Bayesian Hierarchical Models for High-Dimensional Mediation Analysis with Coordinated Selection of Correlated Mediators: Supporting Information

Yanyi Song[1], **Xiang Zhou**[1,*], **Jian Kang**[1,**], **Max T. Aung**[1], **Min Zhang**[1], **Wei Zhao**[2], **Belinda L. Needham**[2], **Sharon L. R. Kardia**[2], **Yongmei Liu**[3], **John D. Meeker**[4], **Jennifer A. Smith**[2], **and Bhramar Mukherjee**[1]

[1]Department of Biostatistics, University of Michigan, Ann Arbor, MI, U.S.A.

[2]Department of Epidemiology, University of Michigan, Ann Arbor, MI, U.S.A.

[3]Division of Cardiology, Department of Medicine, Duke University School of Medicine, Durham, NC, U.S.A.

[4]Department of Environmental Health Sciences, University of Michigan, Ann Arbor, MI, U.S.A.

[*] *email*: xzhousph@umich.edu

[**] *email*: jiankang@umich.edu

1 Identifiability Assumptions for Causal Mediation Analysis

We use the same counterfactual notation as in the main manuscript. To connect potential variables to observed data, we make the Stable Unit Treatment Value Assumption (SUTVA) [1, 2]. Specifically, the SUTVA assumes there is no interference between subjects and the consistency assumption, which states that the observed variables are the same as the potential variables corresponding to the actually observed treatment level, i.e., $\boldsymbol{M_i} = \sum_a \boldsymbol{M_i}(a) I(A_i = a)$, and $Y_i = \sum_a \sum_{\boldsymbol{m}} Y_i(a, \boldsymbol{m}) I(A_i = a, \boldsymbol{M_i} = \boldsymbol{m})$, where $I(\cdot)$ is the indicator function.

Causal effects are formally defined in terms of potential variables which are not necessarily observed, but the identification of causal effects must be based on observed data. Therefore further assumptions regarding the confounders are required for the identification of causal effects in mediation analysis [3]. We will use $A \perp\!\!\!\perp B | C$ to denote that A is independent of B conditional on C. To estimate the average NDE and NIE from observed data, the following assumptions are needed: (1) $Y_i(a, \boldsymbol{m}) \perp\!\!\!\perp A_i | \boldsymbol{C_i}$, no unmeasured confounding for exposure-outcome relationship; (2) $Y_i(a, \boldsymbol{m}) \perp\!\!\!\perp \boldsymbol{M}_i | \{\boldsymbol{C_i}, A_i\}$, no unmeasured confounding for any of mediator-outcome relationship after controlling for the exposure; (3) $\boldsymbol{M}_i(a) \perp\!\!\!\perp A_i | \boldsymbol{C_i}$, no un-

measured confounding for the exposure effect on all the mediators; (4) $Y_i(a, \boldsymbol{m}) \perp\!\!\!\perp \boldsymbol{M}_i(a^\star)|\boldsymbol{C_i}$, no downstream effect of the exposure that confounds any mediator-outcome relationship. The four assumptions are required to hold with respect to the whole set of mediators. Finally, as in all mediation analysis, the temporal ordering assumption also needs to be satisfied, i.e., the exposure precedes the mediators, and the mediators precede the outcome.

2 Posterior Sampling Algorithm Details for GMM-Potts

Sampling $\begin{bmatrix}(\boldsymbol{\beta_m})_j \\ (\boldsymbol{\alpha_a})_j\end{bmatrix}$ *and* γ_j

$$\log p(\begin{bmatrix}(\boldsymbol{\beta_m})_j \\ (\boldsymbol{\alpha_a})_j\end{bmatrix}|\gamma_j = k, .) \propto -\frac{1}{2}\begin{bmatrix}(\boldsymbol{\beta_m})_j \\ (\boldsymbol{\alpha_a})_j\end{bmatrix}^\top (\boldsymbol{W_j} + \boldsymbol{V_k}^{-1})\begin{bmatrix}(\boldsymbol{\beta_m})_j \\ (\boldsymbol{\alpha_a})_j\end{bmatrix} + \boldsymbol{w_j}^\top \begin{bmatrix}(\boldsymbol{\beta_m})_j \\ (\boldsymbol{\alpha_a})_j\end{bmatrix}$$

