# Supporting Information for "A mixed-effects Bayesian regression model for multivariate group testing data"

**Christopher S. McMahan**[1,*], **Chase N. Joyner**[1], **Joshua M. Tebbs**[2], **and Christopher R. Bilder**[3]

[1]School of Mathematical and Statistical Sciences, Clemson University, Clemson, SC 29634, U.S.A.

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

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

[*]*email*: mcmaha2@clemson.edu

**Web Appendix A:** *Full conditional distributions, derivations, and expressions.* We derive the full conditional distributions below and give expressions for the parameters in these distributions:

$$\begin{aligned}
\widetilde{Y}_{id} \mid \widetilde{\mathbf{Y}}_{i(-d)}, \mathbf{Z}, \boldsymbol{\Theta} &\sim \text{Bernoulli}(p^*_{id}) \\
\boldsymbol{\omega}_i \mid \widetilde{\mathbf{Y}}_i, \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{a}, \mathbf{b}_{(i)}, \mathbf{R} &\sim \text{TMN}(\boldsymbol{\eta}_i, \mathbf{R}, \mathbf{L}_i, \mathbf{U}_i) \\
\boldsymbol{\beta}_{\boldsymbol{v}} \mid \boldsymbol{\omega}, \boldsymbol{\lambda}, \mathbf{a}, \mathbf{b}, \mathbf{R}, \boldsymbol{v} &\sim N(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) \\
\lambda_{ld} \mid \boldsymbol{\omega}, \boldsymbol{\beta}, \boldsymbol{\lambda}_{(-\ell)}, \mathbf{a}, \mathbf{b}, \mathbf{R}, w_{ld} &\sim \text{TN}(\mu_{\lambda_{ld}} w_{ld}, \sigma^2_{\lambda_{ld}} w_{ld}, 0, \infty) \\
\mathbf{a} \mid \boldsymbol{\omega}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b}, \mathbf{R} &\sim N(\boldsymbol{\mu}_{\mathbf{a}}, \boldsymbol{\Sigma}_{\mathbf{a}}) \\
\mathbf{b}_k \mid \boldsymbol{\omega}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{a}, \mathbf{R} &\sim N(\boldsymbol{\mu}_{\mathbf{b}_k}, \boldsymbol{\Sigma}_{\mathbf{b}_k}) \\
v_{rd} \mid \boldsymbol{\omega}, \boldsymbol{\lambda}, \mathbf{a}, \mathbf{b}, \mathbf{R}, \boldsymbol{v}_{(-rd)}, \tau_{v_{rd}} &\sim \text{Bernoulli}(p_{v_{rd}}) \\
w_{ld} \mid \boldsymbol{\omega}, \boldsymbol{\beta}, \boldsymbol{\lambda}_{(-\ell)}, \mathbf{a}, \mathbf{b}, \tau_{w_{ld}} &\sim \text{Bernoulli}(p_{w_{ld}}) \\
\tau_{v_{rd}} \mid v_{rd} &\sim \text{beta}(a_v + v_{rd}, b_v + 1 - v_{rd}) \\
\tau_{w_{ld}} \mid w_{ld} &\sim \text{beta}(a_w + w_{rd}, b_w + 1 - w_{rd}) \\
S_{e(m):d} \mid \mathbf{Z}, \widetilde{\mathbf{Y}} &\sim \text{beta}(a^\star_{e(m):d}, b^\star_{e(m):d}) \\
S_{p(m):d} \mid \mathbf{Z}, \widetilde{\mathbf{Y}} &\sim \text{beta}(a^\star_{p(m):d}, b^\star_{p(m):d}).
\end{aligned}$$

We henceforth make use of the following notation: $\mathbf{X}_i = \oplus_{d=1}^D \mathbf{x}'_{id}$, $\mathbf{T}_i = \oplus_{d=1}^D \mathbf{t}'_{id}$, $\boldsymbol{\Lambda} = \oplus_{d=1}^D \boldsymbol{\Lambda}_d$, $\mathbf{A} = \oplus_{d=1}^D \mathbf{A}_d$, $\boldsymbol{v} = (\boldsymbol{v}'_1, ..., \boldsymbol{v}'_D)'$, and $\boldsymbol{v}_d = (v_{1d}, ..., v_{p_d d})'$.

Full conditional of $\widetilde{Y}_{id}$: From the joint distribution of the observed testing outcomes and the individuals' latent statuses, given by

$$\pi(\mathbf{Z}, \widetilde{\mathbf{Y}} \mid \boldsymbol{\Theta}) = \prod_{d=1}^{D} \prod_{m=1}^{M} \prod_{j \in \mathcal{I}_m} \left\{ S^{Z_{jd}}_{e(m):d} (1 - S_{e(m):d})^{1 - Z_{jd}} \right\}^{\widetilde{Z}_{jd}} \left\{ S^{1 - Z_{jd}}_{p(m):d} (1 - S_{p(m):d})^{Z_{jd}} \right\}^{1 - \widetilde{Z}_{jd}} \times \prod_{i=1}^{N} \pi(\widetilde{\mathbf{Y}}_i \mid \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{a}, \mathbf{b}_{(i)}, \mathbf{R}),$$

it is easy to see that the full conditional distribution of $\widetilde{Y}_{id}$ is Bernoulli. In particular, $\widetilde{Y}_{id} \mid \widetilde{\mathbf{Y}}_{i(-d)}, \mathbf{Z}, \boldsymbol{\Theta} \sim \text{Bernoulli}(p^*_{id})$, where $\widetilde{\mathbf{Y}}_{i(-d)}$ is the vector $\widetilde{\mathbf{Y}}_i$ with the $d$th element removed,

$p^*_{id} = p^\star_{id1}/(p^\star_{id0} + p^\star_{id1})$, and

$$\begin{aligned} p^\star_{id1} &= p_{id} \prod_{j \in \mathcal{A}_i} S_{e_j:d}^{Z_{jd}} (1 - S_{e_j:d})^{1-Z_{jd}} \\ p^\star_{id0} &= (1 - p_{id}) \prod_{j \in \mathcal{A}_i} \left\{ S_{e_j:d}^{Z_{jd}} (1 - S_{e_j:d})^{1-Z_{jd}} \right\}^{I\left(s_{ijd}>0\right)} \left\{ (1 - S_{p_j:d})^{Z_{jd}} S_{p_j:d}^{1-Z_{jd}} \right\}^{I(s_{ijd}=0)}. \end{aligned}$$

In the expression above, $p_{id} = \pi(\widetilde{\mathbf{Y}}_{i(d)} \mid \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{a}, \mathbf{b}_{(i)}, \mathbf{R})$, $\widetilde{\mathbf{Y}}_{i(d)} = (\widetilde{Y}_{i1}, ..., \widetilde{Y}_{id} = 1, ..., \widetilde{Y}_{iD})'$, the index set $\mathcal{A}_i = \{j \colon i \in \mathcal{P}_j\}$ keeps track of which pools the $i$th individual belongs to, and $s_{ijd} = \sum_{i' \in \mathcal{P}_j \colon i' \neq i} \widetilde{Y}_{i'd}$. If $j \in \mathcal{I}_m$, then $S_{e_j:d} = S_{e(m):d}$ and $S_{p_j:d} = S_{p(m):d}$.

