# Welfare-Maximizing Pooled Testing

SIMON FINSTER, CREST-ENSAE
MICHELLE GONZÁLEZ AMADOR, UNU-MERIT & Maastricht University
EDWIN LOCK, University of Oxford
FRANCISCO MARMOLEJO-COSSÍO, Harvard University
EVI MICHA, University of Toronto
ARIEL D. PROCACCIA, Harvard University

In an epidemic, how should an organization with limited testing resources safely return to in-person activities after a period of lockdown? We study this question in a setting where the population at hand is heterogeneous in both utility for in-person activities and probability of infection. In such a period of re-integration, tests can be used as a certificate of non-infection, whereby those in negative tests are permitted to return to in-person activities for a designated amount of time. Under the assumption that samples can be pooled, the question of how to allocate a limited testing budget in the population to maximize the aggregate utility (i.e. welfare) of negatively-tested individuals who return to in-person activities is non-trivial, with a large space of potential testing allocations.

We show that non-overlapping testing allocations, which are both conceptually and (crucially) logistically more simple to implement, are approximately optimal, and we design an efficient greedy algorithm for finding non-overlapping testing allocations with approximately optimal welfare. In computational experiments, we highlight the efficacy and viability of our greedy algorithm in practice. To the best of our knowledge, we are also first to implement and provide causal evidence on the benefits of utility-weighted pooled testing in a real-world setting. Our pilot study at a higher education research institute in Mexico demonstrates — surprisingly — no worse performance and mental health outcomes of participants in our testing regime than the first-best counterfactual of full reopening without testing.

## 1 INTRODUCTION

Over the course of the ongoing COVID-19 pandemic, it has become abundantly clear that testing is an indispensable tool for combating the virus. At the same time, the pandemic has underscored the fact that this very resource can be prohibitively constrained in multiple ways (e.g. lack of reagents, trained personnel or lab equipment), making it essential to study how to systematically allocate tests for the benefit of the population at hand [Abera et al., 2020, Binnicker, 2020, De Georgeo et al., 2021, Dhabaan et al., 2020, Dryden-Peterson et al., 2021, Kavanagh et al., 2020]. In this regard, pooled testing has emerged as a promising primitive for expanding the reach of limited testing resources. In a pooled test, the samples of multiple individuals are pooled together, and a single test is applied to the mixed sample. If the test is positive, at least one individual in the pool is infected, and if negative, all individuals in the pool are healthy. Indeed, the latter case is precisely where savings in applying tests occur, as a single test can verify that multiple individuals are healthy.

Ultimately, testing serves two major purposes: it prevents infections from occurring by identifying infected individuals to be isolated, and it provides a means for individuals in a population to return to in-person activities after being cleared with a negative test result. Our point of departure from prior work is the observation that different individuals have different utilities for resuming in-person activities; in a scientific setting, for example, the benefit an experimentalist derives from being able to work in their lab is typically higher than that of a theoretician who is permitted to go to the office. The goal, therefore, should be to

> *... maximize the expected welfare (overall utility) of individuals who are able to resume in-person activities.*

In more detail, we explore the scenario where a population is currently in lockdown and wishes to begin in-person activities, facilitated by a fixed budget of tests. We assume that the population is heterogeneous: each individual has their own probability of being infected and (as mentioned earlier) their own utility for being able to resume in-person activities. Tests are allocated to subsets of the population as pooled tests, and an individual is allowed to return to in-person activities (and hence earn their corresponding utility) if any of the tests to which they are assigned results negative (proving that they are healthy). A testing regime thus obtains a certain expected welfare (overall utility) with respect to the randomized realization of infections in the population.

Our problem setting is fundamentally motivated and informed by a collaboration with the Potosinian Institute of Scientific and Technological Research (IPICYT), a higher education research institution in Mexico, with the goal of providing safe, resource-optimal alternatives to fully easing a virtual work environment. At IPICYT, a heterogeneous population of 130 individuals participated in our pilot study, including students, academics, and administrators, who were largely restricted access to the campus buildings until September 2022. We have worked closely with epidemiologists at IPICYT and in the local state of San Luis Potosí to determine accurate estimates of the probabilities of infection for each member of the population. Furthermore, the National Laboratory of Agricultural, Medical and Environmental Biotechnology (LANBAMA), housed within IPICYT, maintains a qPCR testing facility. Three testing constraints exhibited by the LANBAMA are of crucial importance to our work: i) there are insufficient resources available to regularly test every member of the population individually, ii) the laboratory strongly prefers pooled testing protocols that do not include individuals in more than one test at a time, and iii) the laboratory conducts pooled testing on saliva samples (as opposed to nasopharyngeal sample) with a maximal pool size of 5.

To further elucidate the latter constraint, it is important to note that although pooled testing increases resource efficiency with regards to testing reagents, this can come at a significant logistical cost for laboratory personnel when pools are overlapping, especially if the series of tests to be performed is complicated and requires delicate tab-keeping of results. In this regard, pooled testing regimes in which no individual forms part of more than one pooled test are not only conceptually simpler to study but, more importantly, logistically simpler to implement. We call such testing regimes "non-overlapping" due to the fact that pooled tests in the allocation do not overlap. By contrast, the more general overlapping testing regimes allows for the budget of tests to be allocated arbitrarily, subjecting individuals to arbitrarily many tests.

*Our results.* Despite these practical considerations that favor non-overlapping testing regimes, our first research question is whether they may be vastly outperformed by overlapping regimes. After all, if that is the case then supporting overlaps may be worth the logistical overhead. However, we show that the worst-case ratio between the welfare of the best overlapping testing regime and the best non-overlapping regime is at most 4, and in special cases even smaller. While a factor of 4 is admittedly significant, the worst example we know of gives a ratio of 7/6. Qualitatively, we interpret these results as a justification to focus on non-overlapping testing regimes. This is confirmed by our empirical results, which indicate only small gains from overlaps in practice.

Turning to the challenge of computing testing regimes, even without overlaps, the problem of determining an optimal regime is NP-hard. But we design a greedy polynomial-time algorithm that (roughly speaking) provides a 5-approximation to the optimal non-overlapping testing regime in the worst case. We then compare the performance of our greedy algorithm with optimal non-overlapping testing empirically. In our computational experiments, we choose a population size, pool size constraints and testing budgets that mirror realities at IPICYT. In order to compute (approximately) optimal non-overlapping testing regimes, we model the problem as a mixed-integer linear program (MILP). Our results indicate that the greedy algorithm computes near-optimal testing regimes, and vastly outperforms the MILP approach with respect to running time.

Finally, we provide empirical evidence in support of our approach to welfare-maximizing pooled testing in a resource-constrained environment by evaluating a randomized controlled trial at IPICYT, which we conducted in September 2022. For the pilot, we developed a web application (released as open source) that formed the center point for participants, administrators and the LANBAMA testing laboratory, and implemented our greedy algorithm to compute near-optimal non-overlapping group testing regimes.[1] Our trial results show that, compared to a best-case scenario of free mobility and full access to institutional resources, our testing approach is just as efficient in terms of performance, learning, and mental health outcomes. At the same time, our protocol, which ensures that only negatively qPCR-tested individuals have in-person access, safeguards the population's health within the institution, unlike a full reopening without testing, and at a fraction of the cost of an individual qPCR testing regime.[2]

*Related work.* Pooled testing dates back to the seminal work of Dorfman [1943] and has since become a mature field in its own right with a rich literature of protocols typically aimed at solving the following problem: *precisely ascertain the infection status of all individuals in a population with the minimum number of tests.* As mentioned above, our work departs significantly from this objective, as we instead assume that testing resources are initially limited, and with this provide welfare-optimizing testing allocations. For general references and recent results in this theoretical thread of pooled testing, we refer the reader to an in-depth literature review included in Appendix A.

Resource constraints have motivated recent work aimed at optimally utilizing limited testing capabilities to help local communities. Ely et al. [2021] study a model where a policymaker can employ tests of different types, each with differential costs and sensitivities. The policymaker has an

[1]An anonymised demo of the web app is available at https://ec23demo.pythonanywhere.com.
[2]At the time of reopening, San Luis Potosí had 221,870 cumulative COVID-19 cases, of which 615 were active. KN95 Masks were mandatory for everyone returning to IPICYT.

overall budget, and testing allocations are measured with respect to the rate at which they correctly classify individuals as infected or healthy. Brault et al. [2021] focus on limited pooled tests for early screening at a non-diagnostic level with high penalties associated with false negatives. Gollier and Gossner [2020] study pooled tests as a means to estimate infection prevalence and to find healthy individuals in a population. The main differences between their work and ours is that we consider a heterogeneous population as well as upper bounds on pool sizes imposed by lab constraints.

Most similar to our paper is the work of Lock et al. [2021] and Jonnerby et al. [2020a,b], in which the authors consider a limited testing budget to be used over a heterogeneous population as a means of surveilling and containing viral spread (unlike our work which focuses on testing to find those who are healthy). The testing allocation problem is treated as a multi-objective optimization problem aimed at balancing viral spread with the overall cost of self-isolation. Although not cast as a pooled testing paper, the results of Goldberg and Rudolf [2020] can be interpreted as computing the optimal allocation of a single (arbitrarily large) pooled test to a heterogeneous population as in our model setting. The authors show that computing an optimal single test allocation is NP-hard, and they provide a fully polynomial-time approximation scheme (FPTAS) for finding an approximately optimal single test allocation; we use their FPTAS as a subroutine in our greedy algorithm. Finally, Larremore et al. [2021] and Augenblick et al. [2020] study testing frequency as a crucial factor to limiting viral spread in a pandemic environment, and how pooled testing can increase the reach of a rapid frequency testing regime when test are limited.

## 2 MODEL

Let $[n] = \{1, \ldots, n\}$ denote the population and $B \in \mathbb{N}$ be the testing budget, i.e. the number of available tests. Each individual in the population has an independent probability of infection given by $p_i \in [0, 1]$ and a utility $u_i \geq 0$ capturing their gain of returning to in-person activities.[3] We also let $q_i = 1 - p_i$ denote the probability that an individual is healthy. A population instance $J$ is given by $(q_1, \ldots q_n, u_1, \ldots, u_n)$.

A single test consists of samples of a subset of the individuals, which we identify with a set $t \subseteq [n]$ of the individuals whose samples are included in the test. Test sizes are bounded by a pool size constraint $G \leq n$, so $|t| \leq G$ for all tests $t$.[4] For convenience, we introduce the notation $q_S = \prod_{i \in S} q_i$, for any $S \subseteq [n]$, to express the probability that all individuals in $S$ are healthy; hence, $q_t$ is the probability that test $t$ is negative. A *testing regime* $T = (t_1, \ldots, t_B)$ is a collection of $B$ tests satisfying $|t_j| \leq G$ for each $j \in [B]$.

Individuals are allowed to return to in-person activities only if they are included in a negative test. For a given testing regime $T$, let $P_i^T$ denote the probability that $i \in [n]$ is included in some negative test $t_j \in T$. A testing regime only earns utility from individuals who return to in-person activities as a result of being in a negative test. We let $u(T)$ denote the aggregate expected utility, or *welfare*, earned under testing regime $T$.[5] Linearity of expectation allows us to express the welfare of $T$ as $u(T) = \sum_{i \in [n]} u_i \cdot P_i^T$. In addition, we let $u(t) := u((t)) = q_t \left(\sum_{i \in t} u_i\right)$ for a single pooled test $t$. A testing regime $T$ is *optimal* (for a given population) if it maximizes welfare. Without loss of generality, we assume that $B < n$. If this is not the case, testing every person in the population individually is optimal.

[3]Utility might reflect people's socioeconomic status, the type of occupation, or mental health considerations. See Section 5.1 for details on the utilities in our pilot.

[4]Pool sizes in pooled tests are limited due to biological constraints. Our partners in Mexico have replicated techniques from Sanghani et al. [2021] to achieve a maximal pool size of 5 with saliva samples.

[5]In the following, we will drop the term 'expected' for brevity, and assume that all welfares and utilities are determined in expectation.

*Independence of Infections.* In general it may be the case that infections in a population are correlated. However, we emphasize that our testing model is intended for a regime wherein all individuals in the given population are assumed to be in full lockdown, hence social interactions at the workplace do not contribute to potential correlation of infection for two key reasons: Either individuals who would potentially interact are forcibly at home, and hence no longer interact, or if the individuals are interacting at the workplace, it is because they are both in a negative test and hence cannot infect each other. It may be the case that colleagues interact outside of office hours, but this is not a phenomenon observed often during lockdown with our partners in Mexico.

*Non-overlapping testing regimes.* As discussed in Section 1, we are particularly interested in *non-overlapping* testing regimes that include each individual in at most one test. Formally, testing regime $T$ is *non-overlapping* if $t \cap t' = \emptyset$ for all distinct tests $t, t' \in T$. In general, $P_{i,T}$ can be a complicated expression due to correlation between overlapping tests. In non-overlapping testing regimes $T$, by contrast, test results are independent of one another and the welfare of $T$ is $u(T) = \sum_{t \in T} u(t)$.

*Gain of overlaps.* We are interested in quantifying the relative benefit provided by overlapping testing regimes over non-overlapping regimes, because the latter are not only conceptually simpler but also more feasible to implement in practice, as discussed in Section 1. Given a population instance $J$ and budget $B$, we define the *gain of overlaps* $\mathrm{gain}(B, J)$ as the ratio of the welfare of the optimal testing regime over the welfare of the optimal non-overlapping testing regime. Formally, we let $\mathcal{T}^B$ and $\tilde{\mathcal{T}}^B$ respectively denote the space of all testing regimes and all non-overlapping testing regimes with testing budget $B$, and write

$$\mathrm{gain}(B, J) = \frac{\max_{T^* \in \mathcal{T}^B} u(T^*)}{\max_{T \in \tilde{\mathcal{T}}^B} u(T)}.$$

The *gain of overlaps given a budget* $B$, denoted $\mathrm{gain}(B)$, is the worst-case gain over all possible instances $J$, that is, $\mathrm{gain}(B) = \sup_J \mathrm{gain}(B, J)$.

## 3 THEORETICAL RESULTS

In this section, our goal is to provide upper bounds for the gain of overlaps. In order to develop intuition for cases in which the gain is greater than 1, we first study, as a warm-up, the case where there are only two available tests ($B = 2$), and show that the gain of overlaps is quite small. More generally, we show that for any value of $B$, the gain is at most 4. Motivated by this result and the aforementioned practical constraints, we then focus on non-overlapping testing regimes and present a greedy algorithm that achieves a constant-factor approximation with respect to the optimal non-overlapping testing regime.

### 3.1 Warm-Up: Gain of Overlaps when $B = 2$

We begin by studying the case in which $B = 2$. This case is particularly interesting for two reasons: first, we find the exact value of the gain by providing a lower bound and then showing that it is tight; second, this lower bound is the worst (largest) we know of, for any value of $B$.

Proposition 1. *For* $B = 2$, $\mathrm{gain}(B) \geq 7/6$.

Proof. Consider a population of three individuals $\{1, 2, 3\}$ given by $q_1 = q_2 = 1/2$, $q_3 = 1$ and $u_1 = u_2 = u_3 = 1$. We see that individuals 1 and 2 are identical, and therefore there are only four non-overlapping testing regimes up to symmetries.

- $T^1 = (\{1\}, \{2\})$ yields welfare $u(T^1) = 1$.
- $T^2 = (\{1\}, \{3\})$ yields welfare $u(T^2) = 3/2$.
- $T^3 = (\{1, 2\}, \{3\})$ yields welfare $u(T^3) = 3/2$.

- $T^4 = (\{1,3\},\{2\})$ yields welfare $u(T^4) = 3/2$.

Now consider the overlapping testing regime $T^* = (\{1,3\},\{2,3\})$ in which individual 3 is tested twice. Then we have welfare $u(T^*) = 7/4$. □

It is of particular interest in the proof of Proposition 1 that the optimal welfare is achieved by testing the individual who is definitely healthy twice. This is intuitive, as we can test this individual many times without reducing the chances that any test is negative. In fact, in the proof of the upper bound (given in Appendix B), we use the property that the gain of overlaps is maximized when the probability that the individuals who are tested twice are healthy is maximized.

Proposition 2. *For $B = 2$,* $\text{gain}(B) \leq 7/6$.

### 3.2 Upper Bound on Gain of Overlaps for Any $B \geq 2$

In this section, we show that the gain of overlaps is a small constant; not only for the case that there are two tests available, but for any $B$. Before doing so, we start with some necessary notation. Given a testing regime $T = (t_1, \ldots, t_B)$ and individual $i \in [n]$, we let $T(i) = \{t_j \in T \mid i \in t_j\}$. Furthermore, we denote with $T(i;j)$ the set of tests with index less than $j$ in which $i$ has been tested, i.e. $T(i;j) = \{t_{j'} \in T(i) : j' < j\}$. We say that test $t_j$ is *pivotal* for individual $i$ if: $i$ is included in $t_j$, the result of $t_j$ is negative, and all tests in $T(i;j)$ are positive. Equivalently, test $t_j$ is pivotal for individual $i$ if it is the negative test of smallest index in $T(i)$. We let $P^T_{i,j}$ denote the probability that $t_j$ is pivotal for individual $i$ under random infection realizations. In other words:

$$P^T_{i,j} = \begin{cases} \Pr[\ \forall t_{j'} \in T(i;j), t_{j'} \text{ is positive and } t_j \text{ is negative}] & \text{if } t_j \in T(i), \\ 0 & \text{otherwise.} \end{cases}$$

Individual $i$ is in a negative test if and only if a single test $t_j \in T$ is pivotal for $i$, hence $P^T_i = \sum_{j \in [B]} P^T_{i,j}$. As previously advertized, our main result of this section is that overlapping testing regimes have bounded gain over non-overlapping regimes.

Theorem 1. *For any $B \geq 1$,* $\text{gain}(B) \leq 4$.

To prove the theorem, we will show that given an optimal overlapping testing regime $T^*$, we can find a non-overlapping testing regime $T$ such that $u(T^*)/u(T) \leq 4$ in polynomial time. This does not lead to a polynomial-time algorithm, as it requires access to $T^*$ to begin with.

The proof requires a two lemmas. The first lemma, whose proof appears in Appendix C, is more intuitive: There exists an optimal testing regime $T^* = (t^*_1, \ldots, t^*_B)$ such that if $t^*_j \in T^*(i)$, it must be the case that the probability that $t^*_j$ is pivotal for $i$ is positive.

Lemma 1. *There exists an optimal $T^* = (t^*_1, \ldots, t^*_B)$ such that if $t^*_j \in T^*(i)$, then $P^{T^*}_{i,j} > 0$.*

The second lemma is a non-trivial generalization of Lemma 6 of Goldberg and Rudolf [2020]. At a high level, we show that if a (non-overlapping) testing regime is optimal, no test within this regime can be split into two groups which simultaneously have a "low" probability of being healthy. When the testing regime is non-overlapping, the generalization is straightforward, but for the general case where the testing regime may be overlapping, novel techniques and arguments are used.

Lemma 2. *Suppose that $T^* = (t^*_1, \ldots, t^*_B)$ is an optimal (non-overlapping) testing regime and that $\alpha \in (0,1)$. For any $t^*_j$ and any $S \subset t^*_j$, if $q_S < \alpha$, then $q_{t^*_j \setminus S} \geq 1 - \alpha$.*

Proof. First note that for any $i \in t^*_j$,

$$P^{T^*}_{i,j} = \Pr[\ \forall t^*_{j'} \in T^*(i;j), t^*_{j'} \text{ is positive and } t^*_j \text{ is negative}]$$

$$\begin{aligned}
&= \Pr[\ \forall t^*_{j'} \in T^*(i;j), t^*_{j'} \text{ is positive } \mid t^*_j \text{ is negative}] \cdot \Pr[t^*_j \text{ is negative}] \\
&= \Pr[\ \forall t^*_{j'} \in T^*(i;j), t^*_{j'} \setminus t^*_j \text{ is positive } \mid t^*_j \text{ is negative}] \cdot \Pr[t^*_j \text{ is negative}] \\
&= \Pr[\ \forall t^*_{j'} \in T^*(i;j), t^*_{j'} \setminus t^*_j \text{ is positive } ] \cdot \Pr[t^*_j \text{ is negative}] \\
&= \Pr[\ \forall t^*_{j'} \in T^*(i;j), t^*_{j'} \setminus t^*_j \text{ is positive } ] \cdot q_{t^*_j}
\end{aligned}$$

where the third equality follows since $t^*_{j'}$ can be positive if and only if some individual in $t^*_{j'} \setminus t^*_j$ is infected as given that $t^*_j$ is negative we conclude that any individual in $t^*_j$ is healthy, and the fourth inequality follows since each individual has an independent probability to be infected.

