

Supplementary Material for ‘Assessing simultaneous infection with multiple pathogens via group testing with imperfect multiplex assays’

Web Appendix A

This section contains additional details for the implementation of the EM algorithm and Louis’s method as described in Section 2.

Additional Details for the EM Algorithm

Recall that the E-step in the EM algorithm requires computing

$$E[I(\tilde{\mathbf{Y}}_i = \mathbf{y})|\mathbf{Z}, \boldsymbol{\delta}^{(t)}] = P(\tilde{\mathbf{Y}}_i = \mathbf{y}|\mathbf{Z}, \boldsymbol{\delta}^{(t)}) = d_{\mathbf{y}} \left(\sum_{\mathbf{z} \in \mathcal{Z}} d_{\mathbf{z}} \right)^{-1},$$

for $i \in \mathcal{P}_j$. Since $\tilde{\mathbf{Y}}$ is a discrete random vector, we could compute $P(\tilde{\mathbf{Y}}_i = \mathbf{y}|\mathbf{Z}, \boldsymbol{\delta}^{(t)})$ using the ‘brute force’ approach of summing $f(\tilde{\mathbf{Y}}|\mathbf{Z}, \boldsymbol{\delta}^{(t)})$ over all values of $\tilde{\mathbf{Y}}$ with $\tilde{\mathbf{Y}}_i = \mathbf{y}$. However, this is computationally impractical. Instead, we subset $\tilde{\mathbf{Y}}$ into the components from individuals in pool j and the remaining components, that is, we take $\tilde{\mathbf{Y}} = (\tilde{\mathbf{Y}}_{l:l \in \mathcal{P}_j}, \tilde{\mathbf{Y}}_{l:l \notin \mathcal{P}_j})$. We then note that

$$f(\tilde{\mathbf{Y}}|\mathbf{Z}, \boldsymbol{\delta}^{(t)}) \propto f(\tilde{\mathbf{Y}}, \mathbf{Z}|\boldsymbol{\delta}^{(t)}) = f(\tilde{\mathbf{Y}}_{l:l \in \mathcal{P}_j}|\mathbf{Z}, \boldsymbol{\delta}^{(t)})f(\mathbf{Y}_{l \notin \mathcal{P}_j}|\mathbf{Z}, \boldsymbol{\delta}^{(t)})$$

where

$$f(\tilde{\mathbf{Y}}_{l:l \in \mathcal{P}_j} | \mathbf{Z}, \boldsymbol{\delta}^{(t)}) = \prod_{k=1}^K \{Se_k^{\tilde{Z}_{jk}} (1 - Sp_k)^{1-\tilde{Z}_{jk}}\}^{Z_{jk}} \{(1 - Se_k)^{\tilde{Z}_{jk}} Sp_k^{1-\tilde{Z}_{jk}}\}^{1-Z_{jk}} \\ \times \prod_{l \in \mathcal{P}_j} \left[\left\{ \prod_{\mathbf{z} \in \mathcal{Y} \setminus \mathbf{1}} \delta_{\mathbf{z}}^{(t)I(\tilde{\mathbf{Y}}_l=\mathbf{z})} \right\} \left(1 - \sum_{\mathbf{z} \in \mathcal{Y} \setminus \mathbf{1}} \delta_{\mathbf{z}}^{(t)} \right)^{I(\tilde{\mathbf{Y}}_l=\mathbf{1})} \right]$$

and

$$f(\tilde{\mathbf{Y}}_{l:l \notin \mathcal{P}_j} | \mathbf{Z}, \boldsymbol{\delta}^{(t)}) = \prod_{m=1, m \neq j}^J \prod_{k=1}^K \{Se_k^{\tilde{Z}_{mk}} (1 - Sp_k)^{1-\tilde{Z}_{mk}}\}^{Z_{mk}} \{(1 - Se_k)^{\tilde{Z}_{mk}} Sp_k^{1-\tilde{Z}_{mk}}\}^{1-Z_{mk}} \\ \times \prod_{l \in \mathcal{P}_m} \left[\left\{ \prod_{\mathbf{z} \in \mathcal{Y} \setminus \mathbf{1}} \delta_{\mathbf{z}}^{(t)I(\tilde{\mathbf{Y}}_l=\mathbf{z})} \right\} \left(1 - \sum_{\mathbf{z} \in \mathcal{Y} \setminus \mathbf{1}} \delta_{\mathbf{z}}^{(t)} \right)^{I(\tilde{\mathbf{Y}}_l=\mathbf{1})} \right].$$

Thus we have that $\tilde{\mathbf{Y}}_{l:l \in \mathcal{P}_j}$ is independent from $\tilde{\mathbf{Y}}_{l:l \notin \mathcal{P}_j}$ given $\mathbf{Z}$ and $\boldsymbol{\delta}^{(t)}$. Thus

$$P(\tilde{\mathbf{Y}}_{\mathbf{i}} = \mathbf{y} | \mathbf{Z}, \boldsymbol{\delta}^{(t)}) \propto d_{\mathbf{y}} := \sum_{\tilde{\mathbf{Y}}_{l:l \in \mathcal{P}_j} \text{ with } \tilde{\mathbf{Y}}_{\mathbf{i}}=\mathbf{y}} f(\tilde{\mathbf{Y}}_{l:l \in \mathcal{P}_j} | \mathbf{Z}, \boldsymbol{\delta}^{(t)})$$

where the proportionality constant is given by

$$\sum_{\tilde{\mathbf{Y}}_{l:l \in \mathcal{P}_j}} f(\tilde{\mathbf{Y}}_{l:l \in \mathcal{P}_j} | \mathbf{Z}, \boldsymbol{\delta}^{(t)}) = \sum_{\mathbf{x} \in \mathcal{Y}} d_{\mathbf{x}}.$$

Note $d_{\mathbf{y}}$ depends on j only through $\mathbf{Z}_j$ and therefore $E[I(\tilde{\mathbf{Y}}_{\mathbf{i}} = \mathbf{y}) | \mathbf{Z}, \boldsymbol{\delta}^{(t)}]$ depends only on the observed value of $\mathbf{Z}_j$ and not directly on i or j . It is therefore only necessary to compute $E[I(\tilde{\mathbf{Y}}_{\mathbf{i}} = \mathbf{y}) | \mathbf{Z}, \boldsymbol{\delta}^{(t)}]$ once for each unique $\mathbf{Z}_j$ in the observed data set rather than once for each $i = 1, \dots, N$. This reduces the required number of computations from N to (at most) 2^K .

Additional Details for Louis's Method

Recall that Louis's method defines

$$\widehat{I(\boldsymbol{\delta})} = -\frac{\partial^2 Q(\boldsymbol{\delta}, \hat{\boldsymbol{\delta}})}{\partial \boldsymbol{\delta}, \partial \boldsymbol{\delta}'} - \text{var} \left(\frac{\partial f(\boldsymbol{\delta} | \mathbf{Z}, \tilde{\mathbf{Y}})}{\partial \boldsymbol{\delta}} \middle| \hat{\boldsymbol{\delta}} \right).$$

Let $\mathbf{a}, \mathbf{b} \in \mathcal{Y}^* = \mathcal{Y} \setminus \mathbf{1}$ with $\mathbf{a} \neq \mathbf{b}$ and let $l(\boldsymbol{\delta} | \mathbf{Z}, \tilde{\mathbf{Y}}) = \log\{f(\boldsymbol{\delta} | \mathbf{Z}, \tilde{\mathbf{Y}})\}$. We have that

$$\begin{aligned} \frac{\partial^2 l}{\partial \delta_{\mathbf{a}}^2} &= \sum_{j=1}^J \sum_{i \in \mathcal{P}_j} -I(\tilde{\mathbf{Y}}_i = \mathbf{a}) \delta_{\mathbf{a}}^{-2} - I(\tilde{\mathbf{Y}}_i = \mathbf{1}) \left(1 - \sum_{\mathbf{y} \in \mathcal{Y}^*} \delta_{\mathbf{y}}\right)^{-2} \\ \frac{\partial^2 l}{\partial \delta_{\mathbf{a}} \partial \delta_{\mathbf{b}}} &= \sum_{j=1}^J \sum_{i \in \mathcal{P}_j} -I(\tilde{\mathbf{Y}}_i = \mathbf{1}) \left(1 - \sum_{\mathbf{y} \in \mathcal{Y}^*} \delta_{\mathbf{y}}\right)^{-2} \end{aligned}$$

Therefore

$$\begin{aligned} \frac{\partial^2 Q(\boldsymbol{\delta}, \hat{\boldsymbol{\delta}})}{\partial \delta_{\mathbf{a}}^2} &= \sum_{j=1}^J \sum_{i \in \mathcal{P}_j} -E(\tilde{\mathbf{Y}}_i = \mathbf{a}) \hat{\delta}_{\mathbf{a}}^{-2} - E(\tilde{\mathbf{Y}}_i = \mathbf{1}) \left(1 - \sum_{\mathbf{y} \in \mathcal{Y}^*} \hat{\delta}_{\mathbf{y}}\right)^{-2} \\ \frac{\partial^2 Q(\boldsymbol{\delta}, \hat{\boldsymbol{\delta}})}{\partial \delta_{\mathbf{a}} \partial \delta_{\mathbf{b}}} &= \sum_{j=1}^J \sum_{i \in \mathcal{P}_j} -E(\tilde{\mathbf{Y}}_i = \mathbf{1}) \left(1 - \sum_{\mathbf{y} \in \mathcal{Y}^*} \hat{\delta}_{\mathbf{y}}\right)^{-2}. \end{aligned}$$

Note that the expectations are taken conditional on the observed data $\mathbf{Z}$. However, since pool responses in different pools are independent, we have $E(\tilde{\mathbf{Y}}_i = \mathbf{a}) = E(\tilde{\mathbf{Y}}_k = \mathbf{a})$ for all $i \in \mathcal{P}_j$ and all $k \in \mathcal{P}_l$ for which $\mathbf{Z}_j = \mathbf{Z}_l$. That is, the expectations depend only the observed pool response and not on the individual i or the specific pool j . This is computationally convenient, as it reduces the number of expectations which must be computed for each $\mathbf{a}$ from N to (at most) 2^K . We also note that these expectations are computed as part of the EM algorithm and thus the expectation computed on the final iteration of the algorithm may be used here.

Note that

$$\frac{\partial l}{\partial \delta_{\mathbf{a}}} = \sum_{j=1}^J \sum_{i \in \mathcal{P}_j} I(\tilde{\mathbf{Y}}_i = \mathbf{a}) \delta_{\mathbf{a}}^{-1} - I(\tilde{\mathbf{Y}}_i = \mathbf{y}^*) \left(1 - \sum_{\mathbf{y} \in \mathcal{Y}^*} \delta_{\mathbf{y}}\right)^{-1}$$

And thus

$$\begin{aligned}
\text{var} \left(\frac{\partial l}{\partial \delta_{\mathbf{a}}} \middle| \hat{\boldsymbol{\delta}} \right) &= \sum_{j=1}^J \text{var} \left[\sum_{i \in \mathcal{P}_j} I(\tilde{\mathbf{Y}}_i = \mathbf{a}) \hat{\delta}_{\mathbf{a}}^{-1} - I(\tilde{\mathbf{Y}}_i = \mathbf{1}) (1 - \sum_{\mathbf{y} \in \mathcal{Y}^*} \hat{\delta}_{\mathbf{y}})^{-1} \right] \\
&= \sum_{j=1}^J E \left\{ \left[\sum_{i \in \mathcal{P}_j} I(\tilde{\mathbf{Y}}_i = \mathbf{a}) \hat{\delta}_{\mathbf{a}}^{-1} - I(\tilde{\mathbf{Y}}_i = \mathbf{1}) (1 - \sum_{\mathbf{y} \in \mathcal{Y}^*} \hat{\delta}_{\mathbf{y}})^{-1} \right]^2 \right\} \\
&\quad - \left\{ E \left[\sum_{i \in \mathcal{P}_j} I(\tilde{\mathbf{Y}}_i = \mathbf{a}) \hat{\delta}_{\mathbf{a}}^{-1} - I(\tilde{\mathbf{Y}}_i = \mathbf{1}) (1 - \sum_{\mathbf{y} \in \mathcal{Y}^*} \hat{\delta}_{\mathbf{y}})^{-1} \right] \right\}^2
\end{aligned}$$

