Journal of the American Medical Informatics Association, 28(12), 2021, 2562–2570
doi: 10.1093/jamia/ocab169
Advance Access Publication Date: 11 October 2021
Research and Applications

Research and Applications

DOPE: D-Optimal Pooling Experimental design with application for SARS-CoV-2 screening

Yair Daon,[1,2] Amit Huppert,[1,3] and Uri Obolski [1,2]

[1]School of Public Health, Tel Aviv University, Tel Aviv, Israel, [2]Porter School of the Environment and Earth Sciences, Tel Aviv University, Tel Aviv, Israel, and [3]The Bio-statistical and Bio-mathematical Unit, The Gertner Institute for Epidemiology and Health Policy Research, Tel Hashomer, Israel

Corresponding Author: Uri Obolski, PhD, George Wise Street, Tel Aviv-Yafo 6997801, Israel (uriobols@tauex.tau.ac.il)

Received 29 March 2021; Revised 16 June 2021; Editorial Decision 25 July 2021; Accepted 31 July 2021

ABSTRACT

Objective: Testing individuals for the presence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the pathogen causing the coronavirus disease 2019 (COVID-19), is crucial for curtailing transmission chains. Moreover, rapidly testing many potentially infected individuals is often a limiting factor in controlling COVID-19 outbreaks. Hence, pooling strategies, wherein individuals are grouped and tested simultaneously, are employed. Here, we present a novel pooling strategy that builds on the Bayesian D-optimal experimental design criterion.
Materials and Methods: Our strategy, called DOPE (D-Optimal Pooling Experimental design), is built on a novel Bayesian formulation of pooling. DOPE defines optimal pooled tests as those maximizing the mutual information between data and infection states. We estimate said mutual information via Monte-Carlo sampling and employ a discrete optimization heuristic to maximize it.
Results: We compare DOPE to other, commonly used pooling strategies, as well as to individual testing. DOPE dominates the other strategies as it yields lower error rates while utilizing fewer tests. We show that DOPE maintains this dominance for a variety of infection prevalence values.
Discussion: DOPE has several additional advantages over common pooling strategies: it provides posterior distributions of the probability of infection rather than only binary classification outcomes; it naturally incorporates prior information of infection probabilities and test error rates; and finally, it can be easily extended to include other, newly discovered information regarding COVID-19.
Conclusion: DOPE can substantially improve accuracy and throughput over current pooling strategies. Hence, DOPE can facilitate rapid testing and aid the efforts of combating COVID-19 and other future pandemics.

Key words: Bayesian, Monte-Carlo, epidemiology, COVID-19, RT-PCR

INTRODUCTION

During the current COVID-19 pandemic, large-scale testing efforts for detecting the presence of the SARS-CoV-2 virus, the causative agent of the disease, are crucial. Testing allows isolating infected individuals, thus breaking transmission chains. Testing for SARS-CoV-2 is typically done using RT-PCR (reverse transcription polymerase chain reaction.[1,2] Testing via RT-PCR kits can be a limiting factor, creating a bottleneck in screening and isolation efforts.[3,4] The most common way to increase efficiency and throughput of RT-PCR tests is pooling. Pooling is the act of using samples from several different individuals in a single RT-PCR test, hereby referred to as a pool. Several pooling strategies have been

previously suggested,[5,6] analyzed,[5,6] and applied.[7–10] The modus operandi of pooling is as follows: a result is observed for 1 or several pools, and then further action is taken. Usually, a negative result for a pool means all members of said pool are declared negative without any further testing. A positive result, on the other hand, may render some individuals positive or require further testing.

Pooling Strategies

Pooling originated in the seminal work of Dorfman[5] in 1943. Since then, pooling has evolved into what is known today as group testing.[13] There are several common pooling strategies, and they are outlined below. Implementation details can be found in Supplementary Material A.

Dorfman pooling[5] starts by testing a predetermined number of individuals in a pool. If the pooled test result is negative, all pool members are declared negative. Otherwise, each 1 is tested separately. A large-scale testing effort[9] has shown that Dorfman pooling can save 76% of RT-PCR tests.

In recursive pooling,[7] if the first pooled test is positive, the pool is split into 2 and the process repeats. Otherwise, all pool members are declared negative. Thus, an individual is only declared positive if they are eventually tested separately and the test result is positive. One study showed that a recursive pooling can potentially result in a 7-fold increase in throughput.[12]

Matrix pooling[14] arranges a population of size $N = mn$ in an $m \times n$ matrix. Each row and column are then pooled and individuals in the intersection of positive rows and columns are tested separately. We were not able to find data of a real-world implementation of matrix pooling.

Objective

We developed DOPE (D-Optimal Pooling Experimental design), a novel Bayesian pooling strategy. DOPE identifies which choice of pools maximizes the mutual information between population infection state and pooled test data. This choice of mutual information as an optimization objective categorizes DOPE as a D-optimal experimental design technique[15] and results in superior performance of DOPE compared to common strategies.

Significance

DOPE is a Bayesian strategy and, as such, enjoys the common advantages of Bayesian methods. Assumptions on the population and RT-PCR test error rates are easily incorporated into a prior and a likelihood model, respectively. Furthermore, DOPE allows the probabilities of infection to be naturally quantified via the posterior. These probabilities convey more information and allow greater flexibility compared to a binary test result .

Precise quantification of the aforementioned probabilities of infection allows DOPE to perform trade-offs between error rates and number of tests as required. Most common pooling strategies do not allow for such an adaptive property and hence do not have control over the number of tests or error rates.

Another advantage of DOPE is evident when considering edge cases in common strategies. Consider Dorfman pooling: how should one act if the first pooled test is positive, yet all subsequent tests are negative? Similar events arise for recursive and matrix pooling as well (see implementation details in Supplementary Material A). Such events all have non-negligible probabilities under the empirically estimated test error rates and are likely to result in implementation problems. In contrast, there are no ambiguous events when DOPE is the strategy of choice. All test results are used for updating one's beliefs via Bayes' theorem.

Lastly, DOPE is useful across both high and low infection prevalence scenarios. Some common strategies lose efficiency at high infection prevalence;[5,7,8,10] others may suffer from increased false-negative rates due to unmet assumptions of sparsity.[8] DOPE, in contrast, is inherently adaptable and suitable for a wide range of infection prevalence levels.

MATERIALS AND METHODS

DOPE consists of several components. Briefly, a Bayesian model for pooling is formulated, and a design is defined as a combination of pools. An optimal design is defined as maximizing mutual information between population infection state and pooled test data. Calculating said mutual information proceeds via Monte-Carlo simulations. Then an optimal design is found via discrete optimization, data are collected, and the process repeats.