Assume for the sake of contradiction that for some $j \in [B]$, there exists $S \subset T^*_{j^*}$ such that $q_S < \alpha$ and $q_{T^*_{j^*} \setminus S} < 1 - \alpha$. Without loss of generality, we assume that $j^* = B$. Then, since from Lemma 1 we know that $P^{t^*_j}_{i,j} > 0$ for any individual $i$ such that $i \in T^*_j$, we have that

$$\begin{aligned}
u(T^*) &= \sum_{i \in [n]} P^T_i \cdot u_i \\
&= \sum_{i \in [n]} \sum_{j \in [B]} P^{T^*}_{i,j} \cdot u_i \\
&= \sum_{j \in [B-1]} \sum_{i \in [n]} P^{T^*}_{i,j} \cdot u_i + \sum_{i \in [n]} I^{T^*}_{i,B} \cdot q_{t^*_B} \cdot \Pr[\forall t^*_{j'} \in T^*(i;B), t^*_{j'} \setminus t^*_B \text{ is positive}] \cdot u_i \\
&= \sum_{j \in [B-1]} \sum_{i \in [n]} P^{T^*}_{i,j} \cdot u_i + q_S \cdot q_{t^*_B \setminus S} \cdot \sum_{i \in t^*_B} \Pr[\forall t^*_{j'} \in T^*(i;B), t^*_{j'} \setminus t^*_B \text{ is positive}] \cdot u_i \\
&= \sum_{j \in [B-1]} \sum_{i \in [n]} P^{T^*}_{i,j} \cdot u_i + q_{t^*_B \setminus S} \cdot \left( q_S \cdot \sum_{i \in S} \Pr[\forall t^*_{j'} \in T^*(i;B), t^*_{j'} \setminus t^*_B \text{ is positive}] \cdot u_i \right) \\
&\quad + q_S \left( q_{t^*_B \setminus S} \cdot \sum_{i \in t^*_B \setminus S} \Pr[\forall t^*_{j'} \in T^*(i;B), t^*_{j'} \setminus t^*_B \text{ is positive}] \cdot u_i \right) \\
&\leq \sum_{j \in [B-1]} \sum_{i \in [n]} P^{T^*}_{i,j} \cdot u_i + q_{t^*_B \setminus S} \cdot \left( q_S \cdot \sum_{i \in S} \Pr[\forall t^*_{j'} \in T^*(i;B), t^*_{j'} \setminus S \text{ is positive}] \cdot u_i \right) \\
&\quad + q_S \left( q_{t^*_B \setminus S} \cdot \sum_{i \in t^*_B \setminus S} \Pr[\forall t^*_{j'} \in T^*(i;B), t^*_{j'} \setminus (t^*_B \setminus S) \text{ is positive}] \cdot u_i \right) \\
&< \sum_{j \in [B-1]} \sum_{i \in [n]} P^{T^*}_{i,j} \cdot u_i + \max \Bigg\{ q_S \cdot \sum_{i \in S} \Pr[\forall t^*_{j'} \in T^*(i;B), t^*_{j'} \setminus S \text{ is positive}] \cdot u_i, \\
&\qquad q_{t^*_B \setminus S} \cdot \sum_{i \in t^*_j \setminus S} \Pr[\forall t^*_{j'} \in T^*(i;B), t^*_{j'} \setminus (t^*_B \setminus S) \text{ is positive}] \cdot u_i \Bigg\}
\end{aligned}$$

To justify the first inequality, we begin by showing that

$$\sum_{i \in S} \Pr[\forall t^*_{j'} \in T^*(i;j),\ t^*_{j'} \setminus t^*_B \text{ is positive}] \leq \sum_{i \in S} \Pr[\forall t^*_{j'} \in T^*(i;j),\ t^*_{j'} \setminus S \text{ is positive}].$$

Since $S \subseteq t^*_B$, it follows that for any $t^*_{j'} \in T^*(i;j)$ we have $t^*_{j'} \setminus t^*_B \subseteq t^*_{j'} \setminus S$. This in turn implies that if $t^*_{j'} \setminus t^*_B$ is positive, it must be the case that $t^*_{j'} \setminus S$ is positive. It follows that the event where $\forall t^*_{j'} \in T^*(i:j),\ t^*_{j'} \setminus t^*_B$ is positive implies that $\forall t^*_{j'} \in T^*(i:j),\ t^*_{j'} \setminus S$ is positive, hence the inequality follows. Furthermore, the argumentation above can be replicated with $t_B^* \setminus S$ rather than $S$ to fully justify the inequality from the main derivation. As for the second inequality, in the derivation, this follows from the assumption that $q_S < \alpha$ and $q_{t_B^* \setminus S} < 1 - \alpha$.

To reach a contradiction, let $T$ be a testing regime with $t_j = t_j^*$ for any $j \in [B-1]$ and $t_B = S$ if

$$q_S \cdot \sum_{i \in S} \Pr[\forall t_{j'}^* \in T^*(i;B), t_{j'}^* \setminus S \text{ is positive}] \cdot u_i$$
$$\geq q_{T_B^* \setminus S} \cdot \sum_{i \in T_B^* \setminus S} \Pr[\forall t_{j'}^* \in T^*(i;B), t_{j'}^* \setminus (t_B^* \setminus S) \text{ is positive}] \cdot u_i$$

and $t_B = t_B^* \setminus S$, otherwise. Then,

$$\begin{aligned}
u(T) &= \sum_{i \in [n]} P_i^T \cdot u_i \\
&= \sum_{i \in [n]} \sum_{j \in [B]} P_{i,j}^T \cdot u_i \\
&= \sum_{j \in [B-1]} \sum_{i \in [n]} P_{i,j}^{T^*} \cdot u_i + \sum_{i \in t_B} q_{t_B} \cdot \Pr[\forall t_{j'}^* \in T^*(i;B), t_{j'}^* \setminus t_B \text{ is positive}] \cdot u_i \\
&= \sum_{j \in [B-1]} \sum_{i \in N} P_{i,j}^{T^*} \cdot u_i + \max \Bigg\{ q_S \cdot \sum_{i \in S} \Pr[\forall t_{j'}^* \in T^*(i,B), t_{j'}^* \setminus S \text{ is positive}] \cdot u_i, \\
&\qquad q_{t_B^* \setminus S} \cdot \sum_{i \in T_B \setminus S} \Pr[\forall t_{j'}^* \in T^*(i;B), t_{j'}^* \setminus (t_B^* \setminus S) \text{ is positive}] \cdot u_i \Bigg\} \\
&> u(T^*),
\end{aligned}$$

which contradicts the optimality of $T^*$ and completes the proof of our claim. □

We are now ready to prove the theorem.

Proof of Theorem 1. We begin by constructing an intermediate non-overlapping testing regime $T$ from $T^*$ by simply assigning each individual to only a single test chosen arbitrary among all tests to which that individual is assigned to in $T^*$. Thus, we have that for each $j \in [B]$ $t_j \subseteq t_j^*$ which meana that $q_{t_j} \geq q_{t_j^*}$. Now, for each $t_j$, we let $S_j$ be the smallest subset of $t_j$ such that $q_{S_j} < 1/2$ (if $q_{t_j} \geq 1/2$, then $S_j = \emptyset$). Note that for each $i \in S_j$, we have $q_{S_j \setminus \{i\}} \geq 1/2$, as $S_j$ is the smallest possible set that has probability less than $1/2$ to be negative. In addition, we can show that $q_{t_j \setminus S_j} \geq 1/2$. To see this, notice that $S_j \subseteq t_j \subseteq t_j^*$ and $q_{S_j} < 1/2$. From Lemma 2, we know that $q_{t_j^* \setminus S_j} \geq 1/2$, but since $t_j \setminus S_j \subseteq t_j^* \setminus S_j$, it follows that $q_{t_j \setminus S_j} \geq q_{t_j^* \setminus S_j} \geq 1/2$ as desired.

Next, consider two disjoint testing regimes given by $T^1$ where $t_j^1 = S_j$ and $T^2$ where $t_j^2 = t_j \setminus S_j$. Using the properties of $T$ from the previous paragraph, we wish to show that for each $i \in t_j^\ell$ where $\ell \in \{1,2\}$, we have that $P_i^{T^\ell} \geq q_i \cdot 1/2$. To that end, in the case of $\ell = 1$, we get $P_i^{T^1} = q_{t_j^1} = q_{S_j} < 1/2$. However, we also know that $q_{t_j^1 \setminus \{i\}} = q_{S_j \setminus \{i\}} \geq 1/2$, and hence $q_{t_j^1} = q_i \cdot q_{t_j^1 \setminus \{i\}} \geq q_i \cdot 1/2$ as desired. As for the case where $\ell = 2$, we get $P_i^{T^2} = q_{t_j^2} = q_{t_j \setminus S_j}$. The right hand side can be decomposed as $q_{t_j \setminus S_j} = q_i \cdot q_{t_j^2 \setminus \{i\}}$. However, as we have shown above, the choice of $S_j$ ensures that $q_{t_j \setminus S_j} \geq 1/2$, and it follows that $q_i \cdot q_{t_j^2 \setminus \{i\}} \geq 1/2 \geq q_i \cdot 1/2$. We conclude that $P_i^{T^2} \geq q_i \cdot 1/2$, as desired.

Without loss of generality, assume that $\bigcup_{j \in [B]} t_j^* = [n']$ for some $n' \leq n$, i.e., the first $n'$ individuals are included in some test under $T^*$. Notice that $\bigcup_{j \in [B]} t_j = [n']$, as each individual who is included in some test under $T^*$ is also included in some test under $T$, and no individual who is not included in some test under $T^*$ is included in $T$. Thus, we get that

$$u(T^*) = \sum_{i\in[n']} P_i^{T^*} \cdot u_i = \sum_{j\in[B]} \left( \sum_{i\in S_j} P_i^{T^*} \cdot u_i + \sum_{i\in T_j\setminus S_j} P_i^{T^*} \cdot u_i \right).$$

We now consider two cases, depending on whether $\sum_{j\in[B]} \sum_{i\in S_j} P_i^{T^*} \cdot u_i \geq \sum_{j\in[B]} \sum_{i\in t_j\setminus S_j} P_i^{T^*} \cdot u_i$. If this is the case, we get that

$$u(T^*) \leq 2 \cdot \sum_{j\in[B]} \sum_{i\in S_j} P_i^{T^*} \cdot u_i \leq 2 \cdot \sum_{j\in[B]} \sum_{i\in S_j} q_i \cdot u_i.$$

It also holds that

$$u(T^1) = \sum_{j\in[B]} \sum_{i\in S_j} P_i^{T^1} \cdot u_i \geq \sum_{j\in[B]} \sum_{i\in S_j} 1/2 \cdot q_i \cdot u_i.$$

Thus, we conclude that

$$\frac{u(T^*)}{u(T^1)} \leq \frac{2\sum_{i\in S_j} q_i \cdot u_i}{\sum_{i\in S_j} 1/2 \cdot q_i \cdot u_i} \leq 4.$$

Using similar arguments, we can show that $u(T^*)/u(T^2) \leq 4$ when $\sum_{j\in[B]} \sum_{i\in S_j} P_i^{T^*} \cdot u_i < \sum_{j\in[B]} \sum_{i\in T_j\setminus S_j} P_i^{T^*} \cdot u_i$, and the theorem follows. □

While this proves that the gain of overlaps cannot be larger than 4, the worst known example is the one illustrated in Proposition 1, providing a lower bound of 7/6. Interestingly, after running many simulations we were not able to find a better lower bound, and we believe that the gain of overlaps is less than 4. Moreover, it is possible to show that gain(3) $\leq 7/3$ and gain(4) $\leq 15/4$; the details of these bounds are in Appendix D.

### 3.3 Greedy Algorithm for the Non-Overlapping Testing Regime

In light of the previous result, hereinafter we focus on non-overlapping testing regimes. Consider the case where $B = 1$. If $G$ is a constant, we can efficiently enumerate all $O(n^G)$ potential pooled tests and find the optimal test $t^*$. On the other hand, when $G = n$, it follows from the work of Goldberg and Rudolf [2020] that even when there is only one test, it is NP-hard to find the subset of individuals that maximizes the expected welfare of the test. On the positive side, they provide a fully polynomial-time approximation scheme (FPTAS) for the same case. Here, we show that we can adjust the main ideas of their algorithm to obtain an FPTAS for the case where there is one test with up to $G \in [n]$ individuals in it; the proof is relegated to Appendix E.

Lemma 3. *When $B = 1$, there is an FPTAS for computing approximately optimal $(t^*) \in \mathcal{T}^1$.*

Our goal is to approximate the optimal non-overlapping testing regime when there are $B$ available tests. Given the FPTAS of Lemma 3 for the single test case, a natural greedy approach is the following: design a test in each step by applying the FPTAS for the single test case to the remaining individuals. In other words, in each iteration we greedily find the test that approximates the optimal test over the available individuals using the FPTAS, add this test to the testing regime, disregard all individuals that are included in this test and continue to create greedy tests in the same fashion for the remaining individuals until the budget is exhausted. The above procedure results in an non-overlapping testing regime with at most $B$ tests, as we never consider individuals that have already been included in a test. We refer this algorithm as *$\epsilon$-Greedy*, where $\epsilon$ is the error tolerance used in the FPTAS algorithm[6]. We show that this algorithm gives a $5/(1-\epsilon)$ approximation to the optimal non-overlapping testing regime.

Theorem 2. *$\epsilon$-Greedy returns a $5/(1-\epsilon)$-approximate non-overlapping testing regime.*

Proof. Let $T$ be the testing regime that is returned by $\epsilon$-Greedy and $T^*$ be an optimal non-overlapping testing regime. Without loss of generality, let $N' = \{1, \dots, n'\}$ be the set of individuals that are pooled into a test in $T$. Then,

$$u(T^*) = \sum_{j \in [B]} u(t_j^*) = \sum_{j \in [B]} q_{t_j^*} \cdot \left( \sum_{i \in N'} I_{i,j}^{t^*} \cdot u_i \right) + \sum_{j \in [B]} q_{t_j^*} \cdot \left( \sum_{i \in [n] \setminus N'} I_{i,j}^{t^*} \cdot u_i \right),$$

and

$$u(T) = \sum_{j \in [B]} u(t_j) = \sum_{j \in [B]} q_{t_j} \cdot \left( \sum_{i \in N'} I_{i,j}^{T} \cdot u_i \right) = \sum_{j \in [B]} q_{t_j} \cdot \left( \sum_{i \in t_j} u_i \right).$$

Now, let $T'$ be a testing regime such that $t'_j = t_j^* \setminus (t_j^* \cap N')$. In other words, $T'$ is created by removing from $T^*$ any individual in $N'$. Notice that since every $t'_j$ consists of individuals that are not included in any test in $T$, it means that all the individuals in $t'_j$ are available at the $j$-th iteration of the algorithm, and thus we get that for each $j \in [B]$, $u(t_j) \geq (1-\epsilon) u(t'_j)$, as otherwise the algorithm would have chosen $t'_j$ instead of $t_j$, at the $j$-th iteration. Thus, we get that

$$u(T) = \sum_{j \in [B]} u(t_j) \geq (1-\epsilon) \sum_{j \in [B]} u(t'_j) = (1-\epsilon) \cdot u(T'),$$

and also note that

$$\begin{aligned} u(T') = \sum_{j \in [B]} q_{t'_j} \cdot \left( \sum_{i \in [n] \setminus N'} I_{i,j}^{T'} \cdot u_i \right) &= \sum_{j \in [B]} q_{t'_j} \cdot \left( \sum_{i \in [n] \setminus N'} I_{i,j}^{T^*} \cdot u_i \right) \\ &\geq \sum_{j \in [B]} q_{t_j^*} \cdot \left( \sum_{i \in [n] \setminus N'} I_{i,j}^{T^*} \cdot u_i \right) \end{aligned}$$

where the second equality follows from the fact that $I_{i,j}^{T'} = I_{i,j}^{T^*}$ for any $i \in [n] \setminus N'$ and $j \in [B]$, and the last inequality follows from the fact that $q_{t'_j} \geq q_{t_j^*}$ since $t'_j \subseteq t_j^*$ for any $j \in [B]$. Thus,

$$u(T) \geq (1-\epsilon) \cdot \sum_{j \in [B]} q_{t_j^*} \cdot \left( \sum_{i \in [n] \setminus N'} I_{i,j}^{T^*} \cdot u_i \right).$$

From all the above we have

$$\begin{aligned} \frac{u(T^*)}{u(T)} &= \frac{\sum_{j \in [B]} q_{t_j^*} \cdot \left( \sum_{i \in N'} I_{i,j}^{T^*} \cdot u_i \right) + \sum_{j \in [B]} q_{t_j^*} \cdot \left( \sum_{i \in [n] \setminus N'} I_{i,j}^{T^*} \cdot u_i \right)}{u(T)} \\ &\leq \frac{\sum_{j \in [B]} q_{t_j^*} \cdot \left( \sum_{i \in N'} I_{i,j}^{T^*} \cdot u_i \right) + \frac{u(T)}{1-\epsilon}}{u(T)} \end{aligned}$$

[6]When $G$ is constant, we can efficiently compute optimal $t^*$ at each step of the approach above via brute force. We call this algorithm *Greedy*.

$$= \frac{\sum_{j\in[B]} q_{t_j^*} \cdot \left(\sum_{i\in t_j^*\cap N'} u_i\right)}{u(T)} + \frac{1}{1-\epsilon}$$
$$\leq \frac{\sum_{j\in[B]} \left(\sum_{i\in t_j^*\cap N'} q_i \cdot u_i\right)}{u(T)} + \frac{1}{1-\epsilon}$$
$$= \frac{\sum_{i\in N'} q_i \cdot u_i}{u(T)} + \frac{1}{1-\epsilon} \tag{1}$$

where the second inequality follows since $q_i \geq q_{t_j}$ when $i$ is included in $t_j$.

In what follows, we will show that each test, $t_j \in T$ obtains at least a $\frac{(1-\epsilon)}{4}$ ratio of the maximal possible utility to be gained from individuals in $t_j$. In other words, we show that the following holds: $u(t_j) = q_{t_j} \cdot \sum_{i\in t_j} u_i > \frac{(1-\epsilon)}{4} \sum_{i\in t_j} q_i u_i$. To do so, we first consider the case where there exists $i' \in t_j$ such that $q_{i'} < 1/2$. From the definition of the greedy algorithm, we know that

$$(1-\epsilon) \cdot q_{t_j\setminus\{i'\}} \cdot \sum_{i\in t_j\setminus\{i'\}} u_i \leq q_{t_j} \cdot \sum_{i\in t_j} u_i$$

otherwise the algorithm would return $t_j \setminus \{i'\}$ instead of $t_j$ at step $j$ and also,

$$(1-\epsilon) \cdot q_{i'} \cdot u_{i'} \leq q_{t_j} \cdot \sum_{i\in t_j} u_i$$

as otherwise the algorithm would return $\{i'\}$ instead of $t_j$ at the $j$-th iteration. Moreover, from Lemma 2, we know that $q_{t_j\setminus\{i'\}} \geq 1/2$ since $q_{i'} < 1/2$. Thus, we get that

$$q_{t_j} \cdot \sum_{i\in t_j} u_i \geq \frac{(1-\epsilon)}{2}\left(q_{t_j\setminus\{i'\}} \cdot \sum_{i\in t_j\setminus\{i\}} u_i + q_{i'} \cdot u_{i'}\right) \geq \frac{(1-\epsilon)}{2}\left(\frac{1}{2} \cdot \sum_{i\in t_j\setminus\{i\}} u_i + q_{i'} \cdot u_{i'}\right)$$
$$= \frac{(1-\epsilon)}{4}\left(\sum_{i\in t_j\setminus\{i'\}} u_i + q_{i'} \cdot u_{i'}\right)$$
$$\geq \frac{(1-\epsilon)}{4} \sum_{i\in t_j} q_i u_i.$$

As a second case, assume that for any $i \in t_j$, $q_i \geq 1/2$. We show that for each $i \in t_j$, $q_{t_j\setminus\{i\}} \geq 1/4$. If $q_{t_j} \geq 1/2$, then indeed $q_{t_j\setminus\{i\}} \geq 1/4$. Otherwise, we do the following: In a set $S$ we add individuals that are included in $t_j$, except for $i$, one at a time until $q_S \geq 1/2$ and $q_{S\cup\{i\}} < 1/2$ (as $q_{t_j} < 1/2$ notice that such an $S$ should exist). Then, from Lemma 2, we get that $q_{t_j\setminus\{S\cup i\}} \geq 1/2$, and hence $q_{t_j\setminus\{i\}} = q_S \cdot q_{t_j\setminus\{S\cup i\}} \geq 1/4$. Thus, we have that

$$q_{t_j} \cdot \sum_{i\in t_j} u_i = \sum_{i\in t_j} q_{t_j\setminus\{i\}} \cdot q_i \cdot u_i \geq \frac{1}{4} \cdot \sum_{i\in t_j} q_i \cdot u_i \geq \frac{(1-\epsilon)}{4} \sum_{i\in t_j} q_i \cdot u_i.$$

We see that in either case, $q_{t_j} \cdot \sum_{i\in t_j} u_i > \frac{(1-\epsilon)}{4} \cdot \sum_{i\in t_j} q_i \cdot u_i$ hence we get:

$$u(T) = \sum_{j\in[B]} q_{t_j} \cdot \left(\sum_{i\in t_j} u_i\right) > \sum_{j\in[B]} \left(\frac{(1-\epsilon)}{4} \cdot \sum_{i\in t_j} q_i \cdot u_i\right) = \frac{(1-\epsilon)}{4} \sum_{i\in N'} q_i u_i$$

Along with Equation (1), we get that $u(T^*)/u(T) \leq 5/(1-\epsilon)$ and the theorem follows. □

Note that one can combine Theorem 1 and Theorem 2 to see that $\epsilon$-Greedy gives a constant-factor approximation to the optimal *overlapping* testing regime. Furthermore, we can also show that Greedy is optimal in instances where testing budgets are low and individuals can only take utilities and probabilities from a finite set of values. Specifically, assume that the population at hand can be partitioned into $C$ clusters, where the $i$-th cluster has $n_i$ individuals with identical utility $u_i$ and probability of infection $p_i$[7]. Moreover, suppose that $(t^*) \in \mathcal{T}^1$ is an optimal single pooled test for this clustered population. In Appendix F we prove the following result, which identifies a natural setting where greedy is optimal — even with respect to the optimal *overlapping* testing regime.

Proposition 3. *For testing budget $B > 0$, if $B \cdot |t^* \cap C_i| \leq n_i$, Greedy returns an optimal allocation that applies $B$ distinct copies (in terms of composition) of $t^*$.*

Finally, in Appendix G, we also consider the case where all the individuals have the same utility. First, we show that when $B$ is a constant we can find the optimal testing regime. For general $B$, we design a variant of Greedy which sorts the individuals in decreasing order with respect to the probability of being healthy and in each step adds individuals to the current test as long as the expected utility of the test decreases. We show that this algorithm returns an $e$-approximate non-overlapping testing regime and this result is tight.