Full conditional of $\boldsymbol{\omega}_i$: From the joint distribution

$$\begin{aligned} \pi(\mathbf{Z}, \widetilde{\mathbf{Y}}, \boldsymbol{\omega} \mid \boldsymbol{\Theta}) \propto \prod_{d=1}^{D} \prod_{m=1}^{M} \prod_{j \in \mathcal{I}_m} & \left\{ S_{e(m):d}^{Z_{jd}} (1 - S_{e(m):d})^{1-Z_{jd}} \right\}^{\widetilde{Z}_{jd}} \left\{ S_{p(m):d}^{1-Z_{jd}} (1 - S_{p(m):d})^{Z_{jd}} \right\}^{1-\widetilde{Z}_{jd}} \\ & \times \prod_{i=1}^{N} |\mathbf{R}|^{-1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{\omega}_i - \boldsymbol{\eta}_i)' \mathbf{R}^{-1} (\boldsymbol{\omega}_i - \boldsymbol{\eta}_i) \right\} \prod_{i=1}^{N} f(\boldsymbol{\omega}_i), \end{aligned}$$

one can see the full conditional distribution of $\boldsymbol{\omega}_i$ is multivariate truncated normal with mean $\boldsymbol{\eta}_i$, covariance matrix $\mathbf{R}$, lower truncation limits $\mathbf{L}_i = (L_{i1}, ..., L_{iD})'$, and upper truncation limits $\mathbf{U}_i = (U_{i1}, ..., U_{iD})'$. The truncation region for the $d$th dimension is $L_{id} = 0$ and $U_{id} = \infty$ if $\widetilde{Y}_{id} = 1$ and $L_{id} = -\infty$ and $U_{id} = 0$ if $\widetilde{Y}_{id} = 0$; i.e.,

$$\boldsymbol{\omega}_i \mid \widetilde{\mathbf{Y}}_i, \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{a}, \mathbf{b}_{(i)}, \mathbf{R} \sim \text{TMN}(\boldsymbol{\eta}_i, \mathbf{R}, \mathbf{L}_i, \mathbf{U}_i).$$

Full conditional of $\boldsymbol{\beta}$: The full conditional distribution of $\beta_{rd}$ is degenerate at 0 if $v_{rd} = 0$, while the nonzero elements of $\boldsymbol{\beta}$, say $\boldsymbol{\beta}_{\boldsymbol{v}}$, have the following normal full conditional distribution

$$\boldsymbol{\beta}_{\boldsymbol{v}} \mid \boldsymbol{\omega}, \boldsymbol{\lambda}, \mathbf{a}, \mathbf{b}, \mathbf{R}, \boldsymbol{v}, \sim N(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}).$$

The mean and covariance matrix are

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{\beta}} &= \left\{ \boldsymbol{\Phi}(\boldsymbol{v})^{-1} + \sum_{i=1}^{N} \mathbf{X}_i(\boldsymbol{v})' \mathbf{R}^{-1} \mathbf{X}_i(\boldsymbol{v}) \right\}^{-1} \sum_{i=1}^{N} \mathbf{X}_i(\boldsymbol{v})' \mathbf{R}^{-1} \boldsymbol{\omega}^\star_{\boldsymbol{\beta} i} \\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}} &= \left\{ \boldsymbol{\Phi}(\boldsymbol{v})^{-1} + \sum_{i=1}^{N} \mathbf{X}_i(\boldsymbol{v})' \mathbf{R}^{-1} \mathbf{X}_i(\boldsymbol{v}) \right\}^{-1}, \end{aligned}$$

where $\boldsymbol{\Phi}(\boldsymbol{v})$ is the matrix that is formed by retaining the rows and columns of $\boldsymbol{\Phi} = \text{diag}(\phi^2_{rd}; r = 1, ..., p_d, d = 1, ..., D)$ that correspond to the non-zero elements of $\boldsymbol{v}$. Also, $\mathbf{X}_i(\boldsymbol{v})$ is the matrix that is formed by retaining the columns of $\mathbf{X}_i$ corresponding to the non-zero elements of $\boldsymbol{v}$, and $\boldsymbol{\omega}^\star_{\boldsymbol{\beta} i} = \boldsymbol{\omega}_i - \mathbf{T}_i \boldsymbol{\Lambda} \mathbf{A} \mathbf{b}_{(i)}$.

Full conditional of $\lambda_{ld}$: We introduce new notation. For the $i$th individual, define a $q_d \times 1$ vector $\mathbf{e}_{id}$ whose $l$th element is $t_{idl} b_{(i)dl} + t_{idl} \sum_{m=1}^{l-1} b_{(i)dm} a_{dlm}$, where $t_{idl}$ is the $l$th element of $\mathbf{t}_{id}$,

$b_{(i)dl}$ is the $l$th element of $\mathbf{b}_{(i)d}$, and $a_{dlm}$ is the $(l,m)$th entry of $\mathbf{A}_d$. Construct $\mathbf{E}_i = \oplus_{d=1}^{D} \mathbf{e}'_{id}$. Based on this notation, we can succinctly express the full conditional distribution of $\lambda_{ld}$, which is the $\ell$th element of $\boldsymbol{\lambda}$. In particular, the full conditional of $\lambda_{ld}$ is degenerate at 0 if $w_{ld} = 0$. When $w_{ld} = 1$, the full conditional is

$$\lambda_{ld} \mid \boldsymbol{\omega}, \boldsymbol{\beta}, \boldsymbol{\lambda}_{(-\ell)}, \mathbf{a}, \mathbf{b}, \mathbf{R}, w_{ld} \sim \text{TN}(\mu_{\lambda_{ld}}, \sigma^2_{\lambda_{ld}}, 0, \infty),$$

where

$$\begin{aligned} \mu_{\lambda_{ld}} &= \left( \frac{1}{\boldsymbol{\Psi}_{\ell\ell}} + \sum_{i=1}^{N} \mathbf{E}_i^{\ell\prime} \mathbf{R}^{-1} \mathbf{E}_i^{\ell} \right)^{-1} \sum_{i=1}^{N} \mathbf{E}_i^{\ell\prime} \mathbf{R}^{-1} \boldsymbol{\omega}^{\star}_{\lambda_\ell i} \\ \sigma^2_{\lambda_{ld}} &= \left( \frac{1}{\boldsymbol{\Psi}_{\ell\ell}} + \sum_{i=1}^{N} \mathbf{E}_i^{\ell\prime} \mathbf{R}^{-1} \mathbf{E}_i^{\ell} \right)^{-1}. \end{aligned}$$

In the expressions above, $\mathbf{E}_i^{\ell}$ is the $\ell$th column of $\mathbf{E}_i$, $\boldsymbol{\Psi}_{\ell\ell}$ is the $\ell$th diagonal element of $\boldsymbol{\Psi} = \text{diag}(\psi^2_{ld}; l = 1, ..., q_d, d = 1, ..., D)$, $\boldsymbol{\omega}^{\star}_{\lambda_\ell i} = \boldsymbol{\omega}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{E}_i^{(\ell)}\boldsymbol{\lambda}_{(-\ell)}$, $\mathbf{E}_i^{(-\ell)}$ is the matrix that remains after removing the $\ell$th column of $\mathbf{E}_i$, and $\boldsymbol{\lambda}_{(-\ell)}$ is the vector that remains after removing $\lambda_{ld}$ from $\boldsymbol{\lambda}$.

