Supporting Information for "Group testing regression model estimation when case identification is a goal" by B. Zhang, C. Bilder, and J. Tebbs

TABLE OF CONTENTS

WEB APPENDIX A

This appendix shows how to find ω_{ik} for Dorfman's protocol when $Z_k = 1$. We first express the conditional mean ω_{ik} as

$$P(\tilde{Y}_{ik}=1 \mid Y_{1k}=y_{1k},\ldots,Y_{I_k k}=y_{I_k k},Z_k=1)=\frac{P(\tilde{Y}_{ik}=1,Y_{1k}=y_{1k},\ldots,Y_{I_k k}=y_{I_k k},Z_k=1)}{P(Y_{1k}=y_{1k},\ldots,Y_{I_k k}=y_{I_k k},Z_k=1)}$$

$$=\frac{\sum_{\tilde{\boldsymbol{y}}_{-i,k}} P(\tilde{Y}_{ik}=1,\tilde{\boldsymbol{Y}}_{-i,k}=\tilde{\boldsymbol{y}}_{-i,k},Y_{1k}=y_{1k},\ldots,Y_{I_k k}=y_{I_k k},Z_k=1)}{\sum_{\tilde{\boldsymbol{y}}_k} P(\tilde{\boldsymbol{Y}}_k=\tilde{\boldsymbol{y}}_k,Y_{1k}=y_{1k},\ldots,Y_{I_k k}=y_{I_k k},Z_k=1)}, \tag{1}$$

where $\tilde{\boldsymbol{Y}}_k=(\tilde{Y}_{1k},\ \ldots,\ \tilde{Y}_{I_k k})'$ and $\tilde{\boldsymbol{Y}}_{-i,k}$ is the same as $\tilde{\boldsymbol{Y}}_k$ but without $\tilde{Y}_{i',k}$. Examining the summand of the numerator in (1), we have

$$\begin{aligned}
&P(\tilde{Y}_{ik}=1,\tilde{\boldsymbol{Y}}_{-i,k}=\tilde{\boldsymbol{y}}_{-i,k},Y_{1k}=y_{1k},\ldots,Y_{I_k k}=y_{I_k k},Z_k=1)\\
&=P(Y_{1k}=y_{1k},\ldots,Y_{I_k k}=y_{I_k k},Z_k=1 \mid \tilde{Y}_{ik}=1,\tilde{\boldsymbol{Y}}_{-i,k}=\tilde{\boldsymbol{y}}_{-i,k})P(\tilde{Y}_{ik}=1)P(\tilde{\boldsymbol{Y}}_{-i,k}=\tilde{\boldsymbol{y}}_{-i,k})\\
&=\tilde{p}_{ik}P(Y_{ik}=y_{ik} \mid \tilde{Y}_{ik}=1)P(Z_k=1 \mid \tilde{Y}_{ik}=1,\tilde{\boldsymbol{Y}}_{-i,k}=\tilde{\boldsymbol{y}}_{-i,k})\prod_{i'\neq i} P(Y_{i'k}=y_{i'k} \mid \tilde{Y}_{i'k}=\tilde{y}_{i'k})P(\tilde{Y}_{i'k}=\tilde{y}_{i'k})\\
&=\eta\tilde{p}_{ik}P(Y_{ik}=y_{ik} \mid \tilde{Y}_{ik}=1)\prod_{i'\neq i} P(Y_{i'k}=y_{i'k} \mid \tilde{Y}_{i'k}=\tilde{y}_{i'k})P(\tilde{Y}_{i'k}=\tilde{y}_{i'k}),
\end{aligned} \tag{2}$$

where we use the standard assumption that test outcomes are conditionally independent given the true outcomes (Litvak et al. 1994). Similarly, in the denominator of (1), we have

$$\begin{aligned}
&P(\tilde{\boldsymbol{Y}}_k=\tilde{\boldsymbol{y}}_k,Y_{1k}=y_{1k},\ldots,Y_{I_k k}=y_{I_k k},Z_k=1)\\
&=P(Y_{1k}=y_{1k},\ldots,Y_{I_k k}=y_{I_k k},Z_k=1 \mid \tilde{\boldsymbol{Y}}_k=\tilde{\boldsymbol{y}}_k)P(\tilde{\boldsymbol{Y}}_k=\tilde{\boldsymbol{y}}_k)\\
&=P(Z_k=1 \mid \tilde{\boldsymbol{Y}}_k=\tilde{\boldsymbol{y}}_k)\prod_{i'=1}^{I_k} P(Y_{i'k}=y_{i'k} \mid \tilde{Y}_{i'k}=\tilde{y}_{i'k})P(\tilde{Y}_{i'k}=\tilde{y}_{i'k}).
\end{aligned} \tag{3}$$

Substituting (2) and (3) into (1) results in

$$w_{ik}=\frac{\eta\tilde{p}_{ik}P(Y_{ik}=y_{ik} \mid \tilde{Y}_{ik}=1)\sum_{\tilde{\boldsymbol{y}}_{-i,k}}\prod_{i'\neq i} P(Y_{i'k}=y_{i'k} \mid \tilde{Y}_{i'k}=\tilde{y}_{i'k})P(\tilde{Y}_{i'k}=\tilde{y}_{i'k})}{\sum_{\tilde{\boldsymbol{y}}_k} P(Z_k=1 \mid \tilde{\boldsymbol{Y}}_k=\tilde{\boldsymbol{y}}_k)\prod_{i'=1}^{I_k} P(Y_{i'k}=y_{i'k} \mid \tilde{Y}_{i'k}=\tilde{y}_{i'k})P(\tilde{Y}_{i'k}=\tilde{y}_{i'k})}. \tag{4}$$

Equation (4) is very difficult to compute for large group sizes, because the number of summands within it increases exponentially as the group size increases (e.g., there are 2^{I_k} terms to sum in the denominator for group k). Fortunately, we can reformulate the numerator and denominator to make Equation (4) computationally feasible for large group sizes. The denominator can be written as

$$
\begin{aligned}
&\sum_{\tilde{\boldsymbol{y}}_k} P(Z_k = 1 \mid \tilde{\boldsymbol{Y}}_k = \tilde{\boldsymbol{y}}_k) \prod_{i'=1}^{I_k} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}) \\
&= P(Z_k = 1 \mid \tilde{Z}_k = 0) \prod_{i'=1}^{I_k} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = 0)(1 - \tilde{p}_{i'k}) + \\
&\quad P(Z_k = 1 \mid \tilde{Z}_k = 1) \sum_{\tilde{\boldsymbol{y}}_k \neq \mathbf{0}} \prod_{i'=1}^{I_k} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}) \\
&= (1 - \delta) \prod_{i'=1}^{I_k} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = 0)(1 - \tilde{p}_{i'k}) + \\
&\quad \eta \left[\left(\prod_{i'=1}^{I_k} \sum_{\tilde{y}_{i'k}=0}^{1} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}) \right) - \left(\prod_{i'=1}^{I_k} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = 0)(1 - \tilde{p}_{i'k}) \right) \right] \\
&= \eta \prod_{i'=1}^{I_k} \sum_{\tilde{y}_{i'k}=0}^{1} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}) + \varphi \prod_{i'=1}^{I_k} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = 0)(1 - \tilde{p}_{i'k}),
\end{aligned}
$$

where $\varphi = 1 - \eta - \delta$ and we make use of the relation

$$
\sum_{\tilde{\boldsymbol{y}}_k} \prod_{i'=1}^{I_k} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}) = \prod_{i'=1}^{I_k} \sum_{\tilde{y}_{i'k}=0}^{1} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}).
$$

Using the same technique, the numerator of (4) can be re-written in a similar manner leading to

$$
w_{ik} = \frac{\eta \tilde{p}_{ik} P(Y_{ik} = y_{ik} \mid \tilde{Y}_{ik} = 1) \prod_{i' \neq i} \sum_{\tilde{y}_{i'k}=0}^{1} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k})}{\varphi \prod_{i'=1}^{I_k} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = 0)(1 - \tilde{p}_{i'k}) + \eta \prod_{i'=1}^{I_k} \sum_{\tilde{y}_{i'k}=0}^{1} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k})}. \tag{5}
$$

The denominator (numerator) of Equation (5) is the product of only I_k ($I_k - 1$) terms. Therefore, this formula makes finding ω_{ik} possible even for large group sizes.

