# Online Appendix to “Pool Testing with Dilution Effects and Heterogeneous Priors” (Not for publication)

Gustavo Quinderé Saraiva
Business School, Pontificia Universidad Católica de Chile, gsaraiva@uc.cl

April 11, 2022
Latest draft: click here. This draft: March 28th, 2022.
First draft: February 19th, 2021.

---

## Contents

## Appendix

### A. When ordered pooling minimizes the expected number of tests

For any arbitrary group $G_g \subseteq S$, we define

$$T_{G_g} \equiv \begin{cases} 1, & \text{if } |G_g| = 1 \\ 1 + |G_g| \sum_{I=0}^{|G_g|} h(I,k) P_{G_g}(I), & \text{if } |G_g| > 1 \end{cases},$$

which corresponds to the expected number of tests for group $G_g$.[1]

**Lemma 1** *If $h(\cdot,k)$ is concave, then for any arbitrary group $G_g \subseteq S$ such that $|G_g| = k$, and any $l \in G_g$,*

$$\sum_{I=0}^{k-1} P_{G_g \setminus \{l\}}(I)\left[h(I+1,k) - h(I,k)\right]$$

*is decreasing in the probability of infection from each subject in $G_g \setminus \{l\}$.*

**Proof:** Notice that

$$\sum_{I=0}^{k-1} P_{G_g \setminus \{l\}}(I)\left[h(I+1,k) - h(I,k)\right]$$

corresponds to a weighted average of $h(I+1,k) - h(I,k)$, where the weights are determined by the probability mass function $P_{G_g \setminus \{l\}}(\cdot)$. Clearly, increasing the probability of infection from a subject in $G_g \setminus \{l\}$ causes this average to put more weight on higher values of $I$ (formally, letting $Y$ be the random variable associated with the probability mass function $P_{G_g \setminus \{l\}}(\cdot)$ and $Y'$ be its transformed version after the probability of infection from a subject in $G_g \setminus \{l\}$ is increased, we have that $Y'$ first-order stochastically dominates $Y$). Therefore, because concavity of $h(\cdot,k)$ implies that $h(I+1,k) - h(I,k)$ is decreasing in $I$, we have that increasing the probability of infection from a subject in $G_g \setminus \{l\}$ causes $\sum_{I=0}^{k-1} P_{G_g \setminus \{l\}}(I)[h(I+1,k) - h(I,k)]$ to decrease. ■

**Lemma 2** *Let $G_1$ and $G_2$ be two disjoint subsets of $S$ such that $|G_1| = k_1 \geq 2$ and $|G_2| = k_2 \geq 2$. Then consider the following ordered partitions of $G_1 \cup G_2$:*

$$\{G_1^*, G_2^*\},$$

*and*

$$\{G_2^{**}, G_1^{**}\},$$

*where $|G_1^*| = |G_1^{**}| = k_1$, $|G_2^*| = |G_2^{**}| = k_2$, $i < j$ for all $i \in G_1^*$ and all $j \in G_2^*$ and $i > j$ for all $i \in G_1^{**}$ and all $j \in G_2^{**}$.*

*If $h(\cdot,k_1)$ and $h(\cdot,k_2)$ are both concave, then*

$$\min\{T_{G_1^*} + T_{G_2^*}, T_{G_1^{**}} + T_{G_2^{**}}\} \leq T_{G_1} + T_{G_2}.$$

[1] The constant 1 corresponds to the pooled test performed for the group, while the second term corresponds to the probability that the pooled test detects infection times the number of subjects in the group (who are each tested individually in the event the pooled test detects infection). If the group is comprised of a single subject, then only one test is performed for that subject.

**Proof:** If $\{G_1, G_2\} = \{G_1^*, G_2^*\}$ or $\{G_1, G_2\} = \{G_1^{**}, G_2^{**}\}$, the proof is trivial. So suppose that $\{G_1, G_2\} \neq \{G_1^*, G_2^*\}$ and $\{G_1, G_2\} \neq \{G_1^{**}, G_2^{**}\}$, so that

$$\min_{i \in G_1} i < \max_{j \in G_2} j$$

and

$$\min_{j \in G_2} j < \max_{i \in G_1} i.$$

For an arbitrary $i \in G_1$ and $j \in G_2$ either one of the following inequalities must hold

$$\sum_{I=0}^{k_1-1} P_{G_1 \setminus \{i\}}(I)[h(I+1,k_1) - h(I,k_1)] \leq \sum_{I=0}^{k_2-1} P_{G_2 \setminus \{j\}}(I)[h(I+1,k_2) - h(I,k_2)] \tag{1}$$

or

$$\sum_{I=0}^{k_2-1} P_{G_2 \setminus \{j\}}(I)[h(I+1,k_2) - h(I,k_2)] \leq \sum_{I=0}^{k_1-1} P_{G_1 \setminus \{i\}}(I)[h(I+1,k_1) - h(I,k_1)]. \tag{2}$$

1. Suppose that inequality 2 holds. Then define

$$\bar{i} \equiv \max_{i \in G_1} i$$

and

$$\underline{j} \equiv \min_{j \in G_2} j.$$

Because $q_{\bar{i}} \geq q_i$, lemma 1 implies that

$$\sum_{I=0}^{k_1-1} P_{G_1 \setminus \{i\}}(I)[h(I+1,k_1) - h(I,k_1)] \leq \sum_{I=0}^{k-1} P_{G_1 \setminus \{\bar{i}\}}(I)[h(I+1,k_1) - h(I,k_1)]. \tag{3}$$

Analogously, $q_{\underline{j}} \leq q_j$ and lemma 1 imply that

$$\sum_{I=0}^{k_2-1} P_{G_2 \setminus \{\underline{j}\}}(I)[h(I+1,k_2) - h(I,k_2)] \leq \sum_{I=0}^{k_2-1} P_{G_2 \setminus \{j\}}(I)[h(I+1,k_2) - h(I,k_2)]. \tag{4}$$

Together inequalities 2, 3 and 4 imply that

$$\sum_{I=0}^{k_2-1} P_{G_2 \setminus \{\underline{j}\}}(I)[h(I+1,k_2) - h(I,k_2)] \leq \sum_{I=0}^{k_1-1} P_{G_1 \setminus \{\bar{i}\}}(I)[h(I+1,k_1) - h(I,k_1)]. \tag{5}$$

Now let us exchange the position of subjects $\underline{i}$ and $\bar{j}$ to create the new groups

$$\tilde{G}_1 = G_1 \cup \{\underline{j}\} \setminus \{\bar{i}\}$$

and

$$\tilde{G}_2 = G_2 \cup \{\bar{i}\} \setminus \{\underline{j}\}.$$

We will show that $T_{\tilde{G}_1} + T_{\tilde{G}_2} \leq T_{G_1} + T_{G_2}$. First, notice that

$$\begin{aligned} & T_{\tilde{G}_1} + T_{\tilde{G}_2} \leq T_{G_1} + T_{G_2} \\ \Longleftrightarrow & \sum_{I=0}^{k_1} h(I,k_1) P_{\tilde{G}_1}(I) + \sum_{I=0}^{k_2} h(I,k_2) P_{\tilde{G}_2}(I) \leq \sum_{I=0}^{k_1} h(I,k_1) P_{G_1}(I) + \sum_{I=0}^{k_2} h(I,k_2) P_{G_2}(I) \end{aligned} \tag{6}$$

Now notice that

$$\sum_{I=0}^{k_1} h(I,k_1)P_{\tilde{G}_1}(I) = q_{\underline{j}} \sum_{I=0}^{k_1-1} h(I+1,k_1)P_{\tilde{G}_1\setminus\{\underline{j}\}}(I) + (1-q_{\underline{j}}) \sum_{I=0}^{k_1-1} h(I,k_1)P_{\tilde{G}_1\setminus\{\underline{j}\}}(I)$$
$$= q_{\underline{j}} \sum_{I=0}^{k_1-1} P_{\tilde{G}_1\setminus\{\underline{j}\}}(I)[h(I+1,k_1)-h(I,k_1)] + \sum_{I=0}^{k_1} h(I,k_1)P_{\tilde{G}_1\setminus\{\underline{j}\}}(I).$$

Because $\tilde{G}_1\setminus\{\underline{j}\} = G_1\setminus\{\bar{i}\}$, the above expression can be rewritten as

$$\sum_{I=0}^{k_1} h(I,k_1)P_{\tilde{G}_1}(I) = q_{\underline{j}} \sum_{I=0}^{k_1-1} P_{G_1\setminus\{\bar{i}\}}(I)[h(I+1,k_1)-h(I,k_1)] + \sum_{I=0}^{k_1-1} h(I,k_1)P_{G_1\setminus\{\bar{i}\}}(I,k_1).$$

Analogously,

$$\sum_{I=0}^{k_2} h(I,k_2)P_{\tilde{G}_2}(I) = q_{\bar{i}} \sum_{I=0}^{k_2-1} P_{G_2\setminus\{\underline{j}\}}(I)[h(I+1,k_2)-h(I)] + \sum_{I=0}^{k_1-1} h(I,k_2)P_{G_2\setminus\{\underline{j}\}}(I),$$

$$\sum_{I=0}^{k_1} h(I,k_1)P_{G_1}(I) = q_{\bar{i}} \sum_{I=0}^{k_1-1} P_{G_1\setminus\{\bar{i}\}}(I)[h(I+1,k_1)-h(I,k_1)] + \sum_{I=0}^{k_1-1} h(I,k_1)P_{G_1\setminus\{\bar{i}\}}(I),$$

$$\sum_{I=0}^{k_2} h(I,k_2)P_{G_2}(I) = q_{\underline{j}} \sum_{I=0}^{k_2-1} P_{G_2\setminus\{\underline{j}\}}(I)[h(I+1,k_2)-h(I,k_2)] + \sum_{I=0}^{k_2-1} h(I,k_2)P_{G_2\setminus\{\underline{j}\}}(I).$$

Replacing these expressions into 6 and rearranging the terms, we get

$$T_{\tilde{G}_1} + T_{\tilde{G}_2} \le T_{G_1} + T_{G_2} \tag{7}$$
$$\Longleftrightarrow (q_{\bar{i}} - q_{\underline{j}}) \sum_{I=0}^{k_2-1} P_{G_2\setminus\{\underline{j}\}}(I)[h(I+1,k_2)-h(I)] \le (q_{\bar{i}} - q_{\underline{j}}) \sum_{I=0}^{k_1-1} P_{G_1\setminus\{\bar{i}\}}(I)[h(I+1,k_1)-h(I,k_1)]$$
$$\Longleftrightarrow \sum_{I=0}^{k_2-1} P_{G_2\setminus\{\underline{j}\}}(I)[h(I+1,k_2)-h(I,k_2)] \le \sum_{I=0}^{k_1-1} P_{G_1\setminus\{\bar{i}\}}(I)[h(I+1,k_1)-h(I,k_1)], \tag{8}$$

which, from inequality 5, holds.

Now, defining

$$\bar{i}' \equiv \max_{i\in\tilde{G}_1} i$$

and

$$\underline{j}' \equiv \min_{j\in\tilde{G}_2} j,$$

we have that either one of the following conditions must hold:

(a) $q_{\underline{j}} \ge q_{\bar{i}'}$ (i.e., $j > i$ for all $i \in \tilde{G}_1$ and all $j \in \tilde{G}_2$), in which case $\tilde{G}_1 = G_1^*$ and $\tilde{G}_2 = G_2^*$, so that $T_{G_1{}^*} + T_{G_2{}^*} \le T_{G_1} + T_{G_2}$.

