Web-based Supplementary Materials for "Informative Dorfman Screening"

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, NE 68583, U.S.A.

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

Web Appendix A: *Derivation of* $E(T)$ *and* $\text{var}(T)$ *for ID.* Let $\widetilde{G}_j = 1$, if the jth pool is truly positive; i.e., $\mathcal{P}_j$ contains at least one positive individual, $\widetilde{G}_j = 0$, otherwise, for $j = 1, 2, ..., J$. Because the individual statuses are independent,

$$\text{pr}(\widetilde{G}_j = 1) = 1 - \prod_{k=1}^{c_j}(1 - p_{j(k)})$$

and $\text{pr}(\widetilde{G}_j = 0) = 1 - \text{pr}(\widetilde{G}_j = 1)$. Let $G_j = 1$, if $\mathcal{P}_j$ tests positive, $G_j = 0$, otherwise. By the Law of Total Probability,

$$\begin{aligned}
\text{pr}(G_j = 1) &= \text{pr}(G_j = 1|\widetilde{G}_j = 1)\text{pr}(\widetilde{G}_j = 1) + \text{pr}(G_j = 1|\widetilde{G}_j = 0)\text{pr}(\widetilde{G}_j = 0) \\
&= S_e\text{pr}(\widetilde{G}_j = 1) + (1 - S_p)\text{pr}(\widetilde{G}_j = 0) \\
&= S_e + (1 - S_e - S_p)\text{pr}(\widetilde{G}_j = 0) \\
&= S_e + (1 - S_e - S_p)\prod_{k=1}^{c_j}(1 - p_{j(k)}),
\end{aligned}$$

where S_e and S_p denote the test sensitivity and test specificity, respectively. Let $T_{\mathcal{P}_j}$ denote the number of tests needed to decode $\mathcal{P}_j$. If $c_j > 1$, the support of $T_{\mathcal{P}_j}$ is $\{1, c_j + 1\}$, where $\text{pr}(T_{\mathcal{P}_j} = 1) = \text{pr}(G_j = 0)$, and $\text{pr}(T_{\mathcal{P}_j} = c_j + 1) = \text{pr}(G_j = 1)$. Note that if $c_j = 1$, then $\text{pr}(T_{\mathcal{P}_j} = 1) = 1$. Therefore,

$$\begin{aligned}
E(T) = \sum_{j=1}^{J} E(T_{\mathcal{P}_j}) &= \sum_{j=1}^{J}[I(c_j > 1)\{(c_j + 1)\text{pr}(G_j = 1) + \text{pr}(G_j = 0)\} + I(c_j = 1)] \\
&= \sum_{j=1}^{J}[I(c_j > 1)\{c_j\text{pr}(G_j = 1) + 1\} + I(c_j = 1)] \\
&= J + \sum_{j=1}^{J} c_j I(c_j > 1)\text{pr}(G_j = 1),
\end{aligned}$$

as claimed. The variance of $T_{\mathcal{P}_j}$ is $\text{var}(T_{\mathcal{P}_j}) = E(T_{\mathcal{P}_j}^2) - \{E(T_{\mathcal{P}_j})\}^2$, where

$$\begin{aligned} E(T_{\mathcal{P}_j}^2) &= I(c_j > 1)\{(c_j+1)^2 \text{pr}(G_j = 1) + \text{pr}(G_j = 0)\} + I(c_j = 1) \\ &= I(c_j > 1)\{(c_j^2 + 2c_j)\text{pr}(G_j = 1) + 1\} + I(c_j = 1) \end{aligned}$$

and

$$\begin{aligned} \{E(T_{\mathcal{P}_j})\}^2 &= [I(c_j > 1)\{(c_j+1)\text{pr}(G_j = 1) + \text{pr}(G_j = 0)\} + I(c_j = 1)]^2 \\ &= \{I(c_j > 1)c_j \text{pr}(G_j = 1) + 1 + I(c_j = 1)\}^2 \\ &= I(c_j > 1)\{c_j \text{pr}(G_j = 1) + 1\}^2 + I(c_j = 1). \end{aligned}$$

Therefore,

$$\begin{aligned} \text{var}(T_{\mathcal{P}_j}) &= I(c_j > 1)[(c_j^2 + 2c_j)\text{pr}(G_j = 1) + 1 - \{c_j \text{pr}(G_j = 1) + 1\}^2] \\ &= I(c_j > 1)[c_j^2 \text{pr}(G_j = 1) - c_j^2\{\text{pr}(G_j = 1)\}^2] \\ &= I(c_j > 1)c_j^2 \text{pr}(G_j = 1)\{1 - \text{pr}(G_j = 1)\} \\ &= c_j^2 I(c_j > 1)\text{pr}(G_j = 1)\text{pr}(G_j = 0). \end{aligned}$$

Since the J pools are independent, $\text{var}(T) = \sum_{j=1}^{J} \text{var}(T_{\mathcal{P}_j})$, and the result follows. □

Web Appendix B: *Informative Dorfman algorithms.* The notation from Section 3.1 (OD), Section 3.2 (TOD), and Section 3.3 (PSOD) is used. An R program that implement these algorithms is available from the authors.

OD Algorithm:

1. Order the N individuals as in ID.
2. Specify a maximum allowable pool size M.
3. Let $c_{opt} = 1$ and $E(T^{(c_{opt})}) = N$.
4. Let $c = \min\{c_{opt} + 1, M\}$, $J = \lceil N/c \rceil$, $c_J = N - c(J-1)$, and calculate

$$\begin{aligned} E(T^{(c)}) &= J + c\sum_{j=0}^{J-2}\left\{S_e + (1 - S_e - S_p)\prod_{k=1}^{c}(1 - p_{(jc+k)})\right\} \\ &\quad + c_J\left\{S_e + (1 - S_e - S_p)\prod_{k=1}^{c_J}(1 - p_{(\{J-1\}c+k)})\right\}. \end{aligned}$$

5. IF $E(T^{(c_{opt})}) \leq E(T^{(c)})$, return c_{opt} as the optimal pool size for OD.
   ELSE let $c_{opt} = c$, $E(T^{(c_{opt})}) = E(T^{(c)})$, and go to Step 4.

TOD Algorithm:

1. Order the N individuals as in ID.
2. Specify a common pool size $c_0 > 1$. Set $j = J$, $j^* = J + 1$, $p_{J+1(1)} = 1$, and $p^* = 0$.
3. IF $j \geq 1$, calculate $E(T_{\mathcal{P}_j}) = 1 + c_0 \text{pr}(G_j = 1)$.
   ELSE go to Step 5.
4. IF $E(T_{\mathcal{P}_j}) \geq c_0$, let $j^* = j$, $j = j - 1$, and go to Step 3.
   ELSE set $p^* = \frac{1}{2}(p_{j^*-1(c_0)} + p_{j^*(1)})$ and go to Step 5.
5. Return p^* as the thresholding value.

Proof of claim that $E(T_{\mathcal{P}_J}) \geq E(T_{\mathcal{P}_{J-1}}) \geq \cdots \geq E(T_{\mathcal{P}_1})$ *for TOD.* It suffices to show that $E(T_{\mathcal{P}_j}) \geq E(T_{\mathcal{P}_{j'}})$, for all $j > j'$, for ID whenever $c_j \geq c_{j'}$. This covers the equal pool size case c_0 specified in TOD and the case when $c_1 < c_0$. Assume that $S_e + S_p \geq 1$. This is not prohibitive since assay tests usually possess values of S_e and S_p both close to 1. For ID,

$$\begin{aligned} E(T_{\mathcal{P}_j}) &= 1 + c_j I(c_j > 1)\text{pr}(G_j = 1) \\ E(T_{\mathcal{P}_{j'}}) &= 1 + c_{j'} I(c_{j'} > 1)\text{pr}(G_{j'} = 1). \end{aligned}$$

If $c_j = 1$, the result is obvious. When $c_j > 1$, it suffices to show that $\text{pr}(G_j = 1) \geq \text{pr}(G_{j'} = 1)$. From Web Appendix A,

$$\begin{aligned} \text{pr}(G_j = 1) &= S_e - (S_e + S_p - 1) \prod_{k=1}^{c_j} (1 - p_{j(k)}) \\ \text{pr}(G_{j'} = 1) &= S_e - (S_e + S_p - 1) \prod_{k'=1}^{c_{j'}} (1 - p_{j'(k')}). \end{aligned}$$

Therefore, it suffices to show that $\prod_{k=1}^{c_j}(1 - p_{j(k)}) \leq \prod_{k'=1}^{c_{j'}}(1 - p_{j'(k')})$ since $S_e + S_p - 1 \geq 0$. For $j > j'$, $0 \leq p_{j'(k')} \leq p_{j(k)} \leq 1$, for all k, k', which implies $0 \leq 1 - p_{j(k)} \leq 1 - p_{j'(k')} \leq 1$. Therefore, for all $j > j'$, $\prod_{k=1}^{c_j}(1 - p_{j(k)}) \leq \prod_{k=1}^{c_{j'}}(1 - p_{j(k)}) \leq \prod_{k'=1}^{c_{j'}}(1 - p_{j'(k')})$. $\square$

PSOD Algorithm:

1. Order the N individuals as in ID.
2. Specify a maximum allowable pool size M. Let $j = 1$, $N_j^P = 0$, and $N_j^R = N$.
3. IF $N_j^P < N$, go to Step 4.

   ELSE END.
4. Let $c_j^* = 1$ and $E(T_{\mathcal{P}_j;c_j^*}) = 1$.
5. Let $c_j = \min\{c_j^* + 1, N_j^R, M\}$ and calculate

$$E(T_{\mathcal{P}_j;c_j}) = 1 + c_j \left\{ S_e + (1 - S_e - S_p) \prod_{k=1}^{c_j} (1 - p_{(N_j^P + k)}) \right\}.$$

6. IF $c_j^{*-1} E(T_{\mathcal{P}_j;c_j^*}) \leq c_j^{-1} E(T_{\mathcal{P}_j;c_j})$, return c_j^* as the optimal pool size for the jth pool.

   ELSE let $c_j^* = c_j$, $E(T_{\mathcal{P}_j;c_j^*}) = E(T_{\mathcal{P}_j;c_j})$, and go to Step 5.
7. Let $j = j + 1$, $N_j^P = N_{j-1}^P + c_{j-1}$, $N_j^R = N_{j-1}^R - c_{j-1}$, and go to Step 3.

Web Appendix C: *Simulation results for Section* 4. This appendix contains our complete simulation results comparing D, OD, TOD, PSOD, H, and A. In all tables and figures listed below, we use $N = 1000$ individuals.

METHODOLOGY: We define the pooling sensitivity (specificity), PS_e (PS_p), to be the probability that a randomly selected individual is classified as positive (negative) given that the individual is truly positive (negative). Similarly, we define the pooling positive (negative) predictive value, PPV (NPV), as the probability that a randomly selected individual is truly positive (negative) given that the individual is classified as positive (negative). For the simulations in Section 4, we calculate PS_e, PS_p, PPV, and NPV as measures of classification accuracy for the six procedures. For D and A, these are calculated using the relevant equations in Kim et al. (2007) under the assumption of a common prevalence p. For OD, TOD, PSOD, and H we calculate

$$\begin{aligned} PS_e &= \frac{1}{N}\sum_{j=1}^{J}\sum_{k=1}^{c_j} PS_e^{\mathcal{I}_{j(k)}} \\ PS_p &= \frac{1}{N}\sum_{j=1}^{J}\sum_{k=1}^{c_j} PS_p^{\mathcal{I}_{j(k)}} \\ PPV &= \frac{1}{N}\sum_{j=1}^{J}\sum_{k=1}^{c_j} PPV^{\mathcal{I}_{j(k)}} \\ NPV &= \frac{1}{N}\sum_{j=1}^{J}\sum_{k=1}^{c_j} NPV^{\mathcal{I}_{j(k)}}. \end{aligned}$$

In contrast to the individual-specific measures derived in Section 2.2, reporting PS_e, PS_p, PPV, and NPV in this way (for OD, TOD, PSOD, and H) allows us to characterize global performance.

SUMMARY: Tables 1-3 list values of PS_e, PS_p, PPV, and NPV for D, OD, TOD, PSOD, H, and A, when $M = 10$ and $N = 1000$, for different values of p, and $S_e = S_p$. Using larger values of M produced similar results. In summary, note that $PS_e = S_e^2$ for all Dorfman procedures when $p < 0.1$. When $p \geq 0.1$, PS_e can be larger than S_e^2 for TOD, PSOD, and H. This is caused by the propensity to specify that high risk subjects are to be screened individually. It is also worth noting that when $p \geq 0.3$, the noninformative D algorithm reverts to individual testing so that $PS_e = S_e$. In general, all Dorfman procedures tend to have identical or larger values of PS_e and NPV when compared to A. In contrast, A tends to have larger values of PS_p when compared to all Dorfman procedures, and D and A tend to produce larger values of PPV.

Table 1: Measures of accuracy for D, OD, TOD, PSOD, H, and A when $M = 10$, $N = 1000$, and $S_e = S_p = 0.90$.

Measures of Accuracy							
	p	D	OD	TOD	PSOD	H	A
PS_e	0.0001	0.810	0.810	0.810	0.810	0.810	0.791
	0.001	0.810	0.810	0.810	0.810	0.810	0.787
	0.01	0.810	0.810	0.810	0.810	0.810	0.757
	0.02	0.810	0.810	0.810	0.810	0.810	0.742
	0.05	0.810	0.810	0.810	0.810	0.810	0.730
	0.10	0.810	0.810	0.813	0.814	0.814	0.731
	0.20	0.810	0.810	0.826	0.832	0.832	0.729
	0.30	0.900	0.810	0.843	0.849	0.849	0.729
PS_p	0.0001	0.990	0.990	0.990	0.990	0.990	0.992
	0.001	0.989	0.989	0.989	0.989	0.989	0.992
	0.01	0.983	0.984	0.984	0.984	0.985	0.994
	0.02	0.978	0.979	0.979	0.981	0.982	0.993
	0.05	0.972	0.970	0.970	0.975	0.977	0.984
	0.10	0.968	0.965	0.965	0.966	0.969	0.977
	0.20	0.951	0.951	0.947	0.949	0.952	0.959
	0.30	0.900	0.934	0.935	0.936	0.938	0.925
PPV	0.0001	0.008	0.008	0.008	0.008	0.008	0.010
	0.001	0.070	0.062	0.062	0.062	0.062	0.094
	0.01	0.326	0.261	0.261	0.269	0.273	0.560
	0.02	0.430	0.361	0.361	0.376	0.385	0.688
	0.05	0.603	0.508	0.508	0.525	0.540	0.709
	0.10	0.740	0.633	0.633	0.627	0.644	0.781
	0.20	0.805	0.726	0.724	0.728	0.741	0.817
	0.30	0.794	0.754	0.777	0.781	0.791	0.806
NPV	0.0001	1.000	1.000	1.000	1.000	1.000	1.000
	0.001	1.000	1.000	1.000	1.000	1.000	1.000
	0.01	0.998	0.998	0.998	0.998	0.998	0.998
	0.02	0.996	0.996	0.996	0.996	0.996	0.995
	0.05	0.990	0.989	0.989	0.989	0.989	0.986
	0.10	0.979	0.976	0.978	0.979	0.979	0.970
	0.20	0.952	0.945	0.956	0.959	0.959	0.934
	0.30	0.955	0.894	0.926	0.929	0.929	0.888

Table 2: Measures of accuracy for D, OD, TOD, PSOD, H, and A when $M = 10$, $N = 1000$, and $S_e = S_p = 0.95$.