Note that these expectations again depend only on the observed value $\mathbf{Z}_j$ and not directly on j or i , and thus we need only compute 2^K expectations for each $\mathbf{a}$. Once again, partition $\tilde{\mathbf{Y}} = (\tilde{\mathbf{Y}}_{l:l \in \mathcal{P}_j}, \tilde{\mathbf{Y}}_{l:l \notin \mathcal{P}_j})$ and let $\mathbf{y}_{l:l \in \mathcal{P}_j} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_c)'$ for $\mathbf{y}_i \in \mathcal{Y}$. We have that

$$P(\tilde{\mathbf{Y}}_{l:l \in \mathcal{P}_j} = \mathbf{y}_{l:l \in \mathcal{P}_j} | \mathbf{Z}_j, \hat{\boldsymbol{\delta}}) = \frac{p_{\mathbf{y}_{l:l \in \mathcal{P}_j}}}{\sum_{\mathbf{x} \in \mathcal{Y}^c} p_{\mathbf{x}}}$$

where $\mathcal{Y}^c$ denotes the set product of $\mathcal{Y}$ with itself c times and

$$\begin{aligned}
p_{\mathbf{y}_{l:l \in \mathcal{P}_j}} &= \prod_{k=1}^K \{ Se_k^{\tilde{z}_k} (1 - Sp_k)^{1-\tilde{z}_k} \}^{Z_{jk}} \{ (1 - Se_k)^{\tilde{z}_k} Sp_k^{1-\tilde{z}_k} \}^{1-Z_{jk}} \\
&\quad \times \prod_{l=1}^c \left\{ \left[\prod_{\mathbf{z} \in \mathcal{Y}^*} \hat{\delta}_{\mathbf{z}}^{I(\mathbf{y}_l=\mathbf{z})} \right] \left(1 - \sum_{\mathbf{z} \in \mathcal{Y}^*} \hat{\delta}_{\mathbf{z}} \right)^{I(\mathbf{y}_l=\mathbf{1})} \right\}
\end{aligned}$$

and $\tilde{\mathbf{z}}$ is the pool status corresponding to $\mathbf{y}_{l:l \in \mathcal{P}_j}$. Thus we can compute the expectations in (2) and (3) by noting that they are a function of $\mathbf{Y}_{l:l \in \mathcal{P}_j}$ and using $P(\mathbf{Y}_{l:l \in \mathcal{P}_j} | \mathbf{Z}_j, \hat{\boldsymbol{\delta}})$.

For $\mathbf{a} \neq \mathbf{b}$, we have that

$$\text{cov} \left(\frac{\partial l}{\partial \delta_{\mathbf{a}}} \middle| \hat{\boldsymbol{\delta}}, \frac{\partial l}{\partial \delta_{\mathbf{b}}} \middle| \hat{\boldsymbol{\delta}} \right)$$

$$\begin{aligned}
&= \sum_{j=1}^J \text{cov} \left(\sum_{i \in \mathcal{P}_j} I(\tilde{\mathbf{Y}}_i = \mathbf{a}) \delta_{\mathbf{a}}^{-1} - I(\tilde{\mathbf{Y}}_i = \mathbf{y}^*) (1 - \sum_{\mathbf{y} \in \mathcal{Y}^*} \delta_{\mathbf{y}})^{-1}, \right. \\
&\quad \left. \sum_{i \in \mathcal{P}_j} I(\tilde{\mathbf{Y}}_i = \mathbf{b}) \delta_{\mathbf{b}}^{-1} - I(\tilde{\mathbf{Y}}_i = \mathbf{y}^*) (1 - \sum_{\mathbf{y} \in \mathcal{Y}^*} \delta_{\mathbf{y}})^{-1} \right) \\
&= \sum_{j=1}^J E \left\{ \left(\sum_{i \in \mathcal{P}_j} I(\tilde{\mathbf{Y}}_i = \mathbf{a}) \delta_{\mathbf{a}}^{-1} - I(\tilde{\mathbf{Y}}_i = \mathbf{y} \setminus \mathbf{1}) (1 - \sum_{\mathbf{y} \in \mathcal{Y}^*} \delta_{\mathbf{y}})^{-1} \right) \right. \\
&\quad \times \left. \left(\sum_{i \in \mathcal{P}_j} I(\tilde{\mathbf{Y}}_i = \mathbf{b}) \delta_{\mathbf{b}}^{-1} - I(\tilde{\mathbf{Y}}_i = \mathbf{y}^*) (1 - \sum_{\mathbf{y} \in \mathcal{Y}^*} \delta_{\mathbf{y}})^{-1} \right) \right\} \\
&\quad - \left\{ E \left(\sum_{i \in \mathcal{P}_j} I(\tilde{\mathbf{Y}}_i = \mathbf{a}) \delta_{\mathbf{a}}^{-1} - I(\tilde{\mathbf{Y}}_i = \mathbf{y}^*) (1 - \sum_{\mathbf{y} \in \mathcal{Y}^*} \delta_{\mathbf{y}})^{-1} \right) \right. \\
&\quad \times \left. E \left(\sum_{i \in \mathcal{P}_j} I(\tilde{\mathbf{Y}}_i = \mathbf{b}) \delta_{\mathbf{b}}^{-1} - I(\tilde{\mathbf{Y}}_i = \mathbf{y}^*) (1 - \sum_{\mathbf{y} \in \mathcal{Y}^*} \delta_{\mathbf{y}})^{-1} \right) \right\}.
\end{aligned}$$

Again, these expectations depend only on the value of $\mathbf{Z}_j$ and not directly on j or i and thus we need only compute 2^K expectations for each pair of $\mathbf{a}$ and $\mathbf{b}$. Note that we can compute these expectations using the same $P(\tilde{\mathbf{Y}}_{\mathcal{P}_j} | \mathbf{Z}_j, \hat{\boldsymbol{\delta}})$ used to compute $\text{var}(\frac{\partial l}{\partial \mathbf{a}})$.

Additional Details for the Independent Model

In order to fit the model which assumes independent co-infections from Section 2.3, we must compute

$$E \left[I(\tilde{Y}_{ik} = l | \mathbf{Z}, \boldsymbol{\pi}^{(t)}) \right] = P(\tilde{Y}_{ik} = l | \mathbf{Z}, \boldsymbol{\pi}^{(t)})$$

for $i \in \mathcal{P}_j$ and $l \in \{0, 1\}$. Just as in the full multinomial model, we can show $\tilde{\mathbf{Y}}_{l:l \in \mathcal{P}_j}$ and $\tilde{\mathbf{Y}}_{l:l \notin \mathcal{P}_j}$ are independent given $\mathbf{Z}$ and $\boldsymbol{\pi}$. Take

$$\begin{aligned}
f(\tilde{\mathbf{Y}}_{l:l \in \mathcal{P}_j} | \mathbf{Z}, \boldsymbol{\delta}^{(t)}) &= \prod_{k=1}^K \{ Se_k^{\tilde{Z}_{jk}} (1 - Sp_k)^{1 - \tilde{Z}_{jk}} \}^{Z_{jk}} \{ (1 - Se_k)^{\tilde{Z}_{jk}} Sp_k^{1 - \tilde{Z}_{jk}} \}^{1 - Z_{jk}} \\
&\quad \times \prod_{l \in \mathcal{P}_j} \prod_{\mathbf{z} \in \mathcal{Y}} \delta_{\mathbf{z}}^{(t)I(\tilde{\mathbf{Y}}_i = \mathbf{z})},
\end{aligned}$$

where $\delta_{\mathbf{z}} = \prod_{k=1}^K \pi_k^{z_k} (1 - \pi_k)^{1-z_k}$. Therefor

$$P(\tilde{Y}_{ik} = l | \mathbf{Z}, \boldsymbol{\pi}^{(t)}) \propto c_{kl} := \sum_{\tilde{\mathbf{Y}}_{l:l \in \mathcal{P}_j} \text{ with } \tilde{Y}_{ik}=l} f(\tilde{\mathbf{Y}}_{l:l \in \mathcal{P}_j} | \mathbf{Z}, \boldsymbol{\delta}^{(t)})$$

with proportionality constant $c_{k0} + c_{k1}$.

Web Appendix B

This section provides additional details about the relationship between individual-level and pool-level infection probabilities.

Given a pool size c and individual infection probabilities $\boldsymbol{\delta}$, we may wish to calculate $\theta_{\mathbf{z}} = P(\tilde{\mathbf{Z}}_j = \mathbf{z} | \boldsymbol{\delta})$, the probability that a pool of size c has true infection status $\mathbf{z}$ given $\boldsymbol{\delta}$. We first note that $\theta_{\mathbf{0}} = \delta_{\mathbf{0}}^c$. For $1 \leq k \leq K$, define $\mu_{\mathbf{z}} = P(\tilde{Z}_{jk} = 0 \forall k : z_k = 0)$, that is $\mu_{\mathbf{z}}$ is the probability that a pool of size c contains no individuals infected with any pathogens k for which $z_k = 0$. We have that $\mu_{\mathbf{z}} = (\sum_{\mathbf{x}: x_k=0 \forall k: z_k=0} \delta_{\mathbf{x}})^c$. We also have that $\mu_{\mathbf{z}} = \sum_{\mathbf{x}: x_k=0 \forall k: z_k=0} \theta_{\mathbf{x}}$. This produces a system of 2^K non-linear equations in 2^K unknowns. The system can be easily solved for $\boldsymbol{\theta} = (\theta_{\mathbf{z}})_{\mathbf{z} \in \mathcal{Z}}$ by starting with $\theta_{\mathbf{0}} = \delta_{\mathbf{0}}^c$. The equations derived from $\mu_{\mathbf{z}}$ for $\mathbf{z}$ with exactly one entry equal to 1 each involve only $\theta_{\mathbf{0}}$ and $\theta_{\mathbf{z}}$, which allows us to next solve for $\theta_{\mathbf{z}}$ for all $\mathbf{z}$ with exactly one 1. Each equation involving $\mu_{\mathbf{z}}$ with exactly two entries equal to 1 each involve only $\theta_{\mathbf{x}}$ s previously solved for and $\theta_{\mathbf{z}}$. This pattern continues, allowing us to solve for all components of $\boldsymbol{\theta}$.

Web Appendix C

Web Figures 1-6 provide enlarged versions of the panels from Figure 1; Web Figures 7-12 provide enlarged versions of the panes from Figure 2.

We conducted a simulation study to assess the performance of the test for mutual independence described in Section 2.3. We generated data using $c = 3$ for each of the $\boldsymbol{\delta}$ values in Table 1, with $Se_k = Sp_k = 0.99$ for all k . Supplementary Table 1 provides the percent

Scenario	1	2	3	4	5	6
Rejection Rate	7%	100%	100%	10%	100%	100%

Web Table 1: The table displays the empirical rejection rate of the null hypothesis of independence using data generated with each value of δ shown in Table 1 and $c = 3$.

of datasets for which the likelihood ratio test statistic exceeded the 95th quantile of a chi-squared distribution with 4 degrees of freedom. For the independent scenarios (1 and 4), this quantity is the empirical type I error rate. For the other scenarios, this quantity is the empirical power.

Web Appendix D

This section provides additional details about the likelihood ratio test used in Section 5.

Recall that we wish to test $H_0 : \delta_{1100} = \delta_{1010} = \delta_{1110} = \delta_{0101} = \delta_{1101} = \delta_{0011} = \delta_{1011} = \delta_{0111} = \delta_{1111}$ versus $H_A : \delta_{1100}, \delta_{1010}, \delta_{1110}, \delta_{0101}, \delta_{1101}, \delta_{0011}, \delta_{1011}, \delta_{0111}, \delta_{1111}$ not all 0. Note that since under H_A , $0 \leq \delta_z$ for all z , the null hypothesis exists on the boundary of the parameter space for the alternative hypothesis, and thus the assumptions of the usual likelihood ratio test are violated. As such, we have no guarantee that the null distribution of the likelihood ratio test statistic follows the usual chi-squared distribution. To circumvent this, we use a Monte Carlo procedure to estimate the null distribution of the test statistic. We generated 100 datasets by independently generating $n = 573$ individual infection statuses from a multinomial distribution with $\delta = \hat{\delta}$ (Reduced) from Table 2. We randomly assigned these individuals to 191 pools of size 3 and generated pool testing responses as described in Section 3. We then fit both the full and reduced model to each datasets and computed the likelihood ratio test statistics. We then took the 95th quantile of the sample of 100 test statistics as the critical value for our hypothesis test.