Full conditional of $\mathbf{a}$: We introduce new notation. Define the $q_d \times (q_d - 1)/2$ vector $\mathbf{u}_{id} = (b_{(i)dl}\lambda_{dm}t_{idm}; l = 1, ..., q_d - 1, m = l + 1, ..., q_d)'$ and construct $\mathbf{U}_i = \oplus_{d=1}^{D} \mathbf{u}'_{id}$, where $b_{(i)dl}$ is the $l$th element of $\mathbf{b}_{(i)d}$, $\lambda_{dm}$ is the $m$th element of $\boldsymbol{\lambda}_d$, and $t_{idm}$ is the $m$th element of $\mathbf{t}_{id}$. The linear predictor in our model can then be re-expressed as $\eta_{id} = \mathbf{x}'_{id}\boldsymbol{\beta} + \mathbf{t}'_{id}\boldsymbol{\Lambda}_d\mathbf{b}_{(i)d} + \mathbf{u}'_{id}\mathbf{a}_d$. It is easy to see the full conditional distribution of $\mathbf{a}$ is

$$\mathbf{a} \mid \boldsymbol{\omega}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{b}, \mathbf{R} \sim N(\boldsymbol{\mu}_{\mathbf{a}}, \boldsymbol{\Sigma}_{\mathbf{a}}).$$

The mean and covariance matrix are

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{a}} &= \left( \mathbf{C}^{-1} + \sum_{i=1}^{N} \mathbf{U}'_i \mathbf{R}^{-1} \mathbf{U}_i \right)^{-1} \left( \mathbf{C}^{-1}\mathbf{m} + \sum_{i=1}^{N} \mathbf{U}'_i \mathbf{R}^{-1} \boldsymbol{\omega}^{\star}_{\mathbf{a}i} \right) \\ \boldsymbol{\Sigma}_{\mathbf{a}} &= \left( \mathbf{C}^{-1} + \sum_{i=1}^{N} \mathbf{U}'_i \mathbf{R}^{-1} \mathbf{U}_i \right)^{-1}, \end{aligned}$$

where $\boldsymbol{\omega}^{\star}_{\mathbf{a}i} = \boldsymbol{\omega}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{T}_i\boldsymbol{\Lambda}\mathbf{b}_{(i)}$, $\mathbf{C} = \text{diag}(\mathbf{C}_1, ..., \mathbf{C}_D)$, and $\mathbf{m} = (\mathbf{m}'_1, ..., \mathbf{m}'_D)'$. Recall $\mathbf{m}_d$ and $\mathbf{C}_d$ are hyperparameters defined in Section 2 of the manuscript.

Full conditional of $\mathbf{b}_k$: Define $\mathcal{S}_k = \{i : \mathbf{b}_{(i)} = \mathbf{b}_k\}$ to be the index set of individuals who visited site $k$. The full conditional distribution of $\mathbf{b}_k$ is

$$\mathbf{b}_k \mid \boldsymbol{\omega}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{a}, \mathbf{R} \sim N(\boldsymbol{\mu}_{\mathbf{b}_k}, \boldsymbol{\Sigma}_{\mathbf{b}_k}),$$

where the mean and covariance matrix are

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{b}_k} &= \left( \mathbf{I} + \sum_{i \in \mathcal{S}_k} \mathbf{A}'\boldsymbol{\Lambda}\mathbf{T}'_i\mathbf{R}^{-1}\mathbf{T}_i\boldsymbol{\Lambda}\mathbf{A} \right)^{-1} \sum_{i \in \mathcal{S}_k} \mathbf{A}'\boldsymbol{\Lambda}\mathbf{T}'_i\mathbf{R}^{-1}\boldsymbol{\omega}^{\star}_{\mathbf{b}_k i} \\ \boldsymbol{\Sigma}_{\mathbf{b}_k} &= \left( \mathbf{I} + \sum_{i \in \mathcal{S}_k} \mathbf{A}'\boldsymbol{\Lambda}\mathbf{T}'_i\mathbf{R}^{-1}\mathbf{T}_i\boldsymbol{\Lambda}\mathbf{A} \right)^{-1} \end{aligned}$$

and $\boldsymbol{\omega}^{\star}_{\mathbf{b}_k i} = \boldsymbol{\omega}_i - \mathbf{X}_i\boldsymbol{\beta}$.

Full conditional of $v_{rd}$: Under the Dirac spike, $\boldsymbol{v}$ should be sampled from its marginal posterior, which is obtained after integrating over $\boldsymbol{\beta}$; i.e.,

$$\begin{aligned}\pi(\boldsymbol{v} \mid \boldsymbol{\omega}, \boldsymbol{\lambda}, \mathbf{a}, \mathbf{b}, \mathbf{R}, \boldsymbol{\tau}_v) &\propto \pi(\boldsymbol{v}|\boldsymbol{\tau}_v) \int \pi(\mathbf{Z}, \widetilde{\mathbf{Y}}, \boldsymbol{\omega} \mid \boldsymbol{\Theta})\pi(\boldsymbol{\beta} \mid \boldsymbol{v})d\boldsymbol{\beta} \\ &\propto \pi(\boldsymbol{v}|\boldsymbol{\tau}_v)\pi(\boldsymbol{\omega} \mid \boldsymbol{\lambda}, \mathbf{a}, \mathbf{b}, \mathbf{R}, \boldsymbol{v}),\end{aligned}$$

where $\boldsymbol{\tau}_v = (\tau_{v_{rd}}; r = 1, ...p_d, , d = 1, ..., D)'$ and

$$\pi(\boldsymbol{\omega} \mid \boldsymbol{\lambda}, \mathbf{a}, \mathbf{b}, \mathbf{R}, \boldsymbol{v}) \propto |\boldsymbol{\Phi}(\boldsymbol{v})|^{-1/2}|\boldsymbol{\Sigma}_{\boldsymbol{\beta}}|^{1/2} \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{N} \boldsymbol{\omega}^{\star'}_{\boldsymbol{\beta} i}\mathbf{R}^{-1}\boldsymbol{\omega}^{\star}_{\boldsymbol{\beta} i} - \boldsymbol{\mu}'_{\boldsymbol{\beta}}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\beta}}\boldsymbol{\mu}_{\boldsymbol{\beta}}\right)\right\},$$

where $\boldsymbol{\Phi}(\boldsymbol{v})$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$, $\boldsymbol{\mu}_{\boldsymbol{\beta}}$, and $\boldsymbol{\omega}^{\star}_{\boldsymbol{\beta} i}$ are defined in the full conditional derivation of $\boldsymbol{\beta}$ above. If $\boldsymbol{v} = \mathbf{0}$, then this marginalized likelihood reduces to

$$\exp\left(-\frac{1}{2}\sum_{i=1}^{N} \boldsymbol{\omega}^{\star'}_{\boldsymbol{\beta} i}\mathbf{R}^{-1}\boldsymbol{\omega}^{\star}_{\boldsymbol{\beta} i}\right).$$