Measures of Accuracy							
	p	D	OD	TOD	PSOD	H	A
PS_e	0.0001	0.903	0.902	0.902	0.902	0.902	0.914
	0.001	0.903	0.902	0.902	0.902	0.902	0.910
	0.01	0.903	0.902	0.902	0.902	0.902	0.882
	0.02	0.903	0.903	0.903	0.903	0.903	0.868
	0.05	0.903	0.903	0.903	0.903	0.903	0.858
	0.10	0.903	0.902	0.905	0.905	0.905	0.858
	0.20	0.903	0.903	0.914	0.915	0.915	0.859
	0.30	0.950	0.902	0.924	0.924	0.924	0.948
PS_p	0.0001	0.997	0.997	0.997	0.997	0.997	0.997
	0.001	0.997	0.997	0.997	0.997	0.997	0.997
	0.01	0.994	0.994	0.994	0.994	0.994	0.998
	0.02	0.992	0.991	0.991	0.993	0.993	0.997
	0.05	0.989	0.989	0.989	0.990	0.990	0.993
	0.10	0.985	0.983	0.983	0.985	0.985	0.989
	0.20	0.981	0.978	0.975	0.976	0.976	0.983
	0.30	0.950	0.969	0.968	0.969	0.969	0.995
PPV	0.0001	0.034	0.034	0.034	0.034	0.034	0.029
	0.001	0.237	0.197	0.197	0.197	0.197	0.238
	0.01	0.588	0.492	0.492	0.504	0.506	0.809
	0.02	0.686	0.590	0.590	0.605	0.609	0.875
	0.05	0.814	0.723	0.723	0.713	0.718	0.860
	0.10	0.872	0.811	0.810	0.805	0.810	0.895
	0.20	0.923	0.878	0.875	0.866	0.869	0.927
	0.30	0.891	0.892	0.901	0.897	0.898	0.988
NPV	0.0001	1.000	1.000	1.000	1.000	1.000	1.000
	0.001	1.000	1.000	1.000	1.000	1.000	1.000
	0.01	0.999	0.999	0.999	0.999	0.999	0.999
	0.02	0.998	0.998	0.998	0.998	0.998	0.997
	0.05	0.995	0.995	0.995	0.995	0.995	0.993
	0.10	0.989	0.987	0.989	0.989	0.989	0.984
	0.20	0.976	0.968	0.977	0.977	0.977	0.965
	0.30	0.978	0.937	0.960	0.960	0.960	0.978

Table 3: Measures of accuracy for D, OD, TOD, PSOD, H, and A when $M = 10$, $N = 1000$, and $S_e = S_p = 0.99$.

Measures of Accuracy							
	p	D	OD	TOD	PSOD	H	A
PS_e	0.0001	0.980	0.980	0.980	0.980	0.980	0.988
	0.001	0.980	0.980	0.980	0.980	0.980	0.987
	0.01	0.980	0.980	0.980	0.980	0.980	0.978
	0.02	0.980	0.980	0.980	0.980	0.980	0.973
	0.05	0.980	0.980	0.980	0.980	0.980	0.970
	0.10	0.980	0.980	0.980	0.980	0.980	0.971
	0.20	0.980	0.980	0.983	0.983	0.983	0.971
	0.30	0.990	0.980	0.985	0.984	0.984	0.990
PS_p	0.0001	1.000	1.000	1.000	1.000	1.000	1.000
	0.001	1.000	1.000	1.000	1.000	1.000	1.000
	0.01	0.999	0.999	0.999	0.999	0.999	1.000
	0.02	0.999	0.999	0.999	0.999	0.999	1.000
	0.05	0.998	0.998	0.998	0.998	0.998	0.999
	0.10	0.997	0.997	0.997	0.998	0.998	0.998
	0.20	0.996	0.996	0.995	0.995	0.995	0.997
	0.30	0.990	0.994	0.994	0.994	0.994	1.000
PPV	0.0001	0.474	0.379	0.379	0.379	0.379	0.354
	0.001	0.839	0.734	0.734	0.734	0.734	0.854
	0.01	0.913	0.886	0.886	0.891	0.892	0.984
	0.02	0.935	0.921	0.921	0.919	0.919	0.984
	0.05	0.964	0.948	0.948	0.942	0.942	0.974
	0.10	0.975	0.963	0.963	0.958	0.959	0.980
	0.20	0.985	0.977	0.976	0.972	0.972	0.986
	0.30	0.977	0.978	0.981	0.979	0.979	1.000
NPV	0.0001	1.000	1.000	1.000	1.000	1.000	1.000
	0.001	1.000	1.000	1.000	1.000	1.000	1.000
	0.01	1.000	1.000	1.000	1.000	1.000	1.000
	0.02	1.000	1.000	1.000	1.000	1.000	0.999
	0.05	0.999	0.999	0.999	0.999	0.999	0.998
	0.10	0.998	0.998	0.998	0.998	0.998	0.997
	0.20	0.995	0.993	0.996	0.995	0.995	0.993
	0.30	0.996	0.987	0.992	0.992	0.992	0.996

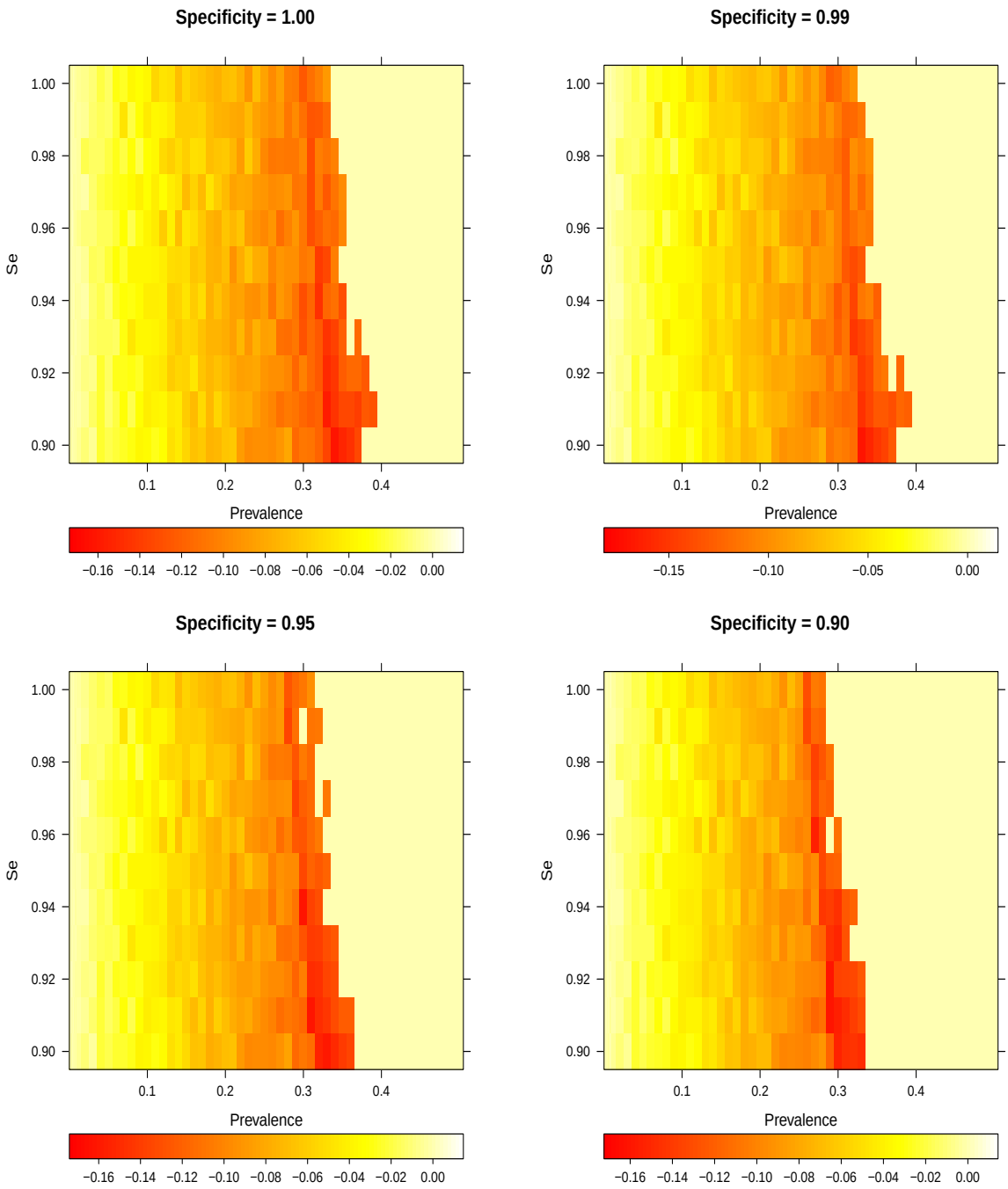

Figure 1: Comparison of OD and D, when $M = 10$ and $N = 1000$, using different levels of S_e, S_p, and p. Each figure displays values of $\{E(T|\text{OD}) - E(T|\text{D})\}/N$, the difference in expected individual expenditure between OD and D. Negative values mean that OD is more efficient; the optimal size for D has been used throughout.

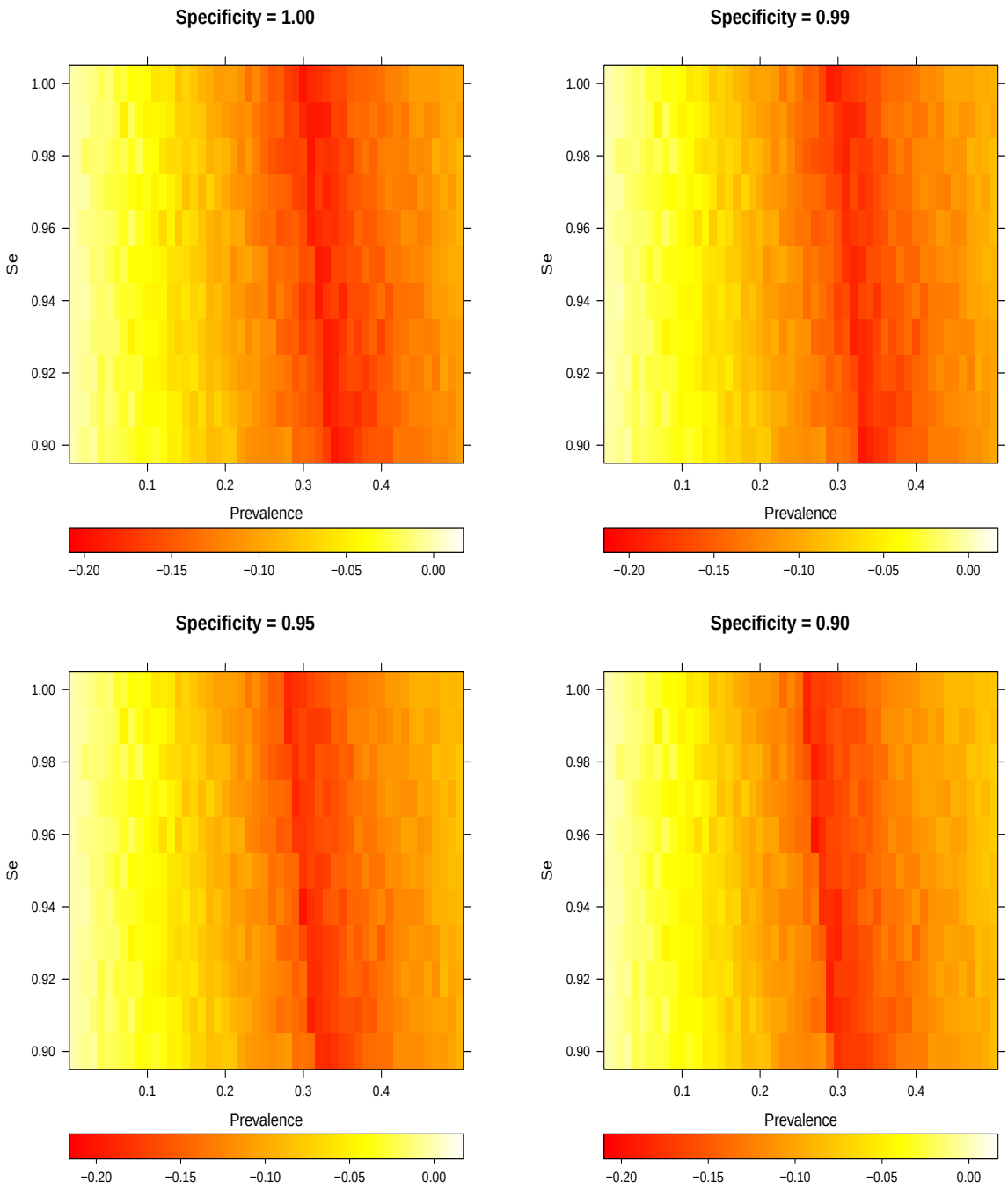

Figure 2: Comparison of TOD and D, when $M = 10$ and $N = 1000$, using different levels of S_e, S_p, and p. Each figure displays values of $\{E(T|\text{TOD}) - E(T|\text{D})\}/N$, the difference in expected individual expenditure between TOD and D. Negative values mean that TOD is more efficient; the optimal size for D has been used throughout.