(b) $q_{\underline{j}} < q'_{\bar{i}}$. In this case, notice that when we moved from $G_1$ to $\tilde{G}_1$ we decreased the probability of infection from one subject in this group while not altering the probability of infection from the remaining subjects within the group. From lemma 1 this implies that

$$\sum_{I=0}^{k_1-1} P_{G_1\setminus\{\bar{i}\}}(I)[h(I+1,k_1)-h(I,k_1)] \le \sum_{I=0}^{k_1-1} P_{\tilde{G}_1\setminus\{\bar{i}'\}}(I)[h(I+1,k_1)-h(I,k_1)]. \tag{9}$$

Analogously, when moving from $G_2$ to $\tilde{G}_2$ we increased the probability of infection from one subject in this group while not altering the probability of infection from the remaining subjects in this group. From lemma 1 this implies that

$$\sum_{I=0}^{k_1-1} P_{\tilde{G}_2\setminus\{\underline{j}'\}}(I)[h(I+1,k_1)-h(I,k_1)] \leq \sum_{I=0}^{k_1-1} P_{G_2\setminus\{\underline{j}'\}}(I)[h(I+1,k_1)-h(I,k_1)]. \tag{10}$$

Together, inequalities 8, 9 and 10 imply that

$$\sum_{I=0}^{k-1} P_{\tilde{G}_2\setminus\{\underline{j}'\}}(I)[h(I+1)-h(I)] \leq \sum_{I=0}^{k-1} P_{\tilde{G}_1\setminus\{\bar{i}'\}}(I)[h(I+1)-h(I)].$$

So we can redefine $G_1 = \tilde{G}_1$, $G_2 = \tilde{G}_2$, $\bar{i} = \bar{i}'$ and $\underline{j} = \underline{j}'$ and repeat the previous steps iteratively, until the final pair of groups is given by $G_1^*$ and $G_2^*$ such that $|G_1^*| = |G_1| = k_1$, $|G_2^*| = |G_2| = k_2$, $q_i \leq q_j$ for all $i \in G_1^*$ and all $j \in G_2^*$.

Because at each step of this algorithm we reduce the expected number of tests for subjects within these groups, and because at the end of this process we obtain the groups $G_1^*$ and $G_2^*$, we have that

$$T_{G_1^*} + T_{G_2^*} \leq T_{G_1} + T_{G_2},$$

as we wanted to show.

2. If inequality 1 holds, the proof is analogous to case 1, when inequality 2 holds, instead. Indeed, if 1 holds, we define $\underline{i} = \min_{i\in G_1} i$ and $\bar{j} = \max_{j\in G_m} j$, and then repeat the steps in case 1 to show that switching subject $\underline{i}$ with subject $\bar{j}$ weakly reduces the expected number of tests, provided that $h(\cdot,k_1)$ and $h(\cdot,k_2)$ are both concave. Then, we iteratively switch subjects from the new groups in the same fashion until we reach the ordered partition $\{G_2^{**}, G_1^{**}\}$. Because at each step of the algorithm the expected number of tests diminishes, we obtain $T_{G_1^{**}} + T_{G_2^{**}} \leq T_{G_1} + T_{G_2}$.

■

**Lemma 3** *Let $G_1$ and $G_2$ be two disjoint subsets of $S$ such that $|G_1| = k_1 \geq 1$ and $|G_2| = 1$. Let $\bar{j} = \max(G_1 \cup G_2)$ (i.e., $\bar{j}$ is the subject with highest probability of infection in group $G_1 \cup G_2$). Then consider the following ordered partition of $G_1 \cup G_2$:*

$$\{G_1^*, G_2^*\},$$

*where*

$$G_1^* \equiv G_1 \cup G_2 \setminus \{\bar{j}\},$$

$$G_2^* \equiv \left\{\bar{j}\right\},$$

*i.e., $\{G_1^*, G_2^*\}$ is the ordered partition that groups the $k_1$ subjects with lowest probability of infection together, and the subject with highest probability of infection alone.*

*If $h(\cdot,k_1)$ is increasing, then*

$$T_{G_1^*} + T_{G_2^*} \leq T_{G_1} + T_{G_2}.$$

**Proof:** Let $G_1$ and $G_2$ be two disjoint subsets of $S$. The proof when $|G_1| = |G_2| = 1$ is trivial. So suppose that $|G_1| = k_1 \geq 2$ and $G_2 = \{j\}$, with $j < \overline{j} = \max(G_1 \cup G_2)$. Let

$$G_1^* \equiv G_1 \cup G_2 \setminus \{\overline{j}\},$$

$$G_2^* \equiv \left\{\overline{j}\right\}.$$

Because $|G_2| = |G_2^*| = 1$, we have that $T_{G_2} = T_{G_2^*} = 1$. Therefore,

$$\begin{aligned}
& T_{G_1{}^*} + T_{G_2{}^*} \leq T_{G_1} + T_{G_2} \\
\iff & \left(1 + k_1 \sum_{I=0}^{k_1} P_{G_1^*}(I) h(I, k_1)\right) + 1 \leq \left(1 + k_1 \sum_{I=0}^{k_1} P_{G_1}(I) h(I, k_1)\right) + 1 \\
\iff & k_1 \left(q_j \sum_{I=0}^{k_1-1} P_{G_1 \setminus \{\overline{j}\}}(I) h(I+1, k_1) + (1 - q_j) \sum_{I=0}^{k_1-1} P_{G_1 \setminus \{\overline{j}\}}(I) h(I, k_1)\right) \leq \\
& k_1 \left(q_{\overline{j}} \sum_{I=0}^{k_1-1} P_{G_1 \setminus \{\overline{j}\}}(I) h(I+1, k_1) + (1 - q_{\overline{j}}) \sum_{I=0}^{k_1-1} P_{G_1 \setminus \{\overline{j}\}}(I) h(I, k_1)\right) \\
\iff & 0 \leq (q_{\overline{j}} - q_j) \left(\sum_{I=0}^{k_1-1} P_{G_1 \setminus \{\overline{j}\}}(I) (h(I+1, k_1) - h(I, k_1))\right) \\
\iff & 0 \leq q_{\overline{j}} - q_j,
\end{aligned}$$

which is satisfied, since $q_{\overline{j}} \geq q_j$ for all $j < \overline{j}$. ■

**Proof of theorem 1:**

Let $\{G_1, G_2, \cdots, G_m\}$ be an arbitrary partition of $S = \{1, 2, \cdots, n\}$, where $m$ is the number of groups from the partition (e.g., if all of the groups from the partition have the same size $k$, then $m = \frac{n}{k}$). Also suppose that $h(\cdot, |G_g|)$ is concave for every $G_g \in \{G_1, G_2, \cdots, G_m\}$. Then implement the following algorithm:

**Algorithm 1**

1. Initialize the partition $\tilde{\Omega} = \{\tilde{G_1}, \tilde{G_2}, \cdots, \tilde{G_m}\}$, where $\tilde{G_g} = G_g$ for all $g \in \{1, 2, \cdots, m\}$.
2. Initialize $g = 1$.
3. Set $w = g + 1$.
4. Pick groups $\tilde{G_g}$ and $\tilde{G_w}$ from $\tilde{\Omega}$.
   (a) If $|\tilde{G_g}| \geq 2$ and $|\tilde{G_w}| \geq 2$, consider the following ordered partitions of $\tilde{G_g} \cup \tilde{G_w}$:

   $$\{G_g^*, G_w^*\},$$

   and

   $$\{G_w^{**}, G_g^{**}\},$$

   where $|G_g^*| = |G_g^{**}| = |\tilde{G_g}|$, $|G_w^*| = |G_w^{**}| = |\tilde{G_w}|$, $i < j$ for all $i \in G_g^*$ and all $j \in G_w^*$, and $q_i > q_j$ for all $i \in G_g^{**}$ and all $j \in G_w^{**}$.

If $T_{G_g^*} + T_{G_w^*} \leq T_{G_g^{**}} + T_{G_w^{**}}$, redefine

$$\tilde{G_g} = G_g^* \quad \text{and} \quad \tilde{G_w} = G_w^*$$

else, redefine

$$\tilde{G_g} = G_w^{**} \quad \text{and} \quad \tilde{G_w} = G_g^{**}.$$

(b) If $|\tilde{G_g}| = 1$ or $|\tilde{G_w}| = 1$, redefine

$$\tilde{G_g} = \tilde{G_g} \cup \tilde{G_w} \setminus \{\max(\tilde{G_g} \cup \tilde{G_w})\},$$

and

$$\tilde{G_w} = \{\max(\tilde{G_g} \cup \tilde{G_w})\}.$$

5. If $w = m$, proceed to the next step. Else, redefine $w = w + 1$ and repeat step 4.
6. If $g = m - 1$, stop the algorithm, else, redefine $g = g + 1$ and repeat step 3.

Let $\Omega = \{G_1, G_2, \cdots, G_m\}$ be the original partition and $\tilde{\Omega} = \{\tilde{G_1}, \tilde{G_2}, \cdots, \tilde{G_m}\}$ be the final partition obtained obtained after implementing algorithm 1. From lemmas 2 and 3, at each step of this algorithm the overall expected number of tests weakly diminishes, which implies that $\mathbb{E}[T(\tilde{\Omega})] \leq \mathbb{E}[T(\Omega)]$. Because at the end of each step 6 of algorithm 1 we have that, for each $w > m$, $q_i \leq q_j$ for all $i \in \tilde{G_g}$ and all $j \in \tilde{G_w}$, we have that $\tilde{\Omega}$ is an ordered partition of $S$. Moreover, because changes made at teps 4a and 4b of algorithm 1 preserve pool sizes, there exists a permutation $p$ of the indices $(1, 2, \cdots, m)$ such that $|G_{p(g)}| = |G_g|$ for all $g \in \{1, 2, \cdots, m\}$.

As to the second part of the theorem, notice that, at each step 4b of algorithm 1 we are allocating the subject with highest probability of infection to be tested individually. This implies that if a subject $i$ is tested individually under $\tilde{\Omega}$, then a subject $j$ with $q_j > q_i$ is also tested individually under $\tilde{\Omega}$. ■

## B. When ordered pooling minimizes the expected number of false negatives

For any arbitrary group $G_g \subseteq S$, we define

$$FN_{G_g} \equiv \begin{cases} (1 - S_e) q_i, & \text{if } G_g = \{i\}, \\ \sum_{I=0}^{|G_g|} P_{G_g}(I) I \left[1 - h(I, |G_g|) S_e\right], & \text{if } |G_g| > 1 \end{cases},$$

which corresponds to the expected number of false negatives from group $G_g$.

For any arbitrary group $G_g \subseteq S$ such that $|G_g| \geq 2$ and any $j \in G_g$ we define

$$A_{G_g, j} \equiv \sum_{I=0}^{|G_g|-1} (I + 1)(1 - h(I + 1, |G_g|) S_e) P_{G_g \setminus j}(I),$$

$$B_{G_g, j} \equiv \sum_{I=0}^{|G_g|-1} I(1 - h(I, |G_g|) S_e) P_{G_g \setminus j}(I).$$

From the above expressions, $A_{G_g,j}$ is the expected number of false negatives in group $G_g$ conditional that $j \in G_g$ is infected. Similarly, $B_{G_g,j}$ is the expected number of false negatives in group $G_g$ conditional that $j \in G_g$ is not infected.

Now notice that, for any $j \in G_g$,

$$\begin{aligned} FN_{G_g} &= q_j A_{G_g,j} + (1-q_j) B_{G_g,j} \\ &= q_j (A_{G_g,j} - B_{G_g,j}) + B_{G_g,j}. \end{aligned} \tag{11}$$

**Lemma 4** *If hypothesis 1 holds for a given $k \geq 2$, then for any arbitrary group $G_g \subseteq S$ such that $|G_g| = k$, and any $l \in G_g$,*

$$A_{G_g,l} - B_{G_g,l}$$

*is decreasing in the probability of infection from each subject in $G_g \backslash \{l\}$.*

**Proof:** Notice that

$$A_{G_g,l} - B_{G_g,l} = \sum_{I=0}^{k-1} P_{G_g \backslash \{l\}}(I)[(I+1)(1-S_e h(I+1,k)) - I(1-S_e h(I,k))],$$

which corresponds to a weighted average of $(I+1)(1-S_e h(I+1,k)) - I(1-S_e h(I,k))$, where the weights are determined by the probability mass function $P_{G_g \backslash \{l\}}(\cdot)$. Clearly, increasing the probability of infection from a patient in $G_g \backslash \{l\}$ causes this average to put more weight on higher values of $I$ (formally, letting $Y$ be the random variable associated with the probability mass function $P_{G_g \backslash \{l\}}(\cdot)$ and $Y'$ be its transformed version after the probability of infection from a patient in $G_g \backslash \{l\}$ is increased, we have that $Y'$ first-order stochastically dominates $Y$). So it suffices to show that $(I+1)(1-S_e h(I+1,k)) - I(1-S_e h(I,k))$ is decreasing in $I$, a property that is satisfied if, for any $I \in \{1, 2, \cdots, k-1\}$,

$$\begin{aligned} &(I+1)(1-S_e h(I+1,k)) - I(1-S_e h(I,k)) - \\ &- [I(1-S_e h(I,k)) - (I-1)(1-S_e h(I-1,k))] \leq 0 \\ \iff &\frac{I+1}{2I} h(I+1,k) + \frac{I-1}{2I} h(I-1,k) \geq h(I,k). \end{aligned}$$

