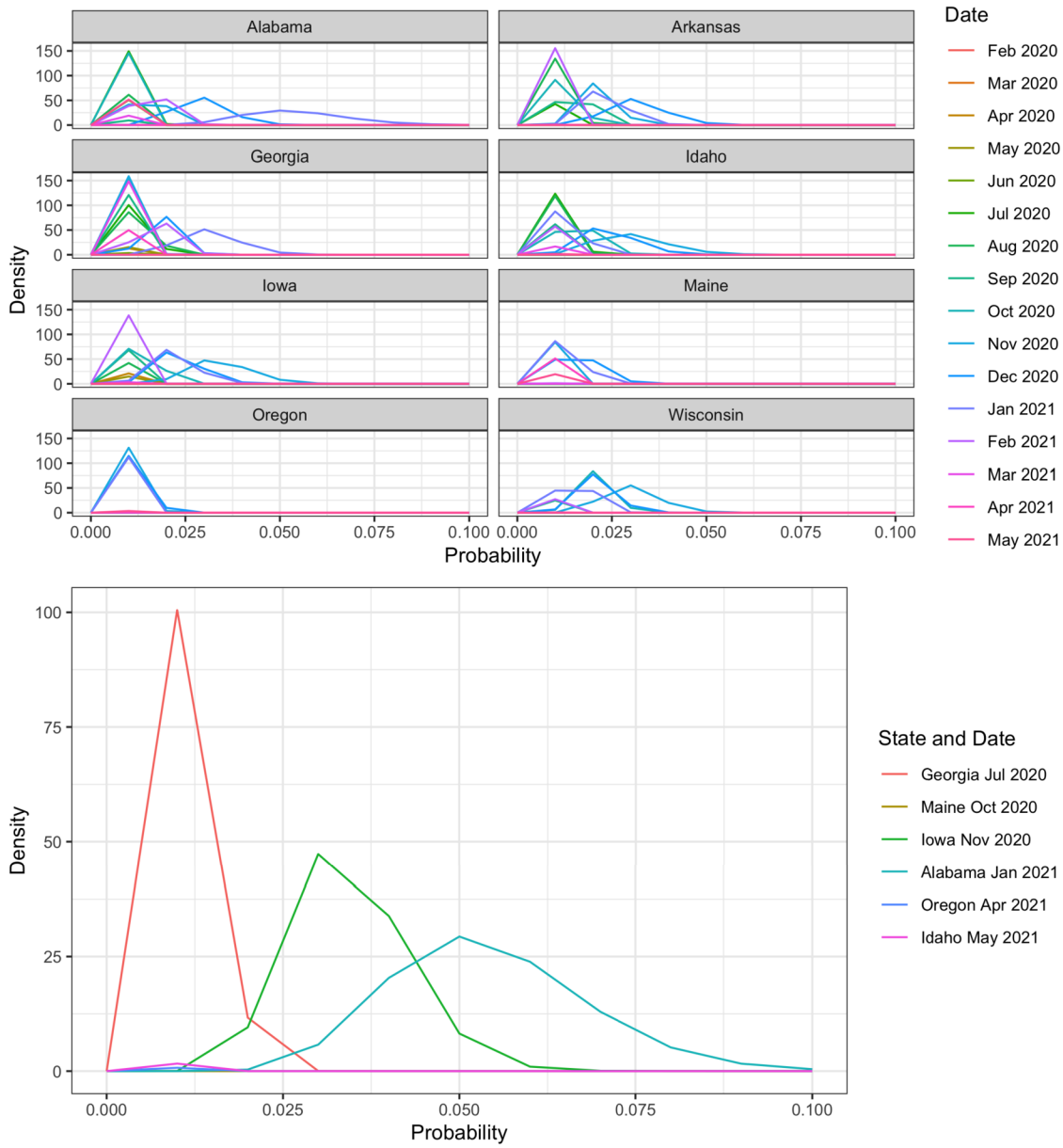


FIG. 8. *Density plots of the Beta prior distributions for π, corresponding to all fitted models (top) and the parameters and settings selected for the simulation study and included in Table 5.*

set to consider a single representative state from each geographic division as defined by the U.S. census bureau. Prior to fitting the distribution, we normalize raw infection counts by the state population for 2019 (Census. gov, 2021) and sum over the past ten days of data to estimate active cases. Finally, we average the number of active cases by month for each state. After fitting Beta distributions to the resulting data, we observed that many prior distributions have very similar α, β parameters. Thus, we select six representative distributions corresponding to different stages in the pandemic (e.g., large surge, small surge, peak of surge, low cases, declining cases), and varying time points and geographic regions. The selected time points and regions, as well as corresponding fitted distributions, are presented in Table 5; density plots for all fitted distributions and those selected for use in the simulations are presented in Figure 8.