Thus, the full conditional distribution of $v_{rd}$, after marginalizing over $\boldsymbol{\beta}$, is Bernoulli with success probability $p_{v_{rd}}$; i.e.,

$$v_{rd} \mid \boldsymbol{\omega}, \boldsymbol{\lambda}, \mathbf{a}, \mathbf{b}, \mathbf{R}, \boldsymbol{v}_{(-rd)}, \tau_{v_{rd}} \sim \text{Bernoulli}(p_{v_{rd}}),$$

where $\boldsymbol{v}_{(-rd)}$ is the vector $\boldsymbol{v}$ after removing the $r$th element of $\boldsymbol{v}_d$ and

$$p_{v_{rd}} = \frac{\pi(\boldsymbol{\omega} \mid \boldsymbol{\lambda}, \mathbf{a}, \mathbf{b}, \mathbf{R}, \boldsymbol{v}_{(-rd)}, v_{rd} = 1)\tau_{v_{rd}}}{\pi(\boldsymbol{\omega} \mid \boldsymbol{\lambda}, \mathbf{a}, \mathbf{b}, \mathbf{R}, \boldsymbol{v}_{(-rd)}, v_{rd} = 0)(1 - \tau_{v_{rd}}) + \pi(\boldsymbol{\omega} \mid \boldsymbol{\lambda}, \mathbf{a}, \mathbf{b}, \mathbf{R}, \boldsymbol{v}_{(-rd)}, v_{rd} = 1)\tau_{v_{rd}}}.$$

Full conditional of $w_{ld}$: Under the Dirac spike, $w_{ld}$ should be sampled from its marginal posterior, which is obtained after integrating over $\lambda_{ld}$ the $\ell$th element of $\boldsymbol{\lambda}$; that is, sample from

$$\begin{aligned}\pi(w_{ld} \mid \boldsymbol{\omega}, \boldsymbol{\beta}, \boldsymbol{\lambda}_{(-\ell)}, \mathbf{a}, \mathbf{b}, \tau_{w_{ld}}) &\propto \pi(w_{ld}|\tau_{w_{ld}}) \int \pi(\mathbf{Z}, \widetilde{\mathbf{Y}}, \boldsymbol{\omega} \mid \boldsymbol{\Theta})\pi(\lambda_{ld} \mid w_{ld})d\lambda_{ld} \\ &\propto \pi(w_{ld}|\tau_{w_{ld}})\pi(\boldsymbol{\omega} \mid \boldsymbol{\beta}, \boldsymbol{\lambda}_{(-\ell)}, \mathbf{a}, \mathbf{b}, w_{ld}),\end{aligned}$$

where $\boldsymbol{\lambda}_{(-\ell)}$ is the vector $\boldsymbol{\lambda}$ with $\lambda_{ld}$ removed and

$$\pi(\boldsymbol{\omega} \mid \boldsymbol{\beta}, \boldsymbol{\lambda}_{(-\ell)}, \mathbf{a}, \mathbf{b}, w_{ld}) \propto \frac{\sigma_{\lambda_{ld}}\{1 - \Phi(-\mu_{\lambda_{ld}}/\sigma_{\lambda_{ld}})\}}{\psi_{ld}/2} \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{N} \boldsymbol{\omega}^{\star'}_{\lambda_\ell i}\mathbf{R}^{-1}\boldsymbol{\omega}^{\star}_{\lambda_\ell i} - \mu^2_{\lambda_{ld}}/\sigma^2_{\lambda_{ld}}\right)\right\}.$$

All notational conventions developed to express the full conditional distribution of $\boldsymbol{\lambda}$ are adopted. When $w_{ld} = 0$, this marginalized likelihood reduces to

$$\exp\left(-\frac{1}{2}\sum_{i=1}^{N} \boldsymbol{\omega}^{\star'}_{\lambda_\ell i}\mathbf{R}^{-1}\boldsymbol{\omega}^{\star}_{\lambda_\ell i}\right).$$

Thus, the full conditional distribution of $w_{ld}$, after marginalizing over $\lambda_{ld}$, is Bernoulli with probability $p_{w_{ld}}$; i.e.,

$$w_{ld} \mid \boldsymbol{\omega}, \boldsymbol{\beta}, \boldsymbol{\lambda}_{(-\ell)}, \mathbf{a}, \mathbf{b}, \tau_{w_{ld}} \sim \text{Bernoulli}(p_{w_{ld}}),$$

where $w_{ld} \mid \boldsymbol{\omega}, \boldsymbol{\beta}, \boldsymbol{\lambda}_{(-\ell)}, \mathbf{a}, \mathbf{b}, \tau_{w_{ld}} \sim \text{Bernoulli}(p_{w_{ld}})$, where

$$p_{w_{ld}} = \frac{\pi(\boldsymbol{\omega} \mid \boldsymbol{\beta}, \boldsymbol{\lambda}_{(-\ell)}, \mathbf{a}, \mathbf{b}, w_{ld} = 1)\tau_{w_{ld}}}{\pi(\boldsymbol{\omega} \mid \boldsymbol{\beta}, \boldsymbol{\lambda}_{(-\ell)}, \mathbf{a}, \mathbf{b}, w_{ld} = 0)(1 - \tau_{w_{ld}}) + \pi(\boldsymbol{\omega} \mid \boldsymbol{\beta}, \boldsymbol{\lambda}_{(-\ell)}, \mathbf{a}, \mathbf{b}, w_{ld} = 1)\tau_{w_{ld}}}.$$

