An Upper Bound to the Benefits of Implementing Positive Assortative Matching in Pooled Testing

Gustavo Quinderé Saraiva*

March 23, 2026

Abstract

Dorfman pooled testing combines individual specimens (e.g., blood samples) into one test; if the pooled sample tests positive for infection, each specimen is tested separately. Under small prevalence levels, this method is known to reduce the expected number of tests required to screen a population, as individual tests only occur when a pooled test detects an infection. When conducting Dorfman testing, studies often recommend implementing positive assortative matching, i.e., pooling together samples with a similar risk of infection, as this tends to minimize the expected number of tests, the expected number of false negatives, and the expected number of false positives. However, because the logistics of collecting data and assorting samples from lowest to highest probability of infection can be costly, one may ask if implementing this procedure is indeed cost-effective. This article provides easy-to-compute upper bounds to the benefits of implementing Dorfman testing with positive assortative matching instead of matching samples randomly. Testers can then compare these upper bounds with the costs of estimating the probabilities of infection from each sample and then matching together those with similar risk of infection, to aid their decision on whether or not to implement this method.

Highlights

- This study derives a simple-to-compute upper bound to the benefits of implementing positive assortative matching when conducting Dorfman testing.
- Healthcare professionals can use these upper bounds to assess whether the additional effort required to implement positive assortative matching is justified by its potential gains.
- The upper bounds show that the benefits of implementing ordered pooling per patient tends to increase as the batch size increases.
- To further facilitate the usage of these upper bounds, a *Shiny app* (`https://jcbrnd.shinyapps.io/shiny_app_anonymous/`) was developed that allows users to input relevant parameters - such as disease prevalence, test sensitivity, test specificity and batch size - and instantly obtain the maximum benefits achievable by implementing positive assortative matching.

1 Introduction

Pooled testing is a technique commonly utilized for screening various infectious diseases like HIV, Hepatitis B, and COVID-19. It involves amalgamating samples from different individuals (e.g., multiple samples of blood) into pools and testing them together. In the approach conceptualized by Dorfman, often referred to as *Dorfman testing*, whenever the pooled test detects an infection, each sample used to form the pool is individually tested. When the prevalence of the disease is sufficiently low, this technique often reduces the

*School of Management, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Macul, Santiago, Chile. E-mail address: gsaraiva@uc.cl

total number of tests needed to effectively screen a population, as individual tests are only conducted when infection is detected in the pooled sample.[1]

The foundational work in the field of pooled testing is often traced back to the seminal work of Dorfman [1943], which introduced the concept of combining blood samples from American soldiers to identify syphilis during the Second World War. This area of study has since expanded, yielding numerous new applications, such as detecting other infectious diseases, including Chlamydia [McMahan, Tebbs and Bilder, 2012], HIV [Nguyen et al., 2019] and, more recently, COVID-19 [Basso et al., 2022, 2023; Grobe et al., 2020]. Research has also suggested implementing this procedure in industrial settings for pinpointing manufacturing defects [Sobel and Groll, 1959] and in the food industry to detect salmonella [Price, Olsen and Hunter, 1972; Adams et al., 2013].

Several different criteria can be used to form the pools. The literature has shown that *ordered pooling*, i.e., the case in which subjects with similar probability of infection are grouped together, is optimal, as this partition simultaneously minimizes the expected number of tests, the expected number of false positives, and the expected number of false negatives, provided that certain testable conditions hold [Aprahamian, Bish and Bish, 2019, 2018; Saraiva, 2023*a*].

Despite these results, in practice, pools are often formed randomly or by the order in which samples arrive at the laboratories [Barak et al., 2021; Agoti et al., 2021]. I conjecture that this may be due to logistics costs associated with implementing ordered pooling. Indeed, to effectively implement ordered pooling, laboratories may need detailed patient information that is not readily accessible within laboratory information systems. In addition, if pooled testing is not fully automated, ordering samples from lowest to highest probability of infection may place additional strain on laboratory technicians.

Motivated by these considerations, this article devises a simple method to estimate an upper bound to the potential benefits that can be achieved by implementing ordered pooling assuming all pools have the same size, as opposed to matching subjects randomly. Testers can then weigh these potential gains against the logistical expenses involved in implementing ordered pooling to devise an optimal policy. If the upper bounds indicate that the savings in costs associated with ordered pooling are too small, testers may find it more practical to simply match samples randomly or by the order they arrive for testing.

A desirable feature from the results presented in this article is their reliance on minimal information regarding the distribution of agents' infection probabilities. Indeed, my upper bounds only require information regarding the average probability of infection (i.e., the prevalence of the disease), and the maximum and minimum possible probabilities of infection. This is useful as, in practice, one may not have reliable estimates of the full distribution of probabilities of infection.

Even if the tester does not have information on the minimum and maximum probabilities of infection, the tester can conservatively set the minimum probability at 0 and the maximum at a very high level to get a more conservative upper bound. Of course, what should be considered as a conservative upper bound to the maximum probability of infection is relative: it should depend on the disease we are analyzing, the targeted population, and the data we have about those being screened. For example, if we are screening donated blood from the general US population, and the only data that we have from donors is their age and gender, setting an upper bound to the maximum probability of HIV infection at 10% is quite conservative. But if we are testing whether patients *infected* with HIV undergoing antiretroviral therapy are undetectable, as in Fosah Tayong et al. [2025], then this upper bound becomes quite unrealistic.

Though the benefits of implementing ordered pooling do not vary monotonically with the prevalence of the disease, for a given upper and lower limit to the probabilities of infection, this article shows how a tester can easily set a conservative level for the prevalence so that this benefit is maximized.

Pooled testing is usually conducted in batches: at each period a laboratory receives a batch of N samples to be grouped into disjoint pools of size $K \leq N$. I show that, not only the pool size K has a significant impact on the potential benefits of implementing ordered pooling, but the batch size N as well. Indeed, when the batch size N is small, there are usually not enough samples with a high probability of infection that can fill an entire group, which tends to reduce the benefits of implementing ordered pooling.

The article presents examples illustrating how these upper bounds can be applied to real data through

[1]Whether the prevalence of the disease is sufficiently low or not for pooled testing to be cost-effective relative to individual testing depends on the dilution effect, sensitivity and specificity of the test and the costs associated with classification errors. Using one of the features from the Shiny app `https://jcbrnd.shinyapps.io/shiny_app_anonymous/` developed as a companion material to this paper, readers can make this assessment for different assays.

the analysis of two case studies. The first case study explores the potential advantages of implementing ordered pooling for chlamydia detection in the United States. It highlights how the effectiveness of ordered pooling is significantly influenced by both the batch size and pool size used. Specifically, smaller batches and smaller pool sizes tend to yield considerably lower benefits.

The second case study analyzes the potential benefits that could have been achieved by implementing ordered pooling to detect SARS-CoV-2 (the pathogen responsible for COVID-19) in Chile during the beginning of 2022, using data on patients' age and vaccination status. Using a dataset that was made publicly available by the Chilean government during the pandemic, my upper bounds predict small gains from implementing ordered pooling as opposed to matching subjects randomly when pools are of size 5 or 4 (the pool sizes mostly used in Chile to detect SARS-CoV-2 during the pandemic). It is possible, however, that with more detailed data on patients (e.g., if, in addition to age and vaccination status, one also had data on patients' symptoms) one would observe a higher dispersion in the probabilities of infection, thus increasing the benefits of implementing ordered pooling in this context.

The derived formulas for the upper bounds can be easily implemented in other real-world applications. To further facilitate their use, I developed a *Shiny app* (`https://jcbrnd.shinyapps.io/shiny_app_anonymous/`) that allows users to input relevant parameters - such as disease prevalence, test sensitivity, teste specificity and batch size - and instantly obtain the maximum benefits achievable through ordered pooling.

2 Related Literature

It is well known since the seminal work from Dorfman [1943] that the optimum pool size associated with Dorfman testing depends heavily on the prevalence of the disease: the more prevalent the disease, the lower the pool size should be. If the prevalence is too high, testing every specimen individually is optimal. Based on this simple result, scholars like Saraniti [2006] proposed matching together samples with *equal* risk of infection and assigning those with a higher risk of infection into smaller pools. By construction, one can easily see that this approach can substantially reduce the expected number of tests required to screen a population.

One limitation of this result, however, is that it requires each pool to be comprised of individuals who have the exact same probability of infection. This cannot always be accomplished, especially if the batch size in question is small and we have a high diversity in risks of infection. Aprahamian, Bish and Bish [2019] addressed this issue by creating an algorithm that finds the optimal partition when implementing Dorfman testing. To accomplish this goal, the authors first showed that, for any given partition, we can always find another partition, with the same pool size configuration, that yields better outcomes by matching together those with *similar* probabilities of infection, without requiring every sample from each pool to have the same risk of infection. This result allows the authors to narrow down their search for the optimal partition. Focusing on this narrower set, the authors show that they can implement an optimal-path algorithm to find the optimal partition in polynomial time.

Like in Dorfman [1943] and Saraniti [2006], Aprahamian, Bish and Bish [2019] assume that pooled testing is not subject to *dilution effects*, i.e., they assume that adding more non-infected specimens into an infected pool does not reduce the probability that infection is detected. Saraiva [2023*a*] derives conditions under which the results from Aprahamian, Bish and Bish [2019] still hold when the precision of the test *is* affected by dilution effects. Saraiva [2023*a*] finds that implementing ordered pooling can greatly reduce testing costs by as much as 25% (a result similar to the one obtained in Aprahamian, Bish and Bish [2019]), but that these savings become far less impressive when pool sizes are required to be small and have the same size.

Aprahamian, Bish and Bish [2018] derived similar results as the ones in Saraiva [2023*a*], but assuming a different testing protocol, and requiring all pools to have the same size (i.e., that pool sizes are homogeneous). One of the benefits of working with homogeneous pool sizes is that this procedure tends to be easier to implement, especially when testing is not fully automated or in situations where we do not know patients' exact probabilities of infection, only the order of their probabilities of infection. This may explain why many laboratories implement Dorfman testing with homogeneous pool sizes. For this reason, and because the literature has already established that implementing ordered partitions tends to greatly reduce costs when pool sizes are allowed to be heterogeneous, throughout the paper, I assume that all pools must have the

same size.

One potential limitation of implementing ordered partitions is that subjects may have incentives to misreport their probability of infection to authorities to influence the pool they get assigned into and, hence, the accuracy of their test. But in an environment in which lying is costly and all pools must have the same size, Saraiva [2023*b*] shows that, in spite of these incentives, ordered pooling is still optimal. This happens because, in equilibrium, there is no reversal of probabilities of infection, i.e., those who report having a higher probability of infection are indeed more likely to be infected, so the signals from subjects' reports are still informative, albeit less informative than they would be if all agents reported their types truthfully. To prove this result, Saraiva [2023*b*] requires that, for each possible risk of infection, there exists a sufficiently large number of subjects who do not misreport their type. This ensures that, if someone reports having probability of infection $\hat{q}_i$, one is most likely matched with others who have made the same report.

In a different setting, Lipnowski and Ravid [2021] also show that implementing positive assortative matching is optimal when the planner has a limited number of test kits, and wishes to implement pooled testing without the retesting phase to determine who should be quarantined and who should be released. In their setting, the test kits are error-free (i.e., they have perfect sensitivity and perfect specificity) and the objective of the planner is to minimize a linear combination of the number of individuals quarantined and the number of infected individuals who are released.

One of the few studies that find potential benefits of not matching subjects according to ordered partitions is the work of Bobkova, Chen and Eraslan [2023], who show that *negative* assortative matching may actually reduce the expected number of tests. The testing scheme they study, however, differs from Dorfman testing. In their proposed method, if a pool tests positive for an infection, each subject from the pool, but one, is retested individually. If no infection is detected in this phase, then one can "safely assume" (if the test is error-free) that the catalyst for the previous detection was the subject who was left out from the retesting phase, so this subject does not need to be retested; otherwise, she is retested, as the original pool may have been comprised of more than one infected specimen. While this testing scheme can marginally reduce the expected number of tests compared to Dorfman testing, in practice, it can generate more false positives if the test does not have perfect specificity. Moreover, this testing scheme may be perceived as unfair by those who are excluded from the individual tests from the second phase of screening, as they are not as thoroughly tested as the remaining subjects. Another disadvantage of this method is that it is slower to implement than Dorfman testing. Indeed, under Dorfman testing, only two testing phases occur in parallel: the pooling phase and the retesting phase. The mechanism proposed by Bobkova, Chen and Eraslan [2023], on the other hand, requires three testing phases: the pooling phase, the first part of the retesting phase, and then individually testing some of the subjects who were left out from the first part of the retesting phase.

Ghosh et al. [2021] propose a combinatorial testing scheme in which all patients are tested in one go, using a novel "Tapestry" method. Using information on patients' viral loads, they show that this testing scheme has the potential to substantially reduce the expected number of tests compared with the random pooling version of Dorfman testing while, at the same time, improving the speed with which results are generated, as all tests occur in one phase, as opposed to two phases. This testing protocol is, however, computationally more difficult to implement, and it is not very flexible to accommodate different batch sizes. In addition, the test tends to require large batch sizes to guarantee its effectiveness (at least 105 samples). This testing scheme also tends to generate more false negatives and false positive results on average.[2] These limitations may help to explain why Dorfman testing is still widely popular.

Though several studies dating back to the seminal work of Dorfman [1943] have documented the importance that the pool size and the prevalence of a disease has on the efficacy of pooled testing (e.g., Hwang [1976], Wein and Zenio [1996] and Saraniti [2006]), to the best of my knowledge, researchers have not yet analyzed the impact that the batch size (i.e., the total number of samples processed in a given period) may have on the benefits of implementing the Dorfman testing protocol. This paper fills this gap by showing that the benefits of grouping subjects according to positive assortative matching tends to be smaller when the batch size is small, as, in this case, the tester is less likely to form pools comprised entirely of samples that have a high risk of infection.

[2]With the assumptions invoked in Ghosh et al. [2021], Dorfman testing never generates false negative or false positive results, but their Tapestry method can generate some false classifications, as infected patients are not as thoroughly tested under this method compared to the case in which they are individually tested in a second phase.

3 Environment

Table 1: Notation

Notation	Description
$\mathcal{S}$	Batch of subjects to be tested on a given day.
N	Batch size: $N = \lvert\mathcal{S}\rvert$.
K	Pool size.
n	Number of pools: $n = N/K$.
I	Number of infected within a pool.
$h(I, K)$	Dilution function: it returns the probability that infection is detected in a pool with K specimens with exactly $I \leq K$ of them infected.
S_e	Sensitivity of individual tests: $S_e = h(1, 1)$.
S_p	Specificity of individual tests: $S_p = 1 - h(0, 1)$.
q_i	ith ordered statistic of probabilities of infection: $q_1 \leq q_2 \leq \cdots \leq q_N$.
q	Vector of risks of infection ordered from lowest to highest risk: $q = (q_1, q_2, \cdots, q_n)$.
$\overline{q}$	Average risk of infection from the batch: $\overline{q} = \sum_i q_i/N$.
μ	Prevalence of the disease.
a	Minimum risk of infection.
b	Maximum risk of infection.
G_g	$\{(g-1)K + 1, (g-1)K + 2, \cdots, gK\}$, where $g \in \{1, 2, \cdots, n\}$.
α	$(b - \mu)/(b - a)$.
$UB_o^T(\mu, n, K)$	An upper bound to the reduction in the expected number of tests per subject obtained by implementing ordered pooling conditional that $\overline{q} = \mu$.
$UB_o^{FP}(\mu, n, K)$	An upper bound to the reduction in the expected number of false positives per subject obtained by implementing ordered pooling conditional that $\overline{q} = \mu$.
$UB_o^{FN}(\mu, n, K)$	An upper bound to the reduction in the expected number of false negatives per subject obtained by implementing ordered pooling conditional that $\overline{q} = \mu$.