To estimate distributions for network transmission, we followed a similar procedure as described above for community prevalence. Beta distributions were fit to the 95% confidence intervals of SAR estimates reported in published meta-analyses for different settings. Again, many fitted distributions were similarly specified and so six representative distributions were selected (Figure 9). The settings and corresponding parameters of the selected distributions are presented in Table 6; density plots of all fitted distributions and the representative selected distributions for use in the simulations are presented in Figure 9.

TABLE 6

Parameters for select Beta prior distributions for network transmission τ in select settings. The Beta parameters (α, β) are estimated from the 95% confidence intervals of secondary attack rate (SAR) values in the literature

Setting	SAR (95% CI)	Beta(α, β)	Citation
Child Index Case	13.40% (5.7–21.1)	Beta(8.38, 59.43)	Spielberger et al. (2021)
Healthcare Setting	0.7% (0.4–1.0)	Beta(8.3, 359.61)	Koh et al. (2020)
Household (Spouses)	37.8% (25.8–50.5)	Beta(21.78, 35.92)	Madewell et al. (2020)
Household (Asymptomatic Index Case)	0.7% (0–4.9)	Beta(0.74, 62.23)	Madewell et al. (2020)
Household (Symptomatic Index Case)	18.0% (14.2–22.1)	Beta(64.95, 296.26)	Madewell et al. (2020)
Household (General)	30% (0–67)	Beta(0.45, 2.37)	Curmei et al. (2020)

A.4 Implementing Simulations

We perform simulations under four distinct settings:

- Fixed (point estimates of π and τ): Deterministic π and τ, probability calculated as in Equation (1)
- τ Graph Effect (Random τ, Fixed π): Sample τ from a Beta prior, probability calculated as in Equation (2)
- π Graph Effect (Random π, Fixed τ): Sample π from a Beta prior, probability calculated as in Equation (3)
- All (τ and π) Graph Effect (Random τ, Random π): Sample both π and τ, combining Equations (2) and (3).

Results for All Graph Effect and Fixed model are presented in the main paper; results for the τ Graph Effect and π Graph Effect are presented in Figures 10(a) and 10(b).