Full conditionals of $S_{e(m):d}$ and $S_{p(m):d}$: Based on the form of $\pi(\mathbf{Z}, \widetilde{\mathbf{Y}} \mid \boldsymbol{\Theta})$ in Section 3 of the manuscript, it is easy to establish the full conditionals

$$\begin{aligned} S_{e(m):d} \mid \mathbf{Z}, \widetilde{\mathbf{Y}} &\sim \text{beta}(a^{\star}_{e(m):d}, b^{\star}_{e(m):d}) \\ S_{p(m):d} \mid \mathbf{Z}, \widetilde{\mathbf{Y}} &\sim \text{beta}(a^{\star}_{p(m):d}, b^{\star}_{p(m):d}), \end{aligned}$$

where

$$\begin{aligned} a^{\star}_{e(m):d} &= a_{e(m):d} + \sum_{j \in \mathcal{I}_m} Z_{jd}\widetilde{Z}_{jd}, \\ b^{\star}_{e(m):d} &= b_{e(m):d} + \sum_{j \in \mathcal{I}_m} (1 - Z_{jd})\widetilde{Z}_{jd}, \\ a^{\star}_{p(m):d} &= a_{p(m):d} + \sum_{j \in \mathcal{I}_m} (1 - Z_{jd})(1 - \widetilde{Z}_{jd}), \\ b^{\star}_{p(m):d} &= b_{p(m):d} + \sum_{j \in \mathcal{I}_m} Z_{jd}(1 - \widetilde{Z}_{jd}). \end{aligned}$$

**Web Appendix B:** *Additional simulation results.* As noted in the manuscript, we performed three additional simulation studies to illustrate our methodology. In the order described in Section 4, these studies examine

**B.1.** Single-stage group testing protocol. This study illustrates how our regression and model selection methods perform for a single-stage group testing protocol where specimens are placed in arrays.

**B.2.** Comparison with Joyner et al. (2020). This study compares our multivariate modeling approach with the marginal modeling methods in Joyner et al. (2020), which is referenced in the manuscript.

**B.3.** Robustness to model misspecification. This study examines the performance of our methods when the linear predictor in the multivariate probit model is misspecified.

In this Web Appendix, we describe each study, present the results, and offer discussions. All references are cited in the manuscript.

**B.1:** *Single-stage group testing protocol.* This study illustrates the performance of our methods when using a non-adaptive group testing protocol; i.e., a protocol where positive pools are not resolved adaptively. We consider single-stage array testing; see Hou et al. (2020). This protocol first assigns individuals to an array and then proceeds to test pools formed by combining individuals who share a common row or column of the array. No further testing is performed regardless of the outcome of the row and column pool tests. Therefore, from an estimation perspective, this protocol presents a more challenging scenario than the two-stage Dorfman algorithm used at the SHL. For the two-stage protocol, additional testing results are available when positive pools are resolved. This is not the case with single-stage protocols.

We randomly assign individuals to $5 \times 5$ arrays and consider one stratum for the assay accuracy probabilities; i.e., the testing stratum ($m = 1$) applies to all row and column pools. We set $S_{e(1):d} = 0.95$ and $S_{p(1):d} = 0.98$, for $d = 1, 2$. We simulate the execution of this single-stage protocol to produce 500 group testing data sets analogously to the study in Section 4 of the manuscript. All prior distributions and model fitting specifications are the same as those described in Section 4. The results from this study are shown in Web Table 1 (see next page).

Web Table 1 provides the average bias and the sample standard deviation of the 500 posterior mean estimates. Also provided are the average estimated posterior probabilities of inclusion for the fixed and random effects in $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$, respectively. The results from this study convey the same findings we reached in Section 4 for the two-stage Dorfman protocol. Estimation is accurate and we identify nonzero fixed and random effects in this more challenging situation.

**B.2:** *Comparison with Joyner et al.* (2020). We seek to benchmark our multivariate modeling approach against the corresponding marginal modeling approach in Joyner et al. (2020), which also adopts a probit link. We simulate the execution of the two-stage Dorfman protocol as described in Section 4 of the manuscript. However, marginal models are used to estimate the relationship between disease statuses and covariates instead. The results from this study are shown in Web Table 2 (see page 8). The reader should compare Table 1 in the manuscript with Web Table 2 to compare the approaches.

Web Table 2 provides the same quantities as Table 1 in the manuscript, except for the correlation matrix $\mathbf{R}_{12}$, which cannot be estimated using marginal methods. Overall, the approach in Joyner et al. (2020) does fairly well, but there are clear gains from joint modeling. For example, intercepts for fixed effects and site-specific random effects are 2-5 times more variable when estimating with marginal models and suffer from much larger bias. Similarly, estimates for the non-zero covariate effects (both fixed and random) have larger bias and are less precise when modeling the disease statuses marginally. Finally, although marginal models perform satisfactorily in model selection (as judged by the posterior probabilities of inclusion, PI), the selection of real effects and the exclusion of null effects is noisier than with a joint model.

Web Table 1: Simulation results from one-stage array testing. Average bias (Bias) of the posterior mean estimates, sample standard deviation (SSD) of the estimates, and average estimated posterior probability of inclusion (PI) for the associated fixed and random effects. Averaged posterior mean estimates of the elements of $\mathbf{a}_d$, $d = 1, 2$, the assay accuracy probabilities, and the correlation matrix element $\mathbf{R}_{12}$ are also shown.