WEB APPENDIX B

This appendix shows how to find ω_{ik} for scenarios 1) to 5) of the halving protocol. The derivation is very similar to that of Dorfman's protocol.

(i) $Z_k = 0$:

This is the same as for Dorfman.

(ii) $Z_k = 1,\ Z_{k1} = 0,\ Z_{k2} = 0$:

We need to find

$$\omega_{ik} = P(\tilde{Y}_{ik} = 1 \mid Z_k = 1, Z_{k1} = 0, Z_{k2} = 0) = \frac{P(Z_k = 1, Z_{k1} = 0, Z_{k2} = 0, \tilde{Y}_{ik} = 1)}{P(Z_k = 1, Z_{k1} = 0, Z_{k2} = 0)}.$$

For the denominator, we begin by including $\tilde{\boldsymbol{Y}}_k = (\tilde{Y}_{1k},\ \ldots,\ \tilde{Y}_{I_k k})' = (\tilde{\boldsymbol{Y}}_{k1}',\ \tilde{\boldsymbol{Y}}_{k2}')'$ in the expression to obtain

$$\begin{aligned}
&P(Z_k = 1, Z_{k1} = 0, Z_{k2} = 0) \\
&= \sum_{\tilde{\boldsymbol{y}}_k} P(Z_k = 1, Z_{k1} = 0, Z_{k2} = 0 \mid \tilde{\boldsymbol{Y}}_{k1} = \tilde{\boldsymbol{y}}_{k1}, \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) P(\tilde{\boldsymbol{Y}}_{k1} = \tilde{\boldsymbol{y}}_{k1}) P(\tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) \\
&= \sum_{\tilde{\boldsymbol{y}}_{k2}} \sum_{\tilde{\boldsymbol{y}}_{k1}} P(Z_k = 1 \mid \tilde{\boldsymbol{Y}}_{k1} = \tilde{\boldsymbol{y}}_{k1}, \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) P(Z_{k1} = 0 \mid \tilde{\boldsymbol{Y}}_{k1} = \tilde{\boldsymbol{y}}_{k1}) P(\tilde{\boldsymbol{Y}}_{k1} = \tilde{\boldsymbol{y}}_{k1}) \\
&\qquad \times P(Z_{k2} = 0 \mid \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) P(\tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) \\
&= \delta^2 (1-\delta) \prod_{i'=1}^{I_k} (1 - \tilde{p}_{i'k}) + \delta\eta(1-\eta) \prod_{i' \in k1} (1 - \tilde{p}_{i'k}) \left[1 - \prod_{i' \in k2} (1 - \tilde{p}_{i'k})\right] + \\
&\quad \delta\eta(1-\eta) \left[1 - \prod_{i' \in k1} (1 - \tilde{p}_{i'k})\right] \prod_{i' \in k2} (1 - \tilde{p}_{i'k}) + \eta(1-\eta)^2 \left[1 - \prod_{i' \in k1} (1 - \tilde{p}_{i'k})\right]\left[1 - \prod_{i' \in k2} (1 - \tilde{p}_{i'k})\right],
\end{aligned}$$

where we let $i \in kj$ denote those individuals within the j^{th} subgroup (j = 1, 2) and we again use the standard assumption that test outcomes are conditionally independent given the true outcomes (Litvak et al. 1994).

Without loss of generality, we assume here and throughout this appendix that individual i is within the first subgroup ($i \in k1$). The numerator can be written as

$$\begin{aligned}
&P(\tilde{Y}_{ik} = 1, Z_k = 1, Z_{k1} = 0, Z_{k2} = 0) \\
&= \sum_{\tilde{\boldsymbol{y}}_{-i,k}} P(Z_k = 1, Z_{k1} = 0, Z_{k2} = 0 \mid \tilde{Y}_{ik} = 1, \tilde{\boldsymbol{Y}}_{-i,k} = \tilde{\boldsymbol{y}}_{-i,k}) P(\tilde{Y}_{ik} = 1, \tilde{\boldsymbol{Y}}_{-i,k} = \tilde{\boldsymbol{y}}_{-i,k}) \\
&= \eta(1-\eta) \sum_{\tilde{\boldsymbol{y}}_{-i,k}} P(Z_{k2} = 0 \mid \tilde{Y}_{ik} = 1, \tilde{\boldsymbol{Y}}_{-i,k} = \tilde{\boldsymbol{y}}_{-i,k}) P(\tilde{\boldsymbol{Y}}_{-i,k} = \tilde{\boldsymbol{y}}_{-i,k}) \tilde{p}_{ik}
\end{aligned}$$

$$
\begin{aligned}
&= \eta(1-\eta)\tilde{p}_{ik} \sum_{\tilde{\boldsymbol{y}}_{-i,k1}} P(\tilde{\boldsymbol{Y}}_{-i,k1} = \tilde{\boldsymbol{y}}_{-i,k1}) \sum_{\tilde{\boldsymbol{y}}_{k2}} P(Z_{k2} = 0 \mid \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) P(\tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) \\
&= \eta(1-\eta)\tilde{p}_{ik} \sum_{\tilde{\boldsymbol{y}}_{k2}} P(Z_{k2} = 0 \mid \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) P(\tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) \\
&= \eta(1-\eta)\tilde{p}_{ik} \left\{ \delta \prod_{i' \in k2} (1-\tilde{p}_{i'k}) + (1-\eta)\left[1 - \prod_{i' \in k2} (1-\tilde{p}_{i'k})\right] \right\},
\end{aligned}
$$

where $\tilde{\boldsymbol{Y}}_{-i,k} = \{\tilde{Y}_{i',k} : i' = 1,\ldots, I_k, i' \neq i\}$ is the vector of all true individual statuses excluding the i^{th} subject in group k and $\tilde{\boldsymbol{Y}}_{-i,k1} = \{\tilde{Y}_{i',k1} : i' \in k1, i' \neq i\}$ is the vector of all true individual statuses excluding the i^{th} subject in subgroup $k1$.

(iii) $Z_k = 1,\ Z_{k1} = 1,\ Z_{k2} = 0$:

We need to find

$$
\omega_{ik} = P(\tilde{Y}_{ik} = 1 \mid Z_k = 1, Z_{k1} = 1, \boldsymbol{Y}_{k1} = \boldsymbol{y}_{k1}, Z_{k2} = 0) = \frac{P(Z_k = 1, Z_{k1} = 1, \boldsymbol{Y}_{k1} = \boldsymbol{y}_{k1}, Z_{k2} = 0, \tilde{Y}_{ik} = 1)}{P(Z_k = 1, Z_{k1} = 1, \boldsymbol{Y}_{k1} = \boldsymbol{y}_{k1}, Z_{k2} = 0)}.
$$

Using the same technique as above, we calculate the denominator to be

$$
\begin{aligned}
&P(Z_k = 1, Z_{k1} = 1, \boldsymbol{Y}_{k1} = \boldsymbol{y}_{k1}, Z_{k2} = 0) \\
&= \sum_{\tilde{\boldsymbol{y}}_k} P(Z_k = 1, Z_{k1} = 1, \boldsymbol{Y}_{k1} = \boldsymbol{y}_{k1}, Z_{k2} = 0 \mid \tilde{\boldsymbol{Y}}_{k1} = \tilde{\boldsymbol{y}}_{k1}, \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) P(\tilde{\boldsymbol{Y}}_{k1} = \tilde{\boldsymbol{y}}_{k1}) P(\tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) \\
&= \sum_{\tilde{\boldsymbol{y}}_{k2}} \sum_{\tilde{\boldsymbol{y}}_{k1}} P(Z_k = 1 \mid \tilde{\boldsymbol{Y}}_{k1} = \tilde{\boldsymbol{y}}_{k1}, \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) P(Z_{k1} = 1 \mid \tilde{\boldsymbol{Y}}_{k1} = \tilde{\boldsymbol{y}}_{k1}) \prod_{i' \in k1} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) \times \\
&\qquad P(\tilde{\boldsymbol{Y}}_{k1} = \tilde{\boldsymbol{y}}_{k1}) P(Z_{k2} = 0 \mid \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) P(\tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) \\
&= \delta(1-\delta)^2 \prod_{i' \in k1} \Big(P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = 0)(1-\tilde{p}_{i'k}) \Big) \prod_{i' \in k2} (1-\tilde{p}_{i'k}) + \\
&\quad (1-\delta)\eta(1-\eta) \prod_{i' \in k1} \Big(P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = 0)(1-\tilde{p}_{i'k}) \Big) \left[1 - \prod_{i' \in k2} (1-\tilde{p}_{i'k})\right] + \\
&\quad \delta\eta^2 \sum_{\tilde{\boldsymbol{y}}_{k1} \neq \boldsymbol{0}} \left(\prod_{i' \in k1} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}) \right) \prod_{i' \in k2} (1-\tilde{p}_{i'k}) + \\
&\quad \eta^2(1-\eta) \sum_{\tilde{\boldsymbol{y}}_{k1} \neq \boldsymbol{0}} \left(\prod_{i' \in k1} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}) \right) \left[1 - \prod_{i' \in k2} (1-\tilde{p}_{i'k})\right].
\end{aligned}
$$