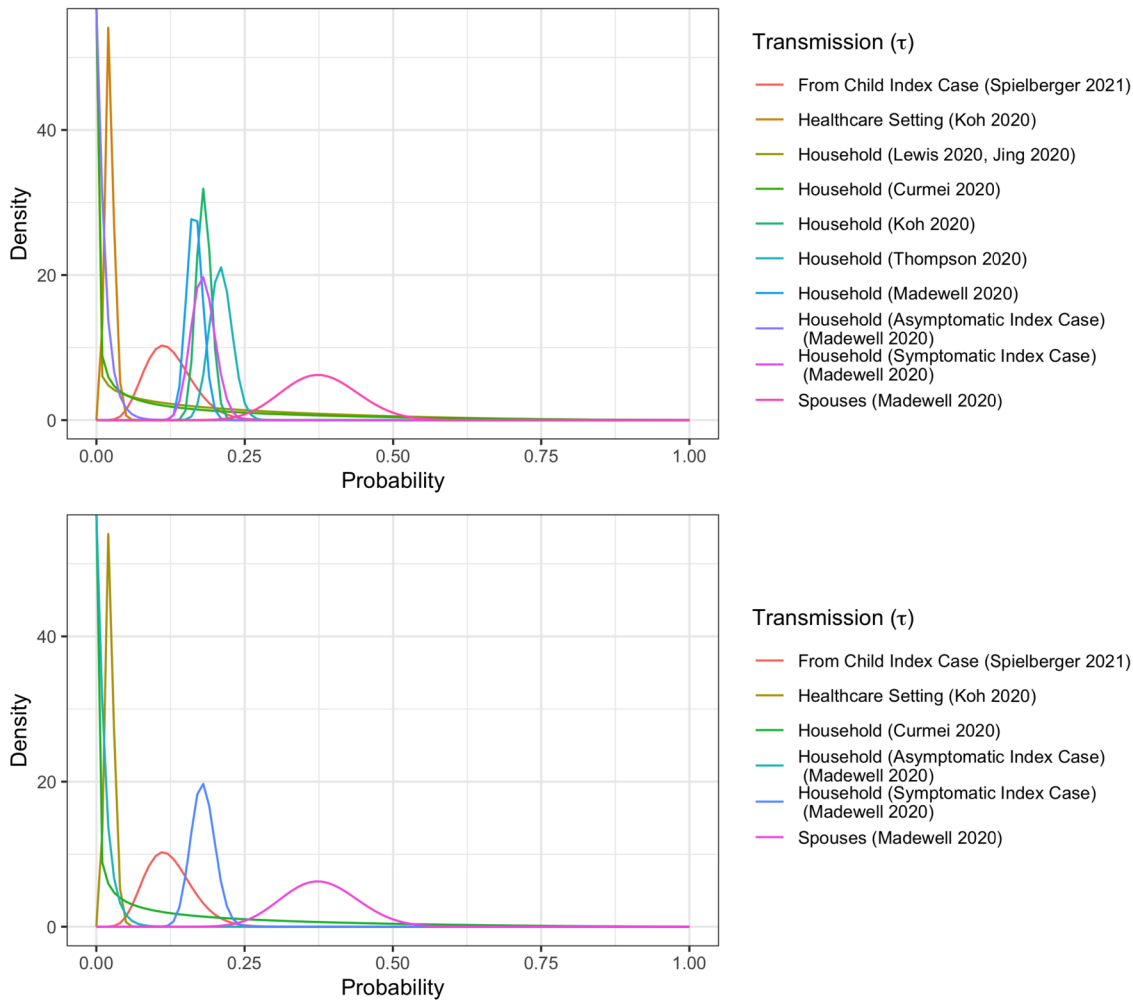
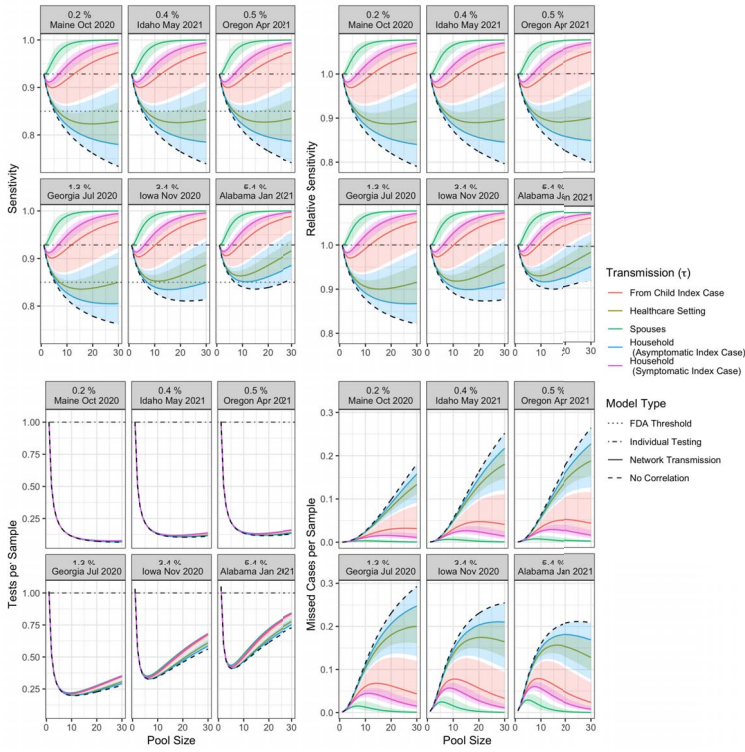
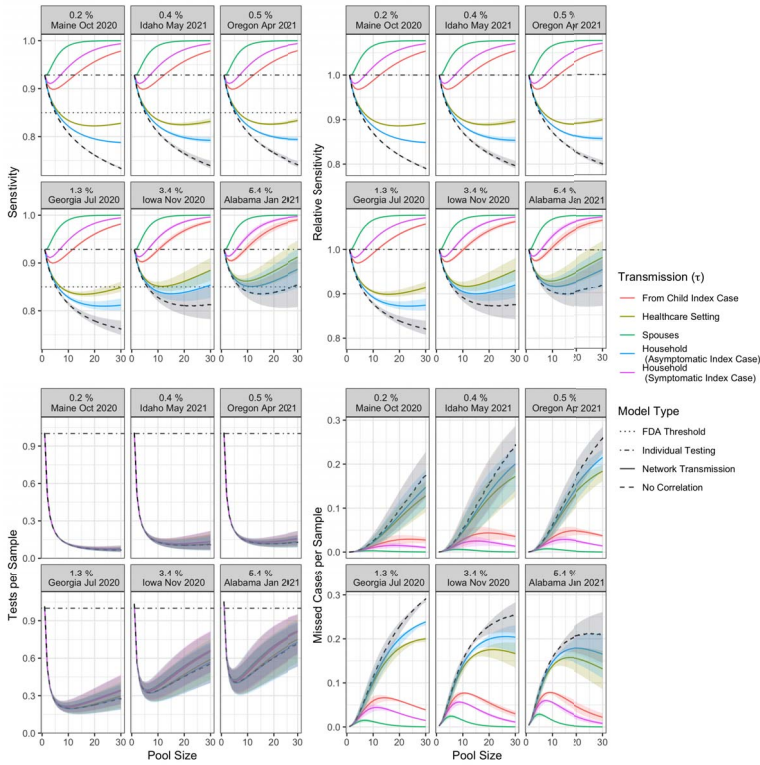


