Web-based Supplementary Materials for "Two-Dimensional Informative Array Testing"

Christopher S. McMahan,[1] Joshua M. Tebbs,[1,*] and Christopher R. Bilder[2]

[1]Department of Statistics, University of South Carolina, Columbia, SC 29208, U.S.A.

[2]Department of Statistics, University of Nebraska-Lincoln, Lincoln, NE 68583, U.S.A.

[*]*email*: tebbs@stat.sc.edu

Web Appendix A: *Derivation of* $E(T)$. Let T_{jk} denote the number of tests required to classify individual $\mathcal{I}_{jk}$ after row and column testing has been completed. Using the classification methodology in Kim et al. (2007),

$$T_{jk} = \begin{cases} 1, & \text{if } R_j = 1 \text{ and } C_k = 1 \\ 1, & \text{if } R_j = 1 \text{ and } \sum_{k'=1}^{K} C_{k'} = 0 \\ 1, & \text{if } \sum_{j'=1}^{J} R_{j'} = 0 \text{ and } C_k = 1 \\ 0, & \text{otherwise.} \end{cases}$$

Subsequently, for a two-dimensional array testing algorithm,

$$\begin{aligned} E(T_{jk}) &= \text{pr}\left(R_j = 1, C_k = 1\right) + \text{pr}\left(R_j = 1, \sum_{k'=1}^{K} C_{k'} = 0\right) \\ &\qquad + \text{pr}\left(\sum_{j'=1}^{J} R_{j'} = 0, C_k = 1\right). \end{aligned} \tag{A.1}$$

Under the assumptions stated in Section 2, the first probability in (A.1), by the Law of Total Probability, is

$$\begin{aligned} \text{pr}(R_j = 1, C_k = 1) &= \sum_{\widetilde{r}=0}^{1} \sum_{\widetilde{c}=0}^{1} \text{pr}(R_j = 1, C_k = 1 | \widetilde{R}_j = \widetilde{r}, \widetilde{C}_k = \widetilde{c}) \text{pr}(\widetilde{R}_j = \widetilde{r}, \widetilde{C}_k = \widetilde{c}) \\ &= S_e^2 \text{pr}(\widetilde{R}_j = 1, \widetilde{C}_k = 1) & \text{(A.2)} \\ &\quad + S_e(1 - S_p) \text{pr}(\widetilde{R}_j = 0, \widetilde{C}_k = 1) & \text{(A.3)} \\ &\quad + S_e(1 - S_p) \text{pr}(\widetilde{R}_j = 1, \widetilde{C}_k = 0) & \text{(A.4)} \\ &\quad + (1 - S_p)^2 \text{pr}(\widetilde{R}_j = 0, \widetilde{C}_k = 0). & \text{(A.5)} \end{aligned}$$

The jth row and kth column share only individual $\mathcal{I}_{jk}$. Under the assumption that the individual statuses are independent, $\widetilde{R}_j$ and $\widetilde{C}_k$ are independent, conditional on $\widetilde{g}_{jk}$. Thus, the probability expression in (A.2) is

$$\begin{aligned}\mathrm{pr}(\widetilde{R}_j = 1, \widetilde{C}_k = 1) &= \mathrm{pr}(\widetilde{R}_j = 1 | \widetilde{g}_{jk} = 1)\mathrm{pr}(\widetilde{C}_k = 1 | \widetilde{g}_{jk} = 1)\mathrm{pr}(\widetilde{g}_{jk} = 1) \\ &\quad + \mathrm{pr}(\widetilde{R}_j = 1 | \widetilde{g}_{jk} = 0)\mathrm{pr}(\widetilde{C}_k = 1 | \widetilde{g}_{jk} = 0)\mathrm{pr}(\widetilde{g}_{jk} = 0).\end{aligned}$$

Clearly, $\mathrm{pr}(\widetilde{C}_k = 1|\widetilde{g}_{jk} = 1) = \mathrm{pr}(\widetilde{R}_j = 1|\widetilde{g}_{jk} = 1) = 1$. Recall that $\widetilde{R}_j = 0$ ($\widetilde{C}_k = 0$) if all individuals in the jth row (kth column) are truly negative. With $\pi_R(j) = \mathrm{pr}(\widetilde{R}_j = 0) = \prod_{k'=1}^{K}(1 - p_{jk'})$ and $\pi_C(k) = \mathrm{pr}(\widetilde{C}_k = 0) = \prod_{j'=1}^{J}(1 - p_{j'k})$, we have

$$\mathrm{pr}(\widetilde{R}_j = 0|\widetilde{g}_{jk} = 0) = \frac{\mathrm{pr}(\widetilde{R}_j = 0, \widetilde{g}_{jk} = 0)}{\mathrm{pr}(\widetilde{g}_{jk} = 0)} = \frac{\mathrm{pr}(\widetilde{R}_j = 0)}{\mathrm{pr}(\widetilde{g}_{jk} = 0)} = \frac{\pi_R(j)}{1 - p_{jk}}$$

and

$$\mathrm{pr}(\widetilde{C}_k = 0|\widetilde{g}_{jk} = 0) = \frac{\mathrm{pr}(\widetilde{C}_k = 0, \widetilde{g}_{jk} = 0)}{\mathrm{pr}(\widetilde{g}_{jk} = 0)} = \frac{\mathrm{pr}(\widetilde{C}_k = 0)}{\mathrm{pr}(\widetilde{g}_{jk} = 0)} = \frac{\pi_C(k)}{1 - p_{jk}}.$$

Subsequently,

$$\begin{aligned}\mathrm{pr}(\widetilde{R}_j = 1, \widetilde{C}_k = 1) &= p_{jk} + \left\{1 - \frac{\pi_C(k)}{1 - p_{jk}}\right\}\left\{1 - \frac{\pi_R(j)}{1 - p_{jk}}\right\}(1 - p_{jk}) \\ &= 1 - \pi_C(k) - \pi_R(j) + \frac{\pi_C(k)\pi_R(j)}{1 - p_{jk}}.\end{aligned}$$

Expressions for the three probabilities in (A.3), (A.4), and (A.5) are found similarly and are therefore given without derivation:

$$\begin{aligned}\mathrm{pr}(\widetilde{R}_j = 0, \widetilde{C}_k = 1) &= \left\{1 - \frac{\pi_C(k)}{1 - p_{jk}}\right\}\pi_R(j) \\ \mathrm{pr}(\widetilde{R}_j = 1, \widetilde{C}_k = 0) &= \left\{1 - \frac{\pi_R(j)}{1 - p_{jk}}\right\}\pi_C(k) \\ \mathrm{pr}(\widetilde{R}_j = 0, \widetilde{C}_k = 0) &= \frac{\pi_C(k)\pi_R(j)}{1 - p_{jk}}.\end{aligned}$$

After extensive algebra, the first probability in (A.1) reduces to

$$\mathrm{pr}(R_j = 1, C_k = 1) = S_e^2 + (1 - S_e - S_p)^2\left\{\frac{\pi_C(k)\pi_R(j)}{1 - p_{jk}}\right\} + \{S_e(1 - S_p) - S_e^2\}\{\pi_C(k) + \pi_R(j)\}.$$

We now turn our attention to the second probability in (A.1). By the Law of Total Probability,

$$\begin{aligned}\mathrm{pr}\left(R_j = 1, \sum_{k'=1}^{K} C_{k'} = 0\right) &= \sum_{\widetilde{r}=0}^{1}\sum_{\widetilde{c}_1=0}^{1}\cdots\sum_{\widetilde{c}_K=0}^{1}\mathrm{pr}\left(R_j = 1, \bigcap_{k=1}^{K}\{C_k = 0\}\middle|\widetilde{R}_j = \widetilde{r}, \bigcap_{k=1}^{K}\{\widetilde{C}_k = \widetilde{c}_k\}\right) \\ &\quad \times \mathrm{pr}\left(\widetilde{R}_j = \widetilde{r}, \bigcap_{k=1}^{K}\{\widetilde{C}_k = \widetilde{c}_k\}\right).\end{aligned}$$

Define $\mathcal{B}_c$, for $c = 1, 2, ..., K$, to be the set of all c-combinations of $\mathcal{K}_0 = \{1, 2, ..., K\}$ and let $\mathcal{B}_0 = \emptyset$. For all $\mathcal{B} \in \mathcal{B}_c$, $c = 0, 1, ..., K$, define the events

$$\begin{aligned} \widetilde{C}(\mathcal{B}) &= \bigcap_{k=1}^{K} \{\widetilde{C}_k = I(k \in \mathcal{B})\} \\ C(\mathcal{B}) &= \bigcap_{k=1}^{K} \{C_k = I(k \in \mathcal{B})\}, \end{aligned}$$

where $I(\cdot)$ denotes the usual indicator function. Using this notation, the previous probability can be written as

$$\text{pr}\left(R_j = 1, \sum_{k'=1}^{K} C_{k'} = 0\right) = \sum_{\widetilde{r}=0}^{1}\sum_{c=0}^{K}\sum_{\mathcal{B}\in\mathcal{B}_c} \text{pr}\{R_j = 1, C(\mathcal{B}_0)|\widetilde{R}_j = \widetilde{r}, \widetilde{C}(\mathcal{B})\}\text{pr}\{\widetilde{R}_j = \widetilde{r}, \widetilde{C}(\mathcal{B})\}. \tag{A.6}$$

Using the assumptions in Section 2 regarding the test sensitivity and specificity, for all $c \in \{0, 1, ..., K\}$, we have

$$\begin{aligned} \text{pr}\{R_j = 1, C(\mathcal{B}_0)|\widetilde{R}_j = 0, \widetilde{C}(\mathcal{B})\} &= (1 - S_p)(1 - S_e)^c S_p^{K-c} \\ \text{pr}\{R_j = 1, C(\mathcal{B}_0)|\widetilde{R}_j = 1, \widetilde{C}(\mathcal{B})\} &= S_e(1 - S_e)^c S_p^{K-c}. \end{aligned}$$

Changing the order of summation, (A.6) becomes

$$\begin{aligned} \text{pr}\left(R_j = 1, \sum_{k'=1}^{K} C_{k'} = 0\right) &= \sum_{c=0}^{K}\sum_{\mathcal{B}\in\mathcal{B}_c} \Big[(1 - S_p)(1 - S_e)^c S_p^{K-c}\text{pr}\{\widetilde{R}_j = 0, \widetilde{C}(\mathcal{B})\} \\ &\quad + S_e(1 - S_e)^c S_p^{K-c}\text{pr}\{\widetilde{R}_j = 1, \widetilde{C}(\mathcal{B})\}\Big]. \end{aligned} \tag{A.7}$$

We now derive expressions for the two probabilities on the right-hand side of (A.7). Note that

$$\begin{aligned} \text{pr}\{\widetilde{R}_j = 0, \widetilde{C}(\mathcal{B})\} &= \text{pr}\{\widetilde{C}(\mathcal{B})|\widetilde{R}_j = 0\}\text{pr}(\widetilde{R}_j = 0) \\ &= \text{pr}\{\widetilde{C}(\mathcal{B})|\widetilde{R}_j = 0\}\pi_R(j), \end{aligned}$$

where $\pi_R(j) = \text{pr}(\widetilde{R}_j = 0) = \prod_{k'=1}^{K}(1 - p_{jk'})$. Because $\{\widetilde{R}_j = 0\} = \{\widetilde{g}_{jk} = 0,\ k = 1, 2, ..., K\}$,

$$\begin{aligned} \text{pr}\{\widetilde{C}(\mathcal{B})|\widetilde{R}_j = 0\} &= \text{pr}\{\widetilde{C}(\mathcal{B})|\widetilde{g}_{jk} = 0,\ k = 1, 2, ..., K\} \\ &= \prod_{k=1}^{K} \text{pr}\{\widetilde{C}_k = I(k \in \mathcal{B})|\widetilde{g}_{jk} = 0\}, \end{aligned}$$

as the individual statuses $\widetilde{g}_{jk}$ are independent. Simple conditioning shows that

$$\begin{aligned} \mathrm{pr}(\widetilde{C}_k = 0|\widetilde{g}_{jk} = 0) &= \frac{\pi_C(k)}{1 - p_{jk}} \\ \mathrm{pr}(\widetilde{C}_k = 1|\widetilde{g}_{jk} = 0) &= \frac{1 - \pi_C(k)}{1 - p_{jk}}, \end{aligned}$$

where $\pi_C(k) = \mathrm{pr}(\widetilde{C}_k = 0) = \prod_{j'=1}^{J}(1 - p_{j'k})$. Therefore, with $\overline{\mathcal{B}} = \mathcal{K}_0 \backslash \mathcal{B}$,

$$\mathrm{pr}\{\widetilde{C}(\mathcal{B})|\widetilde{R}_j = 0\} = \prod_{k' \in \mathcal{B}} \left\{1 - \frac{\pi_C(k')}{1 - p_{jk'}}\right\} \prod_{k' \in \overline{\mathcal{B}}} \frac{\pi_C(k')}{1 - p_{jk'}},$$

where it is understood that products taken over $\{k' \in \emptyset\}$ are equal to 1. To find the second probability in (A.7), we first write

$$\mathrm{pr}\{\widetilde{R}_j = 1, \widetilde{C}(\mathcal{B})\} = \mathrm{pr}\{\widetilde{C}(\mathcal{B})\} - \mathrm{pr}\{\widetilde{R}_j = 0, \widetilde{C}(\mathcal{B})\},$$

where

$$\mathrm{pr}\{\widetilde{C}(\mathcal{B})\} = \mathrm{pr}\left\{\bigcap_{k=1}^{K}\{\widetilde{C}_k = I(k \in \mathcal{B})\}\right\} = \prod_{k' \in \mathcal{B}} \{1 - \pi_C(k')\} \prod_{k' \in \overline{\mathcal{B}}} \pi_C(k')$$

and

$$\mathrm{pr}\{\widetilde{R}_j = 0, \widetilde{C}(\mathcal{B})\} = \mathrm{pr}\{\widetilde{C}(\mathcal{B})|\widetilde{R}_j = 0\}\mathrm{pr}(\widetilde{R}_j = 0).$$

Now, to simplify our notation, we define the set function

$$\lambda_{\mathcal{C}}(\mathcal{B}|\mathcal{S}, j) = \prod_{k' \in \mathcal{B}} \left\{1 - \frac{\pi_C(k')}{(1 - p_{jk'})^{I(k' \in \mathcal{S})}}\right\} \prod_{k' \in \overline{\mathcal{B}}} \frac{\pi_C(k')}{(1 - p_{jk'})^{I(k' \in \mathcal{S})}}.$$

It follows directly that $\mathrm{pr}\{\widetilde{C}(\mathcal{B})|\widetilde{R}_j = 0\} = \lambda_{\mathcal{C}}(\mathcal{B}|\mathcal{K}_0, j)$ and $\mathrm{pr}\{\widetilde{C}(\mathcal{B})\} = \lambda_{\mathcal{C}}(\mathcal{B}|\emptyset, j)$. Therefore, (A.7) becomes

$$\begin{aligned} \mathrm{pr}\left(R_j = 1, \sum_{k'=1}^{K} C_{k'} = 0\right) &= \sum_{c=0}^{K} \sum_{\mathcal{B} \in \mathcal{B}_c} \left[(1 - S_p)(1 - S_e)^c S_p^{K-c} \pi_R(j) \lambda_{\mathcal{C}}(\mathcal{B}|\mathcal{K}_0, j)\right. \\ &\qquad \left. + S_e(1 - S_e)^c S_p^{K-c} \{\lambda_{\mathcal{C}}(\mathcal{B}|\emptyset, j) - \pi_R(j)\lambda_{\mathcal{C}}(\mathcal{B}|\mathcal{K}_0, j)\}\right] \\ &= \sum_{c=0}^{K} \sum_{\mathcal{B} \in \mathcal{B}_c} \{\gamma_0(c, K)\lambda_{\mathcal{C}}(\mathcal{B}|\emptyset, j) + \gamma_1(c, K)\pi_R(j)\lambda_{\mathcal{C}}(\mathcal{B}|\mathcal{K}_0, j)\}, \end{aligned}$$

where $\gamma_0(c, K) = S_e(1 - S_e)^c S_p^{K-c}$ and $\gamma_1(c, K) = (1 - S_e - S_p)(1 - S_e)^c S_p^{K-c}$. This is our closed-form expression for the second probability in (A.1). The third probability in (A.1) is found in the exact same way as the second probability. Therefore, we give its formula without derivation. Define $\mathcal{A}_r$, for $r = 1, 2, ..., J$, to be the set of all r-combinations of $\mathcal{J}_0 = \{1, 2, ..., J\}$ and let $\mathcal{A}_0 = \emptyset$. Define

$$\lambda_{\mathcal{R}}(\mathcal{A}|\mathcal{S}, k) = \prod_{j' \in \mathcal{A}} \left\{ 1 - \frac{\pi_R(j')}{(1 - p_{j'k})^{I(j' \in \mathcal{S})}} \right\} \prod_{j' \in \overline{\mathcal{A}}} \frac{\pi_R(j')}{(1 - p_{j'k})^{I(j' \in \mathcal{S})}},$$

where $\overline{\mathcal{A}} = \mathcal{J}_0 \backslash \mathcal{A}$. With this definition, the third probability (A.1) equals

$$\text{pr}\left(\sum_{j'=1}^{J} R_{j'} = 0, C_k = 1 \right) = \sum_{r=0}^{J} \sum_{\mathcal{A} \in \mathcal{A}_r} \left\{ \gamma_0(r, J) \lambda_{\mathcal{R}}(\mathcal{A}|\emptyset, k) + \gamma_1(r, J) \pi_C(k) \lambda_{\mathcal{R}}(\mathcal{A}|\mathcal{J}_0, k) \right\}.$$