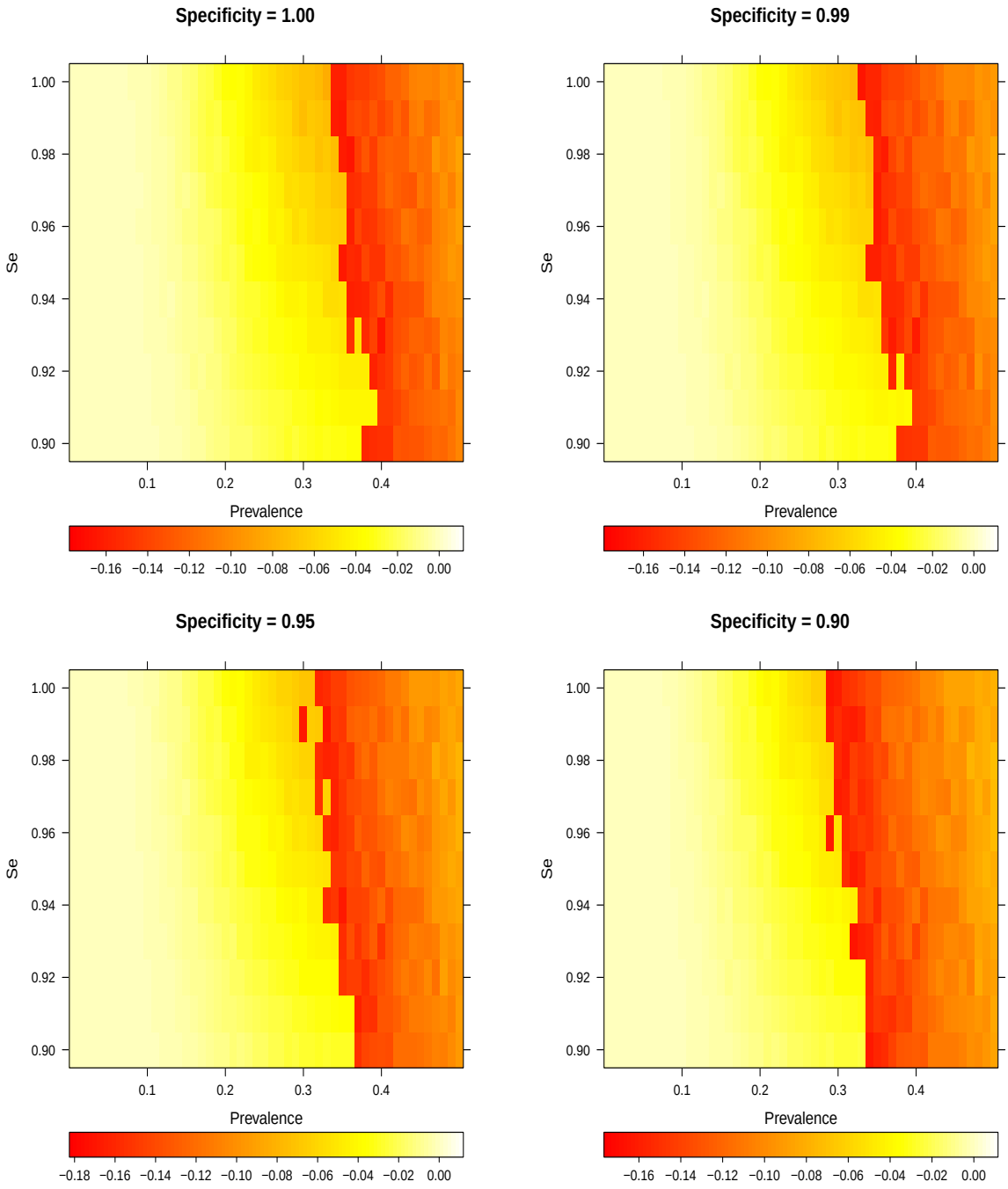

Figure 3: Comparison of TOD and OD, when $M = 10$ and $N = 1000$, using different levels of S_e, S_p, and p. Each figure displays values of $\{E(T|\text{TOD}) - E(T|\text{OD})\}/N$, the difference in expected individual expenditure between TOD and OD. Negative values mean that TOD is more efficient.

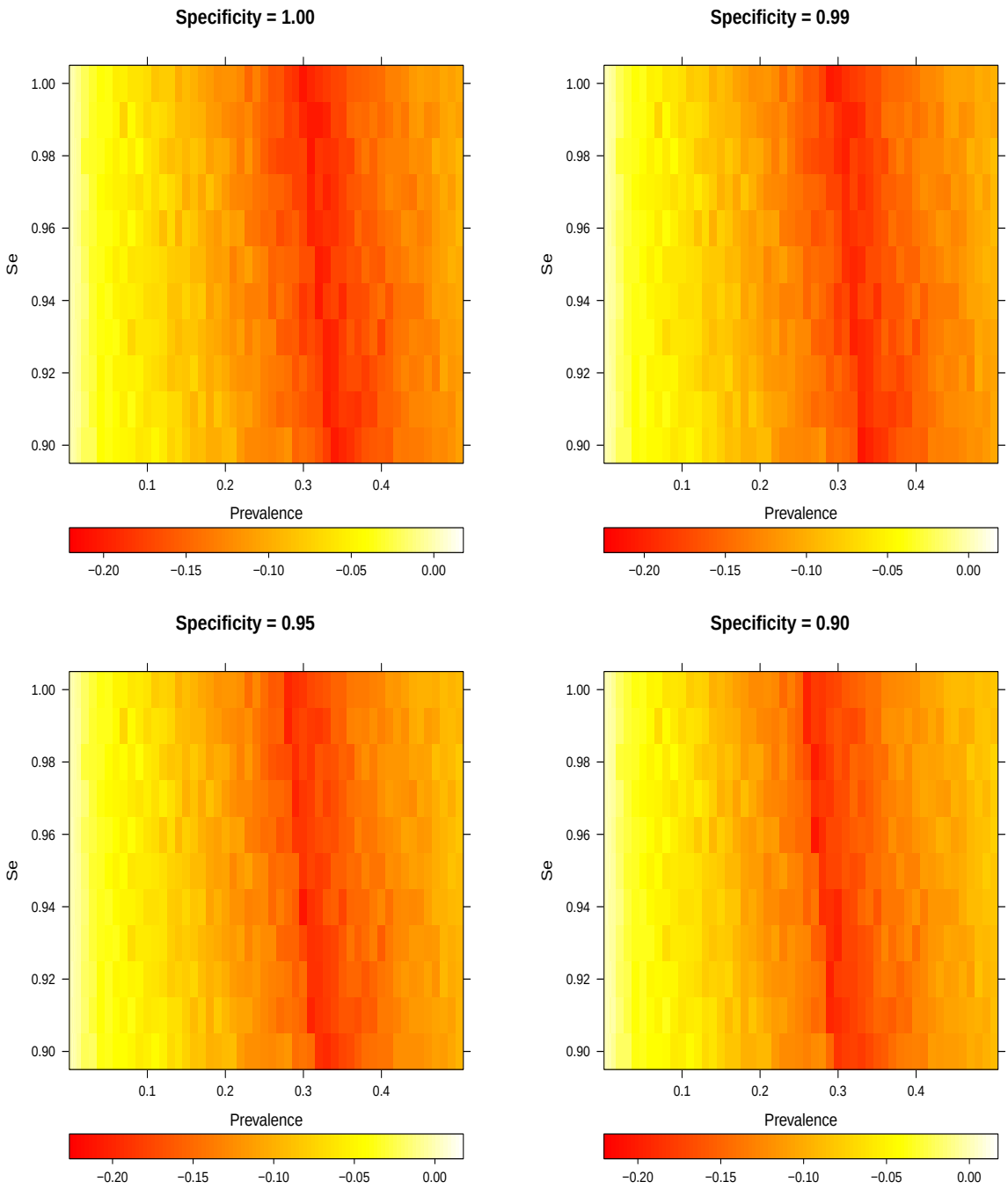

Figure 4: Comparison of PSOD and D, when $M = 10$ and $N = 1000$, using different levels of S_e, S_p, and p. Each figure displays values of $\{E(T|\text{PSOD}) - E(T|\text{D})\}/N$, the difference in expected individual expenditure between PSOD and D. Negative values mean that PSOD is more efficient; the optimal size for D has been used throughout.

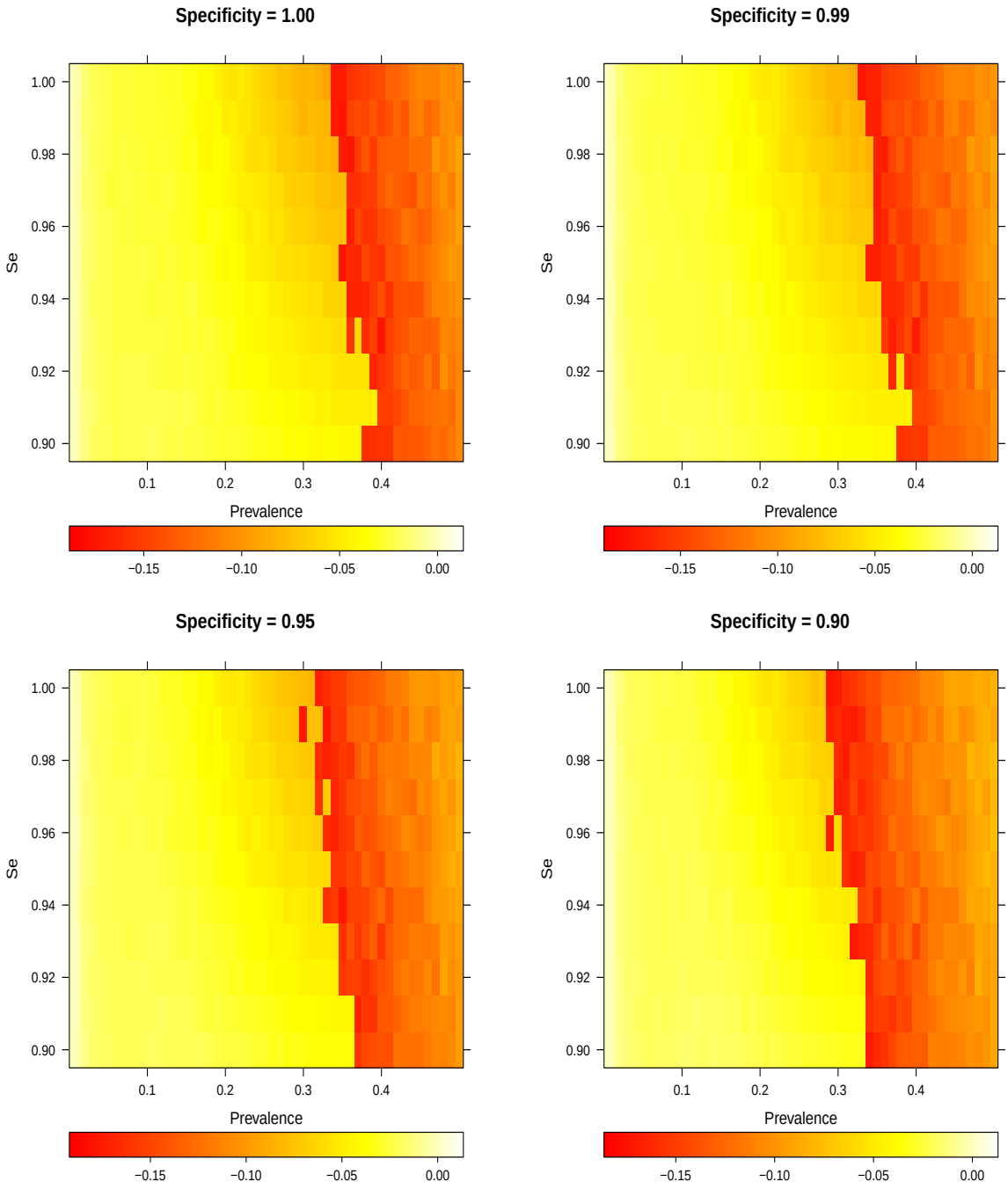

Figure 5: Comparison of PSOD and OD, when $M = 10$ and $N = 1000$, using different levels of S_e, S_p, and p. Each figure displays values of $\{E(T|\text{PSOD}) - E(T|\text{OD})\}/N$, the difference in expected individual expenditure between PSOD and OD. Negative values mean that PSOD is more efficient.

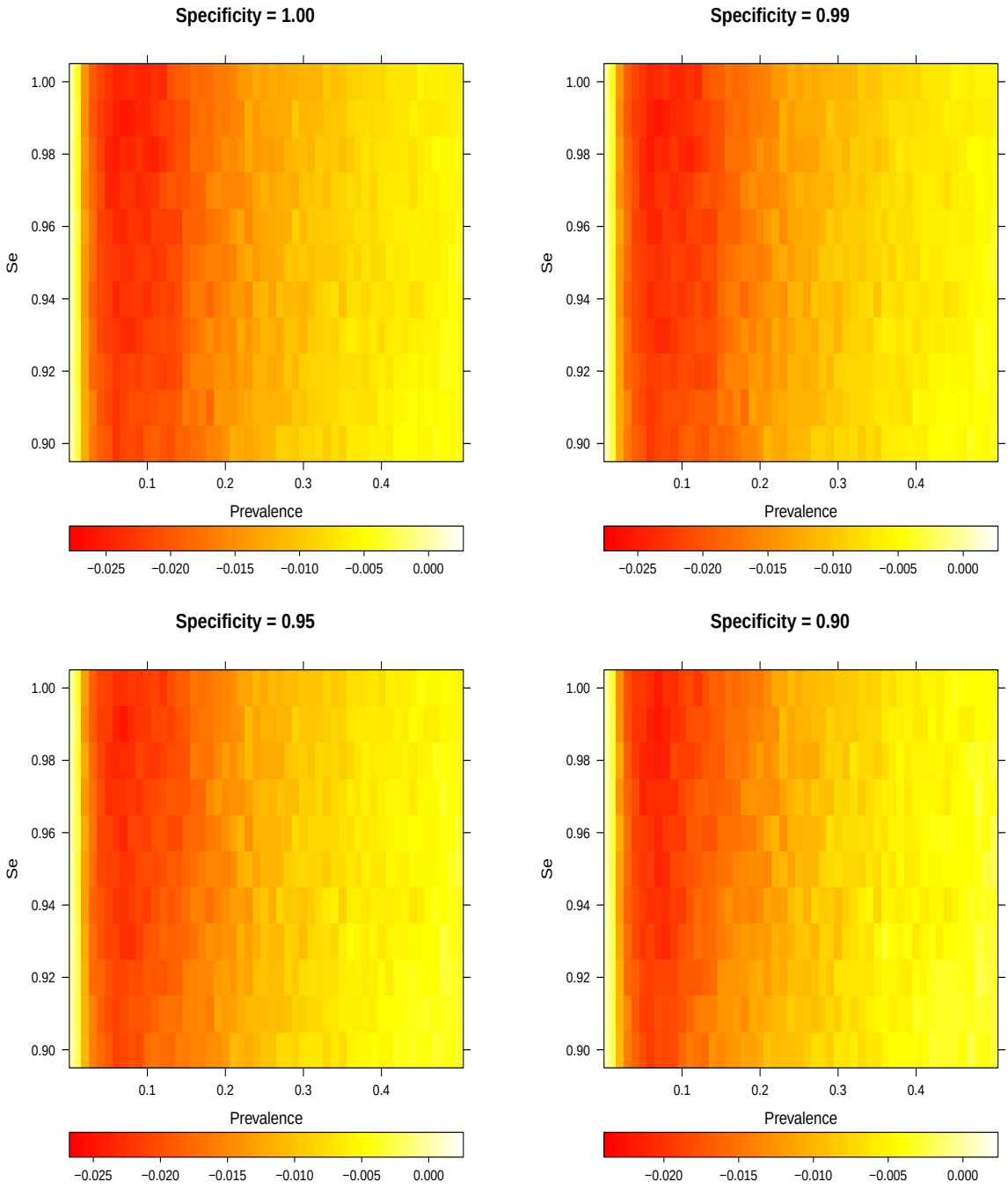

Figure 6: Comparison of PSOD and TOD, when $M = 10$ and $N = 1000$, using different levels of S_e, S_p, and p. Each figure displays values of $\{E(T|\text{PSOD}) - E(T|\text{TOD})\}/N$, the difference in expected individual expenditure between PSOD and TOD. Negative values mean that PSOD is more efficient.