FIG. 9. *Density plots of the Beta prior distributions for τ, corresponding to all fitted models (top) and the parameters and settings selected for the simulation study and included in Table 6.*



(a) Random Network Effect/Fixed Prevalence Model: Tau Graph Effect model, where τ is sampled from and π is held constant at the mean of the respective Beta distributions for the specified scenario.



(b) Random Prevalence/Fixed Network Effect Model: Pi Graph Effect model, where π is sampled from and τ is held constant at the mean of the respective Beta distributions for the specified scenario.

FIG. 10. *Model-estimated parameters* (*sensitivity*, *relative sensitivity*, *expected tests per sample*, *missed cases per sample*) *by pool size*, *prevalence* (π), *and network transmission probability* (τ). *Shaded area indicates the empirical* 95% *credible prediction interval. Proportion of samples with* C_t *above the* 95% *LoD* (*held constant at LoD* $= 35$) *is constant*, *equal to* 25%. *The null* (*no correlation*) *model*, *individual testing*, *and FDA sensitivity threshold are also indicated*, *where relevant*.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Benjamin Pinsky (Associate Professor of Pathology and Medicine at Stanford University) for sharing data and for his feedback on the manuscript.

SUPPLEMENTARY MATERIAL

Supplement to "Statistical Modeling for Practical Pooled Testing During the COVID-19 Pandemic" (DOI: 10.1214/22-STS857SUPP; .pdf). Supplementary calculations and derivations.

REFERENCES

ABDALHAMID, B., BILDER, C. R., MCCUTCHEN, E. L., HINRICHS, S. H., KOEPSELL, S. A. and IWEN, P. C. (2020). Assessment of specimen pooling to conserve SARS CoV-2 testing resources. *Am. J. Clin. Pathol.* **153** 715–718. https://doi.org/10.1093/ajcp/aqaa064

ADAM, D. C., WU, P., WONG, J. Y., LAU, E. H. Y., TSANG, T. K., CAUCHEMEZ, S., LEUNG, G. M. and COWLING, B. J. (2020). Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat. Med.* **26** 1714–1719.

ASM (2021). Supply Shortages Impacting COVID-19 and Non-COVID Testing. [Online; posted 19-January-2021].

BARAK, N., AMI, R. B., SIDO, T., PERRI, A., SHTOYER, A., RIVKIN, M., LICHT, T., PERETZ, A., MAGENHEIM, J. et al. (2021). Lessons from applied large-scale pooling of 133,816 SARS-CoV-2 RT-PCR tests. *Sci. Transl. Med.* **13** 1–8.