Prior

The prior encodes the probability of every possible infection state of the tested population. We assume the following structure: The population is divided into disjoint clusters (eg, families, work places, classrooms), each contains a (potential) initial source of primary infection, which occurs with probability P_p. A secondary infection of other members of the cluster occurs with probability P_s for each. If no primary infection occurs, the probability that nonprimary members of the cluster are infected is the infection prevalence in the general population P_b. Our assumptions are given below, with their corresponding notation:

- Population members are denoted $\{1, \ldots, N\}$.
- The population state is captured in $\theta \in \{0,1\}^N$. Individual $h \in \{1, \ldots, N\}$ is either infected or not, with $\theta_h = 1$ or $\theta_h = 0$, respectively.
- The population is partitioned into M disjoint clusters $C_1, \ldots, C_M$. A single cluster represents, for example, a household.
- A cluster C is a tuple : $C = (h_0, h_1, \ldots, h_n)$. We assume here, for the sake of notation only, that all clusters contain the same number of members $n+1$.
- For cluster C denote $\theta^{(C)} := (\theta_{h_0}, \ldots, \theta_{h_n})$.
- A primary infection of h_0 occurs with probability P_p.
- A secondary infection of any of $h_1, \ldots, h_n$ by h_0 occurs independently with probability P_s.
- If no primary infection occurs, $h_1, \ldots, h_n$ are infected with the basal prevalence of infection in the general population P_b.

Since clusters are disjoint, their prior probabilities are independent:

$$\mathbb{P}(\theta) = \prod_{C \in C_1, \ldots, C_M} \mathbb{P}\left(\theta^{(C)}\right). \tag{1}$$

Turning our attention to cluster C:

$$\mathbb{P}(\theta^{(C)}) = \mathbb{P}(\theta_{h_0}) \prod_{j=1}^{n} \mathbb{P}(\theta_{h_j} | \theta_{h_0}) \tag{2}$$

$$= [P_p \prod_{j=1}^{n} P_s^{\theta_{h_j}} (1-P_s)^{1-\theta_{h_j}}]^{\theta_{h_0}} [\left(1-P_p\right) \prod_{j=1}^{n} P^{\theta_{h_j}} (1-P_b)^{1-\theta_{h_j}}]^{1-\theta_{h_0}}.$$

An explicit expression for $P(\theta)$ is easily found from equations (1) and (2). Our model easily accommodates a common source of infec-

tion for 2 clusters by taking the union of said clusters. We note that the above-mentioned prior does not allow simultaneous initial infection between nonprimary household members (eg, h_2 and h_3). However, the difference in probabilities when including such a scenario is negligible (Supplementary Material A).

Beta Binomial Prior

We also consider a different prior model, which we refer to as the "beta binomial prior."[16] Contrary to the model presented above (the Prior section), the beta binomial prior does not assume any infection mechanism but merely enforces an infection prevalence p and a given correlation ρ between infection states of cluster members. For given p and ρ, we set $\alpha := p(\rho^{-1}-1)$ and $\beta := (1-p)(\rho^{-1}-1)$, and draw $q \sim \text{Beta}(\alpha, \beta)$ and $\theta_i \sim \text{Ber}(q)$ iid. This procedure ensures $E[\theta_i] = p$ for $i = 1, \ldots, N$ and $Corr(\theta_i, \theta_j) = \rho$ for $i \neq j$ (more details can be found in[16] Section 3).

Likelihood

The likelihood encodes our assumptions on pooled SARS-CoV-2 RT-PCR tests and how they can err. A detailed exposition of RT-PCR can be found in.[1,2]

Failed detection of SARS-CoV-2 RNA in pooled RT-PCR testing is referred to below as a false-negative. Previous studies of group testing strategies assumed that the false-negative probability does not depend on the number of infected samples but merely on the existence of at least 1 such sample in a pool.[9,11] Current studies of pooling in the context of SARS-CoV-2 also employ similar assumptions.[19,20] Specifically, these studies assume the same probability of a negative result for a pool with either a single or multiple samples originating from infected individuals. However, in a previous study, we have shown that this assumption does not align with experimental data.[21] Thus, we assume that viral RNA from each positive individual in a pool undergoes the RT-PCR amplification process independently. Consequently, the probability of (failed) amplification and/or detection for every sample whose source was an infected individual is considered separately.

Erroneous detection of SARS-CoV-2 RNA (a false-positive) in pooled RT-PCR testing can also occur. A common assumption[7,8,17,18] is that the false-positive probability does not depend on the number of negative samples in a pool. We incorporate this assumption in our likelihood model with a small modification. We assume that an erroneous amplification can occur in any pool. Specifically, it is possible that correct amplifications fail and an erroneous one occurs simultaneously. This assumption is relatively specific for the current application of screening for SARS-CoV-2 via RT-PCR. For example, cross-reactivity with other coronaviruses would have violated this assumption, but it was ruled out in.[20]

To summarize, we assume that for a single pool, a positive test result is generated in 1 of 2 paths: either SARS-CoV-2 RNA from an infected individual is correctly amplified and detected independently for each positive sample in a pool; or, some erroneous amplification occurs (at most once per pool). Our model is illustrated in Figure 1. We proceed to formulate the likelihood, so we require some definitions and notations:

- A pool is a collection of individuals $\{h_1, \ldots, h_m\} \subseteq \{1, \ldots, N\}$.
- A design T is a collection of pools. The k^{th} pool is denoted T_k.
- Data are denoted $\boldsymbol{d} \in \{0,1\}^T$. We let $\boldsymbol{d}_k = 1$ if upon testing T_k a positive result was observed, and let $\boldsymbol{d}_k = 0$ otherwise.
- The probability that the detection process fails for 1 sample taken from an infected individual is P_{fn}.

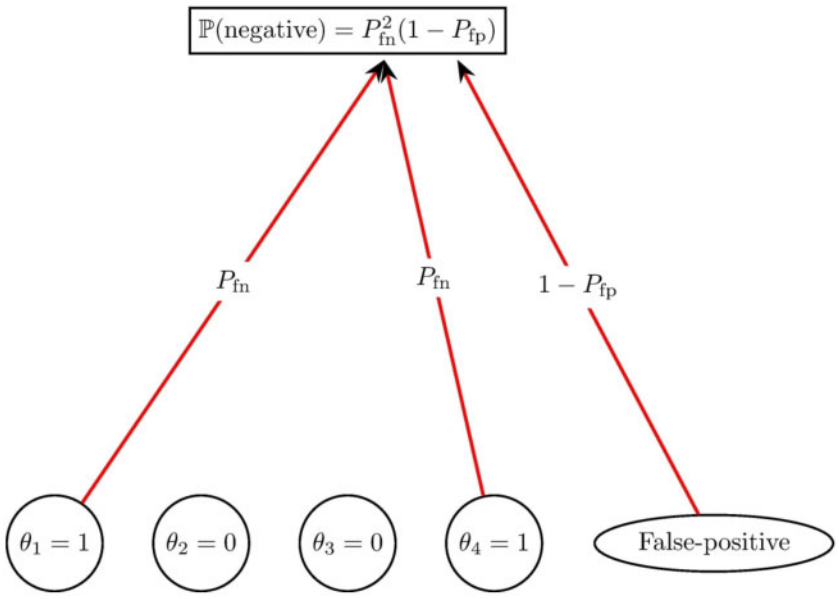

Figure 1. Illustration of the likelihood model of a pooled test result. A pool contains individuals {1,2,3,4} with state $\theta = (1,0,0,1)$ (ie, individuals 1 and 4 are infected). A negative pooled test implies that 3 detection paths failed. A false-negative occurred for 1 and 4, each with probability P_{fn}. Additionally, no erroneous detection of SARS-CoV-2 occurred, with probability 1-P_{fp}. Individuals 2 and 3 are not infected and do not contribute to the probability of the pooled test result. A positive pooled test arises if any one of the above-mentioned paths results in a detection.

