medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

BAYESIAN GROUP TESTING WITH DILUTION EFFECTS

Curtis Tatsuoka* and Weicong Chen

January 2021

Abstract

A Bayesian framework for group testing under dilution effects is introduced. This work has particular relevance given the pressing public health need to enhance testing capacity for COVID-19, and the need for wide-scale and repeated testing for surveillance. The proposed Bayesian approach allows for dilution effects in group testing and for general test response distributions beyond just binary outcomes. It is shown that even with strong dilution effects, an intuitive and simple-to-implement group testing selection rule, referred to as the Bayesian halving algorithm, has attractive optimal properties. A web-based calculator is introduced to assist and guide decisions on when and how to pool under various conditions.

KEY WORDS: group testing, surveillance, Bayesian, dilution effects, optimal rates of convergence.

1 Introduction

The group testing formulation originated by Dorfman [1] has found use in a diverse array of applications, including COVID-19 testing. The motivating idea of group testing for a disease like COVID-19 is that, say if biomarker samples from $N, N > 1$ subjects are pooled, and if the prevalence is low, most likely the test result for the pooled sample will be negative, indicating that all N subjects in the pool are negative, using only in one test. On the other hand, if the result indicates a positive test, and hence that there is at least one positive sample present among the pool, then further testing can be conducted to identify the positive subjects. Dorfman suggested that all samples which comprise a positively-tested pool should subsequently be tested individually. This approach has been adopted broadly. Nonetheless, in the presence of testing error, this approach can be problematic, in that classification error thresholds may not be reached, and it may not be comparatively efficient, as pooling subsets can still be useful. An important potential source of testing error is through dilution, which may occur when pooled samples contain few positives relative to

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

a larger number of negatives. Described here is a Bayesian framework for systematically addressing these important issues surrounding pooled testing such as for COVID-19.

Recently, the Food and Drug Administration (FDA) has released guidelines for allowing group testing of patients suspected of having COVID-19. There is thus a renewed flurry of interest in group testing for COVID-19 [2, 3, 4, 5, 6], as it can play a fundamental role in efficient disease surveillance. As there is a large percentage of asymptomatic cases, widespread testing is important for understanding and controlling spread. There also is a need for monitoring through repeated testing of large swaths of the population, such as for frontline and essential workers, and those at high risk. A potential drawback to such type of surveillance is the reliance on imprecise testing. PCR-based assays based on samples taken with nasal swabs can be highly accurate, although issues with sample collection or mis-priming can lead to errors. False negative rates with PCR-based assays with COVID-19 have for instance been estimated to range from 0.02 to 0.29 [7, 8]. Fast, less expensive, and non-intrusive approaches for testing, such as those based on saliva or antigens, can broaden the scope of testing, but may be even less accurate.

A basis for the proposed group testing methods are models known as lattices, in which states in the model follow a partial ordering. This order structure is used to guide next stage pool selection, and to gain insight into the statistical properties of group testing with dilution effects. The proposed Bayesian group testing approach explicitly acknowledges testing error and heterogeneity in individual risk levels through prior distribution specification, and can be applied with either quantitative or categorical test responses. Importantly, individual-level classification error can be systematically be reduced to any error threshold, as our proposed approach has attractive optimality properties in which correct classification can efficiently be made for all test subjects. A framework for sequential Bayesian classification on lattices has been described earlier [9]. This current work differs from the previous formulation as we now consider test response distributions that depend on pool size and the number of positives in the pool. In terms of theoretical work involving partially ordered classification models, rules that rely on Kullback-Leibler information of response distributions that are specific to the experiment, such as in cognitive and educational adaptive testing, has been studied [10], as well as in group testing and partially ordered classification models [11, 12].

Group testing has a rich statistical literature that has expanded upon Dorfman's seminal work. An important result in group testing by Ungar [13] gives a demarcation for when it is optimal to group test when responses are binary and there is no testing error. A Bayesian approach was first described in Sobel and Groll [14], where binary responses without testing error are assumed. In addition to screening for disease, group testing has been used in genetics testing [15, 16], and to identify promising drug compounds [17, 18]. The phenomenon of testing error in group testing has been widely recognized [19, 20, 21, 22, 23], including dilution effects with specific functional forms [20, 24, 25]. The possibility that objects have different probabilities of being positive but with no

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

testing error has been considered [26, 27]. Test response distributions beyond Bernoulli distributions have for instance been employed where responses are assumed to be normally distributed [24], and group testing has been studied in the context of multiplex and high throughput testing [28, 6]. Estimation of proportion such as for disease prevalence is an important application in group testing. Previous work has been based on binary outcomes [14, 29, 30, 31, 32, 33].

Below, we will first introduce the proposed method and related notions, with an example. We then summarize theoretical properties for optimal designs in Bayesian group testing, and for an intuitive pooled test selection rule. Finally, we introduce a web-based tool that implements the proposed algorithms and graphically represents results. Example scenarios are presented, which demonstrate the efficiency gains of group testing. Proofs of theorems and a more thorough technical discussion are given in the Appendix, as well as a brief note on how Bayesian estimation of prevalence can be conducted in the proposed framework.

2 Methods

2.1 Lattice Models for Bayesian Group Testing

We illustrate how lattice-based classification models can be applied in Bayesian group testing with dilution effects.

Example 1. Consider all the possible subsets generated by subjects A and B, as in Figure 2.1. Suppose each subset represents a profile of the subjects that are negative for COVID-19. Hence, one of these subsets represents the true state. The collection of these subsets thus form the classification model, and we henceforth refer to them as states: State AB denotes that both A and B are negative, state A denotes that only A is negative, and hence B is positive. Similarly, state B denotes that only B is negative. State $\hat{0}$ represents that both subjects are positive.

Denote the collection of these states as S. In the above example, $S = \{AB, A, B, \hat{0}\}$. These states (i.e. all the possible subsets of subjects that are negative) actually have a specific order structure in that they can be ordered by inclusion, and form what is known as a powerset lattice. When ordering by inclusion, note for instance that state AB is greater than states A, B, and $\hat{0}$ because the set comprised of both A and B contains A alone, B alone, and $\hat{0}$, respectively. On the other hand, states A and B are incomparable, as there is no inclusion relationship between the respective subsets. In this example, there are two subjects, and 4 possible profiles of them being positive or negative. In general, for N subjects, there are 2^N such possibilities, corresponding to the powerset of the N subjects. Further, note that there is essentially a one-to one correspondence between states in lattice model and potential pooled test. There are 2^N - 1 possible way to pool tests, since the empty set, represented as $\hat{0}$, is not a possible test. So, in this example, for state AB, there is also a pooled test

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

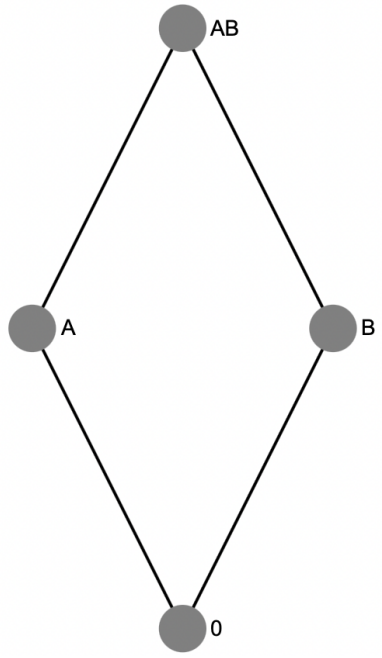

Figure 1: Hasse diagram of lattice model with 2 subjects, A and B

with samples from A and B; for state B, there is a test consisting of a sample only from subject B, etc.

We adopt a Bayesian approach. An advantage is the allowance for individual-level prior information about positivity into the classification process, as well as systematic characterization of uncertainty from potential testing error. In a Bayesian framework, the posterior probability values embody the combined empirical and prior evidence as to which state in the classification model is the true one. Larger posterior probability values for the true state are clearly desirable, with a probability value of 1 indicating certainty in correctly identifying the true status of positivity for all the subjects. A *stage* of testing is an iteration of preparing a collection of (pooled) samples for testing and observing the results of the test(s). The posterior distribution after n stages of testing on the collection of classification states S will be denoted as π_n, and for a state $j \in S$, the posterior probability value that state j is the true state is $\pi_n(j)$. Corresponding prior distribution values are denoted as π_0 and $\pi_0(j)$.

For instance, prior to testing for COVID-19, each subject could be assigned a prior probability of positivity based on logistic regression modeling, with explanatory variables such as age, gender, location of residence, travel history, status of household contacts, self-reported exposure, community infection rates, etc. For a given profile of the N subjects in terms of their positivity status, to obtain its prior probability, we take the product of the respective individual

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

prior probabilities for their respective status in the profile. These values form a prior probability distribution across the possible states. After a (pooled) test result is observed, using Bayes rule, prior probabilities are updated to posterior probabilities. We illustrate these computations in the example below.