| Disease 1 | | | | Disease 2 | | | |
|---|---|---|---|---|---|---|---|
| Parameter | Bias | SSD | PI | Parameter | Bias | SSD | PI |
| $\beta_{11} = -2$ | $-0.01$ | 0.18 | 1.00 | $\beta_{12} = -2.5$ | $-0.01$ | 0.18 | 1.00 |
| $\beta_{21} = -0.75$ | 0.01 | 0.15 | 0.99 | $\beta_{22} = 0$ | 0.00 | 0.04 | 0.03 |
| $\beta_{31} = 0.5$ | 0.01 | 0.08 | 1.00 | $\beta_{32} = 0$ | 0.00 | 0.02 | 0.02 |
| $\beta_{41} = 0$ | 0.00 | <0.01 | 0.01 | $\beta_{42} = 0.5$ | 0.00 | 0.04 | 1.00 |
| $\beta_{51} = 0$ | 0.00 | <0.01 | 0.01 | $\beta_{52} = -0.25$ | 0.00 | 0.04 | 1.00 |
| $\lambda_{11} = 1$ | 0.04 | 0.16 | 1.00 | $\lambda_{12} = 1$ | 0.06 | 0.16 | 1.00 |
| $\lambda_{21} = 0.75$ | 0.02 | 0.09 | 1.00 | $\lambda_{22} = 0.75$ | 0.02 | 0.10 | 1.00 |
| $\lambda_{31} = 0.25$ | $-0.02$ | 0.09 | 0.91 | $\lambda_{32} = 0.25$ | $-0.01$ | 0.07 | 0.95 |
| $\lambda_{41} = 0$ | 0.00 | <0.01 | 0.01 | $\lambda_{42} = 0$ | 0.00 | <0.01 | 0.01 |
| $\lambda_{51} = 0$ | 0.00 | <0.01 | 0.01 | $\lambda_{52} = 0$ | 0.00 | <0.01 | 0.01 |
| $a_{211} = 0.5$ | $-0.02$ | 0.20 | – | $a_{212} = 0.5$ | $-0.02$ | 0.20 | – |
| $a_{311} = 0.2$ | 0.02 | 0.28 | – | $a_{312} = 0.2$ | $-0.02$ | 0.24 | – |
| $a_{321} = 0.5$ | $-0.03$ | 0.28 | – | $a_{322} = 0.5$ | $-0.01$ | 0.26 | – |
| $a_{411} = 0.1$ | $-0.10$ | 0.02 | – | $a_{412} = 0.1$ | $-0.10$ | 0.03 | – |
| $a_{511} = 0.0$ | 0.00 | 0.02 | – | $a_{512} = 0.0$ | 0.00 | 0.02 | – |
| $a_{421} = 0.2$ | $-0.20$ | 0.02 | – | $a_{422} = 0.2$ | $-0.20$ | 0.03 | – |
| $a_{521} = 0.1$ | $-0.10$ | 0.02 | – | $a_{522} = 0.1$ | $-0.10$ | 0.02 | – |
| $a_{431} = 0.5$ | $-0.50$ | 0.02 | – | $a_{432} = 0.5$ | $-0.50$ | 0.02 | – |
| $a_{531} = 0.2$ | $-0.20$ | 0.02 | – | $a_{532} = 0.2$ | $-0.20$ | 0.02 | – |
| $a_{541} = 0.5$ | $-0.50$ | 0.02 | – | $a_{542} = 0.5$ | $-0.50$ | 0.02 | – |
| $S_{e(1):1} = 0.95$ | 0.00 | 0.01 | – | $S_{e(1):2} = 0.95$ | 0.00 | 0.01 | – |
| $S_{p(1):1} = 0.98$ | 0.00 | 0.01 | – | $S_{p(1):2} = 0.98$ | 0.00 | 0.01 | – |
| $\mathbf{R}_{12} = 0.6$ | $-0.41$ | 0.05 | | | | | |

Web Table 2: Simulation results from marginal modeling using Joyner et al. (2020). Average bias (Bias) of the posterior mean estimates, sample standard deviation (SSD) of the estimates, and average estimated posterior probability of inclusion (PI) for the associated fixed and random effects. Averaged posterior mean estimates of the elements of $\mathbf{a}_d$, $d = 1, 2$ and the assay accuracy probabilities are shown. The correlation matrix element $\mathbf{R}_{12}$ cannot be estimated using a marginal approach.

| Disease 1 | | | | Disease 2 | | | |
|---|---|---|---|---|---|---|---|
| Parameter | Bias | SSD | PI | Parameter | Bias | SSD | PI |
| $\beta_{11} = -2$ | −0.11 | 0.49 | 1.00 | $\beta_{12} = -2.5$ | −0.12 | 0.71 | 0.99 |
| $\beta_{21} = -0.75$ | −0.02 | 0.27 | 0.98 | $\beta_{22} = 0$ | −0.01 | 0.05 | 0.08 |
| $\beta_{31} = 0.5$ | 0.07 | 0.13 | 1.00 | $\beta_{32} = 0$ | 0.00 | 0.01 | 0.06 |
| $\beta_{41} = 0$ | 0.00 | <0.01 | 0.01 | $\beta_{42} = 0.5$ | 0.03 | 0.11 | 1.00 |
| $\beta_{51} = 0$ | 0.00 | <0.01 | 0.02 | $\beta_{52} = -0.25$ | −0.03 | 0.07 | 1.00 |
| $\lambda_{11} = 1$ | 0.12 | 0.32 | 1.00 | $\lambda_{12} = 1$ | 0.17 | 0.42 | 1.00 |
| $\lambda_{21} = 0.75$ | 0.04 | 0.16 | 1.00 | $\lambda_{22} = 0.75$ | 0.04 | 0.23 | 1.00 |
| $\lambda_{31} = 0.25$ | −0.02 | 0.09 | 0.89 | $\lambda_{32} = 0.25$ | −0.03 | 0.10 | 0.87 |
| $\lambda_{41} = 0$ | 0.00 | <0.01 | 0.02 | $\lambda_{42} = 0$ | 0.00 | 0.01 | 0.03 |
| $\lambda_{51} = 0$ | 0.00 | 0.01 | 0.02 | $\lambda_{52} = 0$ | 0.00 | 0.01 | 0.03 |
| $a_{211} = 0.5$ | 0.03 | 0.16 | – | $a_{212} = 0.5$ | 0.07 | 0.30 | – |
| $a_{311} = 0.2$ | −0.08 | 0.19 | – | $a_{312} = 0.2$ | −0.06 | 0.24 | – |
| $a_{321} = 0.5$ | −0.06 | 0.20 | – | $a_{322} = 0.5$ | 0.00 | 0.30 | – |
| $a_{411} = 0.1$ | −0.10 | 0.02 | – | $a_{412} = 0.1$ | −0.10 | 0.01 | – |
| $a_{511} = 0.0$ | 0.00 | 0.01 | – | $a_{512} = 0.0$ | 0.01 | 0.05 | – |
| $a_{421} = 0.2$ | −0.20 | 0.02 | – | $a_{422} = 0.2$ | −0.20 | 0.02 | – |
| $a_{521} = 0.1$ | −0.10 | 0.02 | – | $a_{522} = 0.1$ | −0.10 | 0.05 | – |
| $a_{431} = 0.5$ | −0.50 | 0.01 | – | $a_{432} = 0.5$ | −0.50 | 0.02 | – |
| $a_{531} = 0.2$ | −0.20 | 0.02 | – | $a_{532} = 0.2$ | −0.20 | 0.02 | – |
| $a_{541} = 0.5$ | −0.50 | 0.01 | – | $a_{542} = 0.5$ | −0.50 | 0.01 | – |
| $S_{e(1):1} = 0.95$ | −0.01 | 0.02 | – | $S_{e(1):2} = 0.95$ | −0.01 | 0.01 | – |
| $S_{e(2):1} = 0.98$ | −0.01 | 0.01 | – | $S_{e(2):2} = 0.98$ | −0.01 | 0.01 | – |
| $S_{p(1):1} = 0.98$ | 0.00 | 0.01 | – | $S_{p(1):2} = 0.98$ | 0.00 | <0.01 | – |
| $S_{p(2):1} = 0.99$ | 0.00 | <0.01 | – | $S_{p(2):2} = 0.99$ | 0.00 | <0.01 | – |

**B.3:** *Robustness to model misspecification.* Although the multivariate probit model is a common choice for correlated binary data, as a parametric model, it is certainly not immune from criticism due to potential misclassification. We therefore assess the impact of misspecifying the model using our estimation and model selection methods for group testing data from multiplex assays.