This completes the derivation of $E(T_{jk})$. Finally, the efficiency is

$$E(T) = J + K + \sum_{j=1}^{J} \sum_{k=1}^{K} E(T_{jk}).$$

Web Appendix B: *Derivation of $PS_e^{\mathcal{I}_{jk}}$ and $PS_p^{\mathcal{I}_{jk}}$.* We first present the derivation of the pooling sensitivity, $PS_e^{\mathcal{I}_{jk}}$. Let $g_{jk} = 1$, if $\mathcal{I}_{jk}$ tests positive (if tested individually) and let $g_{jk} = 0$, otherwise. By definition, the pooling sensitivity for $\mathcal{I}_{jk}$ is

$$\begin{aligned} PS_e^{\mathcal{I}_{jk}} = \mathrm{pr}(\mathcal{I}_{jk}^{+}|\widetilde{g}_{jk} = 1) &\equiv \mathrm{pr}(\mathcal{I}_{jk} \text{ classified positive}|\widetilde{g}_{jk} = 1) \\ &= \mathrm{pr}(g_{jk} = 1, R_j = 1, C_k = 1|\widetilde{g}_{jk} = 1) \qquad \text{(B.1)} \\ &\quad + \mathrm{pr}\left(g_{jk} = 1, R_j = 1, \sum_{k'=1}^{K} C_{k'} = 0 \middle| \widetilde{g}_{jk} = 1\right) \qquad \text{(B.2)} \\ &\quad + \mathrm{pr}\left(g_{jk} = 1, \sum_{j'=1}^{J} R_{j'} = 0, C_k = 1 \middle| \widetilde{g}_{jk} = 1\right). \qquad \text{(B.3)} \end{aligned}$$

If $\widetilde{g}_{jk} = 1$, then $\widetilde{R}_j = 1$ and $\widetilde{C}_k = 1$. This fact, together with the conditional independence assumption, implies that (B.1) equals

$$\mathrm{pr}(g_{jk} = 1|\widetilde{g}_{jk} = 1)\mathrm{pr}(R_j = 1|\widetilde{g}_{jk} = 1)\mathrm{pr}(C_k = 1|\widetilde{g}_{jk} = 1) = S_e^3.$$

Similarly, (B.2) can be written as

$$\mathrm{pr}(g_{jk} = 1|\widetilde{g}_{jk} = 1)\mathrm{pr}(R_j = 1|\widetilde{g}_{jk} = 1)\mathrm{pr}\left(\sum_{k'=1}^{K} C_{k'} = 0 \middle| \widetilde{g}_{jk} = 1\right).$$

Because S_e does not depend on the pool size, $\mathrm{pr}(g_{jk} = 1|\widetilde{g}_{jk} = 1) = \mathrm{pr}(R_j = 1|\widetilde{g}_{jk} = 1) = S_e$. For all $k \neq k'$, both (a) C_k is independent of $C_{k'}$ and (b) $C_{k'}$ is independent of g_{jk}. Thus,

$$\mathrm{pr}\left(\sum_{k'=1}^{K} C_{k'} = 0 \middle| \widetilde{g}_{jk} = 1\right) = (1 - S_e) \prod_{k' \neq k} \mathrm{pr}(C_{k'} = 0),$$

where $\mathrm{pr}(C_{k'} = 0) = 1 - S_e - (1 - S_e - S_p)\pi_C(k')$ and $\pi_C(k) = \mathrm{pr}(\widetilde{C}_k = 0) = \prod_{j'=1}^{J}(1 - p_{j'k})$. Therefore, (B.2) equals

$$\mathrm{pr}\left(g_{jk} = 1, R_j = 1, \sum_{k'=1}^{K} C_{k'} = 0 \middle| \widetilde{g}_{jk} = 1\right) = S_e^2(1 - S_e) \prod_{k' \neq k} \mathrm{pr}(C_{k'} = 0).$$

By a completely analogous argument, (B.3) equals

$$\mathrm{pr}\left(g_{jk} = 1, \sum_{j'=1}^{J} R_{j'} = 0, C_k = 1 \middle| \widetilde{g}_{jk} = 1\right) = S_e^2(1 - S_e) \prod_{j' \neq j} \mathrm{pr}(R_{j'} = 0),$$

where $\mathrm{pr}(R_{j'} = 0) = 1 - S_e - (1 - S_e - S_p)\pi_R(j')$ and $\pi_R(j) = \mathrm{pr}(\widetilde{R}_j = 0) = \prod_{k'=1}^{K}(1 - p_{jk'})$. Combining these results, we obtain

$$PS_e^{\mathcal{I}_{jk}} = S_e^3 + S_e^2(1 - S_e)\left\{\prod_{k' \neq k}\mathrm{pr}(C_{k'} = 0) + \prod_{j' \neq j}\mathrm{pr}(R_{j'} = 0)\right\}.$$

This completes the derivation of $PS_e^{\mathcal{I}_{jk}}$. We now present the derivation of the pooling specificity, $PS_p^{\mathcal{I}_{jk}}$. By definition,

$$\begin{aligned} PS_p^{\mathcal{I}_{jk}} = \mathrm{pr}(\mathcal{I}_{jk}^-|\widetilde{g}_{jk} = 0) \quad &\equiv \quad \mathrm{pr}(\mathcal{I}_{jk} \text{ classified negative}|\widetilde{g}_{jk} = 0) \\ &= \quad 1 - \mathrm{pr}(\mathcal{I}_{jk} \text{ classified positive}|\widetilde{g}_{jk} = 0) \\ &= \quad 1 - \mathrm{pr}(\mathcal{I}_{jk}^+|\widetilde{g}_{jk} = 0), \end{aligned}$$

and the probability

$$\begin{aligned} \mathrm{pr}(\mathcal{I}_{jk}^+|\widetilde{g}_{jk} = 0) \quad = \quad & \mathrm{pr}(g_{jk} = 1, R_j = 1, C_k = 1|\widetilde{g}_{jk} = 0) && \text{(B.4)} \\ & + \mathrm{pr}\left(g_{jk} = 1, R_j = 1, \sum_{k'=1}^{K} C_{k'} = 0 \,\middle|\, \widetilde{g}_{jk} = 0\right) && \text{(B.5)} \\ & + \mathrm{pr}\left(g_{jk} = 1, \sum_{j'=1}^{J} R_{j'} = 0, C_k = 1 \,\middle|\, \widetilde{g}_{jk} = 0\right). && \text{(B.6)} \end{aligned}$$

Using the conditional independence assumption, (B.4) is equal to

$$\mathrm{pr}(g_{jk} = 1, R_j = 1, C_k = 1|\widetilde{g}_{jk} = 0) = \mathrm{pr}(g_{jk} = 1|\widetilde{g}_{jk} = 0)\mathrm{pr}(R_j = 1, C_k = 1|\widetilde{g}_{jk} = 0).$$

Clearly, $\mathrm{pr}(g_{jk} = 1|\widetilde{g}_{jk} = 0) = 1 - S_p$. Using the Law of Total probability, the second probability

$$\begin{aligned} \mathrm{pr}(R_j = 1, C_k = 1|\widetilde{g}_{jk} = 0) \quad = \quad & \sum_{r=0}^{1}\sum_{c=0}^{1}\Big\{\mathrm{pr}(\widetilde{R}_j = r, \widetilde{C}_k = c|\widetilde{g}_{jk} = 0) \\ & \qquad \times \mathrm{pr}(R_j = 1, C_k = 1|\widetilde{R}_j = r, \widetilde{C}_k = c, \widetilde{g}_{jk} = 0)\Big\} \\ = \quad & (1 - S_p)^2\mathrm{pr}(\widetilde{R}_j = 0|\widetilde{g}_{jk} = 0)\mathrm{pr}(\widetilde{C}_k = 0|\widetilde{g}_{jk} = 0) \\ & + S_e(1 - S_p)\mathrm{pr}(\widetilde{R}_j = 0|\widetilde{g}_{jk} = 0)\mathrm{pr}(\widetilde{C}_k = 1|\widetilde{g}_{jk} = 0) \\ & + S_e(1 - S_p)\mathrm{pr}(\widetilde{R}_j = 1|\widetilde{g}_{jk} = 0)\mathrm{pr}(\widetilde{C}_k = 0|\widetilde{g}_{jk} = 0) \\ & + S_e^2\mathrm{pr}(\widetilde{R}_j = 1|\widetilde{g}_{jk} = 0)\mathrm{pr}(\widetilde{C}_k = 1|\widetilde{g}_{jk} = 0). \end{aligned}$$

Simple conditioning shows that

$$\begin{aligned}\mathrm{pr}(\widetilde{R}_j = 0|\widetilde{g}_{jk} = 0) &= \frac{\pi_R(j)}{1-p_{jk}} \\ \mathrm{pr}(\widetilde{C}_k = 0|\widetilde{g}_{jk} = 0) &= \frac{\pi_C(k)}{1-p_{jk}}.\end{aligned}$$

After extensive algebra, (B.4) becomes

$$\begin{aligned}\mathrm{pr}(g_{jk} = 1, R_j = 1, C_k = 1|\widetilde{g}_{jk} = 0) &= (1-S_p)\left[S_e^2 + (1-S_e-S_p)^2\frac{\pi_C(k)\pi_R(j)}{(1-p_{jk})^2}\right. \\ &\qquad \left. + \{S_e(1-S_p) - S_e^2\}\frac{\pi_R(j)+\pi_C(k)}{1-p_{jk}}\right].\end{aligned}$$

Using conditional independence, we can write (B.5) as

$$\mathrm{pr}\left(g_{jk} = 1, R_j = 1, \sum_{k'=1}^{K} C_{k'} = 0 \middle| \widetilde{g}_{jk} = 0\right) = \mathrm{pr}(g_{jk} = 1|\widetilde{g}_{jk} = 0)\mathrm{pr}\left(R_j = 1, \sum_{k'=1}^{K} C_{k'} = 0 \middle| \widetilde{g}_{jk} = 0\right). \tag{B.7}$$

Conditioning on the true statuses of the rows and columns, the second probability on the right-hand side of (B.7) can be written as

$$\begin{aligned}\mathrm{pr}\left(R_j = 1, \sum_{k'=1}^{K} C_{k'} = 0 \middle| \widetilde{g}_{jk} = 0\right) &= \sum_{\widetilde{r}=0}^{1}\sum_{c=0}^{K}\sum_{\mathcal{B}\in\mathcal{B}_c} \mathrm{pr}\{R_j = 1, C(\mathcal{B}_0)|\widetilde{R}_j = \widetilde{r}, \widetilde{C}(\mathcal{B}), \widetilde{g}_{jk} = 0\} \\ &\qquad \times \mathrm{pr}\{\widetilde{R}_j = \widetilde{r}, \widetilde{C}(\mathcal{B})|\widetilde{g}_{jk} = 0\},\end{aligned}$$

where

$$\begin{aligned}\mathrm{pr}\{R_j = 1, C(\mathcal{B}_0)|\widetilde{R}_j = 0, \widetilde{C}(\mathcal{B}), \widetilde{g}_{jk} = 0\} &= (1-S_p)(1-S_e)^c S_p^{K-c} \\ \mathrm{pr}\{R_j = 1, C(\mathcal{B}_0)|\widetilde{R}_j = 1, \widetilde{C}(\mathcal{B}), \widetilde{g}_{jk} = 0\} &= S_e(1-S_e)^c S_p^{K-c}.\end{aligned}$$

This allows us to rewrite

$$\begin{aligned}\mathrm{pr}\left(R_j = 1, \sum_{k'=1}^{K} C_{k'} = 0 \middle| \widetilde{g}_{jk} = 0\right) &= \sum_{c=0}^{K}\sum_{\mathcal{B}\in\mathcal{B}_c}\left[(1-S_p)(1-S_e)^c S_p^{K-c}\mathrm{pr}\{\widetilde{R}_j = 0, \widetilde{C}(\mathcal{B})|\widetilde{g}_{jk} = 0\}\right. \\ &\qquad \left. + S_e(1-S_e)^c S_p^{K-c}\mathrm{pr}\{\widetilde{R}_j = 1, \widetilde{C}(\mathcal{B})|\widetilde{g}_{jk} = 0\}\right].\end{aligned}$$

Note that

$$\mathrm{pr}\{\widetilde{R}_j = 0, \widetilde{C}(\mathcal{B})|\widetilde{g}_{jk} = 0\} = \mathrm{pr}(\widetilde{R}_j = 0|\widetilde{g}_{jk} = 0)\mathrm{pr}\{\widetilde{C}(\mathcal{B})|\widetilde{R}_j = 0, \widetilde{g}_{jk} = 0\},$$

where

$$\begin{aligned}
\mathrm{pr}(\widetilde{R}_j = 0 | \widetilde{g}_{jk} = 0) &= \frac{\pi_R(j)}{1 - p_{jk}} \\
\mathrm{pr}\{\widetilde{C}(\mathcal{B}) | \widetilde{R}_j = 0, \widetilde{g}_{jk} = 0\} &= \underbrace{\prod_{k' \in \mathcal{B}} \left\{1 - \frac{\pi_C(k')}{1 - p_{jk'}}\right\} \prod_{k' \in \overline{\mathcal{B}}} \frac{\pi_C(k')}{1 - p_{jk'}}}_{\lambda_{\mathcal{C}}(\mathcal{B}|\mathcal{K}_0, j)}.
\end{aligned}$$

Note also that

$$\mathrm{pr}\{\widetilde{R}_j = 1, \widetilde{C}(\mathcal{B}) | \widetilde{g}_{jk} = 0\} = \mathrm{pr}\{\widetilde{C}(\mathcal{B}) | \widetilde{g}_{jk} = 0\} - \mathrm{pr}\{\widetilde{R}_j = 0, \widetilde{C}(\mathcal{B}) | \widetilde{g}_{jk} = 0\},$$

where

$$\mathrm{pr}\{\widetilde{C}(\mathcal{B}) | \widetilde{g}_{jk} = 0\} = \prod_{k' \in \mathcal{B}} \left\{1 - \frac{\pi_C(k')}{(1 - p_{jk})^{I(k'=k)}}\right\} \prod_{k' \in \overline{\mathcal{B}}} \frac{\pi_C(k')}{(1 - p_{jk})^{I(k'=k)}} \equiv \lambda_{\mathcal{C}}(\mathcal{B} | \{k\}, j).$$

Therefore, after algebraic manipulation, we can write (B.5) as

$$\begin{aligned}
\mathrm{pr}\left(g_{jk} = 1, R_j = 1, \sum_{k'=1}^{K} C_{k'} = 0 \middle| \widetilde{g}_{jk} = 0\right) &= (1 - S_p) \sum_{c=0}^{K} \sum_{\mathcal{B} \in \mathcal{B}_c} \Bigg[\gamma_0(c, K) \lambda_{\mathcal{C}}(\mathcal{B} | \{k\}, j) \\
&\qquad + \frac{\gamma_1(c, K) \pi_R(j) \lambda_{\mathcal{C}}(\mathcal{B} | \mathcal{K}_0, j)}{1 - p_{jk}}\Bigg].
\end{aligned}$$

Finding an expression for (B.6) follows analogously so we give it without derivation:

$$\begin{aligned}
\mathrm{pr}\left(g_{jk} = 1, \sum_{j'=1}^{J} R_{j'} = 0, C_k = 1 \middle| \widetilde{g}_{jk} = 0\right) &= (1 - S_p) \sum_{r=0}^{J} \sum_{\mathcal{A} \in \mathcal{A}_r} \Bigg[\gamma_0(r, J) \lambda_{\mathcal{R}}(\mathcal{A} | \{r\}, k) \\
&\qquad + \frac{\gamma_1(r, J) \pi_C(k) \lambda_{\mathcal{R}}(\mathcal{A} | \mathcal{J}_0, k)}{1 - p_{jk}}\Bigg].
\end{aligned}$$

Combining all three terms, we have

$$\begin{aligned}
\frac{1 - PS_p^{\mathcal{I}_{jk}}}{1 - S_p} &= S_e^2 + (1 - S_e - S_p)^2 \frac{\pi_C(k) \pi_R(j)}{(1 - p_{jk})^2} + \left\{S_e(1 - S_p) - S_e^2\right\} \frac{\pi_R(j) + \pi_C(k)}{1 - p_{jk}} \\
&\quad + \sum_{c=0}^{K} \sum_{\mathcal{B} \in \mathcal{B}_c} \left[\gamma_0(c, K) \lambda_{\mathcal{C}}(\mathcal{B} | \{k\}, j) + \frac{\gamma_1(c, K) \pi_R(j) \lambda_{\mathcal{C}}(\mathcal{B} | \mathcal{K}_0, j)}{1 - p_{jk}}\right] \\
&\quad + \sum_{r=0}^{J} \sum_{\mathcal{A} \in \mathcal{A}_r} \left[\gamma_0(r, J) \lambda_{\mathcal{R}}(\mathcal{A} | \{j\}, k) + \frac{\gamma_1(r, J) \pi_C(k) \lambda_{\mathcal{R}}(\mathcal{A} | \mathcal{J}_0, k)}{1 - p_{jk}}\right].
\end{aligned}$$

This completes the derivation of $PS_p^{\mathcal{I}_{jk}}$.

Web Appendix C: *Supplemental material for Section* 4.1. We first prove the following proposition, stated in Section 4.1.

PROPOSITION: Suppose that $X^{\alpha,\beta} \sim \text{beta}(\alpha, \beta)$, where $\alpha > 0$ and $\beta = \alpha(1-p)/p$, for $0 < p < 1$. Then, $X^{\alpha,\beta} \xrightarrow{d} X$, as $\alpha \to 0$, where $X \sim \text{Bernoulli}(p)$.

Proof. It suffices to show that the moment-generating function of $X^{\alpha,\beta}$, $m_{X^{\alpha,\beta}}(t)$, converges pointwise to $m_X(t) = (1-p) + pe^t$, for all t, as $\alpha \to 0$. With $\beta = \beta(p) = \alpha(1-p)/p$, we have

$$\begin{aligned} m_{X^{\alpha,\beta}}(t) &= 1 + \sum_{l=1}^{\infty} \left\{ \prod_{s=0}^{l-1} \frac{\alpha + s}{\alpha + \alpha(1-p)/p + s} \right\} \frac{t^l}{l!} \\ &= 1 + \sum_{l=1}^{\infty} \left(\prod_{s=0}^{l-1} \frac{\alpha + s}{\alpha/p + s} \right) \frac{t^l}{l!} \\ &= 1 + \sum_{l=1}^{\infty} p \left(\prod_{s=1}^{l-1} \frac{\alpha + s}{\alpha/p + s} \right) \frac{t^l}{l!}. \end{aligned}$$

Since $p \prod_{s=1}^{l-1} (\alpha + s)/(\alpha/p + s) < 1$ for all $\alpha > 0$, the summand is bounded above by $t^l/l!$, which is integrable with respect to counting measure on $\{1, 2, ..., \}$. Thus, by the Dominated Convergence Theorem, we obtain

$$\lim_{\alpha \to 0} m_{X^{\alpha,\beta}}(t) = 1 + \sum_{l=1}^{\infty} p \lim_{\alpha \to 0} \left(\prod_{s=1}^{l-1} \frac{\alpha + s}{\alpha/p + s} \right) \frac{t^l}{l!} = 1 + p \sum_{l=1}^{\infty} \frac{t^l}{l!} = (1-p) + pe^t,$$

which completes the proof.