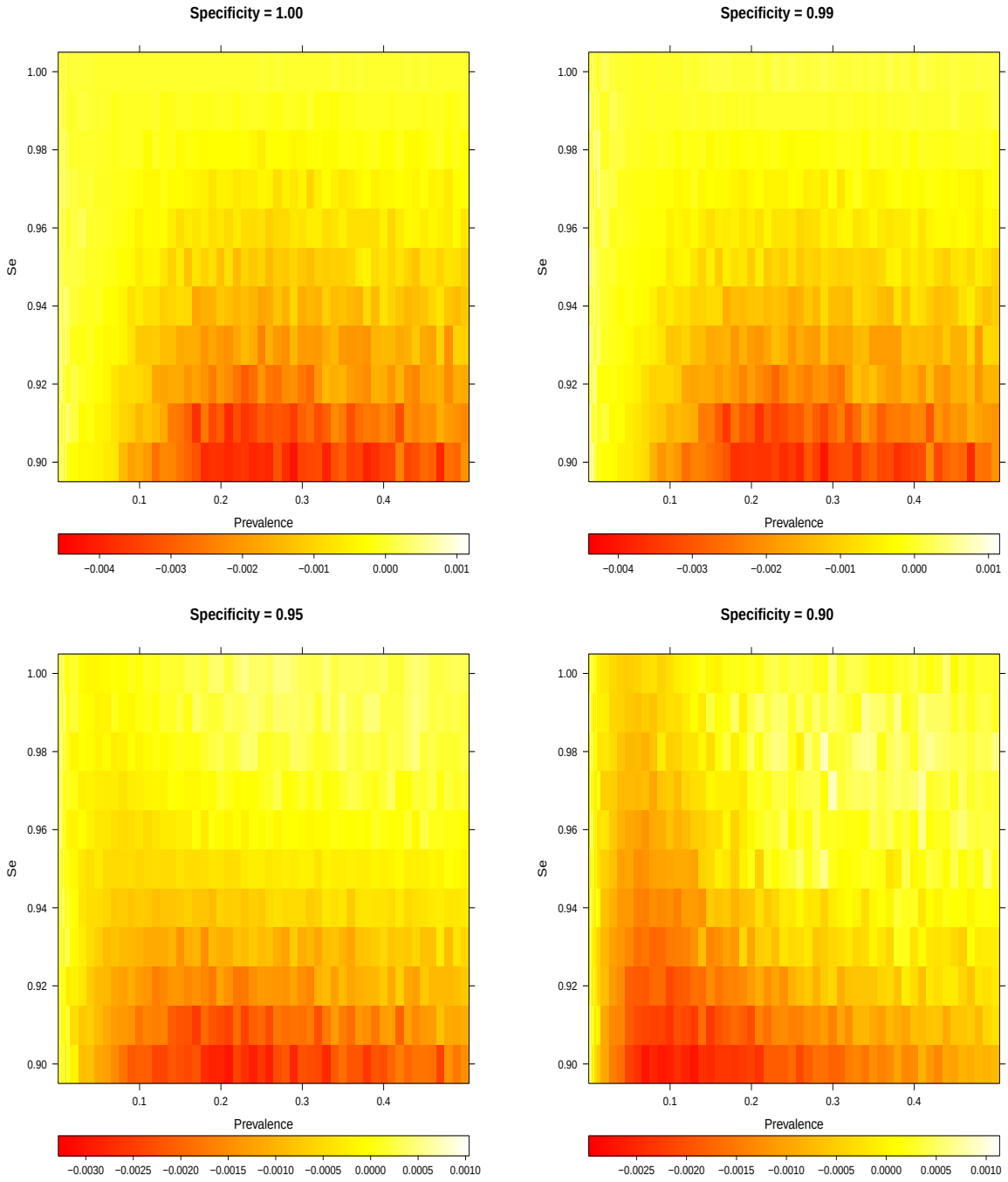

Figure 7: Comparison of PSOD and H, when $M = 10$ and $N = 1000$, using different levels of S_e, S_p, and p. Each figure displays values of $\{E(T|\text{PSOD}) - E(T|\text{H})\}/N$, the difference in expected individual expenditure between PSOD and H. Negative values mean that PSOD is more efficient.

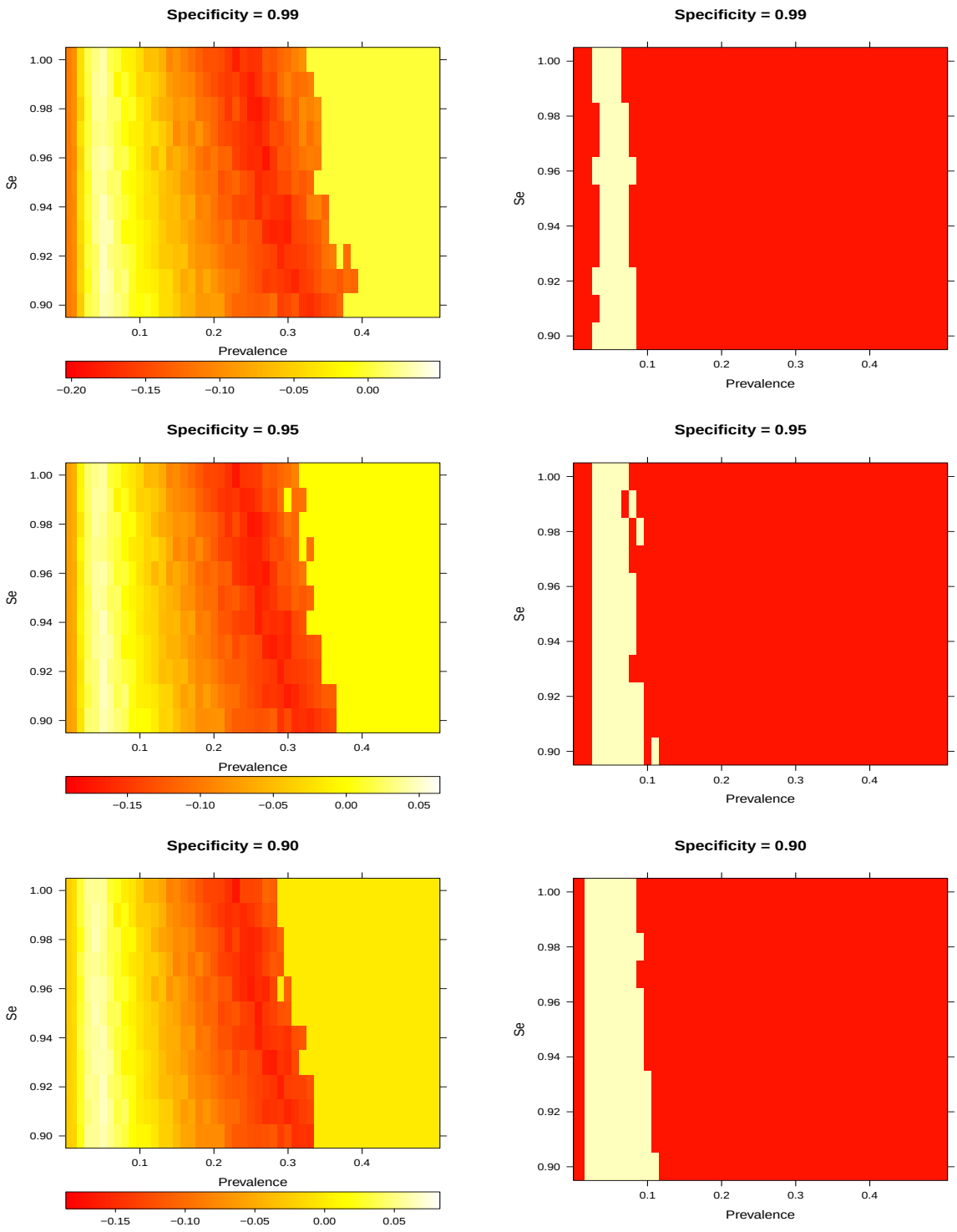

Figure 8: Comparison of OD and A, when $M = 10$ and $N = 1000$, using different levels of S_e, S_p, and p. Left: Values of $\{E(T|\text{OD}) - E(T|\text{A})\}/N$, the difference in expected individual expenditure between OD and A. Negative values mean that OD is more efficient. Right: Binary versions of the corresponding figure on the left; regions in red indicate those settings where OD is more efficient. The optimal array size for A has been used throughout.

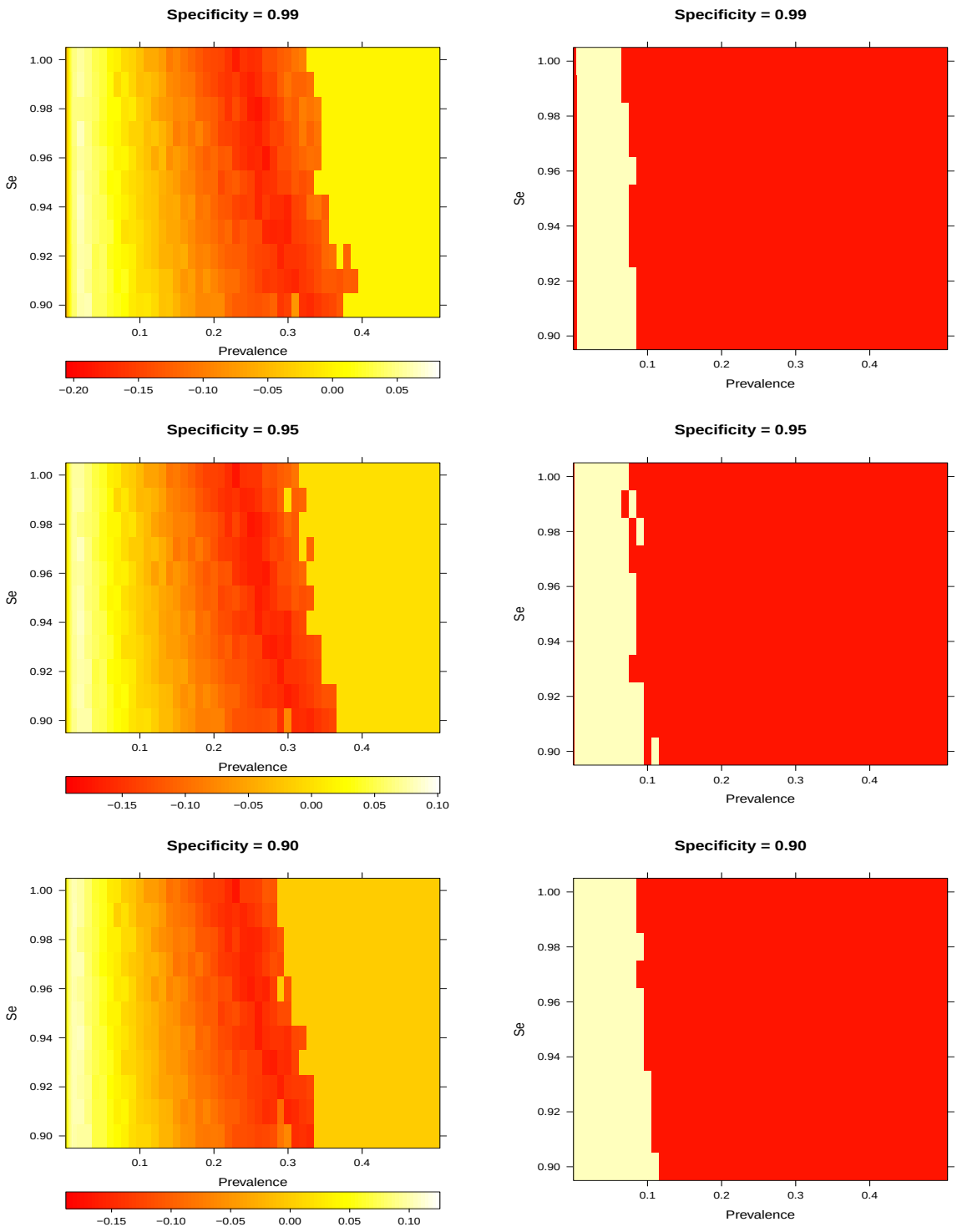

Figure 9: Comparison of OD and A, when $M = 20$ and $N = 1000$, using different levels of S_e, S_p, and p. Left: Values of $\{E(T|\text{OD}) - E(T|\text{A})\}/N$, the difference in expected individual expenditure between OD and A. Negative values mean that OD is more efficient. Right: Binary versions of the corresponding figure on the left; regions in red indicate those settings where OD is more efficient. The optimal array size for A has been used throughout.