We focus on misspecifying the form of the linear predictor $\eta_{id} = \mathbf{x}_{id}'\boldsymbol{\beta}_d + \mathbf{t}_{id}'\boldsymbol{\Lambda}_d\mathbf{A}_d\mathbf{b}_{(i)d}$, $d = 1, 2$. For each individual, we generate the covariate vector

$$\mathbf{x}_i^* = (1, x_1^*, x_2^*, x_3^*, x_4^*, \phi(x_1^* x_3^*))',$$

where $x_1^*, ..., x_4^*$ have the same covariate distributions specified in Section 4 of the manuscript and $\phi(\cdot)$ is the standard normal density. Note that the inclusion of the $\phi(x_1^* x_3^*)$ covariate creates a nonlinear relationship and hence misspecifies the model. In the linear predictor above, we set $\mathbf{x}_{i1} = \mathbf{x}_{i2} = \mathbf{x}_i^*$ and $\mathbf{t}_{i1} = \mathbf{t}_{i2} = (1, x_1^*, x_2^*, x_3^*, x_4^*)'$. However, when we estimate the multivariate probit model

$$P(\widetilde{\mathbf{Y}}_i = \widetilde{\mathbf{y}}_i \mid \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{a}, \mathbf{b}_{(i)}, \mathbf{R}) = \int_{I_{i1}} \int_{I_{i2}} \phi(\boldsymbol{\omega} | \boldsymbol{\eta}_i, \mathbf{R}) d\boldsymbol{\omega},$$

we ignore the nonlinear covariate and assess the resulting impact of estimating a misspecified model.

The results from our simulation study are shown in Web Tables 3 and 4 (see next two pages). In the model above, we set $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$, where

$$\begin{aligned} \boldsymbol{\beta}_1 &= (-2.0, -0.75, 0.5, 0, 0, \beta_{61})' \\ \boldsymbol{\beta}_2 &= (-2.5, 0, 0, 0.5, -0.25, \beta_{62})', \end{aligned}$$

so that $\beta_{6d}$, the regression parameter associated with the nonlinear term for the $d$th disease, $d = 1, 2$, controls the amount of misspecification. Web Tables 3 and 4 give the results when $\beta_{61} = \beta_{62} = 2.5$ (moderate misspecifcation) and $\beta_{61} = \beta_{62} = 5$ (severe misspecifcation), respectively.

Web Tables 3 and 4 show that ignoring the nonlinear relationship can negatively impact the performance of our approach in terms of bias in the fixed effects (most notably the intercepts). Interestingly, the variability in the fixed effects estimates is about the same as it is under no misclassification (Table 1, manuscript), and estimation performance of the random effects is also similar. Even under severe misspecification, our approach continues to reliably identify the nonzero fixed and random effects.

Web Table 3: Robustness study with moderate misspecification ($\beta_{6d} = 2.5$, $d = 1, 2$). Average bias (Bias) of the posterior mean estimates, sample standard deviation (SSD) of the estimates, and average estimated posterior probability of inclusion (PI) for the associated fixed and random effects. Averaged posterior mean estimates of the elements of $\mathbf{a}_d$, $d = 1, 2$ the assay accuracy probabilities, and the correlation matrix element $\mathbf{R}_{12}$ are also shown.

| Disease 1 | | | | Disease 2 | | | |
|---|---|---|---|---|---|---|---|
| Parameter | Bias | SSD | PI | Parameter | Bias | SSD | PI |
| $\beta_{11} = -2$ | 0.26 | 0.15 | 1.00 | $\beta_{12} = -2.5$ | 0.24 | 0.15 | 1.00 |
| $\beta_{21} = -0.75$ | 0.07 | 0.14 | 0.99 | $\beta_{22} = 0$ | 0.00 | 0.03 | 0.03 |
| $\beta_{31} = 0.5$ | 0.00 | 0.05 | 1.00 | $\beta_{32} = 0$ | 0.00 | 0.01 | 0.02 |
| $\beta_{41} = 0$ | 0.00 | <0.01 | 0.01 | $\beta_{42} = 0.5$ | −0.04 | 0.03 | 1.00 |
| $\beta_{51} = 0$ | 0.00 | <0.01 | 0.01 | $\beta_{52} = -0.25$ | 0.00 | 0.03 | 1.00 |
| $\lambda_{11} = 1$ | 0.02 | 0.12 | 1.00 | $\lambda_{12} = 1$ | 0.06 | 0.14 | 1.00 |
| $\lambda_{21} = 0.75$ | −0.02 | 0.09 | 1.00 | $\lambda_{22} = 0.75$ | −0.05 | 0.09 | 1.00 |
| $\lambda_{31} = 0.25$ | 0.00 | 0.05 | 1.00 | $\lambda_{32} = 0.25$ | 0.00 | 0.05 | 0.99 |
| $\lambda_{41} = 0$ | 0.00 | <0.01 | <0.01 | $\lambda_{42} = 0$ | 0.00 | 0.01 | 0.01 |
| $\lambda_{51} = 0$ | 0.00 | <0.01 | 0.01 | $\lambda_{52} = 0$ | 0.00 | <0.01 | 0.01 |
| $a_{211} = 0.5$ | −0.06 | 0.18 | – | $a_{212} = 0.5$ | −0.02 | 0.20 | – |
| $a_{311} = 0.2$ | −0.02 | 0.25 | – | $a_{312} = 0.2$ | −0.02 | 0.24 | – |
| $a_{321} = 0.5$ | −0.02 | 0.23 | – | $a_{322} = 0.5$ | −0.01 | 0.24 | – |
| $a_{411} = 0.1$ | −0.10 | 0.02 | – | $a_{412} = 0.1$ | −0.10 | 0.07 | – |
| $a_{511} = 0.0$ | 0.00 | 0.02 | – | $a_{512} = 0.0$ | 0.00 | 0.02 | – |
| $a_{421} = 0.2$ | −0.20 | 0.02 | – | $a_{422} = 0.2$ | −0.20 | 0.02 | – |
| $a_{521} = 0.1$ | −0.10 | 0.02 | – | $a_{522} = 0.1$ | −0.10 | 0.02 | – |
| $a_{431} = 0.5$ | −0.50 | 0.02 | – | $a_{432} = 0.5$ | −0.50 | 0.02 | – |
| $a_{531} = 0.2$ | −0.20 | 0.02 | – | $a_{532} = 0.2$ | −0.20 | 0.02 | – |
| $a_{541} = 0.5$ | −0.50 | 0.02 | – | $a_{542} = 0.5$ | −0.50 | 0.02 | – |
| $S_{e(1):1} = 0.95$ | 0.00 | 0.01 | – | $S_{e(1):2} = 0.95$ | −0.01 | 0.01 | – |
| $S_{e(2):1} = 0.98$ | 0.00 | 0.01 | – | $S_{e(2):2} = 0.98$ | 0.00 | 0.01 | – |
| $S_{p(1):1} = 0.98$ | 0.00 | 0.01 | – | $S_{p(1):2} = 0.98$ | 0.00 | <0.01 | – |
| $S_{p(2):1} = 0.99$ | 0.00 | <0.01 | – | $S_{p(2):2} = 0.99$ | 0.00 | <0.01 | – |
| $\mathbf{R}_{12} = 0.6$ | −0.17 | 0.04 | | | | | |