We now provide complete results from the second investigation described in Section 4.1.

Note: Some remarks on interpreting the classification accuracy figures (Figures 4-15) are given on the next page.

Remarks:

- For the informative procedures (GA and SA), recall that $PS_e^{\mathcal{I}_{jk}}$, $PS_p^{\mathcal{I}_{jk}}$, $PPV^{\mathcal{I}_{jk}}$ and $NPV^{\mathcal{I}_{jk}}$ are individual-specific measures of accuracy. Figures 4-15 attempt to summarize these $N = K^2$ accuracy measures globally. In all figures, the optimal square array size K (based on A) has been used for each (p, S_e, S_p) configuration.

- In Figures 4-6 (7-9), PS_e (PS_p) is computed as the arithmetic average of $PS_e^{\mathcal{I}_{jk}}$ ($PS_p^{\mathcal{I}_{jk}}$) for all N individuals. Therefore, PS_e (PS_p) can be interpreted as an "estimate" of $PS_e^{\mathcal{I}_{jk}}$ ($PS_p^{\mathcal{I}_{jk}}$) for a randomly-selected individual $\mathcal{I}_{jk}$.

- Both $PPV^{\mathcal{I}_{jk}}$ and $NPV^{\mathcal{I}_{jk}}$ are diagnostic-dependent; i.e., the interpretation of each depends on whether individual $\mathcal{I}_{jk}$ has been diagnosed as positive or negative. Therefore, we calculate PPV (NPV) in Figures 10-12 (13-15) as a weighted average of $PPV^{\mathcal{I}_{jk}}$ ($NPV^{\mathcal{I}_{jk}}$) for all N individuals, where the weight for individual $\mathcal{I}_{jk}$ is taken to be the probability that individual $\mathcal{I}_{jk}$ is diagnosed as positive (negative); the weights are then appropriately scaled to sum to one across all N individuals. In this context, PPV (NPV) can be thought of as an "estimate" of $PPV^{\mathcal{I}_{jk}}$ ($NPV^{\mathcal{I}_{jk}}$) for a randomly-selected individual who has been diagnosed as positive (negative). We can think of no better to way to amalgamate the N values of $PPV^{\mathcal{I}_{jk}}$ ($NPV^{\mathcal{I}_{jk}}$) to produce a single measure.

- Figures 4-9 largely demonstrate that there is no substantial loss in pooling sensitivity and pooling specificity when using GA and SA (compared to A). There is moderate evidence that GA can increase pooling sensitivity (on average) and that both GA and SA can increase pooling specificity (on average), more noticeably when p is larger and when the amount of heterogeneity is larger (i.e., when α is smaller).

- Figures 10-15 show that (on average) there is usually no loss in pooling positive predictive value when using GA and SA (compared to A), unless the amount of heterogeneity is large (e.g., $\alpha = 0.10$), and that there are gains in negative pooling predictive value. These interpretations are based on our use of the weighted measures described above.

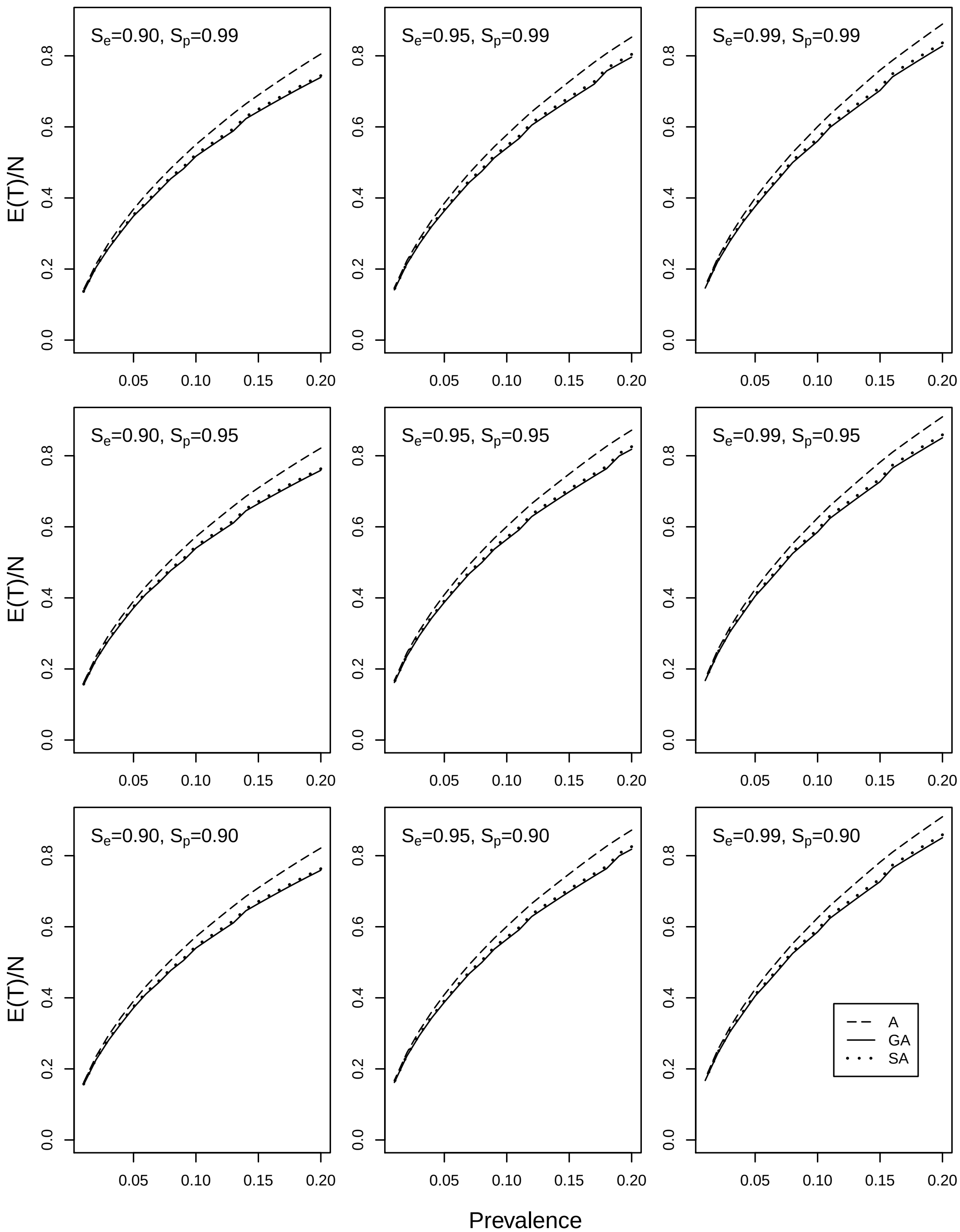

Figure 1: Efficiency comparison with imperfect testing. Per-individual efficiency for GA, SA, and A with $\alpha = 1$. $E(T|\text{A})$ has been approximated using Equation (13) in Kim et al. (2007). The optimal square array size K has been used for each (p, S_e, S_p) configuration.

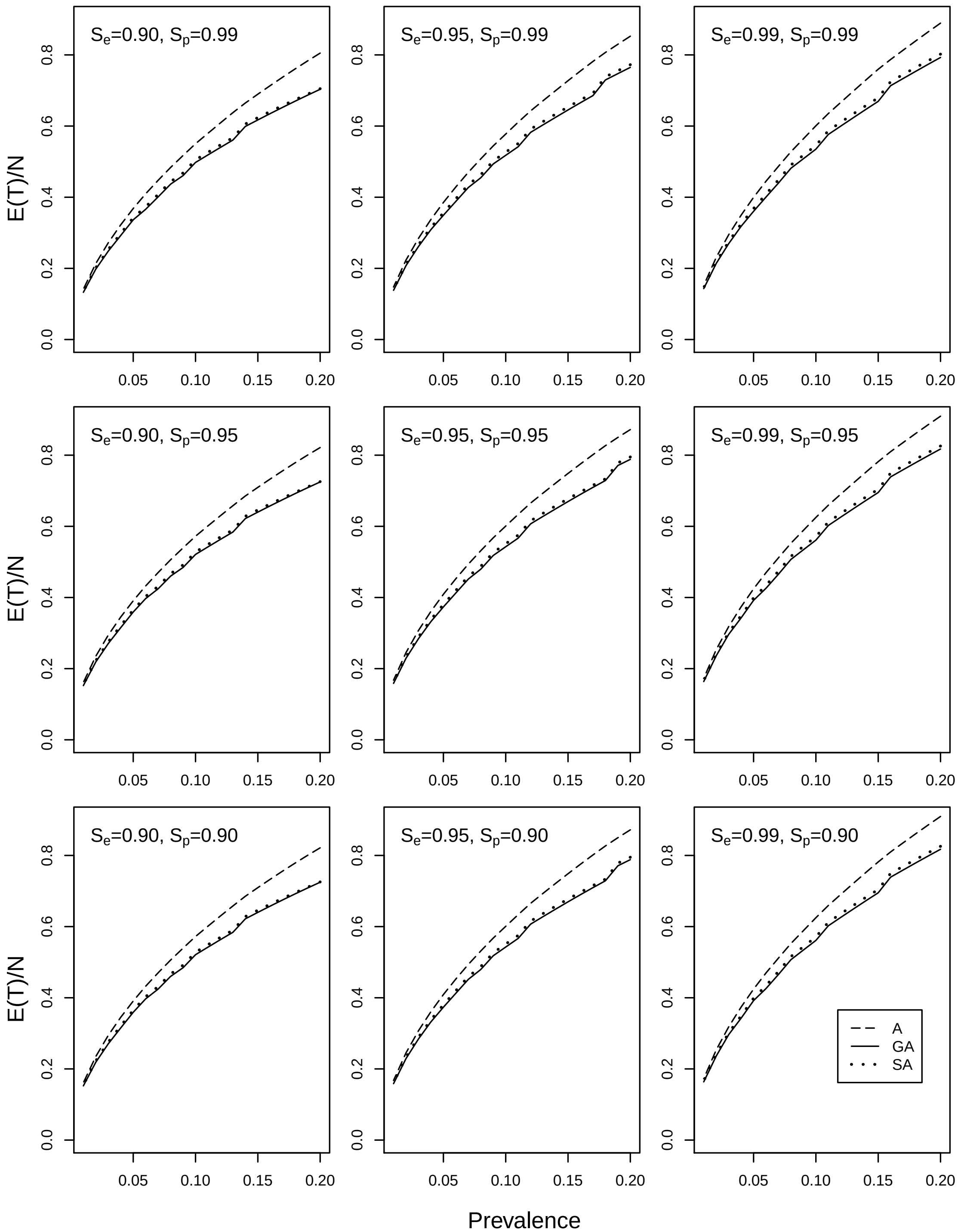

Figure 2: Efficiency comparison with imperfect testing. Per-individual efficiency for GA, SA, and A with $\alpha = 0.50$. $E(T|\mathrm{A})$ has been approximated using Equation (13) in Kim et al. (2007). The optimal square array size K has been used for each (p, S_e, S_p) configuration.

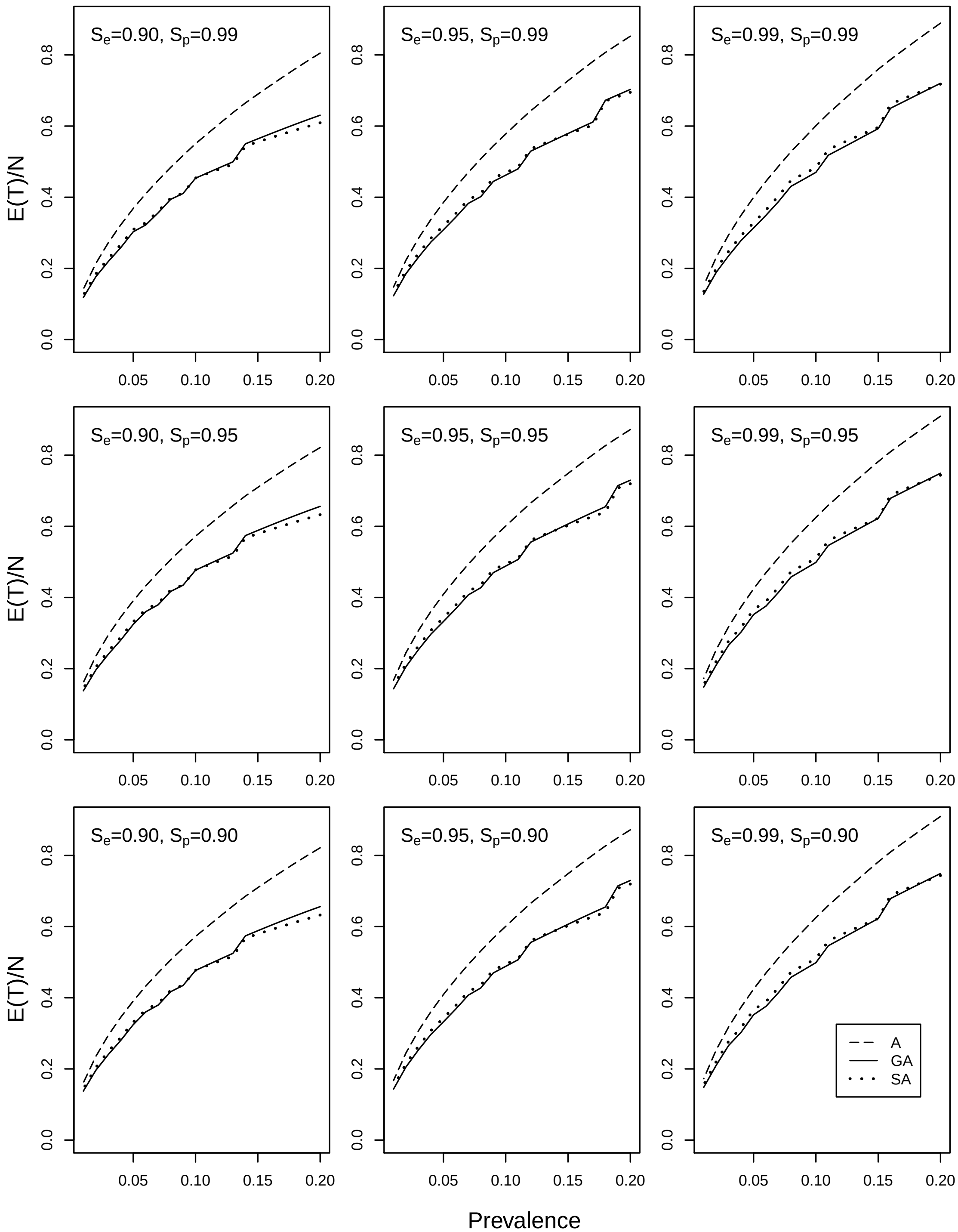

Figure 3: Efficiency comparison with imperfect testing. Per-individual efficiency for GA, SA, and A with $\alpha = 0.10$. $E(T|\text{A})$ has been approximated using Equation (13) in Kim et al. (2007). The optimal square array size K has been used for each (p, S_e, S_p) configuration.

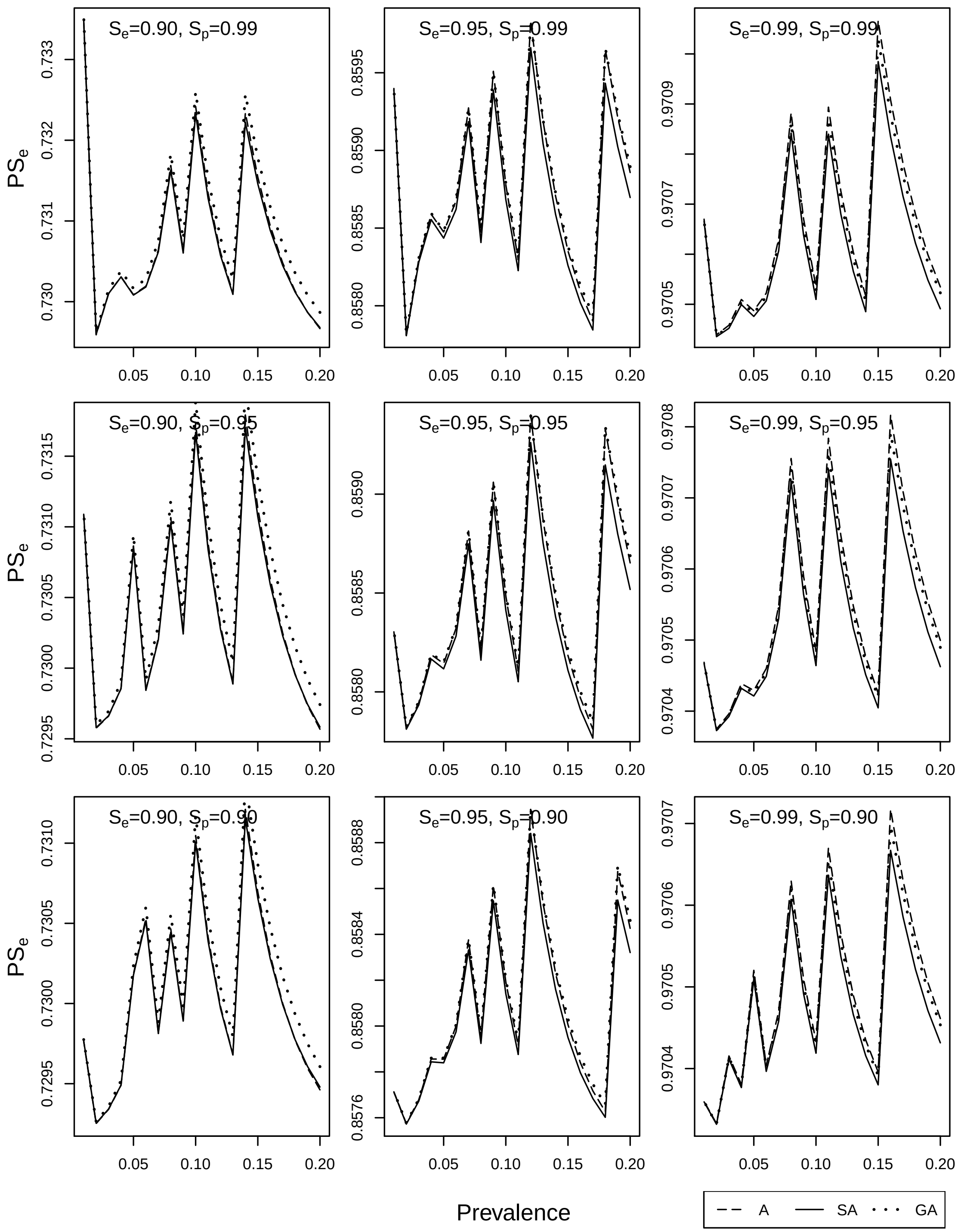

Figure 4: Pooling sensitivity comparison for GA, SA, and A with $\alpha = 1$. The optimal square array size K has been used for each (p, S_e, S_p) configuration.