Web Appendix E

We performed a sensitivity analysis using the data described in Section 4 taking $Se_k = Sp_k = 0.97$. When performing the test for independence, we found a likelihood ratio test statistic

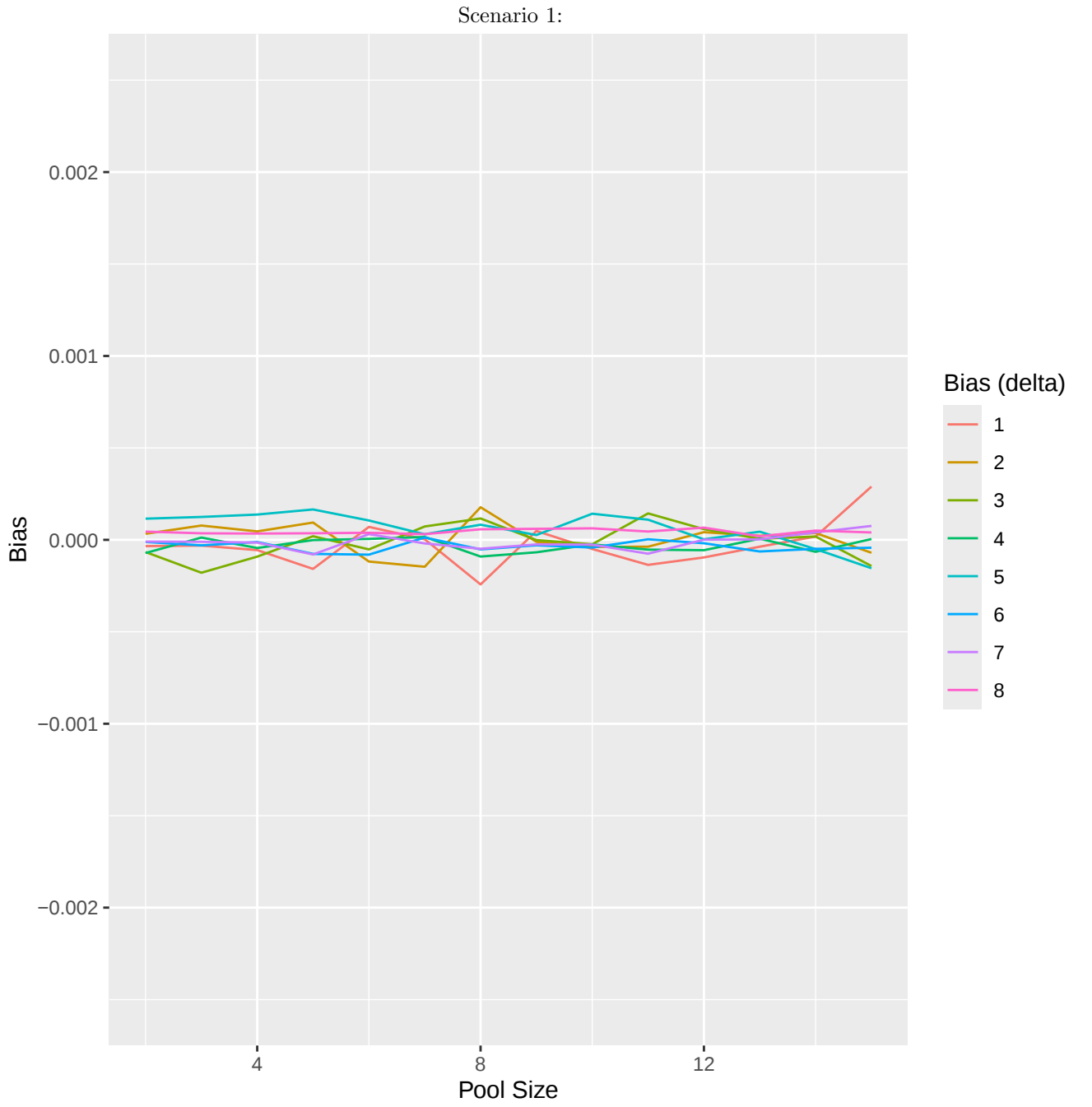
	<i>R. amblyommatis</i>	<i>R. parkeri</i>	<i>PME</i>	<i>E. ewingii</i>	$\hat{\delta}$ (Full)	$\hat{\delta}$ (Reduced)
	0	0	0	0	0.6162	0.6162
	1	0	0	0	0.3153	0.3153
	0	1	0	0	0.0615	0.0615
	1	1	0	0	$<10^{-50}$	0
	0	0	1	0	0.0014	0.0014
	1	0	1	0	$<10^{-50}$	0
	0	1	1	0	0.0016	0.0016
	1	1	1	0	$<10^{-50}$	0
	0	0	0	1	9.6119×10^{-21}	7.8518×10^{-21}
	1	0	0	1	0.0042	0.0042
	0	1	0	1	$<10^{-50}$	0
	1	1	0	1	$<10^{-50}$	0
	0	0	1	1	$<10^{-50}$	0
	1	0	1	1	$<10^{-50}$	0
	0	1	1	1	$<10^{-50}$	0
	1	1	1	1	$<10^{-50}$	0
$\hat{p}$ (Full)	0.3194	0.0630	0.0029	0.0042		
$\hat{p}$ (Reduced)	0.3194	0.0630	0.0029	0.0042		

Web Table 2: The table provides the estimated multinomial probabilities ($\hat{\delta}$) and estimated marginal prevalences ($\hat{p}$) for the South Carolina tick data.

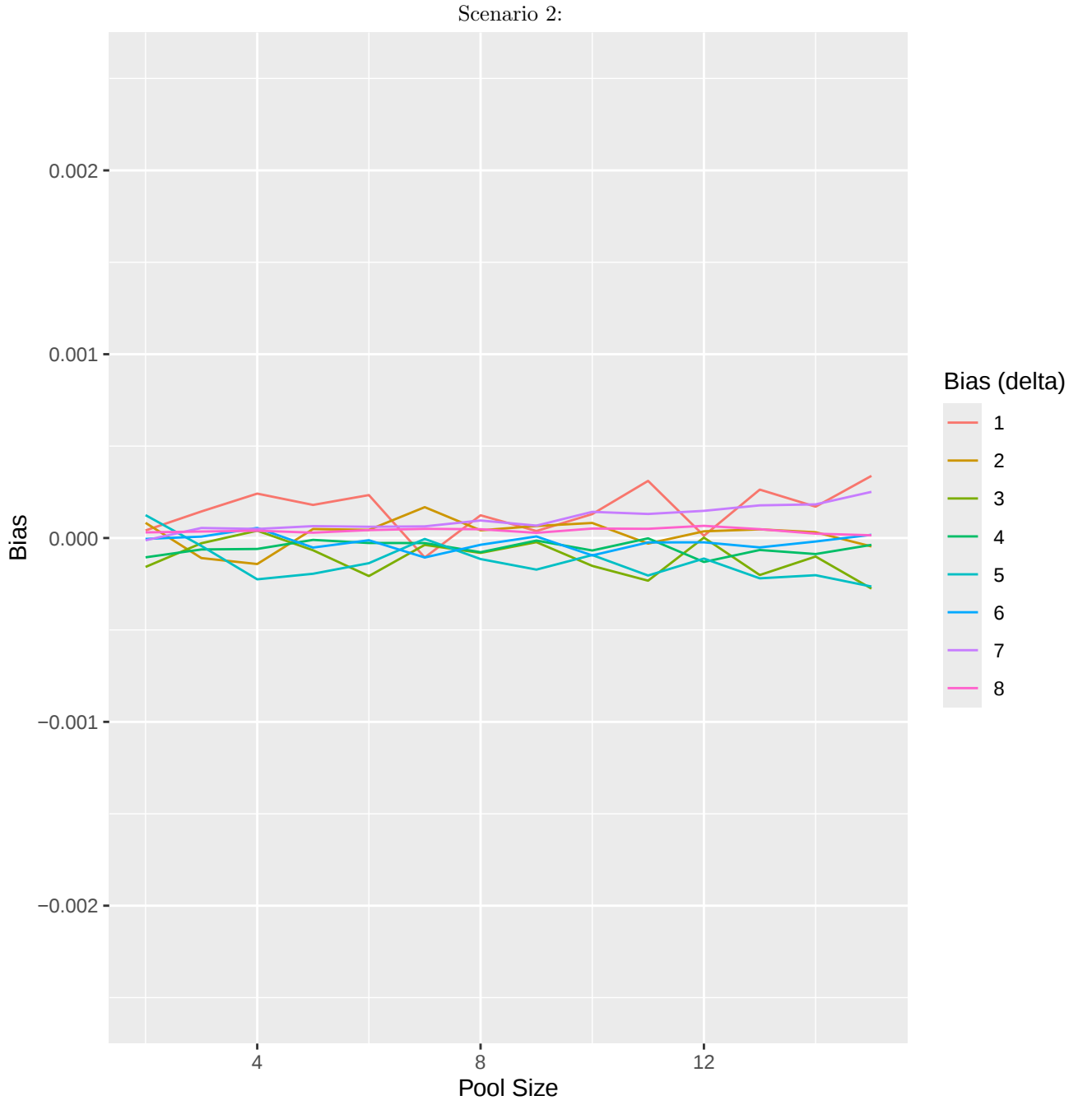
of $17.64 > \chi_{0.95,9} = 16.92$, so we rejected the null hypothesis of independence. The estimate of $\hat{\delta}$ from the full multinomial model is provided in Web Table 2.

We refit the model setting $\delta_{\mathbf{y}} = 0$ for all $\mathbf{y}$ for which $\hat{\delta}_{\mathbf{y}} < 10^{-50}$. We used a likelihood ratio test to compare this reduced model to the full model and failed to reject the null hypothesis. We report $\hat{\delta}$ from the reduced model in Table 2.

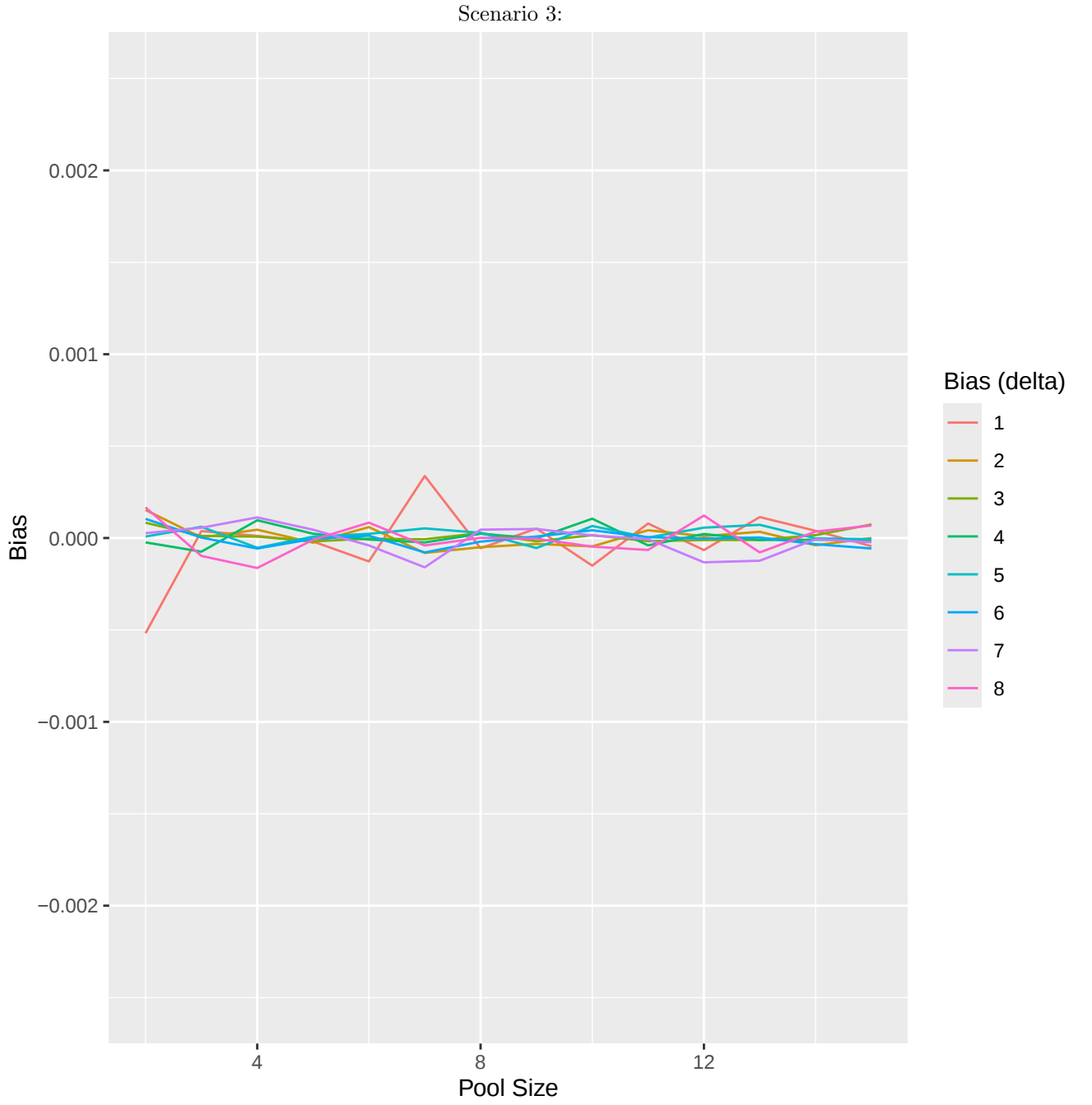
As in the original data application, we applied the delta method to examine *R. parkeri*-PME and *R. amblyommatis*-*E. ewingii* co-infections using the reduced model. We found $\hat{d}(2, 3) = -0.0012$, 95% confidence interval $(-0.0013, -0.0012)$ and thus conclude that co-infection with *R. parkeri* and PME is more likely than expected under independence. We also found $\hat{d}(1, 4) = -0.0046$, 95% confidence interval $(-0.0047, -0.0046)$ and conclude that infection with *R. amblyommatis* and *E. ewingii* are positively associated. These results are consistent with those of the primary analysis.