## 4 PRACTICAL ALGORITHMIC IMPLEMENTATIONS

As discussed above, even the problem of allocating a single test is computationally hard. In order to make the computation of testing regimes tractable for our pilot study, we formulate the problem of non-overlapping testing as optimization problems that we solve using commercial solvers. The problem of allocating a single test can be formulated as a mixed-integer conic optimization program (MICP), and solved using a commercial conic solver. This implementation is used by our Greedy algorithm in our simulations and our pilot study. When multiple tests are to be allocated, we formulate a mixed-integer linear program (MILP) that approximates an optimal non-overlapping solution and can be solved by any MILP solver. The MILP formulation approximates existing exponential constraints with piecewise-linear functions that can be formulated as a collection of mixed integer linear constraints. The accuracy of this approximation can be adjusted by tuning the number $K$ of segments of the piecewise-linear function, at the cost of introducing more (integer) variables and thus the time to solve the program. We provide practical (additive) approximation guarantees for this *Approx algorithm* depending on parameter $K$.

*The optimization program.* We state the non-linear program for determining optimal non-overlapping testing regimes and describe a conic formulation for the single test case as well as a mixed-integer formulation for approximating optimal non-overlapping with one or more tests. We can assume that the testing budget $B$ is at most the population size $n$, and so pool sizes lies between 1 and $G$. For each test $j \in [B]$, we introduce an indicator vector $x^j \in \{0,1\}^n$ with $x_i^j = 1$ if individual $i$ is included in $j$ and $x_i^j = 0$ otherwise, and let variable $w^j$ denote its expected utility $w^j = u \cdot x^j \prod_{i \in [n]} q_i^{x_i^j}$. We impose pool sizes between 1 and $G$ with constraints $1 \leq \sum_{i \in [n]} x_i^j \leq G$ for all $j \in [B]$, and non-overlapping testing with constraints $\sum_{j \in [B]} x_i^j \leq G$ for all $i \in [n]$. Our objective is to maximize welfare $\sum_{j \in [B]} w^j$. In order to isolate the non-linear elements of the optimization problem, we reformulate the convex program with additional variables below; variables $l^j$ denote the log of $w^j$, and variables $y^j$ and $z^j$ allow us to isolate the non-linear elements of the

[7]We highlight that clusters are a very natural constraint in terms of population structure which we intend to use in our pilot with our partner institution in Mexico.

expressions into constraints (2b) and (2d).

$$\max \quad \sum_{j\in[B]} w^j \tag{2a}$$

$$\text{s.t.} \quad w^j = \exp l^j, \qquad \forall j \in [B], \tag{2b}$$

$$l^j = y^j + \sum_{i\in[n]} x_i^j \log q_i, \quad \forall j \in [B], \tag{2c}$$

$$y^j = \log z^j, \qquad \forall j \in [B], \tag{2d}$$

$$z^j = u \cdot x^j, \qquad \forall j \in [B], \tag{2e}$$

$$\sum_{j\in[B]} x_i^j \le 1, \qquad \forall i \in [n], \tag{2f}$$

$$1 \le \sum_{i\in[n]} x_i^j \le G, \qquad \forall j \in [B], \tag{2g}$$

$$x_i^j \in \{0,1\}, \qquad i \in [n], \forall j \in [B] \tag{2h}$$

### 4.1 A Conic Program for a Single Test

Suppose we wish to allocate a single test. In this setting, we can eliminate the exponential constraint (2b) by changing the objective to $\max l^1$. The remaining non-linear constraints (2d) can be relaxed to $y^j \le \log z^j$ without affecting the outcome, and formulated as conic constraints $(z^j, 1, y^j) \in K_{\exp}$, where $K_{\exp}$ is the exponential cone defined as $K_{\exp} = \{(x_1, x_2, x_3) \mid x_1 \ge x_2 e^{x_3/x_2}, x_2 > 0\} \cup \{(x_1, 0, x_3) \mid x_1 \ge 0, x_3 \le 0\}$. The resulting mixed-integer conic optimization program can be solved efficiently[8] with conic solvers such as MOSEK (https://mosek.com). In our practical implementation, the Greedy algorithm repeatedly solves a conic program to allocate a single test.

### 4.2 A Mixed-Integer Linear Programming Approximation

If we wish to allocate more than one test, the problem no longer admits a conic formulation. Instead, we formulate a mixed-integer linear program (MILP) that finds an approximately optimal non-overlapping solution. In order to make the problem tractable, we assume that the utility vector $u$ is integral and non-negative. This assumption is benign, as the problem is invariant to scaling of utilities. We describe how the non-linear constraints (2b) and (2d) can respectively be captured exactly and approximately by a collection of integer linear constraints. In Appendix H, we also state the full mixed-integer linear program with an additional refinement that clusters individuals in the population who have identical utilities and probabilities, speeding up computation in practice.

*Handling the logarithmic constraints.* We can replace (2d) with integer linear constraints as follows. Fix some test $j \in [B]$. Note that $z^j$ takes integral values in the range $[L, U]$, where $L = \min_i u_i$ and $U = G \max_i u_i$. We introduce an indicator vector $\gamma^j \in \{0,1\}^{[L,U]}$ indexed by $k \in [L, U]$ with constraints $\sum_{k\in[L,U]} \gamma_k^j = 1$ and $\sum_{k\in[L,U]} k \cdot \gamma_k^j = z^j$ to encode which value $z$ holds, and ensure $y^j = \log(z^j)$ with the constraint $y^j = \sum_{k\in[L,U]} \log(k) \cdot \gamma_k^j$.

*Approximating the exponential constraints.* We now describe how to approximate (2b) from above by a piecewise-linear function $f$ using integer linear constraints. Fix some test $j \in [B]$. Note first that we can relax the equality in (2b) to $w^j \le \exp(l^j)$ without affecting the outcome. The variable $l^j$ takes values between $A = \min_i(\log u_i) + G \min_i(\log q_i)$ and $B = \log(G \max_i u_i) + \max_i(\log q_i)$ (and these values will be generically non-integral). We approximate exp from above by a piecewise-linear

[8]Example running times are shown in Table 1 (Section 5.2) and Table 4 (Appendix I).

function $f : [A, B] \to \mathbb{R}$ with $K$ linear segments. (Here the parameter $K$ is given exogenously.) Partitioning $[A, B]$ into $K$ parts $[c_k, c_{k+1}], k \in [K]$, we define the $k$-th line segment as the linear function $f_k(x) = a_k x + b_k$ on domain $[c_k, c_{k+1}]$ with slope $a_k = \frac{\exp c_{k+1} - \exp c_k}{c_{k+1} - c_k}$ and residual $b_k = \exp c_{k+1} - a_k c_{k+1}$. Note that the number of integer variables in the MILP increases with $K$, so this parameter must be chosen judiciously. Moreover, given a fixed number of segments $K$, we wish to determine a partitioning of $[A, B]$ that minimizes the approximation error $\varepsilon = \max_{x \in [A,B]}(f(x) - \exp(x))$. In our implementation, we apply binary search techniques to numerically determine the partition of $[A, B]$ such that the error $\max_{x \in [c_k - c_{k+1}]}(f_k(x) - \exp(x))$ is the same for all parts $[c_k, c_{k+1}]$, which minimizes $\varepsilon$. We introduce indicator vectors $\delta^j \in \{0, 1\}^K$ to encode in which part $[c_k, c_{k+1}]$ the value of $l^j$ lies, as well as the vector $v^j \in \mathbb{R}^K$ whose $k$-th entry agrees with $l^t$ if $l^j$ lies in the $k$-th part, and is 0 otherwise. This is guaranteed by constraints $\sum_{k \in [K]} \delta_k^j = 1$, $l^j = \sum_{k \in [K]} v_k^j$ and $c_k \cdot \delta_k^j \leq v_k^j \leq c_{k+1} \cdot \delta_k^j, \forall k \in [K]$. Finally, we require that $w^j \leq f_k(l^j)$ for the $k$-th part $[c_k, c_{k+1}]$ that $l^t$ lies in. This is expressed by constraint $w^j \leq \sum_{k \in [K]} a_k v_k^j + b_k \cdot \delta_k^j$.

*Bounding the approximation error.* Recall that the piecewise-function $f$ with $K$ segments approximates exp on domain $[A, B]$ from above with error $\varepsilon$. Let $\sigma(x) = \sum_{j \in [B]} \exp(l^t)$ and $\sigma'(x) = \sum_{j \in [B]} f(l^t)$ respectively denote the corresponding objective values of the convex program (2) and the MILP described above for testing $x$. Let $x^*$ denote an optimal non-overlapping testing, so $x^*$ maximizes $\sigma$, and $x'$ be an optimal solution for the MILP. Clearly, $x^*$ and $x'$ are both feasible for both programs and satisfy $\sigma(x') \leq \sigma(x^*)$ as well as $\sigma'(x^*) \leq \sigma'(x')$. By construction of $f$, we have $\sigma(x) \leq \sigma'(x)$ and $\sigma(x) \geq \sigma'(x) - \varepsilon B$, which implies $\sigma(x^*) \leq \sigma'(x^*) \leq \sigma'(x') \leq \sigma(x') + T\varepsilon$. Here $\varepsilon$ is the additive approximation error of $f$ with regard to exp. Hence, $0 \leq \sigma(x^*) - \sigma(x') \leq \varepsilon B$. This allows us to compute a bound on the additive gap between the welfare achieved by the optimal solution of our MILP and the optimal non-overlapping testing.

## 5 EMPIRICAL RESULTS

In September 2022, we ran a version of our utility-based pooled testing regime in a randomized control trial at the Potosinian Institute for Scientific and Technological Research (IPICYT), a higher education research institute in San Luis Potosí, Mexico. At IPICYT, a heterogeneous population of 130 individuals participated in the trial of our testing and reopening strategy, including students, academics, and administrative staff.[9]

Shortly before our trial commenced, IPICYT resumed full access for all its members.[10] This enabled us to compare the utility-based pooled testing regime to what one expects to be a normal working environment (the control group), thus providing a ‘first-best’ benchmark with respect to productivity, performance and learning at the workplace, as well as mental health and subjective well-being of individuals. Our trial provides causal evidence that the utility-based pooled testing regime does *no worse* than the ‘first-best’ with respect to the above-mentioned outcomes. Moreover, in contrast with full access, our testing mechanism ensures that everyone with in-person access is guaranteed to be non-infectious, hence providing a much safer work environment.

Section 5.1 describes how health probabilities and utilities were determined for participants in our pilot population. The data thus obtained also allows us to run simulations (see Section 5.2) that demonstrate the performance of the Approx and Greedy algorithms from Section 4.

[9]At the end of the pilot, we collected between 118 and 122 complete data points, depending on the outcome of analysis.
[10]IPICYT mandated individuals who had recently traveled out of IPICYT’s home state of San Luis Potosí to provide a negative lateral flow test upon their return, but in practice this was not enforced.

### 5.1 Determining Utilities and Health Probabilities

The crucial input to the problem described in Section 2 is the individuals' utilities as well as health probabilities. In our pilot study, we constructed the utilities based on multi-dimensional measures of (i) the need for in-person access to work/study resources on campus, (ii) socioeconomic status, and (iii) mental health status.

The premise for developing our utility measure was that an individual's need for in-person work or study heavily depends on the nature of their work: e.g., an experimentalist in a lab must attend their experiments more frequently than a theoretician must visit their office on campus. The participants' need for in-person access due to the nature of their work was inferred from questions about their use of digital media (details in Appendix K.7). These questions were designed so that subjects would not be immediately able to judge how to answer the question in order to be prioritized for in-person work. Furthermore, we considered evidence that the closure of learning environments disproportionately affects vulnerable populations, e.g. individuals with low-income [Azevedo et al., 2021, Bandiera et al., 2019, Gorgen and McAleavy, 2020, Goudeau et al., 2021, Hossain, 2021].[11] While various protected attributes are associated with vulnerability, we chose to not discriminate based on those attributes, but obtain self-reported information on a more simple and direct proxy of vulnerability, socio-economic status.[12] Additionally, mental health is known to be negatively affected by pandemic-induced remote work for younger and and older individuals [Asanov et al., 2021, Bertoni et al., 2022], and students and employees in Mexico have struggled with mental health problems associated with COVID-19 [Limón-Vázquez et al., 2020, Martinez Arriaga et al., 2021].

Our three measures were based on the subjects' answers to a survey they were given before the testing period of our trial. $u_i^{pr}$ captures the utility subject $i$ gains from increased productivity and $u_i^{psy}$ captures the benefit on subject $i$'s mental health from attending in-person; $u_i^{se}$ is a utility bonus for socio-economically disadvantaged individuals, reflecting the fact they are likely to be disproportionately affected by having to work remotely. The overall utility score is a weighted sum given by $u_i = \sum_{k\in\{pr,psy,se\}} w^k u_i^k$.[13] We define the composition of $u_i^k$ for category $k \in \{pr, psy, se\}$. Let $P_{i,z}^k$ denote the number of points "achieved" by the answer of subject $i$ to question $z$, and $Z^k$ the number of questions relevant in category $k$.[14] For each category, the score is $u_i^k = \frac{1}{Z^k}\sum_z P_{i,z}^k$.

Health probabilities are estimated for age and gender categories using Bayesian updates of local public health data. They remain constant throughout the trial. We computed the probability of being infected conditional on being in one of the following 6 groups: {male, female} × {age 15-29, 30-59, ≥60}. The baseline probability of infection for a given age group is determined using Bayesian updates of local public health data and under the guidance of local epidemiologists. More specifically, we used publicly available epidemiological models from the Institute of Health Metrics and Evaluation (IHME) to estimate baseline infection rates in San Luis Potosí.[15] These estimates provided us with values for $\Pr[\text{infection}]$ for all individuals in the population, irrespective of their category. Furthermore, we estimated the probability an individual belongs to a given category

[11] While the cited studies of heterogeneous effects on vulnerable populations are conducted with students (ranging from K-12 to Higher Education), they also document similar issues for teaching staff and may extend to academic research.

[12] Especially in Mexico, having low income is highly correlated with belonging to an ethnic minority, the elderly, or being female [Ordóñez Barba, 2018].

[13] The weights $w^{pr}$, $w^{psy}$, and $w^{se}$ are chosen by the intervention's designer and are $(1/3, 1/3, 1/3)$ in our trial.

[14] Relevance of a question to a specific category is marked in the survey in Appendix K.7 by the corresponding abbreviation just after the question numbering. Not all questions are relevant for the construction of utilities.

[15] Estimated infection rates for SLP with IHME models can be found at their dashboard for different public behaviour regimes: https://covid19.healthdata.org/mexico/san-luis-potosi?view=infections-testing&tab=trend&test=infections.

given an infection via official national data on testing results.[16] These estimates provide us with values for Pr[category | infected] for each category. Finally, we used census data to compute the probability of membership in a given category at the state/national level.[17] This provides us with an estimate for Pr[category] for the population. With Bayes' rule, we compute the desired probability of infection per category as follows: $\Pr[\text{infection} \mid \text{category}] = \frac{\Pr[\text{category}|\text{infection}]\Pr[\text{infection}]}{\Pr[\text{category}]}$. The probabilities of being healthy were approximately 99.5% for each group. The probabilities stay constant throughout this trial, as it only ran for 4 weeks. If applied over a longer period, the health probabilities may also be updated.[18]

## 5.2 Simulations

We evaluate the accuracy and running times of the Greedy and Approx algorithms on populations that reflect real-world scenarios. In our first experiments, we run both algorithms on data from our pilot study with budgets $B \in \{2, 4, \dots, 12\}$.[19] As the health probabilities observed during our pilot study were high, this illustrates the efficacy of Greedy in times of lower disease incidence. In order to study how well Greedy performs when faced with higher infection rates, we also ran experiments on synthetic data in which health probabilities range from 0.5 to 1 and budgets $B \in \{2, 4, \dots, 10\}$. Moreover, while our partnering testing laboratory performed pooled testing with saliva samples, which has a pool size limit of $G = 5$, we also study outcomes when pool sizes are limited to $G = 10$ (the limit for nasopharyngeal swabs).

In our synthetic experiments, we showcase the average-case behavior of Greedy and Approx by generating random populations of size $n = 150$. Health probabilities are drawn independently and uniformly at random from the interval $[0.5, 1]$, and utilities are drawn from a normal distribution that was fitted to the utilities observed in our pilot study. We then run Greedy and Approx on each population for each pool size $G \in \{5, 10\}$ and for all testing budgets $B \in \{2, 4, \dots, 10\}$, recording the welfare achieved for both algorithms, as well as their running times (in milliseconds).[20]

For all experiments, we record the true welfares achieved by the testing regimes returned by both algorithms, and not the objective values of the underlying MILP and conic optimization problems, as the latter will be an approximation of the true welfare. For Approx, we tune the parameter $K$ of the MILP formulation so that the additive approximation guarantee (cf. Section 4.2) is small ($K = 25$ for the experiments on pilot data, and $K = 20$ for the experiments on synthetic data). The code used to run these experiments can be found at redacted for anonymity. We also refer to Appendix J for preliminary experiments that compare non-overlapping with overlapping testing on small populations; these give additional evidence that the average-case gain from overlaps is small.

*Results.* Table 1 lists the welfares achieved by Approx and Greedy on our pilot study data for pool size constraint $G = 5$, as well as the running times for both algorithms and the approximation guarantee achieved by Approx. Table 3 in Appendix I shows analogous results for pool size constraint $G = 10$. We observe that Greedy achieves near-optimal welfare for budgets up to 10 for both pool sizes. Moreover, while the running time of Approx appears to increase exponentially with

[16]National testing aggregates can be found at https://datos.covid-19.conacyt.mx and https://covid19.healthdata.org/mexico/san-luis-potosi?view=infections-testing&tab=trend&test=infections.

[17]Census data can be found at https://www.inegi.org.mx/programas/ccpv/2020/.

[18]If the random element of time spent not onsite can be controlled for, Bayesian updating according to test results may be preferred.

[19]As our mixed-integer program formulation is designed to admit integral utilities only, and the problem of computing testing regimes is invariant to scaling utilities, we first scale up the utilities of all individuals in the population by a factor of 50, and then round the resulting number to the nearest integer. Choosing a larger scaling factor increases the running time, as the number of variables in the MILP increases (cf. Section 4.2).

[20]The experiments were run on a 2022 MacBook Pro. Gurobi 9.5.0 was used for the MILP, and MOSEK 10 for the MICP.

| | Approx | | | Greedy | |
|---|---|---|---|---|---|
| Budget | Welfare | Guarantee | Time | Welfare | Time |
| 2 | 461.24 | 0.17865 | 191 ms | 461.23 | 26 ms |
| 4 | 886.08 | 0.35730 | 524 ms | 886.01 | 36 ms |
| 6 | 1292.04 | 0.53594 | 968 ms | 1291.91 | 23 ms |
| 8 | 1684.67 | 0.71459 | 2145 ms | 1684.52 | 26 ms |
| 10 | 2070.82 | 0.89320 | 11515 ms | 2070.58 | 32 ms |
| 12 | 2446.64 | 1.07188 | 8868 ms | 2446.32 | 50 ms |

Table 1. Summary showing welfare and computation time for Approx and Greedy on our pilot study data with a population of $n = 130$ and pool size constraint $G = 5$, with testing budgets $B \in \{2, 4, \dots, 10\}$. We also state the additive approximation guarantee of Approx (compared to optimal non-overlapping welfare).

the testing budget $B$, Greedy scales linearly and is extremely fast (even for larger populations and testing budgets). This makes Greedy attractive for implementations that rely on a quick turnaround, run on 'budget hardware' or wish to avoid costly cloud computing services.

In our synthetic experiments with lower health probabilities, Greedy performs as well as Approx when $G = 5$, and remains competitive also when $G = 10$. Figures 1 and 2 in Appendix I plot the mean welfares achieved by both algorithms for pool size constraints $G = 5$ and $G = 10$, as well as the welfare ratios. For the latter, we divide the welfare achieved by Approx by the welfare of Greedy for each population, and depict the resulting ratios as black dots. In Appendix I, Tables 4 and 5 list the mean welfares and running times of both algorithms, as well as the approximation guarantee of Approx, for $G \in \{5, 10\}$.

Comparing our experiments with different pool sizes $G \in \{5, 10\}$, we see that increasing pool sizes from 5 to 10 significantly increases mean welfare if health probabilities are very high (cf. Tables 1 and 3). This effect is less pronounced in our experiment with synthetic data, in which participants have lower health probabilities on average (cf. Tables 4 and 5). These results suggest that the pool size limit of 5 imposed by saliva sampling, as opposed to the limit of 10 for nasopharyngeal samples, may be considered a limitation in some scenarios, and institutions may wish to weigh the positives and negatives of saliva and nasopharyngeal sampling carefully.

### 5.3 Pilot Study

For our pilot study, we implemented a two-group cluster-randomized design that partitioned the IPICYT participant population into a treatment and a control group, clustered by field and working group.[21] All elements of the trial, including consent and surveys, email invitations for testing, data processing, and computing testing allocations, were coordinated in a web app developed specifically for the trial. A demo of the web app is available at https://ec23demo.pythonanywhere.com.

At the beginning of each week, we computed an optimal testing regime for each day of the week and invited individuals from the treatment group to submit their samples for testing at the LANBAMA testing facility.[22] The treatment group was allowed to freely use the facilities of the university as long as they tested negative, and otherwise were required to work remotely. The control group followed the institutional policy of resuming full access to IPICYT. Participants were

[21]In practice, field and working group are analogous, as only one working group from each participating field volunteered to be a part of the experiment.
[22]LANBAMA has validated and implemented pooled testing with saliva samples for pool sizes up to 5.

assigned to an experimental condition with peers from their working group. Thus, the majority of their social institutional interaction was contained in their experimental treatment arm.[23] We performed daily non-overlapping tests of pooled subjects in the treated population, with a weekly testing budget of 30 tests.