- The probability of an erroneous amplification and detection in a pooled test is P_{fp}.

The probability of a negative pooled test result is presented in (3a), along with its complement (3b) and explained below.

$$\mathbb{P}(\boldsymbol{d}_k = 0|T_k,\theta) \;= (1-P_{fp})\prod\nolimits_{b\in T_k} P_{fn}^{\theta_b}, \tag{3a}$$

$$\mathbb{P}(\boldsymbol{d}_k = 1|T_k,\theta) \;= 1-(1-P_{fp})\prod\nolimits_{b\in T_k} P_{fn}^{\theta_b}. \tag{3b}$$

A negative pooled test result occurs when all detection paths (both correct and erroneous) fail. The probability of no false-detection accounts for the $1-P_{fp}$ term. The probability of no correct detection is P_{fn} per infected individual. The probability that all such paths fail is the product of the above-mentioned terms, displayed in (3a). The probability of a positive result, presented in (3b), is simply the complement. Combining (3a) and (3b), and recalling that $\boldsymbol{d}_k \in \{0,1\}$ yields:

$$\mathbb{P}(\boldsymbol{d}_k|T_k,\theta)=\left[(1-P_{fp})\prod\nolimits_{b\in T_k} P_{fn}^{\theta_b}\right]^{1-\boldsymbol{d}_k}\left[1-(1-P_{fp})\prod\nolimits_{b\in T_k} P_{fn}^{\theta_b}\right]^{\boldsymbol{d}_k}. \tag{4}$$

Since different tests are assumed independent, the full likelihood is the product:

$$\mathbb{P}(\boldsymbol{d}|T,\theta) \;= \prod\nolimits_{k=1}^{|T|}\mathbb{P}(\boldsymbol{d}_k|T_k,\theta). \tag{5}$$

D-Optimal Design

In Bayesian experimental design, a design is called D-optimal if it maximizes any 1 of several equivalent information theoretic design criteria.[15,21–23] For convenience, we consider the mutual information between parameters and data as the optimization criterion. For a given design T, the mutual information between data d and population infection state θ is denoted $\Psi(T)$:

$$\Psi(T): \;= I(\theta;\boldsymbol{d}|T) = \sum\nolimits_{\theta,\boldsymbol{d}}\mathbb{P}(\theta,\boldsymbol{d}|T)\log\frac{\mathbb{P}(\theta,\boldsymbol{d}|T)}{\mathbb{P}(\theta)\mathbb{P}(\boldsymbol{d}|T)}. \tag{6}$$

It is known that maximizing mutual information is equivalent to minimizing the expected posterior entropy and maximizing the expected Kullback-Leibler divergence (henceforth KL divergence)[24] between posterior and prior $E_d[KL(\mathbb{P}(\theta|\boldsymbol{d},T)\parallel\mathbb{P}(\theta))]$.[15,21,23,25] Some details of the D-optimal approach are discussed in the Discussion.

There is no closed form expression for Ψ, and we estimate it via Monte-Carlo sampling. We start with a straightforward calculation:

$$\Psi(T) \;= \sum\nolimits_{\theta,\boldsymbol{d}}\mathbb{P}(\theta,\boldsymbol{d}|T)\log\frac{\mathbb{P}(\theta,\boldsymbol{d}|T)}{\mathbb{P}(\theta)\mathbb{P}(\boldsymbol{d}|T)}$$

$$= \sum\nolimits_{\theta,\boldsymbol{d}}\mathbb{P}(\theta,\boldsymbol{d}|T)\log\frac{\mathbb{P}(\boldsymbol{d}|\theta,T)\mathbb{P}(\theta)}{\mathbb{P}(\theta)\mathbb{P}(\boldsymbol{d}|T)} \tag{7}$$

$$= \sum\nolimits_{\theta,\boldsymbol{d}}\mathbb{P}(\theta,\boldsymbol{d}|T)(\log\mathbb{P}(\boldsymbol{d}|\theta,T)-\log\mathbb{P}(\boldsymbol{d}|T)).$$

Estimating the last sum requires 3 steps:[21,23]

1. Sample $\mathbb{P}(\theta,\boldsymbol{d}|T)$.
2. Evaluate $\log\mathbb{P}(\boldsymbol{d}|\theta,T)$ and estimate $\log\mathbb{P}(\boldsymbol{d}|T)$ for each sample.
3. Average.

We carry out the first step by sampling the prior L times: $\eta_k \sim \mathbb{P}(\theta), k=1,\ldots,L$. Once we obtain the prior samples η_k, the likelihood is sampled $Y_k \sim \mathbb{P}(\boldsymbol{d}|\eta_k,T), k=1,\ldots,L$. This procedure results in L pairs of samples from the joint distribution of states and data: $(\eta_k,Y_k)\sim\mathbb{P}(\theta,\boldsymbol{d}|T)$.

Calculating the left summand $\log\mathbb{P}(Y_k|\eta_k,T)$ is straightforward and only requires evaluating the likelihood. The right summand in the last line of (7) satisfies:

$$\mathbb{P}(Y_k|T) = E_\theta[\mathbb{P}(Y_k|\theta,T)] = \sum\nolimits_\theta \mathbb{P}(\theta)\mathbb{P}(Y_k|\theta,T), \tag{8}$$

and we estimate it via Monte-Carlo, taking advantage of existing samples:

$$\mathbb{P}(\widehat{Y_k|T}): \;= \frac{1}{L}\sum\nolimits_{r=1}^{L}\mathbb{P}(Y_k|\eta_r,T). \tag{9}$$

The third step is realized by first utilizing the samples η_k, Y_k and (9) to define:

$$\widehat{\Psi}(T): \;= \frac{1}{L}\sum\nolimits_{k=1}^{L}\left(\log\mathbb{P}(Y_k|\eta_k,T)-\log\mathbb{P}(\widehat{Y_k|T})\right) \tag{10}$$

$$= \frac{1}{L}\sum\nolimits_{k=1}^{L}(\log\mathbb{P}(Y_k|\eta_k,T)-\log\frac{1}{L}\sum\nolimits_{r=1}^{L}\mathbb{P}(Y_k|\eta_r,T)).$$

Calculating $\widehat{\Psi}$ via equation (10) constitutes 1 of the main computational difficulties in finding an optimal design. The rationale is that the number of likelihood evaluations is L^2, so calculating $\widehat{\Psi}$ is of complexity $O(L^2)$. The estimator $\widehat{\Psi}$ is biased and its bias is $O(L^{-1})$. See[21,23] for a full discussion of convergence and bias of $\widehat{\Psi}$. See Supplementary Material A for a discussion of the choice of number of samples L.