where $\boldsymbol{W_j} = \text{diag}\{\sum_{i=1}^n (\sigma_e^2)^{-1} M_{ij}^2, \sum_{i=1}^n (\sigma_g^2)^{-1} A_i^2\}$, and
$\boldsymbol{w_j} = (\sum_{i=1}^n (\sigma_e^2)^{-1}(Y_i - A_i\beta_a - \sum_{j' \neq j} M_{ij'}(\boldsymbol{\beta_m})_{j'})M_{ij}, \sum_{i=1}^n \boldsymbol{\Sigma}^{-1} M_{ij} A_i)^\top$

$$p(\begin{bmatrix}(\boldsymbol{\beta_m})_j \\ (\boldsymbol{\alpha_a})_j\end{bmatrix}|\gamma_j = k, .) \sim \text{MVN}_2((\boldsymbol{W_j} + \boldsymbol{V_k}^{-1})^{-1}\boldsymbol{w_j}, (\boldsymbol{W_j} + \boldsymbol{V_k}^{-1})^{-1})$$

$$\log p(\gamma_j = k|.) \propto -\frac{1}{2}\log|\boldsymbol{W_j V_k} + \boldsymbol{I_2}| + \frac{1}{2}\boldsymbol{w_j}^\top (\boldsymbol{W_j} + \boldsymbol{V_k}^{-1})^{-1}\boldsymbol{w_j} + \theta_{0k} + \sum_{i \sim j} \theta_{1k} I[\gamma_i = \gamma_j = k]$$

Sampling $\boldsymbol{V_k}$

$$\begin{aligned}\log p(\boldsymbol{V_k}|.) \quad \propto \quad & -\frac{1}{2}(\sum_{j=1}^p I[\gamma_j = k] + df + d + 1)\log|\boldsymbol{V_k}| - \frac{1}{2} tr(\boldsymbol{\Psi_0 V_k}^{-1}) \\ & + \sum_{j=1}^p I[\gamma_j = k](-\frac{1}{2}\begin{bmatrix}(\boldsymbol{\beta_m})_j \\ (\boldsymbol{\alpha_a})_j\end{bmatrix}^\top \boldsymbol{V_k}^{-1}\begin{bmatrix}(\boldsymbol{\beta_m})_j \\ (\boldsymbol{\alpha_a})_j\end{bmatrix})\end{aligned}$$

$$p(\boldsymbol{V_k}|.) \sim \text{Inv-Wishart}(\boldsymbol{\Psi_0} + \sum_{j=1}^p I[\gamma_j = k]\begin{bmatrix}(\boldsymbol{\beta_m})_j \\ (\boldsymbol{\alpha_a})_j\end{bmatrix}\begin{bmatrix}(\boldsymbol{\beta_m})_j \\ (\boldsymbol{\alpha_a})_j\end{bmatrix}^\top, \sum_{j=1}^p I[\gamma_j = k] + df)$$

Sampling β_a

$$\log p(\beta_a|.) \propto -\frac{\beta_a^2}{2\sigma_a^2} - \sum_{i=1}^{n}\{\frac{(A_i\beta_a)^2}{2\sigma_e^2} - \sigma_e^{-2}A_i(Y_i - \boldsymbol{M_i}^\top\boldsymbol{\beta_m} - \boldsymbol{C_i}^\top\boldsymbol{\beta_c})\beta_a\}$$

$$p(\beta_a|.) \sim N(\frac{\sum_{i=1}^{n} A_i(Y_i - \boldsymbol{M_i}^\top\boldsymbol{\beta_m} - \boldsymbol{C_i}^\top\boldsymbol{\beta_c})}{\sigma_e^2/\sigma_a^2 + \sum_{i=1}^{n} A_i^2}, \frac{1}{1/\sigma_a^2 + \sum_{i=1}^{n} A_i^2/\sigma_e^2})$$