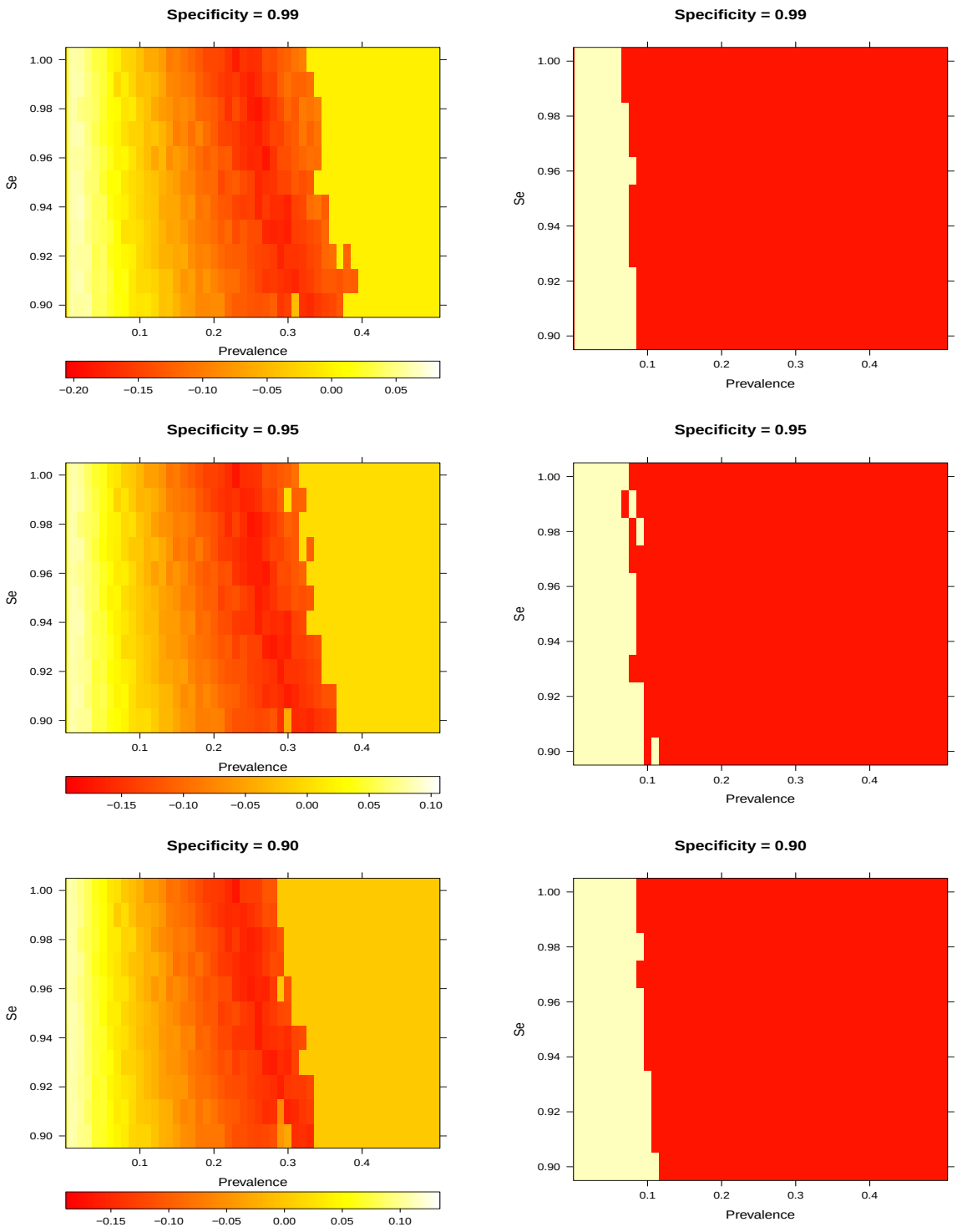

Figure 10: Comparison of OD and A, when $M = 30$ and $N = 1000$, using different levels of S_e, S_p, and p. Left: Values of $\{E(T|\text{OD}) - E(T|\text{A})\}/N$, the difference in expected individual expenditure between OD and A. Negative values mean that OD is more efficient. Right: Binary versions of the corresponding figure on the left; regions in red indicate those settings where OD is more efficient. The optimal array size for A has been used throughout.

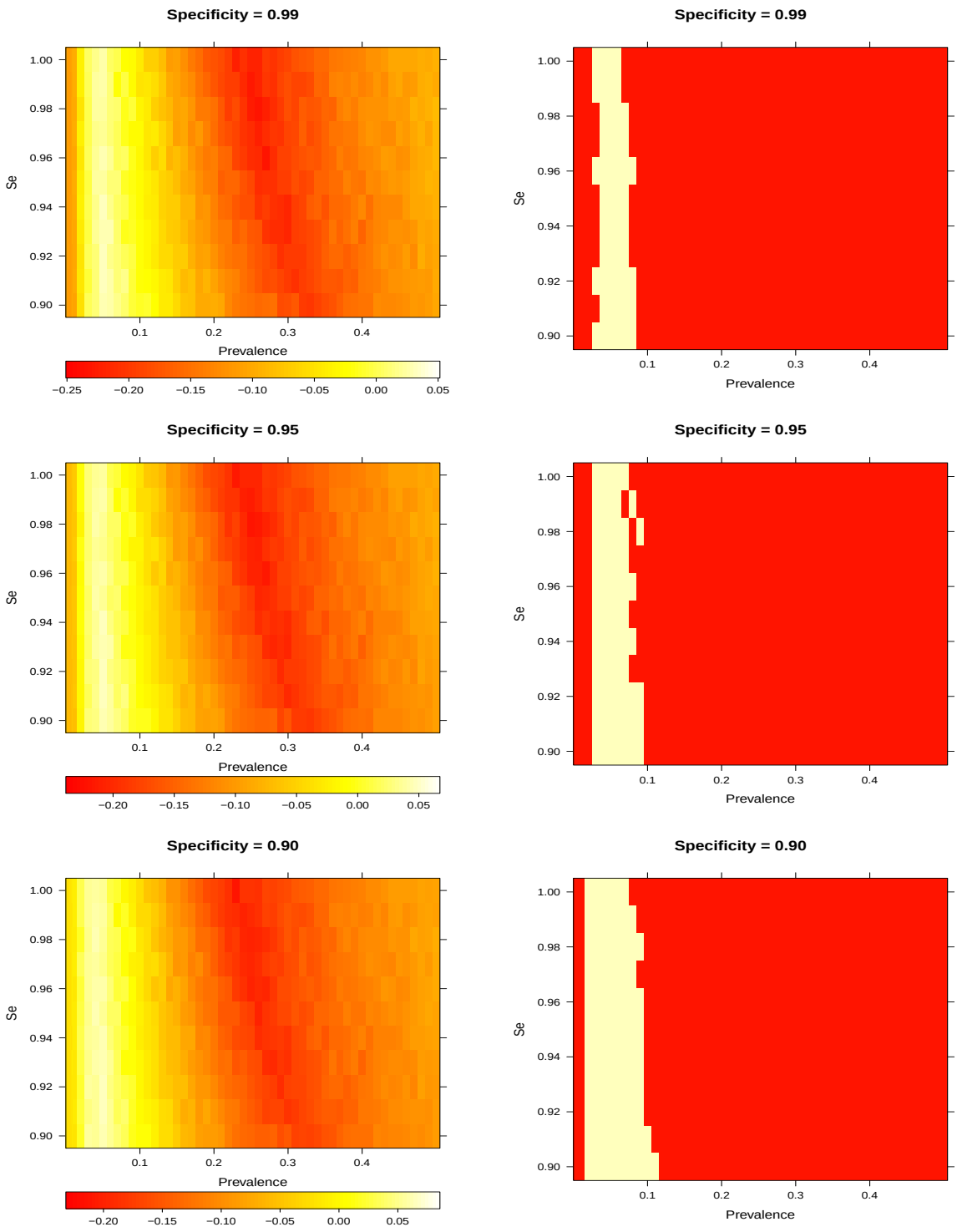

Figure 11: Comparison of TOD and A, when $M = 10$ and $N = 1000$, using different levels of S_e, S_p, and p. Left: Values of $\{E(T|\text{TOD}) - E(T|\text{A})\}/N$, the difference in expected individual expenditure between TOD and A. Negative values mean that TOD is more efficient. Right: Binary versions of the corresponding figure on the left; regions in red indicate those settings where TOD is more efficient. The optimal array size for A has been used throughout.

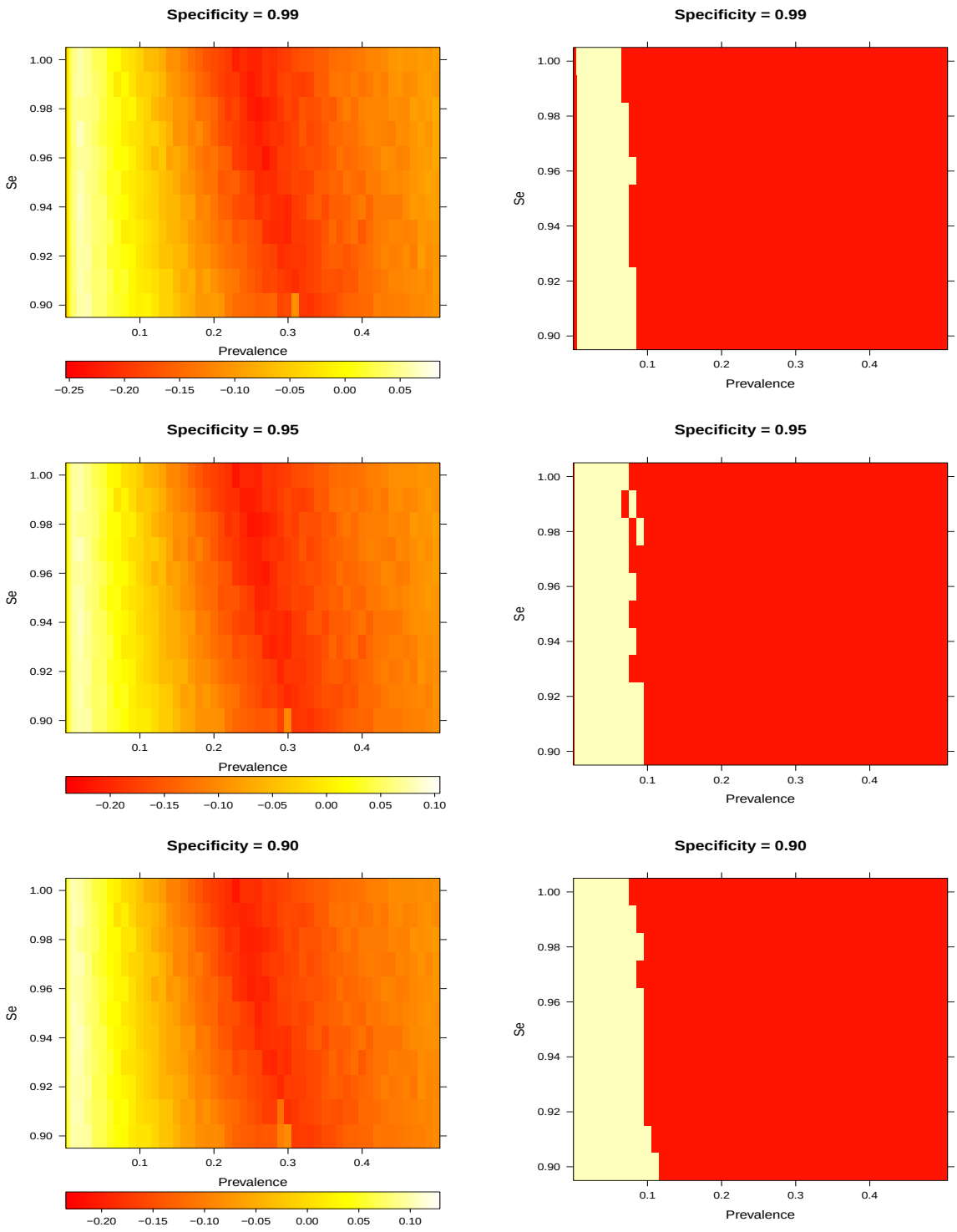

Figure 12: Comparison of TOD and A, when $M = 20$ and $N = 1000$, using different levels of S_e, S_p, and p. Left: Values of $\{E(T|\text{TOD}) - E(T|\text{A})\}/N$, the difference in expected individual expenditure between TOD and A. Negative values mean that TOD is more efficient. Right: Binary versions of the corresponding figure on the left; regions in red indicate those settings where TOD is more efficient. The optimal array size for A has been used throughout.