Example 1, continued. Suppose that for a subject A the prior probability that he/she is a positive (e.g. has COVID-19) is 0.05, and independently for a higher at-risk subject B it is 0.10. The prior probabilities for state membership follows: for state $AB = \hat{1}$, $\pi_0(AB) = (0.95)(0.90) = 0.855$, for state A it is $\pi_0(A) = (0.95)(0.10) = 0.095$, $\pi_0(B) = (0.05)(0.90) = 0.045$, and $\pi_0(\hat{0}) = 0.005$. Possible tests include pooling A and B, and testing each subject individually. Suppose also that a test has two possible outcomes, indicating that either at least one positive is present in the pool, or that only negatives are in the pool. Assume that the specificity, the probability of observing a negative outcome given that the pool does in fact consist only of negative samples, is 0.99. Also, assume that the sensitivity, the probability of observing a positive outcome given that the pool contains at least one positive sample, also is 0.99 when all samples are positive, and reduces slightly to 0.98 when only one of two samples is positive.

Let the first test be comprised of pooled samples from A and B. Then, if a negative outcome is observed, it can be seen by Bayes rule that $\pi_1(AB) = 0.9966$. Hence, combined with the prior information, this test result would lead to strong indication that both of the samples are negative. If instead a positive test outcome is seen, then $\pi_1 = \{\pi_1(AB) = 0.0567, \pi_1(A) = 0.6177, \pi_1(B) = 0.2926, \pi_1(\hat{0}) = 0.0328\}$. Suppose next that samples from A and B are tested individually in the next stage as in Dorfman, and the outcome for A is negative, and for B is positive. If we update the posterior probabilities for the two observed outcomes simultaneously, then, $\pi_2 = \{\pi_2(AB) = 0.0009, \pi_2(A) = 0.9985, \pi_2(B) = 0.0001, \pi_2(\hat{0}) = 0.0005\}$. In particular, the posterior probability that B is positive is the sum of respective state membership probabilities that indicate that B is positive: $\pi_2(A) + \pi_2(\hat{0}) = 0.9990$. This is clear evidence that indeed B is positive. Similarly, note that the probability that A is positive is $\pi_2(B) + \pi_2(\hat{0}) = 0.0006$, which indicates that A is clearly negative.

Making "optimal" pooled test selections becomes more interesting and complex as the number of subjects N being considered for pooling increases, the individual risks of positivity vary, and as testing error increases through dilution effects. We consider the following criterion as a basis for optimality in pool selection, from a theoretical perspective. First, let $s \in S$ be the true state (i.e. representing the correct identification of negatives and positives). Typically, $1 - \pi_n(s)$ converges exponentially to 0 at the order $e^{-\alpha n}$ for some α, and α is called the *rate of convergence*. The mathematical definition of the rate of convergence is taken to be

$$\alpha = \liminf_{n \to \infty} -(1/n) \log(1 - \pi_n(s)).$$

We establish group testing designs and selection rules that attain the optimal rates of convergence (largest possible alpha) with probability one under general

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

dilution effects. Note that rates are in terms of number of tests administered, with one test per stage. We show that they are also characterized in terms of Kullback-Leibler information values, which are discrepancy measures for probability distributions.

2.2 Dilution Effects and Optimal Designs

Note that as pool size increases, it is possible that testing outcomes can become less reliable. For instance, for a given number of positive samples, it may be increasingly harder to detect their presence as pool size increases. Another dilution effect can occur in relation to the number of positive samples in a pool. For instance, for a given pool size, it may be increasingly easier to detect the presence of positive subjects as their number increases. Such dilution effects and their effect on the optimality of pooled test selection in terms of attaining the optimal rate of convergence is characterized by Theorem 1 in the Appendix.

We assume the following conditions that characterize dilution effects, which are formally represented in terms of Kullback-Leibler information values (see Appendix). We assume that i) for a given number of positive samples, as pool size increases, pooled tests become less discriminatory in identifying the presence of positive samples; ii) as the difference in number of positive samples increases for a given pool size, it is relatively easier to discriminate the presence of the positives; (iii) for a given pool size and a fixed difference in the number of positive samples present, as the total number of positive samples increases, the respective test response distributions become harder to discriminate. Finally, we assume a non-restrictive technical condition, (A4) in the Appendix. Theoretical demarcations when it is preferable to group test depend on which state is true. We illustrate examples of demarcations in Appendix, which show that group testing is attractive asymptotically even under strong dilution effects.

Under these conditions, optimal designs can be considered for three case: 1) all subjects are positive (the bottom state in the lattice), 2) all subjects being tested are negative (the top state in the lattice), and 3) there is a mix of negative and positive subjects (a state in the lattice "in-between" the top and bottom states). In Theorem 1, we see that optimal strategies for attaining the fastest possible rates of convergence are: 1) test subjects individually, and 2) to group test samples from all subjects. For the third case, optimal strategies are more complex. The covers and anti-covers of the true state (states directly above and below it) must be considered, as these states pose the most difficulty in discriminating the true state, and hence have posterior probability values that converge to zero the slowest. Under general conditions, the optimal strategy is to group test the negatives and individually test the positives. Optimal rates depend on the statistical discriminatory efficiency of the pooled tests, as measured by Kullback-Leibler information. They also depend on the complexity of the lattice around the true state, as measured by the number of states directly surrounding the true state. Respective tests are selected proportionally so that the rate at which the slowest of the non-true state posterior probability value converges to zero is maximized. In other words, it is desirable that the pos-

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

terior probabilities of the covers and anti-covers converge to zero at the same rate. More technical details are discussed in the Appendix. In practice, the true state identity is not known. Next, we propose a testing selection rule that will eventually identify the true state and adopt the corresponding optimal strategy, hence attaining the optimal rate of convergence, no matter the true state.

2.3 Bayesian Halving Algorithm and Analogous k-Test Look Ahead Rules

A key to obtaining systematic efficiency gains is to base the sequential selection of pooled tests on the Bayesian information provided by the posterior probability values for the states. In Theorem 2 of the Appendix, we will establish that under general conditions for dilution effects, a simple and intuitive selection rule for pooling, referred to as the *Bayesian halving algorithm*, attains the optimal rate for the true state posterior probability to converge to the value 1. While this is an asymptotic property, given that the rates are exponential, attaining optimal rates is practically attractive for finite testing horizons as well, and insures that any error thresholds will be attained eventually with sufficient testing.

A key concept from order theory is the idea of an up-set for a state in the lattice. This concept is useful for motivating the proposed test selection rules. For a state j, denote its up-set as j. This is defined as the subset of states within the lattice that are at least as great as (i.e. contain) the state in question. In Figure 2.1, $AB = \{AB\}$; for $B, B = \{B, AB\}$, for $A, A = \{A, AB\}$, and for $\hat{0}$ it is all 4 of the states. Practical interpretations of these up-sets are as follows: the up-set of AB is all the states for which a pool of A and B would contain only negative subjects, the up-set of B is all the states such that B is negative, etc. Conversely, the complement of the up-set of $AB = \{A, B, \hat{0}\}$, and these are the states for which a pooled test AB would contain at least one positive sample. Similarly, the complement of the up-set of $B = \{A, \hat{0}\}$ is comprised of the states for which an individual test of a sample from subject B would be positive, etc. Hence, the up-set of a state and its complement generate a partition of the classification states according to whether the corresponding pooled test would contain all negatives or not.

Now let us consider rules for pooled test selection based on π_n. It is desirable to have a rule that adopts optimal strategies and attains optimal rates of convergence of the true state posterior probability value to 1, no matter which state is true, and under dilution effects, which are likely to arise especially for large pool sizes. The following intuitive pooled test selection rule has these properties, as we establish in Theorem 2 in the Appendix.

Consider the Bayesian halving algorithm, which selects the pool e that minimizes

$$h(\pi_n, e) = |m_n(e) - 0.5|, \text{ where } m_n(e) = \sum_{j \in upe} \pi_n(j). \tag{1}$$

The Bayesian halving algorithm systematically partitions the lattice model, so that regardless of the actual observed outcome, through Bayes rule, one of the

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

two partitions of states will get increased posterior probability values, while the states in the other partition will get decreased values. This systematically encourages posterior mass to eventually accumulate to one state. This property supports that purposeful and systematic pool choices will be made as test results are observed that will lead to correct classification. From Example 1, we compute $h(\pi_0, AB) = 0.355, h(\pi_0, A) = 0.45, h(\pi_0, B) = 0.40$. Hence, it would be attractive to first select $e = AB$, and to group test.

We also extend the halving algorithm to an analogous k-test look ahead rule, $k > 1$, so that k pools are selected for simultaneous testing, as opposed to the one at a time testing implicit with the proposed halving algorithm. Selecting k pooled samples at a time may be more practical, as this can reduce the back and forth involved in preparing pooled samples for testing based on observed results. We define a stage of testing to be the simultaneous submission of one or more pooled samples for testing. Given π_n at a stage n, the idea behind our proposed k-test version of the halving algorithm is to partition the posterior mass on the lattice model into 2^k partitions according to sets of the following form. We choose k pooled samples $e_1, \ldots, e_k$ that minimize

$$|\Sigma_{j \in e_1 \cap e_2 \ldots e_k} \pi_n(j)^- - 1/2^k| + |\Sigma_{j \in e_1 \cap e_2 \ldots e_k^c} \pi_n(j)^- - 1/2^k| + \ldots$$

$$|\Sigma_{j \in e_1^c \cap e_2^c \ldots e_k^c} \pi_n(j)^- - 1/2^k|.$$

For example, when $k = 2$, this criterion involves selecting e_1, e_2 that splits the posterior mass on the lattice model into $2^k = 4$ parts as equally as possible, minimizing

$$|\Sigma_{j \in e_1 \cap e_2} \pi_n(j)^- - 1/4| + |\Sigma_{j \in e_1 \cap e_2^c} \pi_n(j)^- - 1/4| +$$
$$|\Sigma_{j \in e_1^c \cap e_2} \pi_n(j)^- - 1/4| + |\Sigma_{j \in e_1^c \cap e_2^c} \pi_n(j)^- - 1/4|. \quad (2)$$

We can approximate the minimization in (2) by sequentially selecting the k experiments one at a time. First, given current posterior probabilityπ_n, we choose e_1 with the halving algorithm, e_2 that minimizes (2) given π_n and e_1, etc. This approximation is computationally much faster.