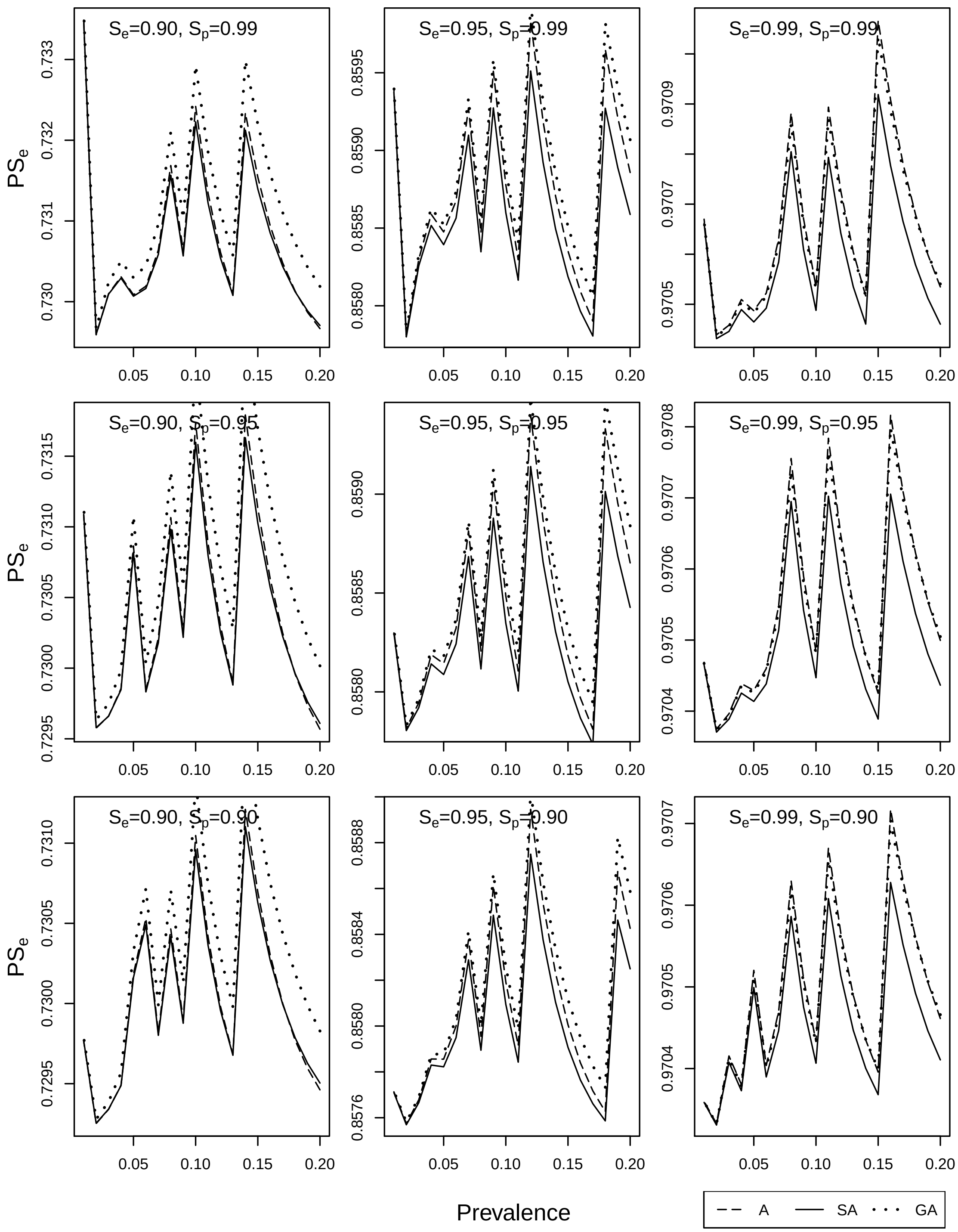

Figure 5: Pooling sensitivity comparison for GA, SA, and A with $\alpha = 0.50$. The optimal square array size K has been used for each (p, S_e, S_p) configuration.

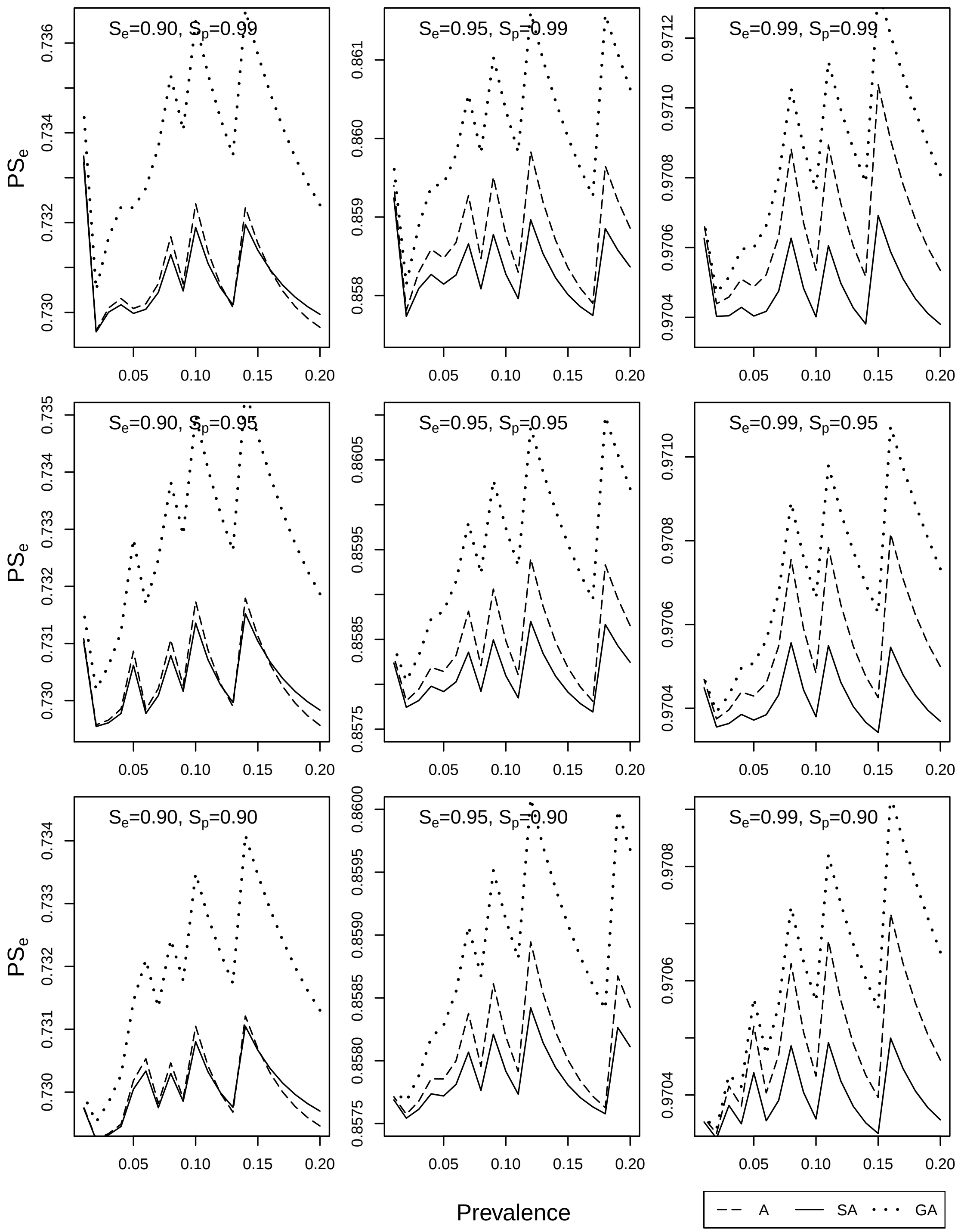

Figure 6: Pooling sensitivity comparison for GA, SA, and A with $\alpha = 0.10$. The optimal square array size K has been used for each (p, S_e, S_p) configuration.

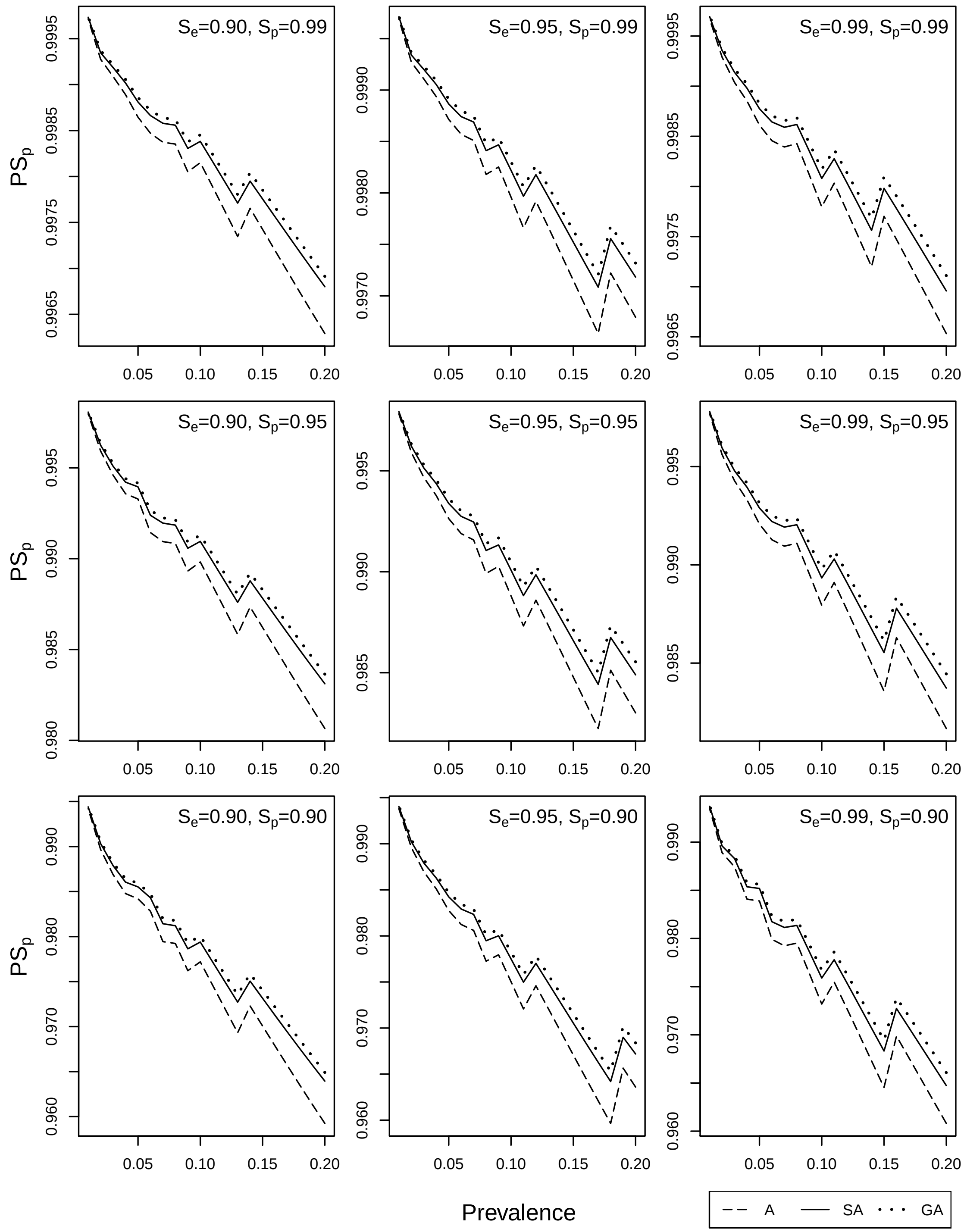

Figure 7: Pooling specificity comparison for GA, SA, and A with $\alpha = 1$. The optimal square array size K has been used for each (p, S_e, S_p) configuration.

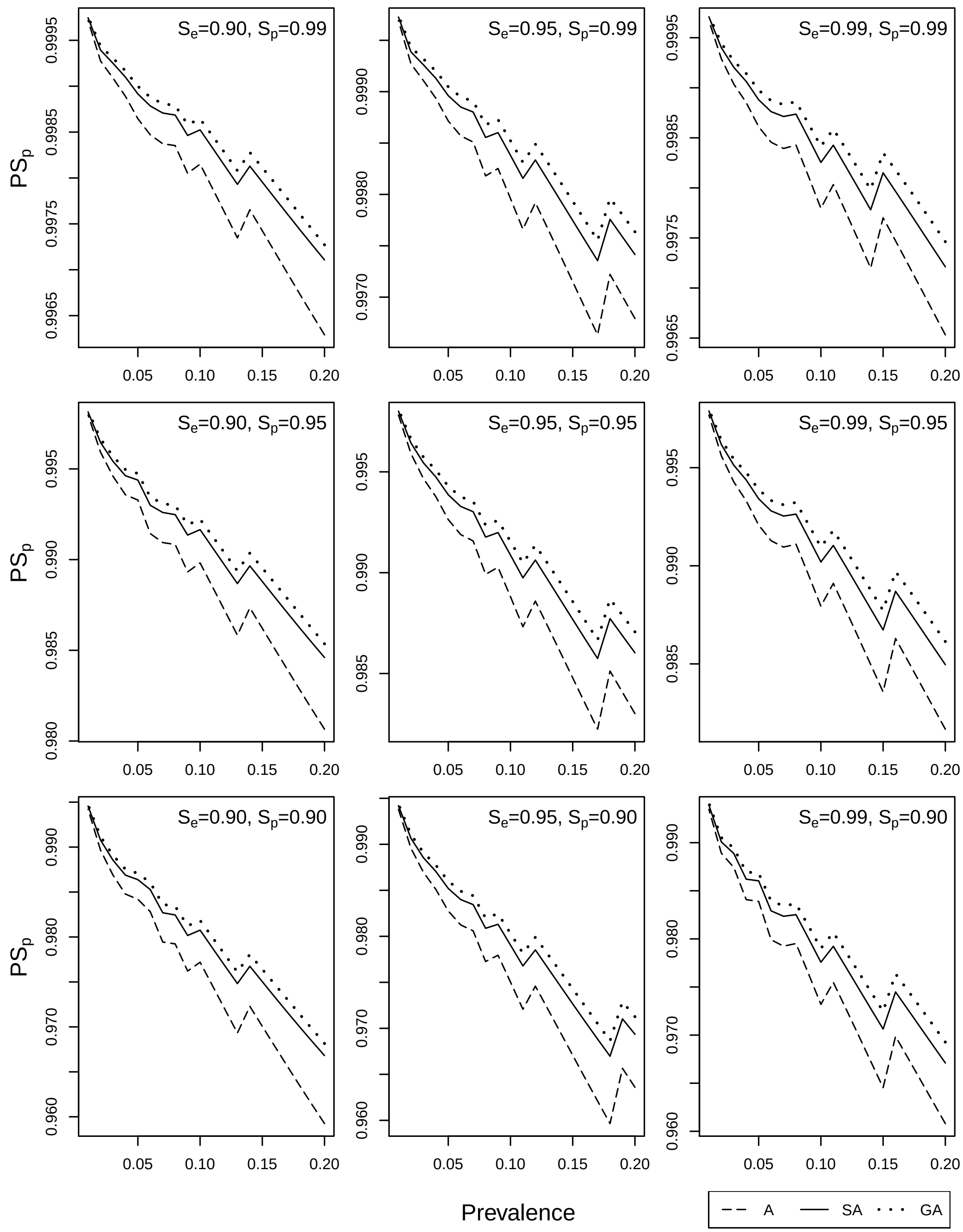

Figure 8: Pooling specificity comparison for GA, SA, and A with $\alpha = 0.50$. The optimal square array size K has been used for each (p, S_e, S_p) configuration.