Posterior

Once data $\boldsymbol{d}'$ for design T' have been observed, we would like to define Ψ for a new design T. The definition is a natural extension of (6), with the posterior $\mathbb{P}(\theta|T',\boldsymbol{d}')$ taking the place of the prior $\mathbb{P}(\theta)$:

$$\Psi(T;T',\boldsymbol{d}'): \;= I(\theta,\boldsymbol{d}|T',\boldsymbol{d}')$$
$$= \sum\nolimits_{\theta,\boldsymbol{d}}\mathbb{P}(\theta,\boldsymbol{d}|T',\boldsymbol{d}',T)\log\frac{\mathbb{P}(\theta,\boldsymbol{d}|T',\boldsymbol{d}',T)}{\mathbb{P}(\boldsymbol{d}|T',\boldsymbol{d}',T)\mathbb{P}(\theta|T',\boldsymbol{d}')}, \tag{11}$$

where d is the data for T. Before data are observed and design generated, we write $T'=\varnothing$ and $\boldsymbol{d}'=\varnothing$. Therefore, $\mathbb{P}(\theta)=\mathbb{P}(\theta|\varnothing,\varnothing)$ and indeed $\Psi(T)=\Psi(T;\varnothing,\varnothing)$.

The calculation of $\widehat{\Psi}(T;T',\boldsymbol{d}')$ proceeds verbatim as in the D-Optimal Design section. The only difference is that instead of sampling $\eta_k\sim\mathbb{P}(\theta), k=1,\ldots,L$, we sample from the posterior $\eta_k\sim\mathbb{P}(\theta|\boldsymbol{d}',T'), k=1,\ldots,L$.

Sampling the posterior is achieved by Gibbs sampling. Denote all θ_j's except the i^{th} by $\theta_{-i}=\{\theta_1,\ldots,\theta_{i-1},\theta_{i+1},\ldots,\theta_N\}$. Gibbs sampling requires repeatedly sampling from $\mathbb{P}(\theta_i|T',\boldsymbol{d}',\theta_{-i})$, which are calculated as follows:

$$\mathbb{P}(\theta_i|T',\boldsymbol{d}',\theta_{-i}) = \frac{\mathbb{P}(\theta_i,\theta_{-i}|T',\boldsymbol{d}')}{\mathbb{P}(\theta_{-i}|T',\boldsymbol{d}')} = \frac{\mathbb{P}(\theta_i,\theta_{-i}|T',\boldsymbol{d}')}{\sum_{x\in\{0,1\}}\mathbb{P}\left(\theta_i=x,\theta_{-i}|T',\boldsymbol{d}'\right)}, \tag{12}$$

and the normalization constant cancels out, making the calculation possible.

Naively utilizing samples from the Gibbs sampler for calculating $\widehat{\Psi}(T;T',\boldsymbol{d}')$ is wasteful. Recall, from the discussion in the D-Optimal Design section, that said calculation is $O(L^2)$, where L is

the number of Monte-Carlo samples utilized. Since Gibbs sampler does not generate independent samples, naively taking L samples from the Gibbs sampler would require huge L to cover all state space for θ, thus rendering the calculation of $\widehat{\Psi}$ prohibitively expensive. A remedy is found in Sokal[26]: "The number of 'effectively independent samples' in a run of length n is roughly $\frac{n}{2\tau_{int,f}}$," where $\tau_{int,f}$ is the integrated autocorrelation time for function f. Thus, we first estimate τ_{int,f_i} for the coordinate projections $f_i(\theta) = \theta_i$ and take $\tau := max_i \tau_{int,f_i}$. The calculation of τ_{int,f_i} is carried out using `emcee's`[27] method autocorr during the chain's burn-in time. We then run the Gibbs sampler for τL steps and discard all but every τ^{th} sample, thus keeping computational costs and variance for $\widehat{\Psi}$ low. Pseudocode for our Gibbs sampler can be found in Supplementary Material A.

Optimization

Given a routine that calculates $\widehat{\Psi}(T; T', d')$ for any design T, we need to find a way to maximize $\widehat{\Psi}$ over all valid designs. Designs are restricted to have a fixed number of pools, denoted K. Optimizing over all valid designs results in a difficult discrete-optimization problem, which we solve via a heuristic hill-climbing approach. Although hill-climbing is a heuristic, we have found it to work sufficiently well. See Supplementary Material A for details.

DOPE

We now present DOPE: D-Optimal Pooling Experimental-design, summarized in Algorithm 1. DOPE requires 2 parameters: first, the number of pooled tests per step K, and second, a decision interval $I \subset [0, 1]$. The decision interval defines the required certainty levels to serve as a stopping criterion for DOPE. The meaning of $P(\theta_i = 1|T, d) \in I$ is that the state of individual i is still uncertain, so further testing is required. DOPE stops when there is no uncertainty regarding the state of any individual (read: $\forall i,\ P(\theta_i = 1|T, d) \notin I$).

DOPE typically proceeds to find K optimal pools, perform the corresponding RT-PCR tests, and repeat the process if any individual's posterior infection probability is in I. However, DOPE can also be executed in a nonsequential manner, where no retesting is allowed. Such a nonsequential implementation can be achieved in Algorithm 1 by letting K be the total number of allotted tests and $I = \varnothing$.