Suppose that there is a finite batch $\mathcal{S}$ of subjects to be tested, with $\lvert\mathcal{S}\rvert = N$ and $\mathcal{S} \cap \mathbb{N} = \emptyset$. Each subject can either be infected or not infected.

Subjects may have heterogeneous probabilities of infection. While many articles adopting this framework assume that the probabilities of infection from each subject are known at the time the tester is choosing which pool size to implement (e.g. Aprahamian, Bish and Bish [2019] and Saraiva [2023*a*]), I instead assume that these probabilities are random at this stage. This assumption is made to take into account that, in practice, risks of infection may vary from batch to batch. Moreover, the objective os this study is to aid testers in implementing an optimal pooling strategy using historical data from the population, before they have engaged in the sunk cost of investing in the infrastructure necessary to implement ordered pooling (e.g., improving information systems, increasing automation, etc.). So, ideally, the tester's choice of which pool size to use should not depend on idiosyncratic variations in the batch, but should be a function of the "expected batch" that the tester gets.

Therefore, I assume that each subject $s \in \mathcal{S}$ is infected with a random probability $Q_s \in [0, 1]$, where $(Q_s)_{s\in\mathcal{S}}$ follows an i.i.d distribution with support $[a, b] \subseteq [0, 1]$. The assumption that probabilities of infection are identically and independently distributed simplifies the analysis and generates a more conservative estimate of the benefits of implementing ordered pooling. Indeed, if the tester pooled samples according to the order in which they arrived and the probabilities of infection had positive serial correlation, matching samples in this manner would already implement some degree of positive assortative matching. In this case, the added benefits of ensuring that samples were grouped according to ordered pooling would diminish.

The realized probability of infection from subject s is denoted by q_s. The realized vector of probabilities of infection is denoted by $(q_s)_{s\in\mathcal{S}}$. For each $i \in \{1, 2, \cdots, N\}$, q_i corresponds to the ith order statistic of the realized vector of probabilities of infection, so that

$$q_1 \leq q_2 \leq \cdots \leq q_N.$$

Similarly, for a given realization of probabilities of infection, $s_i \in \mathcal{S}$ corresponds to the subject with the ith lowest probability of infection.

If a subject is individually tested and she is not infected, the test will correctly classify her as not infected with probability $S_p \in (0, 1]$. An infected subject who is individually tested is correctly classified as infected with probability $S_e \in (0, 1]$. So, S_p corresponds to the *specificity* of the test, while S_e corresponds to its *sensitivity*. It is assumed that $S_e > 1 - S_p$, so that whenever a test detects an infection, the subject has a higher likelihood of infection compared to the case in which no infection is detected.

In a Dorfman testing procedure, the subjects to be tested, $\mathcal{S}$, are pooled into disjoint groups, and the samples from subjects belonging to the same group are amalgamated and tested together. If the test detects the presence of an infection in the pooled sample, each subject within the pool is tested individually.

Throughout this article all pools will be assumed to have the same size $K \leq N$ (i.e., pool sizes are *homogeneous*). Though allowing pool sizes to be heterogeneous (e.g., by allowing those with a higher probability of infection to be polled in smaller groups) tends to substantially increase the potential benefits of implementing ordered pooling, as documented in Saraniti [2006], Aprahamian, Bish and Bish [2019] and Saraiva [2023*a*], in practice testers seem to rely mostly on homogeneous pool sizes,[3] as they are simpler to implement and require less information. Indeed, when pools are allowed to be heterogeneous, finding the optimal partition requires knowing not only the relative ordering of patients in terms of their probability of infection but also the precise magnitudes of these probabilities. Moreover, when laboratories are not equipped with a high level of automation, the reconfiguration of pool sizes on a constant basis may increase the work load from lab technicians.

In order to ensure that all pools indeed have the same size K, we must also have $N = nK$ for some $n \in \mathbb{N}$. When this condition fails in practice, laboratories may form one residual pool containing the remaining N mod K samples, with assignments to this pool made at random.[4]

I denote $h(I, K)$ as the probability of detecting an infection in a pooled sample collected from $K \in \mathbb{N}$ subjects, conditional that exactly $I \in \{0, 1, 2, \cdots, K\}$ of them are infected. From the definition of sensitivity and specificity, we must have $h(1, 1) = S_e$ and $h(0, 1) = 1 - S_p$. I will refer to h as the *dilution function*. Because all pools are assumed to have the same size, I will sometimes write $h(I) = h(I, K)$ to avoid clutter notation. For each $K \in \mathbb{N}$, I assume that $h(I, K)$ is non-decreasing in I, i.e., the more infected subjects there are in the pool, the more likely the pooled test will detect an infection.

Assumption 1 *$h(I, K)$ is increasing in I.*

Letting μ denote the prevalence of the disease, we have that

$$\mu = \mathbb{E}[Q_s].$$

Similarly, the average probability of infection from the batch is denoted by

$$\overline{q} = \sum_{i=1}^{N} \frac{q_i}{N}.$$

Notice that, by the law of large numbers, $\overline{q}$ converges in probability to μ as the batch size N increases. Defining

$$G_g \equiv \{(g-1)K + 1, (g-1)K + 2, \cdots, gK\},$$

[3]For example, to detect SARS-CoV-2, the coastal region of Kenya in 2020 implemented homogeneous pools of size $K = 6$ [Agoti et al., 2021], whereas Chile relied mostly on pools of size $K = 5$ [Basso et al., 2023].

[4]It should be noticed, however, that when implementing ordered pooling with heterogeneous pool sizes, it is usually (but not always) optimal to assign those with higher probabilities of infection into smaller pool sizes [Aprahamian, Bish and Bish, 2019; Saraiva, 2023*a*]. In particular, if N mod $K = 1$, it is always optimal to assign the subject with highest probability of infection to be the one individually tested. Therefore, randomly assigning subjects to be part of the smaller pool and then implementing ordered pooling with the remaining samples will tend to be suboptimal. But this procedure still outperforms random pooling.

and

$$\mathcal{S}_g \equiv \{s_i \in \mathcal{S}; i \in G_g\}$$

for each $g \in \{1, 2, \cdots n\}$, subjects are said to be matched according to *ordered pooling* if they are matched according to the following partition of $\mathcal{S}$:

$$\begin{aligned}\Big\{\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_n\Big\} = \Big\{&\{s_1, s_2, \cdots, s_K\},\\ &\{s_{K+1}, s_{K+2}, \cdots, s_{2K}\}, \cdots,\\ &\{s_{N-K+1}, \cdots, s_N\}\Big\}\end{aligned}$$

In words, subjects are matched according to ordered pooling if they are sorted from lowest to highest risk of infection and then matched with those belonging to the same risk quantile.

Subjects are said to be matched according to *random pooling* if they are randomly assigned to any given group with equal probability. So, under random pooling, subjects' idiosyncratic probabilities of infection do not affect the groups they end up assigned into.

It can be shown that, when all pools have the same size K and the dilution function $h(I, K)$ is concave in I, ordered pooling simultaneously minimizes the expected number of tests and the expected number of false positives [Aprahamian, Bish and Bish, 2018; Saraiva, 2023*a*]. Notice that concavity of $h(\cdot, K)$ is arguably a very mild assumption, as otherwise the dilution function would be extremely strong, in which case pooled testing would not be a viable alternative to individual testing, as it would generate too many false negatives.

Assumption 2 *(The dilution function is concave)* $\frac{\partial^2 h(I,K)}{\partial I^2} < 0$ *for all* $I \geq 0$.[5]

For ordered pooling to also minimize the expected number of false negatives, a more technical assumption is required: the dilution function $h(I, K)$ cannot be "too concave" for positive values of I. Formally, the following assumption must hold:

Assumption 3 *(The dilution function is not "too concave") The dilution function* $h(\cdot, K)$ *is such that, for all* $I \in \{1, 2, \cdots, K-1\}$,

$$\frac{I+1}{2I}h(I+1, K) + \frac{I-1}{2I}h(I-1, K) \geq h(I, K).$$

Therefore, if $h(I, K)$ is increasing in I and assumptions 2 and 3 hold, ordered pooling is optimal in the sense that it simultaneously minimizes the expected number of tests, the expected number of false positives and the expected number of false negatives.

Proposition 1 *[Saraiva, 2023a] If the dilution function* $h(I, K)$ *is increasing in* I *and satisfies assumptions 2 and 3, and all pools have the same size* K*, ordered pooling minimizes the expected number of tests, the expected number of false positives and the expected number of false negatives.*

Example 1 *As an example, one can easily show that the dilution functions*

$$h(I, K) = (1 - S_p) + (S_p + S_e - 1)\left(\frac{I}{K}\right)^{\delta}, \quad \textit{with } \delta \in (0, 1] \tag{1}$$

and

$$h(I, K) = \begin{cases} 1 - S_p, & \textit{if } I = 0 \\ S_e, & \textit{if } I > 0 \end{cases} \tag{2}$$

are increasing, concave and satisfy assumption 3, in which case implementing ordered pooling would be optimal in all of the three attributes mentioned earlier. In practice, one can calibrate the parameters of these dilution functions by using the results from experimental data, as illustrated later on in the case studies.

---

[5] As discussed in Saraiva [2023*a*], this assumption can be relaxed by only requiring "concavity at the discrete level", i.e., all that is needed is that $h(I+1, K) + h(I-1, K) \leq 2h(I, K)$ for all $I \in \{1, 2, \cdots, K-1\}$.

While proposition 1 predicts the optimality of ordered pooling, notice that it does not predict the magnitude of the benefits obtained from assorting subjects this way.

For each $(\overline{q}, n, K) \in [a,b] \times \mathbb{N}^2$, define $q(\overline{q}, n, K) \in [a,b]^N$, where each coordinate i from $q(\overline{q}, n, K)$ is given by

$$q_i(\overline{q}, n, K) = \begin{cases} a, & \text{if } i \leq m_{n,K}(\overline{q}) \\ c_{n,K}(\overline{q}), & \text{if } i = m_{n,K}(\overline{q}) + K \\ b, & \text{if } i > m_{n,K}(\overline{q}) + K \end{cases},$$

where

$$m_{n,K}(\overline{q}) \equiv \begin{cases} \max \begin{Bmatrix} m \in \{0, K, 2K, \ldots, N-K\}; \\ ma + (N-m)b \geq N\overline{q} \end{Bmatrix}, & \text{if } Ka + (N-K)b > N\overline{q} \\ N - K, & \text{else} \end{cases}$$

and

$$c_{n,K}(\overline{q}) \equiv (N\overline{q} - m_{n,K}(\overline{q})a - (N - m_{n,K}(\overline{q}) - K)b)/K$$

It is also useful to define

$$\alpha \equiv \frac{b-\mu}{b-a},$$

since $m_{n,K}(\overline{q})/N$ converges in probability to α as N goes to infinity.

Roughly speaking, $q(\overline{q}, n, K) \in [a,b]^N$ corresponds to the realization of the order statistics $(q_1, q_2, \cdots, q_N)$ that maximize the variability of the probabilities of infection *between* pools, but minimize the variability of the probabilities of infection *within* pools (assuming ordered pooling is implemented), subject to the constraint that the average probability of infection is equal to $\overline{q}$.[6] In the next sections I show that, when ordered pooling is implemented, the expected number of tests, the expected number of false positives and the expected number of false negatives are minimized under this configuration of probabilities of infection, conditional that subjects' average probability of infection must be equal to $\overline{q}$. If the batch size N is sufficiently large, one can then approximate $\overline{q}$ by μ and use $q(\mu, n, K)$ to compute an upper bound to the benefits of implementing ordered pooling as opposed to random pooling.

4 Expected Number of Tests

This section shows how to compute an upper bound to the reduction in the expected number of tests obtained when implementing ordered pooling as opposed to random pooling.

It can be shown that, in infinite populations, the expected number of tests per subject obtained when randomly matching N subjects into n equal groups of size K is given by:

$$T_r^K(\mu, n, K) \equiv \frac{1}{K} + \left[\sum_{I=0}^{K} h(I) \binom{K}{I} \mu^I (1-\mu)^{K-I} \right]. \tag{3}$$

The second term from equation (3) corresponds to the probability that infection is detected in a pool. This term should be multiplied by K because, in the event the pooled test detects infection, the K subjects from that pool are individually tested. But we should also multiply this term by $n/N = \frac{1}{K}$ as this corresponds to the total number of pools divided by the total batch size. Finally, $\frac{1}{K}$ is added to the expression to account for the total number of pools tested per subject.

For a finite batch of size N, we can approximate the expected number of tests of random pooling by computing $T_r^K(\overline{q}, n, K)$ instead, i.e., by replacing the population prevalence, μ, by the prevalence from the batch, $\overline{q}$ in equation (3) (numerical simulations indicate that this is indeed a reliable approximation).

---

[6] Notice that $q(\overline{q}, n, K)$ does not necessarily minimize the overall variance of $(q_1, q_2, \cdots, q_N)$ subject to the constraint that $\sum_i q_i/N = \overline{q}$, as K of the probabilities in $q(\overline{q}, n, K)$ may end up equal to some value $c_{n,K}(\overline{q}) \in (a,b)$, whereas the minimization of variance requires that there can be no more than one probability that differs from the extreme values of the distribution, a or b.

It is straightforward to show that, for a given realization of $(q_1, q_2, \cdots, q_N)$, implementing ordered pooling yields the following expected number of tests per subject:

$$T_o^K(q_1, q_2, \cdots, q_N) \equiv \sum_{g=1}^{N/K} t_g/N,$$

where

$$t_g \equiv 1 + K \sum_{I=0}^{K} h(I, K) P_{G_g}(I),$$

and

$$P_{G_g}(I) \equiv \sum_{\substack{G \subseteq G_g \\ s.t. |G| = I}} \prod_{i \in G} q_i \prod_{j \in G_g \setminus G} (1 - q_j), \tag{4}$$

i.e., t_g is the expected number of test associated with group G_g (i.e., the gth group with lowest probability of infection), and $P_{G_g}(I)$ is the probability that group G_g has exactly I infected subjects.

The first result from this section states that, conditional that $\sum_i q_i/N = \overline{q}$, $T_o^K(q_1, \cdots, q_N)$ is minimized when the dispersion in the probabilities of infection *between* pools is maximized, but the dispersion *within* pools is minimized (so that everyone within the same pool must have the same probability of infection). The intuition as to why the dispersion between pools should be as high as possible is intuitive enough: without dispersion in the probabilities of infection, ordered pooling would be equivalent to implementing random pooling. As to the requirement that the dispersion within pools must be minimal, this comes from the concavity of the dilution function. Indeed, $h(I, K)$ being concave with respect to I implies that the dilution effect is low, which implies that, whenever the pool has at least one infected individual, the probability of detecting infection in the pooled sample is high. So, take, for example, one pool comprised of only two individuals, where one has probability of infection 1, and the other has probability of infection 0. In this case, the pooled test would detect infection with high probability, as the pool is guaranteed to have 1 infected subject. But if both subjects had probability of infection equal to 1/2, the average probability of infection within the pool would be preserved, and yet, there would be a positive probability of 1/4 that none of them was infected, in which case the expected number of tests from this pool would be smaller.

Proposition 2 *Suppose that* $(q_1, \cdots, q_N)$ *is such that* $\sum_i q_i/N = \overline{q}$ *and* $q_i \in [a, b]$ *for all* i. *Then, if* $h(I, K)$ *is increasing and concave in* I *(i.e., assumptions 1 and 2 hold), we must have* $T_o^K(q(\overline{q}, n, K)) \leq T_o^K(q_1, \cdots, q_N)$.