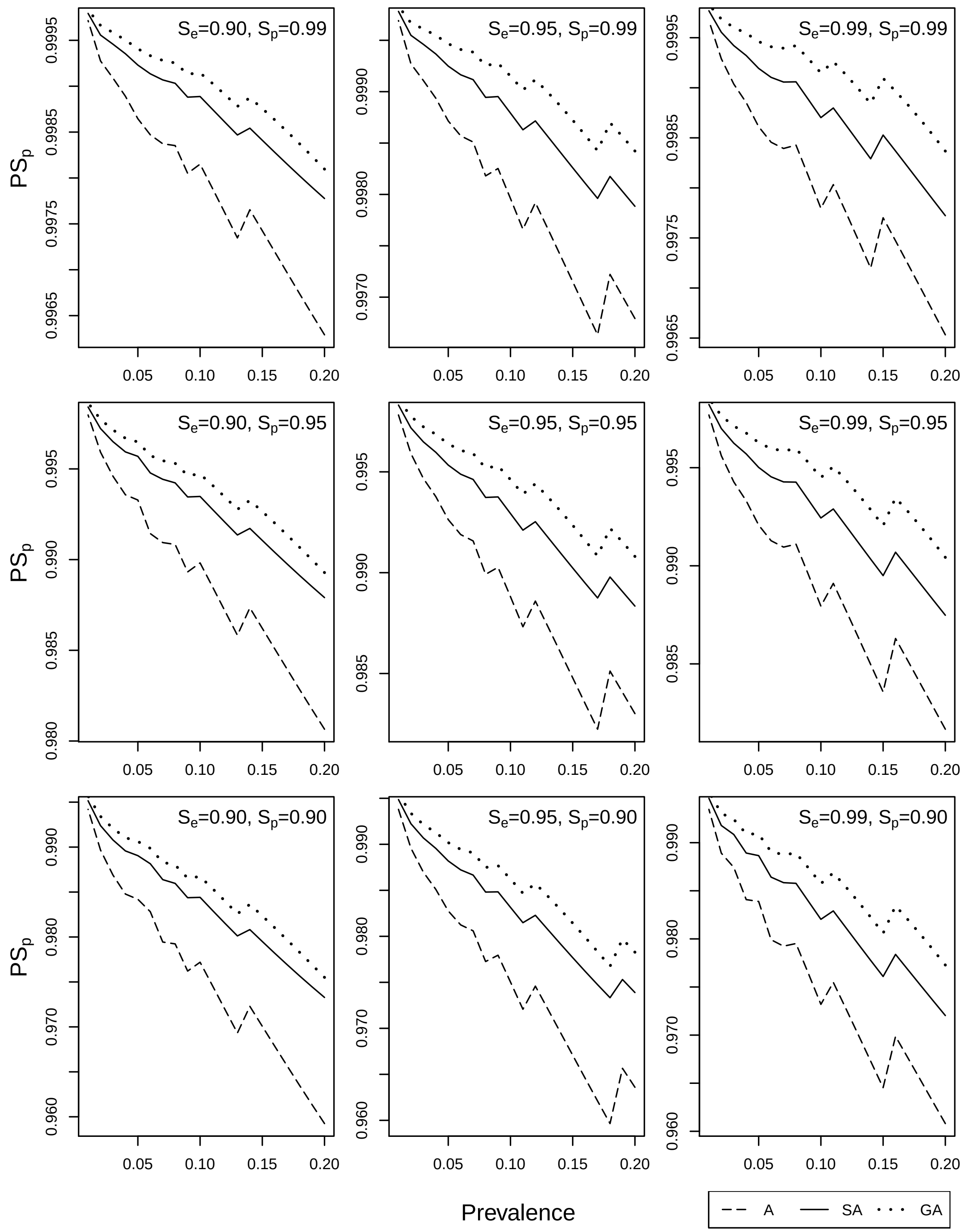

Figure 9: Pooling specificity comparison for GA, SA, and A with $\alpha = 0.10$. The optimal square array size K has been used for each (p, S_e, S_p) configuration.

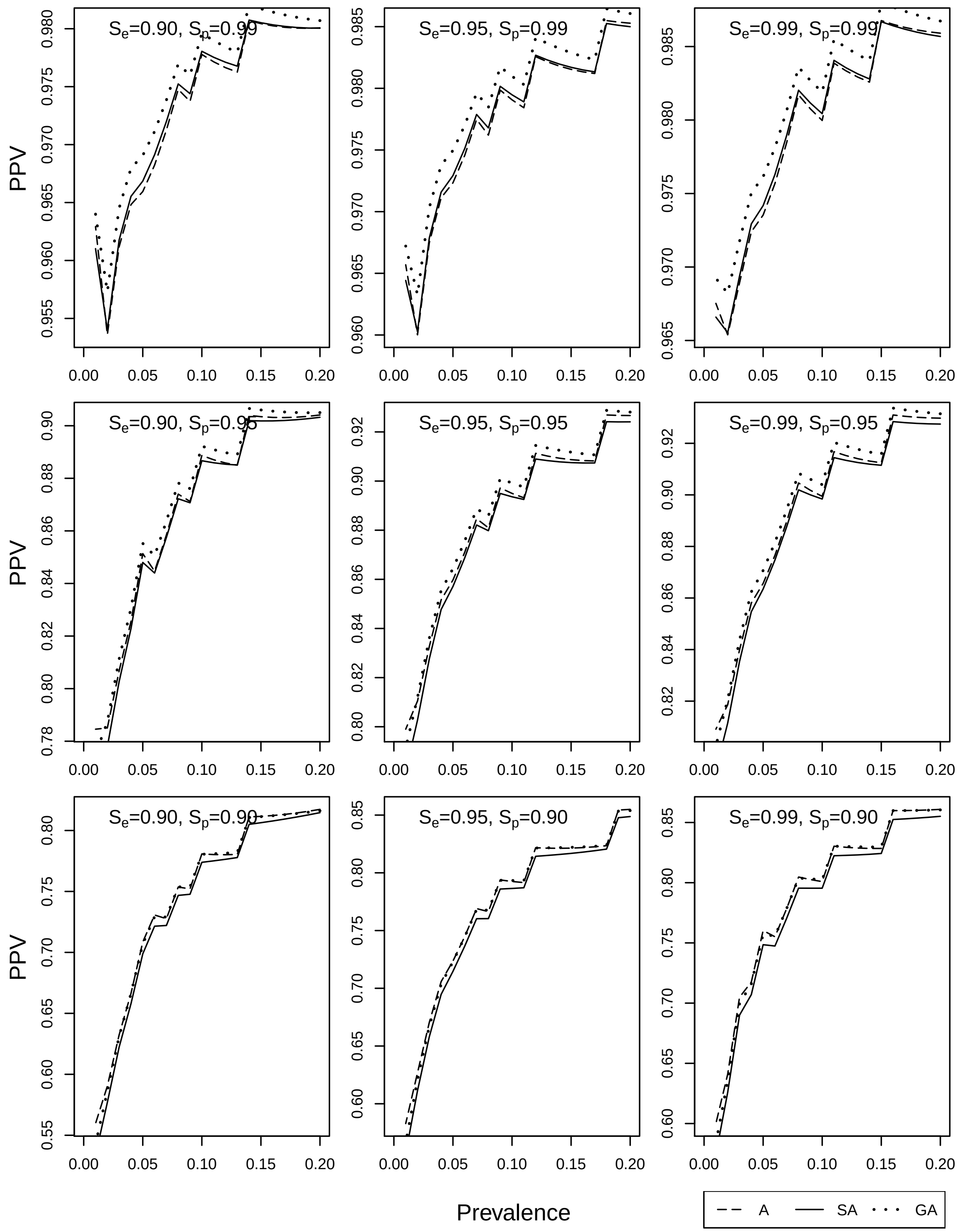

Figure 10: Pooling positive predictive value comparison for GA, SA, and A with $\alpha = 1$. The optimal square array size K has been used for each (p, S_e, S_p) configuration.

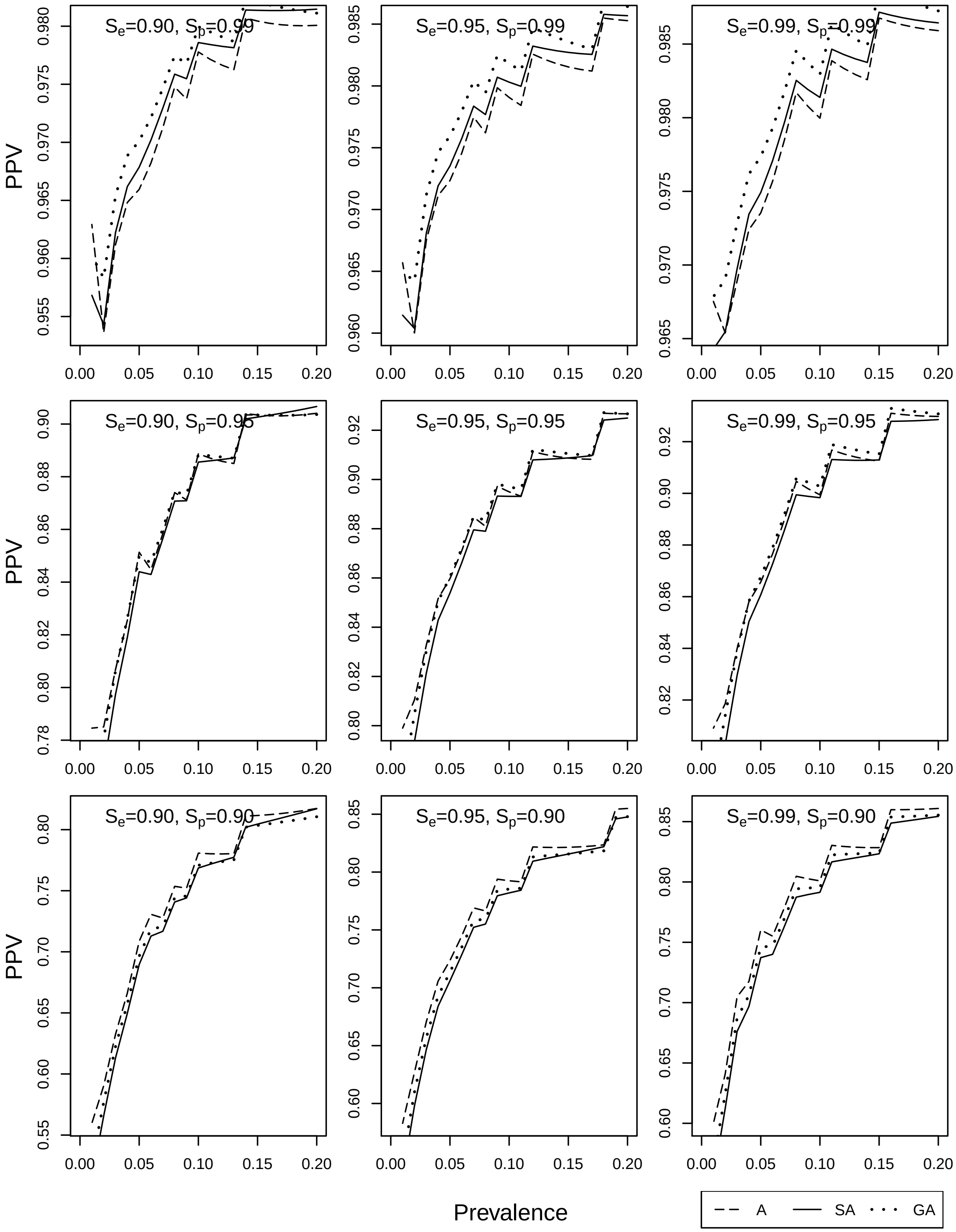

Figure 11: Pooling positive predictive value comparison for GA, SA, and A with $\alpha = 0.50$. The optimal square array size K has been used for each (p, S_e, S_p) configuration.

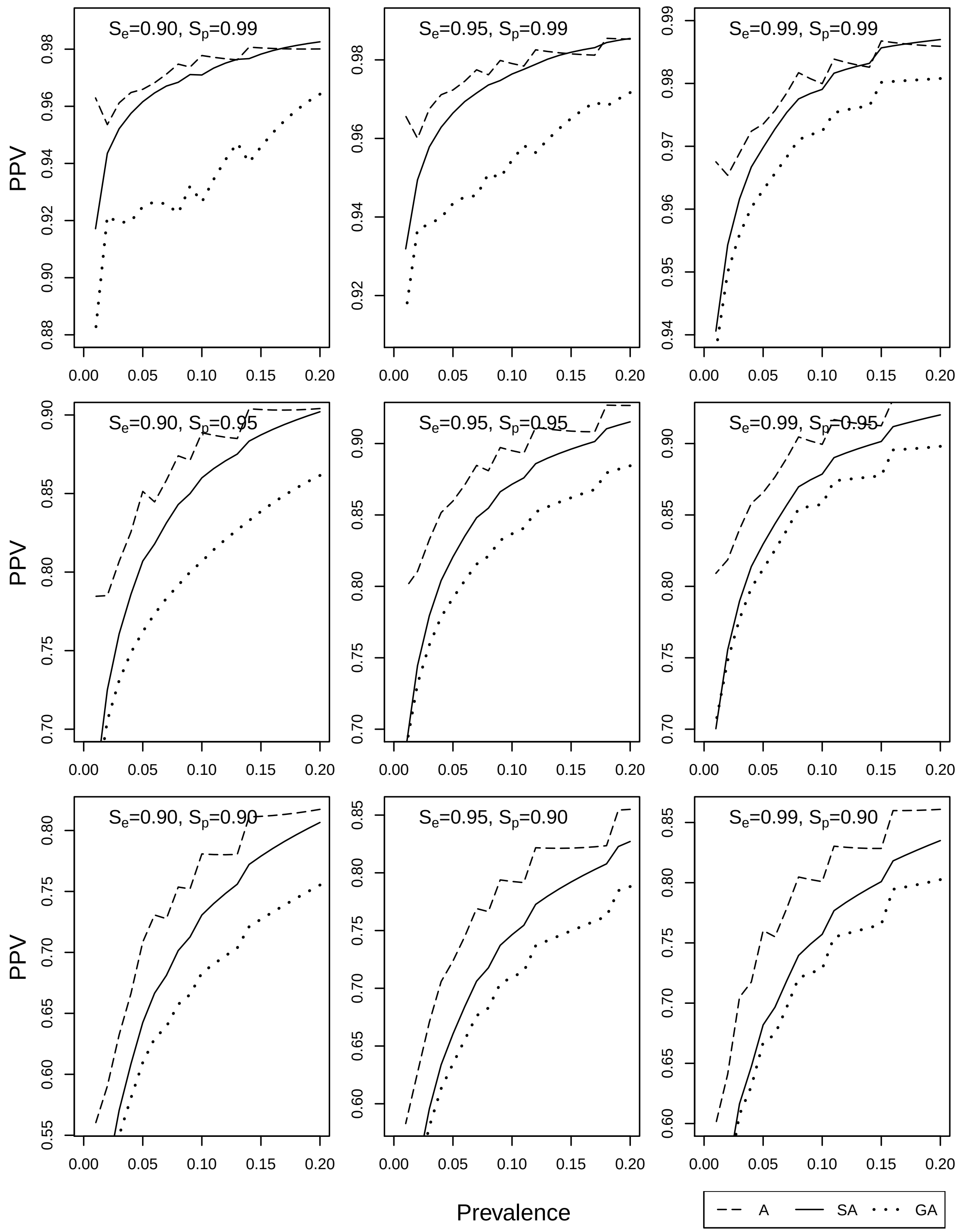

Figure 12: Pooling positive predictive value comparison for GA, SA, and A with $\alpha = 0.10$. The optimal square array size K has been used for each (p, S_e, S_p) configuration.

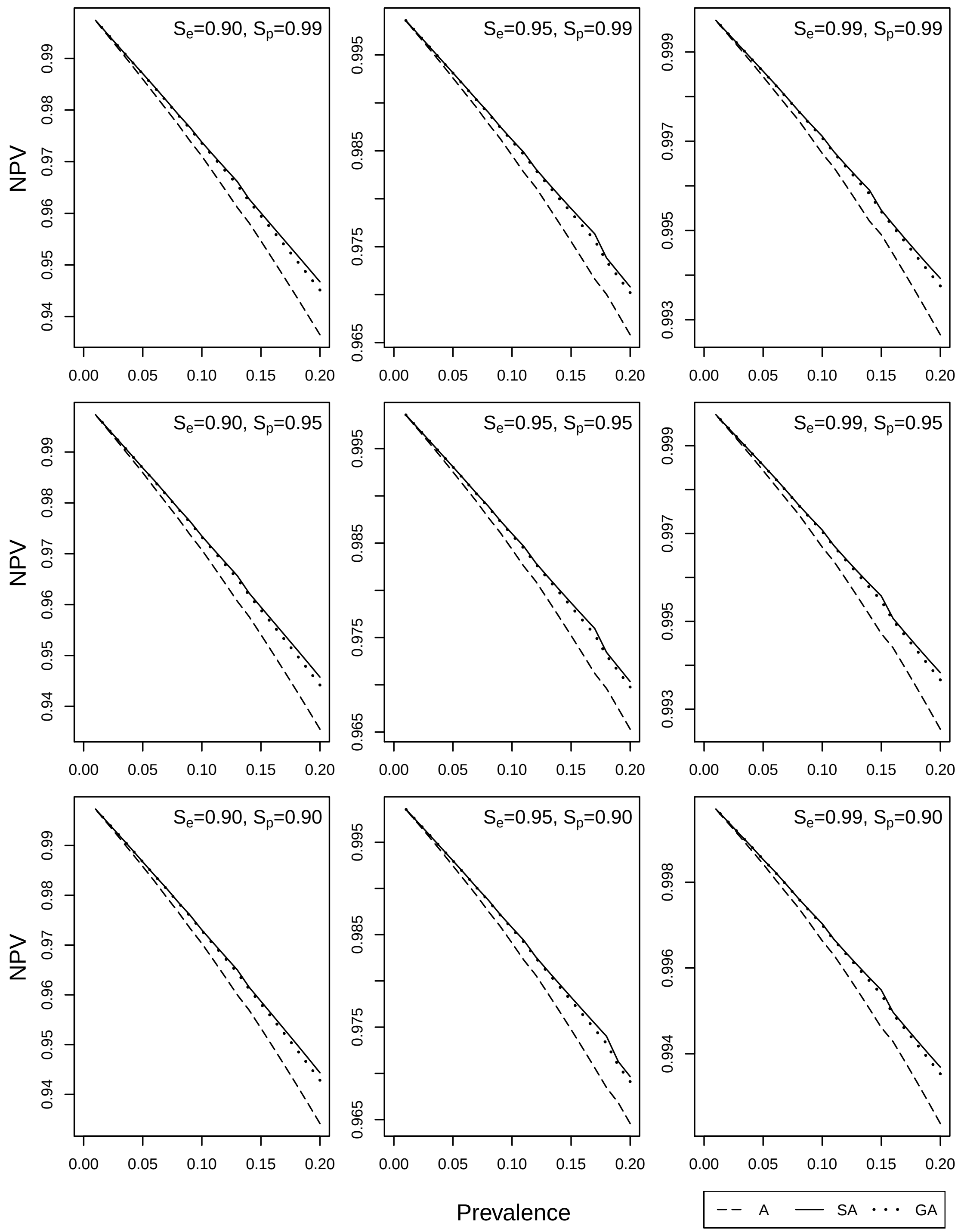

Figure 13: Pooling negative predictive value comparison for GA, SA, and A with $\alpha = 1$. The optimal square array size K has been used for each (p, S_e, S_p) configuration.