■

**Lemma 5** *Let $G_1$ and $G_2$ be two disjoint subsets of $S$ such that $|G_1| = k_1 \geq 2$ and $|G_2| = k_2 \geq 2$. Then consider the following ordered partitions of $G_1 \cup G_2$:*

$$\{G_1^*, G_2^*\},$$

*and*

$$\{G_2^{**}, G_1^{**}\},$$

*where $|G_1^*| = |G_1^{**}| = k_1$, $|G_2^*| = |G_2^{**}| = k_2$, $i < j$ for all $i \in G_1^*$ and all $j \in G_2^*$ and $i > j$ for all $i \in G_1^{**}$ and all $j \in G_2^{**}$.*

*If $h(\cdot, k_1)$ and $h(\cdot, k_2)$ satisfy hypothesis 1, then*

$$\min\{FN_{G_1^*} + FN_{G_2^*}, FN_{G_1^{**}} + FN_{G_2^{**}}\} \leq FN_{G_1} + FN_{G_2}.$$

**Proof:** If $\{G_1, G_2\} = \{G_1^*, G_2^*\}$ or $\{G_1, G_2\} = \{G_1^{**}, G_2^{**}\}$, the proof is trivial. So suppose that $\{G_1, G_2\} \neq \{G_1^*, G_2^*\}$ and $\{G_1, G_2\} \neq \{G_1^{**}, G_2^{**}\}$, so that

$$\min_{i \in G_1} i < \max_{j \in G_2} j$$

and

$$\min_{j \in G_2} j < \max_{i \in G_1} i.$$

For an arbitrary $i \in G_1$ and $j \in G_2$ either one of the following inequalities must hold

$$A_{G_1,i} - B_{G_1,i} \leq A_{G_2,j} - B_{G_2,j} \tag{12}$$

or

$$A_{G_2,j} - B_{G_2,j} \leq A_{G_1,i} - B_{G_1,i} \tag{13}$$

1. Assume that inequality 13 holds, and define

$$\bar{i} \equiv \max_{i \in G_1} i$$

and

$$\underline{j} \equiv \min_{j \in G_2} j.$$

Because $q_{\bar{i}} \geq q_i$, lemma 4 implies that

$$A_{G_1,i} - B_{G_1,i} \leq A_{G_1,\bar{i}} - B_{G_1,\bar{i}}. \tag{14}$$

Analogously, $q_{\underline{j}} \leq q_j$ and lemma 4 imply that

$$A_{G_2,\underline{j}} - B_{G_1,\underline{j}} \leq A_{G_1,j} - B_{G_1,j}. \tag{15}$$

Together inequalities 13, 14 and 15 imply that

$$A_{G_2,\underline{j}} - B_{G_2,\underline{j}} \leq A_{G_1,\bar{i}} - B_{G_1,\bar{i}}. \tag{16}$$

Now let us exchange the position of subjects $\underline{i}$ and $\bar{j}$ to create the new groups

$$\tilde{G}_1 = G_1 \cup \{\underline{j}\} \backslash \{\bar{i}\}$$

and

$$\tilde{G}_2 = G_2 \cup \{\bar{i}\} \backslash \{\underline{j}\}.$$

Then we must have $FN_{\tilde{G}_1} + FN_{\tilde{G}_2} \leq FN_{G_1} + FN_{G_2}$. Indeed, from expression 11, we have that

$$FN_{G_1} = q_{\bar{i}} A_{G_1,\bar{i}} + (1 - q_{\bar{i}}) B_{G_1,\bar{i}},$$

$$FN_{G_2} = q_{\underline{j}} A_{G_2,\underline{j}} + (1 - q_{\underline{j}}) B_{G_2,\underline{j}}.$$

Moreover, because

$$A_{\tilde{G}_1,\underline{j}} = A_{G_1,\bar{i}},$$

$$B_{\tilde{G}_1,\underline{j}} = B_{G_1,\bar{i}},$$

$$A_{\tilde{G}_2,\bar{i}} = A_{G_2,\underline{j}},$$

$$B_{\tilde{G}_2,\bar{i}} = B_{G_2,\underline{j}},$$

we also have that

$$FN_{\tilde{G}_1} = q_{\underline{j}} A_{G_1,\bar{i}} + (1 - q_{\underline{j}}) B_{G_1,\bar{i}},$$

$$FN_{\tilde{G}_2} = q_{\bar{i}} A_{G_2,\underline{j}} + (1 - q_{\bar{i}}) B_{G_2,\underline{j}}.$$

Therefore,

$$\begin{aligned} & FN_{\tilde{G}_1} + FN_{\tilde{G}_2} \leq FN_{G_1} + FN_{G_2} \\ \iff & q_{\underline{j}}(A_{G_1,\bar{i}} - B_{G_1,\bar{i}}) + q_{\bar{i}}(A_{G_2,\underline{j}} - B_{G_2,\underline{j}}) \leq q_{\bar{i}}(A_{G_1,\bar{i}} - B_{G_1,\bar{i}}) + q_{\underline{j}}(A_{G_2,\underline{j}} - B_{G_2,\underline{j}}) \\ \iff & (q_{\bar{i}} - q_{\underline{j}})(A_{G_2,\underline{j}} - B_{G_2,\underline{j}}) \leq (q_{\bar{i}} - q_{\underline{j}})(A_{G_1,\bar{i}} - B_{G_1,\bar{i}}) \\ \iff & (A_{G_2,\underline{j}} - B_{G_2,\underline{j}}) \leq (A_{G_1,\bar{i}} - B_{G_1,\bar{i}}), \end{aligned}$$

which, from inequality 16, holds.

Now, defining

$$\bar{i}' \equiv \max_{i \in \tilde{G}_1} i$$

and

$$\underline{j}' \equiv \min_{j \in \tilde{G}_2} j,$$

we have that either one of the following conditions must hold:

(a) $q_{\underline{j}} \geq q'_{\bar{i}}$, in which case $\tilde{G}_1 = G_1^*$ and $\tilde{G}_2 = G_2^*$, so that $FN_{G_1^*} + FN_{G_2^*} \leq FN_{G_1} + FN_{G_2}$.

(b) $q_{\underline{j}} < q'_{\bar{i}}$. In this case, notice that when we moved from $G_1$ to $\tilde{G}_1$ we decreased the probability of infection from one subject in this group while not altering the probability of infection from the remaining subjects in the group. From lemma 4 this implies that

$$A_{G_1,\bar{i}} - B_{G_1,\bar{i}} \leq A_{\tilde{G}_1,\bar{i}'} - B_{\tilde{G}_1,\bar{i}'}. \tag{17}$$

Analogously, when moving from $G_2$ to $\tilde{G}_2$ we increased the probability of infection from one subject in this group while not altering the probability of infection from the remaining subjects in the group. From lemma 4 this implies that

$$A_{\tilde{G}_2,\underline{j}'} - B_{\tilde{G}_2,\underline{j}'} \le A_{G_2,\underline{j}} - B_{G_2,\underline{j}}. \tag{18}$$

Together inequalities 16, 17 and 18 imply that

$$A_{\tilde{G}_2,\underline{j}'} - B_{\tilde{G}_2,\underline{j}'} \le A_{\tilde{G}_1,\bar{i}'} - B_{\tilde{G}_1,\bar{i}'}$$

So we can redefine $G_1 = \tilde{G}_1$, $G_2 = \tilde{G}_2$, $\bar{i} = \bar{i}'$ and $\underline{j} = \underline{j}'$ and repeat the previous steps iteratively, until the final pair of groups is given by $G_1^*$ and $G_2^*$.

Because at each step of this algorithm we reduce the expected number of false negatives for subjects in these groups, and because at the end of this process we obtain the groups $G_1^*$ and $G_2^*$, we have that

$$FN_{G_1^*} + FN_{G_2^*} \le FN_{G_1} + FN_{G_2},$$

as we wanted to show.

2. If inequality 12 holds, the proof is analogous to case the case in which inequality 13 holds (case 1). Indeed, when 12 holds, we define $\underline{i} = \min_{i\in G_1} i$ and $\bar{j} = \max_{j\in G_m} j$, and then repeat the same steps in case 13 to show that switching the positions of subjects $\underline{i}$ and $\bar{j}$ weakly reduces the expected number of tests, provided that $h(\cdot, k_1)$ and $h(\cdot, k_2)$ both satisfy hypothesis 1. Then, we iteratively switch subjects from the new groups in the same fashion until we reach the ordered partition $\{G_2^{**}, G_1^{**}\}$. Because at each step of the algorithm the expected number of false negatives diminishes, we obtain $FN_{G_1^{**}} + FN_{G_2^{**}} \le FN_{G_1} + FN_{G_2}$.

■

**Proof of theorem 2:**

Let $\Omega = \{G_1, G_2, \cdots, G_m\}$ be an arbitrary partition of $S = \{1, 2, \cdots, n\}$ such that $|G_g| \ge 2$ for all $G_g \in \Omega$, where $m$ is the number of groups from the partition (e.g., if all of the groups from the partition have the same size $k$, then $m = \frac{n}{k}$). Also suppose that hypothesis 1 holds for every $k \in \{|G_1|, |G_2|, \cdots, |G_m|\}$. Then implement the following algorithm:

**Algorithm 2**

1. Initialize the partition $\tilde{\Omega} = \{\tilde{G}_1, \tilde{G}_2, \cdots, \tilde{G}_m\}$, where $\tilde{G}_g = G_g$ for all $g \in \{1, 2, \cdots, m\}$.
2. Initialize $g = 1$.
3. Set $w = g + 1$.
4. Pick groups $\tilde{G}_g$ and $\tilde{G}_w$ from $\tilde{\Omega}$.

5. Then consider the following ordered partitions of $\tilde{G_g} \cup \tilde{G_w}$:

$$\{G_g^*, G_w^*\},$$

and

$$\{G_w^{**}, G_g^{**}\},$$

where $|G_g^*| = |G_g^{**}| = |\tilde{G_g}|$, $|G_w^*| = |G_w^{**}| = |\tilde{G_w}|$, $i < j$ for all $i \in G_g^*$ and all $j \in G_w^*$, and $q_i > q_j$ for all $i \in G_g^{**}$ and all $j \in G_w^{**}$.

If $T_{G_g^*} + T_{G_w^*} \leq T_{G_g^{**}} + T_{G_w^{**}}$, redefine

$$\tilde{G_g} = G_g^* \quad \text{and} \quad \tilde{G_w} = G_w^*$$

else, redefine

$$\tilde{G_g} = G_w^{**} \quad \text{and} \quad \tilde{G_w} = G_g^{**}.$$

6. If $w = m$, proceed to the next step. Else, redefine $w = w + 1$ and repeat step 4.
7. If $g = m - 1$, stop the algorithm, else, redefine $g = g + 1$ and repeat step 3.

Let $\Omega = \{G_1, G_2, \cdots, G_m\}$ be the original partition and $\tilde{\Omega} = \{\tilde{G_1}, \tilde{G_2}, \cdots, \tilde{G_m}\}$ be the final partition obtained obtained after implementing algorithm 2. From lemma 5, at each step of this algorithm the overall expected number of false negatives weakly diminishes, which implies that $\mathbb{E}[FN(\tilde{\Omega})] \leq \mathbb{E}[FN(\Omega)]$. Because at the end of each step 7 of algorithm 2 we have that, for each $w > m$, $q_i \leq q_j$ for all $i \in \tilde{G_g}$ and all $j \in \tilde{G_w}$, we have that $\tilde{\Omega}$ is an ordered partition of $S$. Moreover, because changes made at each step 5 of algorithm 2 preserve pool sizes, there exists a permutation $p$ of the indices $(1, 2, \cdots, m)$ such that $|G_{p(g)}| = |G_g|$ for all $g \in \{1, 2, \cdots, m\}$. ■

**Proposition 1** *For a given partition $\Omega = \{G_1, G_2, \cdots, G_m\}$ of $S$, if*

$$I \frac{\partial^2 h(I, |G_g|)}{\partial I^2} + 2 \frac{\partial h(I, |G_g|)}{\partial I} \geq 0, \quad \forall I \in [0, |G_g|], \quad and \; \forall G_g \in \Omega \tag{19}$$

*then*

$$\frac{I+1}{2I} h(I+1, |G_g|) + \frac{I-1}{2I} h(I-1, |G_g|) \geq h(I, |G_g|), \quad \forall I \in \{1, 2, \cdots, |G_g|\}, \quad and \; \forall G_g \in \Omega, \tag{20}$$

*but the converse is not necessarily true.*

**Proof:** The condition

$$I \frac{\partial^2 h(I, k)}{\partial I^2} + 2 \frac{\partial h(I, k)}{\partial I} \geq 0, \quad \forall I \in [0, k],$$

holds if and only if

$$\frac{\partial^2 I h(I, k)}{\partial I^2} \geq 0, \quad \forall I \in [0, k],$$

i.e., if and only if the function $f:[0,k]\to\mathbb{R}$ such that

$$f(I)=Ih(I,k)\quad \forall I\in[0,k]$$

is convex. Clearly, if $f(\cdot)$ is convex, hypothesis 1 holds.