BERKE, E. M., NEWMAN, L. M., JEMSBY, S., HYDE, B., BHALLA, N., SHEILS, N. E., OOMMAN, N., REPPAS, J., VERMA, P. et al. (2021). Pooling in a pod: A strategy for COVID-19 testing to facilitate a safe return to school. *Public Health Reports* **136** 663–670.

BILDER, C. R. and TEBBS, J. M. (2012). Pooled-testing procedures for screening high volume clinical specimens in heterogeneous populations. *Stat. Med.* **31** 3261–3268. MR3041806 https://doi.org/10.1002/sim.5334

BILDER, C. R., TEBBS, J. M. and CHEN, P. (2010). Informative retesting. *J. Amer. Statist. Assoc.* **105** 942–955. MR2752591 https://doi.org/10.1198/jasa.2010.ap09231

BILDER, C. R., TEBBS, J. M. and MCMAHAN, C. S. (2019). Informative group testing for multiplex assays. *Biometrics* **75** 278–288. MR3953728 https://doi.org/10.1111/biom.12988

CAPS (2020). Pooled Testing: Guidance from the CAP's Microbiology Committee. Available at www.cap.org/covid-19/pooled-testing-guidance-from-cap-microbiology-committee. Accessed: 2021-01-06.

CAVE, E. (2020). COVID-19 super-spreaders: Definitional quandaries and implications. *Asian Bioethics Review* 1.

CENSUS. GOV (2021). State Population Totals and Components of Change: 2010–2019. Available at www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html.

CHEN, P., TEBBS, J. M. and BILDER, C. R. (2009). Group testing regression models with fixed and random effects. *Biometrics* **65** 1270–1278. MR2756515 https://doi.org/10.1111/j.1541-0420.2008.01183.x

CHITWOOD, M. H., RUSSI, M., GUNASEKERA, K., HAVUMAKI, J., KLAASSEN, F., PITZER, V. E., SALOMON, J. A., SWARTWOOD, N. A., WARREN, J. L. et al. (2021). Reconstructing the course of the COVID-19 epidemic over 2020 for US states and counties: Results of a Bayesian evidence synthesis model. *MedRxiv*. https://doi.org/10.1101/2020.06.17.20133983

CIRILLO, P. and TALEB, N. N. (2020). Tail risk of contagious diseases. *Nat. Phys.* **16** 606–613.

COLLINS, K. (2020). Is Your State Doing Enough Coronavirus Testing? Available at www.nytimes.com/interactive/2020/us/coronavirus-testing.html. [Online; posted 1-November-2020].

COMESS, S., WANG, H., HOLMES, S. and DONNAT, C. (2022). Supplement to "Statistical Modeling for Practical Pooled Testing During the COVID-19 Pandemic." https://doi.org/10.1214/22-STS857SUPP

CRAMER, E. Y., LOPEZ, V. K., NIEMI, J., GEORGE, G. E., CEGAN, J. C., DETTWILLER, I. D., ENGLAND, W. P., FARTHING, M. W., HUNTER, R. H. et al. (2021). Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the US. *MedRxiv* 2021.02.03.21250974.

CURMEI, M., ILYAS, A., EVANS, O. and STEINHARDT, J. (2020). Estimating household transmission of SARS-CoV-2. *medRxiv* 1–24. https://doi.org/10.1101/2020.05.23.20111559

DECKERT, A., BÄRNIGHAUSEN, T. and KYEI, N. N. (2020). Simulation of pooled-sample analysis strategies for COVID-19 mass testing. *Bull. World Health Organ.* **98** 590–598. https://doi.org/10.2471/BLT.20.257188

DHILLON, R. S., SRIKRISHNA, D., GARRY, R. F. and CHOWELL, G. (2015). Ebola control: Rapid diagnostic testing. *Lancet Infect. Dis.* **15** 147–148.

DONNAT, C. and HOLMES, S. (2021). Modeling the heterogeneity in COVID-19's reproductive number and its impact on predictive scenarios. *J. Appl. Stat.* 1–29.

DONNAT, C., MIOLANE, N., BUNBURY, F. and KREINDLER, J. (2020). A Bayesian Hierarchical Network for Combining Heterogeneous Data Sources in Medical Diagnoses. Available at arXiv:2007.13847.

DONNAT, C., BUNBURY, F., KREINDLER, J., FILIPPIDIS, F. T., EL-OSTA, A., ESKO, T. and HARRIS, M. (2021). A Predictive Modelling Framework for COVID-19 Transmission to Inform the Management of Mass Events. *MedRxiv* 2021.05.13.21256857.