For each $(\overline{q}, n, K) \in [a, b] \times \mathbb{N}^2$ define

$$UB_o^T(\overline{q}, n, K) \equiv T_r^K(\overline{q}, n, K) - T_o^K(q(\overline{q}, n, K)).$$

In words, $UB_o^T(\overline{q}, n, K)$ corresponds to an upper bound to the reduction in the expected number of tests per subject that one can get by implementing ordered pooling as opposed to random pooling, when assuming the average probability of infection from the batch to be equal to $\overline{q}$. Because the tester cannot know ex-ante the exact average probability of infection from each batch, the tester may approximate $\overline{q}$ by the population prevalence μ by relying on the law of large numbers. The following theorem states that, provided that n is sufficiently large (so that the batch size $N = nK$ is sufficiently large), this is indeed a good approximation, as $UB_o^T(\overline{q}, n, K)$ converges in probability to the limit of $UB_o^T(\mu, n, K)$ as n goes to infinity. The proposition also provides an analytical formula for this upper bound when the batch size is large.

Theorem 1

$$\lim_{n \to \infty} UB_o^T(\mu, n, K) = \sum_{I=0}^{K} h(I) \binom{K}{I} \left[\mu^I (1 - \mu)^{K-I} - \alpha a^I (1 - a)^{K-I} - (1 - \alpha) b^I (1 - b)^{K-I} \right], \tag{5}$$

and

$$UB_o^T(\overline{q}, n, K) \xrightarrow{p} \lim_{n\to\infty} UB_o^T(\mu, n, K)$$

Now, one may ask whether the asymptotic upper bound given by equation (5) is indeed a reliable approximation to the the actual upper bound for small levels of n. The next proposition indicates an affirmative answer to this question, as the upper bound $UB_o^T(\mu, n, K)$ converges to its limit from below, so that computing expression (5) as opposed to $UB_o^T(\mu, n, K)$ yields a looser (and therefore, more conservative) upper bound.

Proposition 3 *For all $n \in \mathbb{N}$ we have that*

$$UB_o^T(\mu, n, K) \leq \lim_{n\to\infty} UB_o^T(\mu, n, K)$$

Not only does $UB_o^T(\mu, n, K)$ converge to its limit from below, but simulations indicate an increasing trend in $UB_o^T(\mu, n, K)$ as n increases, as illustrated in figure 1 below.

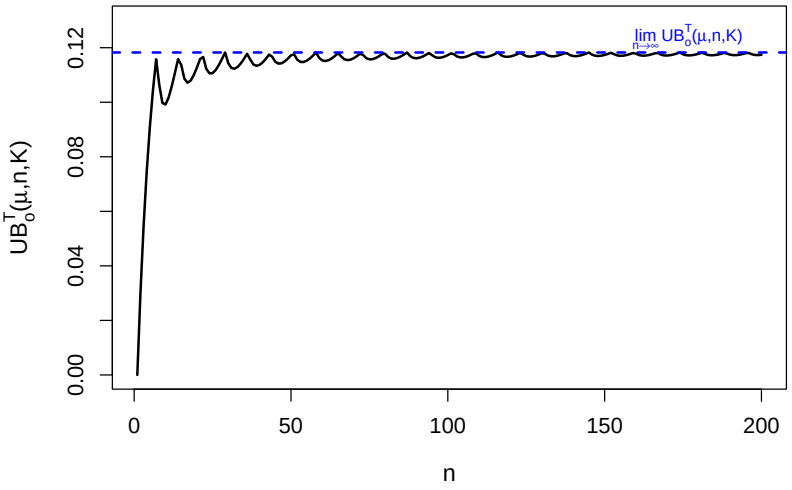


Figure 1: The solid line corresponds to $UB_o^T(\mu, n, K)$ for different values of n, assuming $K = 10$, $\mu = 0.05$, $a = 0.01$ and $b = 0.3$. The dilution function is given by $h(I, K) = (1 - S_p) + (S_e + S_p - 1)(I/K)^\delta$, with parameters $\delta = 0.15$, $S_e = 0.99$ and $S_p = 0.98$. The dashed blue line corresponds to $\lim_{n\to\infty} UB_o^T(\mu, n, K)$ given by expression (5), which is always greater than or equal to $UB_o^T(\mu, n, K)$ for all $n \in \mathbb{N}$.

Intuitively, this increasing trend occurs because, for small values of N and large values of b (relative to μ), computing $q(\overline{q}, n, K)$ yields no groups in which agents have probability of infection equal to the maximum possible probability, b, so that there is less variability between groups, which, in turn, implies that the benefits of implementing ordered pooling are limited. Meanwhile, for large values of N there are enough subjects with a high probability of infection who can fill an entire pool, in which case, implementing ordered pooling becomes more beneficial. This is better illustrated in example 2 below.

Example 2 *Suppose that $a = 0.01$ and $b = 0.3$ and $\mu = 0.05$. Then, if the batch size is equal to $N = 50$ and the pool size is equal to $K = 10$, we have that*

$$q(\overline{q}, n, K) = (\underbrace{0.01, \cdots, 0.01}_{40\ entries}, \underbrace{0.21, \cdots, 0.21}_{10\ entries}),$$

so that no group ends up with subjects with the maximum possible probability of infection $b = 0.3$.

But if I assume $N = 100$*, while not changing the other parameters, then* $q(\overline{q}, n, K)$ *yields one group in which all subjects have probability of infection* $b = 0.3$*, another group in which all subjects have probability of infection* 0.12*, and the remaining groups have subjects with probability of infection* $a = 0.01$.

A similar argument can be made when a is small relative to μ, and the batch size N is small: in this case $q(\overline{q}, n, K)$ yields no groups in which agents have probability of infection equal to the minimum possible probability, a, which diminishes the variability between groups and, thus, the benefits of implementing ordered pooling.

The reason why $UB_o^T(\mu, n, K)$ oscillates for values of n sufficiently high, potentially touching its asymptotic limit, follows from the discrete nature of pool sizes: if $m_{n,K}(\mu)/N$ happens to be exactly equal to α (a possible scenario when N is sufficiently large), all groups formed to create the upper bound must either have subjects with probability of infection a, or subjects with probability of infection b, in which case $UB_o^T(\mu, n, K)$ should be equal to its asymptotic limit. But if $m_{n,K}(\mu)/N < \alpha$, one of the pools will have subjects with a probability of infection on the interval (a, b), reducing the variance of the probabilities of infection between groups, thus causing $UB_o^T(\mu, n, K)$ to fall below $\lim_{n\to\infty} UB_o^T(\mu, n, K)$. But as the batch size N increases (and consequently the number of pools n), this single group with intermediate probability of infection starts having a lower impact on the total average, so that the discreteness of pool sizes starts having a lower impact on $UB_o^T(\mu, n, K)$. This is why the oscillatory behavior of $UB_o^T(\mu, n, K)$ diminishes as the batch size increases.

Notice that theorem 1 is an asymptotic result: in order for it to hold in practical settings, the batch size must be sufficiently large.[7] But, in the absence of dilution effects (i.e., when the dilution function is given by equation (2)), I show that $\mathbb{E}_{\overline{q}}\left[UB_o^T(\overline{q}, n, K)\right] \leq \lim_{n\to\infty} UB_o^T(\mu, n, K)$ for every n, dispensing the necessity for the batch to be large.

Proposition 4 *Suppose that*

$$h(I, K) = \begin{cases} 1 - S_p, & \text{if } I = 0 \\ S_e, & \text{if } I > 0 \end{cases}.$$

Then,

$$\mathbb{E}_{\overline{q}}\left[UB_o^T(\overline{q}, n, K)\right] \leq \lim_{n\to\infty} UB_o^T(\mu, n, K) = \\ (S_e + S_p - 1)\left[\alpha(1-a)^K + (1-\alpha)(1-b)^K - (1-\mu)^K\right] \tag{6}$$

Proposition 4 may be applicable to situations in which dilution effects are believed to be very small. One instance in which this might be the case is when pool sizes are restricted to be very small (e.g., pools cannot have more than 5 samples). In this case, even if only one specimen from the pool is infected, the pooled test should detect an infection with high probability.

5 Expected Number of False Positives

Minimizing the expected number of false positives is arguably as important as minimizing the expected number of tests, as false positives may trigger follow up tests and can cause unnecessary anxiety to patients. This section shows how to compute an upper bound to the reduction in the expected number of false positives obtained when implementing ordered pooling as opposed to random pooling.

One can easily show that, for infinite samples, the expected number of false positives per subject obtained when implementing random pooling with pool sizes equal to K is given by

$$FP_r^K(\mu, n, K) \equiv (1-\mu)\left[\sum_{I=0}^{K-1} h(I)\binom{K-1}{I}\mu^I (1-\mu)^{K-1-I}\right](1-S_p). \tag{7}$$

[7] The Online Appendix describes a method that can be used to assess whether a batch size is sufficiently large.

For any given subject within a pool, the term in brackets from equation (7) corresponds to the probability that the pooled test detects an infection conditional that the subject is *not* infected. This probability is multiplied by $(1 - S_p)$, as this corresponds to the probability that the subject is (incorrectly) detected as infected once retested. This expression must also be multiplied by $(1 - \mu)$ to yield the unconditional probability that the subject receives a false positive classification.

Similar to the expected number of tests, the expected number of false positives obtained when implementing random pooling in a finite batch can be approximated by $FP_r^K(\overline{q}, n, K)$, i.e., by replacing the population prevalence μ by the prevalence from the batch $\overline{q}$.

It is straightforward to show that, for a given realization of $(q_1, q_2, \cdots, q_N)$, implementing ordered pooling yields the following expected number of false positives per subject:

$$FP_o^K(q_1, q_2, \cdots, q_N) \equiv \sum_{g=1}^{N/K} \text{fp}_g / N,$$

where

$$\text{fp}_g \equiv \sum_{I=0}^{K} P_{G_g}(I) h(I, K)(K - I)(1 - S_p) \tag{8}$$

is the expected number of false positives within group G_g (i.e., the gth group with lowest probability of infection), and $P_{G_g}(I)$ is the probability that group G_g has exactly I infected subjects (see equation (4)).

Proposition 5 *Suppose that $(q_1, \cdots, q_N)$ is such that $\sum_i q_i/N = \overline{q}$ and $q_i \in [a, b]$ for all i. Then, if $h(I, K)$ is increasing and concave in I (i.e., assumptions 1 and 2 hold), we must have $FP_o^K(q(\overline{q}, n, K)) \leq FP_o^K(q_1, \cdots, q_N)$.*

For each $(\overline{q}, n, K) \in [a, b] \times \mathbb{N}^2$ define

$$UB_o^{FP}(\overline{q}, n, K) \equiv FP_r^K(\overline{q}, n, K) - FP_o^K(q(\overline{q}, n, K)).$$

In words, $UB_o^{FP}(\overline{q}, n, K)$ corresponds to an upper bound to the reduction in the expected number of false positives per subject that one can get by implementing ordered pooling as opposed to random pooling, when assuming the average probability of infection from the batch is equal to $\overline{q}$. The following theorem states that, provided that n is sufficiently large, one can approximate $UB_o^{FP}(\overline{q}, n, K)$ by $UB_o^{FP}(\mu, n, K)$, since $UB_o^T(\overline{q}, n, K) \xrightarrow{p} UB_o^{FP}(\mu, n, K)$.

Theorem 2

$$\lim_{n\to\infty} UB_o^{FP}(\mu, n, K) = (1 - S_p) \sum_{I=0}^{K-1} h(I) \binom{K-1}{I} \left[\mu^I (1-\mu)^{K-I} - \alpha a^I (1-a)^{K-I} - (1-\alpha) b^I (1-b)^{K-I}\right], \tag{9}$$

and

$$UB_o^{FP}(\overline{q}, n, K) \xrightarrow{p} \lim_{n\to\infty} UB_o^{FP}(\mu, n, K)$$

Like with the expected number of tests, I show that $UB_o^{FP}(\mu, n, K)$ converges to its limit from below, so that, for a given batch size N, expression (9) yields a more conservative (i.e., looser) upper bound to the reduction in the expected number of false positives that stems from implementing ordered pooling, as compared to $UB_o^{FP}(\mu, n, K)$.

Proposition 6 *For all $n \in \mathbb{N}$ we have that*

$$UB_o^{FP}(\mu, n, K) \leq \lim_{n\to\infty} UB_o^{FP}(\mu, n, K)$$

Like $UB_o^T(\mu, n, K)$, $UB_o^{FP}(\mu, n, K)$ oscillates and increases with n until it converges to its limit from below. The intuition for this pattern is the same as the one given for the expected number of tests: an increase in the batch size allows for a greater variability in the probabilities of infection, which, in turn, tends to increase the benefits of implementing ordered pooling. The oscillatory behavior of $UB_o^{FP}(\mu, n, K)$ comes from the fact that, for some values of n, one of the pools has an intermediate probability of infection, which causes $UB_o^{FP}(\mu, n, K)$ to be lower than its asymptotic limit. But as n increases, the effect that this single pool has on the total average dissipates.

Like Theorem 1, Theorem 2 is an asymptotic result: to ensure it holds in practical settings, the batch size must be sufficiently large. But I derive a result similar to Proposition 4 that shows that, in the absence of dilution effects, $\mathbb{E}_{\overline{q}}\big[UB_o^{FP}(\overline{q}, n, K)\big] \leq \lim_{n\to\infty} UB_o^{FP}(\mu, n, K)$ for every n, dispensing the necessity for the batch to be large.

Proposition 7 *Suppose that*

$$h(I, K) = \begin{cases} 1 - S_p, & \text{if } I = 0 \\ S_e, & \text{if } I > 0 \end{cases}.$$

Then,

$$\mathbb{E}_{\overline{q}}\big[UB_o^{FP}(\overline{q}, n, K)\big] \leq \lim_{n\to\infty} UB_o^{FP}(\mu, n, K) = \\ \big[(1 - S_p)S_e - (1 - S_p)^2\big]\big[\alpha(1 - a)^K + (1 - \alpha)(1 - b)^K - (1 - \mu)^K\big] \tag{10}$$

6 Expected Number of False Negatives

The societal costs of false negatives are often substantial, as an infected individual erroneously diagnosed as healthy may not end up receiving the necessary treatment for her recovery. Moreover, in the context of communicable diseases, such as sexually transmitted infections, an undiagnosed individual poses a risk of unwittingly transmitting the infection to others. So, minimizing the expected number of false negatives should be one of the main priorities of the tester. This section provides an upper bound to the reduction in the expected number of false negatives obtained by implementing ordered pooling as opposed to random pooling.

One can easily show that the expected number of false negatives per subject obtained under random pooling with pool sizes equal to K is (approximately) given by:

$$FN_r^K(\overline{q}, n, K) \equiv \overline{q}\Bigg\{\left[\sum_{I=0}^{K-1}(1 - h(I + 1))\binom{K-1}{I}\overline{q}^I(1 - \overline{q})^{K-1-I}\right] \\ + \left[\sum_{I=0}^{K-1} h(I + 1, K)\binom{K-1}{I}\overline{q}^I(1 - \overline{q})^{K-1-I}(1 - S_e)\right]\Bigg\}. \tag{11}$$

For any given subject, the first term inside the braces from equation (11) corresponds to the probability that no infection is detected in the pooled sample conditional that the subject *is* infected. The second term inside the braces corresponds to the probability that infection *is* detected in the pooled sample, but the subject is subsequently misclassified as not infected in the retesting phase. Adding these two expressions yields the probability that an arbitrary subject is incorrectly classified as not infected conditional that she is, in fact, infected. This probability must then be multiplied by $\overline{q}$ to generate the unconditional probability that a subject is incorrectly classified as not infected.