Algorithm 1 DOPE: D-Optimal Pooling Experimental design

procedure DOPE(K, I)
  $\mathbf{T}, \mathbf{d} \leftarrow \varnothing,\ \varnothing$
  repeat
    $\mathbf{T}' \leftarrow \text{OptimalDesign}(K, \mathbf{T}, \mathbf{d})$
    $\mathbf{d}' \leftarrow \text{PCR}\left(\mathbf{T}'\right)$ ▷ Perform RT-PCR tests for T′
    $\mathbf{T} \leftarrow \left(\mathbf{T}_1, \ldots, \mathbf{T}_T, \mathbf{T}'_1, \ldots, \mathbf{T}'_{T'}\right)$ ▷ Concatenate
    $\mathbf{d} \leftarrow \left(\mathbf{d}_1, \ldots, \mathbf{d}_T, \mathbf{d}'_1, \ldots, \mathbf{d}'_{T'}\right)$ ▷ Concatenate
  until $\forall i\ P(\theta_i = 1|\mathbf{T}, \mathbf{d}) \notin I$
  return $\begin{bmatrix} P(\theta_1 = 1|\mathbf{T}, \mathbf{d}) > 0.5 \\ \vdots \\ P(\theta_N = 1|\mathbf{T}, \mathbf{d}) > 0.5 \end{bmatrix}$ ▷ Classification via posterior marginals

RESULTS

We compare DOPE to 3 prominent pooling strategies: Dorfman, recursive, and matrix pooling. We present extensive simulation results and consider a large number of parameter choices for DOPE. We choose $K = 1$, so DOPE always finds a single optimal pool in each step. Decision intervals' lower and upper bounds are chosen from $\{0.01, 0.02, \ldots, 0.15\}$ and $\{0.3, 0.35, \ldots, 0.95\}$, respectively. We take a percentage nomenclature (eg, DOPE α, β utilizes the decision interval $I = [\alpha/100, \beta/100]$).

There are 3 performance metrics with which one can evaluate pooling strategies: false-negative rate, false-positive rate, and number of tests. We plot false-negative rates against an average number of tests and delegate plots of false-positive rates to Supplementary Material A (the reason is explained in the Discussion). In addition, we present the average KL divergence for each strategy. Although the posterior entropy is not a performance metric, per se, we choose to present it in plots. The reason is that presenting posterior entropy shows that indeed DOPE succeeds in maximizing this well-defined statistical criterion.

We say a pooling strategy A dominates another strategy B for false-negative rates if A achieves lower false-positive rates than B, while utilizing a smaller (or equal) number of tests. Similarly, we say that A dominates B for posterior entropy if similar conditions apply for posterior entropy. In the results below, we show that for both false-negative rates, as well as for posterior entropy, there are decision intervals for which DOPE dominates Dorfman, recursive, and matrix pooling.

In all simulations presented in this section, we used the commonly observed test error rates $P_{fn} = 0.2, P_{fp} = 0.01$.[28–30] Infection prevalence in the tested population was realized by the varying population connectivity parameter P_p and P_s, which took values in $[0.05,\ 0.4]$—see Supplementary Material A for a summary of estimates of connectivity parameters and Supplementary Material B for a table of simulation parameters. The basal prevalence in the population was always set to $P_b = 0.01$.

Small and Distinct Clusters

We compare DOPE to Dorfman, recursive, and matrix pooling. Results are shown in Figure 2. We consider a population of size $N = 32$, which is the maximal pool size that can be employed without interfering with the RT-PCR process by sample dilution.[31] Figure 2 shows that a decision interval can be found for which DOPE dominates common strategies for both false-negative rate as well as for posterior entropy. For presentation purposes, choices of decision intervals are grouped according to their lower bound, with colors corresponding to the color bar on the right of Figure 2 (eg, DOPE 60,10, DOPE 70,10 and DOPE 90,10 share the same color). Results are shown for $P_{fn} = 0.2, P_{fp} = 0.01, P_p = 0.2, P_s = 0.2, P_b = 0.01$, which results in average disease prevalence $\approx$7.7%. Number of samples in the Monte-Carlo estimation is 20 000. Each point presented is the average of 123 simulations.

Large Clusters

Scenarios where hundreds of individuals need to be simultaneously tested may also arise during an outbreak. If such populations can be naturally divided into smaller groups (eg, households, classrooms), then we can easily apply DOPE as described in the previous section. Alternatively, if all infections are independently and identically distributed, any subdivision into smaller groups retains the same information regarding infected individuals. This corresponds to applying

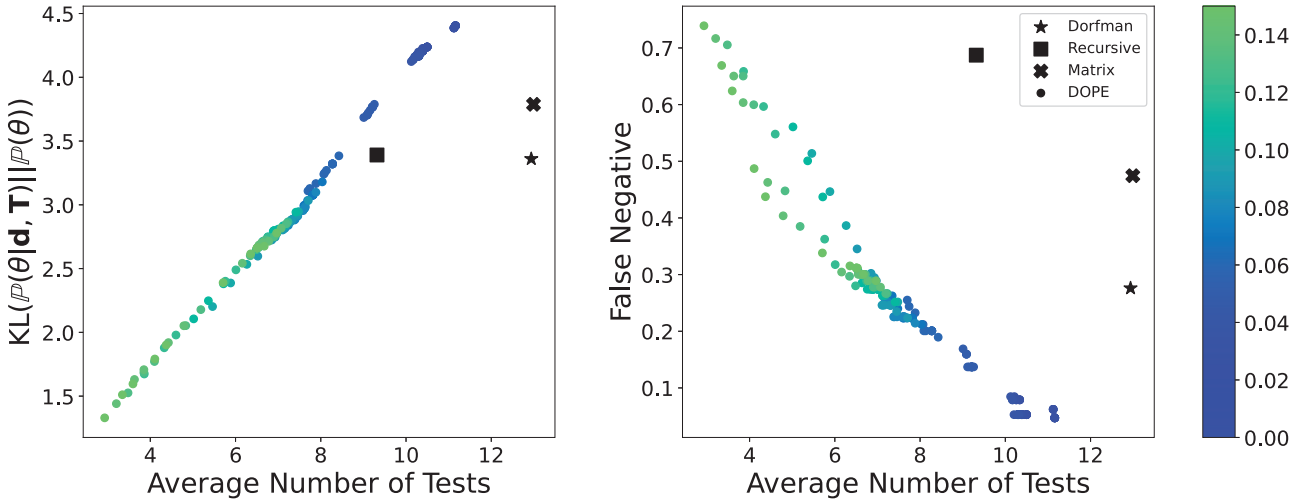

Figure 2. Comparison of DOPE against other pooling strategies. We plot the KL divergence from posterior to prior (left) and false-negative rates (right) achieved by different pooling strategies against the number of tests used (x-axis). The results for DOPE (•) are shown for various decision intervals (colors represent the lower bound of the decision interval). Dorfman (★), recursive (■), and matrix (×) pooling are shown for their only configuration, in black. There are always decision intervals for which DOPE dominates common strategies.

DOPE to a completely disconnected population in which it also dominates (Supplementary Material A).