DORFMAN, R. (1943). The detection of defective members of large populations. *Ann. Math. Stat.* **14** 436–440.

EUNJUNG CHA, A. (2021). The future of coronavirus testing is in Greenville, N.C. Available at www.washingtonpost.com/health/2021/04/28/new-coronavirus-testing-strategy-home-kits/. [Online; posted 28-April-2021].

FDA (2020). Coronavirus (COVID-19) Update: FDA Issues First Emergency Authorization for Sample Pooling in Diagnostic Testing. Available at www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-issues-first-emergency-authorization-sample-pooling-diagnostic. [Online; posted 18-July-2020].

FRAYER, L. (2021). Oxygen Rationing, Test Shortages: India Caught Unprepared In COVID-19 Crisis. Available at www.npr.org/2021/04/24/990544555/oxygen-rationing-test-shortages-india-caught-unprepared-in-covid-19-crisis. [Online; posted 24-April-2021].

GANDHI, M., YOKOE, D. S. and HAVLIR, D. V. (2020). Asymptomatic transmission, the Achilles' heel of current strategies to control COVID-19. *N. Engl. J. Med.* **382** 36–39.

GASTWIRTH, J. L. (2000). The efficiency of pooling in the detection of rare mutations. *Am. J. Hum. Genet.* **67** 1036–1039.

GAYDOS, C. A. (2005). Nucleic acid amplification tests for gonorrhea and chlamydia: Practice and applications. *Infectious Disease Clinics* **19** 367–386.

GIRI, B., PANDEY, S., SHRESTHA, R., POKHAREL, K., LIGLER, F. S. and NEUPANE, B. B. (2021). Review of analytical performance of COVID-19 detection methods. *Anal. Bioanal. Chem.* **413** 35–48. https://doi.org/10.1007/s00216-020-02889-x

GÓMEZ-CARBALLA, A., BELLO, X., PARDO-SECO, J., MARTINÓN-TORRES, F. and SALAS, A. (2020). Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders. *Genome Res.* **30** 1434–1448. https://doi.org/10.1101/gr.266221.120

HE, J., GUO, Y., MAO, R. and ZHANG, J. (2020). Proportion of asymptomatic coronavirus disease 2019: A systematic review and meta-analysis. *Journal of Medical Virology*.

HEILWEIL, R. (2021). How omicron broke Covid-19 testing. Available at www.vox.com/recode/2021/12/21/22848286/omicron-rapid-test-covid-19-antigen. [Online; posted 21 December 2021].

JOACHIM, A., DEWALD, F., SUÁREZ, I., ZEMLIN, M., LANG, I., STUTZ, R., MARTHALER, A., BOSSE, H. M., LÜBKE, N. et al. (2021). Pooled RT-qPCR testing for SARS-CoV-2 surveillance in schools—a cluster randomised trial. *EClinicalMedicine* **39** 101082.

JONES, C. (2021). 'Pool testing' to combat Covid on campus grows popular in California schools. Available at edsource.org/2021/pool-testing-to-combat-covid-on-campus-grows-popular-in-california-schools/661144. [Online; posted 15 September 2021].

JOSEPH, L. and BELISLE, P. (2017). Version 1.3 beta.parms.from.quantiles [R] Computing Beta distribution parameters. Available at www.medicine.mcgill.ca/epidemiology/Joseph/PBelisle/BetaParmsFromQuantiles.html. [Online; updated February 2017].

KIM, H.-Y., HUDGENS, M. G., DREYFUSS, J. M., WESTREICH, D. J. and PILCHER, C. D. (2007). Comparison of group testing algorithms for case identification in the presence of test error. *Biometrics* **63** 1152–1163, 1313. MR2414593 https://doi.org/10.1111/j.1541-0420.2007.00817.x

KOH, W. C., NAING, L., CHAW, L., ROSLEDZANA, M. A., ALIKHAN, M. F., JAMALUDIN, S. A., AMIN, F., OMAR, A., SHAZLI, A. et al. (2020). What do we know about SARS-CoV-2 transmission? A systematic review and meta-analysis of the secondary attack rate and associated risk factors. *PLoS ONE* **15** 1–23.