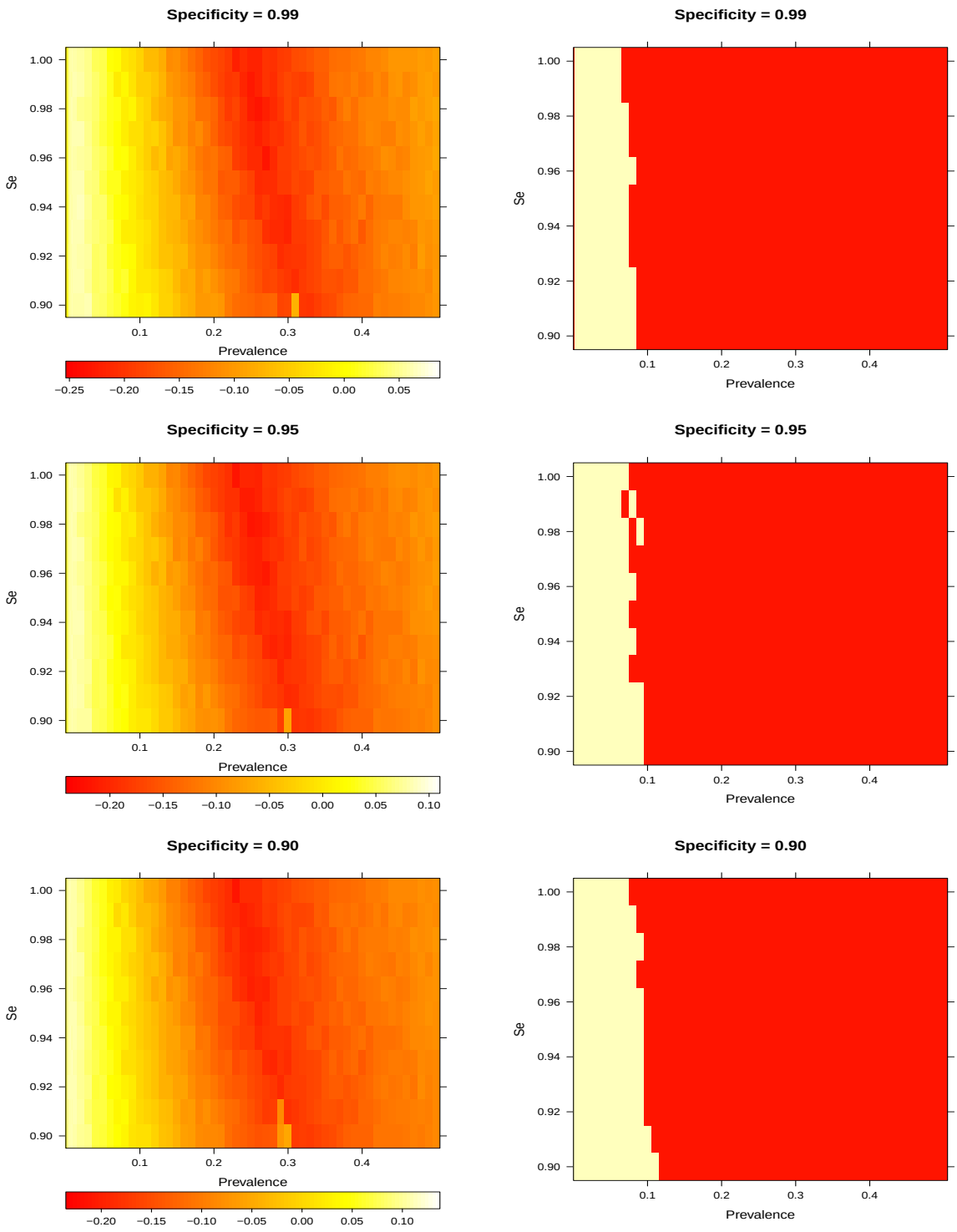

Figure 13: Comparison of TOD and A, when $M = 30$ and $N = 1000$, using different levels of S_e, S_p, and p. Left: Values of $\{E(T|\text{TOD}) - E(T|\text{A})\}/N$, the difference in expected individual expenditure between TOD and A. Negative values mean that TOD is more efficient. Right: Binary versions of the corresponding figure on the left; regions in red indicate those settings where TOD is more efficient. The optimal array size for A has been used throughout.

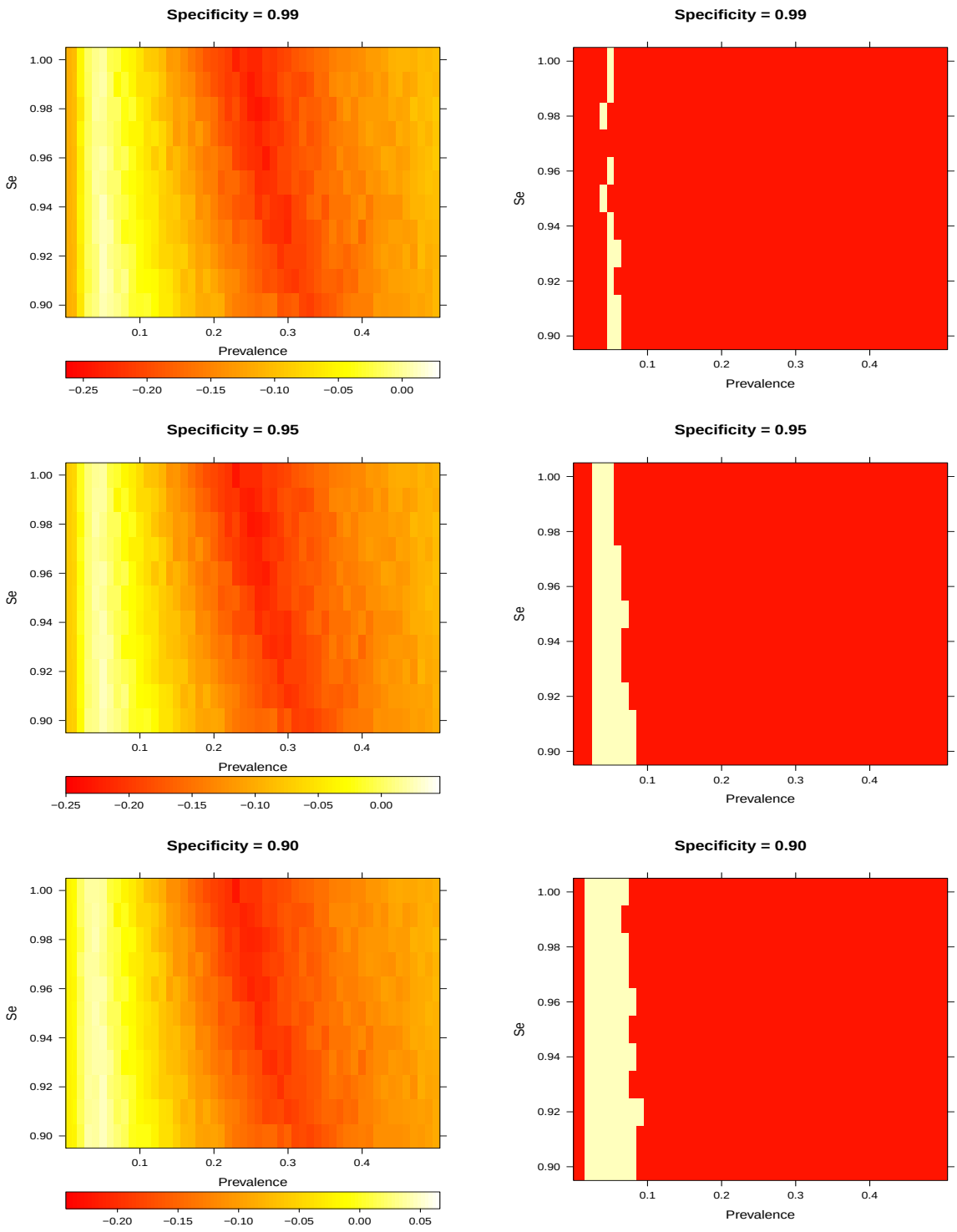

Figure 14: Comparison of PSOD and A, when $M = 10$ and $N = 1000$, using different levels of S_e, S_p, and p. Left: Values of $\{E(T|\text{PSOD}) - E(T|\text{A})\}/N$, the difference in expected individual expenditure between PSOD and A. Negative values mean that PSOD is more efficient. Right: Binary versions of the corresponding figure on the left; regions in red indicate those settings where PSOD is more efficient. The optimal array size for A has been used throughout.

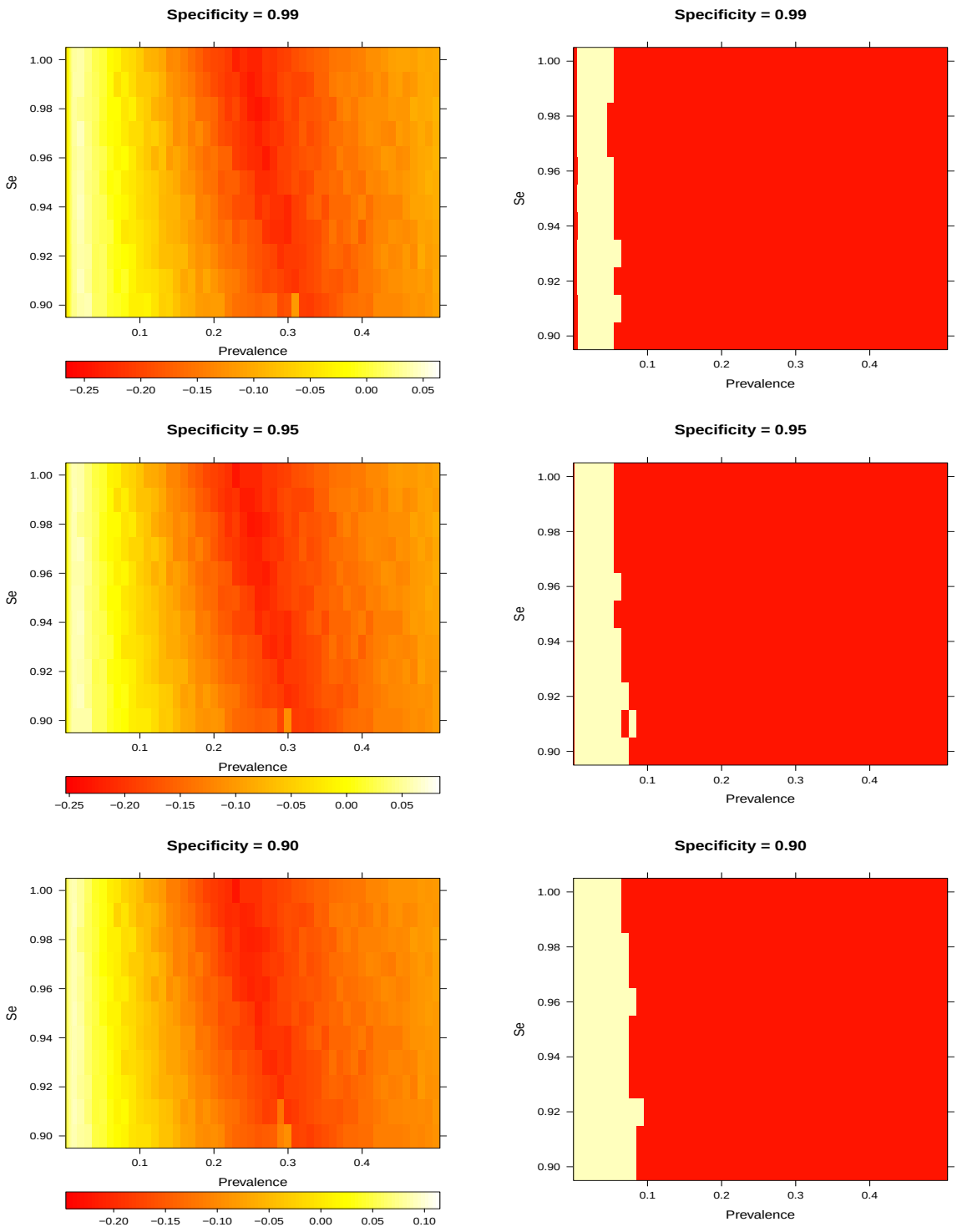

Figure 15: Comparison of PSOD and A, when $M = 20$ and $N = 1000$, using different levels of S_e, S_p, and p. Left: Values of $\{E(T|\text{PSOD}) - E(T|\text{A})\}/N$, the difference in expected individual expenditure between PSOD and A. Negative values mean that PSOD is more efficient. Right: Binary versions of the corresponding figure on the left; regions in red indicate those settings where PSOD is more efficient. The optimal array size for A has been used throughout.

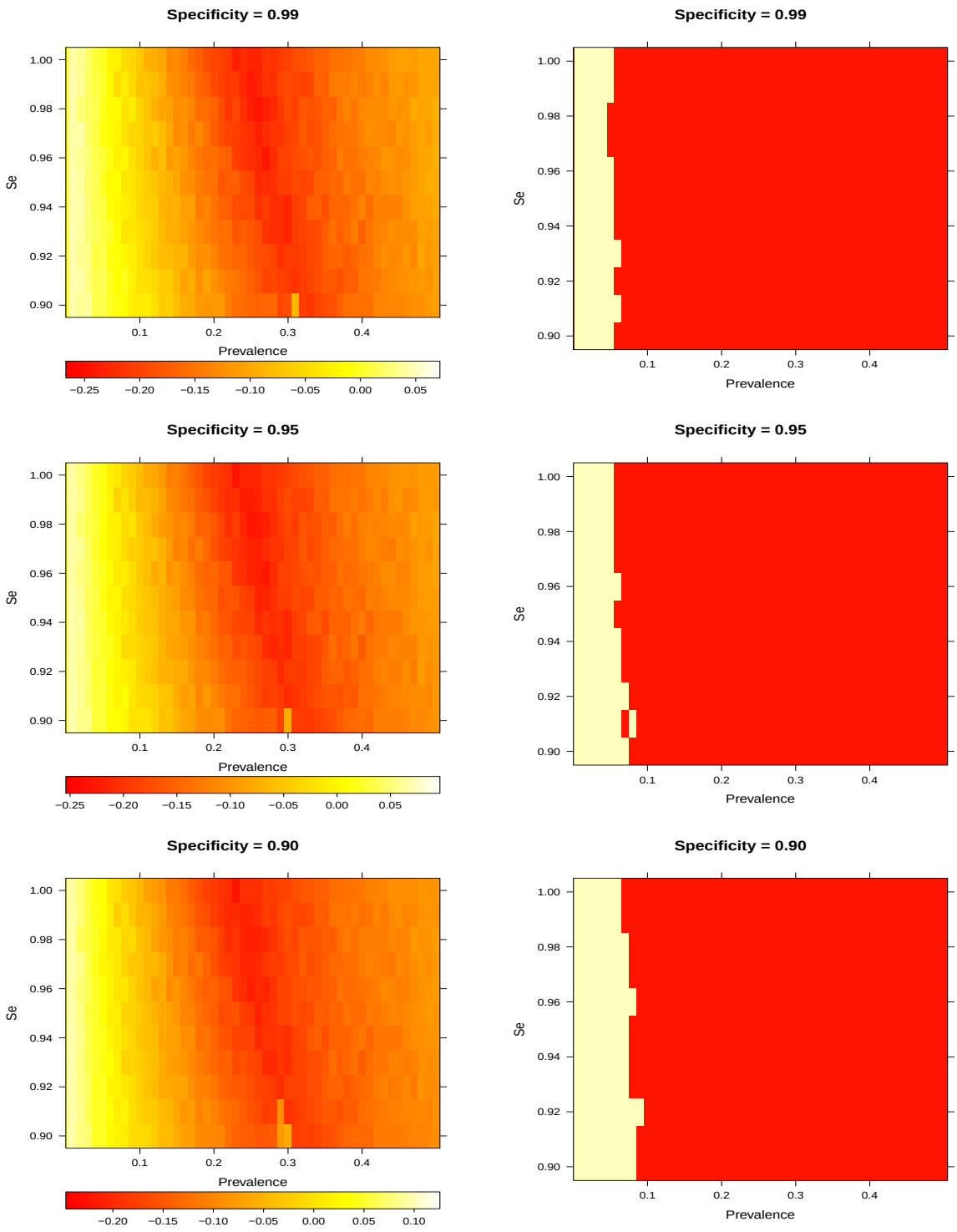

Figure 16: Comparison of PSOD and A, when $M = 30$ and $N = 1000$, using different levels of S_e, S_p, and p. Left: Values of $\{E(T|\text{PSOD}) - E(T|\text{A})\}/N$, the difference in expected individual expenditure between PSOD and A. Negative values mean that PSOD is more efficient. Right: Binary versions of the corresponding figure on the left; regions in red indicate those settings where PSOD is more efficient. The optimal array size for A has been used throughout.