The numerator can be expressed as

$$
\begin{aligned}
&P(\tilde{Y}_{ik} = 1, Z_k = 1, Z_{k1} = 1, \boldsymbol{Y}_{k1} = \boldsymbol{y}_{k1}, Z_{k2} = 0) \\
&= \sum_{\tilde{\boldsymbol{y}}_{-i,k}} P(Z_k = 1, Z_{k1} = 1, \boldsymbol{Y}_{k1} = \boldsymbol{y}_{k1}, Z_{k2} = 0 \mid \tilde{Y}_{ik} = 1, \tilde{\boldsymbol{Y}}_{-i,k} = \tilde{\boldsymbol{y}}_{-i,k}) P(\tilde{Y}_{ik} = 1, \tilde{\boldsymbol{Y}}_{-i,k} = \tilde{\boldsymbol{y}}_{-i,k}) \\
&= \eta^2 \sum_{\tilde{\boldsymbol{y}}_{-i,k}} P(Y_{ik} = y_{ik} \mid \tilde{Y}_{ik} = 1) \prod_{i' \in k1, i' \neq i} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(Z_{k2} = 0 \mid \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) P(\tilde{\boldsymbol{Y}}_{-i,k} = \tilde{\boldsymbol{y}}_{-i,k}) \tilde{p}_{ik} \\
&= \eta^2 \tilde{p}_{ik} P(Y_{ik} = y_{ik} \mid \tilde{Y}_{ik} = 1) \sum_{\tilde{\boldsymbol{y}}_{-i,k}} \left(\prod_{i' \in k1, i' \neq i} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(Z_{k2} = 0 \mid \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) P(\tilde{\boldsymbol{Y}}_{-i,k} = \tilde{\boldsymbol{y}}_{-i,k}) \right)
\end{aligned}
$$

$$= \eta^2 \tilde{p}_{ik} P(Y_{ik} = y_{ik} \mid \tilde{Y}_{ik} = 1) \sum_{\tilde{\boldsymbol{y}}_{-i,k1}} \left(\prod_{i' \in k1, i' \neq i} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}) \right) \times$$
$$\sum_{\tilde{\boldsymbol{y}}_{k2}} \Big(P(Z_{k2} = 0 \mid \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) P(\tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) \Big)$$
$$= \eta^2 \tilde{p}_{ik} P(Y_{ik} = y_{ik} \mid \tilde{Y}_{ik} = 1) \sum_{\tilde{\boldsymbol{y}}_{-i,k1}} \left(\prod_{i' \in k1, i' \neq i} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}) \right) \times$$
$$\left\{ \delta \prod_{i' \in k2} (1 - \tilde{p}_{i'k}) + (1 - \eta) \left[1 - \prod_{i' \in k2} (1 - \tilde{p}_{i'k}) \right] \right\}.$$

(iv) $Z_k = 1,\ Z_{k1} = 0,\ Z_{k2} = 1$:

We need to find

$$\omega_{ik} = P(\tilde{Y}_{ik} = 1 \mid Z_k = 1, Z_{k1} = 0, Z_{k2} = 1, \boldsymbol{Y}_{k2} = \boldsymbol{y}_{k2}) = \frac{P(Z_k = 1, Z_{k1} = 0, Z_{k2} = 1, \boldsymbol{Y}_{k2} = \boldsymbol{y}_{k2}, \tilde{Y}_{ik} = 1)}{P(Z_k = 1, Z_{k1} = 0, Z_{k2} = 1, \boldsymbol{Y}_{k2} = \boldsymbol{y}_{k2})}.$$

The denominator follows immediately from the previous result by interchanging $k1$ and $k2$:

$$P(Z_k = 1, Z_{k1} = 0, Z_{k2} = 1, \boldsymbol{Y}_{k2} = \boldsymbol{y}_{k2})$$
$$= \delta(1 - \delta)^2 \prod_{i' \in k2} \Big(P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = 0)(1 - \tilde{p}_{i'k}) \Big) \prod_{i' \in k1} (1 - \tilde{p}_{i'k}) +$$
$$(1 - \delta)\eta(1 - \eta) \prod_{i' \in k2} \Big(P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = 0)(1 - \tilde{p}_{i'k}) \Big) \left[1 - \prod_{i' \in k1} (1 - \tilde{p}_{i'k}) \right] +$$
$$\delta\eta^2 \sum_{\tilde{\boldsymbol{y}}_{k2} \neq \boldsymbol{0}} \left(\prod_{i' \in k2} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}) \right) \prod_{i' \in k1} (1 - \tilde{p}_{i'k}) +$$
$$\eta^2(1 - \eta) \sum_{\tilde{\boldsymbol{y}}_{k2} \neq \boldsymbol{0}} \left(\prod_{i' \in k2} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}) \right) \left[1 - \prod_{i' \in k1} (1 - \tilde{p}_{i'k}) \right].$$

We can show the numerator has the following form:

$$P(\tilde{Y}_{ik} = 1, Z_k = 1, Z_{k1} = 0, Z_{k2} = 1, \boldsymbol{Y}_{k2} = \boldsymbol{y}_{k2})$$
$$= \sum_{\tilde{\boldsymbol{y}}_{-i,k}} P(Z_k = 1, Z_{k1} = 0, Z_{k2} = 1, \boldsymbol{Y}_{k2} = \boldsymbol{y}_{k2} \mid \tilde{Y}_{ik} = 1, \tilde{\boldsymbol{Y}}_{-i,k} = \tilde{\boldsymbol{y}}_{-i,k}) P(\tilde{Y}_{ik} = 1, \tilde{\boldsymbol{Y}}_{-i,k} = \tilde{\boldsymbol{y}}_{-i,k})$$
$$= \eta(1 - \eta) \sum_{\tilde{\boldsymbol{y}}_{-i,k}} \prod_{i' \in k2} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(Z_{k2} = 1 \mid \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) P(\tilde{\boldsymbol{Y}}_{-i,k} = \tilde{\boldsymbol{y}}_{-i,k}) \tilde{p}_{ik}$$
$$= \eta(1 - \eta) \tilde{p}_{ik} \sum_{\tilde{\boldsymbol{y}}_{-i,k}} \left(\prod_{i' \in k2} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(Z_{k2} = 1 \mid \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) P(\tilde{\boldsymbol{Y}}_{-i,k} = \tilde{\boldsymbol{y}}_{-i,k}) \right)$$
$$= \eta(1 - \eta) \tilde{p}_{ik} \sum_{\tilde{\boldsymbol{y}}_{-i,k1}} P(\tilde{\boldsymbol{Y}}_{-i,k1} = \tilde{\boldsymbol{y}}_{-i,k1}) \sum_{\tilde{\boldsymbol{y}}_{k2}} \left(\prod_{i' \in k2} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(Z_{k2} = 1 \mid \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) P(\tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) \right)$$
$$= \eta(1 - \eta) \tilde{p}_{ik} \times \left[\eta \prod_{i' \in k2} \left(\sum_{\tilde{y}_{i'k} = 0}^{1} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}) \right) + \right.$$
$$\left. \varphi \prod_{i' \in k2} \Big(P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = 0)(1 - \tilde{p}_{i'k}) \Big) \right],$$

where $\varphi = 1 - \eta - \delta$.