But we can find instances in which equation 19 does not hold for some $I\in[0,k]$, and yet equation 20 holds for every $I\in\{1,2,\cdots,k-1\}$. Indeed, suppose that $k=3$ and

$$h(I,k)=1/[1+\exp(-20(I/k-3/4))],$$

then hypothesis 1 holds (for $I=1$ and $I=2$), even though condition 8 does not hold for non-integer values of $I$ greater than 2 and sufficiently close to $k=3$. ■

## C. When ordered pooling minimizes the expected number of false positives

For any arbitrary group $G_g\subseteq S$ s.t. $|G_g|\geq 2$, we define

$$FP_{G_g}\equiv\begin{cases}(1-S_p)(1-q_i), & \text{if } G_g=\{i\},\\ \sum_{I=0}^{|G_g|}P_{G_g}(I)h(I,|G_g|)(|G_g|-I)(1-S_p) & \text{if } |G_g|\geq 2\end{cases}$$

which corresponds to the expected number of false positives from group $G_g$.

For any arbitrary group $G_g\subseteq S$ such that $|G_g|\geq 2$ and any $j\in G_g$ we define

$$C_{G_g,j}\equiv\sum_{I=0}^{|G_g|-1}P_{G_g\setminus j}(I)h(I+1,k)(k-(I+1))(1-S_p),$$

$$D_{G_g,j}\equiv\sum_{I=0}^{|G_g|-1}P_{G_g\setminus j}(I)h(I,k)(k-I)(1-S_p).$$

From the above expressions, $C_{G_g,j}$ is the expected number of false positives in group $G_g$ conditional that $j\in G_g$ is infected. Similarly, $D_{G_g,j}$ is the expected number of false positives in group $G_g$ conditional that $j\in G_g$ is not infected.

Now notice that, for any $j\in G_g$,

$$\begin{aligned}FP_{G_g}&=q_jC_{G_g,j}+(1-q_j)D_{G_g,j}\\ &=q_j(C_{G_g,j}-D_{G_g,j})+D_{G_g,j}.\end{aligned}\tag{21}$$

**Lemma 6** *If hypothesis 2 holds for a given $k\geq 2$, then for any arbitrary group $G_g\subseteq S$ such that $|G_g|=k$, and any $l\in G_g$,*

$$C_{G_g,l}-D_{G_g,l}$$

*is decreasing in the probability of infection from each subject in $G_g\setminus\{l\}$.*

**Proof:** Notice that

$$C_{G_g,l} - D_{G_g,l} = \sum_{I=0}^{k-1} P_{G_g \setminus \{l\}}(I)(1-S_p)[(k-I-1)h(I+1,k) - (k-I)h(I,k)],$$

which corresponds to a weighted average of $(k-I-1)(1-S_p)h(I+1,k) - (k-I)(1-S_p)h(I,k)$, where the weights are determined by the probability mass function $P_{G_g \setminus \{l\}}(\cdot)$. Clearly, increasing the probability of infection from a subject in $G_g \setminus \{l\}$ causes this average to put more weight on higher values of $I$ (formally, letting $Y$ be the random variable associated with the probability mass function $P_{G_g \setminus \{l\}}(\cdot)$ and $Y'$ be its transformed version after the probability of infection from a subject in $G_g \setminus \{l\}$ is increased, we have that $Y'$ first-order stochastically dominates $Y$). So it suffices to show that $(k-I-1)(1-S_p)h(I+1,k) - (k-I)(1-S_p)h(I,k)$ is decreasing in $I$, a property that is satisfied if, for any $I \in \{1,2,\cdots,k-1\}$,

$$\begin{aligned}
&(k-I-1)h(I+1,k) - (k-I)h(I,k) - \\
&- [(k-I)h(I,k) - (k-I+1)h(I-1,k)] \leq 0 \\
\iff & \frac{k-I-1}{2(k-I)} h(I+1,k) + \frac{k-I+1}{2(k-I)} h(I-1,k) \leq h(I,k).
\end{aligned}$$

■

**Lemma 7** *Let $G_1$ and $G_2$ be two disjoint subsets of $S$ such that $|G_1| = k_1 \geq 2$ and $|G_2| = k_2 \geq 2$. Then consider the following ordered partitions of $G_1 \cup G_2$:*

$$\{G_1^*, G_2^*\},$$

*and*

$$\{G_2^{**}, G_1^{**}\},$$

*where $|G_1^*| = |G_1^{**}| = k_1$, $|G_2^*| = |G_2^{**}| = k_2$, $i < j$ for all $i \in G_1^*$ and all $j \in G_2^*$ and $i > j$ for all $i \in G_1^{**}$ and all $j \in G_2^{**}$.*

*If $h(\cdot, k_1)$ and $h(\cdot, k_2)$ satisfy hypothesis 2, then*

$$\min\{FP_{G_1^*} + FP_{G_2^*}, FP_{G_1^{**}} + FP_{G_2^{**}}\} \leq FP_{G_1} + FP_{G_2}.$$

**Proof:** If $\{G_1, G_2\} = \{G_1^*, G_2^*\}$ or $\{G_1, G_2\} = \{G_1^{**}, G_2^{**}\}$, the proof is trivial. So suppose that $\{G_1, G_2\} \neq \{G_1^*, G_2^*\}$ and $\{G_1, G_2\} \neq \{G_1^{**}, G_2^{**}\}$, so that

$$\min_{i \in G_1} i < \max_{j \in G_2} j$$

and

$$\min_{j \in G_2} j < \max_{i \in G_1} i.$$

For an arbitrary $i \in G_1$ and $j \in G_2$ either one of the following inequalities must hold

$$C_{G_1,i} - D_{G_1,i} \leq C_{G_2,j} - D_{G_2,j} \tag{22}$$

or

$$C_{G_2,j} - D_{G_2,j} \leq C_{G_1,i} - D_{G_1,i} \tag{23}$$

1. Assume that inequality 23 holds, and define

$$\bar{i} \equiv \max_{i \in G_1} i$$

and

$$\underline{j} \equiv \min_{j \in G_2} j.$$

Because $q_{\bar{i}} \geq q_i$, lemma 6 implies that

$$C_{G_1,i} - D_{G_1,i} \leq C_{G_1,\bar{i}} - D_{G_1,\bar{i}}. \tag{24}$$

Analogously, $q_{\underline{j}} \leq q_j$ and lemma 4 imply that

$$C_{G_2,\underline{j}} - D_{G_1,\underline{j}} \leq C_{G_1,j} - D_{G_1,j}. \tag{25}$$

Together inequalities 23, 24 and 25 imply that

$$C_{G_2,\underline{j}} - D_{G_2,\underline{j}} \leq C_{G_1,\bar{i}} - D_{G_1,\bar{i}}. \tag{26}$$

Now let us exchange the position of subjects $\underline{i}$ and $\bar{j}$ to create the new groups

$$\tilde{G}_1 = G_1 \cup \{\underline{j}\} \backslash \{\bar{i}\}$$

and

$$\tilde{G}_2 = G_2 \cup \{\bar{i}\} \backslash \{\underline{j}\}.$$

Then we must have $FP_{\tilde{G}_1} + FP_{\tilde{G}_2} \leq FP_{G_1} + FP_{G_2}$. Indeed, from expression 11, we have that

$$FP_{G_1} = q_{\bar{i}} C_{G_1,\bar{i}} + (1 - q_{\bar{i}}) D_{G_1,\bar{i}},$$
$$FP_{G_2} = q_{\underline{j}} C_{G_2,\underline{j}} + (1 - q_{\underline{j}}) D_{G_2,\underline{j}}.$$

Moreover, because

$$C_{\tilde{G}_1,\underline{j}} = C_{G_1,\bar{i}},$$

$$D_{\tilde{G}_1,\underline{j}} = D_{G_1,\bar{i}},$$

$$C_{\tilde{G}_2,\bar{i}} = C_{G_2,\underline{j}},$$

$$D_{\tilde{G}_2,\bar{i}} = D_{G_2,\underline{j}},$$

we also have that

$$FP_{\tilde{G}_1} = q_{\underline{j}} C_{G_1,\bar{i}} + (1 - q_{\underline{j}}) D_{G_1,\bar{i}},$$
$$FP_{\tilde{G}_2} = q_{\bar{i}} C_{G_2,\underline{j}} + (1 - q_{\bar{i}}) D_{G_2,\underline{j}}.$$

Therefore,

$$\begin{aligned}
&FP_{\tilde{G}_1} + FP_{\tilde{G}_2} \leq FP_{G_1} + FP_{G_2} \\
\Longleftrightarrow\ & q_{\underline{j}}(C_{G_1,\bar{i}} - D_{G_1,\bar{i}}) + q_{\bar{i}}(C_{G_2,\underline{j}} - D_{G_2,\underline{j}}) \leq q_{\bar{i}}(C_{G_1,\bar{i}} - D_{G_1,\bar{i}}) + q_{\underline{j}}(C_{G_2,\underline{j}} - D_{G_2,\underline{j}}) \\
\Longleftrightarrow\ & (q_{\bar{i}} - q_{\underline{j}})(C_{G_2,\underline{j}} - D_{G_2,\underline{j}}) \leq (q_{\bar{i}} - q_{\underline{j}})(C_{G_1,\bar{i}} - D_{G_1,\bar{i}}) \\
\Longleftrightarrow\ & (C_{G_2,\underline{j}} - D_{G_2,\underline{j}}) \leq (C_{G_1,\bar{i}} - D_{G_1,\bar{i}}),
\end{aligned}$$

which, from inequality 26, holds.

Now, defining

$$\bar{i}' \equiv \max_{i \in \tilde{G}_1} i$$

and

$$\underline{j}' \equiv \min_{j \in \tilde{G}_2} j,$$

we have that either one of the following conditions must hold:

(a) $q_{\underline{j}} \geq q'_{\bar{i}}$, in which case $\tilde{G}_1 = G_1^*$ and $\tilde{G}_2 = G_2^*$, so that $FP_{G_1{}^*} + FP_{G_2{}^*} \leq FP_{G_1} + FP_{G_2}$.

(b) $q_{\underline{j}} < q'_{\bar{i}}$. In this case, notice that when we moved from $G_1$ to $\tilde{G}_1$ we decreased the probability of infection from one subject in this group while not altering the probability of infection from the remaining subjects in the group. From lemma 6 this implies that

$$C_{G_1,\bar{i}} - D_{G_1,\bar{i}} \leq C_{\tilde{G}_1,\bar{i}'} - D_{\tilde{G}_1,\bar{i}'}. \tag{27}$$

Analogously, when moving from $G_2$ to $\tilde{G}_2$ we increased the probability of infection from one subject in this group while not altering the probability of infection from the remaining subjects in the group. From lemma 6 this implies that

$$C_{\tilde{G}_2,\underline{j}'} - D_{\tilde{G}_2,\underline{j}'} \leq C_{G_2,\underline{j}} - D_{G_2,\underline{j}}. \tag{28}$$

Together inequalities 26, 27 and 28 imply that

$$C_{\tilde{G}_2,\underline{j}'} - D_{\tilde{G}_2,\underline{j}'} \leq C_{\tilde{G}_1,\bar{i}'} - D_{\tilde{G}_1,\bar{i}'}$$

So we can redefine $G_1 = \tilde{G}_1$, $G_2 = \tilde{G}_2$, $\bar{i} = \bar{i}'$ and $\underline{j} = \underline{j}'$ and repeat the previous steps iteratively, until the final pair of groups is given by $G_1^*$ and $G_2^*$.

Because at each step of this algorithm we reduce the expected number of false negatives for subjects in these groups, and because at the end of this process we obtain the groups $G_1^*$ and $G_2^*$, we have that

$$FP_{G_1^*} + FP_{G_2^*} \leq FP_{G_1} + FP_{G_2},$$

as we wanted to show.