Web Appendix D: *Nebraska IPP results in Section* 5. This appendix contains all of the Nebraska IPP simulation results. Values of S_e and S_p are given in Table 1 of the manuscript. All measures are averaged over $B = 1000$ implementations, as described in Section 5.

FIS: We summarize the salient features of the Full Informative Sterrett (FIS) algorithm from Bilder et al. (2010). FIS starts by ordering individuals in a positive pool $\mathcal{P}_j$ in terms of the risk probabilities, i.e., $p_{j(1)} \leq p_{j(2)} \leq \cdots \leq p_{j(c_j)}$, and begins by testing the highest risk individual $\mathcal{I}_{j(c_j)}$. If $\mathcal{I}_{j(c_j)}$ tests positive, the remaining individuals are pooled and tested. If this subpool tests negative, decoding is complete. If $\mathcal{I}_{j(c_j)}$ tests negative or if the subpool $\{\mathcal{I}_{j(1)}, \mathcal{I}_{j(2)}, ..., \mathcal{I}_{j(c_j-1)}\}$ tests positive, FIS moves to the next highest risk individual $\mathcal{I}_{j(c_j-1)}$. If $\mathcal{I}_{j(c_j-1)}$ tests positive, the $c_j - 2$ lowest risk individuals are pooled and tested. If $\mathcal{I}_{j(c_j-1)}$ tests negative or if the subpool $\{\mathcal{I}_{j(1)}, \mathcal{I}_{j(2)}, ..., \mathcal{I}_{j(c_j-2)}\}$ tests positive, FIS proceeds to test $\mathcal{I}_{j(c_j-2)}$. This process continues until a subpool, possibly consisting of only $\mathcal{I}_{j(1)}$, tests negative or until $\mathcal{I}_{j(1)}$ tests positive. The binary trees in Bilder et al. (2010) are helpful in depicting the FIS algorithm visually. To ensure the fairest possible comparison for 2009, we use the FIS (common) pool size that produced the smallest number of tests when decoding the 2008 individuals (averaged over $B = 1000$ implementations). Under FIS (for 2009), individuals are assigned to pools chronologically, based on the specimen arrival date. Executing the algorithm then proceeds (as described above) on each pool using the year-2009 estimated probabilities.

3SH: The noninformative 3-stage halving algorithm, denoted by 3SH, proceeds by splitting each positive pool $\mathcal{P}_j$, at random, into equal-sized subpools (or as close to equal as possible) and testing the subpools. A subpool that tests negative is declared to be free of positive individuals. A subpool that tests positive is decoded by using individual testing. Under 3SH, individuals are assigned to pools chronologically, based on the specimen arrival date. In our analysis of the Nebraska IPP data, we consider initial (master) pool sizes of $c_j = 4, 6, 8$ and 10 for 3SH. To ensure the fairest possible comparison for 2009, we use the master pool size that produced the smallest number of tests when decoding the 2008 individuals (averaged over $B = 1000$ implementations). Executing the algorithm then proceeds (as described above) on each pool.

SUMMARY MEASURES: To estimate the efficiency $E(T)$, we compute the mean number of tests $\overline{T}$, the averaged value of T over $B = 1000$ implementations. For classification accuracy, we compute

$$\begin{aligned}
\overline{PS}_e &= \frac{1}{B}\sum_{b=1}^{B}\frac{1}{\widetilde{N}^+}\sum_{i:\widetilde{g}_i=1} I(g_{i,b}=1) \\
\overline{PS}_p &= \frac{1}{B}\sum_{b=1}^{B}\frac{1}{\widetilde{N}^-}\sum_{i:\widetilde{g}_i=0} I(g_{i,b}=0) \\
\overline{PPV} &= \frac{1}{B}\sum_{b=1}^{B}\frac{1}{N_b^+}\sum_{i:g_{i,b}=1} I(\widetilde{g}_i=1) \\
\overline{NPV} &= \frac{1}{B}\sum_{b=1}^{B}\frac{1}{N_b^-}\sum_{i:g_{i,b}=0} I(\widetilde{g}_i=0),
\end{aligned}$$

where $g_{i,b} = 1(0)$, if the ith individual is classified as positive (negative) on the bth implementation, $\widetilde{N}^+$ ($\widetilde{N}^-$) denotes the number of positive (negative) individuals, and N_b^+ (N_b^-) denotes the number of individuals that test positive (negative) on the bth implementation. In contrast to the individual-specific measures derived in Section 2.2, reporting values of $\overline{PS}_e$, $\overline{PS}_p$, $\overline{PPV}$, and $\overline{NPV}$ in this way allows us to describe the global performance of the algorithms for the year-2009 Nebraska IPP data.

Table 1: Optimal pool sizes for Nebraska IPP screening in 2009. Note that OD, TOD, PSOD, and H determine their own algorithm-specific pool size(s).

			Optimal c			
Infection	Gender	Specimen	D	A	FIS	3SH
Chlamydia	Female	Urine	5	9	10	10
		Swab	5	8	8	8
	Male	Urine	5	8	8	8
		Swab	4	6	7	6
Gonorrhea	Female	Urine	8	10	10	10
		Swab	10	10	10	10
	Male	Urine	10	10	10	10
		Swab	5	8	10	8

Table 2: Summary measures for 2009 Nebraska IPP data. Male subjects. Mean number of tests and classification accuracy measures, averaged over $B = 1000$ implementations. Blocks of size $N = 50$ are used for OD, TOD, PSOD, and H. There are 6139 (1910) urine (swab) specimens for each infection.

	Method	$\overline{T}$	$\overline{PS}_e$	$\overline{PS}_p$	$\overline{PPV}$	$\overline{NPV}$
Chlamydia Urine	D	3417.3	0.865	0.985	0.833	0.988
	OD	3301.6	0.865	0.986	0.842	0.988
	TOD	3301.2	0.864	0.986	0.842	0.988
	PSOD	3221.2	0.867	0.987	0.858	0.988
	H	3218.2	0.865	0.987	0.858	0.988
	FIS	2715.8	0.843	0.990	0.887	0.986
	A	3136.4	0.806	0.988	0.857	0.983
	3SH	2721.1	0.695	0.992	0.881	0.974
Chlamydia Swab	D	1395.7	0.855	0.980	0.889	0.973
	OD	1299.1	0.854	0.982	0.897	0.973
	TOD	1290.8	0.867	0.983	0.903	0.976
	PSOD	1308.4	0.868	0.982	0.901	0.976
	H	1289.1	0.870	0.983	0.904	0.976
	FIS	1215.8	0.823	0.986	0.918	0.968
	A	1385.4	0.796	0.982	0.893	0.963
	3SH	1202.4	0.677	0.990	0.925	0.943
Gonorrhea Urine	D	1956.3	0.941	0.992	0.716	0.999
	OD	1767.9	0.941	0.994	0.768	0.999
	TOD	1768.6	0.942	0.994	0.770	0.999
	PSOD	1716.7	0.942	0.995	0.791	0.999
	H	1691.7	0.941	0.994	0.787	0.999
	FIS	1364.8	0.937	0.997	0.859	0.999
	A	1630.5	0.919	0.998	0.917	0.998
	3SH	1396.2	0.859	0.997	0.864	0.997
Gonorrhea Swab	D	1025.3	0.970	0.989	0.864	0.998
	OD	829.4	0.970	0.992	0.899	0.998
	TOD	787.6	0.977	0.991	0.894	0.998
	PSOD	750.0	0.976	0.993	0.909	0.998
	H	741.7	0.978	0.992	0.905	0.998
	FIS	627.3	0.966	0.996	0.949	0.997
	A	963.8	0.957	0.992	0.895	0.997
	3SH	918.5	0.927	0.992	0.902	0.995

Table 3: Summary measures for 2009 Nebraska IPP data. Female subjects. Mean number of tests and classification accuracy measures, averaged over $B = 1000$ implementations. Blocks of size $N = 50$ are used for OD, TOD, PSOD, and H. There are 4972 (14530) urine (swab) specimens for each infection.

	Method	$\overline{T}$	$\overline{PS}_e$	$\overline{PS}_p$	$\overline{PPV}$	$\overline{NPV}$
Chlamydia Urine	D	2483.6	0.649	0.990	0.848	0.970
	OD	2482.0	0.648	0.990	0.846	0.970
	TOD	2483.2	0.649	0.990	0.847	0.970
	PSOD	2458.7	0.653	0.990	0.857	0.970
	H	2472.4	0.654	0.991	0.865	0.970
	FIS	2049.4	0.598	0.991	0.856	0.966
	A	2124.6	0.530	0.993	0.876	0.960
	3SH	1558.9	0.338	0.995	0.863	0.945
Chlamydia Swab	D	7354.9	0.861	0.990	0.860	0.990
	OD	7214.6	0.861	0.990	0.864	0.990
	TOD	7213.4	0.861	0.990	0.864	0.990
	PSOD	7076.9	0.864	0.991	0.877	0.990
	H	7035.2	0.863	0.991	0.878	0.990
	FIS	5864.7	0.841	0.993	0.901	0.988
	A	6550.9	0.801	0.993	0.890	0.985
	3SH	5801.0	0.688	0.994	0.902	0.977
Gonorrhea Urine	D	1225.2	0.722	0.998	0.851	0.995
	OD	1200.5	0.721	0.998	0.849	0.995
	TOD	1203.8	0.721	0.998	0.849	0.995
	PSOD	1212.3	0.724	0.998	0.864	0.995
	H	1170.4	0.720	0.998	0.869	0.995
	FIS	929.0	0.703	0.999	0.908	0.995
	A	1229.2	0.659	0.999	0.943	0.994
	3SH	820.5	0.440	0.999	0.916	0.991
Gonorrhea Swab	D	3401.8	0.934	0.998	0.834	0.999
	OD	3179.8	0.933	0.998	0.871	0.999
	TOD	3184.8	0.934	0.998	0.871	0.999
	PSOD	3189.4	0.935	0.998	0.880	0.999
	H	3128.7	0.932	0.998	0.879	0.999
	FIS	2437.6	0.930	0.999	0.934	0.999
	A	3467.2	0.918	0.999	0.958	0.999
	3SH	2565.2	0.842	0.999	0.928	0.998

Table 4: Summary measures for 2009 Nebraska IPP data. Male subjects. Mean number of tests and classification accuracy measures, averaged over $B = 1000$ implementations. Blocks of size $N = 100$ are used for OD, TOD, PSOD, and H. There are 6139 (1910) urine (swab) specimens for each infection.

	Method	$\overline{T}$	$\overline{PS}_e$	$\overline{PS}_p$	$\overline{PPV}$	$\overline{NPV}$
Chlamydia Urine	D	3417.3	0.865	0.985	0.833	0.988
	OD	3267.4	0.864	0.986	0.844	0.988
	TOD	3267.6	0.865	0.986	0.843	0.988
	PSOD	3210.5	0.867	0.987	0.855	0.988
	H	3218.0	0.864	0.987	0.858	0.988
	FIS	2715.8	0.843	0.990	0.887	0.986
	A	3136.4	0.806	0.988	0.857	0.983
	3SH	2721.1	0.695	0.992	0.881	0.974
Chlamydia Swab	D	1395.7	0.855	0.980	0.889	0.973
	OD	1295.3	0.857	0.982	0.900	0.974
	TOD	1288.7	0.865	0.983	0.902	0.975
	PSOD	1285.4	0.868	0.982	0.902	0.976
	H	1303.5	0.868	0.982	0.901	0.975
	FIS	1215.8	0.823	0.986	0.918	0.968
	A	1385.4	0.796	0.982	0.893	0.963
	3SH	1202.4	0.677	0.990	0.925	0.943
Gonorrhea Urine	D	1956.3	0.941	0.992	0.716	0.999
	OD	1771.4	0.941	0.994	0.778	0.999
	TOD	1767.7	0.939	0.994	0.778	0.999
	PSOD	1672.8	0.942	0.995	0.796	0.999
	H	1680.9	0.941	0.994	0.787	0.998
	FIS	1364.8	0.937	0.997	0.859	0.999
	A	1630.5	0.919	0.998	0.917	0.998
	3SH	1396.2	0.859	0.997	0.864	0.997
Gonorrhea Swab	D	1025.3	0.970	0.989	0.864	0.998
	OD	811.4	0.970	0.991	0.886	0.998
	TOD	812.8	0.976	0.990	0.881	0.998
	PSOD	752.7	0.977	0.992	0.904	0.998
	H	746.1	0.976	0.992	0.903	0.998
	FIS	627.3	0.966	0.996	0.949	0.997
	A	963.8	0.957	0.992	0.895	0.997
	3SH	918.5	0.927	0.992	0.902	0.995

Table 5: Summary measures for 2009 Nebraska IPP data. Female subjects. Mean number of tests and classification accuracy measures, averaged over $B = 1000$ implementations. Blocks of size $N = 100$ are used for OD, TOD, PSOD, and H. There are 4972 (14530) urine (swab) specimens for each infection.