Finally, a larger population of potentially infected individuals with no a priori structure but potential dependencies between the infected (eg, a hospital ward, military unit, or a cruise ship) might present itself. Thus, we also present results for a large population with a single large cluster that behaves according to the beta binomial prior model (the Beta Binomial Prior section and[16]) Disease prevalence in the population was set to 7% whereas the correlation of infection state between cluster members was set to $\rho = 0.2$. Test error rates were set as $P_{fn} = 0.2, P_{fp} = 0.01$. We consider a population of size 256 and apply the competing strategies as follows: Dorfman pooling is utilized with an optimal pool size chosen following.[8] Recursive pooling was utilized with a pool size of 32, following.[31] Matrix pooling was implemented with a 16×16 square matrix. DOPE was implemented by partitioning the population into subpopulations of size $N = 32$. Since the optimization DOPE utilizes is now restricted to each separate pool, such a partitioning can only degrade the performance of DOPE. However, as we observe in Figure 3, the results are qualitatively similar to those presented in the Small and Distinct Clusters section, and DOPE still dominates. Calculating the expected posterior entropy simultaneously (in contrast to the computation performed by DOPE for each subpopulation separately) for such large clusters is computationally very demanding. Hence, we do not present it here, as we did in Figures 2 and 3.

Varying Prevalence

We examine the performance of DOPE under a wide range of disease prevalence values. Performance is demonstrated for a population of size $N = 10$ with infection prevalences in $[0.02,\ 0.18]$. Connectivity parameters generating these prevalences can be found in Supplementary Material B. Test error rates of $P_{fn} = 0.2, P_{fp} = 0.01$ were used, with $L = 12000$ samples for the Monte-Carlo estimation.

In Figure 4, we show the performance of DOPE for 4 decision intervals. Each decision interval was chosen so that DOPE's expected number of tests was closest to 1 of the common pooling strategies (Dorfman, recursive, and matrix). We also show such a comparison for separate testing. For each choice of decision interval, DOPE outperformed other strategies in terms of false-negative rates, mostly using substantially fewer tests. The only exceptions occurred when we were not able to find decision intervals that closely match the behavior of matrix pooling and separate testing. We could not find such decision intervals since the number of tests used in matrix pooling and separate testing are constant and near-constant, respectively, in the range of disease prevalence rates we consider. DOPE is more adaptive than these strategies, hence the number of tests it utilizes increases with the disease prevalence rate. Consequently, it was difficult to find a decision interval for which DOPE utilizes a number of tests close enough to the number of tests utilized by either separate testing or matrix pooling throughout the prevalence range considered. Our choice of decision intervals resulted in DOPE utilizing fewer tests compared to matrix pooling and individual testing. Thus, we can expect even lower error rates when utilizing the same number of tests as matrix pooling or separate testing.

DISCUSSION

In this manuscript we have presented DOPE, a novel pooling strategy that has the potential to substantially improve the performance of RT-PCR pooling in terms of number of tests and error rates. We have demonstrated DOPE's superiority over competing pooling strategies under different scenarios—including varying prevalence rates, test error rates, population infection models, and population sizes.

DOPE was developed with the aim of maximizing the information gained from pooled tests, precisely defined in the D-Optimal Design section. DOPE is a Bayesian method and, as such, enjoys many of the advantages Bayesian analysis has to offer. For example, DOPE offers seamless integration of probabilistic assumptions on population connectivity and test errors into its underlying probabilistic model. Thus, we can apply DOPE to tests with different error models/rates (eg, COVID-19 antigen test[32,33]). Moreover, DOPE can make the trade-off between the number of tests and test error rates explicit. Lastly, DOPE's error rates are lower compared to common pooling methods, while utilizing the same amount of, or

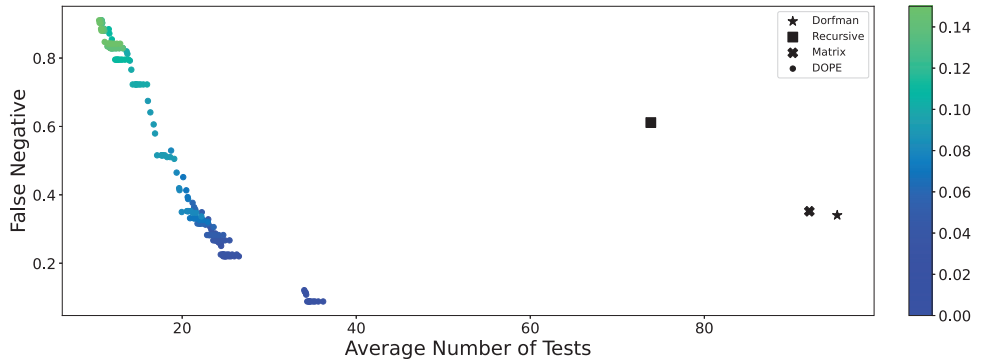

Figure 3. Comparison of DOPE against other pooling strategies, assuming the infections follow the beta binomial model of (Basso et al 2020) for a population size of 256. We plot the false-negative rates achieved by different pooling strategies against the number of tests utilized (x-axis). The results for DOPE (•) are shown for various decision intervals (colors represent the lower bound of the decision interval). Dorfman (★), recursive (■), and matrix (×) pooling are shown in black. There are always decision intervals for which DOPE dominates common strategies.

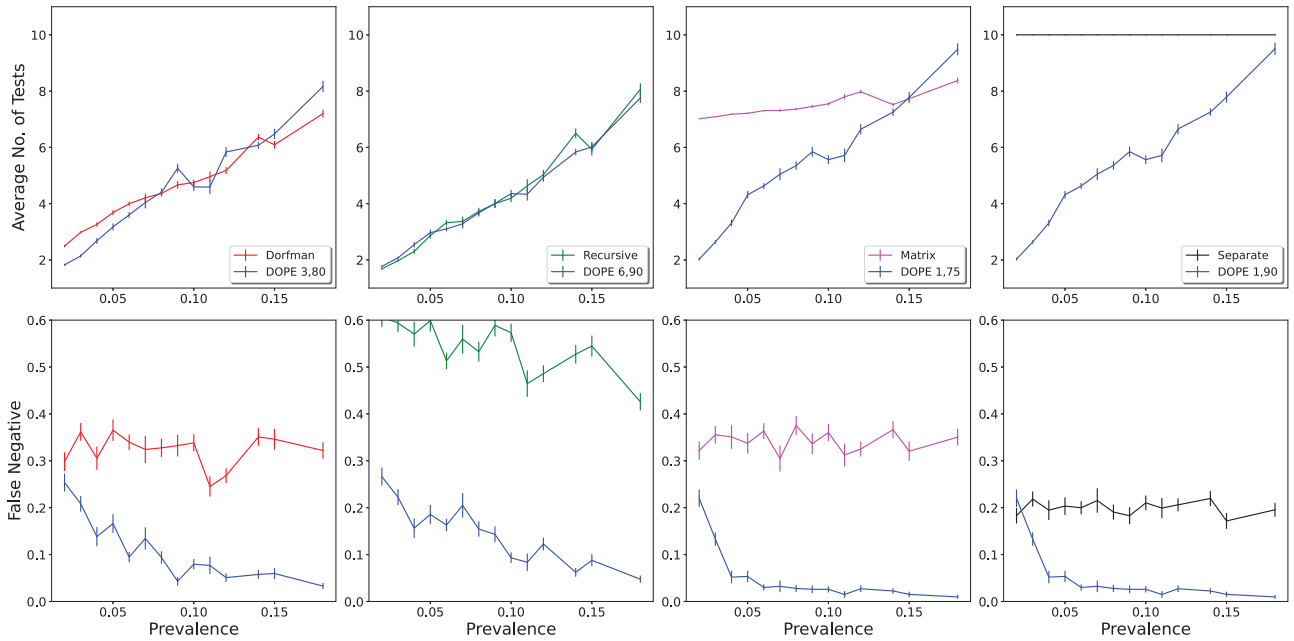

Figure 4. Comparison of DOPE to other strategies under varying disease prevalence values. DOPE decision intervals were chosen so that the number of tests taken was as close as possible to the common strategies. Rows, top, and bottom: number of tests and false-negative rate, respectively, plotted against disease prevalence. Columns show performance for each pair of strategies (see legend in the top row).

fewer, tests. Last but not least, DOPE can return posterior infection probabilities, giving a very refined tool for decision-making under uncertainty.