A halving algorithm that uses prior probabilities is previously described in Black et al.[27], where samples are successively split in half for positive pooled tests while also accounting for heterogeneity in risk. This approach is related to what we propose here, but our approach is fully Bayesian, with posterior probability computation that serves as the basis for selection and stopping. Latvik et al.[21] present a set of several rules that are related to the halving algorithm. However, the user has to select a rule from this set. These rules differ in the amount of re-testing of pools that is conducted. In contrast, the proposed Bayesian halving algorithm is an automated rule. It also can be used with general distributions for test results.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

3 Results

3.1 A Web-Based Calculator and Simulations

For the Bayesian halving algorithm and its k-test look ahead counterparts, $k > 1$, we have created a web-based calculator that generates important performance statistics relating to expected number of tests and stages, and correctness rates, as well as false positive and false negative rates. The website address is *www.bayesgrouptest.case.edu*. Expectations are taken with respect to π_n. Required inputs are a vector of prior probabilities of positivity for each subject, a matrix with the probabilities indicating presence of a positive that depend on pool size and the number of positives in the pool, and a target posterior probability error threshold for individual classification. Pooled test sequences are also generated. Note that this tool currently only employs binary outcomes for pooled tests ("positive is present", or not).

Example 2 We next consider examples with $N = 12$ as an illustration, with varying individual prior probability risk levels: 1) all prior probability values of positivity p_0 for each subject is 0.02, 2) all such values are 0.20, 3) mixed prior values, with one subject having the value 0.20, while the rest having the value 0.02. We also assume a dilution effect model modified from Hung and Swallow [25], which depends on constants $0 < \alpha < 1$ and $h > 0$, the number of positives in a pool e, denoted as r, and pool size $|e|$. Suppose responses are binary, with specificity $p_{r=0,|e|} = \alpha$, and one minus the sensitivity value

$$q_{r,|e|} = \alpha \cdot r/((|e| - r) \cdot h + r), 1 \leq r \leq |e|.$$

For our examples, we assume $\alpha = 0.99$, and $h = 0.005$. As an illustration, when pool size is 12, sensitivity ranges from 0.9384 when only one subject is positive, to 0.99 when all of the 12 subjects in the pool are positive. The webtool can generate matrices of dilution effects for varying α and h. The dilution effect conditions required in Theorem 2, which relate to optimality of the Bayesian halving algorithm, are satisfied with this example. Finally, stopping of testing is assessed at the individual level. Note that the posterior probability that a subject is negative (positive) is equal to the sum of the posterior probability values of all the states for which the subject is represented as negative (positive). In our example, an individual subject is classified and no longer considered for further testing once this posterior classification error of being negative or positive is at or below a given error threshold, in this case 0.01. In these settings, we consider performance of individual testing, k-test look ahead rules, and the Bayesian halving algorithm (denoted as $k = 1$).

Simulations are based on exhaustive analyses of possible test response sequences, up to a number stages that varies by algorithm. This variation was due in part due to numerical challenges as k increases. Hence, the reported averages and error levels are approximations. Note for $k = 1$: 16 stages; $k = 2$: 12 stages; $k = 3$: 8 stages; $k = 4$: 5 stages for $_0 = 0.02$ and for the case with mixed $_0$ values, 6 stages for $_0 = 0.2$; Individual: 5 stages. Overall classification up to these stages is generally above 99

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Simulation	$p_0 = 0.02$	$p_0 = 0.20$	mixed p_0 cases
k=1	2.071	8.648	3.313
k=2	3.238	10.72	3.554
k=3	4.305	13.161	4.926
k=4	5.508	14.092	6.188
Individual	15.948	22.5	16.884

Table 1: Average number of tests to reach classification threshold

Simulation	$p_0 = 0.02$	$p_0 = 0.20$	mixed p_0
k=1	2.071	8.648	3.313
k=2	1.619	5.36	1.777
k=3	1.435	4.387	1.642
k=4	1.377	3.523	1.547
Individual	1.329	1.875	1.407

Table 2: Average number of stages to reach classification threshold

$_0 = 0.02$	False positive	false negative rates
k=1	0.022 %	2.186%
k=2	0.011 %	1.933 %
k=3	0.007 %	1.873 %
k=4	0.005 %	1.806 %
Individual	0.0 %	0.214 %
$_0 = 0.2$	False positive	false negative rates
k=1	0.618 %	5.002%
k=2	0.721 %	5.125 %
k=3	0.557 %	4.604 %
k=4	0.376 %	3.958 %
Individual	0.0 %	2.102 %
mixed $_0$	False positive	false negative rates
k=1	0.041 %	2.61 %
k=2	0.136 %	1.962 %
k=3	0.163 %	1.7 %
k=4	0.159 %	1.608 %
Individual	0.0 %	0.375 %

Table 3: False positive and false negative rates for N = 12

Simulation	$p_0 = 0.02$	$p_0 = 0.2$-	mixed p_0
k=1	ABCDEFGHIJKL	ABC	ABCDEFGHIJKL
k=2 (add to k=1)	A	DEH	A
k=3 (add to k=1,2)	ABCDE	ABCDE	ABCDEFGH
k=4 (add to k=1,2,3)	ABCDEFGHI	ABCDH	ABCDEFGH

Table 4: First stage test selections

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 1 shows, as expected, that in terms of average number of tests needed to classify all subjects, group testing with the Bayesian halving algorithm ($k = 1$) has the smallest average number of tests and individual testing the largest, across all the scenarios and selection rules considered here. Indeed, the reduction achieved through group testing can be dramatic. Testing efficiency is particularly striking when the prior probability values of positivity for subjects are small. This gain in testing efficiency reduces as k increases among k-test look ahead rules. Conversely, as seen in Table 2, in terms of the number of stages, individual testing has the smallest average while the $k = 1$ case has the largest. This isn't surprising given that more tests are submitted per stage as k increases. There is thus a clear trade off between minimizing number of tests versus number of stages, with the intermediate values of $k > 1$ serving as a compromise between the flexibility of single test at-a-time ($k = 1$) group testing versus the relative logistical ease of individual testing. We only consider rules with $k \leq N/3$ as this allows for sufficient adaptability with respect to test responses, as well as numerical feasibility. Larger values of k require greater computational resources in generating and analyzing the possible testing sequences.

In terms of average false positive and false negative rates, note in Table 3 that false negative rates are higher, particularly as k decreases, and as p_0-values increase. It is possible to adjust error thresholds for negative versus positive classification, making the threshold for negative classification relatively more stringent. Table 4 lists the first stage test selections. Multi-stage lists can be generated with the webtool. In the table, note for instance when k=2 and $p_0 = 0.02$ for all subjects, the first two tests selected for the first stage is to pool all subjects and to individually test subject A. For all prior probability scenarios, group testing is initially selected. In sum, deciding between a specific value of $k \geq 1$ and individual testing depends on the prior probability values, the dilution effect, and the "cost" of preparation per stage versus the cost of testing itself. The webtool provides a way to assess the tradeoffs that are appropriate to specific scenarios.

3.2 Conclusion

In a Bayesian group testing framework, dilution effects are considered that depend on pool size and the number of positives in a pool. It is seen that even when dilution effects are quite significant, asymptotically optimal strategies rely on group testing. Moreover, a simple and intuitive Bayesian halving algorithm is shown to attain optimal rates of convergence. k-test look ahead analogues of this halving algorithm also are proposed, which still can take advantage of pooling and adaptability, while reducing the number of stages in testing. A web-based tool has been developed that allows for assessment of the proposed group testing approaches, given inputted prior probabilities of positivity per subject, dilution effects and classification error level thresholds. The respective expected number of tests and stages, classification error, and comparisons to individual testing are generated. Computation of performance statistics for the

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

lattice models does present combinatorial challenges as N and k increases.

Detailed discussion on the estimation of dilution effects is deferred here, as approaches will depend on the test response distribution modeling. It is possible to consider estimating response probabilities with uncertain diagnoses (in relation to posterior probabilities in classification) using Bayesian latent class estimation approaches that can include nonparametric density estimation [34, 35]. Varied and well-characterized pooled test data should become more accessible as group testing becomes more widespread through population-level surveillance. Also, individual-level prior probabilities of positivity can for instance be estimated through generalized linear regression models based on demographic, geographic, clinical and exposure information. Finally, we conjecture that it will be possible to embed Bayesian halving algorithm-based testing sequences in high-volume, automated workflows.