*Our algorithm in practice.* As the Greedy algorithm as implemented in Section 4 demonstrated favorable trade-offs between speed and accuracy (cf. Section 5.2), we implemented a version of Greedy in our web application for computing testing regimes. In our implementation, we allowed individuals to express preferences for two-day windows through allocation of a virtual token budget in the web app. This allowed us to avoid scheduling individuals for testing on days they did not wish to access IPICYT facilities in the first place, and allocate more tests to particular days that were more popular. Moreover, our partner institute has observed — in an independent pool testing trial — that a small fraction of participants invited to pooled testing fail to attend and submit a sample. In order to optimize pooling in this setting, we perform a second optimization round, in which we compute an (approximately) optimal pooling among the samples that have been submitted. It is immediate that the second optimization round cannot decrease the expected welfare achieved.

*Evaluation and methods.* In our trial we measured subjects' stress levels and subjective well-being (life satisfaction), as well as self-assessed performance, productivity, and learning. We assess these measures through survey questions that subjects are invited to answer before (baseline) and after (endline) the testing period. A detailed description is given in Appendix K.3. The treatment effect is estimated with bivariate linear regressions, using the above-mentioned outcomes as dependent variables, and a binary treatment variable, which takes on the value one if the subject is in the treatment group and zero otherwise, as the regressor. We estimate level effects on endline outcomes as well as the effect on the difference in outcomes (delta models) between our two points of measurement before and after the testing period of the trial. We further collected a number of covariates for robustness checks of our estimations.

*Results.* We present the results on performance from the linear models based on Eqs. (15) and (16) in Table 2. Further results on performance and mental health are shown in Tables 9 and 10 in Appendix K.6. The treatment group outperforms control group participants in self-perceived performance, in productivity, and in learning scores.[24] These results are also not statistically significant. Similarly, participants in the treatment group exhibit higher levels of stress and higher levels of subjective well-being (life satisfaction), but the results are not statistically significant.

We also estimate the delta model, where all time-independent confounds disappear. The positive trend in increased stress in the treatment group disappears, while the trend in life satisfaction increases. All treatment effects related to performance, productivity, and learning are corrected downwards, but remain not statistically significant, with the exception of productivity, where we report a small and borderline[25] statistically significant negative effect of our testing regime.[26]

These findings provide evidence that our testing strategy has no negative effect on participants' work/study performance, learning, or mental health, despite the increased effort in coordination

[23]If non-treated participants were to run into treated participants, possible contagions would be contained within our health protocol: non-infectious participants have a 72-hour window during which they are are given access to the building after submitting a sample and receiving a negative qPCR test result.

[24]This also holds for the measures of achieving their own and their supervisors' goals, see Appendix K.6.

[25]The estimated $p$-value is 0.051, exactly on the cutoff of statistical significance. We consider statistical significance for all values $p < 0.05$, but not for values on or above that cutoff [Zhu, 2016].

[26]This may stem from treated individuals having to exert additional effort attending the testing facility and hence experiencing a loss in productivity. Individuals in positive pools who are required to work from home may also face productivity constraints, but during our trial only two individuals tested positive and were identified in their respective pools.

| | Dependent variable | | | | | |
|---|---|---|---|---|---|---|
| | Performance | Productivity | Learning | Δ Performance | Δ Productivity | Δ Learning |
| Treatment vs. Control | 0.120 | 0.076 | 0.175 | −0.053 | $-0.256^{*}$ | 0.086 |
| | (0.143) | (0.133) | (0.287) | (0.143) | (0.130) | (0.323) |
| Constant | $1.984^{***}$ | $2.097^{***}$ | $8.194^{***}$ | 0.000 | 0.081 | 0.177 |
| | (0.086) | (0.082) | (0.206) | (0.110) | (0.090) | (0.209) |
| Observations | 119 | 120 | 119 | 118 | 119 | 119 |
| $R^2$ | 0.006 | 0.003 | 0.003 | 0.001 | 0.032 | 0.001 |
| Adjusted $R^2$ | −0.002 | −0.006 | −0.005 | −0.007 | 0.024 | −0.008 |

$^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

Table 2. Linear model regressions of performance, productivity, and learning outcomes. Note that regression coefficients are expressed in the unit of the score. Standard errors are in parentheses.

it demands from them compared to a full reopening (the regime followed by control group). At the same time, our strategy ensures greater safety of all participating individuals compared to full reopening. We conjecture that accounting for welfare is the crucial ingredient in our mechanism, enabling in-person access for those who need and benefit from it the most.

## 6 DISCUSSION

This work introduces a novel utility-based approach to pooled testing in resource-constrained environments. In this setting, we provide strong theoretical and empirical performance guarantees that further justify the implementation of non-overlapping testing regimes beyond their essential logistical simplicity. We test a version of our utility-based testing strategy in a real-world experiment at the higher education institute IPICYT in Mexico. The results of our randomized control trial provide causal evidence that our testing regime performs no worse than the 'first-best' benchmark of allowing full access for all individuals with respect to the participants' work or study performance, learning, and mental health. The trial data is also used to evaluate our algorithms Greedy and Approx through simulations. These demonstrate that Greedy performs almost optimally and is significantly faster than our alternative MILP implementation.

There are many directions for future work. On a theoretical level, there is a gap between our upper bound of 4 and lower bound of 7/6 on the gain of overlaps. On a more practical level, the overall testing and re-integration policy we propose is static in nature, as we consider the one-shot setting where a testing budget is to be fully utilized by a policymaker. But testing can be dynamic, with allocations chosen adaptively as a function of previous test results, and it is valuable to understand what potential benefits this extended functionality can bring. Additionally, policymakers potentially have access to different types of tests, each with different associated costs and performance (i.e., pool size and sensitivity), and providing optimal budget-constrained allocations in this heterogeneous test setting is a key open question.

Most importantly, although most countries have eased COVID-19 restrictions, we hope that the valuable insight in performance and efficacy of our welfare-maximizing testing regimes can help better protect resource-constrained communities during future outbreaks of infectious diseases.

## REFERENCES

A. Abera, H. Belay, A. Zewude, B. Gidey, D. Nega, B. Dufera, A. Abebe, T. Endriyas, B. Getachew, H. Birhanu, et al. Establishment of COVID-19 testing laboratory in resource-limited settings: challenges and prospects reported from ethiopia. *Global Health Action*, 13(1):1841963, 2020.

S. Ahn, W.-N. Chen, and A. Özgür. Adaptive group testing on networks with community structure. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 1242–1247. IEEE, 2021.

M. Aldridge, O. Johnson, and J. Scarlett. Group Testing: An Information Theory Perspective. *Foundations and Trends® in Communications and Information Theory*, 15(3-4):196–392, 2019. ISSN 1567-2190. doi: 10.1561/0100000099. URL http://dx.doi.org/10.1561/0100000099.

P. D. Allison. Change scores as dependent variables in regression analysis. *Sociological Methodology*, 20:93–114, 1990. ISSN 00811750, 14679531. URL http://www.jstor.org/stable/271083.

I. Asanov, F. Flores, D. McKenzie, M. Mensmann, and M. Schulte. Remote-learning, time-use, and mental health of ecuadorian high-school students during the covid-19 quarantine. *World development*, 138:105225, 2021.

N. Augenblick, J. T. Kolstad, Z. Obermeyer, and A. Wang. Group testing in a pandemic: The role of frequent testing, correlated risk, and machine learning. Working Paper 27457, National Bureau of Economic Research, July 2020. URL http://www.nber.org/papers/w27457.

J. P. Azevedo, A. Hasan, D. Goldemberg, K. Geven, and S. A. Iqbal. Simulating the potential impacts of covid-19 school closures on schooling and learning outcomes: A set of global estimates. *The World Bank Research Observer*, 36(1):1–40, 2021.

O. Bandiera, N. Buehren, M. Goldstein, I. Rasul, and A. Smurra. The economic lives of young women in the time of ebola: lessons from an empowerment program. *World Bank Policy Research Working Paper*, 1(8760), 2019.

L. Becchetti, L. Corrado, and P. Conzo. Sociability, altruism and well-being. *Cambridge Journal of Economics*, 41(2):441–486, 2017.

M. Bertoni, D. Cavapozzi, G. Pasini, and C. Pavese. Remote working and mental health during the first wave of the covid-19 pandemic. Available at SSRN: https://ssrn.com/abstract=4111999 or http://dx.doi.org/10.2139/ssrn.4111999, 2022.

M. J. Binnicker. Challenges and controversies to testing for COVID-19. *Journal of clinical microbiology*, 58(11):e01695–20, 2020.

V. Brault, B. Mallein, and J.-F. Rupprecht. Group testing as a strategy for COVID-19 epidemiological monitoring and community surveillance. *PLoS computational biology*, 17(3):e1008726, 2021.

B. Cleary, J. A. Hay, B. Blumenstiel, M. Harden, M. Cipicchio, J. Bezney, B. Simonton, D. Hong, M. Senghore, A. K. Sesay, et al. Using viral load and epidemic dynamics to optimize pooled testing in resource-constrained settings. *Science translational medicine*, 13(589):eabf1568, 2021.

A. Cohen, N. Shlezinger, A. Solomon, Y. C. Eldar, and M. Médard. Multi-level group testing with application to one-shot pooled COVID-19 tests. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1030–1034. IEEE, 2021.

S. Cohen, T. Kamarck, R. Mermelstein, et al. Perceived stress scale. *Measuring stress: A guide for health and social scientists*, 10(2):1–2, 1994.

M. Cuturi, O. Teboul, Q. Berthet, A. Doucet, and J.-P. Vert. Noisy adaptive group testing using bayesian sequential experimental design. *arXiv preprint arXiv:2004.12508*, 2020.

M. R. De Georgeo, J. M. De Georgeo, T. M. Egan, K. P. Klee, M. S. Schwemm, H. Bye-Kollbaum, and A. J. Kinser. Containing sars-cov-2 in hospitals facing finite ppe, limited testing, and physical space variability: Navigating resource constrained enhanced traffic control bundling. *Journal of Microbiology, Immunology and Infection*, 54(1):4–11, 2021.

G. N. Dhabaan, W. A. Al-Soneidar, and N. N. Al-Hebshi. Challenges to testing COVID-19 in conflict zones: Yemen as an example. *Journal of global health*, 10(1), 2020.

R. Dorfman. The detection of defective members of large populations. *The Annals of Mathematical Statistics*, 14(4):436–440, 1943.

S. Dryden-Peterson, G. E. Velásquez, T. J. Stopka, S. Davey, S. Lockman, and B. O. Ojikutu. Disparities in SARS-CoV-2 testing in massachusetts during the COVID-19 pandemic. *JAMA network open*, 4(2):e2037067–e2037067, 2021.

D. Du, F. K. Hwang, and F. Hwang. *Combinatorial group testing and its applications*, volume 12. World Scientific, 2000.

J. Ely, A. Galeotti, O. Jann, and J. Steiner. Optimal test allocation. *Journal of Economic Theory*, 193:105236, 2021.

J. Emmanuel, M. Bassett, H. Smith, and J. Jacobs. Pooling of sera for human immunodeficiency virus (hiv) testing: an economical method for use in developing countries. *Journal of clinical pathology*, 41(5):582–585, 1988.

S. Ghosh, A. Rajwade, S. Krishna, N. Gopalkrishnan, T. E. Schaus, A. Chakravarthy, S. Varahan, V. Appu, R. Ramakrishnan, S. Ch, et al. Tapestry: A single-round smart pooling technique for covid-19 testing. *MedRxiv*, 2020.

N. Goldberg and G. Rudolf. On the complexity and approximation of the maximum expected value all-or-nothing subset. *Discrete Applied Mathematics*, 283:1–10, 2020.

C. Gollier and O. Gossner. Group testing against Covid-19. Technical report, EconPol Policy Brief, 2020.

K. Gorgen and T. McAleavy. Best practice in pedagogy for remote teaching (p. 10). *EdTech Hub Report*, 2020.

S. Goudeau, C. Sanrey, A. Stanczak, A. Manstead, and C. Darnon. Why lockdown and distance learning during the covid-19 pandemic are likely to increase the social class achievement gap. *Nature Human Behaviour*, 5(10):1273–1281, 2021.

D. Hong, R. Dey, X. Lin, B. Cleary, and E. Dobriban. Group testing via hypergraph factorization applied to COVID-19. *Nature Communications*, 13(1):1–13, 2022.

M. Hossain. Unequal experience of covid-induced remote schooling in four developing countries. *International Journal of Educational Development*, 85:102446, 2021.

F. K. Hwang. A method for detecting all defective members in a population by group testing. *Journal of the American Statistical Association*, 67(339):605–608, 1972.

J. Jonnerby, P. Lazos, E. Lock, F. Marmolejo-Cossío, C. B. Ramsey, M. Shukla, and D. Sridhar. Maximising the Benefits of an Acutely Limited Number of COVID-19 Tests. *ArXiv (preprint)*, pages 1–12, 2020a. URL http://arxiv.org/abs/2004.13650.

J. Jonnerby, P. Lazos, E. Lock, F. Marmolejo-Cossío, C. B. Ramsey, and D. Sridhar. Test and Contain: A Resource-Optimal Testing Strategy for COVID-19. *AI for Social Good Workshop*, 1(July), 2020b.

M. M. Kavanagh, N. A. Erondu, O. Tomori, V. J. Dzau, E. A. Okiro, A. Maleche, I. C. Aniebo, U. Rugege, C. B. Holmes, and L. O. Gostin. Access to lifesaving medical resources for african countries: COVID-19 testing and response, ethics, and politics. *The Lancet*, 395(10238):1735–1738, 2020.

M. M. King and M. E. Frederickson. The pandemic penalty: The gendered effects of covid-19 on scientific productivity. *Socius*, 7:23780231211006977, 2021.

T. K. Koo and M. Y. Li. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, 15(2):155–163, 2016.

D. Lakens. Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social psychological and personality science*, 8(4):355–362, 2017.

D. Lakens. The practical alternative to the p value is the correctly used p value. *Perspectives on psychological science*, 16(3): 639–648, 2021.

D. B. Larremore, B. Wilder, E. Lester, S. Shehata, J. M. Burke, J. A. Hay, M. Tambe, M. J. Mina, and R. Parker. Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening. *Science advances*, 7(1):eabd5393, 2021.

C. H. Li. A sequential method for screening experimental variables. *Journal of the American Statistical Association*, 57(298): 455–477, 1962.

A. K. Limón-Vázquez, G. Guillén-Ruiz, and E. V. Herrera-Huerta. The social isolation triggered by covid-19: Effects on mental health and education in mexico. *Health and Academic Achievement-New Findings*, 2020.

E. Lock, F. J. Marmolejo-Cossío, J. Jonnerby, N. Rajgopal, H. A. Guzmán-Gutiérrez, L. A. Benavides-Vázquez, J. R. Tello-Ayala, and P. Lazos. Optimal testing and containment strategies for universities in mexico amid covid-19. In *Equity and Access in Algorithms, Mechanisms, and Optimization*, pages 1–9. Proceedings of EAAMO'21, 2021.

R. J. Martinez Arriaga, L. P. Gonzalez Ramirez, J. M. de la Roca-Chiapas, and M. Hernández-González. Psychological distress of covid-19 pandemic and associated psychosocial factors among mexican students: An exploratory study. *Psychology in the Schools*, 58(9):1844–1857, 2021.

L. Mutesa, P. Ndishimye, Y. Butera, J. Souopgui, A. Uwineza, R. Rutayisire, E. L. Ndoricimpaye, E. Musoni, N. Rujeni, T. Nyatanyi, et al. A pooled testing strategy for identifying sars-cov-2 at low prevalence. *Nature*, 589(7841):276–280, 2021.

O. U. Nalbantoglu. Group testing performance evaluation for sars-cov-2 massive scale screening and testing. *BMC Medical Research Methodology*, 20(1):1–11, 2020.

P. Nikolopoulos, T. Guo, S. R. Srinivasavaradhan, C. Fragouli, and S. Diggavi. Community aware group testing. *arXiv preprint arXiv:2007.08111*, 2020.

OECD. *OECD guidelines on measuring subjective well-being*. OECD, 2013.

G. Ordóñez Barba. Discrimination, poverty and vulnerability: the intricacies of social inequality in mexico. *Región y sociedad*, 30(71):0–0, 2018.

H. B. Petersen, B. Bah, and P. Jung. Practical high-throughput, non-adaptive and noise-robust sars-cov-2 testing. *arXiv preprint arXiv:2007.09171*, 2020.

H. R. Sanghani, D. A. Nawrot, F. Marmolejo-Cossío, J. M. Taylor, J. Craft, E. Kalimeris, M. I. Andersson, and S. R. Vasudevan. Concentrating Pooled COVID-19 Patient Lysates to Improve Reverse Transcription Quantitative PCR Sensitivity and Efficiency. *Clinical Chemistry*, 67(5):797–798, 2021. ISSN 0009-9147. doi: 10.1093/clinchem/hvab035.

B. Shan, X. Liu, A. Gu, and R. Zhao. The effect of occupational health risk perception on job satisfaction. *International Journal of Environmental Research and Public Health*, 19(4):2111, 2022.

X. M. Tu, E. Litvak, and M. Pagano. On the informativeness and accuracy of pooled testing in estimating prevalence of a rare disease: application to hiv screening. *Biometrika*, 82(2):287–297, 1995.

M. Versteeg and P. Steendijk. Putting post-decision wagering to the test: a measure of self-perceived knowledge in basic sciences? *Perspectives on Medical Education*, 8(1):9–16, 2019.

L. M. Wein and S. A. Zenios. Pooled testing for hiv screening: capturing the dilution effect. *Operations Research*, 44(4): 543–569, 1996.

W. Zhu. p< 0.05,< 0.01,< 0.001,< 0.0001,< 0.00001,< 0.000001, or< 0.0000001.... *Journal of sport and health science*, 5(1):77, 2016.

## A SUPPLEMENTARY LITERATURE ON POOLED TESTING

Pooled testing dates back to the seminal work of Dorfman [1943] who sought to facilitate syphilis diagnostics during World War II. Dorfman's protocol proceeds in two stages: The first stage tests individuals in disjoint groups of a fixed size. Given the results of the first stage, the second stage involves individually testing individuals from positive groups to precisely find who is infected in the population. Theoretical guarantees are provided when the number of individuals who are infected in the population is known beforehand, which permits computing an optimal pool size in the first stage of the protocol in such a way as to minimize the overall number of tests which are needed to precisely ascertain the health status of all individuals in a population. Pooled testing has since become a mature field in its own right with a rich literature of protocols aimed at solving the same objective: *precisely ascertaining the infection status of all individuals in a population with the minimum number of tests.*

In this vein, pooled testing protocols are typically categorized with respect to two axes: assumptions on infection rates, and whether protocol allows for adaptive testing allocations. For the former, two leading regimes are that of combinatorial pooled testing, where a fixed number of infections are known to exist within the population, however the identity of those infected is unknown, and probabilistic pooled testing, where infections occur according to a well-defined probability distribution. With respect to the latter axis, testing regimes can either be adaptive, where they occur in rounds and the allocation of a given round can depend on the results of tests from previous rounds, or non-adaptive, where all tests are allocated at once. For example, Dorfman's protocol operates in the combinatorial adaptive regime, and it has since been significantly improved starting with the $s$-stage algorithm of Li [1962] and continuing with the asymptotically optimal generalized binary splitting approach of Hwang [1972]. More recent theoretical results include: adaptive methods for combinatorial testing which make use of hypergraph factorization in early stages of testing [Hong et al., 2022], using compressed sensing for non-adaptive combinatorial pooled testing [Cohen et al., 2021, Ghosh et al., 2020, Petersen et al., 2020], pooled testing under network-based (non-i.i.d.) infection models [Ahn et al., 2021, Nikolopoulos et al., 2020], Bayesian infection inference in the adaptive noisy test result regime [Cuturi et al., 2020]. For a general references to pooled testing, we refer the reader to Du et al. [2000] as well as Aldridge et al. [2019] for an information-theoretic focus on the subject.

Beyond theory, pooled testing has been applied to combat various diseases in the past, especially HIV/AIDS [Emmanuel et al., 1988, Tu et al., 1995, Wein and Zenios, 1996]. From the outset of the COVID-19 pandemic it became clear that testing resource constraints would be a large issue for multiple countries, and hence pooled testing became a viable option for combatting the virus, especially as it was shown that qPCR tests can be sensitive enough to pool samples in a pooled test [Mutesa et al., 2021, Nalbantoglu, 2020, Sanghani et al., 2021]. Although having access to a pooled testing primitive is a necessary condition for implementing existing protocols, practical constraints often render these approaches unfeasible. On one hand, complicated pooled testing regimes can be difficult to implement logistically at scale with limited laboratory personnel and workflow infrastructure [Cleary et al., 2021]. On the other hand, the unfortunate reality is that many resource-constrained populations are in a situation where their testing budget falls far below the information theoretic lower bounds required to precisely ascertain the health profile of all individuals as per traditional pooled testing objectives.

## B PROOF OF PROPOSITION 2

Proof. Suppose that given an instance, the optimal testing regime is $T^*$, where $t_1^* \cap t_2^* \neq \emptyset$. Let $A = t_1^* \setminus t_2^*$, $B = t_2^* \setminus t_1^*$ and $C = t_1^* \cap t_2^*$. Note that $t_1^* \cup t_2^* = A \cup B \cup C$. Without loss of generality, we assume that $q_A \geq q_B$ and with a slight abuse of notation, we denote $u_A = \sum_{i \in A} u_i$, $u_B = \sum_{i \in B} u_i$, and $u_C = \sum_{i \in C} u_i$. Moreover, we define the following four different testing regimes:

- $T^1$ with $t_1^1 = A \cup C$ and $t_2^1 = B$
- $T^2$ with $t_1^2 = A$ and $t_2^2 = C$
- $T^3$ with $t_1^3 = B$ and $t_2^3 = C$
- $T^4$ with $t_1^4 = A \cup B$ and $t_2^4 = C$

We start with the following necessary lemma.