2. If inequality 22 holds, the proof is analogous to case the case in which inequality 23 holds (case 1). Indeed, when 22 holds, we define $\underline{i} = \min_{i \in G_1} i$ and $\overline{j} = \max_{j \in G_m} j$, and then repeat the same steps in case 23 to show that switching the positions of subjects $\underline{i}$ and $\overline{j}$ weakly reduces the expected number of false positives, provided that $h(\cdot, k_1)$ and $h(\cdot, k_2)$ both satisfy hypothesis 2. Then, we iteratively switch subjects from the new groups in the same fashion until we reach the ordered partition $\{G_2^{**}, G_1^{**}\}$. Because at each step of the algorithm the expected number of false positives diminishes, we obtain $FP_{G_1^{**}} + FP_{G_2^{**}} \leq FP_{G_1} + FP_{G_2}$.

■

**Lemma 8** *Let $G_1$ and $G_2$ be two disjoint subsets of S such that $|G_1| = k \geq 1$ and $|G_2| = 1$. Let $\overline{j} = \max(G_1 \cup G_2)$ (i.e., $\overline{j}$ is the subject with highest probability of infection in group $G_1 \cup G_2$). Then consider the following ordered partition of $G_1 \cup G_2$:*

$$\{G_1^*, G_2^*\},$$

*where*

$$G_1^* \equiv G_1 \cup G_2 \setminus \{\overline{j}\},$$

$$G_2^* \equiv \left\{\overline{j}\right\},$$

*i.e., $\{G_1^*, G_2^*\}$ is the ordered partition that groups the k subjects with lowest probability of infection together, and the subject with highest probability of infection alone.*

*If $h(\cdot, k)$ is increasing, then*

$$FP_{G_1^*} + FP_{G_2^*} \leq FP_{G_1} + FP_{G_2}.$$

**Proof:** Let $G_1$ and $G_2$ be two disjoint subsets of $S$. The proof when $|G_1| = |G_2| = 1$ is trivial. So suppose that $|G_1| = k_1 \geq 2$ and $G_2 = \{j\}$, with $j < \overline{j} = \max(G_1 \cup G_2)$. Let

$$G_1^* \equiv G_1 \cup G_2 \setminus \{\overline{j}\},$$

$$G_2^* \equiv \left\{\overline{j}\right\}.$$

Because $G_2 = \{j\}$ and $G_2^* = \{\overline{j}\}$, we have that $FP_{G_2} = (1 - S_p)(1 - q_j)$ and $FP_{G_2^*} = (1 - S_p)(1 - q_{\overline{j}})$. Therefore,

$$
\begin{aligned}
& FP_{G_1{}^*} + FP_{G_2{}^*} \leq FP_{G_1} + FP_{G_2} \\
\iff & (1 - S_p) \sum_{I=0}^{k} P_{G_1^*}(I)(k - I) h(I, k) + (1 - S_p)(1 - q_{\overline{j}}) \leq +(1 - S_p) \sum_{I=0}^{k} P_{G_1}(I)(k - I)(h(I, k)) + (1 - S_p)(1 - q_j) \\
\iff & (1 - S_p) \left( q_j \sum_{I=0}^{k-1} P_{G_1 \setminus \{\overline{j}\}}(I)(k - I - 1) h(I + 1, k) + (1 - q_j) \sum_{I=0}^{k-1} P_{G_1 \setminus \{\overline{j}\}}(I) h(I, k)(k - I) \right) \leq
\end{aligned}
$$

$$
\begin{aligned}
&(1-S_p)\left(q_{\bar{j}}\sum_{I=0}^{k-1}P_{G_1\setminus\{\bar{j}\}}(I)(k-I-1)h(I+1,k)+(1-q_{\bar{j}})\sum_{I=0}^{k-1}P_{G_1\setminus\{\bar{j}\}}(I)h(I,k)(k-I)\right)+(1-S_p)(q_{\bar{j}}-q_j)\\
\iff&(q_{\bar{j}}-q_j)\left(\sum_{I=0}^{k-1}P_{G_1\setminus\{\bar{j}\}}(I)\left[(k-I)h(I,k)-(k-I-1)h(I+1,k)\right]\right)\leq(q_{\bar{j}}-q_j)\\
\iff&\sum_{I=0}^{k-1}P_{G_1\setminus\{\bar{j}\}}(I)\left[(k-I)h(I,k)-(k-I-1)h(I+1,k)\right]\leq 1, \qquad (29)
\end{aligned}
$$

Now notice that, since $h(\cdot,k)$ is increasing and since $h(I,k)\leq(1-S_e)\leq 1$, we have that

$$
\begin{aligned}
\sum_{I=0}^{k-1}P_{G_1\setminus\{\bar{j}\}}(I)\left[(k-I)h(I,k)-(k-I-1)h(I+1,k)\right]&\leq\sum_{I=0}^{k-1}P_{G_1\setminus\{\bar{j}\}}(I)\left[(k-I)h(I+1,k)-(k-I-1)h(I+1,k)\right]\\
&=\sum_{I=0}^{k-1}P_{G_1\setminus\{\bar{j}\}}(I)h(I+1,k)\\
&\leq\sum_{I=0}^{k-1}P_{G_1\setminus\{\bar{j}\}}(I)=1,
\end{aligned}
$$

which implies that inequality 29 is satisfied. ■

**Proof of theorem 3:**

Let $\Omega=\{G_1,G_2,\cdots,G_m\}$ be an arbitrary partition of $S=\{1,2,\cdots,n\}$ such that $|G_g|\geq 2$ for all $G_g\in\Omega$, where $m$ is the number of groups from the partition (e.g., if all of the groups from the partition have the same size $k$, then $m=\frac{n}{k}$). Also suppose that hypothesis 2 holds for every $k\in\{|G_1|,|G_2|,\cdots,|G_m|\}$. Then implement the following algorithm:

**Algorithm 3**

1. Initialize the partition $\tilde{\Omega}=\{\tilde{G}_1,\tilde{G}_2,\cdots,\tilde{G}_m\}$, where $\tilde{G}_g=G_g$ for all $g\in\{1,2,\cdots,m\}$.
2. Initialize $g=1$.
3. Set $w=g+1$.
4. Pick groups $\tilde{G}_g$ and $\tilde{G}_w$ from $\tilde{\Omega}$.
   (a) If $|G_g|\geq 2$ and $|G_w|\geq 2$, consider the following ordered partitions of $\tilde{G}_g\cup\tilde{G}_w$:

$$\{G_g^*,G_w^*\},$$

   and

$$\{G_w^{**},G_g^{**}\},$$

   where $|G_g^*|=|G_g^{**}|=|\tilde{G}_g|$, $|G_w^*|=|G_w^{**}|=|\tilde{G}_w|$, $i<j$ for all $i\in G_g^*$ and all $j\in G_w^*$, and $q_i>q_j$ for all $i\in G_g^{**}$ and all $j\in G_w^{**}$.

   If $FP_{G_g^*}+FP_{G_w^*}\leq FP_{G_g^{**}}+FP_{G_w^{**}}$, redefine

$$\tilde{G}_g=G_g^*\quad\text{and}\quad\tilde{G}_w=G_w^*$$

   else, redefine

$$\tilde{G}_g=G_w^{**}\quad\text{and}\quad\tilde{G}_w=G_g^{**}.$$

(b) If $|\tilde{G_g}| = 1$ or $|\tilde{G_w}| = 1$, redefine

$$\tilde{G_g} = \tilde{G_g} \cup \tilde{G_w} \setminus \{\max(\tilde{G_g} \cup \tilde{G_w})\},$$

and

$$\tilde{G_w} = \{\max(\tilde{G_g} \cup \tilde{G_w})\}.$$

5. If $w = m$, proceed to the next step. Else, redefine $w = w + 1$ and repeat step 4.
6. If $g = m - 1$, stop the algorithm, else, redefine $g = g + 1$ and repeat step 3.

Let $\Omega = \{G_1, G_2, \cdots, G_m\}$ be the original partition and $\tilde{\Omega} = \{\tilde{G_1}, \tilde{G_2}, \cdots, \tilde{G_m}\}$ be the final partition obtained obtained after implementing algorithm 3. From lemmas 7 and 8, at each step of this algorithm the overall expected number of false positives weakly diminishes, which implies that $\mathbb{E}[FP(\tilde{\Omega})] \leq \mathbb{E}[FP(\Omega)]$. Because at the end of each step 6 of algorithm 3 we have that, for each $w > m$, $q_i \leq q_j$ for all $i \in \tilde{G_g}$ and all $j \in \tilde{G_w}$, we have that $\tilde{\Omega}$ is an ordered partition of $S$. Moreover, because changes made at steps 4a and 4b of algorithm 3 preserve pool sizes, there exists a permutation $p$ of the indices $(1, 2, \cdots, m)$ such that $|G_{p(g)}| = |G_g|$ for all $g \in \{1, 2, \cdots, m\}$.

As to the second part of the theorem, notice that, at each step 4b of algorithm 3 we are allocating the subject with highest probability of infection to be tested individually. This implies that if a subject $i$ is tested individually under $\tilde{\Omega}$, then a subject $j$ with $q_j > q_i$ is also tested individually under $\tilde{\Omega}$. ■

## D. Proof of proposition 1

Suppose that the dilution function is given by

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

Then,

a) Ordered pooling minimizes the expected number of tests. Indeed this follows directly from theorem 1 and the fact that $h(\cdot, k)$ is concave.

b) Any matching criteria produces the same expected number of false negatives. Indeed, consider any arbitrary infected patient from the population. Then regardless of whom he is matched with in the pooled sample, the probability that his pooled sample tests positive for infection is $(1 - S_e)$. If the pooled test does accuse infection, which happens with probability $S_e$, the infected subject is individually tested and diagnosed as healthy with probability $(1 - S_e)$. So using simple probability theory, the overall probability that an arbitrary infected subject is incorrectly diagnosed as healthy is given by

$$(1 - S_e) + S_e(1 - Se) = 1 - S_e^2,$$

which does not depend on whom the subject is matched with in the pool.

c) Ordered pooling minimizes the expected number of false positives. Indeed, this follows directly from theorem 3 and the fact that this dilution function satisfies hypothesis 2. ■

## E. Proof of proposition 2

Suppose that the dilution function is given by $h(I,k) = a + bI$, where $b \geq 0$.

a) First, we show that the expected number of tests is not affected by the matching criteria used to form the pools.

Let $X_{i,j} = 1$ if individual $i$ from group $j$ is infected. Moreover, assume that $Prob(X_{i,j} = 1) = q_{i,j}$. Then, the expected number of tests is given by

$$T \equiv \frac{n}{k} + \sum_{i=1}^{n/k}\sum_{I=0}^{k} h(I,k) Prob(\sum_{j=1}^{n} X_{i,j} = I).$$

Assuming that $h(I,k) = a + bI$, we have that

$$\begin{aligned} T &= \frac{n}{k} + \sum_{i=1}^{n/k}\sum_{I=0}^{k} \left\{ [a + bI]\, Prob(\sum_{j=1}^{n} X_{i,j} = I) \right\}. \\ &= \frac{n}{k} + \sum_{i=1}^{n/k} \left[ a + b\sum_{I=0}^{k} I Prob(\sum_{j=1}^{n} X_{i,j} = I) \right]. \end{aligned} \quad (30)$$

Because $\sum_{I=0}^{k} IProb(\sum_{j=1}^{n} X_{i,j} = I)$ equals to the expected number of infected in group $i$, and because the number of infected in each group follows a Poisson-binomial distribution with parameters $(k, \{q_{i,j}\}_{j=1}^{k})$, this expectation is given by $\sum_{j=1}^{k} q_{i,j}$. Replacing this into expression 30, we get

$$\begin{aligned} T &= \frac{n}{k} + \sum_{i=1}^{n/k} \left[ a + b\sum_{j=1}^{k} q_{i,j} \right] \\ &= \frac{n}{k} + \frac{n}{k}a + b\sum_{i=1}^{n/k}\sum_{j=1}^{k} q_{i,j}, \end{aligned}$$

which clearly does not depend on how the $q_{i,j}$'s are grouped.

b) As to the proof that ordered pooling achieves a minimum in the expected number of false negatives, it follows directly from theorem 2 and the fact that $h(I,k) = a + bI$ (with $b \geq 0$) satisfies hypothesis 1. Finally, notice that hypothesis 2 is clearly satisfied when $h(I,k)$ is affine, so that ordered pooling also minimizes the expected number of false positives. ■

## F. Proof of proposition 3

Suppose that the dilution function is given by

$$h(l,k) = \begin{cases} 1 - S_p, & \text{if } I < k \\ S_e, & \text{if } I = k. \end{cases}$$

a) First, we will show that ordered pooling *maximizes* the expected number of tests.