Acknowledgements. This work was supported by grants NSF DRL1561716, R01 MH65538 and UL1TR002548. **Author contributions.** CT contributed to statistical theory, numerical implementation, simulation development, manuscript development; WC contributed to software development, numerical implementation, simulation development, manuscript development. **Competing interests.** None **Materials and Correspondence.** Contact Curtis Tatsuoka, cmt66@case.edu.

References

[1] Robert Dorfman. "The Detection of Defective Members of Large Populations". In: *The Annals of Mathematical Statistics* 14.4 (1943), pp. 436–440. ISSN: 00034851. URL: http://www.jstor.org/stable/2235930.

[2] Stefan Lohse et al. "Pooling of samples for testing for SARS-CoV-2 in asymptomatic people". In: *The Lancet Infectious Diseases* 20.11 (2020), pp. 1231–1232. DOI: 10.1016/s1473-3099(20)30362-5.

[3] Farhan Majid, Saad B Omer, and Asim Ijaz Khwaja. "Optimising SARS-CoV-2 pooled testing for low-resource settings". In: *The Lancet Microbe* 1.3 (2020). DOI: 10.1016/s2666-5247(20)30056-2.

[4] Catherine A. Hogan, Malaya K. Sahoo, and Benjamin A. Pinsky. "Sample Pooling as a Strategy to Detect Community Transmission of SARS-CoV-2". In: *Jama* 323.19 (2020), p. 1967. DOI: 10.1001/jama.2020.5445.

[5] Noam Shental et al. "Efficient high-throughput SARS-CoV-2 testing to detect asymptomatic carriers". In: *Science Advances* 6.37 (2020). DOI: 10.1126/sciadv.abc5961. eprint: https://advances.sciencemag.org/content/6/37/eabc5961.full.pdf. URL: https://advances.sciencemag.org/content/6/37/eabc5961.

[6] *The Mathematics of Mass Testing for COVID-19*. https://sinews.siam.org/Details-Page/the-mathematics-of-mass-testing-for-covid-19.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[7] Ingrid Arevalo-Rodriguez et al. "False-Negative Results Of Initial RT-PCR Assays For Covid-19: A Systematic Review". In: *medRxiv* (2020). DOI: 10.1101/2020.04.16.20066787. eprint: https://www.medrxiv.org/content/early/2020/04/21/2020.04.16.20066787.full.pdf. URL: https://www.medrxiv.org/content/early/2020/04/21/2020.04.16.20066787.

[8] Steven Woloshin, Neeraj Patel, and Aaron Kesselheim. "False Negative Tests for SARS-CoV-2 Infection — Challenges and Implications". In: *New England Journal of Medicine* 383 (June 2020). DOI: 10.1056/NEJMp2015897.

[9] Curtis Tatsuoka and Thomas Ferguson. "Sequential classification on partially ordered sets". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65.1 (2003), pp. 143–157. DOI: 10.1111/1467-9868.00377.

[10] Curtis Tatsuoka. "Sequential Classification on Lattices with Experiment-Specific Response Distributions". In: *Sequential Analysis* 33.3 (2014), pp. 400–420. DOI: 10.1080/07474946.2014.916931.

[11] Thomas S. Ferguson and Curtis Tatsuoka. "An optimal strategy for sequential classification on partially ordered sets". In: *Statistics Probability Letters* 68.2 (2004), pp. 161–168. DOI: 10.1016/j.spl.2004.02.007.

[12] Curtis Tatsuoka. "Optimal sequencing of experiments in Bayesian group testing". In: *Journal of Statistical Planning and Inference* 133.2 (2005), pp. 479–488. ISSN: 0378-3758. DOI: https://doi.org/10.1016/j.jspi.2004.01.011. URL: http://www.sciencedirect.com/science/article/pii/S0378375804001636.

[13] Peter Ungar. "The cutoff point for group testing". In: *Communications on Pure and Applied Mathematics* 13.1 (1960), pp. 49–54. DOI: 10.1002/cpa.3160130105.

[14] Milton Sobel and Phyllis A. Groll. "Binomial Group-Testing with an Unknown Proportion of Defectives". In: *Technometrics* 8.4 (1966), p. 631. DOI: 10.2307/1266636.

[15] Joseph L. Gastwirth. "The Efficiency of Pooling in the Detection of Rare Mutations". In: *The American Journal of Human Genetics* 67.4 (2000), pp. 1036–1039. DOI: 10.1086/303097.

[16] Pak Sham et al. "DNA Pooling: A tool for large-scale association studies". In: *Nature reviews. Genetics* 3 (Dec. 2002), pp. 862–71. DOI: 10.1038/nrg930.

[17] Minge Xie et al. "Group Testing With Blockers and Synergism". In: *Journal of the American Statistical Association* 96.453 (2001), pp. 92–102. DOI: 10.1198/016214501750333009.

[18] Katja S Remlinger et al. "Statistical Design of Pools Using Optimal Coverage and Minimal Collision". In: *Technometrics* 48.1 (2006), pp. 133–143. DOI: 10.1198/004017005000000481.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[19] Lois E. Graff and Robert Roeloffs. "Group Testing in the Presence of Test Error; An Extension of the Dorfman Procedure". In: *Technometrics* 14.1 (1972), pp. 113–122. DOI: 10.1080/00401706.1972.10488888.

[20] F. K. Hwang. "Group Testing with a Dilution Effect". In: *Biometrika* 63.3 (1976), pp. 671–673. ISSN: 00063444. URL: http://www.jstor.org/stable/2335750.

[21] Eugene Litvak, Xin M. Tu, and Marcello Pagano. "Screening for the Presence of a Disease by Pooling Sera Samples". In: *Journal of the American Statistical Association* 89.426 (1994), pp. 424–434. DOI: 10.1080/01621459.1994.10476764.

[22] Nicolas Thierry-Mieg. "A new pooling strategy for high-throughput screening: The Shifted Transversal Design". In: *BMC bioinformatics* 7 (Feb. 2006), p. 28. DOI: 10.1186/1471-2105-7-28.

[23] Hae-Young Kim et al. "Comparison of Group Testing Algorithms for Case Identification in the Presence of Test Error". In: *Biometrics* 63.4 (2007), pp. 1152–1163. ISSN: 0006341X, 15410420. URL: http://www.jstor.org/stable/4541470.

[24] Stefanos A. Zenios and Lawrence M. Wein. "Pooled testing for HIV prevalence estimation: exploiting the dilution effect". In: *Statistics in Medicine* 17.13 (1998), pp. 1447–1467. DOI: 10.1002/(sici)1097-0258(19980715)17:13<1447::aid-sim862>3.0.co;2-k.

[25] M. Hung and William H. Swallow. "Robustness of Group Testing in the Estimation of Proportions". In: *Biometrics* 55.1 (1999), pp. 231–237. DOI: 10.1111/j.0006-341x.1999.00231.x.

[26] F. K. Hwang. "A Generalized Binomial Group Testing Problem". In: *Journal of the American Statistical Association* 70.352 (1975), pp. 923–926. DOI: 10.1080/01621459.1975.10480324.

[27] Michael S. Black, Christopher R. Bilder, and Joshua M. Tebbs. "Group testing in heterogeneous populations by using halving algorithms". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 61.2 (2012), pp. 277–290. ISSN: 00359254, 14679876. URL: http://www.jstor.org/stable/41430963.

[28] Christopher R. Bilder, Joshua M. Tebbs, and Christopher S. McMahan. "Informative group testing for multiplex assays". In: *Biometrics* 75.1 (2019), pp. 278–288. DOI: https://doi.org/10.1111/biom.12988. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/biom.12988. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.12988.

[29] Joseph L. Gastwirth and Patricia A. Hammick. "Estimation of the prevalence of a rare disease, preserving the anonymity of the subjects by group testing: application to estimating the prevalence of aids antibodies in blood donors". In: *Journal of Statistical Planning and Inference* 22.1 (1989), pp. 15–27. DOI: 10.1016/0378-3758(89)90061-x.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[30] J. L. Gastwirth and Wesley O. Johnson. "Screening with Cost-Effective Quality Control: Potential Applications to HIV and Drug Testing". In: *Journal of the American Statistical Association* 89.427 (1994), pp. 972–981. DOI: 10.1080/01621459.1994.10476831.

[31] Jacqueline M. Hughes-Oliver and William H. Swallow. "A Two-Stage Adaptive Group-Testing Procedure for Estimating Small Proportions". In: *Journal of the American Statistical Association* 89.427 (1994), pp. 982–993. DOI: 10.1080/01621459.1994.10476832.

[32] Ron Brookmeyer. "Analysis of Multistage Pooling Studies of Biological Specimens for Estimating Disease Incidence and Prevalence". In: *Biometrics* 55.2 (1999), pp. 608–612. DOI: https://doi.org/10.1111/j.0006-341X.1999.00608.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0006-341X.1999.00608.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0006-341X.1999.00608.x.

[33] J. M. Tebbs. "Estimating ordered binomial proportions with the use of group testing". In: *Biometrika* 90.2 (2003), pp. 471–477. DOI: 10.1093/biomet/90.2.471.

[34] Curtis Tatsuoka. "Data analytic methods for latent partially ordered classification models". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 51.3 (2002), pp. 337–350. DOI: 10.1111/1467-9876.00272.

[35] Curtis Tatsuoka, Ferenc Varadi, and Judith Jaeger. "Latent Partially Ordered Classification Models and Normal Mixtures". In: *Journal of Educational and Behavioral Statistics* 38.3 (2013), pp. 267–294. ISSN: 10769986, 19351054. URL: http://www.jstor.org/stable/41999425.