Rearranging the terms, equation (11) can be rewritten as

$$FN_r^K(\overline{q}, n, K) = \overline{q}\left[\sum_{I=0}^{K-1}(1 - S_e h(I + 1))\binom{K-1}{I}\overline{q}^I(1 - \overline{q})^{K-1-I}\right]. \tag{12}$$

It is straightforward to show that, for a given realization of $(q_1, q_2, \cdots, q_N)$, implementing ordered pooling yields the following expected number of false negatives:

$$FN_o^K(q_1, q_2, \cdots, q_N) \equiv \sum_{g=1}^{N/K} \text{fn}_g,$$

where

$$\text{fn}_g \equiv \sum_{I=0}^{K} P_{G_g}(I) I \left[1 - S_e h(I,K)\right] \tag{13}$$

is the expected number of false negatives within group G_g (i.e., the gth group with lowest probability of infection), and $P_{G_g}(I)$ is the probability that group G_g has exactly I infected subjects (see equation (4)).

The next proposition states that, for a given realization of the average probability of infection in a batch, $\overline{q}$, the expected number of false negatives obtained when implementing ordered pooling is minimized when the variability of the probabilities of infection is as high as possible between groups, and as low as possible within groups. That the expected number of false negatives diminishes with an increase in the variability of probabilities of infection between groups is intuitive: in the absence of any variability in the probabilities of infection, implementing ordered pooling would be equivalent to implementing random pooling. The intuition, however, as to why the variability of probabilities of infection must be as low as possible within each group is somewhat less intuitive, but it has to do with assumption 3. Indeed, if assumption 3 holds, then either the dilution function is completely flat for positive values of I (this would represent the case in which pooled testing is not affected by dilution effects), in which case the variability of the probabilities of infection within a group would not alter the expected number of false negatives; or the dilution function $h(I,K)$ is strictly increasing in I, but not "too concave" for positive values of I, which causes the dilution effect to be somewhat strong, in which case false negatives are reduced by reducing the variability of probabilities of infection within the pool. Indeed, consider, for example, one pool comprised of only two individuals, where one has probability of infection 1, and the other has probability of infection 0. In this case, if the dilution effect is sufficiently strong, having only one infected subject reduces the probability that the pooled test detects an infection, thus increasing the probability of a false negative. But if both subjects had probability of infection equal to 1/2, the average probability of infection within the pool would be preserved, and yet, there would be a positive probability of 1/4 that both subjects were infected, in which case the pooled test would be less likely to fail to detect an infection, and there would also be a probability of 1/4 that none was infected, an event that yields zero probability of a false negative.

Proposition 8 *Let $(q_1, \cdots, q_N)$ be such that $\sum_i q_i/N = \overline{q}$ and $q_i \in [a,b]$ for all i. Then, if $h(I,K)$ is increasing in I and satisfies assumption 3, we must have $FN_o^K(q(\overline{q},n,K)) \leq FN_o^K(q_1,\cdots,q_N)$.*

For each $(\overline{q},n,K) \in [a,b] \times \mathbb{N}^2$ define

$$UB_o^{FN}(\overline{q},n,K) \equiv \frac{FN_r^K(\mu,n,K) - FN_o^K(q(\overline{q},n,K))}{N}.$$

In words, $UB_o^{FN}(\overline{q},n,K)$ corresponds to an upper bound to the reduction in the expected number of false negatives that one can get by implementing ordered pooling as opposed to random pooling, when assuming the average probability of infection from the batch is equal to $\overline{q}$. The following theorem ensures that, provided that n is sufficiently large, one can approximate $UB_o^{FP}(\overline{q},n,K)$ by $UB_o^{FP}(\mu,n,K)$.

Theorem 3

$$\begin{aligned}\lim_{n\to\infty} UB_o^{FN}(\mu,n,K) = \sum_{I=0}^{K-1} (1 - S_e h(I+1)) \binom{K-1}{I} [\mu^{I+1}(1-\mu)^{K-1-I} \\ - \alpha a^{I+1}(1-a)^{K-1-I} - (1-\alpha) b^{I+1}(1-b)^{K-1-I}],\end{aligned} \tag{14}$$

and

$$UB_o^{FN}(\overline{q},n,K) \xrightarrow{p} \lim_{n\to\infty} UB_o^{FN}(\mu,n,K)$$

Like with the expected number of tests and the expected number of false positives, I show that $UB_o^{FN}(\mu,n,K)$ converges to its limit from below, so that, for a given batch size N, expression (14) yields a more conservative upper bound to the reduction in the expected number of false negatives that stems from implementing ordered pooling, as compared to $UB_o^{FN}(\mu,n,K)$.

Proposition 9 *For all $n \in \mathbb{N}$ we have that*

$$UB_o^{FN}(\mu, n, K) \leq \lim_{n\to\infty} UB_o^{FN}(\mu, n, K)$$

$UB_o^{FN}(\mu, n, K)$ exhibits the same pattern as $UB_o^T(\mu, n, K)$ and $UB_o^T(\mu, n, K)$: it oscillates and increases with n until it converges to its limit from below. The intuition for this pattern is the same as the ones presented in the previous sections: the oscillatory behavior of $UB_o^{FN}(\mu, n, K)$ stems from the discrete nature of pool sizes, whereas the increasing trend in $UB_o^{FN}(\mu, n, K)$ comes from the fact that, when the batch size is too small, in expectation it does not end up with enough subjects with a high probability of infection to fill out an entire group (or subjects with a very low probability of infection who can fill out an entire group), thus diminishing the benefits of implementing ordered pooling.

It should be noticed that, in the absence of dilution effects, the expected number of false negatives is not affected by how agents are matched to form the pools (e.g., see Aprahamian, Bish and Bish [2019]). This is because, when there are no dilution effects, the likelihood of detecting an infected sample in a pooled test is not influenced by how many other infected specimens there are in the pool. Therefore, when the dilution effect is given by equation (2), $UB_o^{FN}(\overline{q}, n, K)$ collapses to 0.

Proposition 10 *[Aprahamian, Bish and Bish, 2019] Suppose that*

$$h(I, K) = \begin{cases} 1 - S_p, & \text{if } I = 0 \\ S_e, & \text{if } I > 0 \end{cases}.$$

Then,

$$UB_o^{FN}(\overline{q}, n, K) = 0$$

for every $\overline{q}$.

7 Upper bounds with partial information on μ, a, b, S_e and S_p

The previous sections derived asymptotic upper bounds to the benefits of implementing ordered pooling, assuming that information on the maximum and minimum probabilities of infection, a and b, were available. In practice, however, this information may not be available. In these cases, the tester can set very low values for a and very high values of b to get a conservative (i.e., loose) upper bounds to the benefits of implementing ordered pooling. Indeed, for a given μ reducing a or increasing b should either increase or not affect all of the upper bounds $UB_o^T(\mu, n, K)$, $UB_o^{FP}(\mu, n, K)$ and $UB_o^{FN}(\mu, n, K)$.

Corollary 1 *For a given μ, $UB_o^T(\mu, n, K)$, $UB_o^{FP}(\mu, n, K)$ and $UB_o^{FN}(\mu, n, K)$ are non-increasing in "a" and non-decreasing in "b".*

Similarly, there might be instances in which the tester does not know the exact prevalence of the disease, μ. In this case, the tester can still get a conservative upper bound by maximizing equations (5), (9) and (14) with respect to μ, or a linear combination of these equations. Optimizing these expressions with respect to μ can be easily done computationally.

Regardless of the dilution effect, the benefits of implementing ordered pooling will usually have an inverted U-shape relationship with respect to μ. This is because, at the extreme levels, i.e., when $\mu = a$ or $\mu = b$, everyone has the same probability of infection, in which case there is no distinction between random and ordered pooling. It will be shown in proposition 11 that, in the absence of dilution effects (i.e., when the dilution function is given by equation (2)), one can actually get a simple closed form solution to the level of μ that maximizes the benefits of implementing ordered pooling.

In the absence of dilution effects, I show that the difference in the expected number of tests of random vs. ordered pooling will increase with S_e and S_p. Intuitively, this happens because, if S_e is very low, most pools are not retested, regardless of how samples were grouped, in which case the monetary benefits of arranging samples optimally erode. On the other hand, if S_p is very low, we end up with a lot of retests caused by false positives, regardless of how samples are pooled.

As to false positives, increasing S_e will increase the gap of false positives between random and ordered pooling, while increasing S_p will tend to reduce this gap. Intuitively, a lower S_e implies that virtually no one is retested, so we end up with virtually no false positives, regardless of how samples are pooled. On the other hand, at the maximum level of specificity, i.e., when $S_p = 1$, we cannot have any false positive result, regardless of how samples are pooled. Though it is theoretically possible that a marginal increase in S_p may actually increase the difference in false positives of random vs. ordered pooling, this only happens when $2(1 - S_p) > S_e$, an instance that is arguably uncommon as it requires the specificity of the test, S_p, to be very low.

When it comes to false negatives, it has already been shown in proposition 10 that, in the absence of dilution effects, the expected number of false negatives is not influenced by how agents are matched to form the pools. This implies that $UB_o^{FN}(\overline{q}, n, K)$ is always equal to zero regardless of S_e, S_p, a, b or μ.

Proposition 11 *Suppose that*

$$h(I,K) = \begin{cases} 1 - S_p, & \text{if } I = 0 \\ S_e, & \text{if } I > 0 \end{cases}.$$

Then, $\lim_{n\to\infty} UB_o^T(\mu, n, K)$ *is increasing in* S_e *and* S_p*, and* $\lim_{n\to\infty} UB_o^{FP}(\mu, n, K)$ *is increasing in* S_e*. In addition,* $\lim_{n\to\infty} UB_o^T(\mu, n, K)$ *and* $\lim_{n\to\infty} UB_o^{FP}(\mu, n, K)$ *are single peaked with respect to* μ*, achieving a maximum at*

$$\mu^* = 1 - \left[\frac{(1-a)^K - (1-b)^K}{K(b-a)}\right]^{1/(K-1)}.$$

If we add the assumption that $2(1 - S_p) < S_e$*, then* $\lim_{n\to\infty} UB_o^{FP}(\mu, n, K)$ *is decreasing in* S_p*.*

Because $UB_o^{FN}(\overline{q}, n, K)$ *is always equal to zero in the absence of dilution effects,* $UB_o^{FN}(\overline{q}, n, K)$ *is not affected by changes on any of the parameters* S_e*,* S_p*,* a*,* b *or* μ*.*

Example 3 *If*

$$h(I,K) = \begin{cases} 1 - S_p, & \text{if } I = 0 \\ S_e, & \text{if } I > 0 \end{cases},$$

then, conservatively setting $a = 0$ *and* $b = 1$ *and*

$$\mu = 1 - \left[\frac{(1-a)^K - (1-b)^K}{K(b-a)}\right]^{1/(K-1)} = 1 - \left[\frac{1}{K}\right]^{1/(K-1)},$$

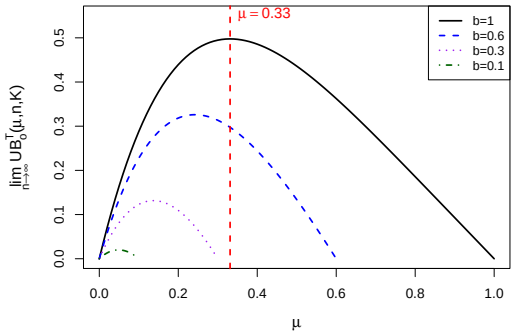
and assuming that $S_p = .98$*,* $S_e = .95$ *and* $K = 5$*, we have that ordered pooling can reduce the expected number of tests by at most 0.5 per person when samples are pooled into groups of* $K = 5$*. Of course, this upper bound is extremely conservative, as it assumes* $\mu \approx 0.33$*, an instance in which the prevalence is so high that pooled testing would actually be less cost-effective than simply testing every specimen individually.*

But if we assumed b *and* μ *to be smaller, we could get much tighter bounds, as displayed in figure 2a. Similarly, assuming a smaller* S_e *(or* S_p*) would also generate tighter upper bounds, as displayed in figure 2b. As to false positives, increasing* S_p *actually reduces the spread between random and ordered pooling, as displayed in figure 2c.*

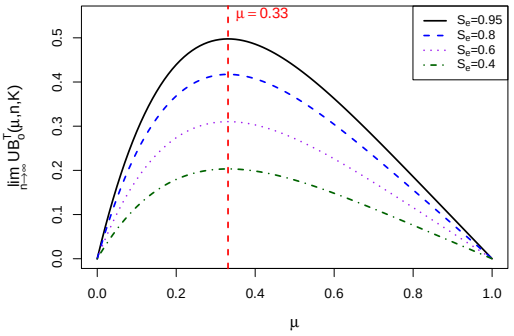
8 Choosing the optimal pool size

In practical settings, a tester usually wishes to implement a pool size K that minimizes a linear combination of the expected number of tests, the expected number of false positives, and the expected number of false negatives. Because the set of possible pool sizes is finite, solving this problem is relatively simple when all pools are restricted to have the same size.[8] This section provides a brief characterization of the optimal pool size as a function of the weights put to each of these attributes.

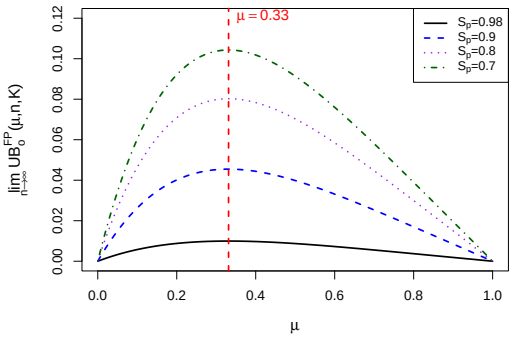
[8]When pools are allowed to have heterogeneous sizes, an optimal path algorithm described in Aprahamian, Bish and Bish [2019] and Saraiva [2023*a*] can be used to find the optimal matching configuration. One of the features from the Shiny app `https://jcbrnd.shinyapps.io/shiny_app_anonymous/` developed as a companion material to this paper implements this algorithm to find the optimal pool configuration.



(a) $\lim_{n\to\infty} UB_o^T$ as a function of μ and b.



(b) $\lim_{n\to\infty} UB_o^T$ as a function of μ and S_e.



(c) $\lim_{n\to\infty} UB_o^{FP}$ as a function of μ and S_p.

Figure 2: The benefits associated with ordered pooling as a function of μ, b, S_e and S_p. For the baseline, I assume $S_p = 0.98$, $S_e = 0.95$, $a = 0$, $b = 1$ and that the dilution function is given by equation (2).

It is well known, since Dorfman [1943], that when one's interest is in minimizing the expected number of tests, the optimal pool size is decreasing in the prevalence of the disease. If the prevalence is sufficiently high, it becomes optimal to just test every sample individually.

The same principle applies when implementing ordered pooling: under smaller prevalences, it becomes optimal to choose a larger pool size. If pooled testing is subject to dilution effects and the tester does not care about classification errors, it becomes even more advantageous to choose larger pool sizes, as doing so reduces the probability that an infected specimen in a pool triggers a retest.

So, if we only cared about the expected number of tests, and strong dilution effects were at play, and the prevalence of the disease was sufficiently low, it would be optimal to just group every sample from the batch into a single pool. Even if there were infected specimens in the pool, there would most likely be very few of them, so that they would not trigger retests due to dilution effects. Such a pooled test would be worthless, of course, as it would invariably generate negative results. For this reason, it is important for the tester to also take into account how the pool size affects classification errors.