(v) $Z_k = 1,\ Z_{k1} = 1,\ Z_{k2} = 1$:

We need to find

$$\begin{aligned}\omega_{ik} &= P(\tilde{Y}_{ik} = 1 \mid Z_k = 1, Z_{k1} = 1, \boldsymbol{Y}_{k1} = \boldsymbol{y}_{k1}, Z_{k2} = 1, \boldsymbol{Y}_{k2} = \boldsymbol{y}_{k2}) \\ &= \frac{P(Z_k = 1, Z_{k1} = 1, \boldsymbol{Y}_{k1} = \boldsymbol{y}_{k1}, Z_{k2} = 1, \boldsymbol{Y}_{k2} = \boldsymbol{y}_{k2}, \tilde{Y}_{ik} = 1)}{P(Z_k = 1, Z_{k1} = 1, \boldsymbol{Y}_{k1} = \boldsymbol{y}_{k1}, Z_{k2} = 1, \boldsymbol{Y}_{k2} = \boldsymbol{y}_{k2})}.\end{aligned}$$

The denominator can be written as

$$\begin{aligned}&P(Z_k = 1, Z_{k1} = 1, \boldsymbol{Y}_{k1} = \boldsymbol{y}_{k1}, Z_{k2} = 1, \boldsymbol{Y}_{k2} = \boldsymbol{y}_{k2}) \\ &= \sum_{\tilde{\boldsymbol{y}}_k} P(Z_k = 1, Z_{k1} = 1, \boldsymbol{Y}_{k1} = \boldsymbol{y}_{k1}, Z_{k2} = 1, \boldsymbol{Y}_{k2} = \boldsymbol{y}_{k2} \mid \tilde{\boldsymbol{Y}}_k = \tilde{\boldsymbol{y}}_k) P(\tilde{\boldsymbol{Y}}_k = \tilde{\boldsymbol{y}}_k) \\ &= \sum_{\tilde{\boldsymbol{y}}_k} P(Z_k = 1, Z_{k1} = 1, \boldsymbol{Y}_{k1} = \boldsymbol{y}_{k1}, Z_{k2} = 1, \boldsymbol{Y}_{k2} = \boldsymbol{y}_{k2} \mid \tilde{\boldsymbol{Y}}_{k1} = \tilde{\boldsymbol{y}}_{k1}, \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) P(\tilde{\boldsymbol{Y}}_{k1} = \tilde{\boldsymbol{y}}_{k1}) P(\tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) \\ &= \sum_{\tilde{\boldsymbol{y}}_{k2}} \sum_{\tilde{\boldsymbol{y}}_{k1}} \Big\{ P(Z_k = 1 \mid \tilde{\boldsymbol{Y}}_{k1} = \tilde{\boldsymbol{y}}_{k1}, \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) P(Z_{k1} = 1 \mid \tilde{\boldsymbol{Y}}_{k1} = \tilde{\boldsymbol{y}}_{k1}) P(Z_{k2} = 1 \mid \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) \\ &\quad \times \prod_{i' \in k1} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) \prod_{i' \in k2} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{\boldsymbol{Y}}_{k1} = \tilde{\boldsymbol{y}}_{k1}) P(\tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) \Big\} \\ &= (1-\delta)^3 \prod_{i'=1}^{I_k} \Big(P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = 0)(1 - \tilde{p}_{i'k}) \Big) + \\ &\quad \eta^2 (1-\delta) \prod_{i' \in k1} \Big(P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = 0)(1 - \tilde{p}_{i'k}) \Big) \sum_{\tilde{\boldsymbol{y}}_{k2} \neq \boldsymbol{0}} \Bigg(\prod_{i' \in k2} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}) \Bigg) + \\ &\quad \eta^2 (1-\delta) \sum_{\tilde{\boldsymbol{y}}_{k1} \neq \boldsymbol{0}} \Bigg(\prod_{i' \in k1} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}) \Bigg) \prod_{i' \in k2} \Big(P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = 0)(1 - \tilde{p}_{i'k}) \Big) + \\ &\quad \eta^3 \sum_{\tilde{\boldsymbol{y}}_{k1} \neq \boldsymbol{0}} \Bigg(\prod_{i' \in k1} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}) \Bigg) \sum_{\tilde{\boldsymbol{y}}_{k2} \neq \boldsymbol{0}} \Bigg(\prod_{i' \in k2} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}) \Bigg), \\ &\quad ,\end{aligned}$$

and the numerator can be expressed as

$$\begin{aligned}&P(\tilde{Y}_{ik} = 1, Z_k = 1, Z_{k1} = 1, \boldsymbol{Y}_{k1} = \boldsymbol{y}_{k1}, Z_{k2} = 1, \boldsymbol{Y}_{k2} = \boldsymbol{y}_{k2}) \\ &= \sum_{\tilde{\boldsymbol{y}}_{-i,k}} P(Z_k = 1, Z_{k1} = 1, \boldsymbol{Y}_{k1} = \boldsymbol{y}_{k1}, Z_{k2} = 1, \boldsymbol{Y}_{k2} = \boldsymbol{y}_{k2} \mid \tilde{Y}_{ik} = 1, \tilde{\boldsymbol{Y}}_{-i,k} = \tilde{\boldsymbol{y}}_{-i,k}) P(\tilde{Y}_{ik} = 1, \tilde{\boldsymbol{Y}}_{-i,k} = \tilde{\boldsymbol{y}}_{-i,k}) \\ &= \sum_{\tilde{\boldsymbol{y}}_{-i,k}} \eta^2 P(Y_{ik} = y_{ik} \mid \tilde{Y}_{ik} = 1) \prod_{i' \in k1, i' \neq i} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) \prod_{i' \in k2} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) \times \\ &\qquad P(Z_{k2} = 1 \mid \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) P(\tilde{\boldsymbol{Y}}_{-i,k} = \tilde{\boldsymbol{y}}_{-i,k}) \tilde{p}_{ik} \\ &= \eta^2 \tilde{p}_{ik} P(Y_{ik} = y_{ik} \mid \tilde{Y}_{ik} = 1) \times \\ &\quad \sum_{\tilde{\boldsymbol{y}}_{-i,k}} \Bigg(\prod_{i' \in k1, i' \neq i} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) \prod_{i' \in k2} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(Z_{k2} = 1 \mid \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) P(\tilde{\boldsymbol{Y}}_{-i,k} = \tilde{\boldsymbol{y}}_{-i,k}) \Bigg) \Bigg) \\ &= \eta^2 \tilde{p}_{ik} P(Y_{ik} = y_{ik} \mid \tilde{Y}_{ik} = 1) \sum_{\tilde{\boldsymbol{y}}_{-i,k1}} \Bigg(\prod_{i' \in k1, i' \neq i} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}) \Bigg) \times \\ &\quad \sum_{\tilde{\boldsymbol{y}}_{k2}} \Bigg(\prod_{i' \in k2} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(Z_{k2} = 1 \mid \tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) P(\tilde{\boldsymbol{Y}}_{k2} = \tilde{\boldsymbol{y}}_{k2}) \Bigg)\end{aligned}$$

$$= \eta^2 \tilde{p}_{ik} P(Y_{ik} = y_{ik} \mid \tilde{Y}_{ik} = 1) \sum_{\tilde{\boldsymbol{y}}_{-i,k1}} \left(\prod_{i' \in k1, i' \neq i} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}) \right) \times$$
$$\left[\eta \prod_{i' \in k2} \left(\sum_{\tilde{y}_{i'k}=0}^{1} P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = \tilde{y}_{i'k}) P(\tilde{Y}_{i'k} = \tilde{y}_{i'k}) \right) + \varphi \prod_{i' \in k2} \Big(P(Y_{i'k} = y_{i'k} \mid \tilde{Y}_{i'k} = 0)(1 - \tilde{p}_{i'k}) \Big) \right].$$

WEB APPENDIX C

This appendix gives the derivation of the expression for γ_{ij} used with the array testing protocol. We can write γ_{ij} as

$$\begin{aligned} &P(\tilde{Y}_{ij}=1 \mid \tilde{\boldsymbol{Y}}_{-i,-j}=\tilde{\boldsymbol{y}}_{-i,-j}, \boldsymbol{R}=\boldsymbol{r}, \boldsymbol{C}=\boldsymbol{c}, \boldsymbol{Y}_Q=\boldsymbol{y}_Q) \\ &= \frac{P(\tilde{Y}_{ij}=1, \tilde{\boldsymbol{Y}}_{-i,-j}=\tilde{\boldsymbol{y}}_{-i,-j}, \boldsymbol{R}=\boldsymbol{r}, \boldsymbol{C}=\boldsymbol{c}, \boldsymbol{Y}_Q=\boldsymbol{y}_Q)}{P(\tilde{\boldsymbol{Y}}_{-i,-j}=\tilde{\boldsymbol{y}}_{-i,-j}, \boldsymbol{R}=\boldsymbol{r}, \boldsymbol{C}=\boldsymbol{c}, \boldsymbol{Y}_Q=\boldsymbol{y}_Q)}. \end{aligned} \quad (1)$$