DOPE is based on an information theoretic experimental design criterion, maximizing mutual information between population infection state θ and pooled tests results d. There are multiple motivations for this definition. If we view the testing procedure as a communication channel,[24] where we wish to transmit θ, then a D-optimal design maximizes the channel capacity. The channel capacity is the upper bound for the amount of information that can be transmitted through the channel with vanishing error probability, so maximizing it is sensible. Alternatively, a quick calculation[15] verifies that $\Psi(T) = H(\theta) - E_{d|T}[H(\theta|d,T)]$, where $H(\cdot)$ is the Shannon entropy. Hence, D-optimal designs minimize the expected posterior entropy. Since entropy is a common measure for uncertainty, minimizing it is reasonable. Yet another calculation[21] shows that $\Psi(T)$ is the expected KL divergence from posterior to prior. KL divergence is a common measure of "distance" between probability distributions. Maximizing it roughly means we have learned as much as possible going from prior to posterior. Our results show that DOPE's performance is considerably superior to common strategies which are based on heuristics.

The Bayesian framework of DOPE also allows us to easily incorporate test error rates into our considerations. Error rates are usually not taken into account in the development of most pooling strategies, and hence such strategies are not adaptive to varying error rates.[7,8,17–19] The Bayesian formulation also allows DOPE to readily incorporate any prior knowledge obtained with regards to infection probabilities of different sub-populations. Although we have only considered connectivity of sub-populations in this manuscript, other covariates can potentially also be incorporated (eg, prior data of the likelihood of infection based on symptoms, age groups, etc).

Another important advantage of DOPE is its potential to inform quarantine decisions in a fine-grained manner. This can be achieved by examining DOPE's posterior infection probabilities $P(\theta|T,d)$ instead of its binary classification. Utilizing this additional information, various quarantine policies can be implemented with respect to the policy makers' utility functions. For example, individuals with higher posterior infection probability can be subject to a strict and prolonged quarantine and vice versa.

By selecting appropriate decision intervals, DOPE can gauge the number of tests it utilizes, giving rise to varying error rates, potentially even lower than a single test's a priori error rate. We find the required decision interval for given $P_{fn}, P_{fn}, P_p, P_s, P_b$ and cluster sizes by first simulating DOPE for many decision intervals (eg, $I = [\alpha, \beta]$ for α and β in $\{0.01, 0.02, \ldots, 0.99\}$). We then choose I that utilizes the minimal number of tests among all decision intervals that achieved error rates lower than the desired error rate.

False-positive rates are omitted from the plots in the main text (but are found in Supplementary Material A) since these are not the main concern in an epidemiological context. A false-negative result has far worse implications than a false-positive result for the spread of an infectious disease in a susceptible population. A false-negative implies an infected individual is not identified as such and consequently can continue to spread the disease. In contrast, a false-positive only implies that a noninfected individual is unnecessarily quarantined or retested. False-positive rates are not entirely meaningless, of course, as superfluous isolation can have economic and social costs. However, the false-positive rates achieved by all strategies are still very low ($\leq 1.5\%$). This is partially because RT-PCR false-positive rates are very low to begin with.[30] Thus, we believe this parameter adds very little to the comparison of common strategies, given the vast discrepancies in the average number of tests and false-negative rates.

DOPE, as any other strategy, has some limitations. First, epidemiological data of population connectivity is not always available. In this case, one can assume the population is disconnected and use this assumption as a prior. Results for such a population are presented in Supplementary Material A. Our simulations show that, even in this case, DOPE dominates common strategies.

It is possible that the iterative steps required by DOPE (find optimal pool, retest, repeat) would be difficult to implement in a real testing scenario. In this case, a nonsequential pooling strategy, where pools are chosen a priori, and no retesting is conducted, can be implemented. We can take K, the number of tests per step, to equal the number of allotted tests. Then, taking the decision interval $I = \varnothing$ makes DOPE nonsequential. This is a potential future research direction which was not in the scope of the current study.

Furthermore, DOPE requires substantial computational efforts, contrary to simple implementation of the Dorfman, recursive, and matrix pooling strategies. We have mitigated most computational obstacles, and currently a full DOPE run, with 10 initial starting points, a population size of $N = 32$, and $L = 20\ 000$ samples takes less than 5 hours to run on 7 Intel(R) Xeon(R) Gold 6252 2.1GHz CPUs. For a population size of $N = 10$, utilizing $L = 12\ 000$ samples, a full DOPE run takes less than half an hour. These times can be considerably reduced if more CPUs are available, as parallelization of DOPE is straightforward. Further improvements to the run time of DOPE can be introduced. For example, it is possible that other approximations for Ψ (eg[25]) could further reduce DOPE's run time. Alternatively, accelerating the optimization is potentially possible via, for example, solving a continuous surrogate optimization problem instead of the current discrete optimization problem. To this end, we employed the ℓ_0-sparsification method of.[34] Although this approach did not yield significant improvements in run time, we believe pursuing continuous optimization techniques should be further explored. Regardless, utilizing DOPE currently requires familiarity with programming. We plan to create a graphical interface for easy use in facilities where frequent testing is performed, so that large-scale use of DOPE is possible for nonprogrammers.

CONCLUSIONS

To summarize, we have shown that Bayesian experimental design holds a great potential for improving RT-PCR pooling. DOPE's ability to drastically increase test throughput, while decreasing testing error rates under various conditions, highlights this potential. We believe further research efforts in this direction can be conducive to mitigating the current pandemic, as well as future ones.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

AUTHOR CONTRIBUTIONS

YD and UO designed the study; YD implemented the analysis; YD, AH, and UO analyzed the results and wrote the manuscript.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

ACKNOWLEDGMENTS

The authors would like to thank Prof. David M. Steinberg for valuable discussions, advice, and support. Yair Daon was supported by a post-doctoral fellowship from the Tel Aviv University Center for Combating Pandemics. Also, Yair Daon would like to acknowledge the generous support of the Raymond and Beverly Sackler Dean's postdoctoral fellowship. The authors would furthermore like to thank Prof. Adi Stern and Prof. Lilach Hadani for granting access to their computing clusters.