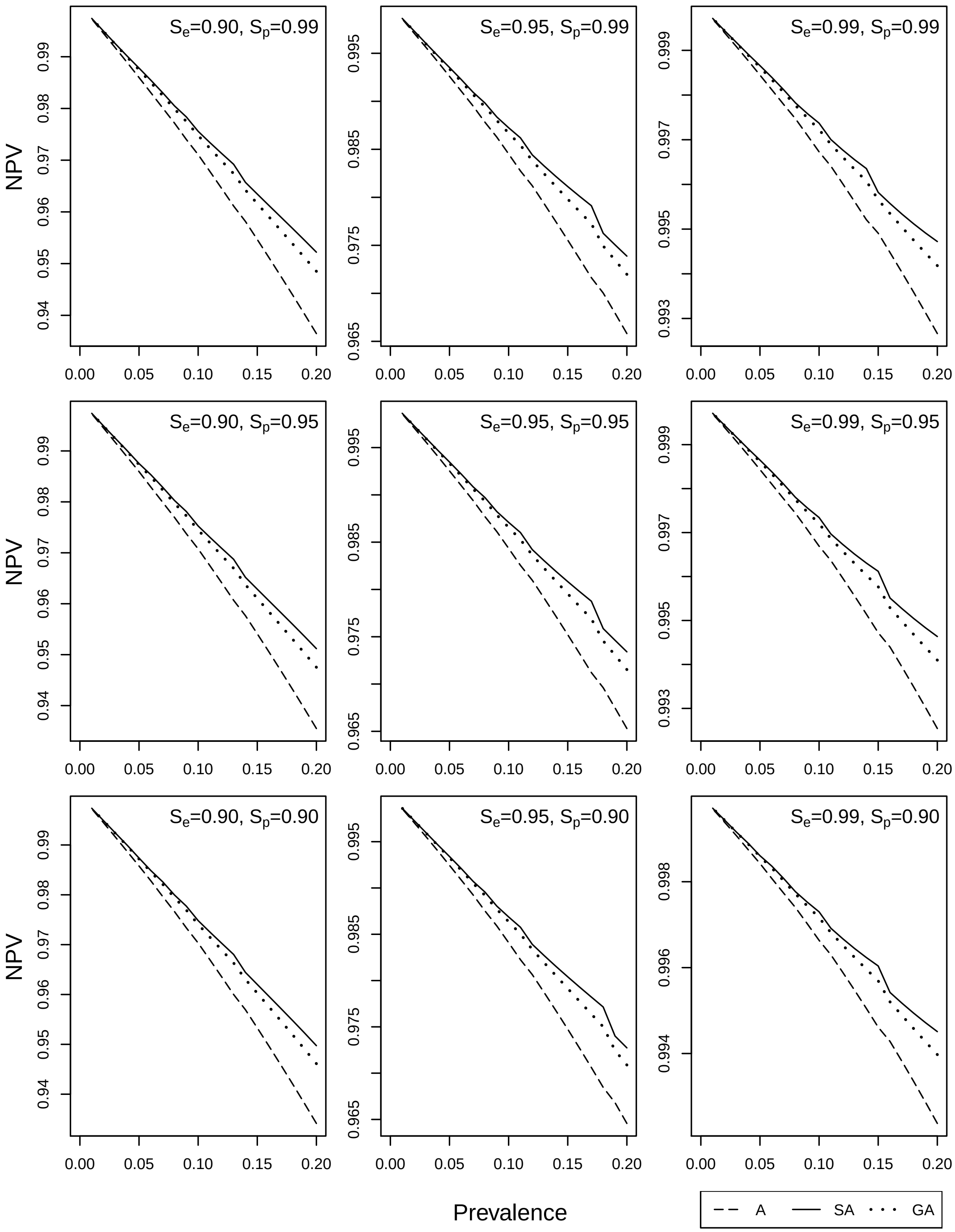

Figure 14: Pooling negative predictive value comparison for GA, SA, and A with $\alpha = 0.50$. The optimal square array size K has been used for each (p, S_e, S_p) configuration.

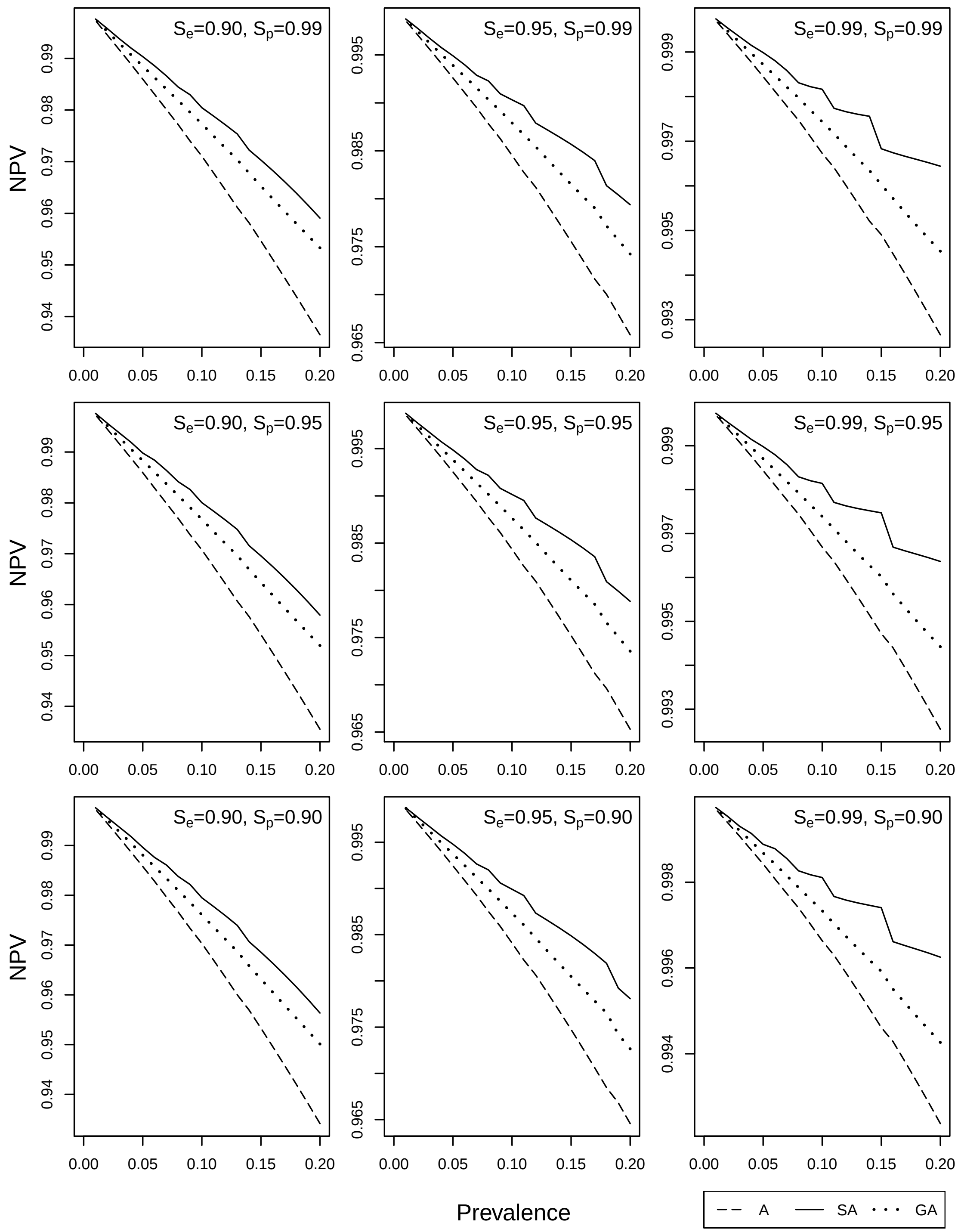

Figure 15: Pooling negative predictive value comparison for GA, SA, and A with $\alpha = 0.10$. The optimal square array size K has been used for each (p, S_e, S_p) configuration.

Web Appendix D: *Supplemental material for Section* 4.2. We include all comparisons among GA, PSOD, and FIS described in Section 4.2.

Remarks:

- GA = Gradient Array; PSOD = Pool-Specific Optimal Dorfman (McMahan et al., 2011); FIS = Full Informative Sterrett (Bilder et al., 2010).

- GA and PSOD are both two-stage procedures. FIS contains at least three stages and at most $2(K-1)$, where K is the pool size.

- The average number of stages needed to decode positive pools is computed exactly for FIS (Table 1) at each (α, p, S_e, S_p) configuration.

- For example, when $\alpha = 0.50$, $p = 0.01$, and $S_e = S_p = 0.95$, FIS needs, on average, 6.8 stages to decode positive pools.

Table 1: Expected number of stages required by FIS to decode positive pools. The optimal pool size is used at each (α, p, S_e, S_p) configuration.

		$S_e = 0.90$			$S_e = 0.95$			$S_e = 0.99$		
		S_p			S_p			S_p		
α	p	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
1	0.01	10.2	10.6	9.8	10.1	10.3	9.3	10.0	10.0	8.9
	0.05	8.5	7.1	6.4	8.0	7.0	6.2	7.6	6.8	6.0
	0.10	7.7	6.5	5.8	6.7	6.0	5.7	6.7	6.0	5.7
	0.20	14.4	14.2	14.1	7.7	6.8	6.0	6.4	6.2	5.4
0.50	0.01	7.4	7.3	6.1	7.1	6.8	5.3	6.8	6.3	4.7
	0.05	6.3	6.3	6.1	6.1	5.8	5.4	5.8	5.5	4.9
	0.10	7.2	7.2	7.1	7.2	7.0	6.7	7.1	6.8	6.3
	0.20	9.8	9.7	9.6	10.4	10.2	9.9	10.9	10.5	10.1
0.10	0.01	7.8	7.7	6.6	7.5	7.2	5.8	7.3	6.8	5.2
	0.05	6.9	6.8	6.6	6.7	6.4	6.0	6.5	6.1	5.5
	0.10	7.8	7.7	7.6	7.8	7.6	7.3	7.8	7.5	7.0
	0.20	10.3	10.2	10.1	11.0	10.8	10.5	11.7	11.3	10.8

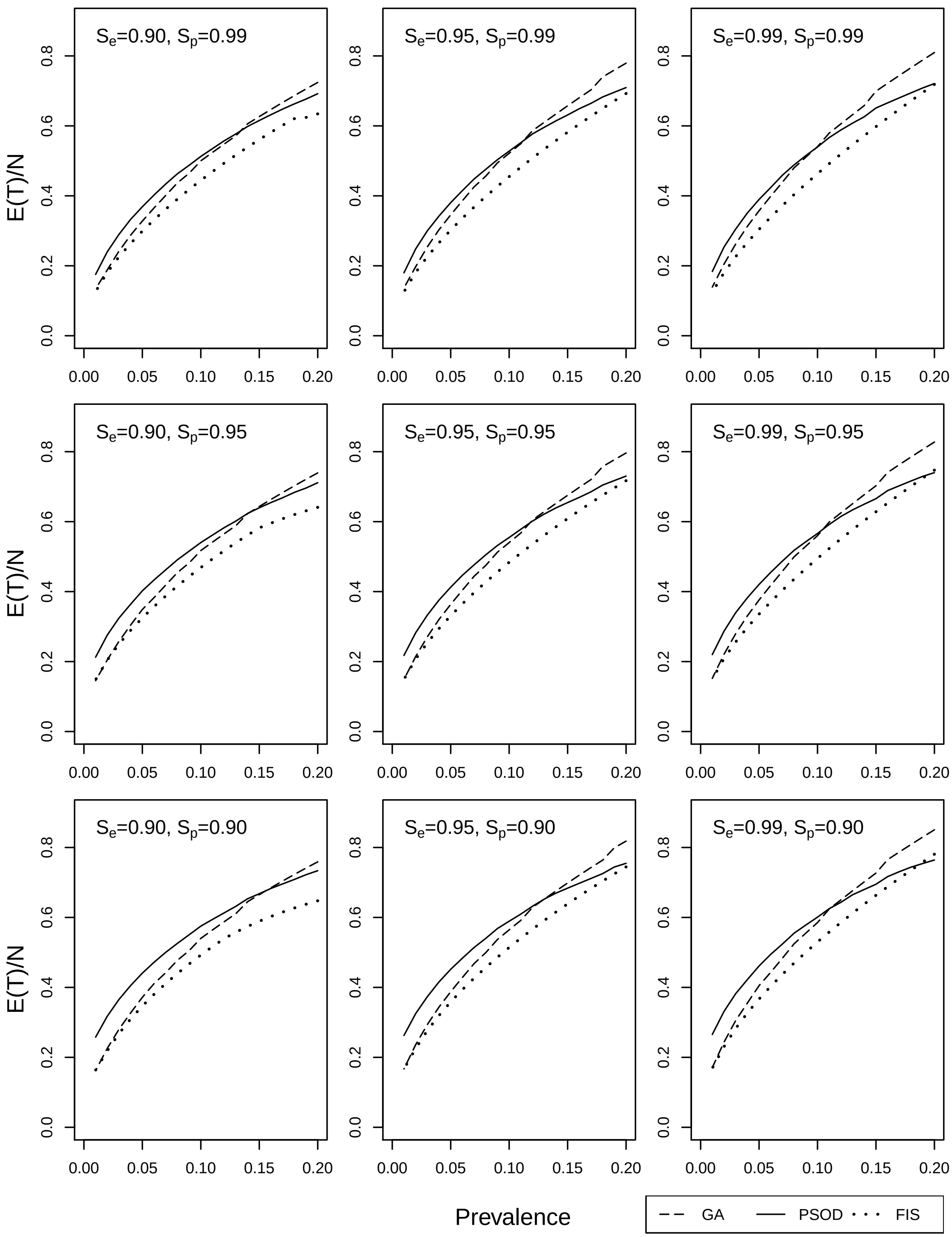

Figure 1: Efficiency comparison with other informative procedures. Per-individual efficiency for GA, PSOD, and FIS with $\alpha = 1$. The optimal pool size has been used for each (p, S_e, S_p) configuration; see Section 4.2.

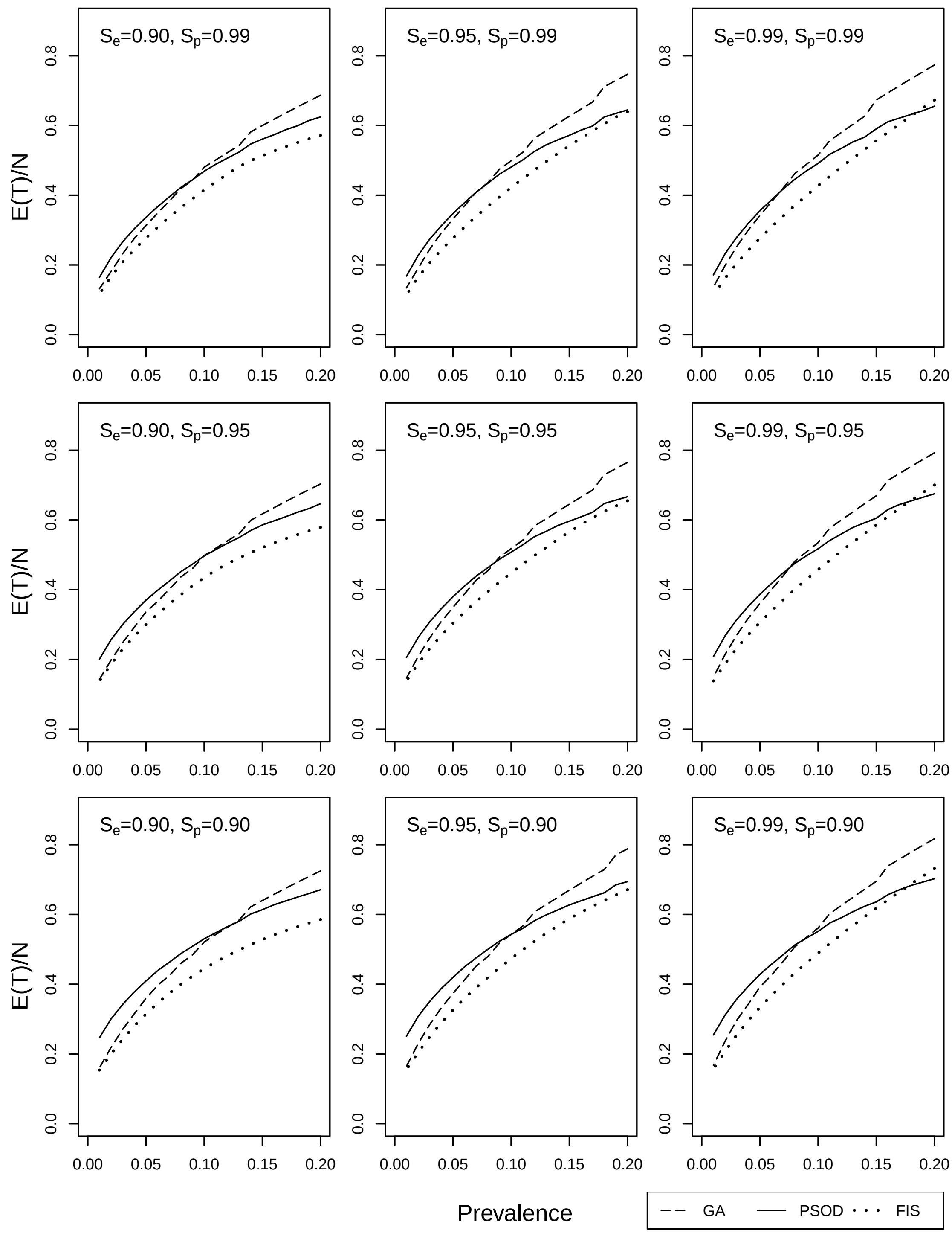

Figure 2: Efficiency comparison with other informative procedures. Per-individual efficiency for GA, PSOD, and FIS with $\alpha = 0.50$. The optimal pool size has been used for each (p, S_e, S_p) configuration; see Section 4.2.