Consider a partition $\Omega \equiv \{G_1, G_2, \cdots, G_{n/k}\}$ of $S = \{1, 2, \cdots, n\}$, where $|G_g| = k$ for all $G_g \in \Omega$ (i.e., each group has the same size). The probability that group $G_g$ has exactly $k$ infected subjects is given by $P_{G_g}(k) = \prod_{j \in G_g} q_j$, and the expected number of tests from this partition is given by

$$\mathbb{E}[T(\Omega)] = \frac{n}{k} + n(1 - S_p) + k(S_p + S_e - 1)\sum_{g \in \Omega} P_{G_g}(k).$$

Because $k(S_p + S_e - 1)$ is a positive constant (recall that we assume that $S_e > 1 - S_p$), we have that the partition that minimizes the expected number of tests is the one that *minimizes* $\sum_{G_g \in \Omega} P_{G_g}(k)$. We now show that ordered pooling *maximizes* $\sum_{G_g \in \Omega} P_{G_g}(k)$.

Suppose by way of contradiction that $\Omega = \{G_1, G_2, \cdots, G_{n/k}\}$ maximizes $\sum_{G_g \in G} P_{G_g}(k)$ among all partitions that have all pool sizes equal to $k$, and suppose that $\Omega$ is not an ordered partition. Then there are $G_l, G_w \in \Omega$ such that $P_{G_l}(k) < P_{G_w}(k)$ and there is a $i \in G_l$ an $j \in G_w$ such that $(1 - q_i) > (1 - q_j)$. Then, denoting $A \equiv \prod_{\tau \in G_l \setminus \{i\}} (1 - q_\tau)$ and $B \equiv \prod_{\tau \in G_w \setminus \{j\}} (1 - q_\tau)$, we must have that $A < B$. Moreover, if $\Omega$ maximizes $\sum_{G_g \in \Omega} P_{G_g}(k)$ we must have

$$(1 - q_i)A + (1 - q_j)B \geq (1 - q_j)A + (1 - q_i)B$$

$$\iff (1 - q_j)(B - A) \geq (1 - q_i)(B - A)$$

$$\iff (1 - q_j) \geq (1 - q_i),$$

a contradiction with $(1 - q_i) > (1 - q_j)$.$\rightarrow\leftarrow$

b) As to the proof that ordered pooling minimizes the expected number of false negatives, it follows directly from theorem 2 and the fact that this dilution function clearly satisfies hypothesis 1.

c) Finally, under this extreme dilution effect, if a healthy subject is pooled tested, the test will detect infection with probability $1 - S_p$ regardless of who the healthy patient is matched with. In this contingency, the subject is tested again as is declared infected with probability $1 - S_p$. Because those events are independent, the overall probability that a healthy subject is diagnosed as infected is $(1 - S_p)^2$. Because this probability does not depend on who the subject is matched with, this implies that any matching mechanism generates the same expected number of false positives. ■

## G. When ordered pooling maximizes social welfare

**Proof of proposition 4**

It suffices to show that ordered pooling maximizes the minimum utility and the sum of utilities. That ordered pooling maximizes the sum of utilities follows immediately from corollary 4. So it only remains to show that ordered pooling maximizes the minimum utility.

Without loss of generality, we can assume that there are only two groups, and that the probability of infection from subjects is given by

$$q_1 \leq q_2 \leq q_3 \leq q_4.$$

Then, the ordered partition is given by

$$\Omega^* = \{\{1, 2\}, \{3, 4\}\}.$$

The only possible alternative partitions are

$$\Omega' = \{\{1, 3\}, \{2, 4\}\}$$

and

$$\Omega'' = \{\{1,4\},\{2,3\}\}.$$

Case 1) Suppose that $\lambda = 1$.

For any $q_i \in [0,1]$, let

$$A(q_i) \equiv q_i[1 - S_e h(2,2)] + (1 - q_i)[1 - S_e h(1,2)],$$

i.e., $A(q_i)$ is the probability that a subject is *not* detected as infected given that this subject is infected and the person this subject is matched with has probability of infection of $q_i$.

In this case, if the ordered partition $\Omega^*$ is implemented, subjects' utilities are given by

$$u_1(\Omega^*) = 1 - q_1 A(q_2), \quad u_2(\Omega^*) = 1 - q_2 A(q_1), \quad u_3(\Omega^*) = 1 - q_3 A(q_4), \quad u_4(\Omega^*) = 1 - q_4 A(q_3).$$

The utilities from the other 2 other possible partitions, $\Omega'$ and $\Omega''$, are given by

$$u_1(\Omega') = 1 - q_1 A(q_3), \quad u_2(\Omega') = 1 - q_2 A(q_4), \quad u_3(\Omega') = 1 - q_3 A(q_1), \quad u_4(\Omega') = 1 - q_4 A(q_2),$$

and

$$u_1(\Omega'') = 1 - q_1 A(q_4), \quad u_2(\Omega'') = 1 - q_2 A(q_3), \quad u_3(\Omega'') = 1 - q_3 A(q_2), \quad u_4(\Omega'') = 1 - q_4 A(q_1).$$

Because $A(q_i) \geq 0$, we have that a subject's utility is decreasing in the subject's probability of infection. Moreover, because $h(\cdot,2)$ is increasing (see assumption 1), we have that $[1 - S_e h(2,2)] \leq [1 - S_e h(1,2)]$, which implies that $A(q_i)$ is decreasing in $q_i$. This implies that a subject's utility is increasing in the probability of infection of the subject he is matched with. This implies that the subject with lowest utility from each group is the subject who has the highest probability of infection from that group.

So suppose that, under ordered pooling, subject 2 is the one with the lowest utility, i.e., suppose that

$$u_2(\Omega^*) = \min_{i \in \{1,2,3,4\}} u_i(\Omega^*).$$

In this case, we have that

$$u_3(\Omega') = 1 - q_3 A(q_1) \leq 1 - q_2 A(q_1) = u_2(\Omega^*) = \min_{i \in \{1,2,3,4\}} u_i(\Omega^*),$$

and

$$u_4(\Omega'') = 1 - q_4 A(q_1) \leq 1 - q_2 A(q_1) = u_2(\Omega^*) = \min_{i \in \{1,2,3,4\}} u_i(\Omega^*),$$

so that ordered pooling maximizes the minimum utility.

Suppose, instead, that subject 4 is the one with lowest utility under ordered pooling, i.e., suppose that

$$u_4(\Omega^*) = \min_{i \in \{1,2,3,4\}} u_i(\Omega^*).$$

In this case, we have that

$$u_4(\Omega') = 1 - q_4 A(q_2) \leq 1 - q_4 A(q_3) = u_4(\Omega^*) = \min_{i\in\{1,2,3,4\}} u_i(\Omega^*),$$

and

$$u_4(\Omega'') = 1 - q_4 A(q_1) \leq 1 - q_4 A(q_3) = u_4(\Omega^*) = \min_{i\in\{1,2,3,4\}} u_i(\Omega^*),$$

so that again, ordered pooling maximizes the minimum utility.

Case 2) Suppose that $\lambda = 0$.

For any $q_i \in [0,1]$, let

$$B(q_i) \equiv q_i(1-S_p)h(1,2) + (1-q_i)(1-S_p)h(0,2),$$

i.e., $B(q_i)$ is the probability that a subject is detected as infected given that this subject is *not* infected and the person this subject is matched with has probability of infection of $q_i$.

In this case, if the ordered partition $\Omega^*$ is implemented, subjects' utilities are given by

$$u_1(\Omega^*) = 1-(1-q_1)B(q_2), \quad u_2(\Omega^*) = 1-(1-q_2)B(q_1), \quad u_3(\Omega^*) = 1-(1-q_3)B(q_4), \quad u_4(\Omega^*) = 1-(1-q_4)B(q_3).$$

The utilities from the other 2 other possible partitions, $\Omega'$ and $\Omega''$, are given by

$$u_1(\Omega') = 1-(1-q_1)B(q_3), \quad u_2(\Omega') = 1-(1-q_2)B(q_4), \quad u_3(\Omega') = 1-(1-q_3)B(q_1), \quad u_4(\Omega') = 1-(1-q_4)B(q_2),$$

and

$$u_1(\Omega'') = 1-(1-q_1)B(q_4), \quad u_2(\Omega'') = 1-(1-q_2)B(q_3), \quad u_3(\Omega'') = 1-(1-q_3)B(q_2), \quad u_4(\Omega'') = 1-(1-q_4)B(q_1).$$

Clearly, a subject's utility is increasing in its own probability of infection. Moreover, because $h(\cdot,2)$ is increasing (see assumption 1), we have that $(1-S_p)h(1,2) \leq (1-S_p)h(2,2)$, which implies that $B(q_i)$ is increasing in $q_i$. Because the term $B(q_i)$ enters negatively in the utility function from each agent, we have that a subject's utility is decreasing in the probability of infection of the subject he is matched with. This implies that the subject with lowest utility from each group is the subject who has the *lowest* probability of infection from that group.

So suppose that, under ordered pooling, subject 1 is the one with the lowest utility, i.e., suppose that

$$u_1(\Omega^*) = \min_{i\in\{1,2,3,4\}} u_i(\Omega^*).$$

In this case, we have that

$$u_1(\Omega') = 1-(1-q_1)B(q_3) \leq 1-(1-q_1)B(q_2) = u_1(\Omega^*) = \min_{i\in\{1,2,3,4\}} u_i(\Omega^*),$$

and

$$u_1(\Omega'') = 1-(1-q_1)B(q_4) \leq 1-(1-q_1)B(q_2) = u_1(\Omega^*) = \min_{i\in\{1,2,3,4\}} u_i(\Omega^*),$$

so that ordered pooling maximizes the minimum utility.

Suppose, instead, that subject 3 is the one with lowest utility under ordered pooling, i.e., suppose that

$$u_3(\Omega^*) = \min_{i\in\{1,2,3,4\}} u_i(\Omega^*).$$

In this case, we have that

$$u_2(\Omega') = 1-(1-q_2)B(q_4) \leq 1-(1-q_3)B(q_4) = u_3(\Omega^*) = \min_{i\in\{1,2,3,4\}} u_i(\Omega^*),$$

and

$$u_1(\Omega'') = 1-(1-q_1)B(q_4) \leq 1-(1-q_3)B(q_4) = u_3(\Omega^*) = \min_{i\in\{1,2,3,4\}} u_i(\Omega^*),$$

so that again, ordered pooling maximizes the minimum utility.

■

**Proof of proposition 5**

For a given pool size of $k \in \mathbb{N}$, we can assume, without loss of generality, that there are only two pools (i.e., that $m = 2$).[2] Consider the following ordered partition of $S = \{1,2,\cdots,n\}$

$$\Omega^* = \{G_1^*, G_2^*\},$$

where

$$G_1^* = \{1,2,\cdots,k\}$$

and

$$G_2^* = \{k+1, k+2, \cdots, n\},$$

and $|G_1^*| = |G_2^*| = k$.

I) Suppose that $\theta = 1$. Consider any partition $\Omega = \{G_1, G_2\}$, such that $|G_1| = |G_2| = k$. For each group $\tilde{G}_g \subseteq \Omega$, define

$$A_{i,G_g} \equiv \sum_{I=0}^{k-1}(1 - S_e h(I+1,k))P_{G\setminus\{i\}(I)},$$

i.e., $A_{i,G_g}$ is the probability that a subject in group $G_g$ is *not* detected as infected given that this subject is infected.

Then, the utility from each subject $i \in \{1,2,\cdots,n\}$ under partition $\Omega$ is given by

$$u_i(\Omega) = \begin{cases} 1 - q_i A_{i,G_1}, & \text{if } i \in G_1 \\ 1 - q_i A_{i,G_2}, & \text{if } i \in G_2 \end{cases}$$

Clearly, the utility from each subject is decreasing in their own probability of infection. Because the dilution function $h(\cdot,k)$ is increasing, we also have that, for any group $G_g$ and any $j \in G_g\setminus\{i\}$,

[2] To extend the result for $m > 2$, one only needs to follow an algorithm similar to the one presented in the proofs of theorems 1, 2 and 3.