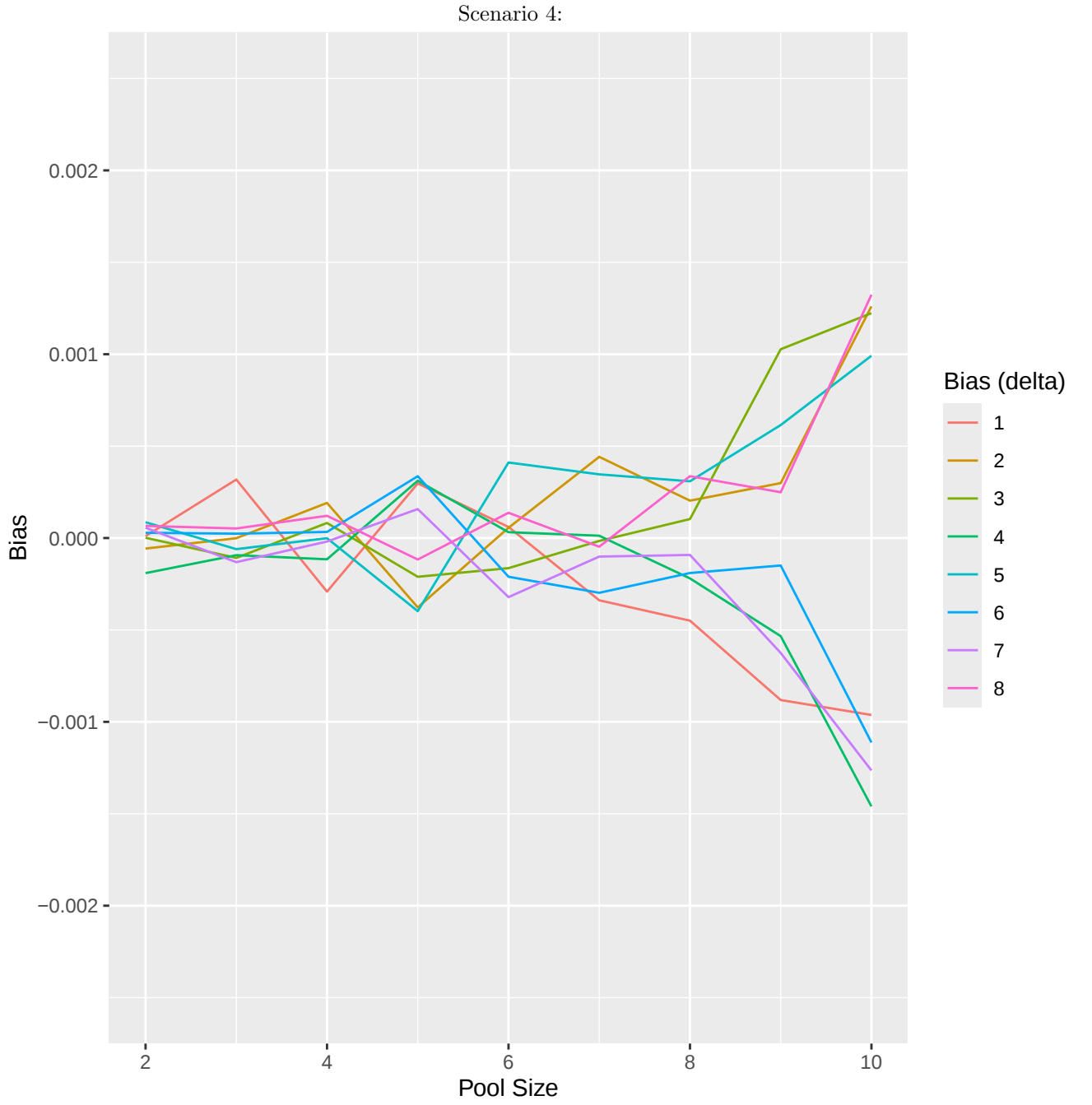
Web Fig. 1: The figure displays the average empirical bias of the estimator $\hat{\delta}$ from scenario 1 as a function of pool size. From 1 to 8, the position numbers correspond to $\mathbf{y} = (0, 0, 0), (1, 0, 0), (0, 1, 0), (1, 0, 1), (1, 1, 0), (0, 1, 1)$ and $(1, 1, 1)$. This is an enlarged version of Figure 1(a).



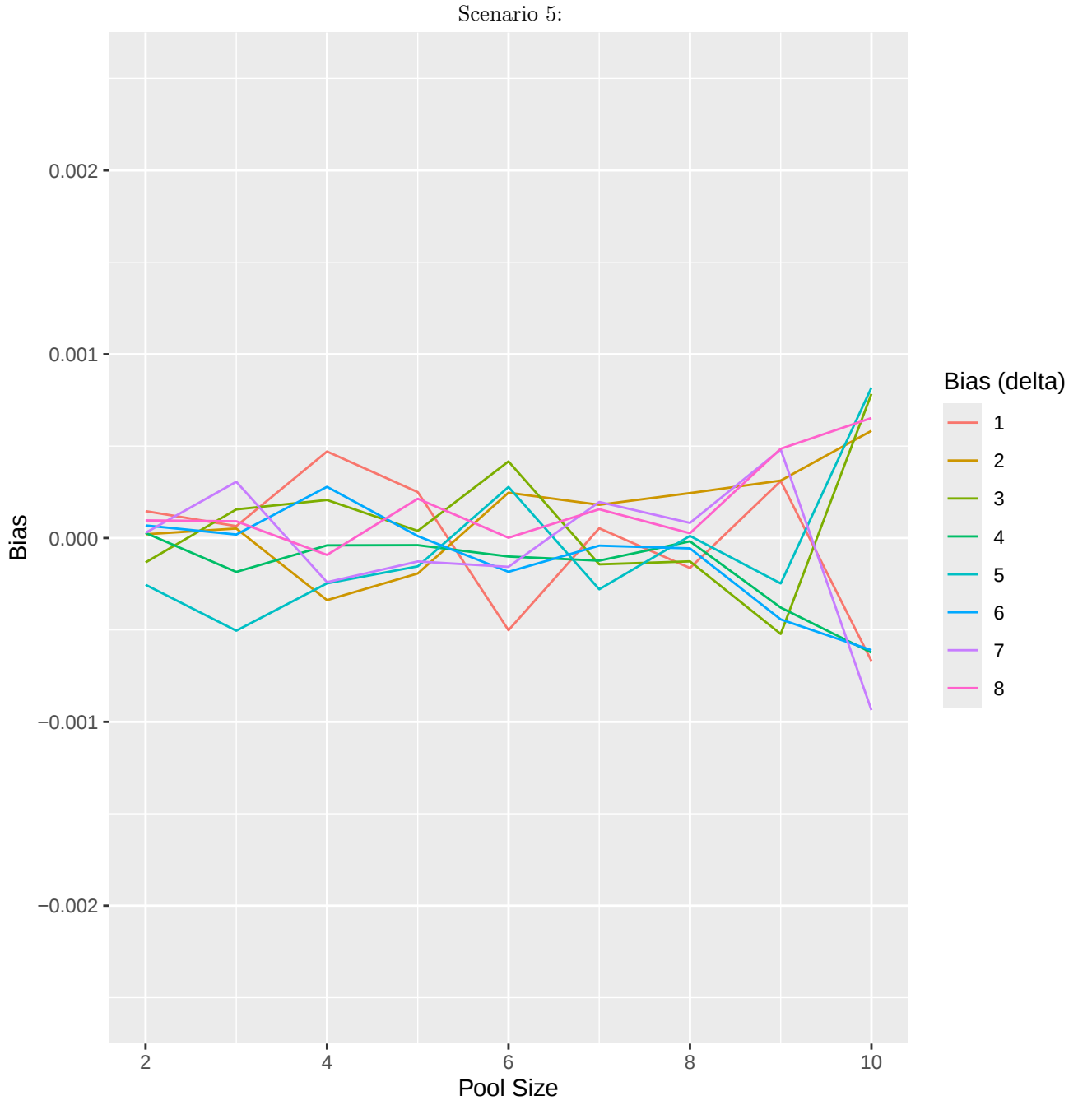
Web Fig. 2: The figure displays the average empirical bias of the estimator $\hat{\delta}$ from scenario 2 as a function of pool size. From 1 to 8, the position numbers correspond to $\mathbf{y} = (0, 0, 0), (1, 0, 0), (0, 1, 0), (1, 0, 1), (1, 1, 0), (0, 1, 1)$ and $(1, 1, 1)$. This is an enlarged version of Figure 1(b).



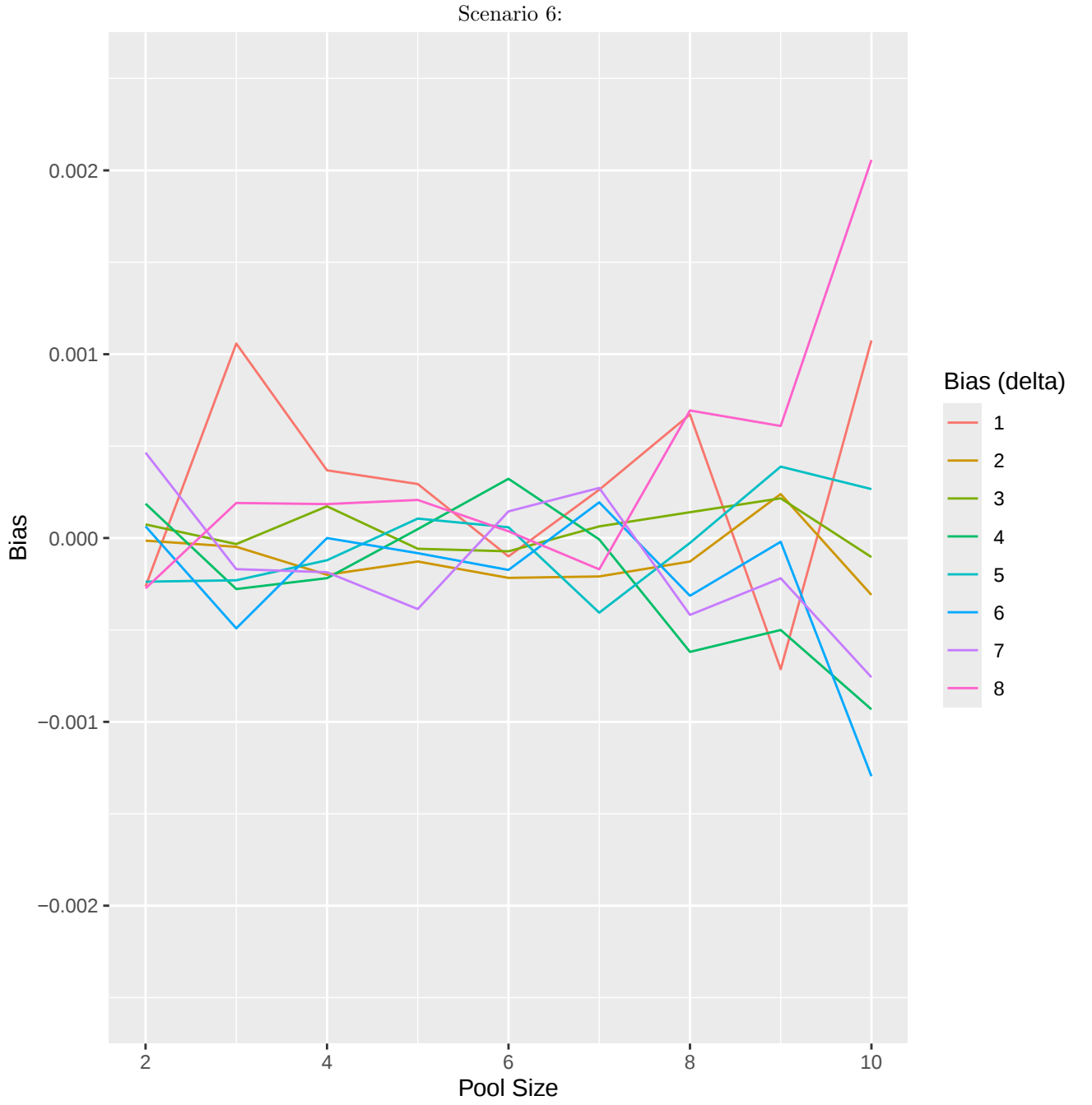
Web Fig. 3: The figure displays the average empirical bias of the estimator $\hat{\delta}$ from scenario 3 as a function of pool size. From 1 to 8, the position numbers correspond to $\mathbf{y} = (0, 0, 0), (1, 0, 0), (0, 1, 0), (1, 0, 1), (1, 1, 0), (0, 1, 1)$ and $(1, 1, 1)$. This is an enlarged version of Figure 1(c).