Appendices

A1. Notation

We now focus on establishing the theoretical properties of Bayesian group testing with lattice models under dilution effects, and the discussion will henceforth be much more technical. We begin by reviewing the statistical formulation more formally. The set of classification states is taken to consist of the powerset of N subject profiles that indicate positivity status for each of N subjects. Each state in the model describes all the subjects in terms of whether each of the subjects is either *positive* or *negative* with respect to an outcome of interest. Note that there are 2^N possible states, corresponding to each element in the powerset. States can be identified through the subjects described as being negative, and can be partially ordered through set inclusion. Note that state $j \geq k$ if the collection of negative subjects associated with state j contains all the negative subjects associated with state k. The collection of these states thus form what will be referred to as a *powerset lattice*.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Let the collection of classification states S be a finite lattice, with an unknown true state. In brief, recall that a *lattice* is a partially ordered set such that any two elements have both a unique least upper bound (*join*) and a unique greatest lower bound (*meet*). For two elements $i, j \in S$, their join and meet are respectively denoted as $i \vee j$ and $i \wedge j$. The up-set of an element i is $i = \{j \in S : i \leq j\}$, and the down-set of i is $i = \{j \in S : j \leq i\}$. More generally, for a set $I \subseteq S$, $i = \{j \in S :$ there exists $i \in I$ such that $i \leq j\}$, and $I = \{j \in S :$ there exists $i \in I$ such that $j \leq i\}$. A top element $\hat{1}$ is an element such that for any $i \in S, i \leq \hat{1}$ Similarly, a bottom element $\hat{0}$ is an element such that for any $i \in S, \hat{0} \leq i$. Both a top and bottom element exist in a finite lattice. For $i \in S$, define C_i to be the set of covers of i, where $y \in S$ is a *cover* of an element i if $i < y$ and there does not exist $z \in S$ such that $i < z < y$. Also, let D_i denote the set of anti-covers, in other words the set of states $y \in S$ such that for $y < i$, there does not exist $z \in S$ such that $y < z < i$. In Figure 2.1, note that the anti-covers for state AB are the states A and B, while state AB is the only cover for both states A and B. In general, for a powerset lattice, the covers of a true state s are the states associated with one more negative subject, while its anti-covers are the states that are associated with one less negative subject. Hence, $|C_s| = N - |s|$ and $|D_s| = |s|$, where $|C_s|, |s|$, and $|D_s|$ denote respective cardinalities.

Let $\mathcal{E}$ be a finite collection of pooled tests. Denoting X as the random variable being observed for an experiment $e \in \mathcal{E}$, let $f(x|e, j)$ be the class conditional probability density for $j \in S$. Note that there is a one-to-one correspondence between elements in $\mathcal{E}$ and $S \backslash \hat{0}$, so that for instance $e = j$ for some $j \in S$ implies that the set of subjects out of N that are represented as negative by j are the corresponding samples being pooled. In establishing asymptotic results, assume each $e \in \mathcal{E}$ can be replicated an unlimited number of times. Also, let $K(f, g)$ represent the Kullback-Leibler information for distinguishing distribution f and g when f is true. Throughout, it will be assumed that for all $e \in \mathcal{E}$, there exists $j_1, j_2 \in S$ such that $K(f(\cdot|e, j_1), f(\cdot|e, j_2)) > 0$. In this case, e is said to *distinguish* states j_1 and j_2. Also, assume that for any $j_1, j_2 \in S$ and $e \in \mathcal{E}$, $K(f(\cdot|e, j_1), f(\cdot|e, j_2))$ is finite.

Denote $s \in S$ as the unknown true state. For $j \in S$, let $\pi_0(j)$ denote the prior probability that $j = s$. Assume $\pi_0(j) > 0$ for all $j \in S$. Given that at the first stage a test $e_1 \in \mathcal{E}$ is selected, and a random variable X_1 with density $f(\cdot|e_1, s)$ has observed value x_1, $\pi_1(j) \propto \pi_0(j) f(x_1|e_1, j)$. Inductively, at stage n for $n > 1$, conditionally on having chosen tests e_1, e_2,..., e_{n-1}, and having observed $X_1 = x_1, X_2 = x_2, \ldots, X_{n-1} = x_{n-1}$, a test $e_n \in \mathcal{E}$ is chosen and X_n with density $f(\cdot|e_n, s)$ is observed. The posterior probability that $j = s$ then becomes

$$\pi_n(j) \propto \pi_0(j) \prod_{i=1}^{n} f(x_i|e_i, j).$$

The posterior probability distribution on S at stage n will be denoted by π_n.

As seen in (Tatsuoka and Ferguson, 2003), it is necessary and sufficient to distinguish s from all other states in order for $\pi_n(s) \to 1$ almost surely. Letting

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

n_e be the number of times e is administered up to stage n, the *limiting proportion* that e is administered is denoted as p_e, with $p_e = \liminf n_e/n$. It is desirable to seek rules that sequentially select tests, $e_1, e_2, \ldots$, in the appropriate limiting proportions so that the posterior probability of the true state s, $\pi_n(s)$, converges almost surely to 1 at the fastest possible, or optimal rate. An *optimal strategy* is comprised of a collection of limiting proportions associated with each $e \in \mathcal{E}$ such that administration of the tests in their respective limiting proportions leads to convergence at the optimal rate. From Theorem 2 in (Tatsuoka and Ferguson, 2003), it is argued that for $k \in S, k \neq s$,

$$\liminf_{n\to\infty} -(1/n)\log(\pi_n(k)) = \sum_{e\in\mathcal{E}} p_e K(f(\cdot|e,s), f(\cdot|e,k)).$$

The right-hand side denotes the rate at which $\pi_n(k) \to 0$ as $n \to \infty$. This rate depends on the Kullback-Leibler information values for tests that distinguish state k from s, as well as the limiting proportions in which they are administered. Hence, larger Kullback-Leiber information values for tests are desirable, as they indicate greater discriminatory efficiency. It can be shown that α is the minimum of these rates among $k \neq s$, and hence maximizing α is equivalent to maximizing the rate of the slowest converging posterior probability terms among $k \neq s$.

A2. Dilution Effects Consider the following statistical formulation. Again, suppose the collection of tests $\mathcal{E}$ corresponds to all possible non-empty subsets of subjects and hence to $S\backslash\{\hat{0}\}$. In other words, every possible combination of the N can be pooled and tested. For $e \in \mathcal{E}$ and $j \in S$, assume that

$$f(x|e,j) = f_{r,|e|}(x), \tag{A1}$$

where $|e|$ denotes the number of samples in pool test e, and

$$r = |e| - |e \cap j|, r \geq 0,$$

where r is the number of positive samples in pool test e given j is the true state.

Note when $e \leq j$, this implies $r = 0$. This formulation allows for response distributions to vary depending on how many positive samples are in a pool, and according to pool size. When $f_{r,|e|}$ is a Bernoulli density, let $p_{r,|e|}$ be the probability that the outcome indicates that no positive samples are present given r positive samples are present and pool size is $|e|$. Note then that for $r = 0$ (no positives), $p_{0,|e|}$ represents the specificity of a test of pool size $|e|$. Also, for $r \geq 1$, $q_{r,|e|} = 1 - p_{r,|e|}$ is the sensitivity of the test when r positive samples are present in a pool of size $|e|$. In this section, let $K(r_1, r_2, |e|) = K(f_{r_1,|e|}, f_{r_2,|e|})$.

The following conditions, based on Kullback-Leibler information, are given to reflect the presence of dilution effects: (i) Suppose that both

$$K(r, r - r', |e|) \text{ and } K(r, r + r', |e|) \tag{A2}$$

are non-increasing in $|e|$ for fixed $r \geq 0$ and respectively for fixed $0 \leq r' \leq r$ and $r' \geq 0$, with $K(0, 1, |e|) > 0$ and $K(1, 0, |e|) > 0$. (ii) Suppose also that the Kullback-Leibler information values are non-decreasing in r' for fixed $r \geq 0$ and

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

$|e|$, with respectively $0 \leq r' \leq r$, and $0 \leq r' \leq |e| - r$. (iii) Moreover, suppose the functions in (A2) are non-increasing in r for fixed r' and $|e|$, with for each $r \geq 1$,

$$K(r, r-1, |e|) \geq K(r, r+1, |e|). \tag{A3}$$

A3. Optimal designs and optimal rates of convergence

The following examples illustrate issues that arise in determining optimal strategies under dilution effects, which are established in Theorem 1. An emphasis is on how the lattice structure determines the difficulty in classification. Given conditions (A1)-(A3), from a lattice-theoretic point of view, it is necessary and sufficient to distinguish a true state from its covers and anti-covers (the states that directly surround it) in order to distinguish the true state from all the others.