At the extreme case in which one's sole purpose is to minimize false negative results, individual testing becomes optimal [Saraiva, 2023*a*]. If individual testing is not allowed and pooled testing is subject to dilution effects, smaller pools tend to reduce the expected number of false negative results (but if pooled testing is not subject to dilution effects, the way that the pools are formed does not affect the incidence of false negatives, as seen in proposition 10.). Indeed, conditional that a subject in a pool of size K is infected, the expected number of infected specimens in the pool is given by $\frac{1+(K-1)\mu}{K}$ (assuming random pooling is implemented), which is decreasing in K. So, conditional that a sample is infected, increasing the pool size decreases the

average concentration of infection in the pool, increasing the probability that the pooled test does not detect an infection. This is arguably the main reason why small pool sizes are often chosen in practical settings.

That reducing the pool size tends to generate fewer false negative results is rather intuitive. What is somewhat less obvious is that smaller pool sizes also tend to reduce the expected number of false positive results. Indeed, in the next proposition, I show that, when pooled testing is not subject to dilution effects, pools of size $K = 2$ minimize the expected number of false positive results.[9]

Proposition 12 *Suppose that*

$$h(I,K) = \begin{cases} 1 - S_p, & \text{if } I = 0 \\ S_e, & \text{if } I > 0 \end{cases},$$

and the batch size N is even. Then, setting $K = 2$ minimizes the expected number of false positives, both when implementing random pooling and when implementing ordered pooling.

The reason why proposition 12 may seem counterintuitive is that one might expect that, by minimizing the expected number of tests, one is automatically minimizing the expected number of false positives. But that is usually not the case, as minimization of the expected number of tests aims to minimize both pooled tests and retests, whereas the minimization of false positives only takes retests into account. When probabilities of infection are small, larger pool sizes tend to reduce the expected number of tests *plus* retests. However, as we increase pool sizes, we also increase the probability that a pool has at least one infected specimen. If dilution effects are negligible, this implies that increasing the pool size increases the probability of retests, thus increasing the probability that false positive results are generated.

Therefore, another reason why a tester may prefer to set smaller pool sizes is that doing so tends to minimize not only false negative results but false positive results as well. But as will be shown in the numerical computations from the case studies, when pool sizes are small, the benefits per subject of implementing ordered pooling also tend to be small.

9 Case study: Chlamydia Screening in the US

According to the CDC, in 2021 alone, 1,644,416 new cases of Chlamydia were reported in the US, making it the most reported sexually transmitted disease for that year countrywide [Louisiana Department of Health, 2021]. Chlamydia mainly affects the young and sexually active, and symptomatic cases are more frequent among women. Though in general, it can be cured with antibiotics, if left untreated, the disease can cause severe sequelae in women, including Pelvic Inflammatory Disease (PID), ectopic pregnancy, and infertility [Centers for Disease Control and Prevention, 2019].

This section analyzes the potential benefits of implementing ordered pooling to detect Chlamydia through Dorfman testing, as opposed to matching subjects randomly. So, letting $C(T)$, $C(FP)$ and $C(FN)$ denote the expected cost of a test, the expected cost of a false positive, and the expected cost of a false negative, respectively, we wish to compute

$$\begin{aligned} \Delta C \equiv\ & C(T)\left(T_r^K - T_o^K\right) + C(FP)\left(FP_r^K - FP_o^K\right) \\ & + C(FN)\left(FN_r^K - FN_o^K\right), \end{aligned} \tag{15}$$

and use the results presented earlier to derive an upper bound to this difference in costs.

For this analysis, I consider the Ligase Chain Reaction (LCR) chlamydia test, a test commonly used in chlamydia screening [Aprahamian, Bish and Bish, 2018]. Following Aprahamian, Bish and Bish [2019], I use the average between the cost of sequelae for men and women, estimated by Owusu-Edusei Jr et al. [2015], to set the cost of a false negative at $C(FN) = \$2,927$.[10] Also following Aprahamian, Bish and Bish [2019], I set the screening cost per test at $C(T) = \$55$, and the cost of a false positive to be equal to the cost of an additional screening test (i.e., I assume $C(FP) = \$55$).

[9] The part of proposition 12 that deals with ordered pooling is an extension of a result derived in Aprahamian, Bish and Bish [2019]. The authors show that, when one's sole objective is to minimize false positives, one cannot have pools of size greater than 3. But I show that, when the batch size is even, it will never be optimal to have a pool of size 3 or 1.

[10] For simplification, this measure does not incorporate a health-related reduction in quality of life, nor costs associated with transmissions that a true positive would have averted.

9.1 The dilution effect for LCR chlamydia tests

Based on the framework proposed by Aprahamian, Bish and Bish [2018], I assume that the dilution function for Chlamydia has the following format:

$$h(I,K) = (1 - S_p) + (S_e + S_p - 1)(I/K)^{\delta}. \tag{16}$$

Clinical studies by Kacena et al. [1998] indicated that pools of size $k = 4$ achieved near-perfect sensitivity and specificity, specifically:

$$h(1,4) = 1,$$

and

$$1 - h(0,4) = 97/99 = 0.98.$$

This leads me to conjecture that $S_e = h(1,1) \approx h(1,4) = 1$ and $S_p = 1 - h(0,1) \approx 1 - h(0,4) = 97/99$. Nevertheless, by contrasting $S_e = 1$ with the findings from studies like Schachter et al. [1994], I suspect that this perfect sensitivity could be due to sampling error; so, I conservatively set $S_e = 0.99$. Assuming constant specificity across pool sizes, I equate $h(0,1)$ to $h(0,4)$, thus setting $S_p = 1 - h(0,4) = 97/99$.

In Kacena et al. [1998], Chlamydia's prevalence for pools of size $K = 10$ was $\mu_r = 62/520$, and the detection rate in such infected pools was 37/38. Adopting the approach of Aprahamian, Bish and Bish [2018], δ is chosen so that

$$37/38 = \frac{1}{1 - (1 - \mu_r)^{10}} \sum_{I=1}^{10} h(I,10) \binom{10}{I} \mu_r^I (1 - \mu_r)^{10-I}, \tag{17}$$

which aligns the sensitivity for pools of size 10 from Kacena et al. [1998] with the dilution function h. This calculation results in $\delta = 0.0089$, suggesting minimal dilution effects. However, considering the potential sampling variability in Kacena et al. [1998], I opt for a higher δ value, more specifically, I assume $\delta = 0.15$, allowing for the possibility of non-trivial dilution effects (in the Online Appendix I show that very similar results are obtained when $\delta = 0.0089$).

9.2 Distribution of the probabilities of infection for Chlamydia

Table 2 reports the prevalence per demographic group extracted from the Centers for Disease and Control Prevention (CDC) website for the year 2014.[11] As in Aprahamian, Bish and Bish [2019], Aprahamian, Bish and Bish [2018] and Saraiva [2023*a*], the probabilities of infection from each group have been multiplied by an underreporting factor of 3, to account for patients who are not properly screened (e.g., because their infection is asymptomatic). Indeed, According to Centers for Disease Control and Prevention [2000], approximately 75% of infected women and 50% of infected men exhibit no symptoms.

The multiplication of the probabilities of infection by a somewhat arbitrary underreporting factor of 3 reflects the lack of information that researchers usually have regarding the true distribution of probabilities of infection in practical settings. This is one of the reasons why the upper bounds derived in this article can be useful: even if one does not know the true distribution of probabilities of infection, one can still simulate the benefits of implementing ordered pooling by assuming conservative values for a, b and μ. In the present example, increasing the underreporting factor from 1 to 3 substantially increases both b and μ, which increases the maximum benefits that can be achieved by implementing ordered pooling (see the Online Appendix for detail).

Figure 3 depicts the distribution of probabilities of infection generated from table 2 assuming an underreporting factor of 3. An advantage of working with this distribution is that, because it corresponds to the same distribution analyzed in Aprahamian, Bish and Bish [2019] and Saraiva [2023*a*], one can compare my numerical results with the ones obtained in these articles.

[11] https://wonder.cdc.gov/std-race-age.html

Table 2: Prevalence of Chlamydia and proportion in population by Gender, Age, and Race/Ethnicity (Centers for Disease Control and Prevention 2014).

Gender	Race/Ethnicity	Age Group (years)	Prevalence (%)	Demographic Composition (%)
Female	Hispanic	15-24	6.54	1.41
		Other	0.65	7.01
	Black	15-24	19.19	1.07
		Other	1.22	5.67
	Other	15-24	4.38	4.29
		Other	0.25	31.31
Male	Hispanic	15-24	1.78	1.53
		Other	0.36	7.16
	Black	15-24	7.45	1.09
		Other	1.05	5.08
	Other	15-24	1.20	4.51
		Other	0.17	29.87
Total	-	-	0.97	100

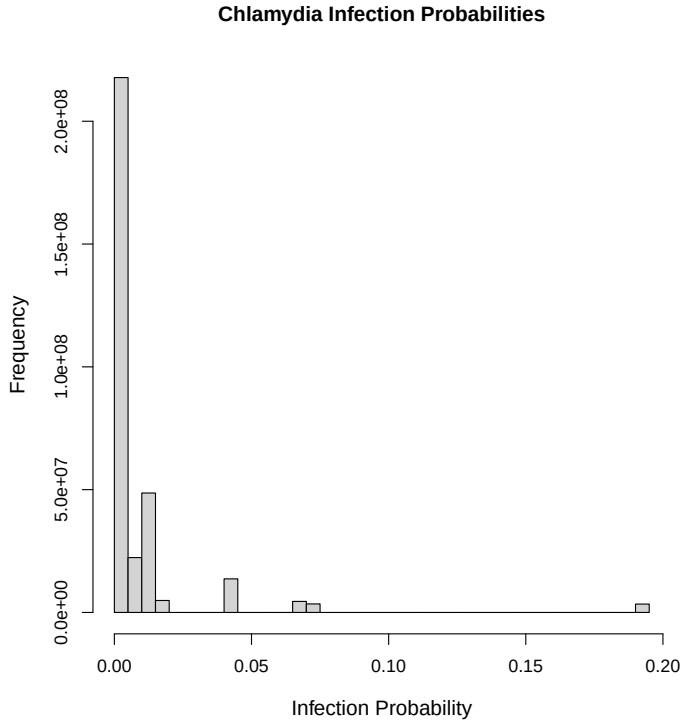


Figure 3: Distribution of probabilities of Chlamydia in the US during 2014 using the data from table 2.

9.3 Benefits of implementing ordered pooling to detect Chlamydia

It can be shown that the calibrated dilution function from equation (16) is concave for any $\delta \in (0, 1]$ and any $S_p, S_e \in [0, 1]$ such that $S_e > 1 - S_p$. Moreover, for any $\delta \geq 0$, this dilution function satisfies assumption 3. Therefore, it follows from proposition 1 that ordered pooling simultaneously minimizes the expected number of tests, the expected number of false positives and the expected number of false negatives.

Figure 4 depicts the simulated benefits of implementing ordered pooling as opposed to random pooling, as well as the upper bounds $UB_o^T(\mu, n, K)$, $UB_o^{FP}(\mu, n, K)$ and $UB_o^{FN}(\mu, n, K)$ with their respective asymptotic limits. To generate these plots, I assumed a batch size of $N = 100$, which has a similar order of magnitude as the batch size used in some practical settings.[12] Whenever N was not divisible by K, I would redefine $N = \lfloor \frac{N}{K} \rfloor \times K$ to ensure that all pools had the same size. This is equivalent to randomly selecting some samples to be assigned to the smaller batch of size N mod K.

From Figure 4a, one can see that the pool size that minimizes the expected number of tests is relatively large: $K = 15$ under random pooling and $K = 16$ under ordered pooling. As discussed in section 8, this happens because the prevalence of the disease is relatively low (less than 1%). Also notice that, consistent with Proposition 12, the pool size that minimizes the expected number of false positives is given by $K = 2$, both when implementing random pooling and when implementing ordered pooling. As the pool size increases, both types of classification errors increase. Taking into account the costs associated with classification errors, the overall costs associated with pooled testing is minimized at $K = 10$ when implementing random pooling and at $K = 11$ when implementing ordered pooling.

Figure 5 zooms in the benefits of implementing ordered pooling displayed in Figure 4. From these figures, one can see that, for batches of size 100, the savings in costs associated with ordered pooling are somewhat modest (less than 1 USD per subject for pool sizes lower than $K = 12$), which is both captured by the simulated differences and the upper bounds $UB_o^T(\mu, n, K)$, $UB_o^{FP}(\mu, n, K)$ and $UB_o^{FN}(\mu, n, K)$. The asymptotic upper bounds, on the other hand, are substantially higher. This is because, as the batch size increases, the tester is able to form pools comprised entirely of risky subjects.

Indeed, as figure 6 illustrates, the benefits of implementing ordered pooling are substantially higher when $N = 300$ (the total cost reduction per subject surpasses 1 USD for pool sizes greater than 8), and the upper bounds $UB_o^T(\mu, n, K)$, $UB_o^{FP}(\mu, n, K)$ and $UB_o^{FN}(\mu, n, K)$ are much closer to their asymptotic limit. It should also be noticed that the simulated benefits of implementing ordered pooling depicted in Figure 6 are almost indistinguishable from the ones obtained in Saraiva [2023*a*] using a very large batch size of $N = 10,000$.

10 Case study: SARS-CoV-2 Screening in Chile

Throughout the COVID-19 pandemic, Chile relied heavily on pooled testing to detect SARS-CoV-2, the pathogen responsible for COVID-19. According to Basso et al. [2023], between 10% and 20% of reverse transcription polymerase chain reaction (RT-PCR) tests conducted in the country used this technique. This section estimates the potential gains of implementing ordered pooling as opposed to random pooling to detect SARS-CoV-2 in Chile during the beginning of 2022, a period characterized by an abnormally high number of new infections.

10.1 The dilution effect for RT-qPCR tests

First, let us start by first estimating the dilution effect for COVID-19 tests. Currently, RT-PCR tests are the most ubiquitous method used to detect the disease due to their high sensitivity and the high speed with which they can generate final results [Habibzadeh et al., 2021]. RT-qPCR (Reverse Transcription Quantitative PCR) is similar in nature to RT-PCR tests. However, it not only provides a classification of infected vs. not infected but also provides a quantitative measure of the viral load in the sample.

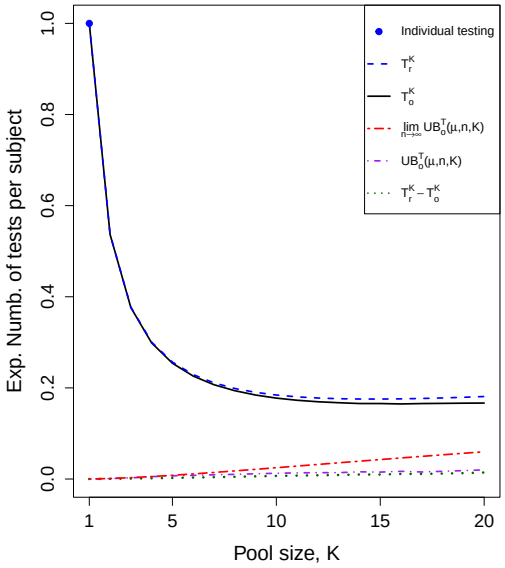
Yelin et al. [2020] conducted clinical studies to determine the likelihood of accurately detecting SARS-CoV-2 in samples using RT-qPCR tests at various dilution levels. Table 3 presents the likelihood of detecting an infection as a function of how strongly the sample was diluted.[13]

As studies show that the specificity from RT-PCR tests to detect SARS-CoV-2 are usually close to 100%

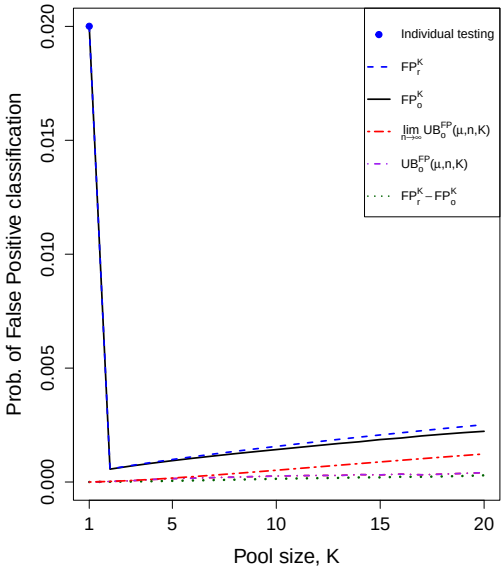
[12] For example, according to Lewis, Lockary and Kobic [2012], the Idaho Bureau of Laboratories used to test approximately 40 specimens per day to detect sexually transmitted diseases such as Chlamydia (15,000 per year), implementing pooled testing on most occasions.