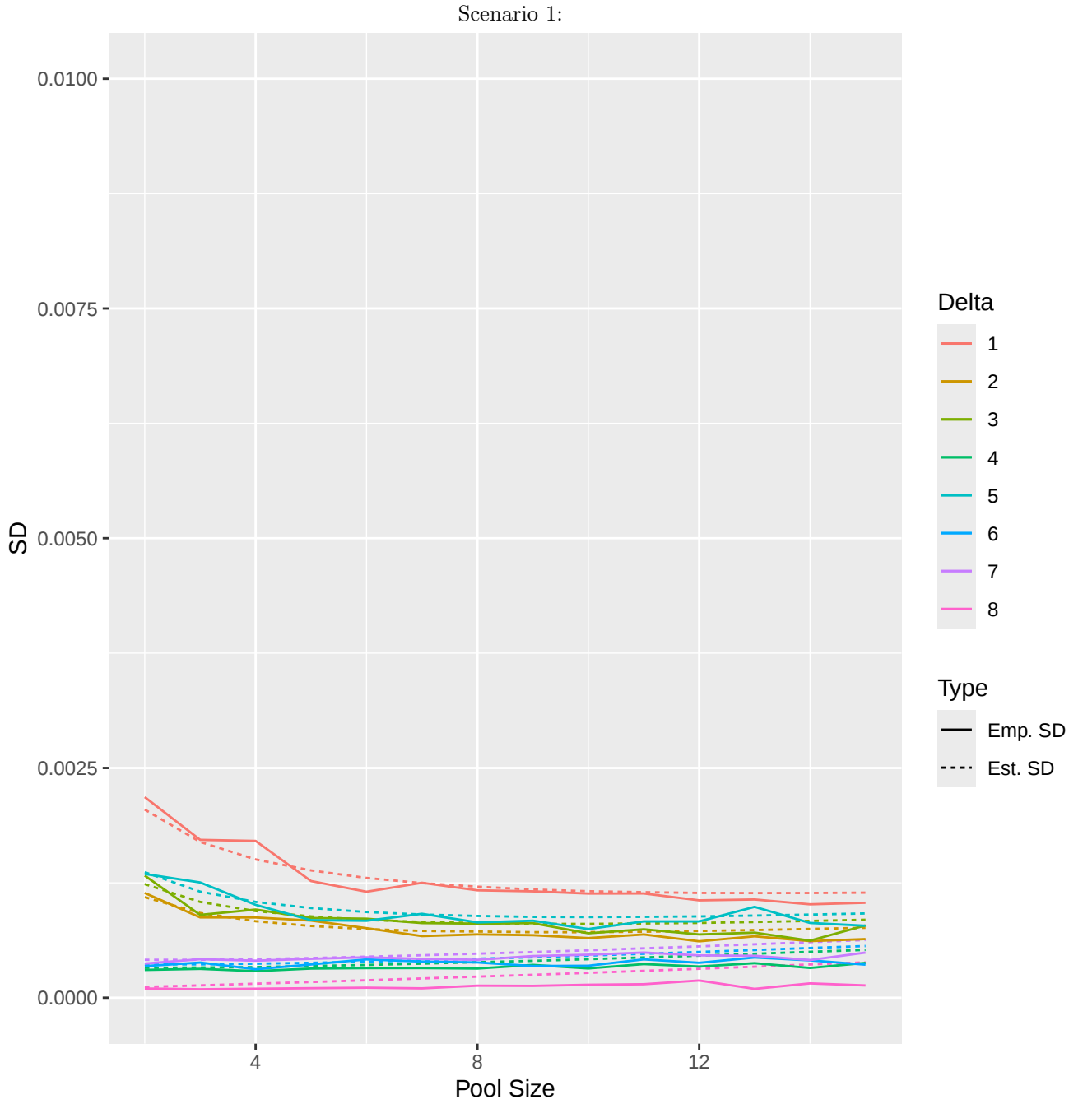
Web Fig. 4: The figure displays the average empirical bias of the estimator $\hat{\delta}$ from scenario 4 as a function of pool size. From 1 to 8, the position numbers correspond to $\mathbf{y} = (0, 0, 0), (1, 0, 0), (0, 1, 0), (1, 0, 1), (1, 1, 0), (0, 1, 1)$ and $(1, 1, 1)$. This is an enlarged version of Figure 1(d).



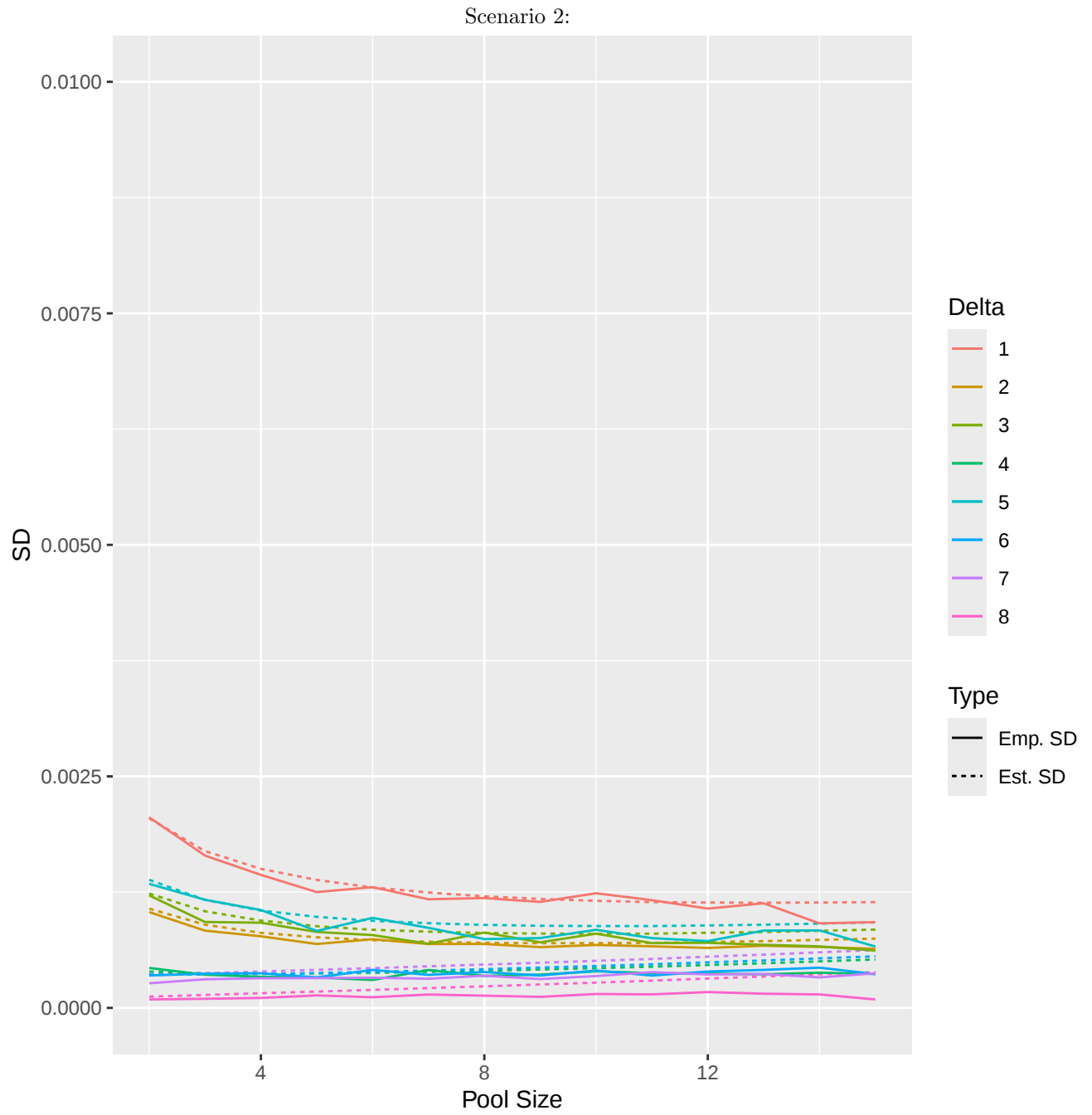
Web Fig. 5: The figure displays the average empirical bias of the estimator $\hat{\delta}$ from scenario 5 as a function of pool size. From 1 to 8, the position numbers correspond to $\mathbf{y} = (0, 0, 0), (1, 0, 0), (0, 1, 0), (1, 0, 1), (1, 1, 0), (0, 1, 1)$ and $(1, 1, 1)$. This is an enlarged version of Figure 1(e).



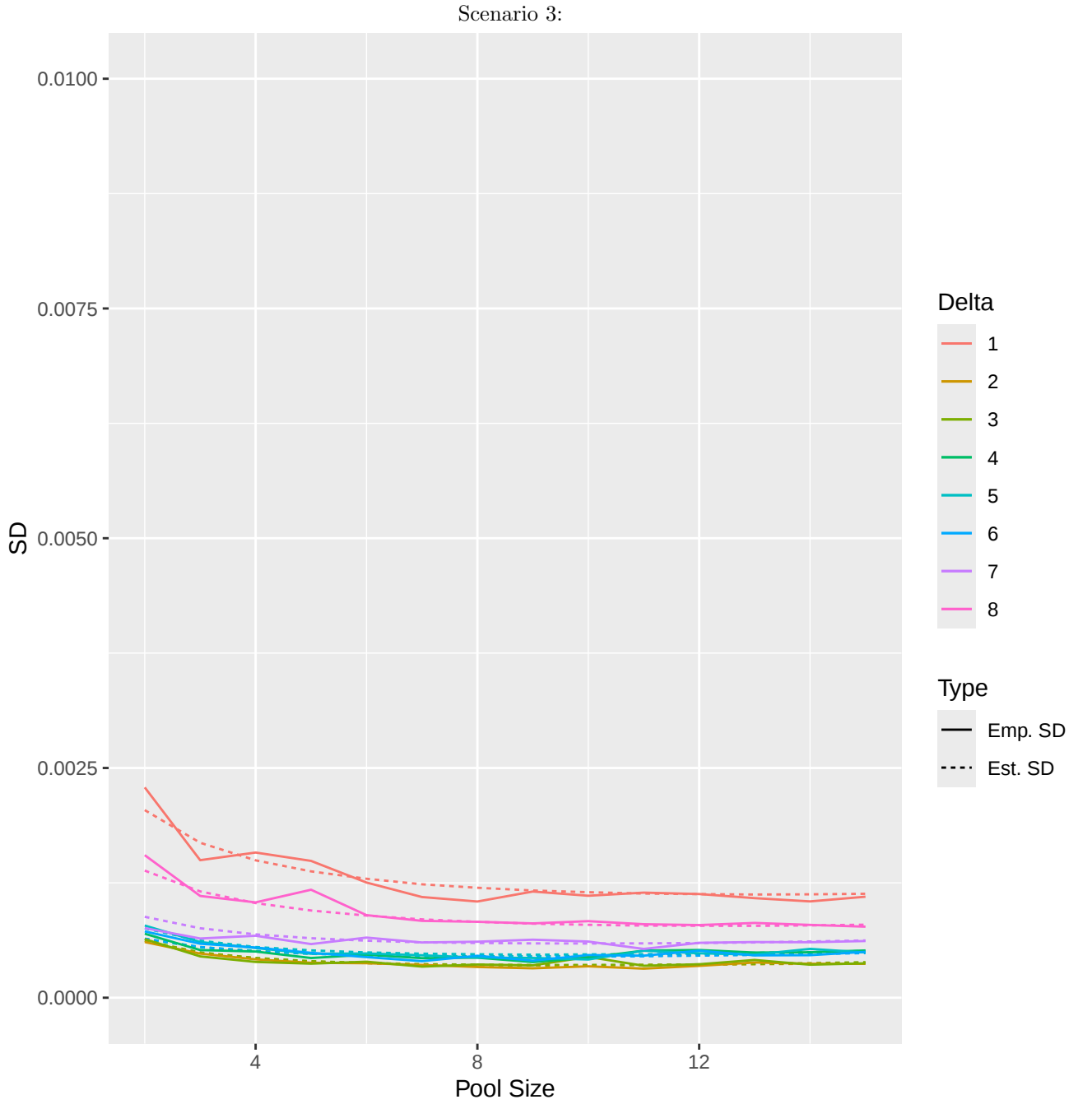
Web Fig. 6: The figure displays the average empirical bias of the estimator $\hat{\delta}$ from scenario 6 as a function of pool size. From 1 to 8, the position numbers correspond to $\mathbf{y} = (0, 0, 0), (1, 0, 0), (0, 1, 0), (1, 0, 1), (1, 1, 0), (0, 1, 1)$ and $(1, 1, 1)$. This is an enlarged version of Figure 1(f).