Lemma 4. *For any $\hat{T} \in \{T^1, T^2, T^3, T^4\}$, the ratio $u(T^*)/u(\hat{T})$ is maximized when $q_C = 1$.*

Proof. We first show that the statement is true for $\hat{T} = T^1$. We need to show that

$$\frac{q_C \left(q_A \cdot u_A + q_B \cdot u_B + (q_A + (1 - q_A) \cdot q_B) \cdot u_C\right)}{q_C \cdot q_A \cdot u_A + q_B \cdot u_B + q_C \cdot q_A \cdot u_C}$$
$$\leq \frac{q_A \cdot u_A + q_B \cdot u_B + (q_A + (1 - q_A) \cdot q_B) \cdot u_C}{q_A \cdot u_A + q_B \cdot u_B + q_A \cdot u_C}$$
$$\Rightarrow q_C \leq \frac{q_C \cdot q_A \cdot u_A + q_B \cdot u_B + q_C \cdot q_A \cdot u_C}{q_A \cdot u_A + q_B \cdot u_B + q_A \cdot u_C}$$

which is true since

$$q_C = q_C \cdot \frac{q_A \cdot u_A + q_B \cdot u_B + q_A \cdot u_C}{q_A \cdot u_A + q_B \cdot u_B + q_A \cdot u_C} \leq \frac{q_C \cdot q_A \cdot u_A + q_B \cdot u_B + q_C \cdot q_A \cdot u_C}{q_A \cdot u_A + q_B \cdot u_B + q_A \cdot u_C}.$$

Now, we consider the case that $\hat{T} = T^2$. Here, we need to show that

$$\frac{q_C \left(q_A \cdot u_A + q_B \cdot u_B + (q_A + (1 - q_A) \cdot q_B) \cdot u_C\right)}{q_A \cdot u_A + q_C \cdot u_C}$$
$$\leq \frac{q_A \cdot u_A + q_B \cdot u_B + (q_A + (1 - q_A) \cdot q_B) \cdot u_C}{q_A \cdot u_A + u_C}$$
$$\Rightarrow q_C \leq \frac{q_A \cdot u_A + q_C \cdot u_C}{q_A \cdot u_A + u_C},$$

which is true since

$$q_C = q_C \cdot \frac{q_A \cdot u_A + u_C}{q_A \cdot u_A + u_C} \leq \frac{q_A \cdot u_A + q_C \cdot u_C}{q_A \cdot u_A + u_C}.$$

With similar arguments as above, we can show that the ratio $u(T^*)/u(T^3)$ is maximized when $q_C = 1$.

Lastly for $\hat{T} = T^4$, we need to show that

$$\frac{q_C \left(q_A \cdot u_A + q_B \cdot u_B + (q_A + (1 - q_A) \cdot q_B) \cdot u_C\right)}{q_A \cdot q_B \cdot (u_A + u_B) + q_C \cdot u_C}$$
$$\leq \frac{q_A \cdot u_A + q_B \cdot u_B + (q_A + (1 - q_A) \cdot q_B) \cdot u_C}{q_A \cdot q_B \cdot (u_A + u_B) + u_C}$$
$$\Rightarrow q_C \leq \frac{q_A \cdot q_B \cdot (u_A + u_B) + q_C \cdot u_C}{q_A \cdot q_B \cdot (u_A + u_B) + u_C},$$

which is true since

$$q_C = q_C \cdot \frac{q_A \cdot q_B \cdot (u_A + u_B) + u_C}{q_A \cdot q_B \cdot (u_A + u_B) + u_C} \leq \frac{q_A \cdot q_B \cdot (u_A + u_B) + q_C \cdot u_C}{q_A \cdot q_B \cdot (u_A + u_B) + u_C}.$$

□

Using Lemma 4, hereinafter, we consider the case that $q_C = 1$.

We distinguish into two cases.

*Case I:* $q_A \geq 5/6$. In this case, note that

$$\frac{(q_A + (1-q_A)\cdot q_B)\cdot u_C}{q_A \cdot u_C} \leq \frac{(q_A + (1-q_A)\cdot q_A)\cdot u_C}{q_A \cdot u_C} = 2 - q_A \leq \frac{7}{6}.$$

where the second transition follows since $q_A \geq q_B$ and the last transition follows since $q_A \geq 5/6$. Hence, we see that

$$\frac{q_A \cdot u_A + q_B \cdot u_B + (q_A + (1-q_A)\cdot q_A)\cdot u_C}{q_A \cdot u_A + q_B \cdot u_B + q_A \cdot u_C} \leq \frac{7}{6},$$

meaning that $u(T^*)/u(T^1) \leq 7/6$.

*Case II:* $q_A < 5/6$. Here, for the sake of contradiction, suppose that for any testing without overlaps the approximation ratio is more that $7/6$. Then, we have that $u(T^*)/u(\hat{T}) > 7/6$, for any $\hat{T} \in \{T^1, T^2, T^3, T^4\}$.

Thus,

$$\begin{aligned}
&\frac{u(T^*)}{u(T^1)} > \frac{7}{6}\\
\Rightarrow &\frac{q_A \cdot u_A + q_B \cdot u_B + (q_A + (1-q_A)\cdot q_B)\cdot u_C}{q_A \cdot u_A + q_B \cdot u_B + q_A \cdot u_C} > \frac{7}{6}\\
\Rightarrow &(6\cdot(q_A + (1-q_A)\cdot q_B) - 7q_A)\cdot u_C > q_A \cdot u_A + q_B \cdot u_B.
\end{aligned} \tag{3}$$

Now, from the fact that $u(T^*)/u(T^2) > 7/6$, we get that

$$\begin{aligned}
\Rightarrow &\frac{q_A \cdot u_A + q_B \cdot u_B + (q_A + (1-q_A)\cdot q_B)\cdot u_C}{q_A \cdot u_A + u_C} > \frac{7}{6}\\
\Rightarrow &q_B \cdot u_B > \frac{1}{6} q_A \cdot u_A + (\frac{7}{6} - (q_A + (1-q_A)\cdot q_B))\cdot u_C\\
\Rightarrow &q_B \cdot u_B > \frac{1}{6} q_A \cdot u_A + (\frac{7}{6} - (q_A + (1-q_A)\cdot q_A))\cdot u_C
\end{aligned} \tag{4}$$

where the last inequality follows from the fact that $q_A \geq q_B$. With a very similar argument we can conclude that

$$q_A \cdot u_A > \frac{1}{6} q_B \cdot u_B + (\frac{7}{6} - (q_A + (1-q_A)\cdot q_A))\cdot u_C \tag{5}$$

from the fact that $u(T^*)/u(T^3) > 7/6$. From Equation (4) and Equation (5), we get that

$$q_A \cdot u_A + q_B \cdot u_B > \frac{6}{5}\cdot 2 \cdot (\frac{7}{6} - (q_A + (1-q_A)\cdot q_A))\cdot u_C$$

and from Equation (3), we conclude that

$$(6\cdot(q_A + (1-q_A)\cdot q_B) - 7q_A) > \frac{6}{5}\cdot 2\cdot(\frac{7}{6} - (q_A + (1-q_A)\cdot q_A))$$

which is true when $1/2 < q_A < 2/3$. Hence, from now one we assume that $q_A$ lies in the interval $(1/2, 2/3)$.

Moreover, we have that

$$
\begin{aligned}
&\frac{u(T^*)}{u(T^2)} > \frac{7}{6} \\
\Rightarrow &\frac{q_A \cdot u_A + q_B \cdot u_B + (q_A + (1-q_A) \cdot q_B) \cdot u_C}{q_A \cdot u_A + u_C} > \frac{7}{6} \\
\Rightarrow &\frac{(6 \cdot (q_A + (1-q_A) \cdot q_B) - 7q_A) \cdot u_C + (q_A + (1-q_A) \cdot q_B) \cdot u_C}{q_A \cdot u_A + u_C} > \frac{7}{6} \\
\Rightarrow &\frac{7 \cdot (1-q_A) \cdot q_B \cdot u_C}{q_A \cdot u_A + u_C} > \frac{7}{6} \\
\Rightarrow &(6 \cdot (1-q_A) \cdot q_B - 1) \cdot u_C > q_A \cdot u_A \qquad (6)
\end{aligned}
$$

where the third inequality follows from Equation (3), and similarly using the fact that $u(T^*)/u(T^3)$, we get

$$(6 \cdot (1-q_A) \cdot q_B - 1) \cdot u_C > q_B \cdot u_B \qquad (7)$$

Lastly,

$$
\begin{aligned}
&\frac{u(T^*)}{u(T^4)} > \frac{7}{6} \\
\Rightarrow &\frac{q_A \cdot u_A + q_B \cdot u_B + (q_A + (1-q_A) \cdot q_B) \cdot u_C}{q_A \cdot q_B \cdot (u_A + u_B) + u_C} > \frac{7}{6} \\
\Rightarrow &q_A \cdot u_A + q_B \cdot u_B > \frac{(7 - 6(q_A + (1-q_A) \cdot q_B)) \cdot u_C}{6 - 7 \cdot q_B} \qquad (8)
\end{aligned}
$$

where the last inequality follows from the fact that $q_B < 6/7$ since $q_B \leq q_A$ and $q_A < 5/6$.

Now, from Equation (6), Equation (7) and Equation (8), we get that

$$
\begin{aligned}
\frac{u(T^*)}{u(T^1)} &= \frac{q_A \cdot u_A + q_B \cdot u_B + (q_A + (1-q_A) \cdot q_B) \cdot u_C}{q_A \cdot u_A + q_B \cdot u_B + q_A \cdot u_C} \\
&\leq \frac{2 \cdot (6 \cdot (1-q_A) \cdot q_B - 1) \cdot u_C + (q_A + (1-q_A) \cdot q_B) \cdot u_C}{\frac{7 - 6(q_A + (1-q_A) \cdot q_B) \cdot u_C}{6 - 7 \cdot q_B} + q_A \cdot u_C}
\end{aligned}
$$

which is maximized when $q_A = q_B = 1/2$, given that $1/2 < q_A < 2/3$ and then we get that $u(T^*)/u(T^1) \leq 7/6$ and reach a contradiction. □

## C PROOF OF Lemma 1

Proof. We re-write $P_{i,j}^{T^*}$ for $i \in t_j^*$ using conditional probabilities:

$$
\begin{aligned}
P_{i,j}^{T^*} &= \Pr[\ \forall t_{j'}^* \in T^*(i;j), t_{j'}^* \text{ is positive and } t_j^* \text{ is negative}]. \\
&= \Pr[t_j^* \text{ negative}] \cdot \Pr[\forall t_{j'}^* \in T^*(i;j), t_{j'}^* \text{ positive} \mid t_j^* \text{ negative}] \\
&= q_{t_j^*} \cdot \Pr[\forall t_{j'}^* \in T^*(i;j)\ \exists i' \in t_{j'}^* \setminus t_j^* \mid i \text{ infected}].
\end{aligned}
$$

First of all, we notice that it must be the case that $q_{t_j^*} = \prod_{i \in t_j^*} q_i > 0$. If this is not so then there must be some individual $i \in t_j^*$ such that $p_i = 1$, and it is straightforward to see that it is always sub-optimal to include such an individual in any testing regime. As for the second term, to show that it is non-zero, we begin by using the fact that $T^*$ is optimal to show that without loss of generality, we can assume that for each $t_{j'}^* \in T^*(i;j)$ it must be the case that there exists an $i \in t_{j'}^* \setminus t_j^*$ such that $p_i > 0$.

Suppose that this is not the case and that there exist $t^*_{j'}, t^*_j \in T^*$ such that every $i \in t^*_{j'} \setminus t^*_j$ has $p_i = 0$ ($q_i = 1$). We show that either $T^*$ is sub-optimal, or we can construct an optimal testing regime where this is no longer the case. To do so, we assume that without loss of generality $j' = 1$ and $j = 2$ (we can arbitrarily re-order test indices), and write the expected utility of $T^*$ as follows:

$$
\begin{aligned}
u(T^*) &= \sum_{i\in[n]} u_i \cdot P_i^{T^*} \\
&= \sum_{i\in[n]} \sum_{j\in[B]} u_i \cdot P_{i,j}^{T^*} \\
&= \sum_{j\in[B]} \sum_{i\in t_j^*} u_i \cdot P_{i,j}^{T^*} \\
&= \left(\sum_{i\in t_1^*} u_i \cdot P_{i,1}^{T^*}\right) + \left(\sum_{i\in t_2^*} u_i \cdot P_{i,2}^{T^*}\right) + \sum_{j=3}^{B} \sum_{i\in t_j^*} u_i \cdot P_{i,j}^{T^*} \\
&= u(t_1^*) + \left(\sum_{i\in t_1^*} u_i \cdot P_{i,2}^{T^*} + \sum_{i\in t_2^*\setminus t_1^*} u_i \cdot P_{i,2}^{T^*}\right) + \sum_{j=3}^{B} \sum_{i\in t_j^*} u_i \cdot P_{i,j}^{T^*} \\
&= u(t_1^*) + \left(\sum_{i\in t_2^*\setminus t_1^*} u_i \cdot P_{i,2}^{T^*}\right) + \sum_{j=3}^{B} \sum_{i\in t_j^*} u_i \cdot P_{i,j}^{T^*} \\
&= u(t_1^*) + q_{t_2^*} \sum_{i\in t_2^*\setminus t_1^*} u_i + \sum_{j=3}^{B} \sum_{i\in t_j^*} u_i \cdot P_{i,j}^{T^*} \\
&\leq u(t_1^*) + u(t_2^* \setminus t_1^*) + \sum_{j=3}^{B} \sum_{i\in t_j^*} u_i \cdot P_{i,j}^{T^*}
\end{aligned}
$$

The first 4 equalities follow from re-ordering terms in the sums. In the fifth equality, we make use of the fact that if $i \in t_1^*$, it must be the case that $P_{i,2}^{T^*} = 0$, for if $t_2^*$ is pivotal for them, then that test must be negative, which implies that $t_1^*$ is negative (for individuals in $t^* \setminus t_2^*$ are guaranteed to be healthy by assumption), contradicting the pivotal nature of $t_2^*$. If $i \in t_2^* \setminus t_1^*$, then $t_2^*$ is pivotal only if it is negative, hence $P_{i,2}^{T^*} = q_{t_2^*}$, which in turn justifies the following equality. Finally, we know that $q_{t_2^*} \leq q_{t_2^*\setminus t_1^*}$, and that $u(t_2^* \setminus t_1^*) = q_{t_2^*\setminus t_1^*} \sum_{i\in t_2^*\setminus t_1^*} u_i$, from which the final equality holds. Putting everything together, let us consider $T'$ where $t'_j = t^*_j$ for $j \neq 2$ and $t'_2 = t^*_2 \setminus t^*_1$. From the above, it follows that either $T^*$ is sub-optimal (as $T'$ achieves more welfare), or we can replace $T^*$ with $T'$, and in either case, we can ensure that our desired property holds.

With this in hand, this means if $T^*$ is optimal, and we consider $i \in t_j^*$ as in the beginning of the proof, we can construct a pool of individuals $S$ by picking one individual from the set $T_j^* \setminus T_{j'}^*$ for each $T_{j'}^* \in T(i;j)$ such that the individual has non-zero probability of infection. From the above, we are guaranteed to be able to construct such an $S$ which is non-empty. Furthermore, if all individuals in $S$ are infected, it follows that each $T_{j'} \in T(i;j)$ is positive without compromising a negative test on $T_j^*$. This in turn implies that $\Pr[\forall t_{j'}^* \in T^*(i;j)\ \exists i' \in t_{j'}^* \setminus t_j^* \mid i \text{ infected}] \geq \Pr[1 - q_S] > 0$. Putting this together with the fact that $q_{t_j^*} > 0$ completes the proof.

□

## D GAIN OF OVERLAPS FOR $B \in \{3, 4\}$

Proposition 4. gain(3) $\leq 7/3$ *and* gain(4) $\leq 15/4$.

Proof. We start from the case that $B = 3$. we partition the individuals, that are pooled into at least one test in an optimal testing regime $T^*$, into seven sets as following: the first three sets, denoted by $S_1$, $S_2$ and $S_3$, consist of individuals that are pooled into only the first, the second and the third test, respectively; the next three sets, denoted by $S_4$, $S_5$ and $S_6$ consist of individuals that are only pooled into the first and the second test, the first and the third test, and the second and the third test, respectively; the last set, denoted by $S_7$, consists of individuals that are included in all three tests. Then, $u(T^*) \leq \sum_{j\in[7]} q_{S_j} \cdot \sum_{i\in S_j} u_i$, as each $i \in S_j$ is always pooled into a test along with all the individuals in $S_j \setminus \{i\}$ and hence her probability to be included in a test that turns negative is at most $q_{S_j}$ which indicates that probability that all the individuals in $S_j$ are healthy. Now, without loss of generality, assume that $q_{S_j} \sum_{i\in S_j} u_i \geq q_{S_{j+1}} \sum_{i\in S_{j+1}} u_i$ for each $j \in [6]$. Then, we define the non-overlapping testing regime $T$ such that $t_1 = S_1$, $t_2 = S_2$ and $t_3 = S_3$. Notice that $u(T) = \sum_{j\in[3]} q_{S_j} \cdot \sum_{i\in S_j} u_i$, and hence $u(T^*)/u(T) \leq 7/3$.

For the case that $B = 4$, we partition the individuals into 15 sets with a very similar way as above, i.e. the first fours sets consist of individuals that are pooled only into one test, the next six tests consist of individuals that are pooled into exactly two tests, the next 4 tests consist of individuals that are pooled into exactly three tests and the last set consists of individuals that are pooled into all four tests. Then, we use the four available tests to pool individuals from the four sets that have the highest utility and we get an approximation of 15/4. □

## E PROOF OF Lemma 3

Proof. Here we show, how the FPTAS, that is introduced in [Goldberg and Rudolf, 2020] and finds an almost optimal test when there is no any size constraint, can be modified for the case that a test can pool up to $G$ samples.

Using similar notation as in [Goldberg and Rudolf, 2020], for $i \in [n]$, we denote with $P(i, C, L)$ the maximum probability of a subset of $[i]$ to be negative with sum of utilities exactly $C$ and size exactly $L$. Then, we modify the dynamic program that was introduced at Equation (6) in [Goldberg and Rudolf, 2020] as following:

$$P(i, C, L) = \begin{cases} \max\{P(i-1, C, L), q_i \cdot P(i-1, C-u_i, L-1)\} & i \geq 2 \text{ and } u_i < C \\ P(i-1, C, L) & i \geq 2 \text{ and } u_i \geq C \\ q_i & i = 1 \text{ and } u_1 = C \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

With a slight abuse of notation, we denote with $t(P(i, C, L))$ the subset of $[i]$ that satisfies $q_{t(P(i,C,L))} = P(i, C, L)$, $\sum_{\ell\in[t(P(i,C,L))]} u_\ell = C$ and $|t(P(i, C, L))| = L$.

If $\bar{C}$ is an upper bound on the sum of utilities of the optimal test ( a straightforward upper bound is $\bar{C} = \sum_{i\in[n]} u_i$), then the optimal test with size at most $G$ that maximizes the utility is given by $t(P(n, C^*, L^*))$ where

$$C^*, L^* = \underset{C\in[\bar{C}], L\in[G]}{\arg\max}\; C \cdot P(n, C, L). \tag{10}$$

Thus, the running time is given by $O(nG\bar{C}) \leq O(n^2\bar{C})$, since $G \leq n$. Then, we see that in order to approximately solve the dynamic programming with a polynomial run-time complexity bound, we should scale down (and round) the utility coefficients whose magnitude determines the running time of the program. We can achieve this by using identical arguments as in Section 3.2 of [Goldberg and Rudolf, 2020]. We present the whole proof here for completeness.

**Algorithm 1**

$\kappa \leftarrow (\epsilon \cdot 1/2 \cdot \max_{i \in N_{1/2}} u_i)/n$
$z^* \leftarrow 0; t \leftarrow \emptyset$
**for** $j = h+1, \dots n$ **do**
  **if** $\hat{z}(h, j) < q_j \cdot u_j$ **then**
    **if** $q_j \cdot u_j > z^*$ **then**
      $z^* \leftarrow q_j \cdot u_j$
      $t \leftarrow \{j\}$
    **end if**
  **else**
    **if** $\hat{z}(h, j) > z^*$ **then**
      $z^* \leftarrow \hat{z}(h, j)$
      $t \leftarrow t(\hat{P}(h, \hat{C}_{i,j}, \hat{L}_{i,j}))$
    **end if**
  **end if**
**end for**
**return** $t$

We scale down the utilities using some factor $\kappa$, by setting $\hat{u}_i = \lfloor u_i/\kappa \rfloor$ for each $i \in [n]$. Before, we choose $\kappa$, we add some more notation. Let $N_{1/2} = \{i \in [n] : q_i \geq 1/2\}$. Without loss of generality, assume that $N_{1/2} = [h]$ and $[n] \setminus N_{1/2} = \{h+1, \dots, n\}$. Let $\hat{P}(i, C, L)$ denote the DP in Equation (9) by replacing $u_i$ with $\hat{u}_i$. Moreover, we assume that there exists a dummy individual $n+1$ with $\hat{u}_{n+1} = 0$ and $q_{n+1} = 1$. Then, for $i \in [n]$ and $j > i$ the scaled DP problem is defined as

$$\hat{z}_\kappa(i, j) = \max_{C \in [\bar{C}(i)], L \in [G]} (\kappa \cdot C + u_j) \cdot \hat{P}(i, C, L) \cdot q_j. \tag{11}$$

where $\bar{C}(i) = \sum_{i' \in [i]} u_{i'}$. Let

$$\hat{C}_{i,j}, \hat{L}_{i,j} = \operatorname*{arg\,max}_{C \in [\bar{C}(i)], L \in [G]} (\kappa \cdot C + u_j) \cdot \hat{P}(i, C, L) \cdot q_j.$$

Note that $t(\hat{P}(i, \hat{C}_{i,j}, \hat{L}_{i,j}) \cup \{j\})$ returns an optimal test by replacing $u_i$ with $\hat{u}_i$ and adding the constraint that for any $\ell \in [i+1, \dots, n] \setminus \{j\}$, $\ell$ is not pooled into the test, while $j$ is pooled into it. From Lemma 2, we know that it suffices to evaluate $\hat{z}(i, j)$ for $i \in [h]$ and $j \in \{h+1, \dots, n\}$ in order to evaluate $\hat{z}_\kappa(n, n+1)$ as at most one individual from $[n] \setminus N_{1/2}$ may be pooled into the test.