Web Table 4: Robustness study with severe misspecification ($\beta_{6d} = 5$, $d = 1, 2$). Average bias (Bias) of the posterior mean estimates, sample standard deviation (SSD) of the estimates, and average estimated posterior probability of inclusion (PI) for the associated fixed and random effects. Averaged posterior mean estimates of the elements of $\mathbf{a}_d$, $d = 1, 2$ the assay accuracy probabilities, and the correlation matrix element $\mathbf{R}_{12}$ are also shown.

| Disease 1 | | | | Disease 2 | | | |
|---|---|---|---|---|---|---|---|
| Parameter | Bias | SSD | PI | Parameter | Bias | SSD | PI |
| $\beta_{11} = -2$ | 0.52 | 0.15 | 1.00 | $\beta_{12} = -2.5$ | 0.50 | 0.15 | 1.00 |
| $\beta_{21} = -0.75$ | 0.12 | 0.13 | 0.98 | $\beta_{22} = 0$ | 0.00 | 0.03 | 0.03 |
| $\beta_{31} = 0.5$ | −0.01 | 0.05 | 1.00 | $\beta_{32} = 0$ | 0.00 | <0.01 | 0.01 |
| $\beta_{41} = 0$ | 0.00 | <0.01 | <0.01 | $\beta_{42} = 0.5$ | −0.08 | 0.03 | 1.00 |
| $\beta_{51} = 0$ | 0.00 | <0.01 | 0.01 | $\beta_{52} = -0.25$ | 0.01 | 0.02 | 1.00 |
| $\lambda_{11} = 1$ | −0.03 | 0.12 | 1.00 | $\lambda_{12} = 1$ | 0.02 | 0.13 | 1.00 |
| $\lambda_{21} = 0.75$ | −0.05 | 0.08 | 1.00 | $\lambda_{22} = 0.75$ | −0.10 | 0.08 | 1.00 |
| $\lambda_{31} = 0.25$ | −0.01 | 0.04 | 1.00 | $\lambda_{32} = 0.25$ | −0.01 | 0.04 | 1.00 |
| $\lambda_{41} = 0$ | 0.00 | <0.01 | <0.01 | $\lambda_{42} = 0$ | 0.00 | <0.01 | 0.01 |
| $\lambda_{51} = 0$ | 0.00 | <0.01 | 0.01 | $\lambda_{52} = 0$ | 0.00 | <0.01 | 0.01 |
| $a_{211} = 0.5$ | −0.11 | 0.18 | – | $a_{212} = 0.5$ | 0.00 | 0.19 | – |
| $a_{311} = 0.2$ | −0.03 | 0.22 | – | $a_{312} = 0.2$ | −0.04 | 0.22 | – |
| $a_{321} = 0.5$ | −0.01 | 0.22 | – | $a_{322} = 0.5$ | 0.01 | 0.23 | – |
| $a_{411} = 0.1$ | −0.10 | 0.02 | – | $a_{412} = 0.1$ | −0.09 | 0.07 | – |
| $a_{511} = 0.0$ | 0.00 | 0.02 | – | $a_{512} = 0.0$ | 0.00 | 0.02 | – |
| $a_{421} = 0.2$ | −0.20 | 0.02 | – | $a_{422} = 0.2$ | −0.20 | 0.04 | – |
| $a_{521} = 0.1$ | −0.10 | 0.02 | – | $a_{522} = 0.1$ | −0.10 | 0.02 | – |
| $a_{431} = 0.5$ | −0.50 | 0.02 | – | $a_{432} = 0.5$ | −0.50 | 0.03 | – |
| $a_{531} = 0.2$ | −0.20 | 0.02 | – | $a_{532} = 0.2$ | −0.20 | 0.03 | – |
| $a_{541} = 0.5$ | −0.50 | 0.02 | – | $a_{542} = 0.5$ | −0.50 | 0.02 | – |
| $S_{e(1):1} = 0.95$ | 0.00 | 0.01 | – | $S_{e(1):2} = 0.95$ | 0.00 | 0.01 | – |
| $S_{e(2):1} = 0.98$ | 0.00 | 0.01 | – | $S_{e(2):2} = 0.98$ | 0.00 | 0.01 | – |
| $S_{p(1):1} = 0.98$ | 0.00 | 0.01 | – | $S_{p(1):2} = 0.98$ | 0.00 | <0.01 | – |
| $S_{p(2):1} = 0.99$ | 0.00 | 0.00 | – | $S_{p(2):2} = 0.99$ | 0.00 | <0.01 | – |
| $\mathbf{R}_{12} = 0.6$ | −0.14 | 0.03 | | | | | |

**Web Appendix C:** *Informative prior selection for assay accuracy probabilities in Section 5.* The Aptima Combo 2 Assay (AC2A, Hologic, Inc.) possesses different levels of sensitivity and specificity depending on the specimen type and the disease. Web Table 5 summarizes pilot data which were collected on female specimens to validate the performance of the AC2A. These data are available from the AC2A product literature (see `www.hologic.com`) and also from Gaydos et al. (2003).

Web Table 5 combines information from Table 5a (chlamydia, CT) and Table 9a (gonorrhea, NG) in the AC2A product literature. The number of true positives (TP), the number of false negatives (FN), the number of true negatives (TN), and the number of false positives (FP) are shown.

Web Table 5: AC2A pilot data.

| Disease | Stratum | TP | FN | TN | FP |
|---|---|---|---|---|---|
| CT | Swab | 195 | 12 | 1154 | 28 |
| | Urine | 197 | 11 | 1170 | 13 |
| NG | Swab | 126 | 1 | 1335 | 17 |
| | Urine | 116 | 11 | 1347 | 10 |

In Section 5 in the manuscript, we build informative prior distributions for $S_{e(m):d}$ and $S_{p(m):d}$, $m = 1, 2, 3$, $d = 1, 2$, using the pilot data above. Informative prior distributions are specified as

$$\begin{aligned} S_{e(m):d} &\sim \text{beta}(\text{TP} + 1, \text{FN} + 1) \\ S_{p(m):d} &\sim \text{beta}(\text{TN} + 1, \text{FP} + 1). \end{aligned}$$

These can be viewed as the posterior distribution estimates of $S_{e(m):d}$ and $S_{p(m):d}$ that would arise from analyzing the pilot data (Web Table 5) under uniform priors. For example, for individual swab specimens tested for chlamydia ($m = 1$, $d = 1$), we use $S_{e(1):1} \sim \text{beta}(196, 13)$ and $S_{p(1):1} \sim \text{beta}(1155, 29)$. Other prior distributions are formed similarly.