[13] These estimates were recovered from a plot published in the online appendix of Yelin et al. [2020] assuming a cutoff of 38 polymerase chain reaction (PCR) cycles.

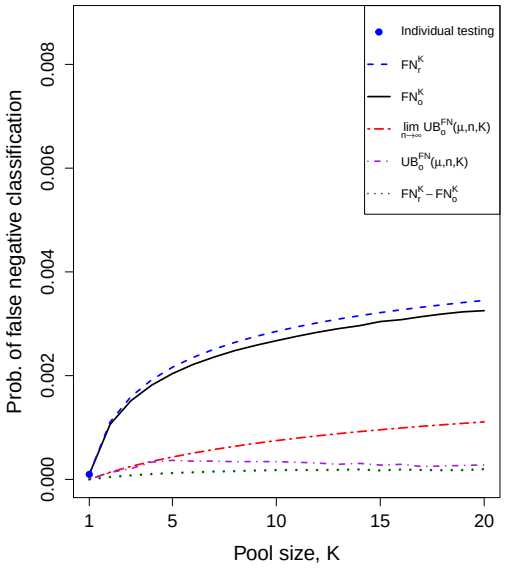
(a) $T_r^K - T_o^K$



(b) $FP_r^K - FP_o^K$



(c) $FN_r^K - FN_o^K$



(d) ΔC

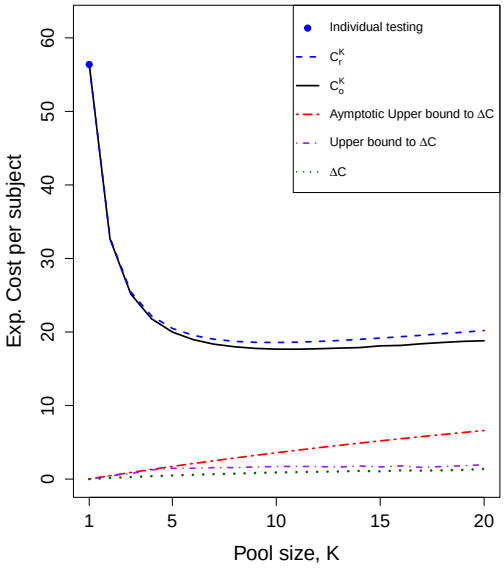


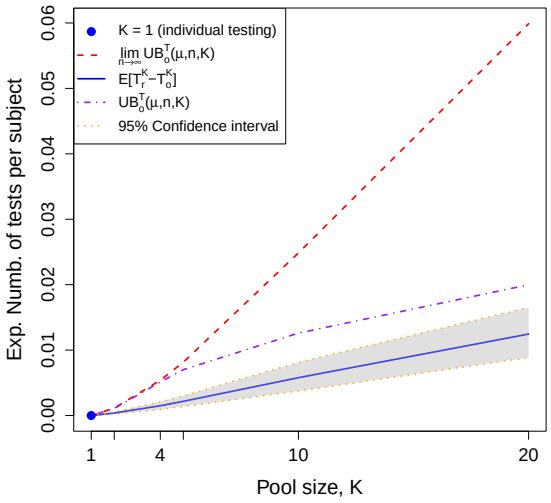
Figure 4: Expected number of tests and classification errors per subject for Chlamydia, assuming $N = 100$. The blue dot corresponds to the case in which subjects are individually tested, i.e. $K = 1$. Parameters for the dilution function (16): $S_e = .99$, $S_p = .98$ and $\delta = 0.15$. Parameters for the upper bounds: $\mu = 0.0097$, $a = 0.0017$ and $b = 0.1919$.

[Litchfield et al., 2022], I set $S_p = 0.99$. Assuming that each observed $h(I, K)$ is given by

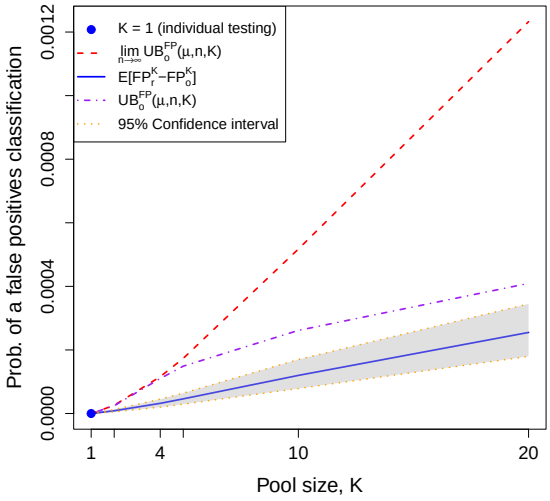
$$(1 - S_p) + (S_p + S_e - 1)\left(\frac{I}{K}\right)^{\delta}$$

plus a random idiosyncratic shock $\varepsilon \sim N(0, \sigma^2)$, one can easily estimate S_e and δ through the method of

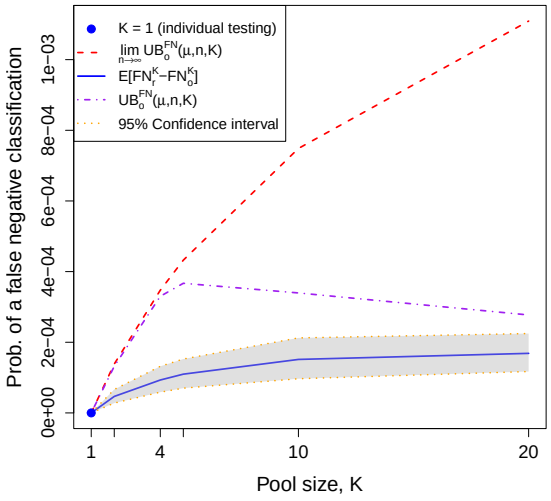
(a) $T_r^K - T_o^K$



(b) $FP_r^K - FP_o^K$



(c) $FN_r^K - FN_o^K$



(d) ΔC

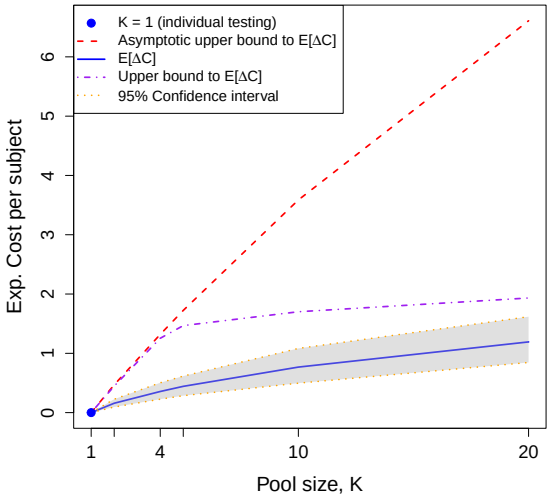
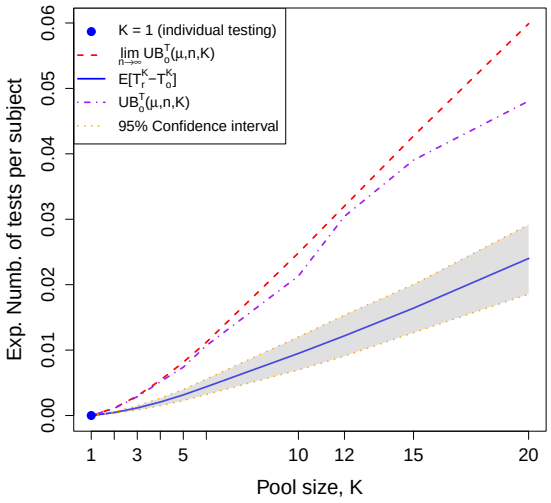


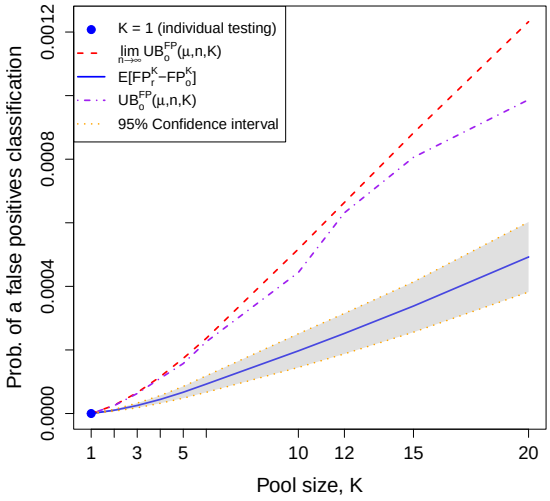
Figure 5: Expected benefits per subject of implementing ordered pooling for Chlamydia, assuming $N = 100$. The gray area corresponds to a 95% confidence interval using non-parametric bootstrap and assuming that 10 separate batches are tested in total. Parameters for the dilution function (16): $S_e = .99$, $S_p = .98$ and $\delta = 0.15$. Parameters for the upper bounds: $\mu = 0.0097$, $a = 0.0017$ and $b = 0.1919$.

maximum likelihood to obtain $S_e = 0.989$ and $\delta = 0.0459$. Because $\delta = 0.0459 \in [0, 1]$, the calibrated dilution function is increasing and satisfies assumptions 2 and 3, which, from proposition 1, implies that ordered pooling simultaneously minimizes the expected number of tests, the expected number of false positives and the expected number of false negatives. Figure 7 depicts the estimated dilution function.

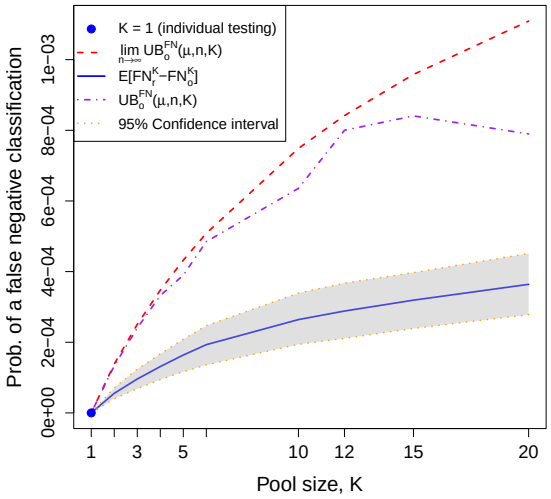
(a) $T_r^K - T_o^K$



(b) $FP_r^K - FP_o^K$



(c) $FN_r^K - FN_o^K$



(d) ΔC

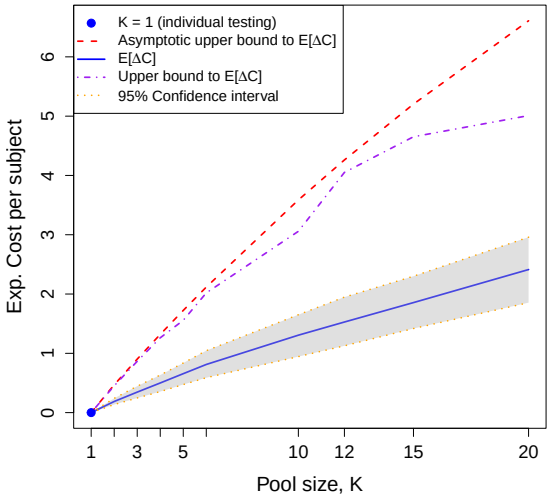


Figure 6: Expected benefits per subject of implementing ordered pooling for Chlamydia, assuming $N = 100$. The gray area corresponds to a 95% confidence interval using non-parametric bootstrap and assuming that 10 separate batches are tested in total. Parameters for the dilution function (16): $S_e = .99$, $S_p = .98$ and $\delta = 0.15$. Parameters for the upper bounds: $\mu = 0.0097$, $a = 0.0017$ and $b = 0.1919$.

10.2 Distribution of the probabilities of infection for SARS-CoV-2

During the COVID-19 pandemic, the Ministry of Science from Chile implemented strict lockdown policies which relied heavily on surveillance data, which were made publicly available [Basso et al., 2023]. This dataset contains the weekly incidence of SARS-CoV-2 for different combinations of age cohort and vaccination

Table 3: Sensitivity of RT-qPCR tests extracted from Yelin et al. [2020] under various dilution levels.

I/K	$h(I,K)$
1	0.9623
1/2	0.9623
1/4	0.9623
1/8	0.9208
1/16	0.8472
1/32	0.8472
1/64	0.8094

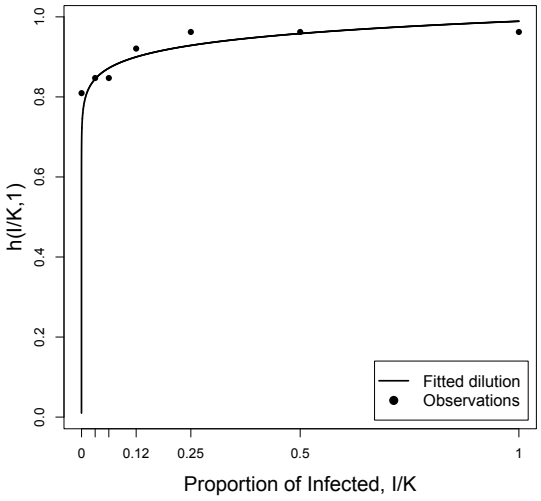


Figure 7: Dilution function $h(I,K) = (1 - S_p) + (S_p + S_e - 1)\left(\frac{I}{K}\right)^{\delta}$, with $S_p = .99$, $S_e = 0.989$ and $\delta = 0.0459$. The dots correspond to the sample observations displayed in table 3.

statuses. The vaccination status indicates whether subjects from the group were: 1) not vaccinated, 2) had received only the first dose, 3) were fully vaccinated, 4) were fully vaccinated and received one vaccine boost, and 5) were fully vaccinated and received two vaccine boosts. The dataset does not specify, however, the type of vaccine received. Groups were also categorized based on the age cohort that they belonged, with 10 age cohorts in total. It should be noticed that this dataset is already aggregated at the vaccination status and age cohort. So, I cannot incorporate additional information from patients, such as the commune where they lived, to get more granular estimates of probabilities of infection.

Using this dataset, I computed the distribution of probabilities of SARS-CoV-2 infection in Chile during the 5th week of 2022. I focus on this week because this is the week from the sample that had the highest number of new confirmed cases. So, the benefits of implementing ordered pooling will tend be higher during this period.

As I have data on *incidence* (i.e., number of *new* infections in a given week) and because the symptomatic phase of the disease can last more than a week, and because many infected individuals often go undetected, I multiply the number of new cases in this week by 2 to get an estimate of the *prevalence* per group during this week.[14]

[14]If, instead of multiplying the incidence by 2, I added the current incidence level with the one from the previous period, I would get a slightly lower prevalence at the peak ($\mu = 0.023$ instead of $\mu = 0.025$) and practically the same bounds on probabilities of infection ($a = 0$ and $b = 0.048$).