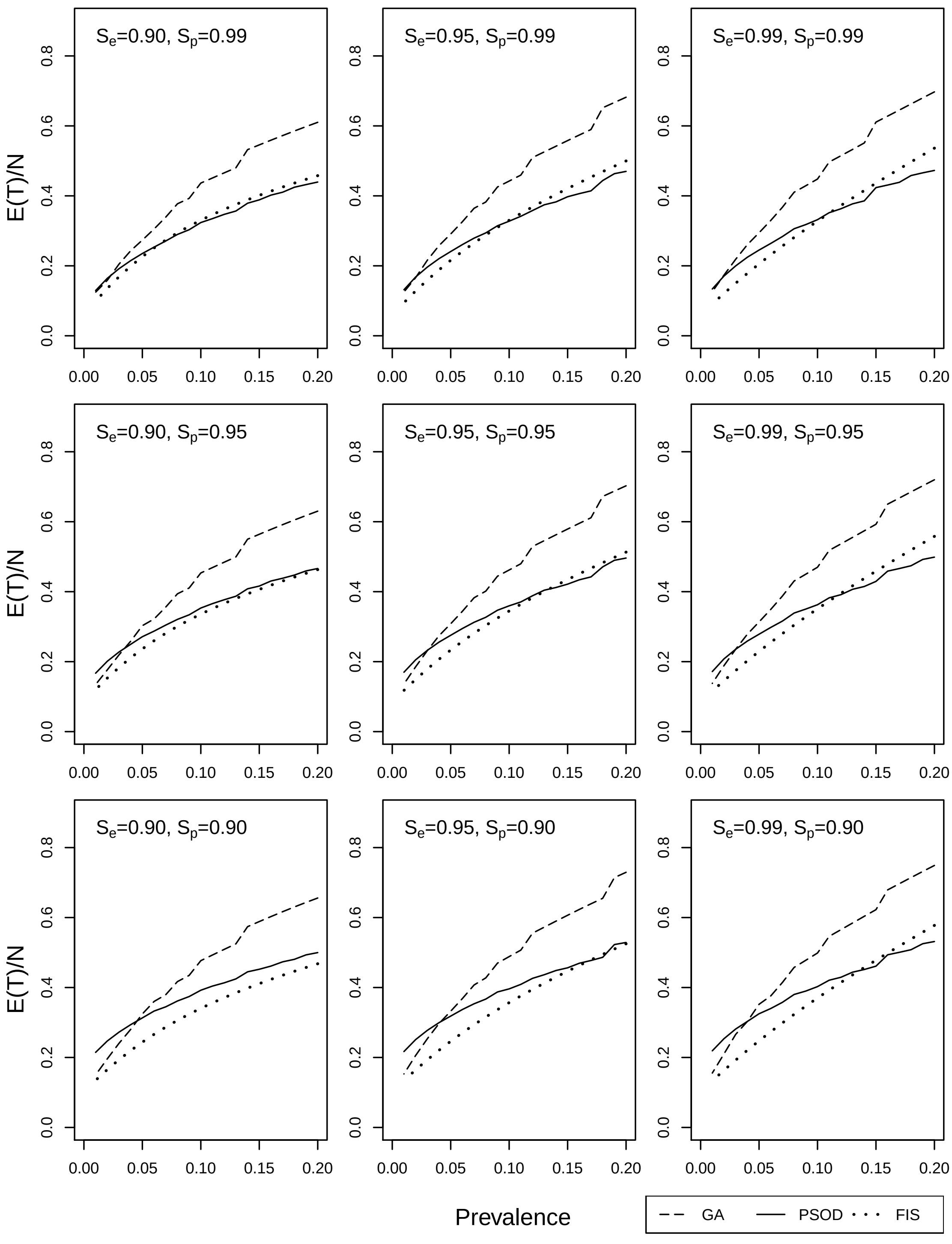

Figure 3: Efficiency comparison with other informative procedures. Per-individual efficiency for GA, PSOD, and FIS with $\alpha = 0.10$. The optimal pool size has been used for each (p, S_e, S_p) configuration; see Section 4.2.

Web Appendix E: *Supplemental material for Section* 5*.* We provide the additional results from the Nebraska IPP analyses.

Remarks:

- Classification accuracy measures in Web Appendix E (and in Section 5 of the paper) do not use the individual-specific formulae in Section 2 of the paper. For each set of diagnoses (among the $B = 1000$ simulated sets), we calculate PS_e, PS_p, PPV and NPV by matching the simulated diagnoses with the 2009 responses provided to us by the NPHL (which we assume to be the true responses). This is done for each of the $B = 1000$ data sets and we then take a simple arithmetic average to produce the measures $\overline{PS}_e$, $\overline{PS}_p$, $\overline{PPV}$, and $\overline{NPV}$.

- In McMahan et al. (2011), 2008 modeling was done within each infection-specimen stratum, and gender was a covariate. In this manuscript, modeling was done within each infection-gender-specimen stratum. Because of this difference, we removed 25 female individuals from the 2009 screening population to carry out the comparisons in this paper. Reason: There was one site that screened no females in 2008 but screened these 25 females in 2009. For the 2008 female models, this site estimate was vacuous.

Table 1: Nebraska IPP screening results for 2009 using best subsets model fits. Mean number of tests ($\overline{T}$) and accuracy measures ($\overline{PS_e}$, $\overline{PS_p}$, $\overline{PPV}$, and $\overline{NPV}$), averaged over 1000 implementations. The average number of stages required to decode positive pools and optimal pool sizes are also given. PSOD does not use a common pool size. Gender/specimen individual counts and values of S_e and S_p are in Table 1 of the manuscript.

Infection	Gend/Spec	Method	Pool Size	$\overline{T}$	$\overline{PS_e}$	$\overline{PS_p}$	$\overline{PPV}$	$\overline{NPV}$	# Stages
Chlamydia	Female/Urine	A	9	2123.3	0.525	0.993	0.875	0.960	2
		SA	9	2143.6	0.524	0.993	0.871	0.960	2
		GA	9	2112.4	0.527	0.994	0.877	0.960	2
		PSOD	-	2478.5	0.649	0.990	0.849	0.970	2
		FIS	13	2141.6	0.579	0.989	0.828	0.964	9.7
	Female/Swab	A	8	6569.0	0.802	0.994	0.906	0.985	2
		SA	8	6535.1	0.801	0.994	0.908	0.985	2
		GA	8	6444.9	0.801	0.994	0.912	0.985	2
		PSOD	-	7136.3	0.862	0.991	0.873	0.990	2
		FIS	8	5965.9	0.842	0.993	0.897	0.988	6.4
	Male/Urine	A	8	3096.9	0.806	0.990	0.876	0.983	2
		SA	8	3005.8	0.806	0.991	0.884	0.983	2
		GA	8	2938.5	0.806	0.991	0.892	0.983	2
		PSOD	-	3264.1	0.865	0.987	0.852	0.988	2
		FIS	8	2744.4	0.843	0.990	0.884	0.986	6.4
	Male/Swab	A	6	1356.0	0.793	0.986	0.911	0.962	2
		SA	6	1330.3	0.792	0.986	0.915	0.962	2
		GA	6	1328.1	0.793	0.987	0.916	0.963	2
		PSOD	-	1308.7	0.866	0.982	0.901	0.975	2
		FIS	7	1218.1	0.822	0.986	0.917	0.968	6.2
Gonorrhea	Female/Urine	A	17	876.1	0.616	0.999	0.918	0.994	2
		SA	17	893.9	0.616	0.999	0.913	0.994	2
		GA	17	905.2	0.615	0.999	0.910	0.994	2
		PSOD	-	1225.1	0.719	0.998	0.852	0.995	2
		FIS	14	959.3	0.699	0.998	0.860	0.995	10.3
	Female/Swab	A	21	2427.5	0.903	0.999	0.908	0.999	2
		SA	21	2332.2	0.904	0.999	0.917	0.999	2
		GA	21	2246.8	0.903	0.999	0.926	0.999	2
		PSOD	-	3090.7	0.933	0.998	0.861	0.999	2
		FIS	21	2176.5	0.926	0.998	0.888	0.999	10.1
	Male/Urine	A	21	1575.5	0.914	0.994	0.775	0.998	2
		SA	21	1387.8	0.913	0.995	0.814	0.998	2
		GA	21	1362.1	0.915	0.996	0.821	0.998	2
		PSOD	-	1695.5	0.942	0.994	0.772	0.999	2
		FIS	22	1257.2	0.928	0.995	0.809	0.998	10.2
	Male/Swab	A	8	930.1	0.956	0.993	0.908	0.997	2
		SA	8	835.8	0.956	0.995	0.932	0.997	2
		GA	8	793.2	0.956	0.996	0.944	0.997	2
		PSOD	-	771.4	0.979	0.991	0.895	0.998	2
		FIS	17	642.9	0.961	0.994	0.925	0.997	8.0

Table 2: Best subsets results. In each gender-infection-specimen stratum, the best subset of covariates was determined by minimizing BIC. Age and Age^2 were included by default.

Infection/Specimen	Male	Female
Chlamydia/Urine	Age Age^2 Symptoms STD Contact	Age Age^2
Chlamydia/Swab	Age Age^2 Urethritis STD Contact	Age Age^2 Family Plan Symptoms Multiple Partners STD Contact
Gonorrhea/Urine	Age Age^2 Symptoms	Age Age^2
Gonorrhea/Swab	Age Age^2 Urethritis STD Contact Symptoms	Age Age^2 Family Plan Symptoms Multiple Partners STD Contact Location

Table 3: Nebraska IPP screening results for 2009 with maximum pool size $K^* = 10$. Mean number of tests ($\overline{T}$) and accuracy measures ($\overline{PS_e}$, $\overline{PS_p}$, $\overline{PPV}$, and $\overline{NPV}$), averaged over 1000 implementations. The average number of stages required to decode positive pools and optimal pool sizes are also given. PSOD does not use a common pool size. Gender/specimen individual counts and values of S_e and S_p are in Table 1 of the manuscript.

Infection	Gend/Spec	Method	Pool Size	$\overline{T}$	$\overline{PS_e}$	$\overline{PS_p}$	$\overline{PPV}$	$\overline{NPV}$	# Stages
Chlamydia	Female/Urine	A	9	2121.6	0.525	0.993	0.873	0.960	2
		SA	9	2126.1	0.524	0.993	0.873	0.960	2
		GA	9	2078.1	0.524	0.994	0.881	0.960	2
		PSOD	-	2479.8	0.648	0.989	0.842	0.970	2
		FIS	10	2034.3	0.597	0.991	0.859	0.966	7.7
	Female/Swab	A	8	6566.4	0.801	0.994	0.906	0.985	2
		SA	8	6470.6	0.802	0.994	0.910	0.985	2
		GA	8	6412.1	0.802	0.994	0.912	0.985	2
		PSOD	-	7026.5	0.863	0.991	0.875	0.990	2
		FIS	8	5913.2	0.842	0.993	0.899	0.988	6.3
	Female/Urine	A	8	3095.6	0.807	0.990	0.877	0.983	2
		SA	8	2973.0	0.807	0.991	0.889	0.983	2
		GA	8	2931.6	0.807	0.991	0.893	0.983	2
		PSOD	-	3237.9	0.870	0.987	0.856	0.989	2
		FIS	8	2694.4	0.843	0.991	0.889	0.986	6.3
	Male/Swab	A	6	1355.4	0.793	0.986	0.911	0.963	2
		SA	6	1342.6	0.793	0.986	0.913	0.963	2
		GA	6	1319.1	0.793	0.987	0.917	0.963	2
		PSOD	-	1278.8	0.870	0.982	0.902	0.976	2
		FIS	7	1213.8	0.825	0.986	0.919	0.968	6.2
Gonorrhea	Male/Urine	A	10	1230.9	0.655	0.999	0.939	0.994	2
		SA	10	1227.8	0.656	0.999	0.940	0.994	2
		GA	10	1220.5	0.668	0.999	0.945	0.994	2
		PSOD	-	1188.7	0.722	0.998	0.861	0.995	2
		FIS	10	944.5	0.706	0.999	0.904	0.995	7.1
	Female/Swab	A	10	3456.1	0.920	0.999	0.959	0.999	2
		SA	10	3451.1	0.920	0.999	0.959	0.999	2
		GA	10	3429.2	0.921	0.999	0.962	0.999	2
		PSOD	-	3136.2	0.936	0.998	0.880	0.999	2
		FIS	10	2452.7	0.929	0.999	0.932	0.999	6.2
	Male/Urine	A	10	1630.7	0.919	0.998	0.915	0.998	2
		SA	10	1594.3	0.918	0.998	0.924	0.998	2
		GA	10	1601.8	0.920	0.998	0.922	0.998	2
		PSOD	-	1666.2	0.942	0.995	0.801	0.999	2
		FIS	10	1334.3	0.937	0.997	0.867	0.999	6.4
	Male/Swab	A	8	934.5	0.956	0.993	0.907	0.997	2
		SA	8	802.3	0.957	0.996	0.943	0.997	2
		GA	8	775.9	0.958	0.996	0.950	0.997	2
		PSOD	-	720.8	0.978	0.993	0.911	0.998	2
		FIS	10	620.6	0.966	0.996	0.951	0.997	5.2

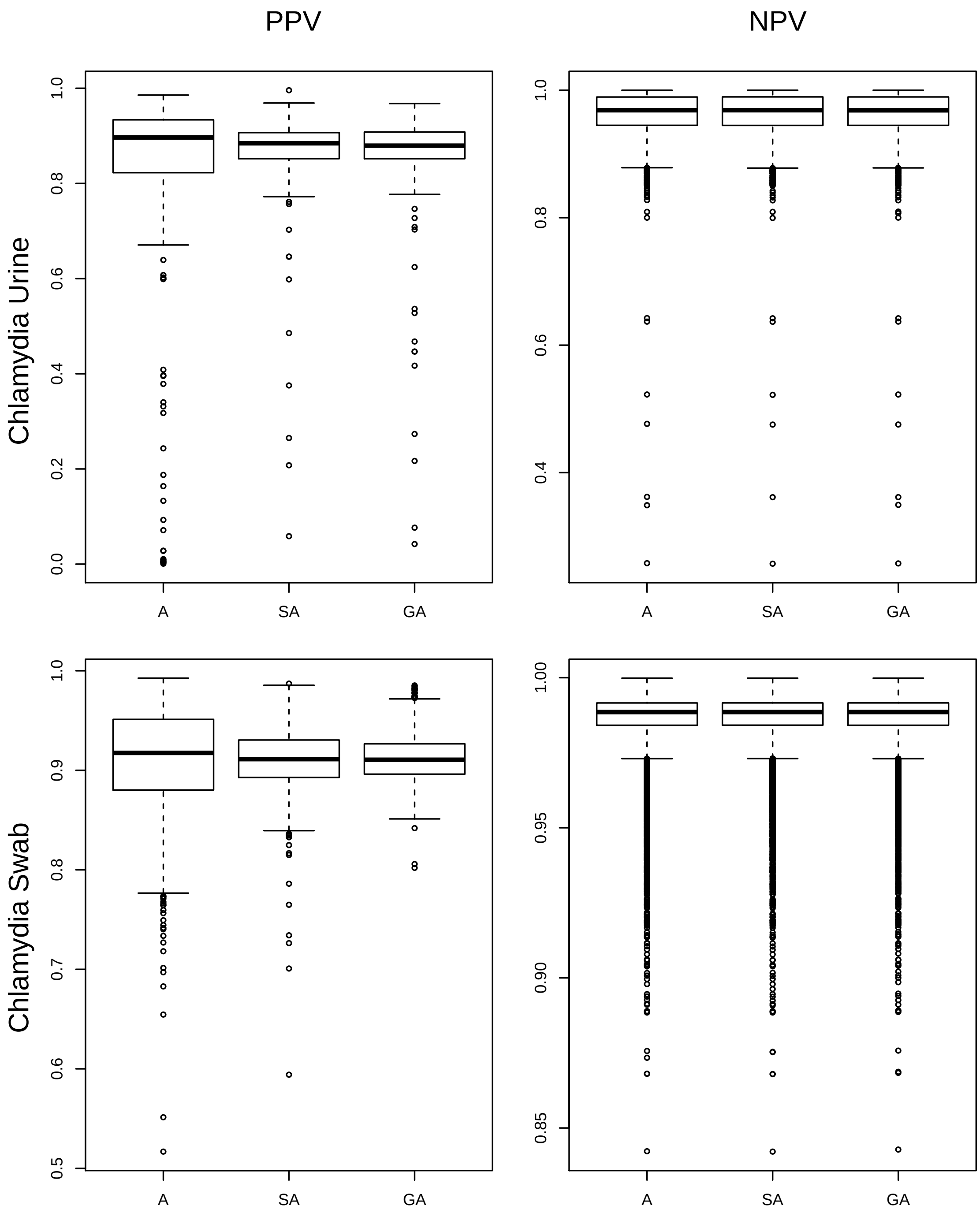

Figure 1: Nebraska IPP screening results for 2009. Estimated values of $PPV^{\mathcal{I}_{jk}}$ and $NPV^{\mathcal{I}_{jk}}$ for female subjects tested for chlamydia. Informative procedures GA and SA are implemented using estimates from a first-order logistic regression model with all covariates included; see Section 5.