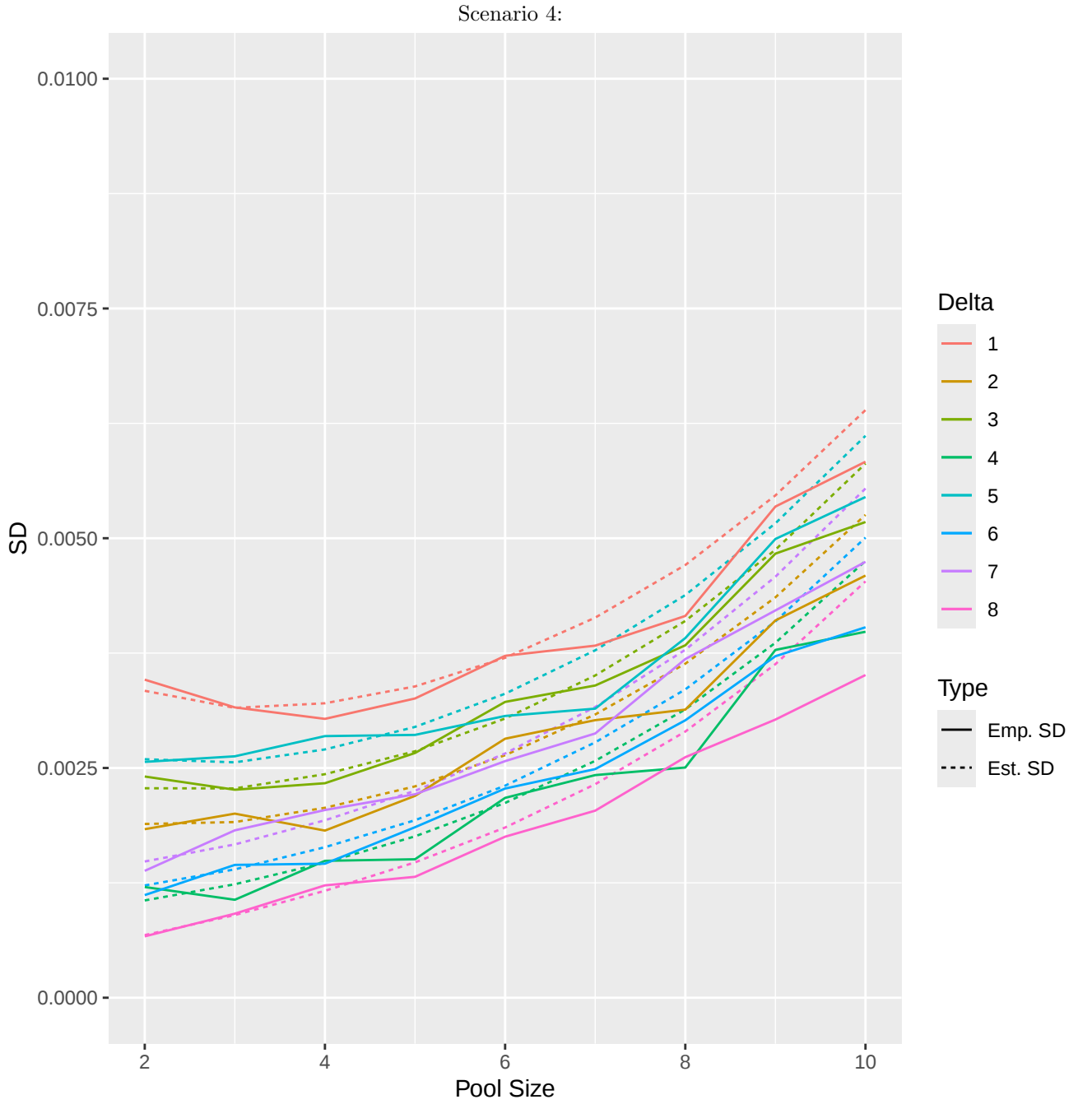
Web Fig. 7: The figure displays the average estimated standard error from Louis's method and the standard deviation of the sample of point estimators for $\hat{\delta}$ from scenario 1 as a function of pool size. From 1 to 8, the position numbers correspond to $\mathbf{y} = (0, 0, 0), (1, 0, 0), (0, 1, 0), (1, 0, 1), (1, 1, 0), (0, 1, 1)$ and $(1, 1, 1)$. This is an enlarged version of Figure 2(a).



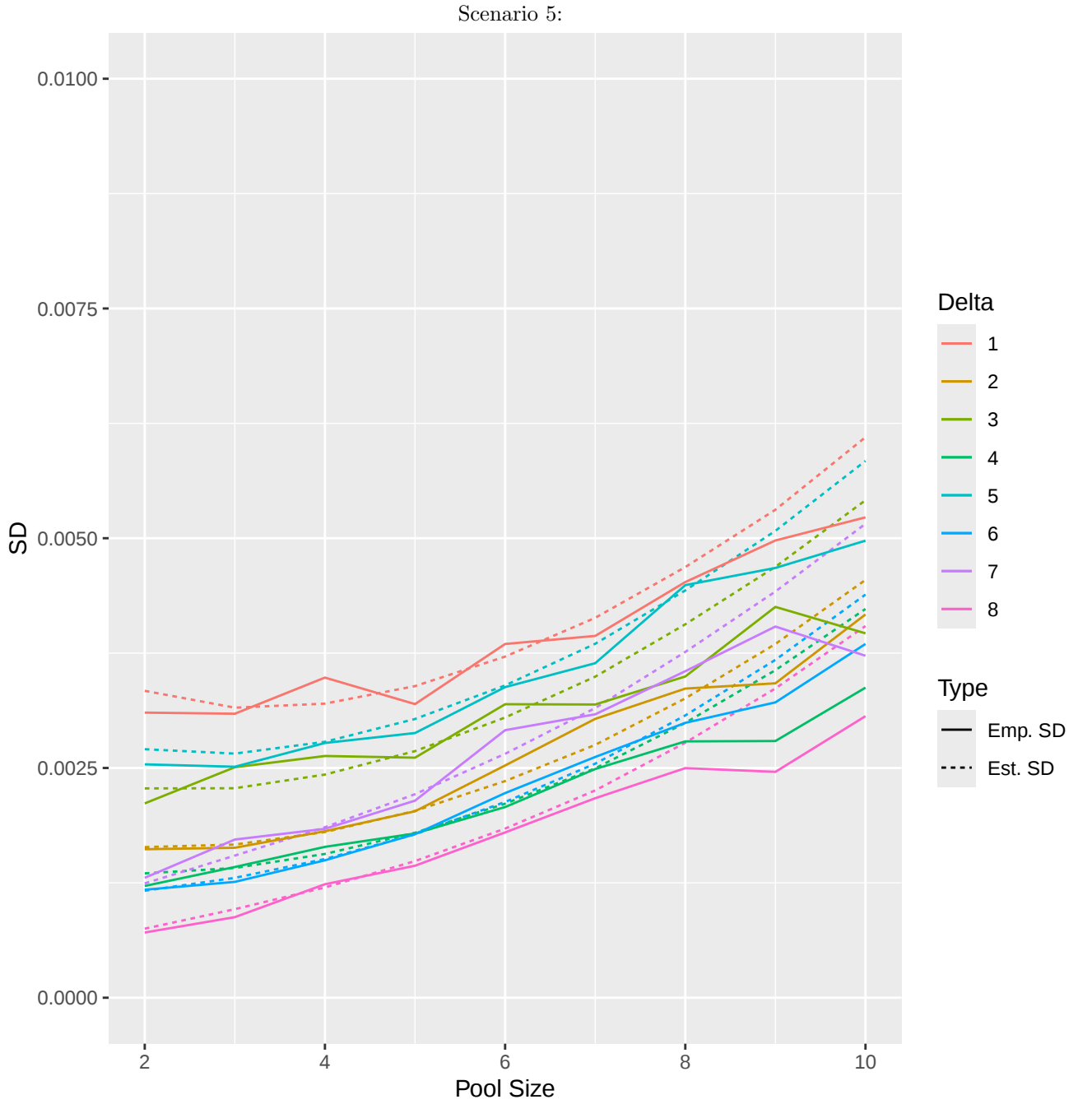
Web Fig. 8: The figure displays the average estimated standard error from Louis's method and the standard deviation of the sample of point estimators for $\hat{\delta}$ from scenario 2 as a function of pool size. From 1 to 8, the position numbers correspond to $\mathbf{y} = (0, 0, 0), (1, 0, 0), (0, 1, 0), (1, 0, 1), (1, 1, 0), (0, 1, 1)$ and $(1, 1, 1)$. This is an enlarged version of Figure 2(b).



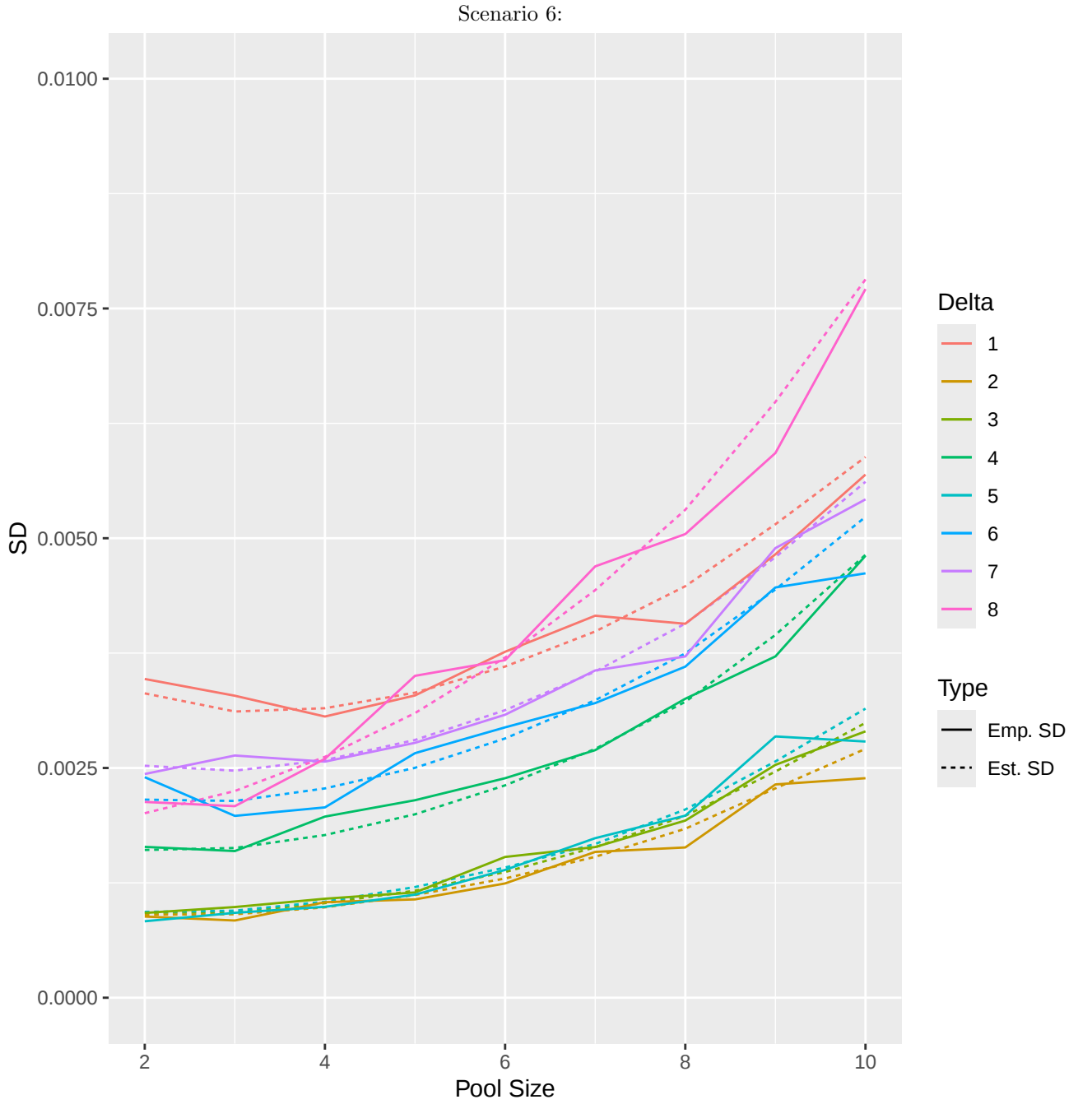
Web Fig. 9: The figure displays the average estimated standard error from Louis's method and the standard deviation of the sample of point estimators for $\hat{\delta}$ from scenario 3 as a function of pool size. From 1 to 8, the position numbers correspond to $\mathbf{y} = (0, 0, 0), (1, 0, 0), (0, 1, 0), (1, 0, 1), (1, 1, 0), (0, 1, 1)$ and $(1, 1, 1)$. This is an enlarged version of Figure 2(c).