LARREMORE, D. B., WILDER, B., LESTER, E., SHEHATA, S., BURKE, J. M., HAY, J. A., TAMBE, M., MINA, M. J. and PARKER, R. (2020). Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening. *Sci. Adv.*

LIN, Y.-J., YU, C.-H., LIU, T.-H., CHANG, C.-S. and CHEN, W.-T. (2020). Positively Correlated Samples Save Pooled Testing Costs. Available at arXiv:2011.09794.

MA, Q., LIU, J., LIU, Q., KANG, L., LIU, R., JING, W., WU, Y. and LIU, M. (2021). Global percentage of asymptomatic SARS-CoV-2 infections among the tested population and individuals with confirmed COVID-19 diagnosis: A systematic review and meta-analysis. *JAMA Network Open* **4**.

MADEWELL, Z.J., YANG, Y., LONGINI, I.M., HALLORAN, M.E. and DEAN, N.E. (2020). Household transmission of SARS-CoV-2: A systematic review and meta-analysis. *JAMA Network Open* **3** e2031756. https://doi.org/10.1001/jamanetworkopen.2020.31756

MASS. GOV (2021). Baker-Polito Administration's First in the Nation COVID-19 Pooled Testing Initiative Finds 0.7% Positivity Rate in Schools Throughout Commonwealth. Available at www.mass.gov/news/baker-polito-administrations-first-in-the-nation-covid-19-pooled-testing-initiative-finds-07-positivity-rate-in-schools-throughout-commonwealth. [Online; posted 29 March 2021].

MASSACHUSSETS, DEPARTMENT OF EDUCATION (2022). COVID-19 Testing Program. Available at www.doe.mass.edu/covid19/testing/default.html. [Online; updated 5 January 2022].

MCMAHAN, C. S., TEBBS, J. M. and BILDER, C. R. (2012a). Informative Dorfman screening. *Biometrics* **68** 287–296. MR2909885 https://doi.org/10.1111/j.1541-0420.2011.01644.x

MCMAHAN, C. S., TEBBS, J. M. and BILDER, C. R. (2012b). Two-dimensional informative array testing. *Biometrics* **68** 793–804. MR3055184 https://doi.org/10.1111/j.1541-0420.2011.01726.x

MINA, M. J., PARKER, R. and LARREMORE, D. B. (2020). Rethinking Covid-19 test sensitivity—a strategy for containment. *N. Engl. J. Med.* **383** e120.

MWAI, P. (2021). Coronavirus in Africa: Concern growing over third wave of Covid-19 infections. Available at bbc.com/news/world-africa-53181555. [Online; posted 7-June-2021].

NOUVELLET, P., GARSKE, T., MILLS, H. L., NEDJATI-GILANI, G., HINSLEY, W., BLAKE, I. M., VAN KERKHOVE, M. D., CORI, A., DORIGATTI, I. et al. (2015). The role of rapid diagnostics in managing Ebola epidemics. *Nature* **528** S109–S116.

ORAN, D. P. and TOPOL, E. J. (2020). Prevalence of asymptomatic SARS-CoV-2 infection : A narrative review. *Ann. Intern. Med.* **173** 362–367. https://doi.org/10.7326/M20-3012

PIETSCH, B. (2021). More coronavirus tests will be available next month, Fauci says, as U.S. struggles with shortage. Available at www.washingtonpost.com/health/2021/12/27/omicron-covid-test-shortage-fauci/. [Online; posted 27 December 2021].

POLLOCK, A. M. and LANCASTER, J. (2020). Asymptomatic transmission of Covid-19. *BMJ* **371** m4851.

POLLOCK, N. R., BERLIN, D., SMOLE, S. C., MADOFF, L. C., BROWN, C., HENDERSON, K., LARSEN, E., HAY, J., GABRIEL, S. et al. (2021). Implementation of SARS-Cov2 screening in K-12 schools using in-school pooled molecular testing and deconvolution by rapid antigen test. *Journal of Clinical Microbiology* **59** 1–7.

PUBLIC HEALTH ENGLAND (2020). Understanding cycle threshold (Ct) in SARS-CoV-2 RT-PCR: A guide for health protection teams. Available at www.gov.uk/government/publications/cycle-threshold-ct-in-sars-cov-2-rt-pcr.