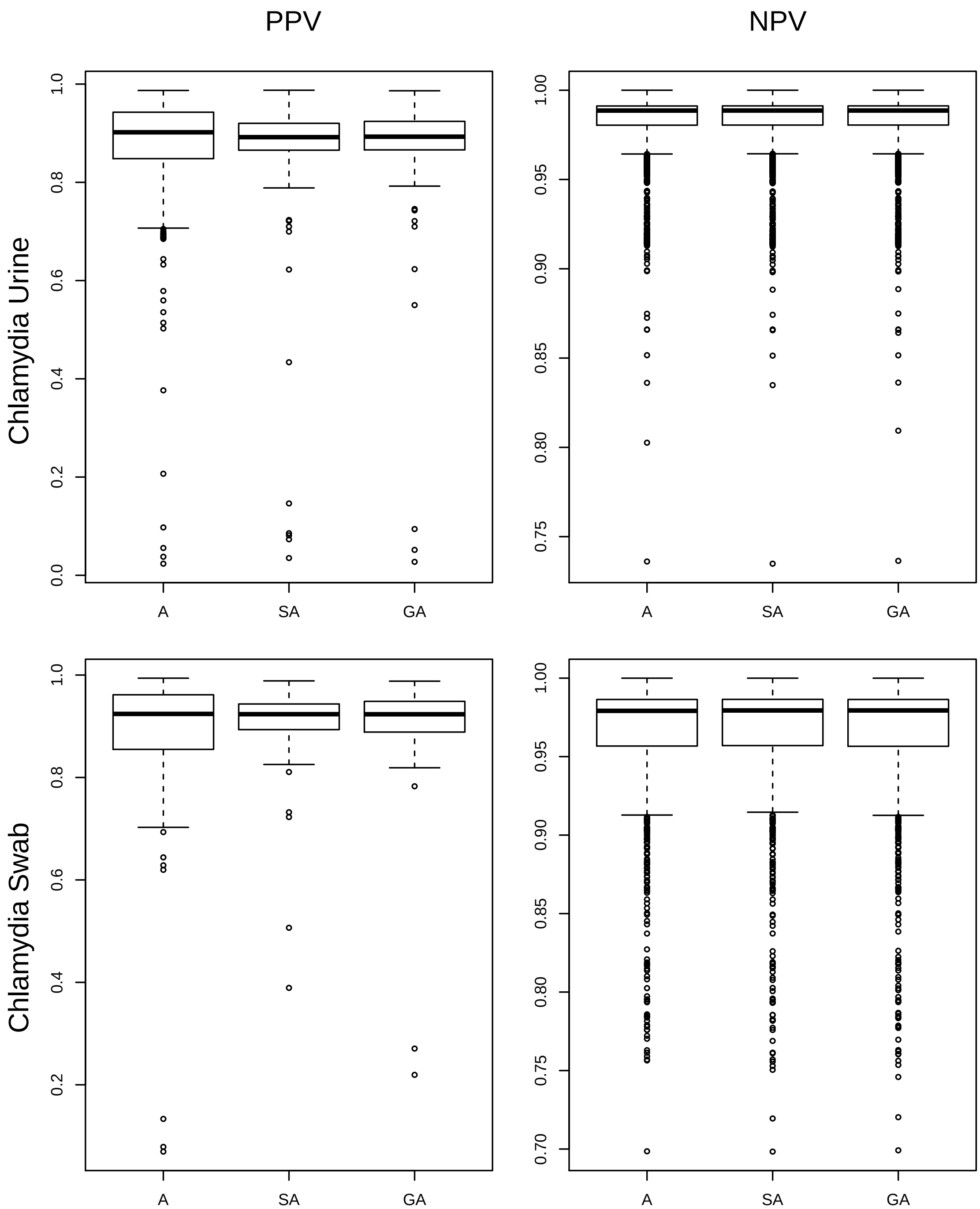

Figure 2: Nebraska IPP screening results for 2009. Estimated values of $PPV^{\mathcal{I}_{jk}}$ and $NPV^{\mathcal{I}_{jk}}$ for male subjects tested for chlamydia. Informative procedures GA and SA are implemented using estimates from a first-order logistic regression model with all covariates included; see Section 5.

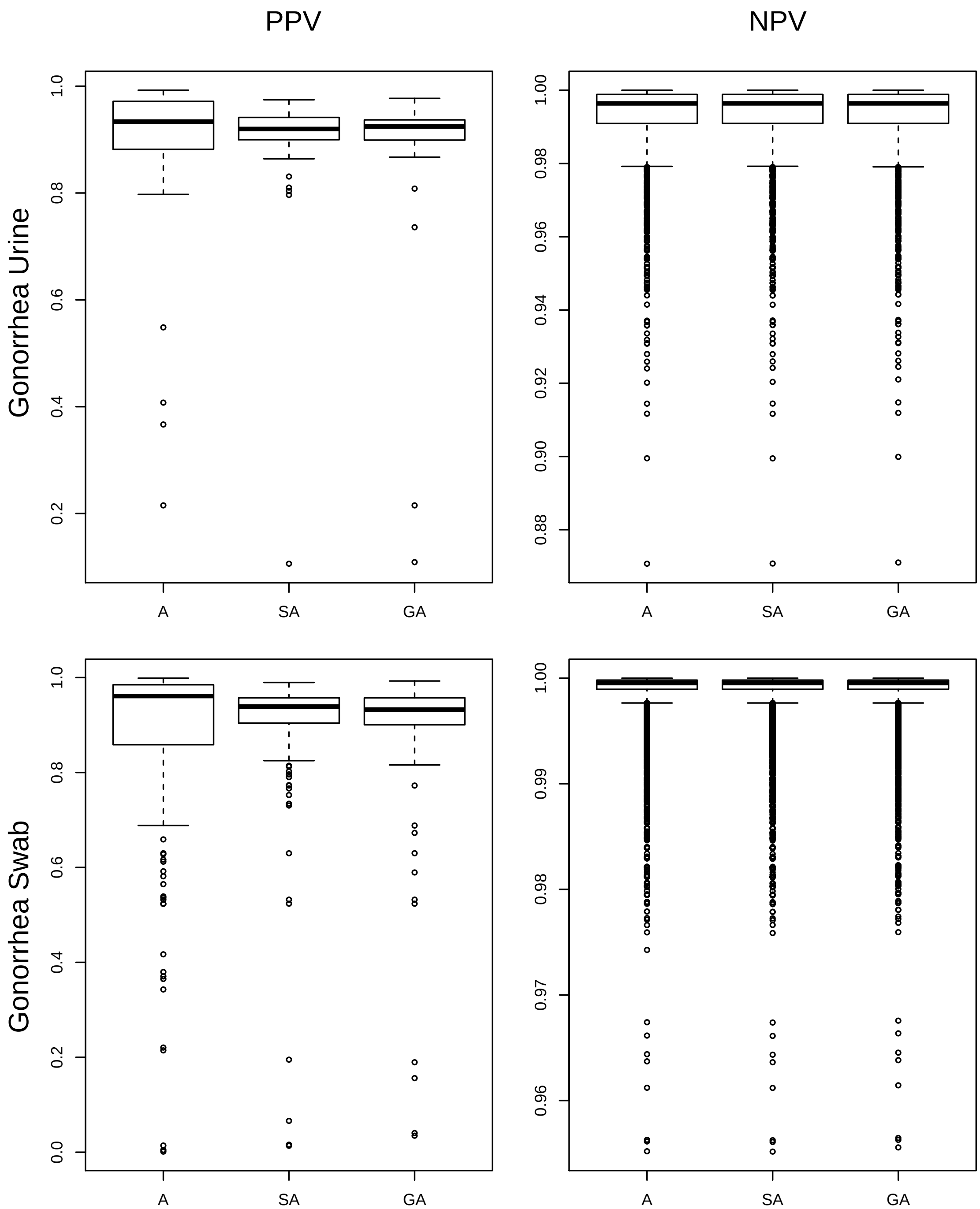

Figure 3: Nebraska IPP screening results for 2009. Estimated values of $PPV^{\mathcal{I}_{jk}}$ and $NPV^{\mathcal{I}_{jk}}$ for female subjects tested for gonorrhea. Informative procedures GA and SA are implemented using estimates from a first-order logistic regression model with all covariates included; see Section 5.

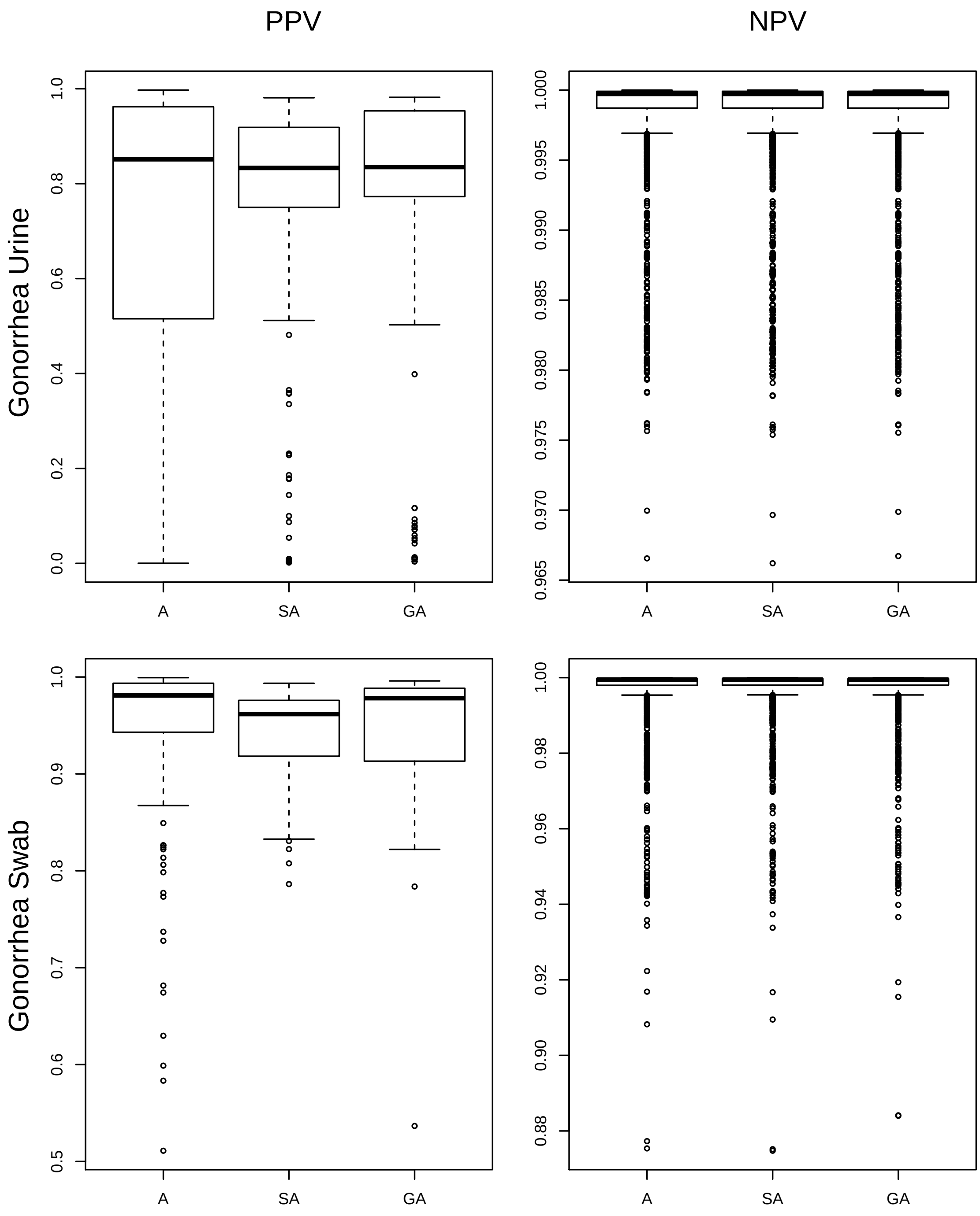

Figure 4: Nebraska IPP screening results for 2009. Estimated values of $PPV^{\mathcal{I}_{jk}}$ and $NPV^{\mathcal{I}_{jk}}$ for male subjects tested for gonorrhea. Informative procedures GA and SA are implemented using estimates from a first-order logistic regression model with all covariates included; see Section 5.

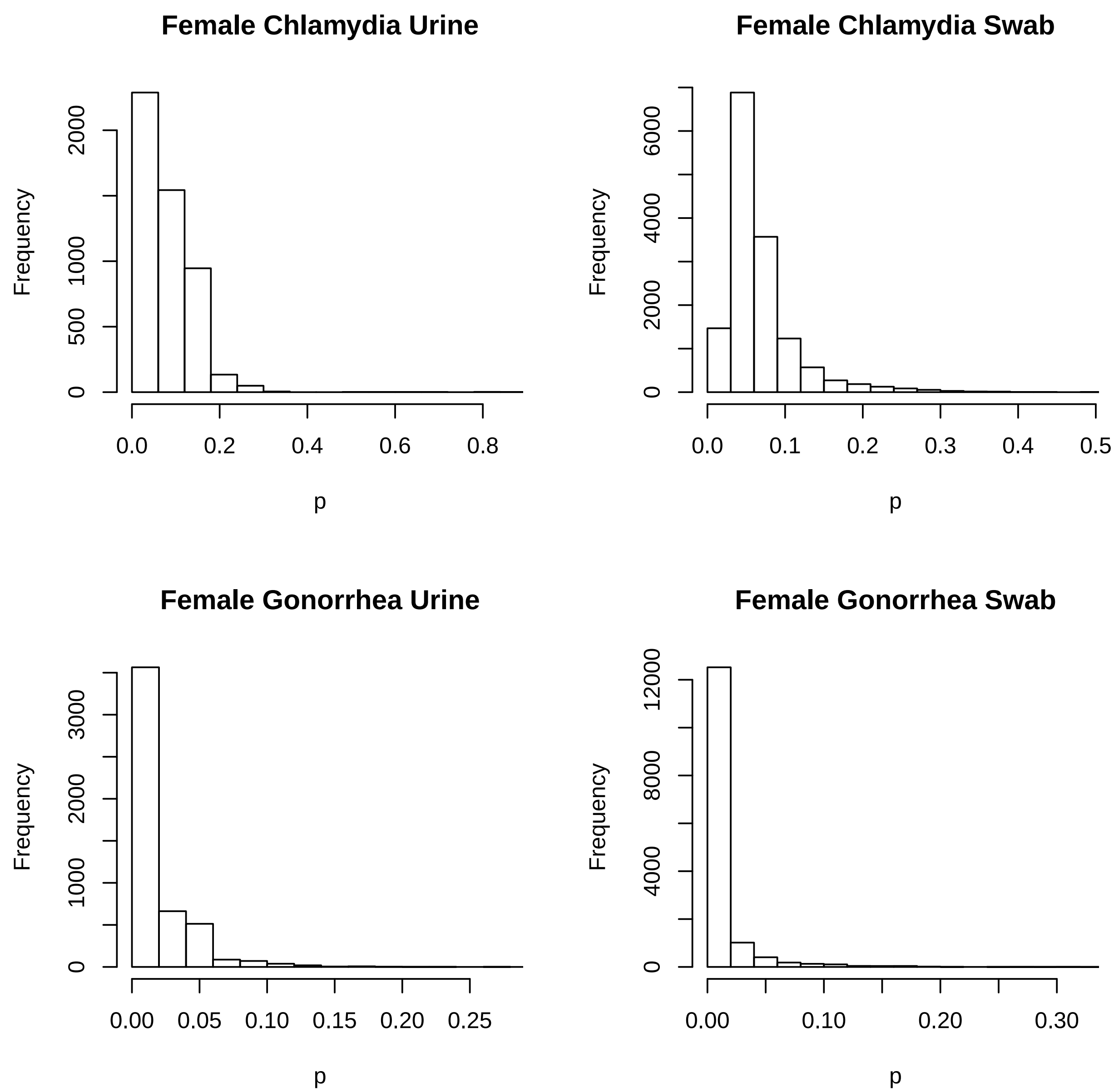

Figure 5: Nebraska IPP data. Histograms of the estimated probabilities $\widehat{p}_i$ from 2009. Female subjects. Estimates are from the first-order logistic regression model fit using all covariates.

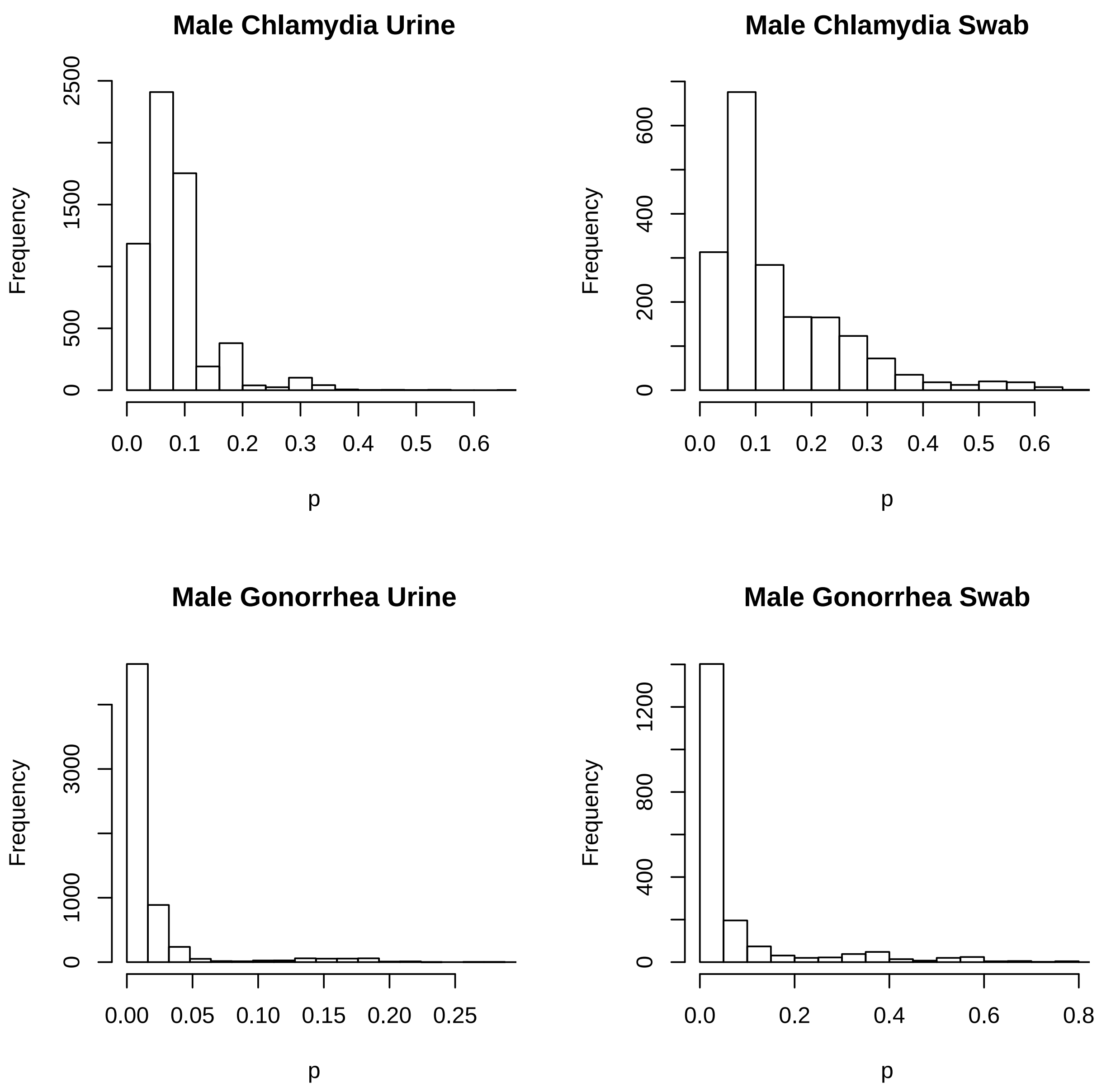

Figure 6: Nebraska IPP data. Histograms of the estimated probabilities $\widehat{p}_i$ from 2009. Male subjects. Estimates are from the first-order logistic regression model fit using all covariates.