The following lemma establishes an upper bound on a value of $\kappa$ that suffices to bound the relative error of solutions of $\hat{z}_\kappa$ in approximating the optimal test within a given $\epsilon > 0$.

Lemma 5. *Let $t^*$ be an optimal test with $\sum_{\ell \in t^*} u_\ell = C^*$ and $|t^*| = L^*$. For a given $\epsilon > 0$, there exist $i \in [h]$, $j \in \{h, \dots, n+1\}$, with $i < j$, and for $\bar{t} = t^* \setminus \{j\}$ and $\kappa \leq \frac{\epsilon \max_{i \in \bar{t}} q_i \cdot u_i}{n}$ such that*

$$\hat{z}_\kappa(i, j) \geq (1 - \epsilon) \cdot C^* \cdot P(n, C^*, L^*).$$

Proof. First note

$$\sum_{i \in t^*} u_i - \kappa \sum_{i \in t^*} \hat{u}_i = \sum_{i \in t^*} u_i - \kappa \sum_{i \in t^*} \lfloor u_i/\kappa \rfloor \leq \sum_{i \in t^*} u_i - \kappa \sum_{i \in t^*} (u_i/\kappa - 1) \leq \kappa n$$

where the last inequality follows since $|t^*| \leq n$.

Let $j$ be the individual with the smallest probability of being healthy in $t^*$ by breaking ties with respect to individuals that have higher index and let $i$ be the individual with the highest index in

$t^* \setminus \{j\}$. We denote with $\hat{t} \subseteq [i]$ the set that maximizes Equation (11). Then, we get that

$$\hat{z}_\kappa(i,j) = q_j\left(\kappa \sum_{\ell \in \hat{t}\setminus\{j\}} \hat{u}_\ell + u_j\right) q_{\hat{t}\setminus\{j\}} = q_j\left(\kappa \sum_{\ell \in t^*\setminus\{j\}} \hat{u}_\ell + u_j\right) q_{t^*\setminus\{j\}}$$
$$\geq \left(1 - \frac{n\kappa}{\sum_{i\in t^*} u_i}\right) \cdot q_{t^*} \sum_{i \in t^*} u_i,$$

where the first inequality follows from optimality of $\hat{t}$ under the scaled utilities. Thus, to ensure an $\epsilon$-approximate solution, we need

$$\frac{n\kappa}{\sum_{\ell\in t^*} u_\ell} \leq \epsilon \Leftrightarrow \kappa \leq \frac{\epsilon \cdot \sum_{\ell \in t^*} u_\ell}{n}.$$

Thus, it suffices to choose

$$\kappa \leq \frac{\epsilon \cdot \max_{\ell\in[t^*\setminus\{j\}]} u_\ell \cdot q_\ell}{n} \leq \frac{\epsilon \cdot q_{t^*} \cdot \sum_{i\in t^*} u_i}{n} \leq \frac{\epsilon \cdot \sum_{\ell\in t^*} u_i}{n}$$

□

Since $t^*$ is not known, we should choose a value for $\kappa$ that satisfies the above lemma. Note that due to Lemma 2, we know that $t^* \cap ([n] \setminus N_{1/2}) \leq 1$.

Algorithm 1, which is an FPTAS for the optimal test, due to Lemma 2, fixes an individual $j \in [n+1] \setminus N_{1/2}$ that is pooled into the optimal test, where $j = n+1$ indicates the case that $t^* \cap ([n] \setminus N_{1/2}) = 0$. Hence, we can apply Lemma 3 by setting $i = h$ and thus $\bar{t} \subseteq N_{1/2}$. Since for each $\ell \in N_{1/2}$, $q_\ell \geq 1/2$, we have that $\max_{\ell\in\bar{t}} q_\ell \cdot u_\ell \geq 1/2 \max_{\ell\in\bar{t}} \cdot u_\ell$. Thus, we can choose $\kappa$ such that

$$\kappa = \frac{\epsilon \cdot 1/2 \cdot \max_{\ell\in[\bar{t}]} u_\ell}{n} \leq \frac{\epsilon \cdot \max_{\ell\in[\bar{t}]} q_\ell \cdot u_\ell}{n} \leq \frac{\epsilon \cdot q_{t^*} \cdot \sum_{i\in t^*} u_i}{n}$$

where the last inequality follows from optimality of $t^*$.

Now, we show that Algorithm 1 is an FPTAS for the optimal test.

First note that if $|t^*| = 1$, then Algorithm 1 finds the optimal test in Lines 5-7. Hence, we focus on the case that $|t^*| > 1$. using Lemma 2, we distinguish into two cases:

*Case I:* $|t^* \setminus N_{1/2}| = 0$. For each given $\epsilon > 0$, $\kappa$ satisfies the supposition of Lemma 5. So following Lemma 5 with $\bar{C} = \bar{C}(h) = \sum_{\ell\in N_{1/2}} \geq \sum_{\ell\in t^*} \hat{u}_\ell$, for $\hat{C}_{h,n+1}$ and $\hat{L}_{h,n+1}$, we have

$$u(t) = \hat{z}_\kappa(h, n+1) = \kappa \cdot \hat{C}_{h,n+1} \cdot \hat{P}(h, \hat{C}_{h,n+1}, \hat{L}_{h,n+1}) \geq (1-\epsilon) \cdot C^* \cdot P(h, C^*, L^*)$$
$$\geq (1-\epsilon) \cdot C^* \cdot P(n, C^*, L^*)$$

*Case II:* $|t^* \setminus N_{1/2}| = 1$. Then, for each $\epsilon > 0$, the choice of $\kappa$ for some $j \in [n] \setminus N_{1/2}$ satisfies

$$\hat{z}_\kappa(h,j) \geq (1-\epsilon) \cdot C^* \cdot P(n, C^*, L^*)$$

where the inequality follows form Lemma 5. The algorithm must determine $j$ since it enumerates all elements of $[n] \setminus N_{1/2}$. in the main loop.

The complexity of the algorithm is determined by at most $n$ evaluations of Equation (11). Hence,

$$O\left(n^3 \bar{C}\right) \subseteq O\left(n^3 \sum_{\ell\in N_{1/2}} \frac{u_\ell}{\kappa}\right) \subseteq O\left(\frac{n^5}{\epsilon}\right).$$

□

## F PROOF OF Proposition 3

We begin by providing a simple upper bound on the attainable welfare for any testing regime $T \in \mathcal{T}^B$.

Lemma 6. *Suppose that $(t^*) \in \mathcal{T}^1$ is optimal. For any $B \leq 1$. If $T \in \mathcal{T}^B$, then it follows that $u(T) \leq B \cdot u(t^*)$.*

Proof. Suppose that $T \in \mathcal{T}^B$. As before, we write its welfare as follows:

$$
\begin{aligned}
u(T^*) &= \sum_{i \in [n]} u_i \cdot P_i^T \\
&= \sum_{i \in [n]} \sum_{j \in [B]} u_i \cdot P_{i,j}^T \\
&= \sum_{j \in [B]} \sum_{i \in [n]} u_i \cdot P_{i,j}^T \\
&= \sum_{j \in [B]} \sum_{i \in t_j} u_i \cdot P_{i,j}^T \\
&\leq \sum_{j \in [B]} q_{t_j} \sum_{i \in [n]} u_i \\
&= \sum_{j \in [B]} u(t_j) \\
&\leq \sum_{j \in [B]} u(t^*) \\
&= B \cdot u(t^*)
\end{aligned}
$$

The initial equalities in the above equations arise from re-arranging the sum and from the fact that $P_{i,j}^T = 0$ when $i \notin t_j$. As for the first inequality, We can express and bound the probability that $t_j$ is pivotal for $i$ in $T$:

$$P_{i,j}^T = q_{t_j} \prod_{t_\ell \in T(i:j)} (1 - q_{t_\ell \setminus t_j}) \leq q_{t_j}.$$

Finally, due to optimality of $t^*$, it follows that $u(t) \leq u(t^*)$ for any feasible pooled test $t \subseteq [n]$. This finishes the proof of the claim. □

Suppose that we consider a population instance where the population to be tested, $[n]$ can be partitioned into $C$ clusters, such that all individuals in the $i$-th cluster have utility given by $u_i$ and probability of infection given by $p_i$. In addition, suppose that the $i$-th cluster contains $n_i > 0$ individuals. Using the above lemma, we can show that if the population sizes of each cluster permit repeating identical copies of the optimal test in $\mathcal{T}^1$, then this is optimal. Notice that this is what result of executing the greedy algorithm with $\epsilon = 0$ (which can be done efficiently by brute force if the number of clusters is not too large).

Proposition 5. *Suppose that $\{t^*\} \in \mathcal{T}^1$ is optimal and that in addition, $B \cdot |t^* \cap C_i| \leq n_i$ for each cluster. Let $T^* \in \mathcal{T}^B$ be a testing regime that simply repeats $t^*$ in disjoint copies $B$ times. It follows that $T^*$ is optimal.*

## G IDENTICAL UTILITIES

In this section, we consider the special case where $u_i = u_{i'}$ for each $i, i' \in [n]$. Without loss of generality, assume that $u_i = 1$ for any $i \in [n]$ and $q_i \geq q_{i+1}$ for any $i \in [n-1]$.

### G.1 Optimal Testing Regime for constant $B$

We start by showing that when $B$ is a constant, we can find the optimal non-overlapping testing regime in polynomial time.

Theorem 3. *When the individuals have identical utilities, we can find an optimal non-overlapping testing regime $T$ in time $O(n^{B+1}/B!)$.*

Proof. We start with the following crucial lemma which indicates that there exists an optimal testing regime where the test that has the largest size pools samples of the first $k_1$ individuals with $k_1 \in [n]$, the test that has the second largest size pools samples of the individuals $k_1 + 1$ to $k_2$ with $k_2 \in \{k_1 + 1, \ldots, n\}$, the test that has the third largest size pools samples of the individuals $k_2 + 1$ to $k_3$ with $k_3 \in \{k_2 + 1, \ldots, n\}$ and so on.

Lemma 7. *Let $T^*$ be an optimal testing regime and without loss of generality let $|t^*_j| \geq |t^*_{j+1}|$ for any $j \in [B-1]$. Then, there exists an optimal testing regime $T'$ such that for each $j \in [B]$, $|t'_j| = |t^*_j|$ and*

$$t'_j = \{\sum_{j' \in [j-1]} |t'_{j'}| + 1, \ldots, \sum_{j' \in [j-1]} |t'_{j'}| + |t'_j|\}.$$

Proof. First, note that for some optimal testing regime, it should hold that for any $i' > i$, if $i'$ is pooled into some test, then $i$ is also pooled into some test, as otherwise the replacement of $i'$ with $i$ cannot worse the expected welfare of the testing regime. Thus, hereinafter, we focus on optimal testing regimes that pool samples of the first $k$ individuals for some $k \in [n]$.

We prove the lemma by induction on the number of tests. Start from the case that $B = 2$. Let $T^*$ be an optimal testing regime with $|t^*_1| \geq |t^*_2|$. Assume that $t^*_1 = S_1 \cup S'_1$, where $S_1 \subset \{1, \ldots, |t^*_1|\}$, and $S'_1 \subset \{|t^*_1| + 1, \ldots, k\}$ and $t^*_2 = S_2 \cup S'_2$, where $S_2 \subset \{1, \ldots, |t^*_1|\}$ and $S'_2 \subset \{|t^*_1| + 1, \ldots, k\}$. Since $t^*_1 \cup t^*_2 = [k]$, we get that $|S'_1| = |S_2|$ and since $|t^*_1| \geq |t^*_2|$, we get that $|S_1| \geq |S'_2|$. Now, consider the testing regime $T$ such that $t_1 = S_1 \cup S_2$ and $t_2 = S'_1 \cup S'_2$. Notice that $t_1 \cup t_2 = [k]$, $|t_1| = |t^*_1|$ and $|t_2| = |t^*_2|$. Then, we have

$$u(T^*) = q_{S_1} \cdot q_{S'_1} \cdot |t^*_1| + q_{S_2} \cdot q_{S'_2} \cdot |t^*_2|$$

and

$$u(T) = q_{S_1} \cdot q_{S_2} \cdot |t^*_1| + q_{S'_1} \cdot q_{S'_2} \cdot |t^*_2|.$$

and hence,

$$u(T) - u(T^*) = \left(q_{S_2} - q_{S'_1}\right) \cdot \left(q_{S_1} \cdot |t^*_1| - q_{S'_2} \cdot |t^*_2|\right). \tag{12}$$

Due to optimality of $T^*$, we have that for any $\hat{S}_1 \subseteq S_1$

$$q_{S_1} \cdot q_{S'_1} \cdot |t^*_1| \geq q_{\hat{S}_1} \cdot q_{S'_1} \cdot (|\hat{S}_1| + |S'_1|)$$

as otherwise if $T'$ is a testing regime with $t'_1 = \hat{S}_1 \cup S'_1$ and $t'_2 = t^*_2$, then it would hold that $u(T') > u(T^*)$ which is a contradiction. Now, choose arbitrary $\hat{S}_1 \subseteq S_1$ such that $|\hat{S}_1| = |S'_2|$. We know that this is feasible since $|S_1| \geq |S'_2|$. Then, we have

$$q_{S_1} \cdot q_{S'_1} \cdot |t^*_1| \geq q_{\hat{S}_1} \cdot q_{S'_1} \cdot (|S'_2| + |S_2|) \geq q_{S'_2} \cdot q_{S'_1} \cdot |t^*_2|$$

where the second transition follows due to optimally of $T^*$, and the facts that $|\hat{S}_1| = |S_2'|$ and $|S_1'| = |S_2|$ and the third transition follows since $|S_2'| + |S_2| = |t_2^*|$ and $q_{\hat{S}_1} \geq q_{S_2'}$ as for each $i \in \hat{S}_1$ and each $i' \in S_2'$ it holds that $q_i \geq q_{i'}$. Hence, we have that

$$q_{S_1} \cdot |t_1^*| \geq q_{S_2'} \cdot |t_2^*|.$$

Now from Equation (12), we have that $u(T) \geq u(T^*)$ since $\left(q_{S_2} - q_{S_1'}\right) \geq 0$ as for each $i \in S_2$ and each $i' \in S_1'$ it holds that $q_i \geq q_{i'}$ and $|S_1'| = |S_2|$. Thus, we conclude in a testing regime $T$, with $|t_1| \geq |t_2|$, $T_1 = \{1 \ldots, |t_1|\}$ and $T_2 = \{|t_1| + 1, \ldots, |t_2|\}$ that is optimal.

Now, suppose that the claim holds for $B - 1$. We will show that it holds for $B$. Let $T^*$ be an optimal testing regime with $|t_j^*| \geq |t_{j+1}^*|$ for any $j \in [B]$. Using the induction hypothesis, we can construct an optimal testing regime $T'$ such that $|t_j'| = |t_j^*|$ for any $j \in [B-1]$, $t_B' = t_B^*$ and there are no $i \in t_j'$ and $i' \in t_{j'}'$ with $i' < i$ and $j' > j$. Then, in round $j$, for any $t_j'$ and $t_B^*$, from induction base we construct $t_j''$ and $t_B^j$ such that $|t_j''| = |t_j'| = |t_j^*|$ and $|t_B^j| = |t_B^*|$ and there are no $i \in t_j''$ and $i' \in t_B^j$ with $i' < i$. Thus, after $n - 1$ rounds, we have $T'' = (t_1'', \ldots, t_{B-1}'', t_B^{n-1})$ which is optimal and satisfies the property of the statement. □

Using Lemma 7, we can find an optimal testing as following. For any $k \in [n]$ and any $k_1 \geq k_2 \ldots \geq k_B$ with $\sum_{\ell \in [B]} k_\ell = k$, calculate the welfare of $T$ such that

$$t_j = \{ \sum_{\ell \in [j-1]} k_\ell + 1, \ldots \sum_{\ell \in [j-1]} k_\ell + k_j \},$$

and return the testing regime that has the highest welfare. Hence, we need time at most $n \cdot n^B / B!$ to find the optimal testing regime as for each $k$, each $k_\ell$ takes up to $n$ values and for each of the cases, we order the $k_\ell$'s in a decreasing order, meaning that among the $B!$ different ways of ordering them we are interested only for the case that $k_1 \geq k_2 \ldots \geq k_B$. □

## G.2 Greedy Algorithm

Here, we show that when the utilities are identical, we can find an $e$-approximate testing regime with respect to the optimal non-overlapping testing regime, for any value $B$. Specifically, we consider a variation of the greedy algorithm that we introduced in Section 3.3 which we denote as *var-Greedy* and is defined as following: *var-Greedy* runs $B$ rounds, and in each round $j$, includes in test $t_j$ individuals that have not been pooled into any other test yet in an decreasing order with respect to their probability of being healthy until the utility of the test is not worsen. Note that *var-Greedy* always returns a testing regime $T$ where samples of the first $n' \in [n]$ individuals are pooled into some test, i.e. $\cup_{j \in [B]} t_j = [n']$.

We start with the following lemma.

Lemma 8. *If var-Greedy returns a testing regime that pools samples of the first $n'$ individuals, then there exists an optimal testing regime that pools samples of the first $n''$ individuals with $n'' \leq n'$.*

Proof. Let $T^*$ be an optimal non-overlapping testing regime that satisfies the properties of Lemma 7, i.e. for each $j \in B$

$$t_j^* = \{i_{j-1}^* + 1, \ldots, i_j^*\}$$

with $i_0^* = 0$ and $i_{j-1}^* < i_j^*$. We denote with $T$ the testing regime that is returned by *var-Greedy*, where for each $j \in [B]$, $t_j = \{i_{j-1} + 1, \ldots, i_j\}$, with $i_0 = 0$ and $i_{j-1} < i_j$. We show that for each $j \in [B]$, $i_j^* \leq i_j$. Suppose for contradiction that $t_j^*$ is the first test such that $i_j^* > i_j$. Due to the structure of $T^*$ and $T$, this means that $\cup_{j' \in [j-1]} t^*_{j'} \subseteq \cup_{j' \in [j-1]} t_{j'}$, and hence $i^*_{j-1} \leq i_{j-1}$. Given that *var-Greedy* did not pool $i_j + 1$ in $t_j$, we have that

$$q_{i_{j-1}+1} \cdot \ldots \cdot q_{i_j} \cdot |t_j| > q_{i_{j-1}+1} \cdot \ldots \cdot q_{i_j} \cdot q_{i_j+1} \cdot \big(|t_j| + 1\big)$$
$$\Rightarrow \frac{|t_j|}{|t_j| + 1} > q_{i_j+1}.$$

as otherwise, from the definition of *var-Greedy*, $i_j + 1$ would have been included in $t_j$.

Note that,

$$q_{i^*_j} < \frac{|t_j|}{|t_j| + 1} \leq \frac{|t^*_j| - 1}{|t^*_j|}$$

where the first transition follows since for any $i' > i$ it holds $q_i \geq q_{i'}$ and $|t_j|/(|t_j| + 1) > q_{i_j+1}$, and the second transition holds since $|t^*_j| - 1 \geq |t_j|$.

Thus, we have that,

$$q_{t^*_j \setminus \{i^*_j\}} \cdot (|t^*_j| - 1) > q_{t^*_j \setminus \{i^*_j\}} \cdot q_{i^*_j} \cdot |t^*_j|.$$

This means that $u(t^*_j \setminus \{i^*_j\}) > u(t^*_j)$ and hence, we have that if $T'$ is the testing regime with $t'_{j'} = t^*_{j'}$ for each $j' \neq j$ and $t'_j = t^*_j \setminus \{i^*_j\}$, then $u(T') > u(T^*)$ which is a contradiction.

We conclude that for each $j \in [B]$, $i^*_j \leq i_j$, and the statement follows. □

Now, we are ready to show that for each instance *var-Greedy* returns an $e$-approximate testing regime.

Theorem 4. *var-Greedy returns an e-approximate testing regime.*

Proof. Let $T$ be the testing that is returned by *var-Greedy* which pools the first $n' \leq n$ individuals. We start by showing that for each $i$, $P^T_i \geq q_i \cdot \frac{1}{e}$.