Example A1. Suppose that S is the lattice in Figure 2.1, and that $s = \hat{0}$ (all subjects are positive), as in part (i) of Theorem 1. The covers of s are the atoms of the lattice, in other words the states that respectively represent only one of the samples being negative, $C_s = \{A, B\}$. Consider now possible strategies for distinguishing s from its covers. One approach would be to individually test each subject. For instance, if sample from A is tested individually, then for states in $A = \{A, AB\}$, since A is negative, the distribution of response is $f_{0,1}$. If the true state is in $\{A\}^c = \{\hat{0}, B\}$, then subject A would be positive, and hence the corresponding distribution would be $f_{1,1}$. Hence, this test distinguishes $s = \hat{0}$ from A and AB. Similarly, states B and AB can be distinguished through individually testing B. In sum, when testing subjects individually, the corresponding atoms are positive distinguished one test at a time. However, when response distributions depend on the number of positives in a pool, the atoms also can be distinguished from the state $s = \hat{0}$ by pooling objects. Note if subjects A and B are pooled, then either state A or B being true would result in the test having distribution $f_{1,2}$, while for $s = \hat{0}$ it would be $f_{2,2}$. This follows since if state A or B are true, there is one positive sample in the pool, while given s is true, both sample are positive. Hence, more than one cover at a time can be distinguished from s through pooled experiments. Because of dilution effects (i) and (iii), $K(1, 0, 1) \geq K(2, 1, 2)$, so individual testing may lead to more efficient discrimination between s and its covers, while pooling A and B has the advantage of distinguishing s from A and B simultaneously. Hence, in the presence of dilution effects, there is a trade-off between these testing approaches. Theorem 1 resolves such trade-offs in terms of optimizing the rate of convergence. Note also that by dilution effect (ii), $K(2, 0, 2) \geq K(2, 1, 2)$, so state AB is distinguished from s at least as efficiently as states A and B when samples from A and B are pooled. Under individual testing, state AB is distinguished whenever states A or B are distinguished. Hence, with either approach, $\pi_n(AB)$ will converge to zero at a rate at least as fast as either $\pi_n(A)$ or $\pi_n(B)$.

Example A2. In part (ii) of Theorem 1, it is supposed that the true state

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

$s = \hat{1}$ (all subjects are negative). In Figure 2.1, state $AB = \hat{1}$. In distinguishing s from its anti-covers, $D_s = \{A, B\}$, again more than one strategy must be considered. When pooling A and B, note that the test has distribution $f_{0,2}$. States A and B each have class conditional distribution $f_{1,2}$. Hence, both of the anti-covers are distinguished simultaneously from s. It is also possible to distinguish states A and B by individually testing the subjects. For instance, when testing subject A, states A and AB have distribution $f_{0,1}$ since A would be negative if either is the true state, while states B and $\hat{0}$ would have distribution $f_{1,1}$, as A is positive if either of those states is true. Hence, the anti-cover B is distinguished. Similarly, the anti-cover A is distinguished when B is tested alone. Due to the increasing pool size, $K(0,1,1) \leq K(0,1,2)$, as reflected by dilution effect (i). Again, because of the presence of a dilution effect, there is a trade-off to consider in terms of distinguishing both anti-covers at once, but perhaps less efficiently than distinguishing them one at a time. As one would expect, it will be seen that as the dilution effect gets stronger, it becomes more attractive to test smaller pools.

When $\hat{0} < s < \hat{1}$, more complex situations can arise. This is because in distinguishing covers, anti-covers can be distinguished as well when positive samples are pooled together with negative ones. However, when it is attractive to do so, there do not exist general closed form solutions of optimal strategies for such cases, as many contingencies can arise. It will instead be assumed that for

$$r \text{ positives}, 1 \leq r \leq N - |s|, \text{ and } k \text{ negatives}, 1 \leq k \leq |s|,$$

$$(r/(N-|s|))K(r, r-1, r) \geq (r/(N-|s|)K(r, r-1, r+k) - (k/|s|)K(r, r+1, r+k). \tag{A4}$$

As will be argued in Theorem 1, this condition insures that for an optimal strategy, it can be assumed that the optimal value of k is $k^* = 0$, and hence that asymptotically, positives and negatives are not tested together. In Example 6, it will be demonstrated that this condition is not restrictive. When $r > 0$ and $k = 0$, a corresponding pooled test distinguishes covers but not anti-covers. Hence, given (A1)-(A4), when S is a powerset lattice and $\hat{0} < s < \hat{1}$, it will be shown that determining optimal rates of convergence involves identifying an optimal value r^*, the number of positives to be pooled to distinguish covers, and j^*, the optimal size of pools comprised only of negatives, to distinguish the anti-covers. In contrast, when there is no dilution effect, as in (Tatsuoka and Ferguson, 2003), the optimal strategy is to distinguish covers by individually testing positives ($r^* = 1, k^* = 0$), and to distinguish anti-covers by pooling negatives all at once ($j^* = |s|$).

In Theorem 1, it is assumed that the true state is known. Clearly, in practice, it is unknown, and in fact determining its identity is the primary objective of classification. Still, the following results have direct practical relevance. After, an optimal rule for selecting pooled tests will be established. In order to be optimal, a rule must first be convergent in the sense that the true state is eventually identified almost surely. In addition, it must then also eventually adopt an optimal strategy for selection of tests, corresponding to whatever state is true.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Thus, Theorem 1 characterizes the pooling sequences that must eventually be adopted almost surely under the various scenarios that can arise.

For $0 \leq |s| < N$, let

$$R_c^* = \sup_{1 \leq r \leq N-|s|} (r/(N-|s|))K(r, r-1, r),$$

with r^* being a value that satisfies $R_c^* = (r^*/(N-|s|))K(r^*, r^*-1, r^*)$. Also, for $0 < |s| \leq N$, let

$$R_d^* = \sup_{1 \leq j \leq |s|} (j/|s|)K(0, 1, j),$$

with $1 \leq j^* \leq |s|$ attaining the value of R_d^*.

Theorem 1. Suppose that S is a powerset lattice of N subjects, and let $s \in S$ denote the true state. Suppose that tests in $\mathcal{E}$ satisfy (A1), dilution effects (i)-(iii), and (A4).

(i) If $s = \hat{0}$, the optimal rate of convergence is R_c^*, with $|s| = 0$. This rate is attained if each possible subset of r^* subjects are pooled in equal limiting proportion $1/\binom{N}{r^*}$.

(ii) If $s = \hat{1}$, the optimal rate of convergence is R_d^*, with $|s| = N$. This rate is attained if each possible subset of j^* subjects are pooled in equal limiting proportion $1/\binom{N}{j^*}$.

(iii) Otherwise, when $\hat{0} < s < \hat{1}$, the optimal rate is

$$R_c^* \cdot R_d^*/(R_c^* + R_d^*).$$

This optimal rate is attained if each subset of r^* positive subjects are pooled and tested in equal limiting proportion

$$(1/\binom{N-|s|}{r^*}) \cdot R_d^*/(R_c^* + R_d^*),$$

and all tests $e \leq s, |e| = j^*$ (pools of size j^* consisting only of negatives) are administered in equal limiting proportion

$$(1/\binom{|s|}{j^*}) \cdot R_c^*/(R_c^* + R_d^*).$$

The respective given optimal strategies are not necessarily unique. When j^* or r^* are not unique, the optimal rate can be attained by any mixture of optimal allocations associated with each of the values. If $\mathcal{E}$ is restricted by bounds on pool size, Theorem 1 can straightforwardly be extended by optimizing j^* and r^* with respect to the corresponding constrained values.

Theorem 1 gives demarcations for when it is optimal to pool subjects under dilution effects. This result is in some sense analogous to that in Ungar (1960), which states that when there is no testing error with binary outcomes, it is preferable to individually test when the proportion of positive subjects is greater than $(1/2)(3-\sqrt{5})$; otherwise, group testing is preferred. In part (i) with $s = \hat{0}$

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

(all subjects are positive), the demarcation for when it is optimal to individually test subjects is that for all $1 < r \leq N - |s|$,

$$K(1,0,1) \geq rK(r, r-1, r). \tag{A5}$$

This condition follows by comparing rates of convergence for the covers of $s = \hat{0}$ when r subjects are tested at a time. It is a regulation on dilution effect (iii). In part (ii) when $s = \hat{1}$, it is optimal to pool all the subjects if for $1 \leq j \leq N - 1$,

$$K(0,1,N) \geq (j/N)K(0,1,j). \tag{A6}$$

More generally, it is optimal to do some form of pooling as long as $1 < j^* \leq N$. Condition (A6) regulates the decrease in efficiency of detecting a single positive subject versus no positive subjects as pool size increases, and hence relates to dilution effect (i). If this decrease does not occur too quickly, it is optimal to pool all subjects, which are all negative when $s = \hat{1}$. These demarcations will be illustrated below, as well as for when pooling is preferred when $\hat{0} < s < \hat{1}$.

Given the presence of dilution effects, it will be of particular interest to demarcate how strong the effects must be to alter the optimal strategies from when there is no-dilution effect. If (A5) and (A6) hold, the respective optimal strategies for when $s = \hat{0}$ and $s = \hat{1}$ correspond to the no-dilution effect case. The following examples indicate that the same optimal strategy as with no dilution effects is still optimal even in the presence of significant dilution effects.