First, to find the numerator of Equation (1), we have

$$\begin{aligned} &P(\tilde{Y}_{ij}=\tilde{y}_{ij}, \tilde{\boldsymbol{Y}}_{-i,-j}=\tilde{\boldsymbol{y}}_{-i,-j}, \boldsymbol{R}=\boldsymbol{r}, \boldsymbol{C}=\boldsymbol{c}, \boldsymbol{Y}_Q=\boldsymbol{y}_Q) \\ &= P(\boldsymbol{R}=\boldsymbol{r}, \boldsymbol{C}=\boldsymbol{c}, \boldsymbol{Y}_Q=\boldsymbol{y}_Q \mid \tilde{Y}_{ij}=\tilde{y}_{ij}, \tilde{\boldsymbol{Y}}_{-i,-j}=\tilde{\boldsymbol{y}}_{-i,-j}) P(\tilde{Y}_{ij}=\tilde{y}_{ij}) P(\tilde{\boldsymbol{Y}}_{-i,-j}=\tilde{\boldsymbol{y}}_{-i,-j}) \\ &= P(\boldsymbol{R}=\boldsymbol{r}, \boldsymbol{C}=\boldsymbol{c} \mid \tilde{Y}_{ij}=\tilde{y}_{ij}, \tilde{\boldsymbol{Y}}_{-i,-j}=\tilde{\boldsymbol{y}}_{-i,-j}, \boldsymbol{Y}_Q=\boldsymbol{y}_Q) P(\boldsymbol{Y}_Q=\boldsymbol{y}_Q \mid \tilde{Y}_{ij}=\tilde{y}_{ij}, \tilde{\boldsymbol{Y}}_{-i,-j}=\tilde{\boldsymbol{y}}_{-i,-j}) \times \\ &\quad P(\tilde{Y}_{ij}=\tilde{y}_{ij}) P(\tilde{\boldsymbol{Y}}_{-i,-j}=\tilde{\boldsymbol{y}}_{-i,-j}) \\ &= P(\boldsymbol{R}=\boldsymbol{r} \mid \tilde{Y}_{ij}=\tilde{y}_{ij}, \tilde{\boldsymbol{Y}}_{-i,-j}=\tilde{\boldsymbol{y}}_{-i,-j}) P(\boldsymbol{C}=\boldsymbol{c} \mid \tilde{Y}_{ij}=\tilde{y}_{ij}, \tilde{\boldsymbol{Y}}_{-i,-j}=\tilde{\boldsymbol{y}}_{-i,-j}) \times \\ &\quad \left[\prod_{(s,t)\in Q} P(Y_{st}=y_{st} \mid \tilde{Y}_{st}=\tilde{y}_{st})\right] \tilde{p}_{ij}^{\tilde{y}_{ij}} (1-\tilde{p}_{ij})^{1-\tilde{y}_{ij}} \prod_{\substack{i'=1 \\ \{i'\neq i,}}^{I} \prod_{\substack{j'=1 \\ j'\neq j\}}}^{J} \tilde{p}_{i'j'}^{\tilde{y}_{i'j'}} (1-\tilde{p}_{i'j'})^{1-\tilde{y}_{i'j'}}, \end{aligned}$$

where we use the same conditional assumptions as in the prior appendices and

$$\prod_{\substack{i'=1 \\ \{i'\neq i,}}^{I} \prod_{\substack{j'=1 \\ j'\neq j\}}}^{J}$$

denotes the product taken over all combinations of $i' = 1,\ldots, I$ and $j' = 1, \ldots, J$ except the (i, j) combination. Then

$$\begin{aligned} &P(\tilde{Y}_{ij}=\tilde{y}_{ij}, \tilde{\boldsymbol{Y}}_{-i,-j}=\tilde{\boldsymbol{y}}_{-i,-j}, \boldsymbol{R}=\boldsymbol{r}, \boldsymbol{C}=\boldsymbol{c}, \boldsymbol{Y}_Q=\boldsymbol{y}_Q) \\ &= \left[\prod_{i=1}^{I} P(R_i=r_i \mid \tilde{Y}_{i1}=\tilde{y}_{i1},\ldots,\tilde{Y}_{iJ}=\tilde{y}_{iJ})\right]\left[\prod_{j=1}^{J} P(C_j=c_j \mid \tilde{Y}_{1j}=\tilde{y}_{1j},\ldots,\tilde{Y}_{Ij}=\tilde{y}_{Ij})\right] \times \\ &\quad \left[\prod_{(s,t)\in Q} P(Y_{st}=y_{st} \mid \tilde{Y}_{st}=\tilde{y}_{st})\right]\left[\prod_{i=1}^{I}\prod_{j=1}^{J} \tilde{p}_{ij}^{\tilde{y}_{ij}} (1-\tilde{p}_{ij})^{1-\tilde{y}_{ij}}\right] \end{aligned} \quad (2)$$

due to the independence among the I row responses and among the J column responses.

Define

$$\begin{aligned} \tau = &\left[\prod_{\substack{i'=1 \\ i'\neq i}}^{I} P(R_{i'}=r_{i'} \mid \tilde{Y}_{i'1}=\tilde{y}_{i'1},\ldots,\tilde{Y}_{i'J}=\tilde{y}_{i'J})\right]\left[\prod_{\substack{j'=1 \\ j'\neq j}}^{J} P(C_{j'}=c_{j'} \mid \tilde{Y}_{1j'}=\tilde{y}_{1j'},\ldots,\tilde{Y}_{Ij'}=\tilde{y}_{Ij'})\right] \times \\ &\left[\prod_{\substack{i'=1 \\ \{i'\neq i,}}^{I} \prod_{\substack{j'=1 \\ j'\neq j\}}}^{J} \tilde{p}_{i'j'}^{\tilde{y}_{i'j'}} (1-\tilde{p}_{i'j'})^{1-\tilde{y}_{i'j'}}\right] \end{aligned}$$

for notational simplicity. Noting that $\tilde{Y}_{ij} = 1$ is in the numerator of Equation (1) and if $(i, j) \in Q$, we find that Equation (2) becomes

$$\begin{aligned}&P(\tilde{Y}_{ij} = 1, \tilde{\boldsymbol{Y}}_{-i,-j} = \tilde{\boldsymbol{y}}_{-i,-j}, \boldsymbol{R} = \boldsymbol{r}, \boldsymbol{C} = \boldsymbol{c}, \boldsymbol{Y}_Q = \boldsymbol{y}_Q) \\ &= \tau \left[\prod_{(s,t)\in Q\setminus\{(i,j)\}} P(Y_{st} = y_{st} \mid \tilde{Y}_{st} = \tilde{y}_{st})\right] P(R_i = r_i \mid \tilde{R}_i = 1) P(C_j = c_j \mid \tilde{C}_j = 1) \times \\ &\quad P(Y_{ij} = y_{ij} \mid \tilde{Y}_{ij} = 1)\tilde{p}_{ij},\end{aligned} \tag{3}$$

where $(s,t) \in Q \setminus \{(i, j)\}$ means all indices in Q except for (i, j) and $\tilde{R}_i$ and $\tilde{C}_j$ are the true values for R_i and C_j, respectively. When $(i, j) \notin Q$, Equation (2) becomes

$$\begin{aligned}&P(\tilde{Y}_{ij} = 1, \tilde{\boldsymbol{Y}}_{-i,-j} = \tilde{\boldsymbol{y}}_{-i,-j}, \boldsymbol{R} = \boldsymbol{r}, \boldsymbol{C} = \boldsymbol{c}, \boldsymbol{Y}_Q = \boldsymbol{y}_Q) \\ &= \tau \left[\prod_{(s,t)\in Q} P(Y_{st} = y_{st} \mid \tilde{Y}_{st} = \tilde{y}_{st})\right] P(R_i = r_i \mid \tilde{R}_i = 1) P(C_j = c_j \mid \tilde{C}_j = 1)\tilde{p}_{ij}.\end{aligned} \tag{4}$$

Equations (3) and (4) help to show the contributions that the individual retests have on the probabilities. Simply, for large sensitivities and specificities, the individual retests contribute values close to 0 or 1.