Consider an individual $i$ that is included in test $t_j$ of size $k$. Note for each $i' \in t_j$, we have that $q_{i'} \geq (k-1)/k$, as otherwise we would have that

$$q_{i'} \prod_{i'' \in t_j \setminus \{i'\}} q_{i''} \cdot k < \prod_{i'' \in t_j \setminus \{i'\}} q_{i''} \cdot (k-1)$$

which is a contradiction. Thus, we get that

$$P^T_i = q_{t_j} = q_i \cdot \prod_{i' \in t_j \setminus \{i\}} q_{i'} \geq q_i \cdot \left(\frac{k-1}{k}\right)^{k-1} \geq q_i \cdot \frac{1}{e}. \tag{13}$$

From Lemma 8, we know that it exists an optimal non-overlapping testing regime $T^*$ that pools the first $n''$ individuals with $n'' \leq n'$. Then, we have

$$\frac{u(T^*)}{u(T)} = \frac{\sum_{i \in [n'']} P^T_i \cdot u_i}{\sum_{i \in [n']} P^T_i \cdot u_i} \leq \frac{\sum_{i \in [n']} q_i \cdot u_i}{\sum_{i \in [n']} q_i \cdot \frac{1}{e} \cdot u_i} \leq e,$$

where the third transition follows from Equation (13). □

## H MILP FORMULATION WITH CLUSTERS

In order to speed up the computation, we can consider groups of individuals with the same utilities and health probabilities as *clusters*. Clusters are particularly pertinent when utilities are integral and health probabilities are discretized, as is the case in our pilot study. Suppose we have $C$ clusters. We introduce a population vector $n \in \mathbb{N}^C_0$ so that $n_i$ denotes the number of individuals in cluster $i \in [C]$. In order to incorporate clustering into the MILP, we now let the index $i$ refer to a cluster (instead of an individual), and allow variables $x_i^j$ to take arbitrary non-negative integral values (instead of binary values in (2h)); these values represent the number of individuals from cluster $i$ that are included in test $j$. Additionally, we relax the non-overlapping test constraint (2f) to $\sum_{j\in[B]} x_i^j \le n_i$. As an aside, it is not difficult to show that if cluster populations are much larger than the testing budget at hand, then non-overlapping tests are optimal. We now state the full MILP with clustering below. Note that constraints (14b)–(14f) capture the exponential constraint (2b), while (14h)–(14j) capture the logarithmic constraint (2d).

$$\max \quad \sum_{j\in[B]} w^j \tag{14a}$$

$$\text{s.t.} \quad w^j \le \sum_{k\in[K]} a_k v_k^j + b_k \cdot \delta_k^j \quad \forall j \in [B], \tag{14b}$$

$$\sum_{k\in[K]} \delta_k^j = 1 \quad \forall j \in [B], \tag{14c}$$

$$\sum_{k\in[K]} v_k^j = l^j \quad \forall j \in [B], \tag{14d}$$

$$c_k \cdot \delta_k^j \le v_k^j \quad \forall j \in [B], k \in [K], \tag{14e}$$

$$c_{k+1} \cdot \delta_k^j \ge v_k^j \quad \forall j \in [B], k \in [K], \tag{14f}$$

$$l^j = y^j + \sum_{i\in[C]} x_i^j \log q_i \quad \forall j \in [B], \tag{14g}$$

$$1 = \sum_{k\in[L,U]} \gamma_k^j \quad \forall j \in [B], \tag{14h}$$

$$z^j = \sum_{k\in[L,U]} k \cdot \gamma_k^j \quad \forall j \in [B], \tag{14i}$$

$$y^j = \sum_{k\in[L,U]} \log(k) \cdot \gamma_k^j \quad \forall j \in [B], \tag{14j}$$

$$z^j = u \cdot x^j, \quad \forall j \in [B], \tag{14k}$$

$$\sum_{j\in[B]} x_i^j \le n_i, \quad \forall i \in [C], \tag{14l}$$

$$\sum_{i\in[C]} x_i^j \ge 1, \quad \forall j \in [B], \tag{14m}$$

$$\sum_{i\in[C]} x_i^j \le G, \quad \forall j \in [B], \tag{14n}$$

$$x_i^j \in \mathbb{N}_0, \quad \forall j \in [B], i \in [C], \tag{14o}$$

$$v_k^j \in \mathbb{R}, \quad \forall i \in [C], k \in [K], \tag{14p}$$

$$\delta_k^j \in \{0,1\}, \quad \forall i \in [C], k \in [K], \tag{14q}$$

$$\gamma_k^j \in \{0,1\}, \quad \forall i \in [C], k \in [L,U] \tag{14r}$$

# I ADDITIONAL FIGURES AND TABLES FROM EXPERIMENTS

Here we show figures and summary tables for our experiments comparing Approx and Greedy on the pilot study data with pool size constraint $G = 10$, and on synthetic populations of size $n = 150$ and pool size constraints $G \in \{5, 10\}$. For details on the experiments, we refer to Section 5.2.

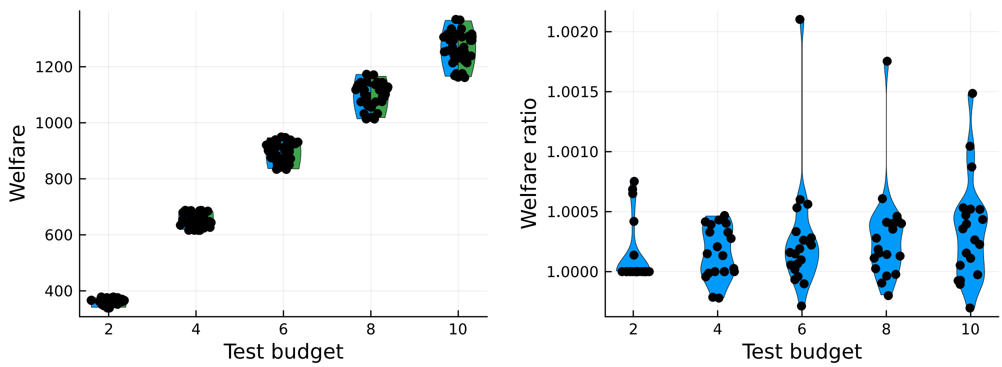



Fig. 1. Outcomes of Greedy and Approx on synthetic data with $n = 150$, pool size bound $G = 5$ and testing budgets $B \in \{2, 4, \ldots, 10\}$. Left: Welfares achieved by Approx (left regions, blue) and Greedy (right regions, red). Right: Ratios between the welfares of Approx and Greedy. In both figures, each black dot corresponds to one of the 20 randomly generated populations.

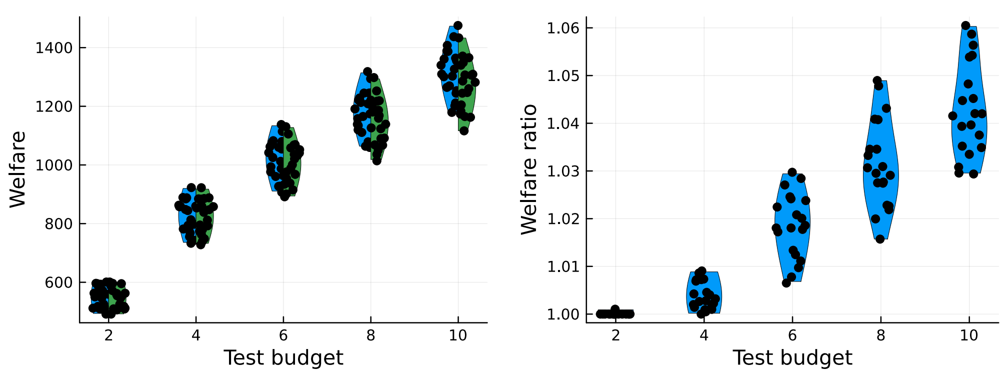



Fig. 2. Outcomes of Greedy and Approx on synthetic data with $n = 150$, pool size bound $G = 10$ and testing budgets $B \in \{2, 4, \ldots, 10\}$. Left: Welfares achieved by Approx (left regions, blue) and Greedy (right regions, red). Right: Ratios between the welfares of Approx and Greedy. In both figures, each black dot corresponds to one of the 20 randomly generated populations.

# J TOWARDS OVERLAPPING TESTING

In Section 3.2, we show that the gain of allowing overlapping testing, compared to non-overlapping testing, is at most 4. Despite extensive computational searching, no example with a gain of more than 7/6 has been found. In order to better understand the average-case gain, we conduct computational experiments in which we generate 20 populations of size 10, with utilities and probabilities

| | Approx | | | Greedy | |
|---|---|---|---|---|---|
| Budget | Welfare | Guarantee | Time | Welfare | Time |
| 2 | 866.60 | 0.47215 | 222 ms | 866.48 | 9 ms |
| 4 | 1646.23 | 0.94430 | 685 ms | 1645.98 | 22 ms |
| 6 | 2391.50 | 1.41645 | 1763 ms | 2391.00 | 17 ms |
| 8 | 3101.97 | 1.88861 | 16380 ms | 3101.18 | 25 ms |
| 10 | 3775.35 | 2.36076 | 6201334 ms | 3774.20 | 75 ms |
| 12 | 4399.37 | 2.83291 | 51746456 ms | 4397.69 | 277 ms |

Table 3. Summary showing welfare and computation time for Approx and Greedy on pilot study data (with a population of $n = 130$) and pool size constraint $G = 10$ with testing budgets $B \in \{2, 4, \dots, 12\}$. We also state the additive approximation guarantee of Approx (compared to optimal non-overlapping welfare).

| | Approx | | | Greedy | |
|---|---|---|---|---|---|
| Budget | Welfare | Guarantee | Time | Welfare | Time |
| 2 | 362.46 | 0.55666 | 187 ms | 362.41 | 12 ms |
| 4 | 652.56 | 1.11332 | 509 ms | 652.46 | 23 ms |
| 6 | 895.78 | 1.66998 | 1643 ms | 895.53 | 36 ms |
| 8 | 1099.03 | 2.22664 | 12170 ms | 1098.70 | 45 ms |
| 10 | 1264.35 | 2.7833 | 38272 ms | 1263.91 | 63 ms |

Table 4. Experiment summary on synthetic data with pool size bound $G = 5$ and testing budgets $B \in \{2, 4, \dots, 10\}$. Welfares and times are averaged over 20 randomly generated populations. We also state the additive approximation guarantee of Approx (compared to optimal non-overlapping welfare).

| | Approx | | | Greedy | |
|---|---|---|---|---|---|
| Budget | Welfare | Guarantee | Time | Welfare | Time |
| 2 | 548.03 | 1.24119 | 246 ms | 547.97 | 12 ms |
| 4 | 828.49 | 2.48239 | 1157 ms | 825.38 | 32 ms |
| 6 | 1022.38 | 3.72358 | 7680 ms | 1004.02 | 65 ms |
| 8 | 1180.68 | 4.96477 | 27963 ms | 1145.06 | 96 ms |
| 10 | 1319.03 | 6.20597 | 243776 ms | 1265.31 | 161 ms |

Table 5. Experiment summary on synthetic data with pool size bound $G = 10$ and testing budgets $B \in \{2, 4, \dots, 10\}$. Welfares and times are averaged over 20 randomly generated populations. We also state the additive approximation guarantee of Approx (compared to optimal non-overlapping welfare).

respectively drawn from $\{1, 2, 3\}$, and $\{0, 0.1, \dots, 1\}$. The pool size is unbounded and testing budgets are $B \in \{2, 3, 4\}$. We note that our choice of population size is constrained by the fact that computing optimal overlapping tests is significantly more computationally intensive. For the same reason, we restrict ourselves to comparing non-overlapping testing with 2-overlapping testing; in the latter case, individuals are permitted to lie in at most two tests. In order to compute optimal non-overlapping and 2-overlapping regimes, we formulate the optimization problems as an (exact) MILP parameterized by the overlap $k$, and refer to the resulting approach as $k$-Overlap. The results of running 1-Overlap and 2-Overlap on our 20 populations are shown in Fig. 3 and Table 6; they indicate that the gain of 2-Overlap is non-negligible but limited.

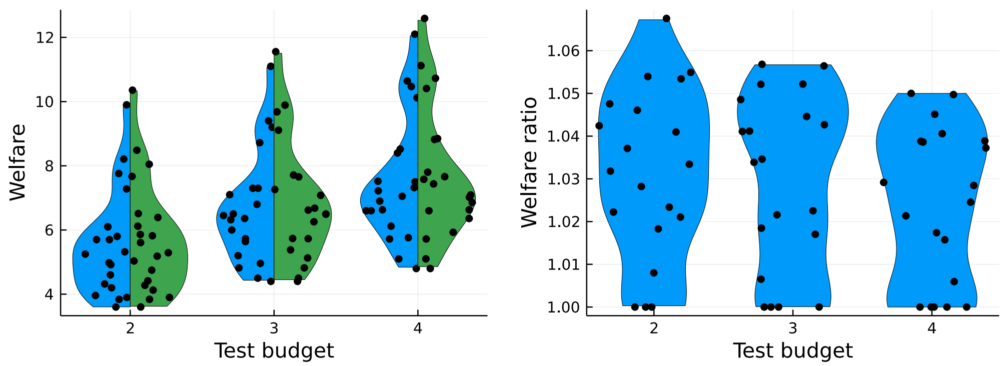



Fig. 3. Outcomes of 1-Overlap and 2-Overlap with population size $n = 10$, pool size constraint $G = 10$ and testing budgets $B \in \{2, 3, 4\}$. Left: Welfares achieved by 1-Overlap (left regions, blue) and 2-Overlap (right regions, red). Right: Ratios between the welfares of 1-Overlap and 2-Overlap. In both figures, each black dot corresponds to one of the 20 randomly generated populations.

| | 1-Overlap | | 2-Overlap | |
|---|---|---|---|---|
| Budget | Welfare* | Time* | Welfare* | Time* |
| 2 | 5.57 | 58 ms | 5.76 | 104 ms |
| 3 | 6.69 | 102 ms | 6.91 | 2107 ms |
| 4 | 7.55 | 136 ms | 7.75 | 94627 ms |

Table 6. Experiment summary showing welfare and computation time for 1-Overlap and 2-Overlap on populations of size $n = 10$ and pool size constraint $G = 10$, with testing budgets $B \in \{2, 3, 4\}$. Starred columns show mean values computed over the 20 random populations.

## K THE RANDOMIZED CONTROL TRIAL

### K.1 Randomization

The separation of treatment and control groups is essential for our protocol to work; in some small part to avoid health spillovers, but primarily to disentangle psychological dynamics. If non-treated participants were to run into treated participants, possible health spillovers would be contained within our health protocol: participants have a non-infectious 48 hour window in which they are allowed into the building after receiving a negative qPCR test result. This is also of relevance because, while we can monitor participants during working hours to some extent, we cannot control their socialization outside of the building.

The IPICYT campus lends itself, for the most part, to a two-group clustered randomization approach. The campus consists of two similar buildings for research in natural sciences, two similar buildings for research in computer science and mathematics, one building for classes, and one building for administration. The individuals participating in our trial belong to a research 'discipline', and within that discipline, to a working group. These working groups are randomly assigned to treatment and control groups. Because only one group from each discipline volunteered to participate in the trial, discipline and working group are henceforth analogous.

Students, researchers and staff are clustered based on their discipline/working group and each cluster is randomly assigned to treatment or control groups. Treatment and control groups work in different buildings or, conditional on the experimental sample size, different building floors or offices. Crucial for this approach to be effective is that individuals across working disciplines are comparable. Given IPICYT's reports about their staff, we know that staff, research students and researchers are assigned to work/study in each of the teams contingent only on their academic discipline, based on no individual characteristics. Hence, we may consider the assignment as pseudo-random. Nevertheless, we further collect a number of covariates to conduct a balance analysis.

### K.2 Mechanism

*K.2.1 Scheduling preferences.* Our mechanism allowed individuals in the treatment group to indicate their own preferences over days on which they wish to access the institute. A negative pooled test on a given day allowed individuals in the pool to access the campus for 48h. In the web app, participants were given a set of 10 virtual tokens that they could distribute arbitrarily among all consecutive two-day windows (Monday & Tuesday, Tuesday & Wednesday, etc.) on which they wished to enter the institute. This distribution of tokens then expressed the agent's relative preferences. (Assigning more tokens to some two-day window indicated a stronger preference for these two days.) The individual's utility for each two-day block is then computed from baseline utilities as described in Section 5.1 and their relative preference for the block.

### K.3 Outcomes and covariates

*K.3.1 Mental health outcomes.* Mental health problems related to social isolation as a consequence of the COVID-19 pandemic have been documented for students and the general population [Martinez Arriaga et al., 2021]. We conjecture that putting in place a safe education protocol decreases stress levels among students, researchers, and staff, by increasing *safe* sociability [Becchetti et al., 2017] and modulating the perception of health risk in the institute [Shan et al., 2022]. Consequently, subjective well-being may also be positively affected. Stress is measured via the validated 4-item Perceived Stress Scale by Sheldon Cohen [Cohen et al., 1994] and we use a variation of the European Quality of Life Survey measure of subjective well-being, using a life/subject evaluation approach [OECD., 2013][27]

*K.3.2 Performance, productivity, and learning.* The pandemic has disrupted learning processes and decreased productivity of Mexican students [Limón-Vázquez et al., 2020, Martinez Arriaga et al., 2021] and female researchers [King and Frederickson, 2021]. A significant portion of this downfall in productivity may be due to remote work with limited access to the necessary resources for work, research, and learning. We conjecture that our testing protocol improves (self-assessed) productivity and performance (in learning environments), and self-assessed learning experiences when compared to a remote work policy. In the presence of an alternative reopening strategy - as

[27]Note that we use baseline Stress and Subjective Well-being in our utilites' computation. On the other hand, we use endline Stress and Subjective Well-being to analyze between-group intervention effects.

is our case - we expect to see no difference between groups. I.e. two competing opening strategies that allow all or some individuals in the population to socialize within the institutional premises should increase productivity. Since our mechanism is stricter than the reopening policy of IPICYT, no difference in performance, productivity and learning is an indication of success.

We use a composite score for the evaluation of performance, productivity. Let $P^{ppa}_{i,z}$ denote the number of points "achieved" by the answer of subject $i$ to question $z$ pertaining to 'Performance, productivity, and sense of achievement'. Let $Z^{ppa}$ denote the number of relevant questions. Then the score is computed as $p_i = \frac{1}{Z^{ppa}} \sum_z P^{ppa}_{i,z}$. For learning, we use a self-assessment likert scale that ranges between 1 and 10, where 1 is poor and 10 is excellent.

*K.3.3 Covariates.* Besides outcome variables, we have collected additional socioeconomic and psychosocial data of participants. These data are used twofold. First, some of these features enter into the utility estimations needed for the testing algorithm. Second, we use relevant features/variables to check for group balance and, as needed for robustness checks and exploration of mechanisms, as covariates in our proposed linear models.

(1) Socio-economic attributes: gender, age, ethnicity, educational affiliation, perceived socio-economic status, financial dependants.
(2) Academic or job resources: access to internet, access to job materials, need to collaborate in person, access to a dedicated working space outside of the office.
(3) Psychosocial attributes: Sociability, fear of the virus, subjective well-being (generalized).

All covariates and their measurement strategy can be found in the baseline survey in Appendix K.7.

### K.4 Power and sample size

We estimate statistical power given the five outcome vectors outlined in Appendices K.3.1 and K.3.2. All our outcome variables are continuous scores, where 'perceived stress' is a non-integer vector ranging from 1 to 4, and 'life satisfaction', 'learning', 'productivity', and 'performance' are integer vectors with ranges 1 to 10 for the first two vectors, and 1 to 5 for the remaining three. We perform two types of tests to determine power: first, we estimate Cohen's $d$ with a two-sample t-test. We specify the true sample size based on post-attrition numbers. We also calculate the power of a two one-sided (TOST) equivalence test, given that we are interested in observing no difference in outcomes between experimental groups. Equivalence tests are usually a good complement to a null hypothesis test to avoid the misinterpretation of $p$-values higher than $\alpha$ being considered as evidence of the absence of an effect [Lakens, 2017]. We set the bounds based on a one unit change from the lower bound on the realized confidence intervals from the null hypothesis tests (t-tests).

| Power $(1-\beta)$ | $\alpha$ | $N$ | Cohen's $d$ |
|---|---|---|---|
| 0.80 | 0.05 | 120 | 0.515 |

Table 7. Two-sample t test power calculation, using the R package 'pwr', assuming ICC $\approx 0$, for all outcome variables: stress, life satisfaction, performance, productivity, learning.

Table 7 presents the results obtained from a two-sample t test power calculation. With a standard power score of 0.80 for the set of outcome vectors of interest at $\alpha = 0.05$ and a conservative $N = 120$ (from a post-attrition experimental sample of 122), we can detect an effect size of 0.515.

Table 8 shows the lower and upper bounds for each outcome variable in our analysis. We set the bounds based on a one unit change from the lower bound on the realized confidence intervals

| | Lower bound | Upper bound | Equivalence $p$ | Result |
|---|---|---|---|---|
| Stress | -0.453 | 0.453 | 2.44e-02 | reject null equivalence |
| Life satisfaction | -0.822 | 0.822 | 0.2.5e-02 | reject null equivalence |
| Performance | -0.371 | 0.371 | 2.3e-02 | reject null equivalence |
| Productivity | -0.403 | 0.403 | 1.61e-02 | reject null equivalence |
| Learning | -0.742 | 0.742 | 2.38e-02 | reject null equivalence |

Table 8. TOST Equivalence test, using the R package 'TOSTER', assuming ICC $\approx 0$.

from the null hypothesis tests. In all cases, the equivalence $p - values$ are statistically significant at $p < 0.02$, and we can reject the null hypothesis of the TOST equivalence test.

Our randomization approach relied on the affiliation of experiment participants to an institutional discipline. The rationale behind this approach was not statistical, but practical. We wanted experiment participants within the same experimental condition to only interact (in the institute) with their direct peers or, at least, not to interact with their experimental counterfactuals. Our power calculations do not include an intra-cluster correlation coefficient (ICC) parameter based on the realized ICCs in our sample. We used the outcome vector 'performance' as the baseline exploration vector and analyzed two possible sources of clustering: first, discipline affiliation, which presented a singular boundary, or close to zero variance by cluster that resulted in no ICC score. Second, the role (researcher, administrative staff, research student or taught student) of participants within the institute. The adjusted ICC was $\rho = 0.205$, too low to be considered reliable [Koo and Li, 2016].

### K.5 Metrics and methods

We propose a two-group experimental design where $[n] = 131$ subjects are semi-randomly assigned to either a treatment or a control group. Group balance in observed and unobserved heterogeneity is a direct result of random assignment, allowing for treatment status to be the only source of exogenous variation. As such, the mean group difference in the outcomes of interest can be presented as the causal effect of our testing strategy in those those dimensions.

We estimate the average treatment effect on the five, previously introduced, outcome variables. As a reminder, we state in our hypotheses that we are interested in non-significant differences between experimental conditions. We therefore refer to the $p$-values in bivariate regressions as evidence of no association.[28]

We previously explain that, while we hope to deliver an ATE, we are likely to deliver ITT results, based on the assumption that some treatment participants may not fully comply with the protocol by, for instance, not attending an invitation for saliva sample submission. The protocol is designed such that opting out of sample submission does not affect results, as the grouping algorithm (for test pools) is ran only on the subsample of compliers. The non-attendee is simply restricted from entering the premises until they get a negative test result, and they are not penalized when generating new sample submission invitations.