$A_{i,G_g}$ is decreasing in $q_j$. This implies that the utility from any subject is *increasing* in the probability of infection of the subjects he is matched with. Therefore, the subject with lowest utility from each group is the subject who has the highest probability of infection in that group.

Because under ordered pooling subject $n$ (i.e., the one with highest probability of infection) is matched with the the $k-1$ subjects with highest probability of infection excluding subject $n$, we have that subject $n$'s utility is maximized under ordered pooling. Therefore,

$$u_n(\Omega^*) \geq u_n(\Omega) \geq \min_{i \in S} u_i(\Omega),$$

as we wanted to show.

II) Suppose that $\theta = 0$. Consider any partition $\Omega = \{G_1, G_2\}$, such that $|G_1| = |G_2| = k$. For each group $G_g \subseteq \Omega$, define

$$B_{i,G_g} \equiv \sum_{I=0}^{k-1} (1 - S_p) h(I,k) P_{G_g \setminus \{i\}(I)},$$

i.e., $B_{i,G_g}$ is the probability that a subject in group $G_g$ *is* detected as infected given that this subject is *not* infected.

Then, the utility from each subject $i \in \{1, 2, \cdots, n\}$ under partition $\Omega$ is given by

$$u_i(\Omega) = \begin{cases} 1 - (1 - q_i) A_{i,G_1}, & \text{if } i \in G_1 \\ 1 - (1 - q_i) A_{i,G_2}, & \text{if } i \in G_2 \end{cases}$$

Clearly, the utility from each subject is *increasing* in their own probability of infection. Because the dilution function $h(\cdot, k)$ is increasing, we also have that, for any group $G_g$ and any $j \in G_g \setminus \{i\}$, $B_{i,G_g}$ is *increasing* in $q_j$. This implies that the utility from any subject is *decreasing* in the probability of infection of the subjects he is matched with. Therefore, the subject with lowest utility from each group is the subject who has the *lowest* probability of infection in that group.

Because under ordered pooling subject 1 (i.e., the one with lowest probability of infection) is matched with the the $k-1$ subjects with lowest probability of infection excluding subject 1, we have that subject 1's utility is maximized under ordered pooling. Therefore,

$$u_n(\Omega^*) \geq u_n(\Omega) \geq \min_{i \in S} u_i(\Omega),$$

as we wanted to show.

■

**Proof of theorem 4**

Theorem 4 follows directly from proposition 5 and theorems 2 and 3, and the fact that, if a partition maximizes the sum of utilities and minimum utility, it also maximizes the utilitarian max-min welfare function for any parameter $\alpha \in [0, 1]$. ■

## H. Proof of proposition 6

For a given pool of size $k$, and for a given patient $i$ belonging to this pool, let $P_{-i}(I)$ denote the probability that exactly $I \leq k-1$ of the other members in the pool are infected. Then the probability that patient $i$ is incorrectly diagnosed as not infected is given by

$$q_i \sum_{I=0}^{k-1} [F(I+1)(1-S_e) + (1-F(I+1))] P_{-i}(I),$$

which is greater than or equal to

$$q_i \sum_{I=0}^{k-1} [F(I+1)(1-S_e) + (1-F(I+1))(1-S_e)] P_{-i}(I) = q_i(1-S_e),$$

the probability the subject is incorrectly diagnosed as not infected if he is tested individually. ■

## I. Calibrating the dilution function for Chlamydia

Following Aprahamian, Bish and Bish (2018), we assume that the dilution function for pooled testing for Chlamydia follows the format

$$\hat{h}(I,k) = (1-S_p) + (S_e + S_p - 1)(I/k)^{\alpha}. \tag{31}$$

In their experiments, Kacena et al. (1998*a*) found that pools of size $k = 4$ exhibited perfect sensitivity and specificity of

$$1 - \hat{h}(0,4) = 97/98 = .98.$$

Because pooled testing with pools of size $k = 4$ exhibited perfect sensitivity, individual testing should also exhibit perfect sensitivity, so we set $S_e = 1$. Assuming that specificity is constant among all pool sizes, we must have $S_p = 1 - \hat{h}(0,4) = 97/98$.

In Kacena et al. (1998*a*), the prevalence of Chlamydia for pools of size $k = 10$ was given by $\mu_r = 62/520$, and the proportion of times infection was detected in infected pooled samples of size $k = 10$ was given by 37/38. So, following Aprahamian, Bish and Bish (2018), we set $\alpha$ so that

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

i.e., we choose $\alpha$ so that the sensitivity for pools of size 10 obtained in Kacena et al. (1998*a*) is consistent with the dilution function $\hat{h}$. This procedure yields $\alpha = 0.014302$.

## J. Simulations

In this section we use simulated data to compare the performance of ordered vs. random pooling for different combinations of dilution functions and distribution of priors. The family of prior distributions used in this section are the beta distributions displayed in figure 1 with their corresponding prevalence rates (i.e., the mean of the distribution). For a given distribution, we assume that subjects' probability of infection are independently drawn from that distribution.

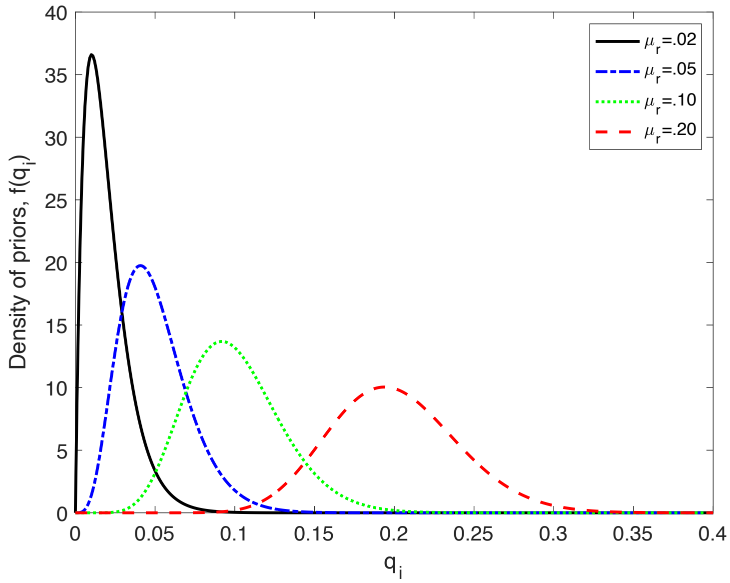



**Figure 1** Possible distribution of priors, and the prevalence associated with each distribution.

Comparisons of random vs. ordered pooling are depicted in figures 2, 3 and 4, which display the simulated percentage reduction in the expected number of tests, expected classification errors, and minimum utility when $\theta = 0$ (figure 3) and minimum utility when $\theta = 1$ (figure 4). In the simulations we varied the dilution effect starting with the case in which there is no dilution (figures 2a, 3a and 4a) and then progressively increasing the dilution effect until we reach the extreme case in which the probability of detecting infection given that there is at lest one healthy subject in the pool equals to one minus the specificity of an individual test (figures 2f, 3f and 4f).

From figure 2 we can see that, consistent with theorem 1, whenever the dilution function is concave, ordered pooling always generates a lower expected number of tests as compared to random pooling, except when the dilution function is linear, in which case any matching criteria produces the same expected number of tests (see proposition 2).

As to classification errors, notice that, consistent with corollary 4, in all of the simulations ordered pooling yields a lower expected number of both types of classification errors when pools are of size $k = 2$.

Moreover, from figure 3a we can see that, in the absence of dilution effects, either matching criteria generates the same expected number of false negatives, which is consistent with the results from proposition 1.

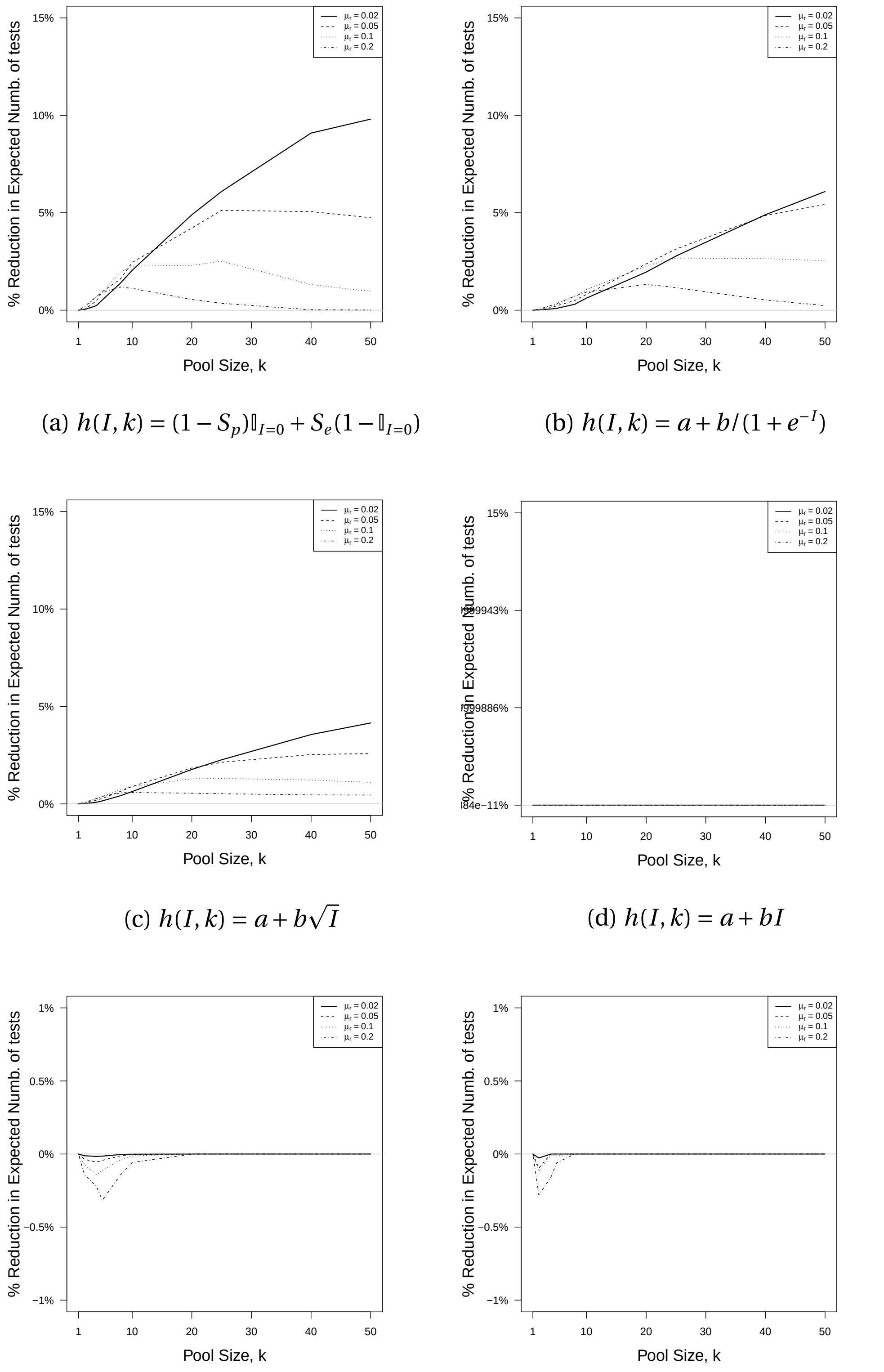
But from figure 3b we can see that when $h(I, k) = a + b/(1 + e^{-I})$, a case in which dilution effects exist but are relatively small (the shape of this dilution function is displayed in figure 1), then random pooling may potentially produce a lower expected number of false negatives than random pooling. This happens because in this case hypothesis 1 is not satisfied, so that other matching criteria, such as random pooling, may potentially produce less false negatives. But notice that in the instances in which ordered pooling produces more false negatives than random pooling, the magnitude of this difference tends to be small, whereas when the opposite holds the difference can be significantly larger. So even if ordered pooling does not minimize the expected number of false negatives under small dilution effects, it may still perform better that other conventional matching criteria, such as random pooling.

Figure 3c illustrates that when $h(\cdot,k)$ displays a mild concavity, ordered pooling performs better than random pooling in terms of minimizing false negatives. That happens because, when $h(\cdot,k)$ is "not too concave", hypothesis 1 holds, in which the ordered partition minimizes the expected number of false negatives.