Second, to find the denominator of Equation (1), note that

$$\begin{aligned}&P(\tilde{\boldsymbol{Y}}_{-i,-j} = \tilde{\boldsymbol{y}}_{-i,-j}, \boldsymbol{R} = \boldsymbol{r}, \boldsymbol{C} = \boldsymbol{c}, \boldsymbol{Y}_Q = \boldsymbol{y}_Q) \\ &= P(\tilde{Y}_{ij} = 0, \tilde{\boldsymbol{Y}}_{-i,-j} = \tilde{\boldsymbol{y}}_{-i,-j}, \boldsymbol{R} = \boldsymbol{r}, \boldsymbol{C} = \boldsymbol{c}, \boldsymbol{Y}_Q = \boldsymbol{y}_Q) + \\ &\quad P(\tilde{Y}_{ij} = 1, \tilde{\boldsymbol{Y}}_{-i,-j} = \tilde{\boldsymbol{y}}_{-i,-j}, \boldsymbol{R} = \boldsymbol{r}, \boldsymbol{C} = \boldsymbol{c}, \boldsymbol{Y}_Q = \boldsymbol{y}_Q).\end{aligned}$$

Using results from Equation (2), we can write the probability for $(i, j) \in Q$ as

$$\begin{aligned}&P(\tilde{\boldsymbol{Y}}_{-i,-j} = \tilde{\boldsymbol{y}}_{-i,-j}, \boldsymbol{R} = \boldsymbol{r}, \boldsymbol{C} = \boldsymbol{c}, \boldsymbol{Y}_Q = \boldsymbol{y}_Q) \\ &= \tau \left[\prod_{(s,t)\in Q\setminus\{(i,j)\}} P(Y_{st} = y_{st} \mid \tilde{Y}_{st} = \tilde{y}_{st})\right] \Big\{ P(R_i = r_i \mid \tilde{Y}_{i1} = \tilde{y}_{i1}, \ldots, \tilde{Y}_{ij} = 0, \ldots, \tilde{Y}_{iJ} = \tilde{y}_{iJ}) \times \\ &\qquad P(C_j = c_j \mid \tilde{Y}_{1j} = \tilde{y}_{1j}, \ldots, \tilde{Y}_{ij} = 0, \ldots, \tilde{Y}_{Ij} = \tilde{y}_{Ij}) P(Y_{ij} = y_{ij} \mid \tilde{Y}_{ij} = 0)(1 - \tilde{p}_{ij}) + \\ &\qquad P(R_i = r_i \mid \tilde{R}_i = 1) P(C_j = c_j \mid \tilde{C}_j = 1) P(Y_{ij} = y_{ij} \mid \tilde{Y}_{ij} = 1)\, \tilde{p}_{ij} \Big\}\end{aligned}$$

and for $(i, j) \notin Q$ as

$$\begin{aligned}&P(\tilde{\boldsymbol{Y}}_{-i,-j} = \tilde{\boldsymbol{y}}_{-i,-j}, \boldsymbol{R} = \boldsymbol{r}, \boldsymbol{C} = \boldsymbol{c}, \boldsymbol{Y}_Q = \boldsymbol{y}_Q) \\ &= \tau \prod_{(s,t)\in Q} P(Y_{st} = y_{st} \mid \tilde{Y}_{st} = \tilde{y}_{st}) \Big\{ P(R_i = r_i \mid \tilde{Y}_{i1} = \tilde{y}_{i1}, \ldots, \tilde{Y}_{ij} = 0, \ldots, \tilde{Y}_{iJ} = \tilde{y}_{iJ}) \times \\ &\qquad P(C_j = c_j \mid \tilde{Y}_{1j} = \tilde{y}_{1j}, \ldots, \tilde{Y}_{ij} = 0, \ldots, \tilde{Y}_{Ij} = \tilde{y}_{Ij})(1 - \tilde{p}_{ij}) + \\ &\qquad P(R_i = r_i \mid \tilde{R}_i = 1) P(C_j = c_j \mid \tilde{C}_j = 1)\tilde{p}_{ij} \Big\}.\end{aligned}$$

Combining all the results, we have for $(i, j) \in Q$:

$$\gamma_{ij} = \frac{P(R_i = r_i \mid \tilde{R}_i = 1)P(C_j = c_j \mid \tilde{C}_j = 1)P(Y_{ij} = y_{ij} \mid \tilde{Y}_{ij} = 1)\tilde{p}_{ij}}{\begin{array}{l}\{P(R_i = r_i \mid \tilde{Y}_{i1} = \tilde{y}_{i1},\ldots,\tilde{Y}_{ij} = 0,\ldots,\tilde{Y}_{iJ} = \tilde{y}_{iJ})P(C_j = c_j \mid \tilde{Y}_{1j} = \tilde{y}_{1j},\ldots,\tilde{Y}_{ij} = 0,\ldots,\tilde{Y}_{Ij} = \tilde{y}_{Ij})\times \\ \quad P(Y_{ij} = y_{ij} \mid \tilde{Y}_{ij} = 0)(1-\tilde{p}_{ij}) + P(R_i = r_i \mid \tilde{R}_i = 1)P(C_j = c_j \mid \tilde{C}_j = 1)P(Y_{ij} = y_{ij} \mid \tilde{Y}_{ij} = 1)\tilde{p}_{ij}\}\end{array}}$$

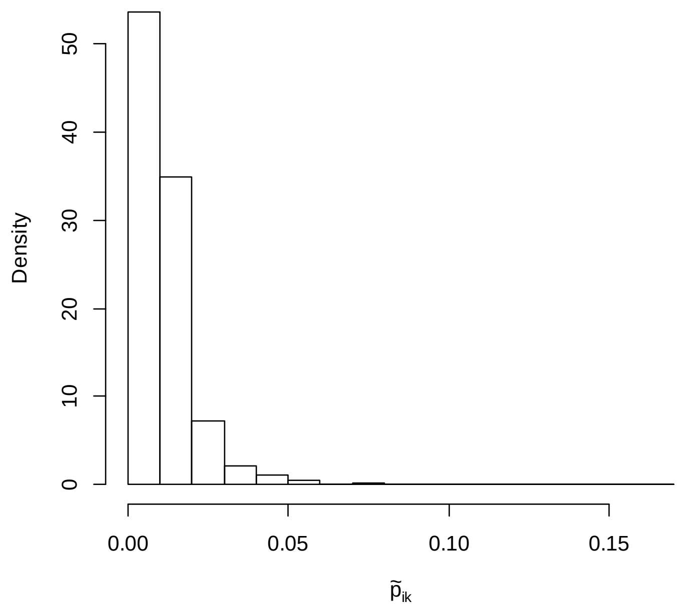
and for $(i, j) \notin Q$:

$$\gamma_{ij} = \frac{P(R_i = r_i \mid \tilde{R}_i = 1)P(C_j = c_j \mid \tilde{C}_j = 1)\tilde{p}_{ij}}{\begin{array}{l}\{P(R_i = r_i \mid \tilde{Y}_{i1} = \tilde{y}_{i1},\ldots,\tilde{Y}_{ij} = 0,\ldots,\tilde{Y}_{iJ} = \tilde{y}_{iJ})P(C_j = c_j \mid \tilde{Y}_{1j} = \tilde{y}_{1j},\ldots,\tilde{Y}_{ij} = 0,\ldots,\tilde{Y}_{Ij} = \tilde{y}_{Ij})\times \\ \quad (1-\tilde{p}_{ij}) + P(R_i = r_i \mid \tilde{R}_i = 1)P(C_j = c_j \mid \tilde{C}_j = 1)\tilde{p}_{ij}\}\end{array}}.$$

Note that for the case of no individual retests, the formula for $(i, j) \notin Q$ should be used for all i and j.

WEB APPENDIX D

A histogram of the true individual probabilities for one simulated data set in Section 3.1.



WEB APPENDIX E

Average number of tests performed by each protocol for 500 simulated data sets, each containing 5000 individuals, with $\eta = \delta = 0.99$ in Section 3.2.