We denote our outcomes as $y_i \in \{s_i, w_i, l_i, p_i, pr_i\}$ :

- The average stress level and subjective well-being, measured by each individual's stress score $s_i$ and life satisfaction score $w_i$

[28]There is an ongoing debate over whether one can use insignificant $p$-values as evidence of no effect [Lakens, 2021]. We resort to equivalence tests as robustness checks for our findings.

- Subjects' self-assessed learning $l_i$, performance $p_i$ and productivity scores $pr_i$.[29]

The treatment effect of endline outcomes is estimated using a linear model, with HC1 standard errors. Let $Y$ denote the stacked vector of outcomes $(y_1, \dots, y_n) \in \{s_i, w_i, l_i, p_i, pr_i\}$. Let $\beta$ denote the vector of parameters to be estimated, and $\tau_i$ the treatment dummy. The independent variables are subsumed in $X = (1^n, \mathrm{T}, C)$, where $1^n$ is an $n-$vector of ones, T is a vector of treatment status $\tau_i$, and $C$ is a matrix of covariates[30]. Let $\epsilon$ denote the vector of error terms $\epsilon_i$. We estimate the model

$$Y = X\beta + \epsilon \tag{15}$$

for $Y \in \{S, W, P, Pr, L\}$ and test the hypothesis $\beta_1 \approx 0$. We collect baseline and endline data for the set of outcome vectors. Let $\Delta Y = Y_{endline} - Y_{baseline}$ denote the change in outcome $Y$ from baseline to endline. We estimate the *delta model*

$$\Delta Y = X\beta' + \epsilon' \tag{16}$$

to identify the effect of our intervention on the change in outcomes throughout the duration of the experiment. This analysis complements the analysis of endline outcomes: it eliminates all observed and unobserved confounds that are constant between our two points of measurement [Allison, 1990]. This allows for the interpretation of results not only as static differences but also in the context of possible outcome trajectories, and adjusted for static unobservables.

### K.6 Results

We present the results from the regression analysis based on Eqs. (15) and (16) in Tables 2, 9 and 10. On a high level, we are able to report for performance outcomes as well as mental health outcomes that our testing protocol has no negative effect, despite the increased effort it demands from participants. The exception is a statistically significant, albeit very small decrease in productivity of the trial group. As individuals in our testing regime have to exert more effort and are more curtailed in their freedom than individuals in the control group, the absence of a negative effect leads us to positively evaluate our trial with respect to the participants' mental health.

*Productivity, performance, learning outcomes.* Table 2 show the linear models' results for self-reported productivity, performance, and learning outcomes. In Table 9 we also report two other outcomes, achieving personally set (work/study) goals and achieving the goals set by your supervisor. We do so with the intention to better understand in which ways our protocol could affect a participants' ability to perform along a varied set of subjective educational/work dimensions. Similar to mental health outcomes, in the endline analysis treated participants outperform their control group counterparts in every outcome variable, albeit with small and non-statistically significant coefficients. For example, participants in the treatment group report, on average, 0.12 points more in the performance score than participants in the control group. At $p = 0.40$, the coefficient is not significant. To check robustness of these outcomes, we also estimate the bivariate delta models for performance, productivity, and learning (also in Table 9). All treatment effects are corrected downwards,[31] but remain non-significant with one exception. At $p = 0.05$, the change in treatment participants' productivity from $t_0$ to $t_1$ is at the border of statistical significance. It shows a downward effect of 0.25 score points. Mean self-reported productivity for treatment participants went from 2.27 at $t_0$, down to 2.17 at $t_1$. This small decrease or 0.1 may be due to the

[29]We adapt a measure based on the fit of 10 and 5 point likert scales, respectively, as per Versteeg and Steendijk [2019]

[30]We present covariate-controlled linear models in the Appendix but it is as robustness checks but are not part of the main analysis.

[31]During the course of the month, treatment participants experience a small and statistically insignificant decrease in performance and goals; however, they still report higher learning, on average.

added coordination effort exerted by treatment participants, and time invested in familiarising themselves with the protocol. On the other hand, control participants experienced an increase in mean self-reported productivity of 0.09 points; they went from 2.00 at $t_0$ up to 2.09 at $t_1$. This small increase in productivity may be benefit from transitioning from remote work to full institutional access. Together, they explain the negative and borderline statistically significant coefficient. This effect is causal, albeit small in magnitude.

*Mental health outcomes.* Table 10 shows that there are no significant effects of our testing strategy on the subjects' stress level or subjective well-being (life satisfaction). On average, subjects in the treatment group report a stress score that is 0.186 points higher than for subjects in the control group. At $p = 0.17$, this difference is not statistically significant. When looking at the change in stress from baseline to endline (or the duration of the trial) in the delta bivariate model, the magnitude of the coefficient decreases to 0.009 at $p = 0.95$. That is, treatment status induces little to no variation in the change in stress between $t_0$ and $t_1$. Treatment participants report higher average life satisfaction scores. At endline, the difference in scores is small at 0.089, and insignificant ($p = 0.81$). However, the magnitude of the coefficient drastically increases for treatment participants by 0.284 points when looking at the change in scores pre and post trial. The change in life satisfaction score is, again, statistically insignificant ($p = 0.37$). As individuals in our pool testing regime had to exert more effort and are curtailed in their freedom compared to individuals in the control group, the absence of statistically significant negative effects leads us to positively evaluate our mechanism with respect to participants' mental health outcomes.

| | Dependent variable: | | | |
|---|---|---|---|---|
| | Own goals | Supervisor goals | Δ Own goals | Δ Supervisor goals |
| Treat vs. control | 0.035 | 0.250 | −0.307 | −0.115 |
| | (0.170) | (0.159) | (0.192) | (0.191) |
| Constant | 2.344*** | 2.129*** | −0.131 | −0.113 |
| | (0.107) | (0.099) | (0.101) | (0.098) |
| Observations | 119 | 120 | 118 | 119 |
| $R^2$ | 0.0004 | 0.021 | 0.022 | 0.003 |
| Adjusted $R^2$ | −0.008 | 0.012 | 0.014 | −0.005 |

*p<0.1; **p<0.05; ***p<0.01

Table 9. Linear model regressions of further performance outcomes.

*Covariate analysis and secondary results.* We also ran covariate-controlled regressions as robustness checks for our bivariate models, and find no contradictions to our main findings in our model specifications. All covariates are taken from the baseline survey, whereas the dependant variables are taken from the endline survey. Further details are available from the authors upon request.

### K.7 Survey

(1) **Socio-demographic attributes**

1.1 Identifier

| | Dependent variable | | | |
|---|---|---|---|---|
| | Overall stress | Life satisfaction | Δ Overall stress | Δ Life satisfaction |
| Treatment vs. control | 0.186 (0.135) | 0.089 (0.371) | 0.009 (0.146) | 0.373 (0.414) |
| Constant | 2.238*** (0.093) | 7.556*** (0.259) | −0.813*** (0.096) | −0.270 (0.276) |
| Observations | 121 | 122 | 121 | 121 |
| $R^2$ | 0.016 | 0.0005 | 0.00004 | 0.007 |
| Adjusted $R^2$ | 0.007 | −0.008 | −0.008 | −0.002 |

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 10. Linear model regressions of mental health outcomes.

Please write down your IPICYT ID number: ____________________

1.2 Role

What is your role at the university?

☐ Taught student
☐ Research student
☐ Researcher
☐ Staff (Administration, maintenance, other employees of IPICYT) ____________

1.3 Affiliation [Only for students and researchers]

Which department are you affiliated with?

☐ Maths and Computer Science
☐ Natural Sciences
☐ Other: ____________________

1.4 Gender

Which gender do you identify yourself with?

☐ Female
☐ Male
☐ Other: ____________________
☐ Prefer not to say

1.5 Age

Please indicate your age in two digits: _____

1.6 Ethnicity

Which ethnic group do you identify most with?

☐ White
☐ Indigenous

☐ Mestizo
☐ Afrolatino
☐ Other: ______________________

(2) **Family, work, and socio-economics features**

2.1 (se) How many dependants do you have?

This could be children, children and partner, other relatives, etc.

☐ Answer: ______________________

☐ 0-1 [1pt.]
☐ 2 [2pt.]
☐ 3 [3pt.]
☐ 4 [4pt.]
☐ 5+ [5pt.]

2.2 (pr) How many people live in the same household as you?

This could be children, children and partner, siblings, other relatives, housemates etc.

☐ Answer: ______________________

☐ 0-1 [1pt.]
☐ 2 [2pt.]
☐ 3 [3pt.]
☐ 4 [4pt.]
☐ 5+ [5pt.]

2.3 (pr) How much of your time during a normal work day do you spend working on a computer?

☐ 0-10% [5pt.]
☐ 11-30% [4pt.]
☐ 31-50% [3pt.]
☐ 51-70% [2pt.]
☐ 70-100% [1pt.]

2.4 (pr) How much of your time during a normal work day do you spend on communication with colleagues?

☐ 0-10% [1pt.]
☐ 11-30% [2pt.]
☐ 31-50% [3pt.]
☐ 51-70% [4pt.]
☐ 70-100% [5pt.]

2.5 (pr) How much of your time during a normal work day do you spend working in a team?

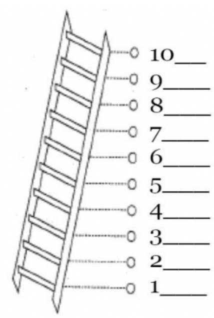
☐ 0-10% [1pt.]
☐ 11-30% [2pt.]
☐ 31-50% [3pt.]

☐ 51-70% [4pt.]

☐ 70-100% [5pt.]

2.6 Socio-economic class

(se) Look at the image of the ladder below. Imagine this ladder pictures how Mexican society is set up:




- At the top of the ladder are the people that are best off - they have the most money, the highest amount of schooling, and the jobs that bring the most respect.
- At the bottom are the people who are the worst off - they have the least money, little or no education, no job or jobs that no one wants or respects.

Now think of your family, please tell us where you think your family would be on this ladder: ________

☐ 10-9 [1pt.]

☐ 8-7 [2pt.]

☐ 6-5 [3pt.]

☐ 4-3 [4pt.]

☐ 2-1 [5pt.]

2.7 (se) Perceived socio-economic status

People sometimes describe themselves as belonging to the working class, the middle class, or the upper or lower class. Would you describe yourself as belonging to the

☐ Upper class [1pt.]

☐ Upper middle class [2pt.]

☐ Lower middle class [3pt.]

☐ Working class [4pt.]

☐ Lower class [5pt.]

☐ Prefer not to answer [0pt.]

(3) **Using digital media**

3.1 (pr) How much of your work time do you spend using the internet?

☐ 0-10% [5pt.]

☐ 11-30% [4pt.]
☐ 31-50% [3pt.]
☐ 51-70% [2pt.]
☐ 70-100% [1pt.]

3.2 (pr) How much of your leisure time do you spend using the internet?

☐ 0-10% [5pt.]
☐ 11-30% [4pt.]
☐ 31-50% [3pt.]
☐ 51-70% [2pt.]
☐ 70-100% [1pt.]

3.3 (pr) How do access the internet from home most of the time?

☐ Through laptop + wifi [1pt.]
☐ Through laptop + mobile connection [2pt.]
☐ Through phone + wifi [3pt.]
☐ Through phone + mobile connection [4pt.]
☐ N/A [5pt.]

(4) **Psychosocial features**

4.1 (psy) Sociability

Please write down the percentage of individuals (in your social circle) who would agree with the following statement about yourself: 'I spend a lot of time visiting friends' ________

☐ 0-10% [5pt.]
☐ 11-30% [4pt.]
☐ 31-50% [3pt.]
☐ 51-70% [2pt.]
☐ 70-100% [1pt.]

4.2 Fear

Please rate the extent to which you experience the following feelings at this moment: Fear because of the COVID-19 disease/ the SARS COV-2 virus.

☐ Not at all
☐ Not really
☐ Neutral
☐ Somewhat
☐ Very much

4.3 (psy) Perceived Stress Scale

| **1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|
| Never | Almost never | Sometimes | Fairly often | Very Often |

Based on the scale above, where zero indicates never experiencing that situation and four indicates experiencing that situation very often, please rate the following statements:

- In the last month, how often have you felt that you were unable to control the important things in your life?__________
- In the last month, how often have you felt confident about your ability to handle your personal problems?__________
- In the last month, how often have you felt that things were going your way?__________
- In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?__________

[Each score translates into the identical number of points. Then the average of the four sub-questions is computed.]

4.4 (psy) Subjective well-being

All things considered, how satisfied would you say you are with your life these days? Please tell me on a scale of 1 to 10, where 1 means very dissatisfied and 10 means very satisfied: __________

☐ 10-9 [1pt.]

☐ 8-7 [2pt.]

☐ 6-5 [3pt.]

☐ 4-3 [4pt.]

☐ 2-1 [5pt.]

4.5 Subjective well-being

Taking all things together on a scale of 1 to 10, how satisfied are you about IPICYT's efforts to keep you safe in the institute throughout the pandemic? __________

(5) **Performance self-assessment**

5.1 (ppa) Self-assessment of performance

How would you rate your overall performance for your job or degree in the past 4 weeks?

☐ Poor [5pt.]

☐ Below average [4pt.]

☐ Average [3pt.]

☐ Above average [2pt.]

☐ High [1pt.]

5.2 Self-assessment of learning

After the COVID-19 pandemic began, the way we learn and interact with our peers drastically changed. How would you say your learning experience has been in the past 4 weeks?

Please rate your learning process and experience between 1 and 10, where 1 is poor and 10 is excellent: __________

5.3 (ppa) Self-assessment of productivity

How would you rate your day-to-day productivity in your work in the past 4 weeks?

☐ Poor [5pt.]

□ Below average [4pt.]

□ Average [3pt.]

□ Above average [2pt.]

□ High [1pt.]

5.4 (ppa) Self-assessment of achievement (supervisor goals)

Considering again the work for your job or degree during the past 4 weeks, please select the statement that fits your situation best.

□ I have struggled to achieve the goals set by my supervisor/employer/course teachers [5pt.]

□ I have managed to achieve some of the goals set by my supervisor/employer/course teachers [4pt.]

□ I have achieved many of the goals set by my supervisor/employer/course teachers [3pt.]

□ I have achieved most of the goals set by my supervisor/employer/course teachers [2pt.]

□ I have achieved all or exceeded the goals set by my supervisor/employer/course teachers [1pt.]

5.5 (ppa) Self-assessment of achievement (own goals)

Considering again the work for your job or degree during the past 4 weeks, please select the statement that fits your situation best.

□ I have struggled to achieve the goals I set for myself [5pt.]

□ I have managed to achieve some of the goals I set for myself [4pt.]

□ I have achieved many of the goals I set for myself [3pt.]

□ I have achieved most of the goals I set for myself [2pt.]

□ I have achieved all or exceeded the goals I set for myself [1pt.]

### K.8 Consent form

You are invited to take part in a research project conducted by researchers from redacted for anonymity in conjunction with IPICYT. This project is funded by IPICYT. In accordance with international standards in the practice of randomized studies, this project has received ethical approval from the Research Ethics Committee at IPICYT and redacted for anonymity.

We ask that you read this form carefully prior to deciding to participate in the study. If you decide you do not want to participate, you may leave at any time without providing a reason and without penalty.

**Purpose:** The purpose of this study is to understand how the implementation of an algorithmic-base safe education protocol influences students and staff well-being and productivity during a pandemic.

**What happens during the study:** This study requires you to follow one of two protocols.

If you are selected to be part of the treatment group, you will participate in COVID-19 pooled testing. Throughout the course of the study, you may receive emails inviting you to submit a saliva sample, which will be pooled with other samples and tested at the LANBAMA laboratory at IPICYT. If your test is negative, then everyone in your pool is healthy and permitted to enter the institute for 48 hours. If your test is positive, then at least one person in your pool is infected, and you (as well as all other individuals in your pool) are not permitted to enter the institute until you are selected for re-testing and the next test result is negative. At no point are you obliged to submit a saliva sample, or to enter the building.

If you are selected to be part of the control group, you will be asked to follow the same remote working policy that is currently in place at IPICYT. If you would like to access the institute, you must contact the head of your department for permission.

We also ask all participants to respond to a short survey at the beginning and at the end of the trial - within a month's time - where you will be asked sociodemographic questions, alongside a set of psychological questions. You are not required to answer any questions that you may find uncomfortable. Furthermore, for the purpose of COVID-19 testing, you may be asked to give a saliva sample to the technicians at LANBAMA if you are selected for pooled testing. The sample will be used directly on the day of reception and will be destroyed after being processed for a qPCR test. The sample(s) will not be stored. You will be informed about the result of all pooled tests that contain your sample.

**Participation:** The trial is expected to run for a month, throughout August 2022, during which participants in the treatment group will receive free COVID-19 testing. Participants are asked to fill in a survey at the beginning and end of the study. In addition, participants in the treatment group are able to indicate their preference for which days they wish to be tested. Throughout the course of the month, the principal investigators will link health data (i.e. COVID-19 test results) to survey data (collected at the beginning and end of the trial). However, at the end of the trial all gathered data will be anonymized. If you wish to withdraw consent on the use of your data at any point during the study, please contact redacted for anonymity. You always have the option of stopping your participation in the study and you may leave at any time during the study (4 weeks from the start of the trial) without providing a reason and without penalty. If you decide to leave, the data you have provided up to this point will be anonymized immediately and deleted after attrition analysis.

**Potential risks:** If you choose not to participate in the study, or you participate and are selected into the control group, you will not be exposed to any additional risk. If you choose to participate and are selected into the treatment group, there is a risk that you will be infected if you are permitted to enter the institute and decide to do so. This risk is small, as all individuals must test negative in order to enter the institute. In particular, [our protocols, redacted for anonymity are much safer than reopening without monitoring for infections. While the probability of infection can be minimized and contained, it is not guaranteed to be zero. There is always a very small chance to get infected when participating in social activities, and COVID-19 comes with small and major consequences; among which, fever, cough, loss of taste and smell, respiratory problems and, in some cases, death.

Your survey responses are strictly confidential and will only be accessible to the researchers. Below, we describe the steps we are taking to protect your privacy. In addition, your decision on whether to participate will not adversely affect your relationship with IPICYT or any other institution to which the researchers are affiliated.

**Benefits:** Participating in this study means that you are aiding further development of science. Additionally, a successful trial would allow IPICYT to reformulate the institutional policy regarding work and study during the current and future waves of the pandemic into one that gives you more social interactions and flexibility with a minimized risk of contagion.

**Data protection and privacy:** The information collected during the study will be kept private. In concordance with the redacted for anonymity, is the data controller with respect to your personal data, and as such will determine how your personal data is used in the research. The University will process your personal data for the purpose of the research outlined above. Research is a task that is performed in the public interest. Further information about your rights with respect to your personal data is available at redacted for anonymity.

Responsible members of redacted for anonymity and IPICYT may be given access to data for monitoring and or audit of the study to ensure we are complying with the guidelines or as otherwise required by law. Moreover, in concordance with the signed Memorandum of Understanding, the Potosinian Institute of Scientific Research and Technology (IPICYT) will store and anonymize the original data in a secure server. During the trial, no one other than the head of the IPICYT Supercomputing Centre and responsible members of redacted for anonymity will have access to any records of this trial. The data will be stored in electronic form, encrypted and password protected. At the conclusion of the trial, all data will be anonymized, and none of the records will identify you. A copy of the anonymized data will be provided to the primary investigators of the trial. The data that we collect from you may be transferred to, and stored or processed at a destination outside Mexico. Archived/stored data, once anonymized, is available for research purposes upon request (primarily for peer-review replication processes). By submitting your personal data, you agree to this transfer, storing, or processing. After completion of the study, you cannot withdraw your personal information. Your individual privacy will be maintained in all publications or presentations resulting from this study. No information about you provided by you during this research will be disclosed to others without your written permission, except:

- if necessary to protect your rights or welfare (for example, if you are injured and need emergency care); or
- if required by law.

**Additional information:** If you are interested in receiving additional information about the results of the study, please contact the study authors.

**Concerns:** If you have any questions or concerns about any aspect of this project, you can contact the study authors at redacted for anonymity who will do their best to answer your query. The researcher(s) should acknowledge reception of your concern within 10 working days and give you an indication of how they intend to address it. If you fail to receive a response, are dissatisfied with the response you receive, or desire to report an aspect of how the study is being conducted, please contact the relevant Chair of Research Ethics Committee at the redacted for anonymity:

[Address, redacted for anonymity]

The Chair will seek to resolve the matter in a reasonably expeditious manner.

**Please confirm the following by marking each of the boxes next to the statements.**

*Please Mark Each Box*

- I confirm that I have read and understand the information for the above study and have had the opportunity to properly consider the information provided. □
- I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason and without any adverse consequences. □
- I understand the risks associated with participating in this study as explained in the information sheet. □
- I understand that a saliva sample will be taken during the study and that this sample will be tested for COVID-19. I understand that the sample will be destroyed after completion of this test or if I withdraw my consent. □
- I consider these samples a gift to redacted for anonymity and the LANBAMA laboratory and I understand I will not gain any direct personal benefit from this. □
- I understand that research data collected during the study may be looked at by designated individuals from redacted for

anonymity and IPICYT where it is relevant to my taking part in this study. I give permission for these individuals to access my data. I give permission for anonymized data to be made publicly available at the end of the research. □
- I understand that this project has been reviewed by, and received ethics clearance through, the Research Ethics Committee at IPICYT and redacted for anonymity □
- I understand who will have access to the personal data provided, how the data will be stored, and what will happen to the data at the end of the project. □
- I understand how this research will be written up and published. □
- I understand how to raise a concern or make a complaint. □
- I agree to take part in the study. □

By selecting "Yes, I agree to participate" below you are signifying that you have read and understood the above information and are agreeing to have the data that you provide during the course of the study to be processed accordingly.

□ Yes, I agree to participate

□ No, I do not agree to participate