Example A4. Suppose $s = \hat{1}$, let N be the number of objects to be classified, and suppose sensitivity decreases with pool size, but specificity stays constant. Let $f_{r,|e|}$ be Bernoulli density functions, with $p_{0,|e|} = 0.99$ for all $1 \leq |e| \leq N$, and $q_{1,1} = 0.99$. Following (A6), note that if $N = 15$, in terms of the rate of convergence, group testing all fifteen subjects is preferred if $q_{1,15} > 0.294$ and (A6) holds for the other values of $j < 15$. Further, (A6) would still be satisfied at $j = 30$ if $q_{1,30} > 0.174$. For $j = 100$, (6) would still hold if $q_{1,100} > 0.073$. Hence, under these conditions, (A6) should hold in most practical applications.

Next, supposed that specificity is affected by pool size as well, and in equal magnitude to the sensitivity, with $q_{1,|e|} = p_{0,|e|}$. With $q_{1,1} = p_{0,1} = 0.99$ and assuming (A6) holds for all other values of $j < N$, it is still preferable to group test all subjects when $N = 15$ if $q_{1,15} = p_{0,15} > 0.689$, when $N = 30, q_{1,30} = p_{0,30} > 0.635$, and when $N = 100$ if $q_{1,100} = p_{1,100} > 0.575$.

Suppose $f_{0,|e|}$ is the density for the standard normal distribution for all e, and $f_{1,|e|}$ is the density for a normal distribution with mean $\mu_{1,|e|}$ and variance 1. Assume that (A6) holds for $j < N$. Letting $\mu_{1,1} = 3.0$, it can be seen that when $N = 15$, it is still more attractive to group test all objects if $\mu_{1,15} > 0.775$, when $N = 30, \mu_{1,30} > 0.548$, and when $N = 100$, if $\mu_{1,100} > 0.300$.

Example A5. Consider when $s = \hat{0}$. The demarcation in (A5) can similarly be illustrated numerically for when it is more attractive to individually test subjects under this case. As described in Example A2, the trade-off is between discriminatory efficiency as measured by Kullback-Leibler information versus the number of covers of $s = \hat{0}$ being distinguished per test. Pooling allows for more covers to be distinguished, but is less efficient in discrimination than

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

individual testing. For instance, sensitivity $q_{r,r} = 0.99$ for $r \geq 1$, and specificity $p_{0,|e|} = 0.99$ for all $0 < |e| \leq N$. For this case, the demarcation is the same as in Example A4. Thus, assuming (A5) holds for all $j < N$, (A5) is also satisfied for $N = 15, 30$ and 100 when respectively $q_{14,15} < 0.294, q_{29,30} < 0.174$, and $q_{99,100} < 0.174$. This example illustrates that the dilution effect (iii) can be quite strong, and yet the no-dilution effect strategy of eventually individually testing positive subjects is not affected.

Given (A1), dilution effects (i)-(iii), and $\hat{0} < s < \hat{1}$, a simple condition that is sufficient for insuring that $r^* = 1$ is

$$K(1, 0, 1) \geq rK(r, r-1, r) \tag{A7}$$

for $2 \leq r \leq N - |s|$. This condition is essentially the same as (A5), except (A7) applies to a smaller range of r. Hence, if (A5) is already established, (A7) follows as well. The numerical demarcation in Example A5 is thus applicable to (A7), and indicates that (A7) can hold generally. If the following condition holds, along with (A7), testing positives individually is optimal: for all $1 \leq k \leq |s|$,

$$(1/(N-|s|))K(1, 0, 1) \geq (1/(N-|s|))K(1, 0, k+1) - (k/|s|)K(1, 2, k+1). \tag{A8}$$

Note that (A8) is a special case of (A4) with $r = 1$. Finally, given (A7) and (A8), it follows that distinguishing the anti-covers is conducted by pooled tests consisting of negatives of size j^*. For $1 \leq j \leq |s| - 1$, $j^* = |s|$ when

$$K(0, 1, |s|) \geq (j/|s|)K(0, 1, j). \tag{A9}$$

Note that (A6) implies (A9). Indeed, substituting N with $|s|$, note that Example A4 illustrates that $j^* = |s|$ except under strong dilution effects. In practice, checking conditions (A5) and (A6), and (A8) with $|s| = N$, is sufficient for determining for all states whether or not corresponding strategies are altered.

In sum, Examples A4 and A5 illustrate that dilution effects, unless severe, do not generally alter the optimal strategy of eventually pooling all negative subjects, while individually testing each of the positive subjects. One practical ramification that these results suggest is that, given low prevalence of positive samples, it is attractive to initially pool as many samples as possible, in spite of dilution effects, since the top element is most likely the true state a priori. In the next section, it will be seen that when (A5), (A6) and (A8) hold, a simple and intuitive rule for dynamically selecting pools will be optimal in the sense that optimal strategies will be selected eventually, corresponding to the unknown true state, with probability one.

A4. Optimal Pooling Selection Under Dilution Effects Now let us consider rules for pooling selection that, given the observed outcomes to previously administered pools and prior information, select the composition of the next stage pool to test. It is desirable to have a rule that adopts optimal strategies and attains optimal rates almost surely, no matter which state is true. It will be seen in this section that an intuitive and simple rule, a halving algorithm,

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

attains optimal rates of convergence when the optimal strategy coincides with the no-dilution effect case, and hence even under strong dilution effects. For instance, if (A5), (A6) and (A8) hold and $s = \hat{1}$, then it is desirable for a pooling selection rule to eventually pool all objects. If $\hat{0} < s < \hat{1}$, then we would want the rule to lead to sequences of pools that eventually consist of all the negative subjects, or individually tests the positive subjects.

Theorem 2. Under (A1), dilution effects (i)-(iii), and (A5), (A6), and (A8), for S being a powerset lattice and any $s \in S$ being the true state, the Bayesian halving algorithm described by (1) will attain the optimal rate almost surely.

A5. Bayesian Estimation of Prevalence Through Group Testing

Based on the above-described group testing frameworks, *prevalence* can be estimated in a Bayesian manner. Suppose that it is of interest to estimate the proportion of positives in a target population, denoted by θ, based on group testing data. Given there is uncertainty as to the positivity status of subjects, a natural estimate could rely the information provided by π_n. For instance, given a conjugate Beta prior distribution for $\theta, f(a, b)$, the posterior distribution for θ given the observed group testing outcomes is a mixture of posterior Beta densities with respect to π_n, where the Beta densities are updated depending on which state is true:

$$\sum_{j \in S} f(a + |j|, b + N - |j|) \cdot \pi_n(j).$$

For the powerset lattice, $|j|$ is the number of negatives out of the N objects given state j is true. Note that this mixture will converge to the posterior distribution for θ that would be obtained if the correct diagnoses for all the N subjects were known exactly, given $\pi_n(s)$ converges to 1 almost surely for true state s.

Example A6. As a simple illustration, suppose as in Example 1 that N=2. We are interested in estimating θ, and suppose a beta prior distribution, $\beta(2, 2)$. Prior mean and standard deviation are thus 0.50 and 0.22. For $n = 2$ stages, recall $\pi_2 = \{\pi_2(AB) = 0.0009, \pi_2(A) = 0.9985, \pi_2(B) = 0.0001, \pi_2(\hat{0}) = 0.0005\}$. The posterior density for θ is thus the mixture

$$(0.0009)f(4, 2) + (0.9985)f(3, 3) + (0.0043)f(3, 3) + (0.0005)f(2, 4).$$

The mixture posterior mean of θ is 0.5022, and the mixture posterior standard error is 0.1888.

A7. Proofs

Proof of Theorem 1. Again, let p_e denote the limiting proportion that $e \in \mathcal{E}$ is administered. From (Tatsuoka and Ferguson, 2003), the optimal rate of convergence of $\pi_n(s)$ to 1 is thus related to the following linear program: find probability vector $\{p_e\}_{e \in \mathcal{E}}$ and v to

$$\text{maximize } v$$

subject to

$$p_e \geq 0, e \in \mathcal{E}, \text{ and } \textstyle\sum_{e \in \mathcal{E}} p_e = 1,$$

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

and

$$v \leq \sum_{e \in \mathcal{E}} p_e K(|e| - |e \cap s|, |e| - |e \cap k|, |e|) \qquad \text{for } k \in S \backslash \{s\}. \tag{A10}$$

A probability vector, $\{p_e\}$, is sought to maximize the minimum of the right side of (A10) over $k \in S \backslash \{s\}$. Note that the right side of (11) is the rate of convergence for state k given the allocation $\{p_e\}$.

Consider first when $s = \hat{1}$. The case when $s = \hat{0}$ follows similarly. For $z \in D_s \backslash D_s$, note that since S is a powerset lattice, there exists $x, y \in D_s$ such that $z \leq x \wedge y$. Hence, z is distinguished from s whenever x and y are distinguished. Further, by (2), the rate of convergence for terms in D_s is thus slowest for terms in D_s. We thus need only consider the terms in D_s in establishing the optimal rates. Moreover, it can be established that if all experiments in $\{e : |e| = j, e \leq s\}$, $1 \leq j \leq N$, are each administered in limiting proportion $1/\binom{N}{j}$, the rate of convergence is $(j/N)K(0, 1, j)$. Denote such an allocation as $a(j, N)$.