Group Size	IG	Dorfman	Halving	Array w/o retesting	Array w/ retesting
4	1250	1522	1500	2502	2652
6	834	1214	1129	1669	1812
8	625	1111	968	1254	1398
10	500	1087	891	1000	1144
12	417	1107	857	837	993
14	358	1149	848	720	897
16	313	1196	844	633	831
18	278	1251	854	564	790
20	250	1321	875	510	771

WEB APPENDIX F

This appendix discusses additional simulations used to reinforce the findings in Section 3 of the paper. We simulate data for each testing protocol according to the model $\text{logit}(\tilde{p}_{ik}) = \beta_0 + \beta_1 x_{ik}$ ($\text{logit}(\tilde{p}_{ij}) = \beta_0 + \beta_1 x_{ij}$ for the array testing protocols), where β_0 = -6 and β_1 = 4.7. The covariates are generated from a Uniform(0, 1) distribution. These configurations provide an overall mean prevalence of about 0.05. The sensitivity and specificity are set to be $\eta = \delta = 0.99$. The range of the group sizes included in this study is reasonable given the prevalence level.

Figures 1-3 give the results for 1000 individuals and Figures 4-6 give the results for 5000 individuals. Overall, we see that the results from the paper continue to hold true here. Note that ψ of IG begins to increase with the group sizes in Figures 3 and 6, which it did not for the simulations in the paper. This occurs due to the larger overall prevalence that leads to some of the larger group sizes not being ideal for IG.

Note that Figure 7 provides a histogram of the true individual probabilities for one simulated data set of 5000 individuals under the simulation settings.

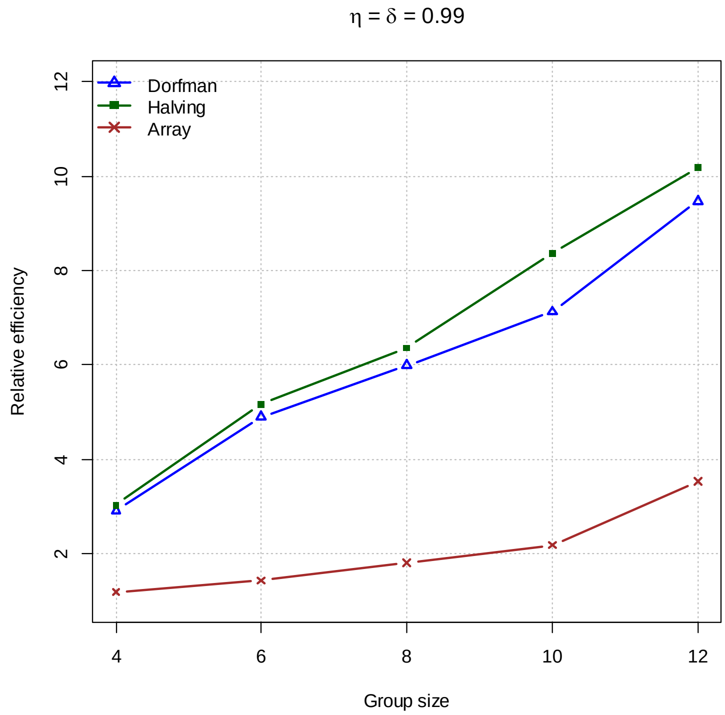


Figure 1. Estimated relative efficiencies calculated by Equation (4) in Section 3.2 based on 500 simulated data sets and 1000 individuals. Dorfman and halving are compared to IG. Array testing is compared with and without retests.

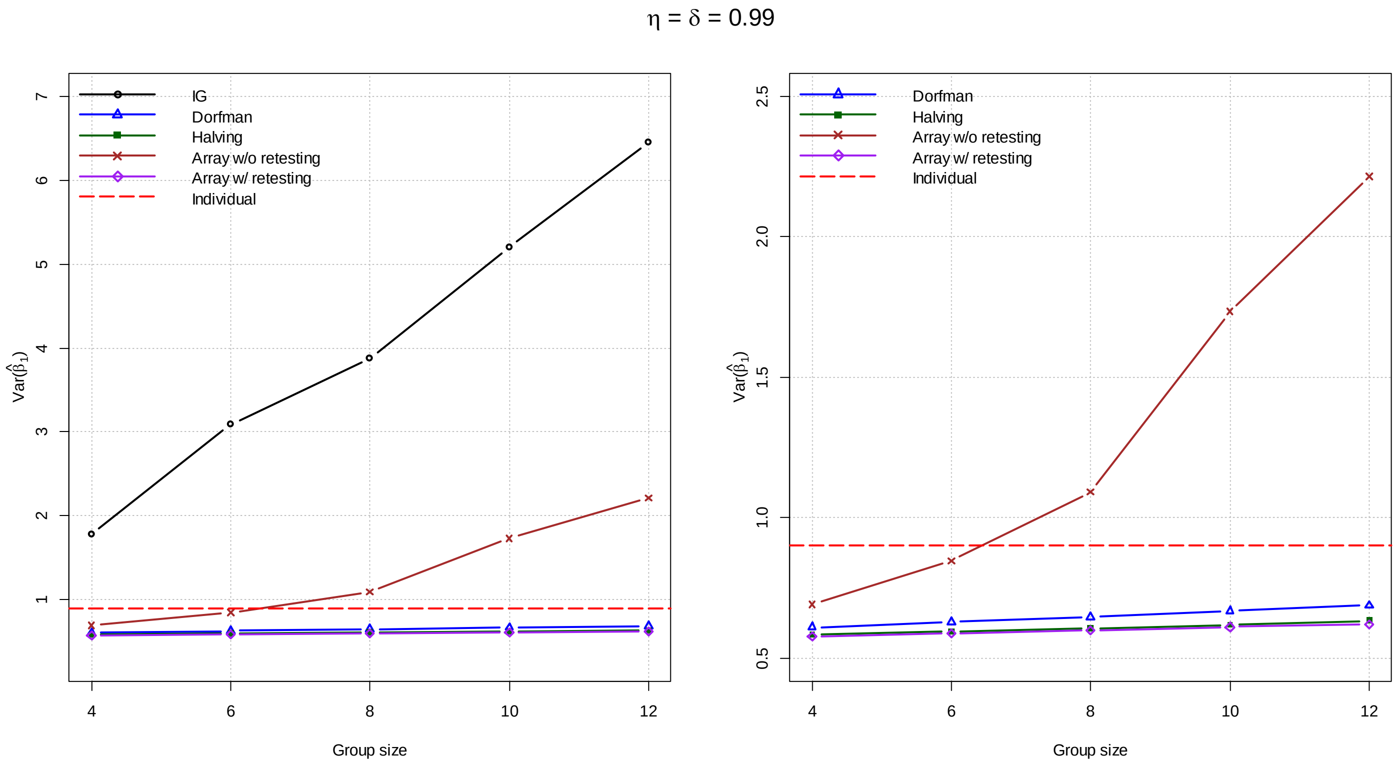


Figure 2. Averaged $Var(\hat{\beta}_1)$ for 500 simulated data sets and 1000 individuals. The horizontal dashed line corresponds to $Var(\hat{\beta}_1)$ from individual testing. The right-side plot is the same as on the left-side except we omit IG in order to reduce the y-axis scale.

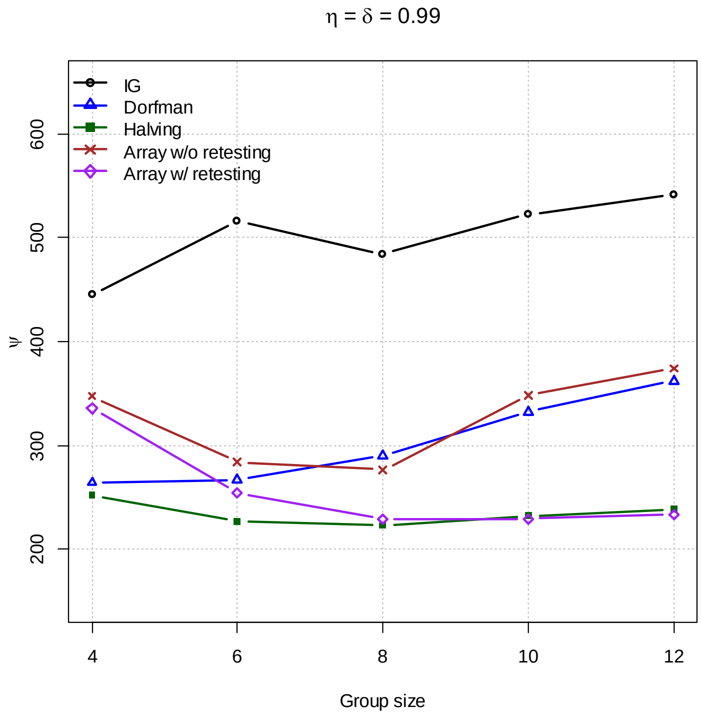


Figure 3. Average number of tests per unit of information calculated by Equation (5) in Section 3.3 based on 500 simulated data sets and 1000 individuals. Note that $\psi = 901$ for individual testing.