For figures 3d, 3e and 3f, the dilution function is convex, which is why random pooling generates a lower expected number of false negatives (see theorem 2).

Now looking at the expected number of false positives, we can see from figure 4 that this attribute goes in the opposite direction of the expected number of false negatives, which is consistent with theorems 2 and **??**. The only instances in which ordered pooling is guaranteed to minimize both the expected number of false negatives and false positives is when the dilution function $h(\cdot,k)$ is not too concave and not too convex for a given $k$, i.e., when it satisfies both hypothesis 1 and hypothesis 2, such as when the dilution function is linear (figure 4e).

As to the minimum utility we can see that in general the minimum utility is higher under ordered pooling. In the instances the minimum utility is higher under random pooling, the differences in utilities are usually small, suggesting that ordered pooling performs well in terms of equity.




(a) $h(I,k) = (1 - S_p)\mathbb{I}_{I=0} + S_e(1 - \mathbb{I}_{I=0})$

(b) $h(I,k) = a + b/(1 + e^{-I})$

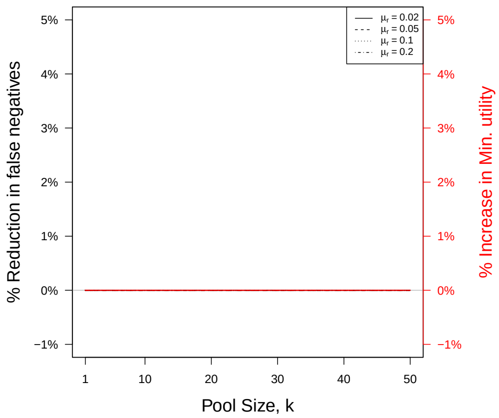
(c) $h(I,k) = a + b\sqrt{I}$

(d) $h(I,k) = a + bI$

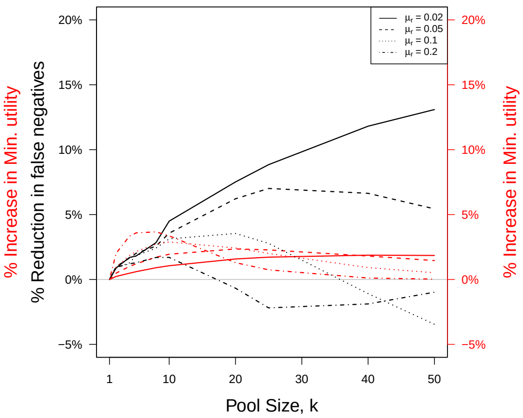
(e) $h(I,k) = a + b\exp(I)$

(f) $h(I,k) = (1 - S_p)\mathbb{I}_{I<k} + S_e(1 - \mathbb{I}_{I=k})$

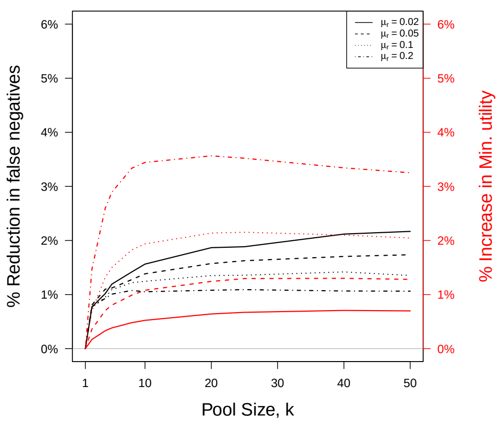
**Figure 2** Percentage reduction of the expected number of tests and expected number of false negatives obtained by performing ordered pooling as opposed to random pooling. The simulations used $S_e = .97$, $S_p = .95$ and $n = 1{,}000$, and the probabilities of infection, $\{q_i\}_{i=1}^n$, were drawn from one of the beta distributions depicted in figure 1. Each plot is based on a different dilution function $h$. The parameters $a$ and $b$ of each dilution function were chosen so that $h(0,k) = 1 - S_p$ and $h(k,k) = S_e$.



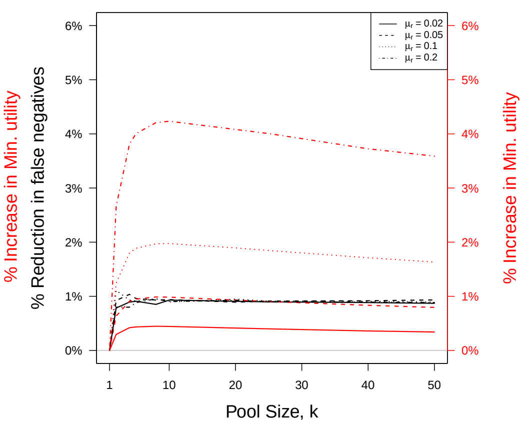

(a) $h(I,k) = (1-S_p)\mathbb{I}_{I=0} + S_e(1-\mathbb{I}_{I=0})$



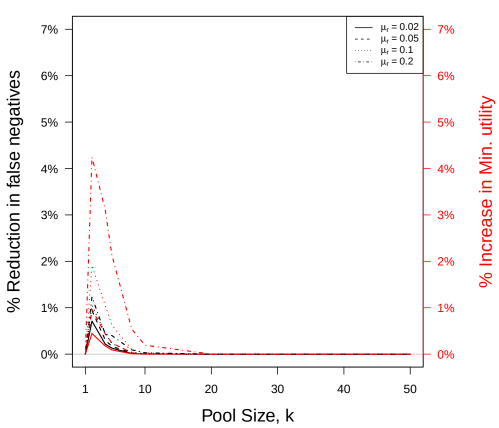

(b) $h(I,k) = a + b/(1+e^{-I})$



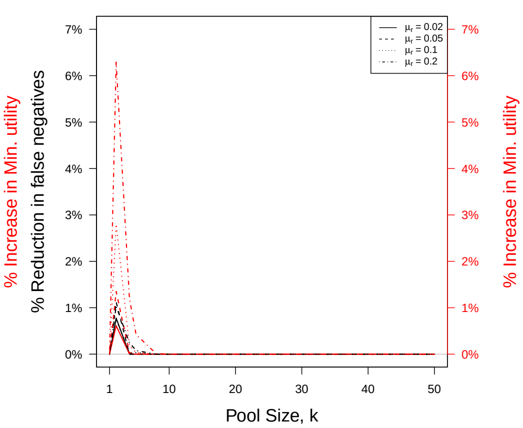

(c) $h(I,k) = a + b\sqrt{I}$




(d) $h(I,k) = a + bI$

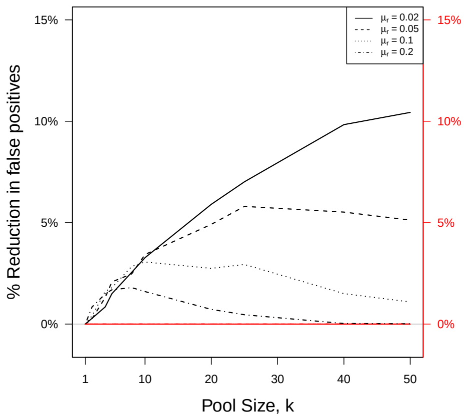



(e) $h(I,k) = a + b\exp(I)$

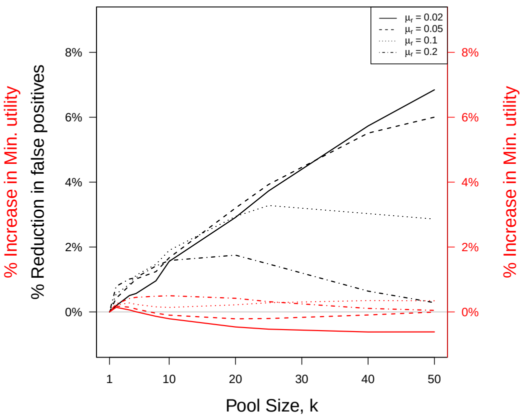



(f) $h(I,k) = (1-S_p)\mathbb{I}_{I<k} + S_e(1-\mathbb{I}_{I=k})$

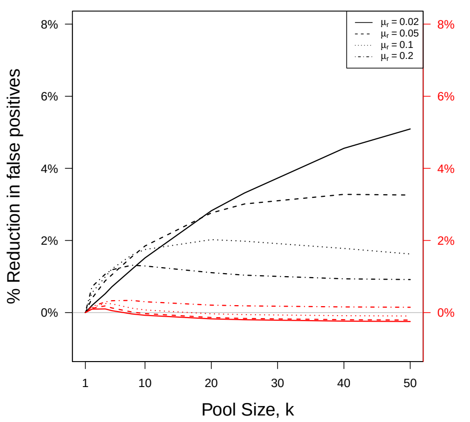
**Figure 3** Percentage reduction of the expected number of false negatives and percentage increase in the minimum utility when all weight on misclassification is put on false negative errors (i.e., $\theta = 0$). The simulations used $S_e = .97$, $S_p = .95$ and $n = 1{,}000$, and the probabilities of infection, $\{q_i\}_{i=1}^n$, were drawn from a beta distribution using different $\alpha$ and $\beta$ parameters. Each graph uses a different dilution function $h(\cdot,\cdot)$. The parameters $a$ and $b$ from each dilution function were chosen so that $h(0,k) = 1 - S_p$ and $h(k,k) = S_e$.



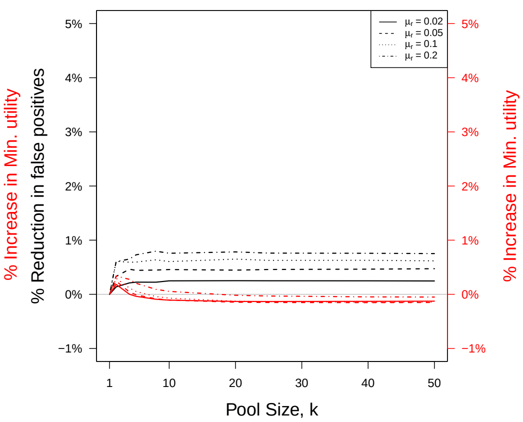

(a) $h(I,k) = (1-S_p)\mathbb{I}_{I=0} + S_e(1-\mathbb{I}_{I=0})$



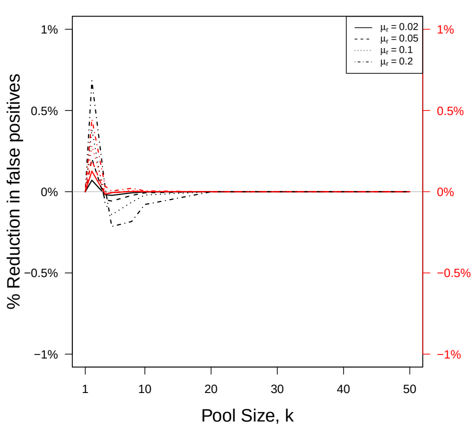

(b) $h(I,k) = a + b/(1+e^{-I})$



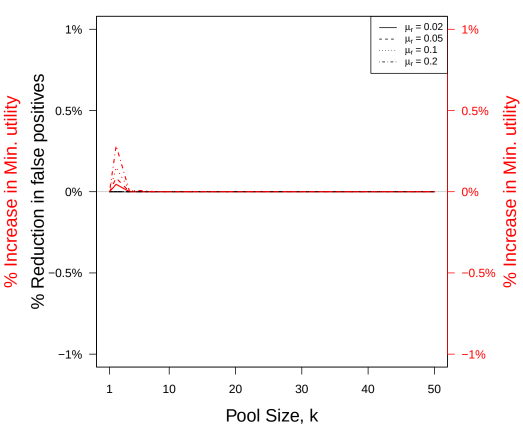

(c) $h(I,k) = a + b\sqrt{I}$




(d) $h(I,k) = a + bI$




(e) $h(I,k) = a + b\exp(I)$




(f) $h(I,k) = (1-S_p)\mathbb{I}_{I<k} + S_e(1-\mathbb{I}_{I=k})$

**Figure 4** Percentage reduction in the expected number of false positives and percentage increase in the minimum utility when all weight on misclassification is put on false positive errors (i.e., $\theta = 1$). The simulations used $S_e = .97$, $S_p = .95$ and $n = 1{,}000$, and the probabilities of infection, $\{q_i\}_{i=1}^n$, were drawn from a beta distribution using different $\alpha$ and $\beta$ parameters. Each graph uses a different dilution function $h(\cdot,\cdot)$. The parameters $a$ and $b$ from each dilution function were chosen so that $h(0,k) = 1 - S_p$ and $h(k,k) = S_e$.