	Method	$\overline{T}$	$\overline{PS}_e$	$\overline{PS}_p$	$\overline{PPV}$	$\overline{NPV}$
Chlamydia Urine	D	2483.6	0.649	0.990	0.848	0.970
	OD	2469.4	0.649	0.990	0.846	0.970
	TOD	2468.1	0.648	0.990	0.845	0.970
	PSOD	2445.7	0.651	0.990	0.854	0.970
	H	2450.7	0.654	0.991	0.867	0.970
	FIS	2049.4	0.598	0.991	0.856	0.966
	A	2124.6	0.530	0.993	0.876	0.960
	3SH	1558.9	0.338	0.995	0.863	0.945
Chlamydia Swab	D	7354.9	0.861	0.990	0.860	0.990
	OD	7168.6	0.862	0.990	0.867	0.990
	TOD	7167.8	0.861	0.990	0.866	0.990
	PSOD	7107.6	0.863	0.991	0.872	0.990
	H	7103.3	0.863	0.990	0.874	0.989
	FIS	5864.7	0.841	0.993	0.901	0.988
	A	6550.9	0.801	0.993	0.890	0.985
	3SH	5801.0	0.688	0.994	0.902	0.977
Gonorrhea Urine	D	1225.2	0.722	0.998	0.851	0.995
	OD	1194.4	0.718	0.998	0.860	0.995
	TOD	1197.6	0.719	0.998	0.861	0.995
	PSOD	1176.9	0.724	0.998	0.867	0.995
	H	1155.4	0.723	0.998	0.871	0.995
	FIS	929.0	0.703	0.999	0.908	0.995
	A	1229.2	0.659	0.999	0.943	0.994
	3SH	820.5	0.440	0.999	0.916	0.991
Gonorrhea Swab	D	3401.8	0.934	0.998	0.834	0.999
	OD	3279.1	0.933	0.998	0.870	0.999
	TOD	3279.1	0.932	0.998	0.871	0.999
	PSOD	3128.0	0.935	0.998	0.881	0.999
	H	3127.1	0.932	0.998	0.874	0.999
	FIS	2437.6	0.930	0.999	0.934	0.999
	A	3467.2	0.918	0.999	0.958	0.999
	3SH	2565.2	0.842	0.999	0.928	0.998

Table 6: Summary measures for 2009 Nebraska IPP data. Male subjects. Mean number of tests and classification accuracy measures, averaged over $B = 1000$ implementations. Blocks of size $N = 200$ are used for OD, TOD, PSOD, and H. There are 6139 (1910) urine (swab) specimens for each infection.

	Method	$\overline{T}$	$\overline{PS}_e$	$\overline{PS}_p$	$\overline{PPV}$	$\overline{NPV}$
Chlamydia Urine	D	3417.3	0.865	0.985	0.833	0.988
	OD	3271.5	0.864	0.986	0.844	0.988
	TOD	3271.5	0.865	0.986	0.845	0.988
	PSOD	3218.1	0.867	0.987	0.853	0.988
	H	3240.8	0.865	0.987	0.855	0.988
	FIS	2715.8	0.843	0.990	0.887	0.986
	A	3136.4	0.806	0.988	0.857	0.983
	3SH	2721.1	0.695	0.992	0.881	0.974
Chlamydia Swab	D	1395.7	0.855	0.980	0.889	0.973
	OD	1311.4	0.857	0.982	0.899	0.974
	TOD	1301.2	0.865	0.981	0.896	0.975
	PSOD	1316.3	0.867	0.981	0.895	0.975
	H	1309.0	0.869	0.981	0.899	0.975
	FIS	1215.8	0.823	0.986	0.918	0.968
	A	1385.4	0.796	0.982	0.893	0.963
	3SH	1202.4	0.677	0.990	0.925	0.943
Gonorrhea Urine	D	1956.3	0.941	0.992	0.716	0.999
	OD	1704.5	0.941	0.994	0.759	0.999
	TOD	1704.7	0.940	0.994	0.761	0.999
	PSOD	1624.8	0.941	0.995	0.799	0.999
	H	1618.2	0.940	0.994	0.799	0.998
	FIS	1364.8	0.937	0.997	0.859	0.999
	A	1630.5	0.919	0.998	0.917	0.998
	3SH	1396.2	0.859	0.997	0.864	0.997
Gonorrhea Swab	D	1025.3	0.970	0.989	0.864	0.998
	OD	806.3	0.970	0.990	0.882	0.998
	TOD	816.9	0.977	0.989	0.874	0.998
	PSOD	766.3	0.976	0.992	0.898	0.998
	H	742.8	0.976	0.992	0.903	0.998
	FIS	627.3	0.966	0.996	0.949	0.997
	A	963.8	0.957	0.992	0.895	0.997
	3SH	918.5	0.927	0.992	0.902	0.995

Table 7: Summary measures for 2009 Nebraska IPP data. Female subjects. Mean number of tests and classification accuracy measures, averaged over $B = 1000$ implementations. Blocks of size $N = 200$ are used for OD, TOD, PSOD, and H. There are 4972 (14530) urine (swab) specimens for each infection.

	Method	$\overline{T}$	$\overline{PS}_e$	$\overline{PS}_p$	$\overline{PPV}$	$\overline{NPV}$
Chlamydia Urine	D	2483.6	0.649	0.990	0.848	0.970
	OD	2451.3	0.648	0.990	0.851	0.970
	TOD	2453.1	0.649	0.990	0.851	0.970
	PSOD	2434.5	0.648	0.990	0.852	0.970
	H	2469.9	0.652	0.991	0.964	0.970
	FIS	2049.4	0.598	0.991	0.856	0.966
	A	2124.6	0.530	0.993	0.876	0.960
	3SH	1558.9	0.338	0.995	0.863	0.945
Chlamydia Swab	D	7354.9	0.861	0.990	0.860	0.990
	OD	7189.1	0.861	0.990	0.866	0.990
	TOD	7190.6	0.862	0.990	0.866	0.990
	PSOD	7091.1	0.863	0.991	0.871	0.990
	H	7054.2	0.862	0.991	0.877	0.989
	FIS	5864.7	0.841	0.993	0.901	0.988
	A	6550.9	0.801	0.993	0.890	0.985
	3SH	5801.0	0.688	0.994	0.902	0.977
Gonorrhea Urine	D	1225.2	0.722	0.998	0.851	0.995
	OD	1178.6	0.722	0.998	0.837	0.995
	TOD	1177.8	0.722	0.998	0.836	0.995
	PSOD	1169.6	0.720	0.998	0.861	0.995
	H	1176.4	0.719	0.998	0.867	0.995
	FIS	929.0	0.703	0.999	0.908	0.995
	A	1229.2	0.659	0.999	0.943	0.994
	3SH	820.5	0.440	0.999	0.916	0.991
Gonorrhea Swab	D	3401.8	0.934	0.998	0.834	0.999
	OD	3215.3	0.932	0.998	0.851	0.999
	TOD	3216.7	0.933	0.998	0.851	0.999
	PSOD	3125.9	0.933	0.998	0.878	0.999
	H	3108.4	0.932	0.998	0.875	0.999
	FIS	2437.6	0.930	0.999	0.934	0.999
	A	3467.2	0.918	0.999	0.958	0.999
	3SH	2565.2	0.842	0.999	0.928	0.998

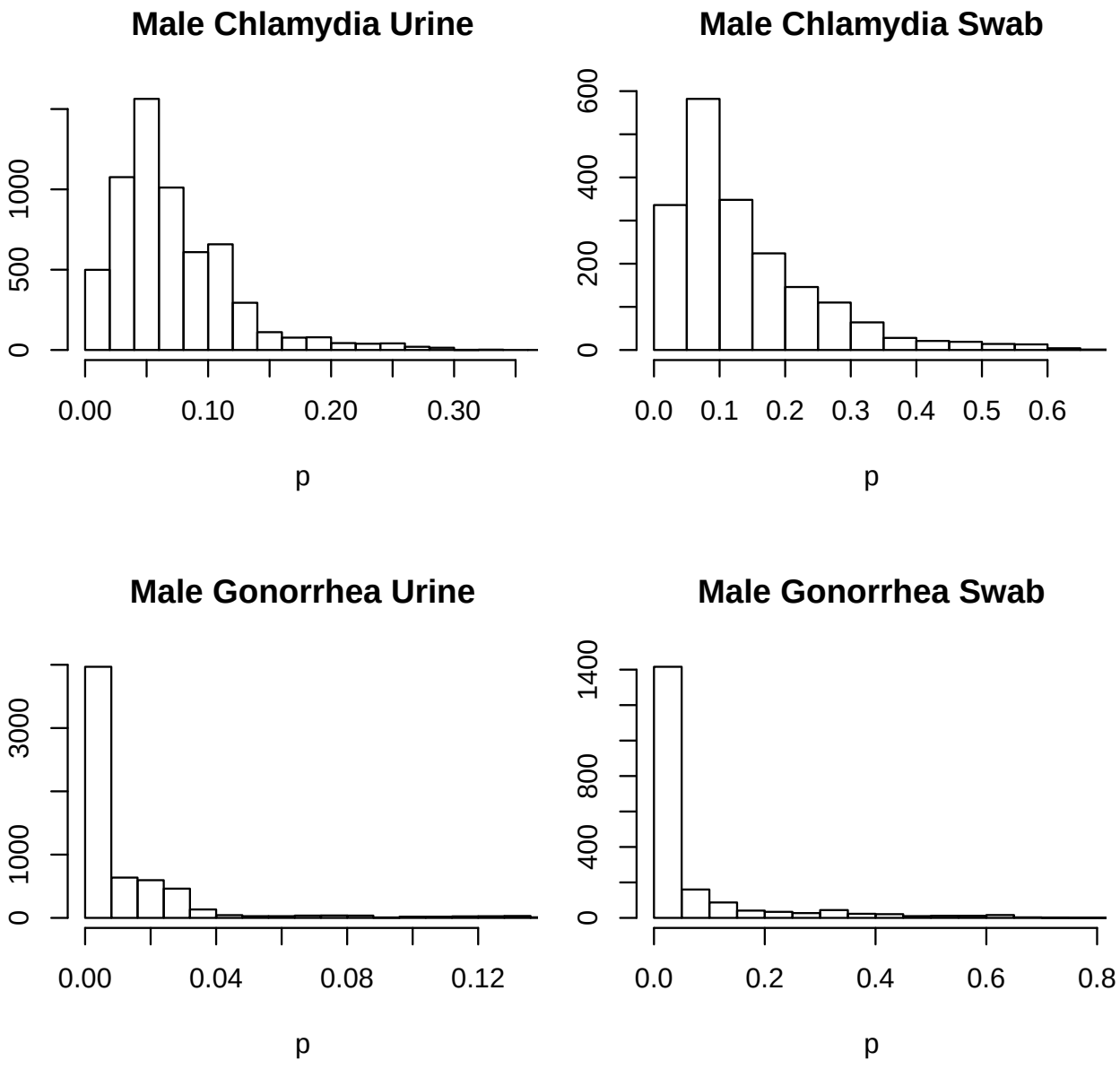

Figure 1: Nebraska IPP data. Histograms of the estimated probabilities $\widehat{p}_i$ from 2009. Male subjects.

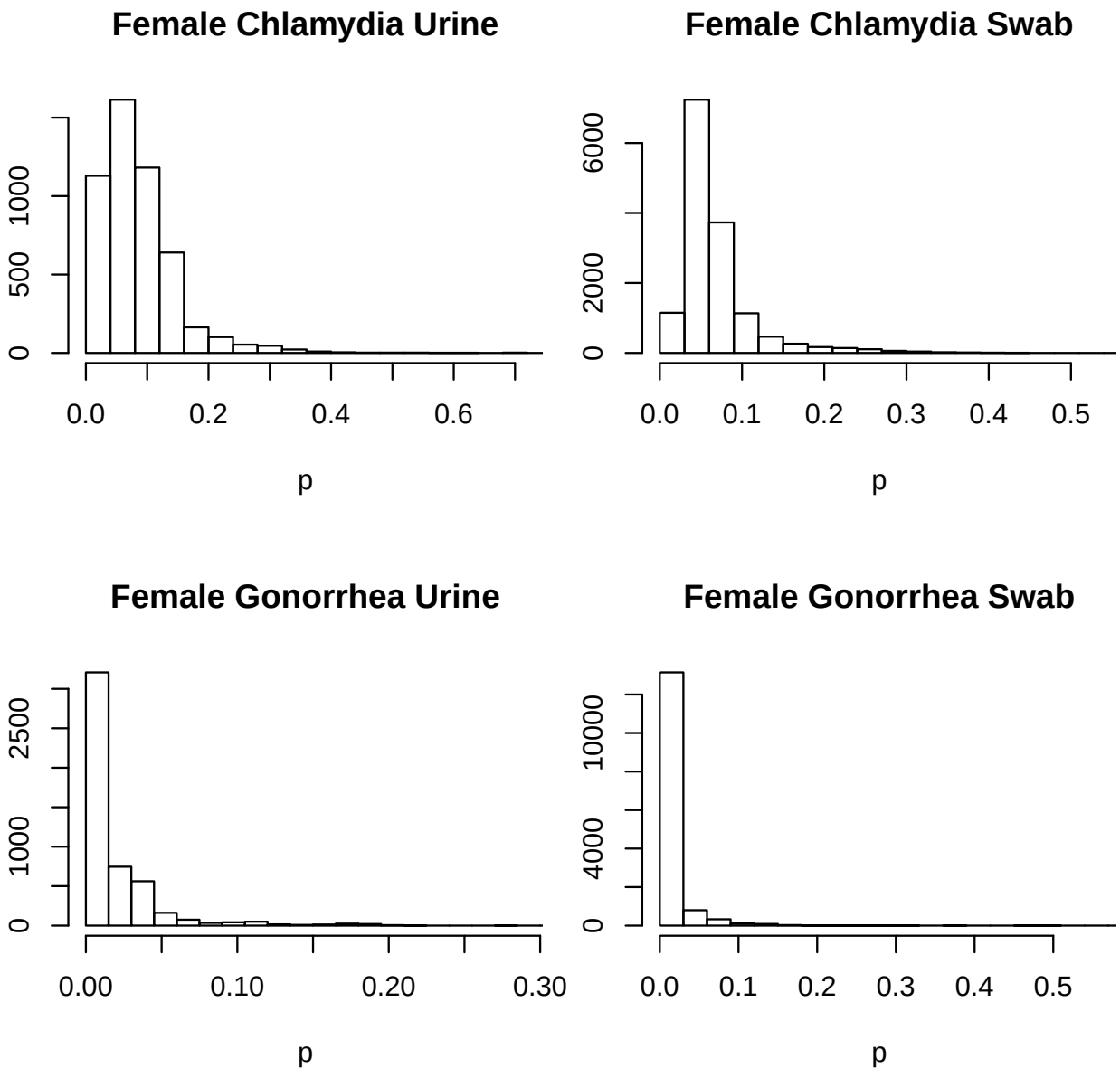

Figure 2: Nebraska IPP data. Histograms of the estimated probabilities $\widehat{p}_i$ from 2009. Female subjects.