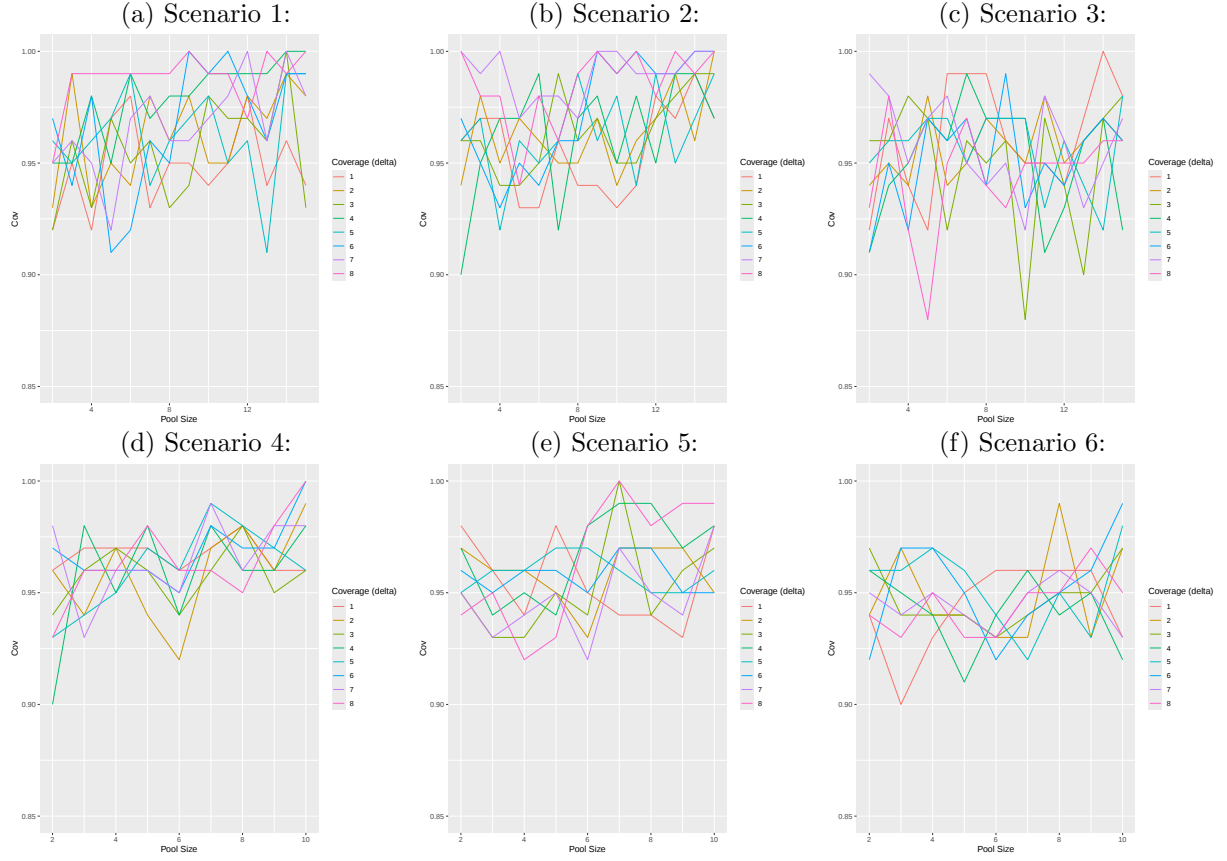
Web Fig. 10: The figure displays the average estimated standard error from Louis's method and the standard deviation of the sample of point estimators for $\hat{\delta}$ from scenario 4 as a function of pool size. From 1 to 8, the position numbers correspond to $\mathbf{y} = (0, 0, 0), (1, 0, 0), (0, 1, 0), (1, 0, 1), (1, 1, 0), (0, 1, 1)$ and $(1, 1, 1)$. This is an enlarged version of Figure 2(d).



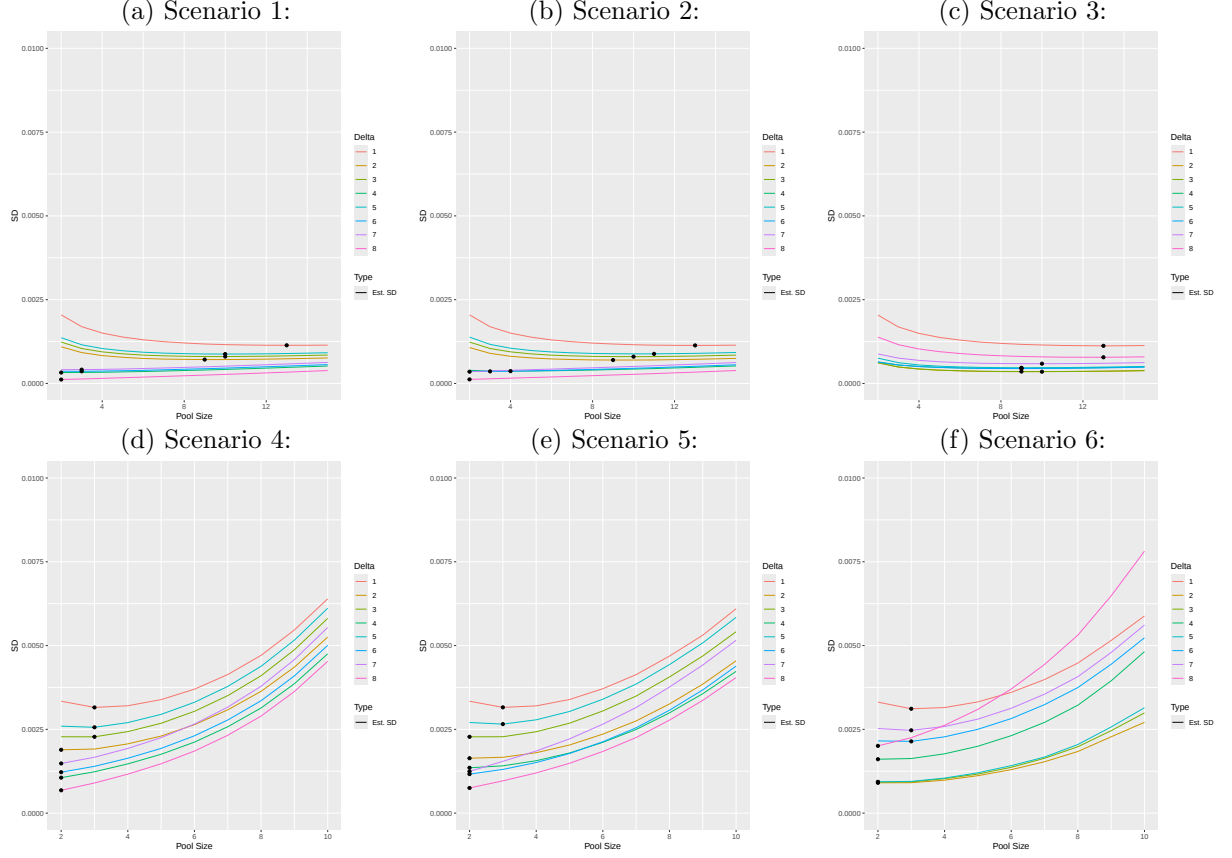
Web Fig. 11: The figure displays the average estimated standard error from Louis's method and the standard deviation of the sample of point estimators for $\hat{\delta}$ from scenario 5 as a function of pool size. From 1 to 8, the position numbers correspond to $\mathbf{y} = (0, 0, 0), (1, 0, 0), (0, 1, 0), (1, 0, 1), (1, 1, 0), (0, 1, 1)$ and $(1, 1, 1)$. This is an enlarged version of Figure 2(e).



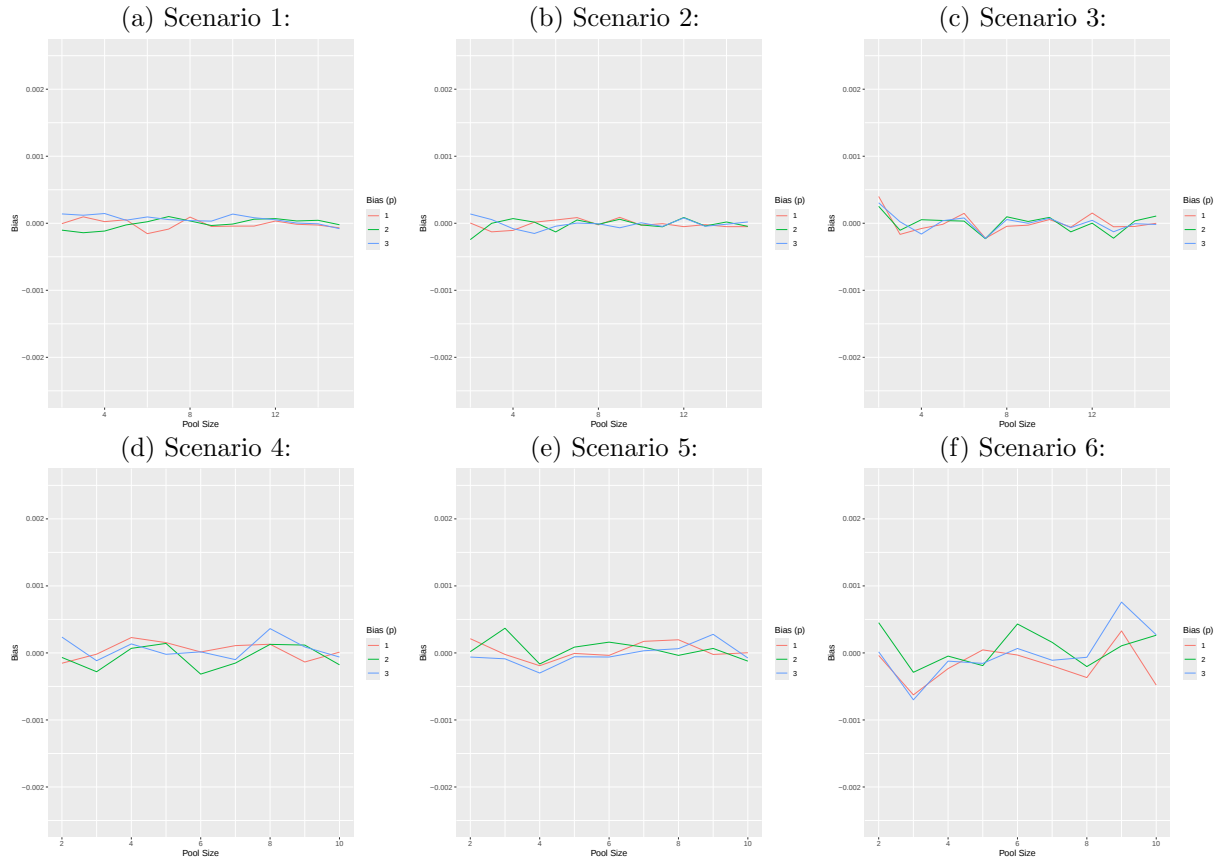
Web Fig. 12: The figure displays the average estimated standard error from Louis's method and the standard deviation of the sample of point estimators for $\hat{\delta}$ from scenario 6 as a function of pool size. From 1 to 8, the position numbers correspond to $\mathbf{y} = (0, 0, 0), (1, 0, 0), (0, 1, 0), (1, 0, 1), (1, 1, 0), (0, 1, 1)$ and $(1, 1, 1)$. This is an enlarged version of Figure 2(f).