DATA AVAILABILITY

No data was used in this study, only simulation results, which can be reproduced from Supplementary Material B.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. Bustin SA, Nolan T. RT-qPCR testing of SARS-CoV-2: a primer. *Int J Mol Sci* 2020; 21: 3004.
2. Bustin SA. *AZ of Quantitative PCR*. La Jolla, CA: International University Line; 2004.
3. Mina MJ, Parker R, Larremore DB. Rethinking Covid-19 test sensitivity—a strategy for containment. *N Engl J Med* 2020; 383 (22): e120.

4. Mina MJ, Andersen KG. COVID-19 testing: one size does not fit all. *Science* 2021; 371 (6525): 126–7.
5. Dorfman R. The detection of defective members of large populations. *Ann Math Statist* 1943; 14 (4): 436–40.
6. Shental N, Levy S, Wuvshet V, *et al.* Efficient high-throughput SARS-CoV-2 testing to detect asymptomatic carriers. *Sci Adv* 2020; 6 (37): eabc5961.
7. Kim H-Y, Hudgens MG, Dreyfuss JM, *et al.* Comparison of group testing algorithms for case identification in the presence of test error. *Biometrics* 2007; 63 (4): 1152–63.
8. Aprahamian H, Bish DR, Bish EK. Optimal group testing: structural properties and robust solutions, with application to public health screening. *INFORMS J Comput* 2020; 32: 895–911.
9. Barak N, Ben-Ami R, Sido T, *et al.*; Hebrew University-Hadassah COVID-19 Diagnosis Team. Lessons from applied large-scale pooling of 133,816 SARS-CoV-2 RT-PCR tests. *Sci Transl Med* 2021; 13 (589): eabf2823.
10. Mutesa L, Ndishimye P, Butera Y, *et al.* A pooled testing strategy for identifying SARS-CoV-2 at low prevalence. *Nature* 2021; 589 (7841): 276–80.
11. Ben-Ami R, Klochendler A, Seidel M, *et al.*; Hebrew University-Hadassah COVID-19 Diagnosis Team. Large-scale implementation of pooled RNA extraction and RT-PCR for SARS-CoV-2 detection. *Clin Microbiol Infect* 2020; 26 (9): 1248–53.
12. Eberhardt JN, Breuckmann NP, Eberhardt CS. Multi-stage group testing improves efficiency of large-scale COVID-19 screening. *J Clin Virol* 2020; 128: 104382.
13. Aldridge M, Johnson O, Scarlett J. Group testing: an information theory perspective. arXiv190206002, doi: 10.1561/0100000099, 2019, preprint: not peer reviewed.
14. Kwiatkowski TJ Jr, Zoghbi HY, Ledbetter SA, *et al.* Rapid identification of yeast artificial chromosome clones by matrix pooling and crude lysate PCR. *Nucleic Acids Res* 1990; 18 (23): 7191–2.
15. Chaloner K, Verdinelli I. Bayesian experimental design: a review. *Stat Sci* 1995; 10 (3): 273–304.
16. Basso LJ, Salinas V, Saure D, *et al.* The effect of correlation and false negatives in pool testing strategies for COVID-19. Available SSRN 3732829 2020. 10.2139/ssrn.3732829
17. Pikovski A, Bentele K. Pooling of coronavirus tests under unknown prevalence. *Epidemiol Infect* 2020; 148: e183.
18. Cherif A, Grobe N, Wang X, *et al.* Simulation of pool testing to identify patients with coronavirus disease 2019 under conditions of limited test availability. *JAMA Netw Open* 2020; 3 (6): e2013075.
19. Daon Y, Huppert A, Obolski U. An accurate model for SARS-CoV-2 pooled RT-PCR test errors. *medRxiv* 2021. 10.1101/2020.12.02.20242651
20. van Kasteren PB, van Der Veer B, van den Brink S, *et al.* Comparison of seven commercial RT-PCR diagnostic kits for COVID-19. *J Clin Virol* 2020; 128: 104412.
21. Huan X, Marzouk YM. Simulation-based optimal Bayesian experimental design for nonlinear systems. *J Comput Phys* 2013; 232 (1): 288–317.
22. Lindley DV. On a measure of the information provided by an experiment. *Ann Math Statist* 1956; 27 (4): 986–1005.
23. Ryan KJ. Estimating expected information gains for experimental designs with application to the random fatigue-limit model. *J Comput Graph Stat* 2003; 12 (3): 585–603.
24. Cover TM, Thomas, JA. *Elements of Information Theory*. New York, NY: John Wiley & Sons; 1999.
25. Foster A, Jankowiak M, Bingham E, *et al.* Variational bayesian optimal experimental design. arXiv190305480, 2019, preprint: not peer reviewed.
26. Sokal A. Monte Carlo methods in statistical mechanics: foundations and new algorithms In: *Functional Integration*. Springer; 1997: 131–92.
27. Foreman-Mackey D, Hogg DW, Lang D, *et al.* emcee: the MCMC hammer. *Publ Astron Soc Pacific* 2013; 125 (925): 306–12.
28. Kucirka LM, Lauer SA, Laeyendecker O, *et al.* Variation in false-negative rate of reverse transcriptase polymerase chain reaction–based SARS-CoV-2 tests by time since exposure. *Ann Intern Med* 2020; 173 (4): 262–7.
29. Wikramaratna PS, Paton RS, Ghafari M, *et al.* Estimating the false-negative test probability of SARS-CoV-2 by RT-PCR. *Eurosurveillance* 2020; 25 (50): 2000568.
30. Cohen AN, Kessel B. False positives in reverse transcription PCR testing for SARS-CoV-2. *MedRxiv* 2020. 10.1101/2020.04.26.20080911
31. Yelin I, Aharony N, Tamar ES, *et al.* Evaluation of COVID-19 RT-qPCR test in multi sample pools. *Clin Infect Dis* 2020; 71 (16): 2073–8.
32. World Health Organization. Antigen-detection in the diagnosis of SARS-CoV-2 infection using rapid immunoassays. 2020. https://www.who.int/publications/i/item/antigen-detection-in-the-diagnosis-of-sars-cov-2infection-using-rapid-immunoassays. Accessed August 24, 2021.
33. Centers for Disease Control and Prevention. Interim guidance for antigen testing for SARS-CoV-2. 2020. https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/antigen-tests-guidelines.html. Accessed August 24, 2021.
34. Alexanderian A, Petra N, Stadler G, *et al.* A-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems with regularized $\ell 0$-sparsification. *SIAM J Sci Comput* 2014; 36 (5): A2122–A2148.