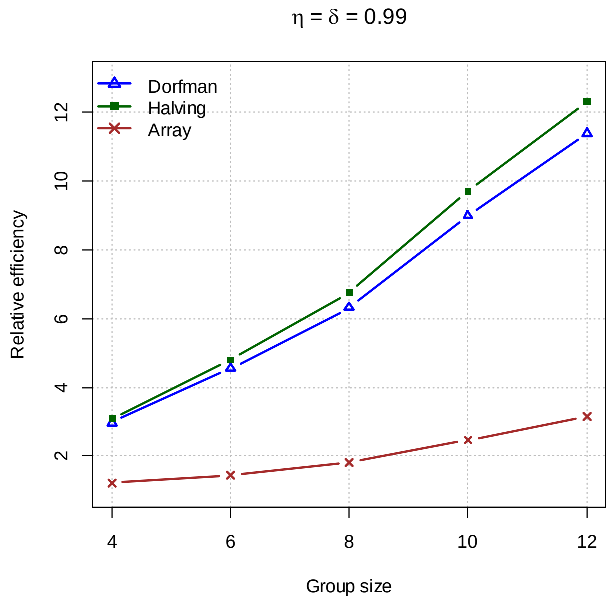


Figure 4. Estimated relative efficiencies calculated by Equation (4) in Section 3.2 based on 500 simulated data sets and 5000 individuals. Dorfman and halving are compared to IG. Array testing is compared with and without retests.

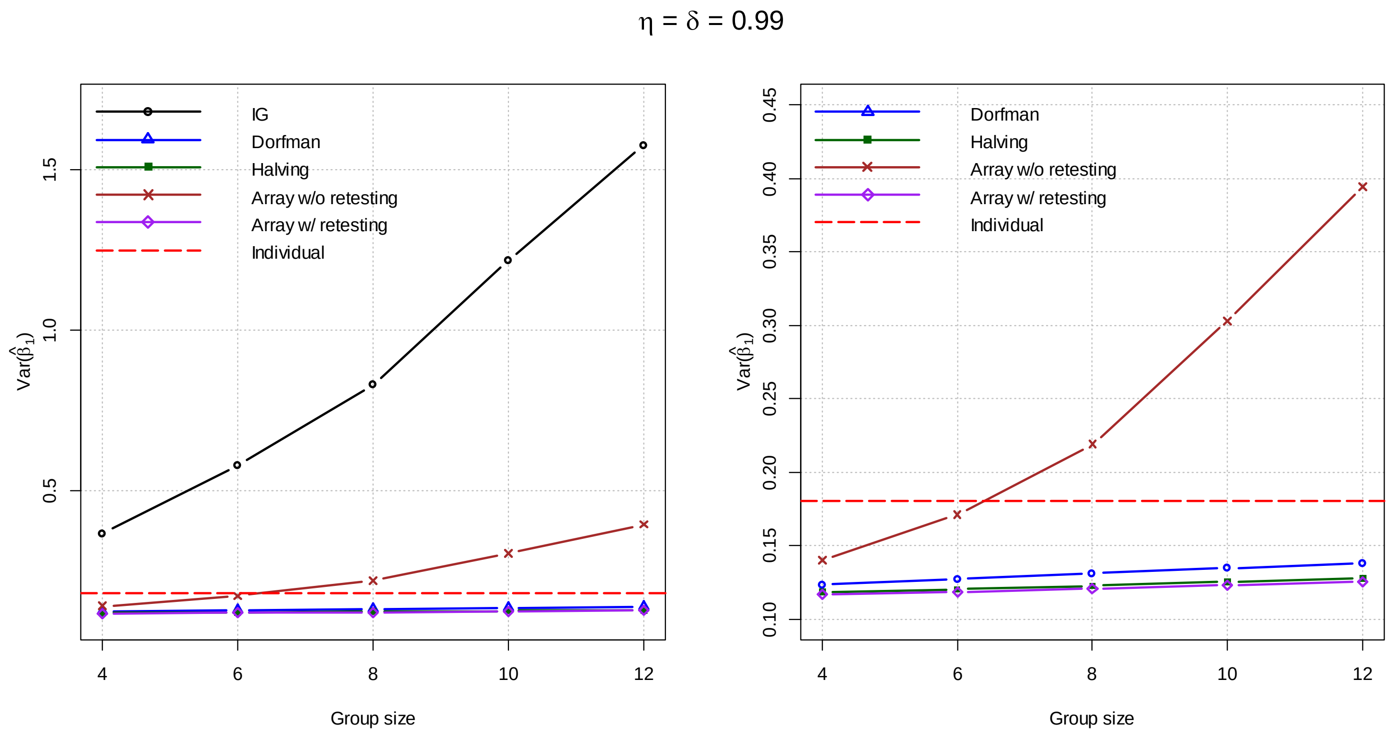


Figure 5. Averaged $Var(\hat{\beta}_1)$ for 500 simulated data sets and 5000 individuals. The horizontal dashed line corresponds to $Var(\hat{\beta}_1)$ from individual testing. The right-side plot is the same as on the left-side except we omit IG in order to reduce the y-axis scale.

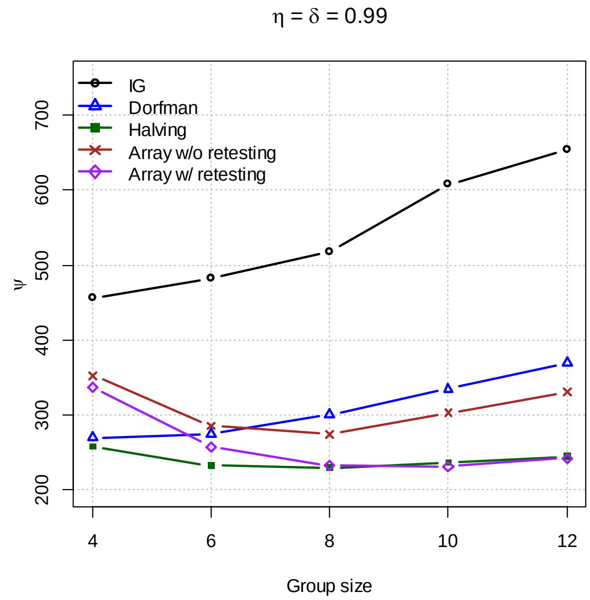


Figure 6. Average number of tests per unit of information calculated by Equation (5) in Section 3.3 based on 500 simulated data sets and 5000 individuals. Note that $\psi = 903$ for individual testing.

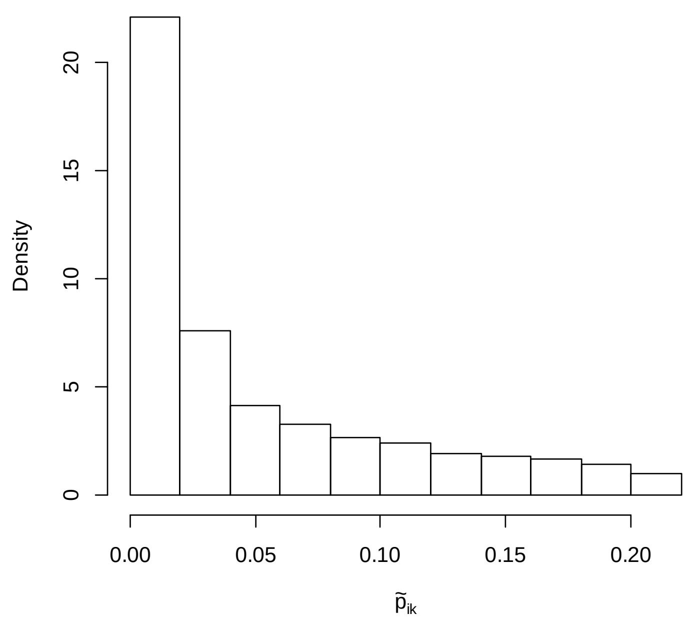


Figure 7. A histogram of the true individual probabilities for one simulated data set of 5000 individuals.