Overall, I find that in this week there is considerable dispersion in the probabilities of infection per group, as illustrated in the histogram displayed in figure 8, though there is not as much dispersion as in the chlamydia case study. The average probability of infection during this period was 0.0251. The maximum probability of infection from this distribution was 0.048, and the minimum was 0, with variance equal to 5.039e-05.

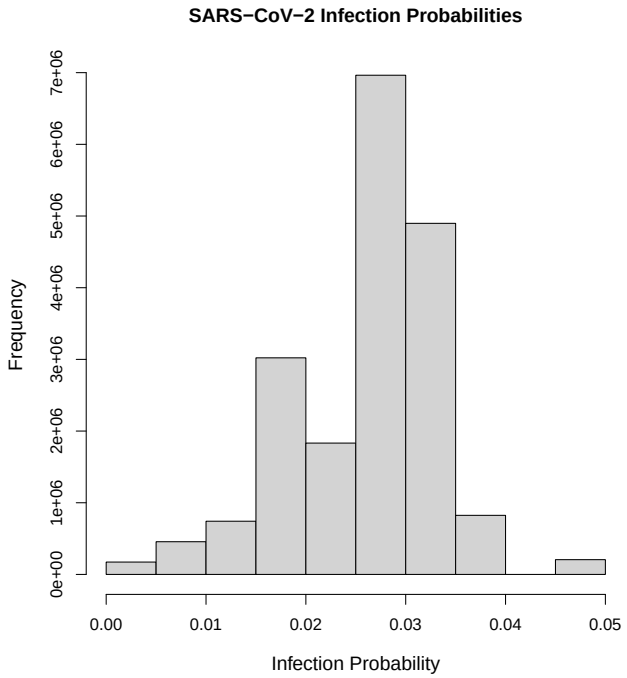


Figure 8: The distribution of probabilities of SARS-CoV-2 infection in Chile during the 5th week of 2022. Prevalence per group was obtained by multiplying the incidence (number of new cases) by two.

10.3 Benefits of implementing ordered pooling to detect COVID-19

Now the estimated dilution function

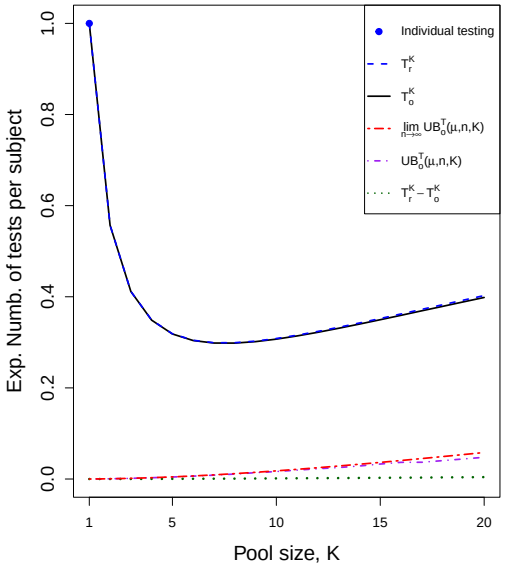
$$h(I, K) = (1 - 0.99) + (0.99 + 0.989 - 1)\left(\frac{I}{K}\right)^{0.0459}$$

can be used in conjunction with the distribution of probabilities of infection to estimate the benefits of implementing ordered pooling as opposed to random pooling, as well as the upper bounds to these benefits by computing $UB_o^T(\mu, n, K)$, $UB_o^{FP}(\mu, n, K)$ and $UB_o^{FN}(\mu, n, K)$. These benefits are displayed in figure 9. These plots were generated using a batch of size $N = 100$, which has a similar order of magnitude as the batch size used in some practical settings – for example, the Hadassah Medical Center in Jerusalem implemented Dorfman testing in batches of size 64 using a pool size of $K = 8$ to detect SARS-CoV-2 using RT-PCR tests [Barak et al., 2021].

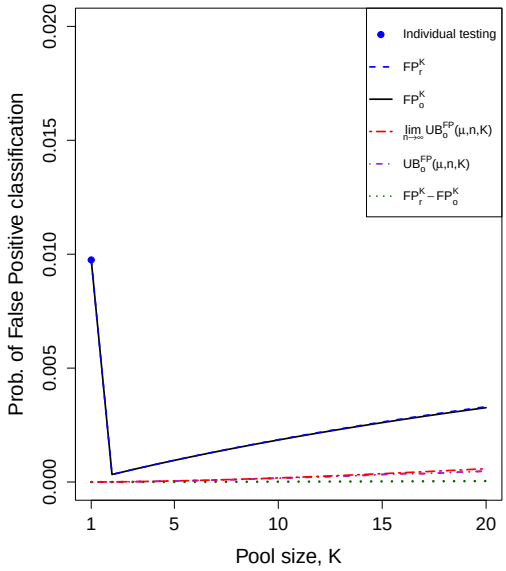
Though I do not estimate the cost of a false negative and the cost of a false positive due to the complexity involved in measuring these volatile costs, from the plots displayed in figure 9, one can see that the benefits of implementing ordered pooling are quite small for each of the three attributes considered. Moreover, if I follow Basso et al. [2023] and assume the average cost of a RT-PCR test to be 31.25 USD, then the savings in costs associated exclusively with the expected number of tests is always below 0.1 USD per subject for pool sizes less than or equal to 13.

For pools of size $K \leq 6$, the small gains associated with implementing ordered pooling are captured by the upper bounds $UB_o^T(\mu, n, K)$, $UB_o^{FP}(\mu, n, K)$ and $UB_o^{FN}(\mu, n, K)$ displayed in figure 9. Given that Chile

(a) $T_r^K - T_o^K$



(b) $FP_r^K - FP_o^K$



(c) $FN_r^K - FN_o^K$

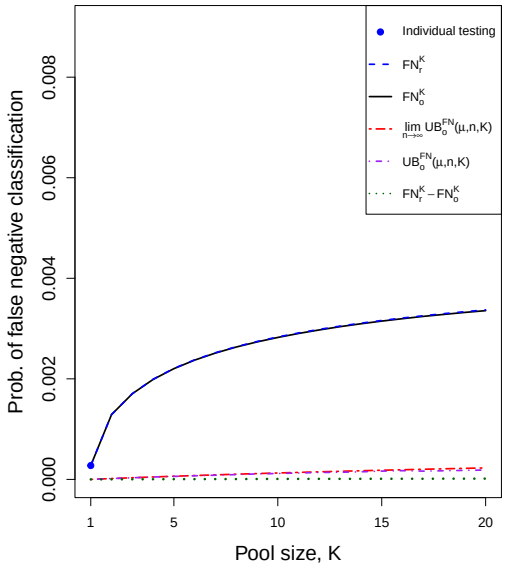


Figure 9: Expected number of tests and classification errors per subject for COVID-19, assuming $N = 100$. The blue dot corresponds to the case in which subjects are individually tested, i.e. $K = 1$. Parameters for the dilution function (16): $S_e = .99$, $S_p = .989$ and $\delta = 0.0459$. Parameters for the upper bounds: $\mu = 0.0251$, $a = 0$ and $b = 0.048$.

relied mostly on pools no greater than $K = 5$ [Basso et al., 2023], whereas the coastal region of Kenya relied mostly on pools of size 6 [Agoti et al., 2021], my upper bounds could have been used as an early indication that the added benefits of implementing ordered pooling in this context would have been modest at best. Increasing the batch size to $N = 300$ yields similar results.

The reason why these benefits are so small in this example compared to what was observed in the

chlamydia case study, is that here there is less variability in the probabilities of infection. Indeed, comparing the histograms from figures 8 and 3, one can observe a much greater dispersion in the probabilities of infection for chlamydia than for COVID-19 (even though the overall prevalence for chlamydia is lower).

One could argue that perhaps the benefits of implementing ordered partitions could be greater during other weeks of the pandemic, say, because there might be more dispersion in the probabilities of infection during these periods. But doing this same numerical exercise in other periods results in equally small savings in the expected number of tests, as illustrated in figure 10. This graph includes savings in testing costs obtained when implementing ordered pooling as opposed to random pooling for each week from my dataset, assuming a cost per test of 31.25 USD and ignoring the savings obtained from the reduction in the expected number of false negatives and false positives. From the figure, one can see that the reduction in costs is never above 20 cents per subject. Moreover, when implementing pools of size 5, the pool size mostly used in Chile to detect SARS-CoV-2 [Basso et al., 2023], the benefits are never above 2 cents per patient. Perhaps with more granular data, one could observe greater dispersion in the probabilities of infection for COVID-19, in which case implementing ordered pooling could prove to be more beneficial. But at least for the dataset that I have, these benefits appear to be rather small.

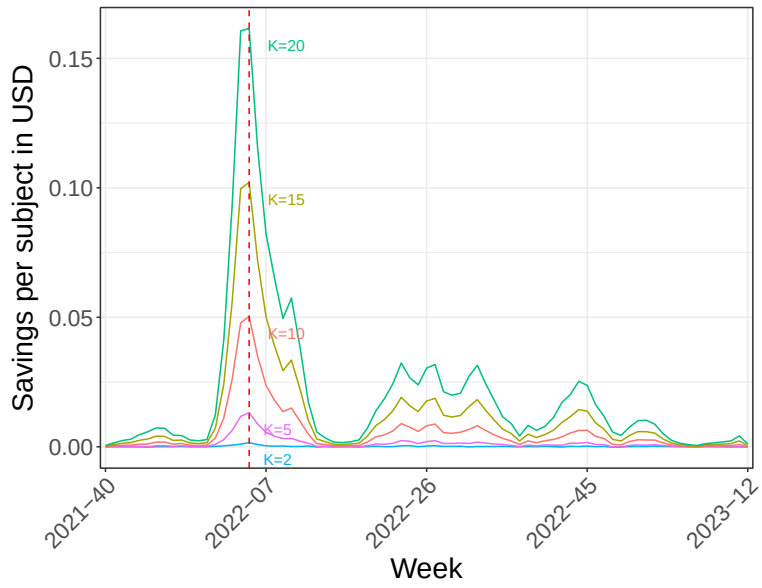


Figure 10: Evolution of the savings in costs in USD related to the expected number of tests obtained when implementing ordered pooling as opposed to random pooling to screen for SARS-CoV-2 infections in Chile using different pool sizes. These estimates assume a cost per test of 31.25 USD. The red vertical dashed line corresponds to 5th week of 2022, the week with the highest prevalence in the sample.

One can also get looser upper bounds to the benefits of implementing ordered pooling by setting more conservative values for μ, b and a. Figure 11, for example, depicts how the upper bounds to the monetary benefits of implementing ordered pooling as opposed to random pooling would increase if we had $b = 0.1$ as opposed to $b = 0.048$, for different prevalence levels μ, assuming a cost per test of 31.25 USD. As it can be seen from the figures, the benefits from implementing ordered pooling increase substantially as b increases.

One practical element that can influence the magnitude of b is how rich the dataset is, e.g., whether or not it provides information on patients' symptoms. The next section explores how the upper bounds derived in this paper can be used to assess the benefits of collecting richer surveillance data when conducting pooled testing.

(a) $b = 0.048$ (b) $b = 0.1$

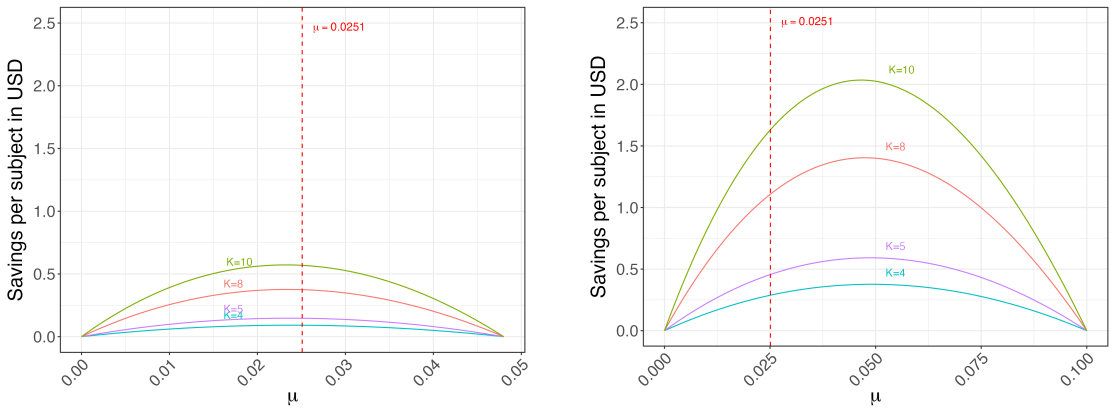


Figure 11: Maximum savings in testing costs in USD obtained from equation (5) for different pool sizes and assuming $a = 0$ and different levels of μ and b, and assuming a cost per test of 31.25 USD.

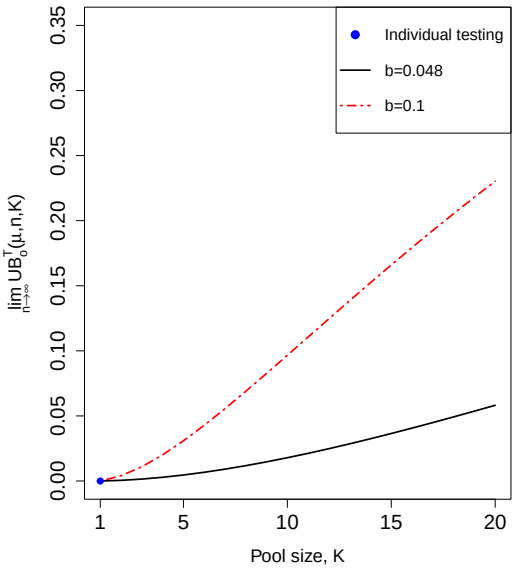
10.4 Improving the information collected from patients

The COVID-19 dataset that I analyzed did not have information on patients' probability of infection conditional on their reported symptoms, as information regarding their symptoms was probably not collected by most laboratories at the time of screening.

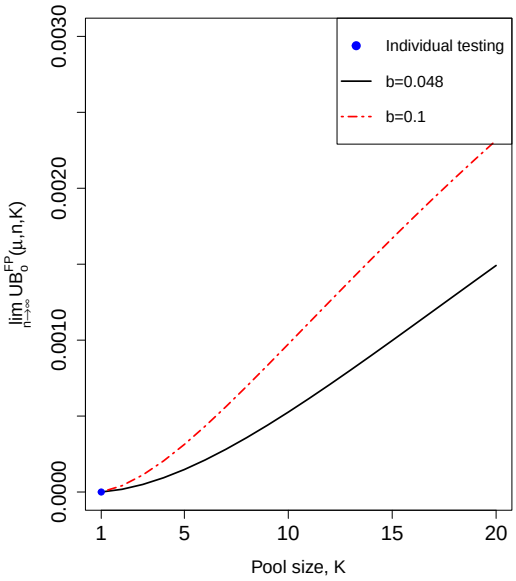
However, having data on symptoms could increase the variability in observed prior probabilities of infection, which, in turn, could enhance the benefits of implementing ordered pooling as opposed to random pooling. A desirable feature of the upper bounds derived in this article is that they can be used to assess the benefits of collecting this additional information, requiring minimal experimentation. For instance, if historical data indicates that unvaccinated patients with symptoms have a 10% infection probability, one can simply replace $b = 0.048$ with $b = 0.1$ in the expressions for the upper bounds (equations (5), (9) and (14)). This approach enables one to estimate the benefits of collecting data on symptoms even if one does not know, ex-ante, the full distribution of infection probabilities that one would get by collecting this data. Indeed, to assess the maximum benefits of gathering this additional information, one does not need to know, for example, the proportion of the population that exhibits symptoms *and* is unvaccinated, nor the infection probability of someone who *is* vaccinated *and* symptomatic. The only critical information required is the highest possible probability of infection, b, which can be assessed by estimating the probability that an individual is infected conditional that he is unvaccinated *and* exhibits symptoms.