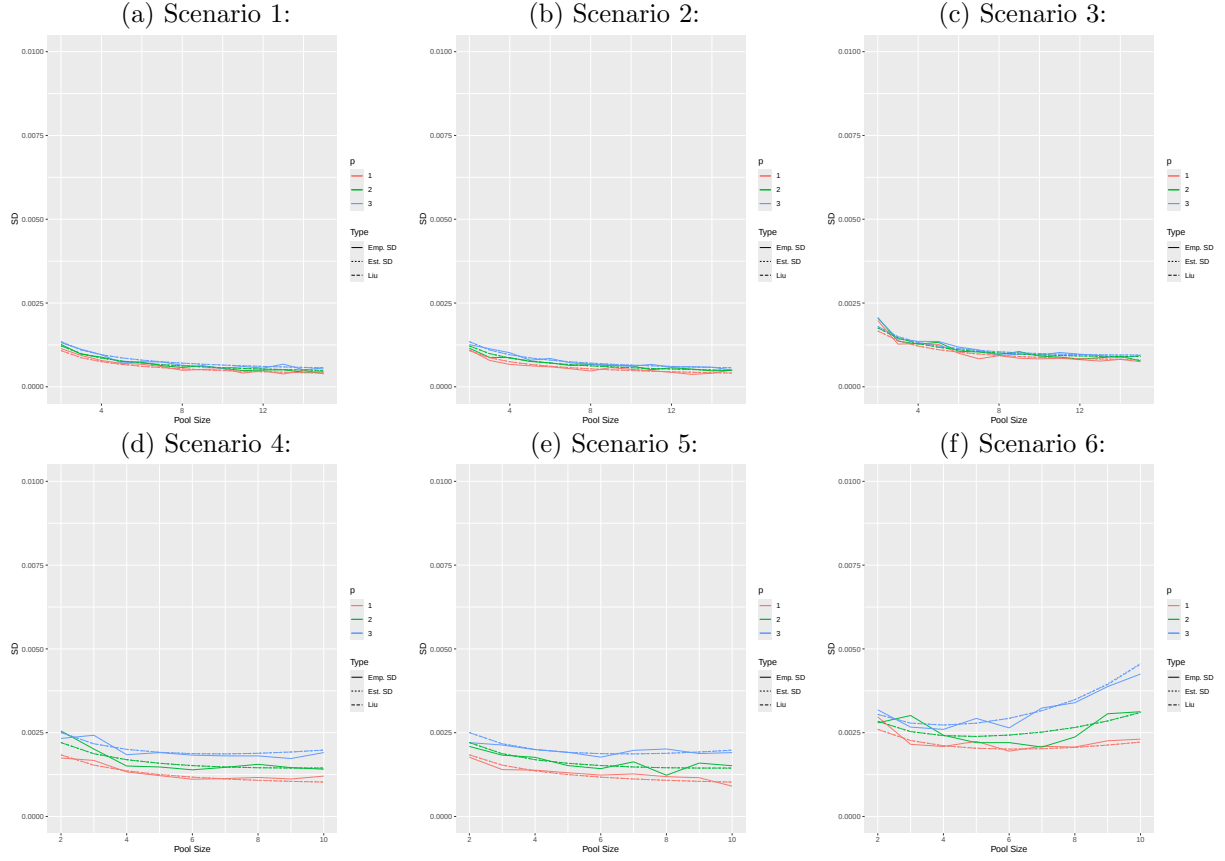
Web Fig. 13: The figure displays the empirical coverage probabilities for 95% confidence intervals for $\hat{\delta}$ as a function of pool size. From 1 to 8, the position numbers correspond to $\mathbf{y} = (0, 0, 0), (1, 0, 0), (0, 1, 0), (1, 0, 1), (1, 1, 0), (0, 1, 1)$ and $(1, 1, 1)$. The figure panels correspond to scenarios 1-6 from top left (1) to bottom right (6).



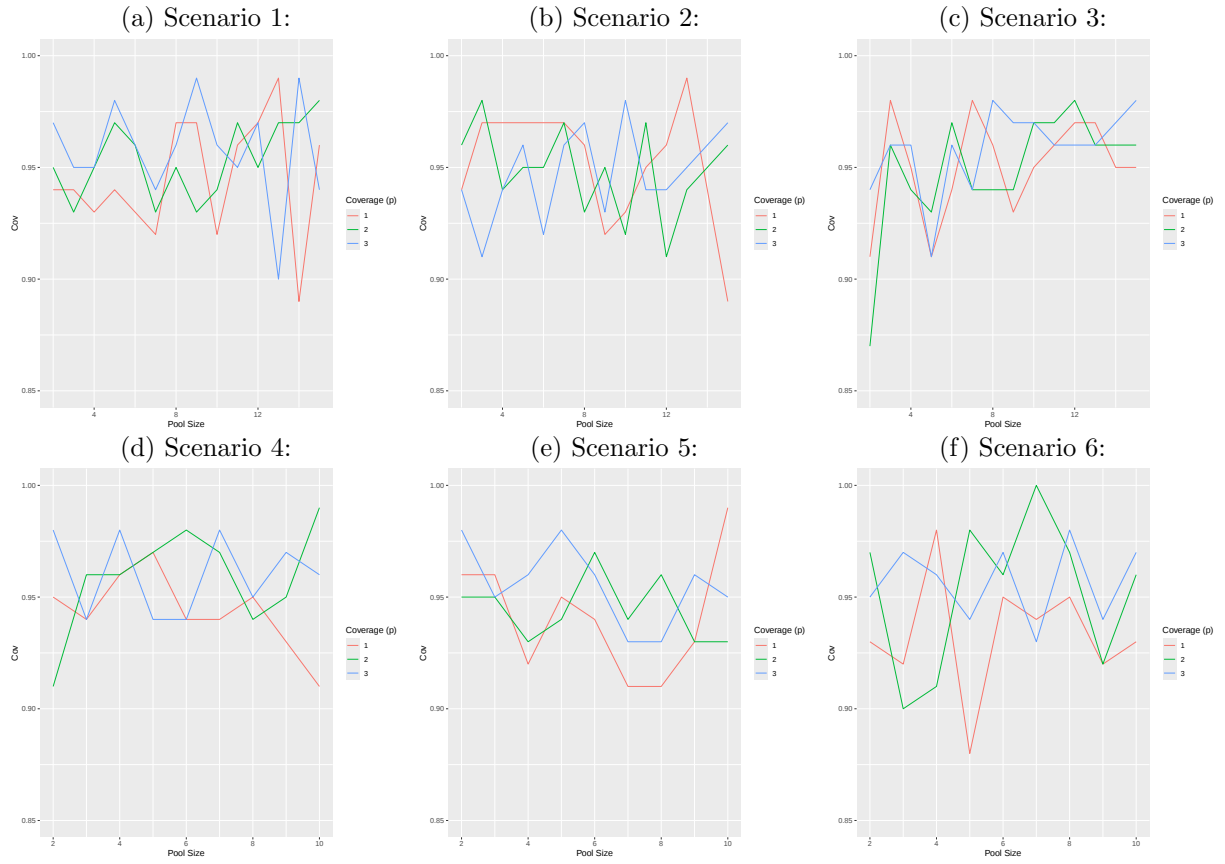
Web Fig. 14: The figure displays the average standard errors from Louis's method for $\hat{\delta}$ as a function of pool size. Black dots indicate the pool size which minimizes the standard error. From 1 to 8, the position numbers correspond to $\mathbf{z} = (0, 0, 0), (1, 0, 0), (0, 1, 0), (1, 1, 0), (0, 0, 1), (1, 0, 1), (0, 1, 1)$ and $(1, 1, 1)$. The figure panels correspond to scenarios 1-6 from top left (1) to bottom right (6).



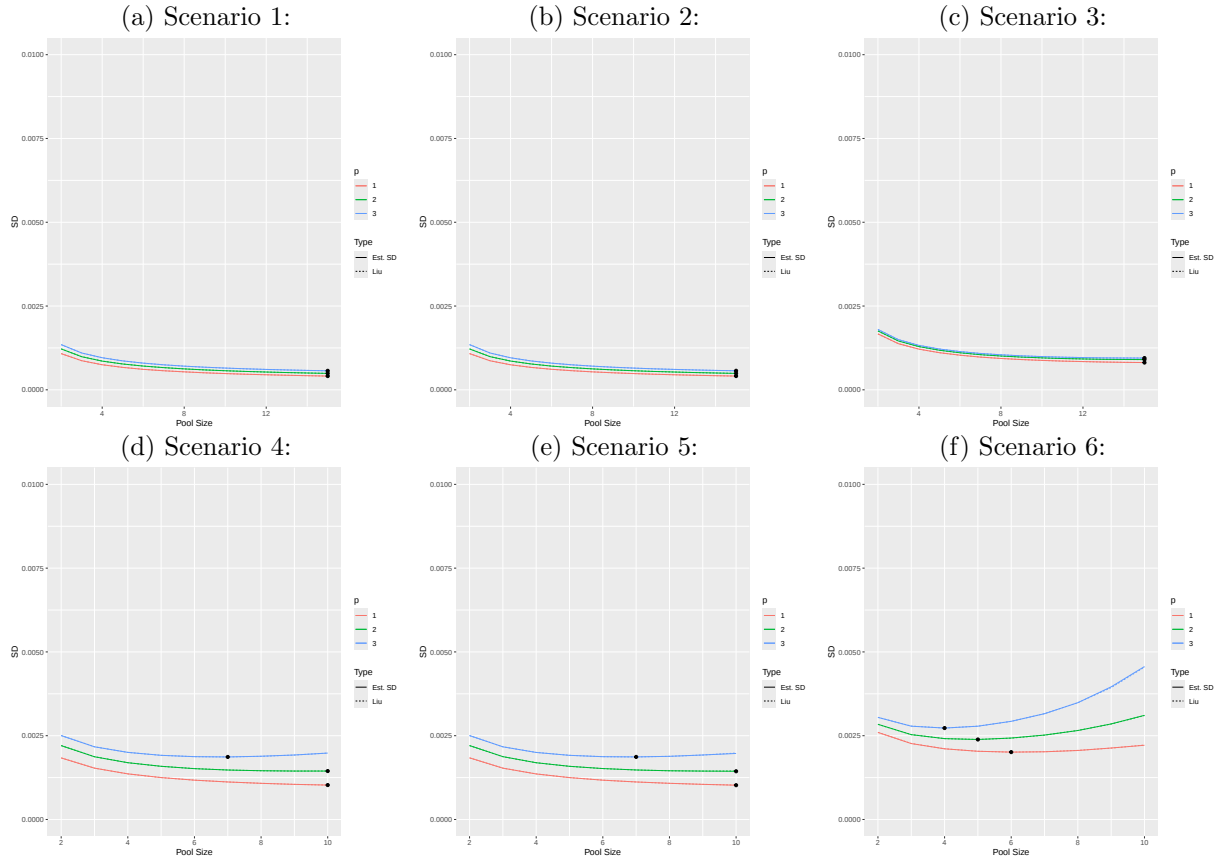
Web Fig. 15: The figure displays the average empirical bias of the estimator $\hat{p}$ as a function of pool size. The figure panels correspond to scenarios 1-6 from top left (1) to bottom right (6).



Web Fig. 16: The figure displays the average estimated standard error from Louis's method, the sample standard deviation of the point estimators for $\hat{\delta}$ as a function of pool size and the standard error calculated using Liu et. al's method for singlexplex assays. From 1 to 8, the position numbers correspond to $\mathbf{z} = (0, 0, 0)$, $(1, 0, 0)$, $(0, 1, 0)$, $(1, 0, 1)$, $(1, 1, 0)$, $(0, 1, 1)$ and $(1, 1, 1)$. The figure panels correspond to scenarios 1-6 from top left (1) to bottom right (6).



Web Fig. 17: The figure displays the empirical coverage probabilities for 95% confidence intervals for $\hat{p}$ as a function of pool size. The figure panels correspond to scenarios 1-6 from top left (1) to bottom right (6).



Web Fig. 18: The figure displays average estimated standard error from Louis's method for $\hat{p}$ as a function of pool size. Black dots indicate the pool sizes which minimizes the standard error. The figure panels correspond to scenarios 1-6 from top left (1) to bottom right (6).