It will now be shown that if

$$|e'|K(0, 1, |e'|) > |e|K(0, 1, |e|) \text{ and } |e'| < |e|,$$

then it is optimal for $p_e = 0$. Suppose e is administered with limiting proportion $p_e > 0$, and let $\delta = \{p_k\}_{k \in \mathcal{E}}$ be an optimal allocation. Note that the rate of convergence contribution to each state distinguished from s by administration of e is $p_e K(0, 1, |e|)$. However, by instead administering all $k \in \mathcal{E}$ contained in $\{e\}^c \cap D_s, |k| = |e'|$, in equal limiting proportion $p_e(1/\binom{|e|}{|e'|})$, the corresponding rate contribution to each state in $\{e\}^c \cap D_s$ is $p_e(|e'|/|e|)K(0, 1, |e'|)$, which is greater than $p_e K(0, 1, |e|)$. Hence, allocation δ can be improved, and would not be optimal.

Suppose now there exists e such that $|e| < j^*, |e| = \inf_{k: p_k \in \delta, p_k > 0} |k|$. From above, there exists $e' \in \delta$ such that $p_{e'} > 0, |e'| > |e|, e \not< e'$, and $|e'|K(0, 1, |e'|) > |e|K(0, 1, |e|)$, or else the allocation is dominated. Consider

$$D' = \{\{e\}^c \ \cup \{e'\}^c\} \cap D_s = D_s \backslash \{e \cap e'\}.$$

It will now be established that the rate can be improved by administering all tests in

$$\mathcal{E}(e, e') = \{k : |k| = |e'|, \{k\}^c \cap D_s \subseteq D', \{e\}^c \cap D_s \subseteq \{k\}^c \cap D_s\},$$

in a certain equal limiting proportion given below instead of administering e with proportion p_e. Note that there are $n' = \binom{|D'| - |e|}{|e'| - |e|}$ such experiments.

Consider two cases: First suppose

$$\textit{(i)}\ p_{e'} - p_e((K(0, 1, |e|)/K(0, 1, |e'|) - 1) > 0.$$

For terms in $\{e\}^c \cap D_s$, note that if tests in $\mathcal{E}(e, e')$ are administered in equal proportion

$$(p_e + p_e((K(0, 1, |e|)/K(0, 1, |e'|)) - 1))/n' = p_e((K(0, 1, |e|)/K(0, 1, |e'|))/n',$$

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

then the corresponding rate contribution for each distinguished state is still

$$p_e K(0,1,|e|),$$

the same as if e were administered in proportion p_e. Moreover, administering e' with proportion

$$p_{e'} - p_e((K(0,1,|e|)/K(0,1,|e'|)) - 1),$$

note that terms in $\{e'\}^c \cap e \cap D_s$ are distinguished in proportion

$$(p_e + p_e((K(0,1,|e|)/K(0,1,|e'|)) - 1))(|e'| - |e|)/(|D'| - |e|) +$$

$$p_{e'} - p_e((K(0,1,|e|)/K(0,1,|e'|)) - 1).$$

The corresponding rate contribution is greater than $p_{e'}K(0,1,|e'|)$ when $|D'| > |e'| > |e| > 0, |e'| + |e| > |D'|$, and $|e'|K(0,1,|e'|) > |e|K(0,1,|e|)$, as is assumed. Finally, note the rate contribution to states in $\{e'\}^c \cap \{e\}^c \cap D_s$ by administration of tests in $\mathcal{E}(e,e')$ and e' is even greater.

Suppose now

$$\textit{(ii)}\ p_{e'} - p_e((K(0,1,|e|)/K(0,1,|e'|)) - 1) \leq 0.$$

Solving for p_e^* such that

$$p_{e'} - p_e^*((K(0,1,|e|)/K(0,1,|e'|)) - 1) = 0,$$

it can be found that

$$p_e^* = p_{e'}K(0,1,|e'|)/(K(0,1,|e|) - K(0,1,|e'|)).$$

The rate contribution to terms in $\{e\}^c \cap D_s$ from administration of experiments in $\mathcal{E}(e,e')$ in equal proportion $(p_{e'} + p_e^*)/n'$, and administration of e in proportion $p_e - p_e^*$ is still

$$(p_{e'} + p_e^*)K(0,1,|e'|) + (p_e - p_e^*)K(0,1,|e|) = p_e K(0,1,|e|).$$

Further, note the rate contribution to each state in $\{e'\}^c \cap e \cap D_s$ by experiments in $\mathcal{E}(e,e')$is

$$(p_{e'} + p_e^*)((|e'| - |e|)/(|D'| - |e|))K(0,1,|e'|),$$

which is greater than $p_{e'}K(0,1,|e'|)$ under the given conditions. The rate contribution to states in $\{e'\}^c \cap \{e\}^c \cap D_s$ is thus also greater. Hence any optimal allocation will administer only tests such that $|e| = j^*$. Moreover, the optimal rate is attained when all terms in D_s converge to zero at the same rate. The allocation $a(j^*, j^*)$ is thus optimal.

Suppose now that $\hat{0} < s < \hat{1}$. The condition in (4) results from comparing the respective maximin rate among states in $C_s \cup D_s$ that would result when covers are distinguished by pools comprised of r positives with no positives, versus pools with r positives and k negativess, and anti-covers are distinguished

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

by $e \in s, |e| = j^*$. By (3), even when anti-covers are distinguished by pools containing both positives and negatives, it is necessary to administer tests in s in positive limiting proportion. Using the arguments above, in an optimal strategy, among experiments in s, only $e \in s$ with $|e| = j^*$ will be administered. Given (4), then, among states in $C_s \cup D_s$, the maximin rate is obtained by administering r positives with $k = 0$ negatives rather than if $k > 0$. Hence, for states in $C_s \cup D_s$, it is optimal to consider only strategies with $k = 0$.

If j is incomparable to s, then there must be an object k_1 that is negative if s is true but that is positive if j is true. Let $d_1 \in D_s$ be the anti-cover associated with k_1 in the sense that the objects associated as negative for d_1 correspond to all the negatives for s except k_1. When d_1 is so distinguished, so is j. Further, if $k = 0$ when distinguishing covers, j converges at least as fast as d_1. Hence, only states in $C_s \cup D_s$ would determine the optimal rate, and so it is optimal to only consider strategies with $k = 0$. Following as above, only pools with r^* positives and $k = 0$ negatives will be administered to distinguish covers, along with pools comprised only of j^* negatives to distinguish anti-covers. Respective limiting proportions that insure that posterior probability terms for states in $C_s \cup D_s$ converge at the same rate, such as stated in (iii), are optimal.

Proof of Theorem 2. Define $p_j(n) = n_j/n$, where n_j represents the number of times $j \in \mathcal{E}$ is administered through stage n, and $p_{js}(n) = n_{js}/n$, where n_{js} represents the number of times $j \in S \backslash \{s\}$ is distinguished from s through stage n. Also, let G_{cs} be the minimal elements in K_{cs}. For a powerset lattice, G_{cs} corresponds to the atom associated with the subject that is negative for c but positive for s.

First note that the halving algorithm is convergent in the sense that $\pi_n(s) \to 1$ almost surely, and that all states are distinguished from s infinitely often. Note

$$\{c\} = c \cap \{ \bigcup_{c' \in C_s \backslash \{c\}} G_{c's} \}^c \text{ for } c \in C_s, \text{ and } D_s = \{s\}^c \cap \{ \bigcup_{c \in C_s} G_{cs} \}^c.$$

Also, let

$$M_n^c = |S| \cdot \pi_n(c) \text{ for } c \in C_s, \text{ and } M_n^s = |S| \cdot \sup_{j \in D_s} \pi_n(j).$$

Define two test selection rules, procedures A_I and A_{II}. Procedure A_I is defined as follows:

$$\begin{cases} e_n = s \text{ if } M_n^s \geq \pi_n(c) \text{ for all } c \in C_s; \\ \text{otherwise, select } j \in \bigcup_{c \in C_s} G_{cs} \text{ which maximizes } h(j, \pi_n). \end{cases}$$

Consider now procedure A_{II}:

$$\begin{cases} e_n = s \text{ if } \pi_n(\hat{0}) > M_n^c \text{ for all } c \in C_s; \\ \text{otherwise, select } j \in \bigcup_{c \in C_s} G_{cs} \text{ which maximizes } h(j, \pi_n). \end{cases}$$

Both of these procedures attain the optimal rate of convergence for s being the true state in a general lattice model. This can be seen by applying the arguments

medRxiv preprint doi: https://doi.org/10.1101/2021.01.15.21249894; this version posted January 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

of Theorem 6 in (Tatsuoka and Ferguson, 2003) (in Proposition 1 of (Tatsuoka and Ferguson, 2003) it should instead be stated that $a(n)/n \to B/(A+B)$). Note the slowest converging states in $S \backslash \{s\}$ are those in C_s and those in D_s.

When n is large and hence $\pi_n(s) > 0.5$, each cover c is distinguished by the corresponding element $c^* \in G^*_{cs}$. This follows since $m_n(c^*) > m_n(j)$ for any $j \in \{j' : c^* < j' \leq c\}$. Note also that for n large, $1 - m_n(s) > 1 - m_n(j)$ for $j \in D_s$. This implies that $h(s, \pi_n) > h(j, \pi_n)$ eventually for all $j \in D_s$.

Again following as in Theorem 6 of (Tatsuoka and Ferguson, 2003), it can be argued that eventually $j \in \{C_s\}^c$ is not administered. Further, the halving algorithm can be compared with procedures A_I and A_{II}, establishing that the halving algorithm administers the respective optimal limiting proportions.