Figure 12 depicts how the upper bounds would change if one collected data on patients' symptoms and *if* previous trials indicated that unvaccinated and symptomatic patients had a 10% chance of being infected, thus lifting the upper limit of the support of the distribution of Q_s from $b = 0.048$ to $b = 0.1$. From the figures, one can see that collecting this data has a great potential of improving the benefits of implementing ordered pooling. These benefits should, of course, be weighed against the costs associated with this greater surveillance.

(a) $\lim_{n\to\infty} UB_o^T(\mu, n, K)$



(b) $\lim_{n\to\infty} UB_o^{FP}(\mu, n, K)$



(c) $\lim_{n\to\infty} UB_o^{FN}(\mu, n, K)$

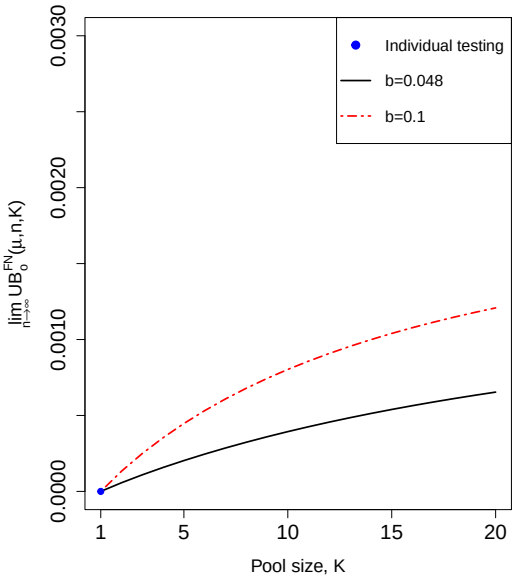


Figure 12: Asymptotic upper bounds to the benefits of implementing ordered pooling instead of random pooling for COVID-19 screening, assuming $a = 0$, $\mu = 0.0251$ and $b \in \{0.048, 0.1\}$, for different levels of K. The blue dot corresponds to the case in which subjects are individually tested, i.e. $K = 1$. Parameters for the dilution function (16): $S_e = .99$, $S_p = .989$ and $\delta = 0.0459$.

11 Discussion

This study has provided a simple method to compute an upper bound to the savings gained by implementing ordered pooling as opposed to matching subjects randomly when conducting pooled testing. The results presented have potential practical applications in healthcare management, as the upper bounds are easy to

compute and require very little information regarding the distribution of the probabilities of infection within the population.

Another desirable property of these results is that they pertain to a test scheme that is highly ubiquitous due to its simplicity and high speed with which it can generate final results. Indeed, because tests can be conducted in parallel in each testing phase (in the pooling phase and in the retesting phase), Dorfman screening allows testers to obtain quick and reliable results compared to more complex forms of adaptive testing [Hwang, 1975].

An interesting insight that emerges from these upper bounds is that they show that the benefits of implementing ordered pooling tend to diminish as the number of patients tested per batch diminishes. Therefore, for small-scale operations in which very few patients are tested each day, the benefits per patient of implementing ordered pooling will tend to be limited.

In practice, laboratories usually pool samples based on the consecutive order in which they arrive for testing. However, this method may already implement some degree of positive assortative matching if the probabilities of infection are serially correlated. This may happen, for example, if samples that arrive at the same time tend to be collected from the same region, and the probabilities of infection vary considerably across regions. If this is the case, the added benefits of ordering samples from lowest to highest probability of infection should be even lower than what the upper bounds derived in this article predict. As an example, Barak et al. [2021] conducted Dorfman testing to detect COVID-19 infections, and they ended up testing fewer samples than what the theory predicted. They conjectured that this happened because they grouped specimens based on the consecutive order in which they arrived for testing, which may have unintentionally generated some degree of positive assortative matching, thus reducing the expected number of tests below the levels obtained when implementing random pooling.

It is also not uncommon for laboratories to implement individual testing on samples that have a high risk of infection (e.g., Lewis, Lockary and Kobic [2012], Agoti et al. [2021] and Barak et al. [2021]). In these cases, testers can still rely on the formulas derived in this article by adjusting the maximum probability of infection from b to the threshold $\tilde{b}$ beyond which samples are individually tested, and by adjusting the prevalence μ to the expected probability of infection conditional that the subject has probability of infection less than or equal to $\tilde{b}$. Notice that, because this implies a reduction in b, μ and the average batch size N that is actually pooled tested, this selective approach tends to reduce the maximum potential benefits of implementing ordered pooling.

Notice that, to compute the upper bounds derived in this article, the tester must have some estimate of the dilution function $h(I, K)$. Because pooled testing is only effective when the dilution effect is not too strong (otherwise, pooled testing generates too many false negatives), the requirement that the tester must have a reliable estimate of $h(I, K)$ is arguably not unreasonable. But if, for some reason, this information is not readily available, I show in the Appendix that the tester may still compute alternative upper bounds that do not require knowing $h(I, K)$. The trade-off is that these alternative upper bounds are not nearly as tight as the ones presented in the main text (though, for very small prevalence levels and small pool sizes, these upper bounds can still be very close to zero, and, therefore, tight).

As improvements in data collection and surveillance tend to generate a higher dispersion in the probabilities of infection, the derived upper bounds can serve as a tool to simulate the potential benefits of implementing ordered pooling under improved surveillance conditions. For instance, in the COVID-19 case study, significant benefits from ordered pooling were not observed for small pool sizes. This might be attributed to the dataset lacking critical individual-level information, such as whether individuals exhibited COVID-19 symptoms. By using the upper bounds, it is possible to estimate the maximum benefits of implementing ordered pooling if, in addition to recording vaccination status, testing facilities also had information on patients' symptoms. Notably, this estimation does not require prior knowledge of the complete distribution of infection probabilities resulting from collecting this additional information.

As shown in Aprahamian, Bish and Bish [2019] and Saraiva [2023*a*], the benefits of implementing ordered pooling tend to be much larger when pool sizes are allowed to be heterogeneous. If, however, samples exhibit low variability in probabilities of infection or if the maximum pool size is restricted not to surpass a certain threshold (e.g., if pools are not allowed to have more than 5 samples), then, even if pool sizes are allowed to be heterogeneous, the benefits of implementing ordered pooling should still be limited. Indeed, consider a small batch of only 10 samples, where the probabilities of infection are given by $q = (.01, .01, .02, .02, .03, .03, .03, .03, .03, .03)$ with a dilution function given by equation (1) with parameters

$S_e = .99$, $S_p = .98$ and $\delta = .15$. Also suppose that both the cost of a test and the cost of a false positive are given by \$55, whereas the cost of a false negative is given by \$2,927. In this case, implementing the optimal partition allowing for any pool size – a computation that can be done using one of the features from the Shiny app `https://jcbrnd.shinyapps.io/shiny_app_anonymous/` – reduces the expected costs per subject by only 0.6% compared to matching patients randomly into one of two pools of size 5. Therefore, future research could attempt to extend the results from this article by deriving upper bounds to the benefits of implementing the optimal partition without imposing the restriction that all pools must have the same size.

Acknowledgements

I am grateful to friends and colleagues for their comments and feedback. Special thanks go to Eduardo Azevedo, Justin Burkett, Rafael R. Guthmann, Rosario Macera, and Caio Machado. I am also grateful for the invaluable feedback provided by two anonymous referees.

References

Adams, Derek R., Wendy R. Stensland, Chong H. Wang, Annette M. O'Connor, Darrell W. Trampel, Karen M. Harmon, Erin L. Strait, and Timothy S. Frana. 2013. "Detection of Salmonella Enteritidis in Pooled Poultry Environmental Samples Using a Serotype-Specific Real-TimePolymerase Chain Reaction Assay." *Avian Diseases*, 57(1): 22–28.

Agoti, Charles N., Martin Mutunga, Arnold W. Lambisia, et al. 2021. "Pooled testing conserves SARS-CoV-2 laboratory resources and improves test turn-around time: experience on the Kenyan Coast." *Wellcome Open Research*, 5: 186.

Aprahamian, Hrayer, Douglas R. Bish, and Erub K. Bish. 2019. "Optimal Risk-Based Group Testing." *Management Science*, 65(9): 4365–4384.

Aprahamian, Hrayer, Ebru K. Bish, and Douglas R. Bish. 2018. "Adaptive risk-based pooling in public health screening." *IISE Transactions*, 50(9): 753–766.

Barak, Netta, Roni Ben-Ami, Tal Sido, et al. 2021. "Lessons from applied large-scale pooling of 133,816 SARS-CoV-2 RT-PCR tests." *Science Translational Medicine*, 13(589): eabf2823.

Basso, Leonardo J., Marcel Goic, Marcelo Olivares, et al. 2023. "Analytics Saves Lives During the COVID-19 Crisis in Chile." *INFORMS Journal on Applied Analytics*, 53(1): 9–31.

Basso, Leonardo, Vicente Salinas, Denis Sauré, Charles Thraves, and Natalia Yankovic. 2022. "The Effect of Correlation and False Negatives in Pool Testing Strategies for COVID-19." *Healthcare Management Science*, 25: 146–165.

Bobkova, Nina, Ying Chen, and Hülya Eraslan. 2023. "Optimal group testing with heterogeneous risks." *Economic Theory*.

Centers for Disease Control and Prevention. 2000. "Tracking the hidden epidemics, trends in STDs in the United States." *https://www.cdc.gov/std/trends2000/trends2000.pdf*.

Centers for Disease Control and Prevention. 2019. "Sexually Transmitted Infections Treatment Guidelines, 2021." *https://www.cdc.gov/std/treatment-guidelines/chlamydia.htm*.

Dorfman, Robert. 1943. "The Detection of Defective Members of Large Populations." *The Annals of Mathematical Statistics*, 14: 436–440.

Fosah Tayong, Gladys E., Comfort Vuchas, Ngha Ndze Mbuh, Ubaida N. Suiteng, Cyrille Mbuli, Kingsley Che Soh, Albert Franck Zeh Meka, Serge C. Billong, Rina Djubgang, Gloria Ashuntantang, Vera Kum, Joseph Fokam, Moses Samje, and Melissa Sander. 2025. "Implementation of pooled testing to increase access to routine viral load monitoring for people living with HIV on antiretroviral therapy." *Scientific Reports*, 15(1): 14713.

Ghosh, Sabyasachi, Rishi Agarwal, Mohammad Ali Rehan, Shreya Pathak, Pratyush Agarwal, Yash Gupta, Sarthak Consul, Nimay Gupta, Ritika, Ritesh Goenka, Ajit Rajwade, and Manoj Gopalkrishnan. 2021. "A Compressed Sensing Approach to Pooled RT-PCR Testing for COVID-19 Detection." *IEEE Open Journal of Signal Processing*, 2: 248–264.

Grobe, Nadja, Alhaji Cherif, Xiaoling Wang, Zijun Dong, and Peter Kotanko. 2020. "Sample pooling: burden or solution?" *Clinical Microbiology and Infection*.

Habibzadeh, Parham, Mohammad Mofatteh, Mohammad Silawi, Saeid Ghavami, and Mohammad Ali Faghihi. 2021. "Molecular diagnostic assays for COVID-19: an overview." *Crit Rev Clin Lab Sci*, 58(6): 385–398.

Hwang, F. K. 1975. "A Generalized Binomial Group Testing Problem." *Journal of the American Statistical Association*, 70: 923–926.

Hwang, F. K. 1976. "Group Testing with a Dilution Effect." *Biometrika*, 63(3): 671–673.

Kacena, Katherine A., Sean B. Quinn, René Howell, Guillermo E. Madico, Thomas C. Quin, and Charlotte A. Gaydos. 1998. "Pooling Urine Samples for Ligase Chain Reaction Screening for Genital Chlamydia trachomatis Infection in Asymptomatic Women." *Journal of Clinical Microbiology*, 36(2): 481–485.

Lewis, Joanna Lynn, Vivian Marie Lockary, and Sadika Kobic. 2012. "Cost Savings and Increased Efficiency Using a Stratified Specimen Pooling Strategy for Chlamydia trachomatis and Neisseria gonorrhoeae." *Sexually Transmitted Diseases*, 39(1): 46–48.

Lipnowski, Elliot, and Doron Ravid. 2021. "Pooled testing for quarantine decisions." *Journal of Economic Theory*, 198: 105372.

Litchfield, Mark, Paul Brookes, Agnieszka Ojrzynska, Janki Kavi, and Richard Dawood. 2022. "Comparison of the clinical sensitivity and specificity of two commercial RNA SARS-CoV-2 assays." *International Journal of Infectious Diseases*, 118: 194–196.

Louisiana Department of Health. 2021. "2021 STI Data and Rankings, Louisiana." *https://ldh.la.gov/assets/oph/HIVSTD/Tables-Profiles/2021-STI-Rankings-Release.pdf*.

McMahan, Christopher S., Joshua M. Tebbs, and Christopher R. Bilder. 2012. "Informative Dorfman Screening." *Biometrics*, 68(1): 287–296.

Nguyen, Ngoc T., Hrayer Aprahamian, Ebru K. Bish, and Douglas R. Bish. 2019. "A Methodology for Deriving the Sensitivity of Pooled Testing, Based on Viral Load Progression and Pooling Dilution." *Journal of Translational Medicine*, 17(1): 1152–1163.

Owusu-Edusei Jr, Kwame, Harrell W. Chesson, Thomas L. Gift, Robert C. Brunham, and Gail Bolan. 2015. "Cost-effectiveness of Chlamydia Vaccination Programs for Young Women." *Emerging infectious diseases*, 21(6): 960–968.

Price, W. R., R. A. Olsen, and J. E. Hunter. 1972. "Salmonella Testing of Pooled Pre-Enrichment Broth Cultures for Screening Multiple Food Samples." *Applied Microbiology*, 23(4): 679–682.

Saraiva, Gustavo Quinderé. 2023*a*. "Pool testing with dilution effects and heterogeneous priors." *Health Care Management Science*, 651–672.

Saraiva, Gustavo Quinderé. 2023*b*. "Strategic incentives when implementing Dorfman testing with assortative matching." *Economics Letters*, 232: 111314.

Saraniti, Brett A. 2006. "Optimal pooled testing." *Health Care Management Science*, 9(2): 143–149.

Schachter, J, W E Stamm, T C Quinn, W W Andrews, J D Burczak, and H H Lee. 1994. "Ligase chain reaction to detect Chlamydia trachomatis infection of the cervix." *Journal of Clinical Microbiology*, 32(10): 2540–2543.

Sobel, M., and P. A. Groll. 1959. "Group testing to eliminate efficiently all defectives in a binomial sample." *The Bell System Technical Journal*, 38(5): 1179–1252.

Wein, Lawrence M., and Stefanos A. Zenio. 1996. "Pooled testing for HIV screening: capturing the dilution effect." *Operations Research*, 44(4): 543–569.

Yelin, Idan, Noga Aharony, Einat Shaer Tamar, Amir Argoetti, Esther Messer, Dina Berenbaum, Einat Shafran, Areen Kuzli, Nagham Gandali, Omer Shkedi, Tamar Hashimshony, Yael Mandel-Gutfreund, Michael Halberthal, Yuval Geffen, Moran Szwarcwort-Cohen, and Roy Kishony. 2020. "Evaluation of COVID-19 RT-qPCR Test in Multi sample Pools." *Clinical Infectious Diseases*, 71(16): 2073–2078.