RANNAN-ELIYA, R. P., WIJEMUNIGE, N., GUNAWARDANA, J., AMARASINGHE, S. N., SIVAGNANAM, I., FONSEKA, S., KAPUGE, Y. and SIGERA, C. P. (2021). Increased intensity of PCR testing reduced COVID-19 transmission within countries during the first pandemic wave: Study examines increased intensity of reverse transcription–polymerase chain reaction (PCR) testing and its impact on COVID-19 transmission. *Health Aff.* 10–1377.

REWLEY, J. (2020). Specimen pooling to conserve additional testing resources when persons' infection status is correlated: A simulation study. *Epidemiology* **31** 832–835. https://doi.org/10.1097/EDE.0000000000001244

RITCHIE, H., MATHIEU, E., RODÉS-GUIRAO, L., APPEL, C., GIATTINO, C., ORTIZ-OSPINA, E., HASELL, J., MACDONALD, B., BELTEKIAN, D. et al. (2020). Coronavirus pandemic (COVID-19). *Our World in Data*. ourworldindata.org/coronavirus.

SIMAS, A. M., CROTT, J. W., SEDORE, C., ROHRBACH, A., MONACO, A. P., GABRIEL, S. B., LENNON, N., BLUMENSTIEL, B. and GENCO, C. A. (2021). Pooling for SARS-CoV2 surveillance: Validation and strategy for implementation in K-12 schools. *Frontiers in Public Health* **9** 1–7.

SPIELBERGER, B. D., GOERNE, T., GEWENIGER, A., HENNEKE, P. and ELLING, R. (2021). Intra-household and close-contact SARS-CoV-2 transmission among children—A systematic review. *Frontiers in Pediatrics* **9**. https://doi.org/10.3389/fped.2021.613292

STEVENS, J. P., HORNG, S., O'DONOGHUE, A., MORAVICK, S. and WEISS, A. (2021). How one Boston hospital built a Covid-19 forecasting system. *Harvard Business Review*.

TOM, M. R. and MINA, M. J. (2020). To interpret the SARS-CoV-2 test, consider the cycle threshold value. *Clin. Infect. Dis.* **71** 2252–2254. https://doi.org/10.1093/cid/ciaa619

TSO, C. F., GARIKIPATI, A., GREEN-SAXENA, A., MAO, Q. and DAS, R. (2021). Correlation of population SARS-CoV-2 cycle threshold values to local disease dynamics: Exploratory observational study. *JMIR Public Health and Surveillance* **7**.

TU, X. M., LITVAK, E. and PAGANO, M. (1995). On the informativeness and accuracy of pooled testing in estimating prevalence of a rare disease: Application to HIV screening. *Biometrika* **82** 287–297. MR1354229 https://doi.org/10.1093/biomet/82.2.287

WANG, H., HOGAN, C. A., MILLER, J. A., SAHOO, M. K., HUANG, C. H., MFUH, K. O., SIBAI, M., ZEHNDER, J., HICKEY, B. et al. (2021). Performance of nucleic acid amplification tests for detection of severe acute respiratory syndrome coronavirus 2 in prospectively pooled specimens. *Emerg. Infec. Dis.* **27** 92–103.

WEIN, L. M. and ZENIOS, S. A. (1996). Pooled testing for HIV screening: Capturing the dilution effect. *Oper. Res.* **44** 543–569.

YAMAMURA, K. and HINO, A. (2007). Estimation of the proportion of defective units by using group testing under the existence of a threshold of detection. *Comm. Statist. Simulation Comput.* **36** 949–957. MR2415696 https://doi.org/10.1080/03610910701539278

ZHANG, Y., LI, Y., WANG, L., LI, M. and ZHOU, X. (2020). Evaluating transmission heterogeneity and super-spreading event of COVID-19 in a Metropolis of China. *Int. J. Environ. Res. Public Health* **17** 3705.

ZHANG, J., TIAN, S., LOU, J. and CHEN, Y. (2020). Familial cluster of COVID-19 infection from an asymptomatic. *Critical Care* **24** 7–9.

ZHOU, Y., WANG, L., ZHANG, L., SHI, L., YANG, K., HE, J., ZHAO, B., OVERTON, W., PURKAYASTHA, S. et al. (2020). A spatiotemporal epidemiological prediction model to inform county-level COVID-19 risk in the United States. *Harvard Data Science Review*.