Sampling σ_a^2

$$\log p(\sigma_a^2|.) \propto -(\frac{1}{2} + h_a + 1)\log(\sigma_a^2) - (\frac{\beta_a^2}{2} + l_a)\sigma_a^{-2}$$

$$p(\sigma_a^2|.) \sim \text{inverse-gamma}(\frac{1}{2} + h_a, \frac{\beta_a^2}{2} + l_a)$$

Sampling σ_e^2

$$\log p(\sigma_e^2|.) = -(\frac{n}{2} + h_1 + 1)\log(\sigma_e^2) - (\frac{\sum_{i=1}^{n}(Y_i - \boldsymbol{M_i}^\top\boldsymbol{\beta_m} - A_i\beta_a - \boldsymbol{C_i}^\top\boldsymbol{\beta_c})^2}{2} + l_1)\sigma_e^{-2}$$

$$p(\sigma_e^2|.) \sim \text{inverse-gamma}(\frac{n}{2} + h_1, \frac{\sum_{i=1}^{n}(Y_i - \boldsymbol{M_i}^\top\boldsymbol{\beta_m} - A_i\beta_a - \boldsymbol{C_i}^\top\boldsymbol{\beta_c})^2}{2} + l_1)$$

Sampling σ_g^2

$$\begin{aligned}\log p(\sigma_g^2|.) &= -(\frac{pn}{2} + h_2 + 1)\log(\sigma_g^2) \\ &\quad -(\frac{\sum_{i=1}^{n}(\boldsymbol{M_i}^\top - A_i\boldsymbol{\alpha_a} - \boldsymbol{C_i}^\top\boldsymbol{\alpha_c})(\boldsymbol{M_i}^\top - A_i\boldsymbol{\alpha_a} - \boldsymbol{C_i}^\top\boldsymbol{\alpha_c})^\top}{2} + l_2)\sigma_g^{-2}\end{aligned}$$

$$p(\sigma_g^2|.) \sim \text{inverse-gamma}(\frac{pn}{2}+h_2, \frac{\sum_{i=1}^{n}(\boldsymbol{M_i}^\top - A_i\boldsymbol{\alpha_a} - \boldsymbol{C_i}^\top\boldsymbol{\alpha_c})(\boldsymbol{M_i}^\top - A_i\boldsymbol{\alpha_a} - \boldsymbol{C_i}^\top\boldsymbol{\alpha_c})^\top}{2}+l_2)$$

Sampling $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1$

We update each of the $\theta_{0k}, \theta_{1k}, k \in \{1, 2, 3, 4\}$ using a double Metropolis-Hastings (DMH) algorithm [4]. For example, for updating θ_{0k}, we first propose a new $\theta_{0k}^\star$ from $\text{N}(\theta_{0k}, \tau_\theta^2)$ and then simulate an auxiliary variable $\boldsymbol{\gamma}^\star$ starting from $\boldsymbol{\gamma}$ based on the new $\boldsymbol{\theta}_0^\star, \boldsymbol{\theta}_1$ where all the elements are the same as $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1$, excluding θ_{0k}. The proposed value $\theta_{0k}^\star$ will be accepted

with probability $\min(1, r_\theta)$ and the Hastings ratio is,

$$r_\theta = \frac{\phi(\theta_{0k}^\star; \mu_{0k}, \sigma_{0k}^2) p(\boldsymbol{\gamma}^\star|\boldsymbol{\theta}_0, \boldsymbol{\theta}_1) p(\boldsymbol{\gamma}|\theta_0^\star, \boldsymbol{\theta}_1)}{\phi(\theta_{0k}; \mu_{0k}, \sigma_{0k}^2) p(\boldsymbol{\gamma}|\boldsymbol{\theta}_0, \boldsymbol{\theta}_1) p(\boldsymbol{\gamma}^\star|\theta_0^\star, \boldsymbol{\theta}_1)}$$

where $\phi(\theta_{0k}; \mu_{0k}, \sigma_{0k}^2)$ is the pdf for the normal distribution $\mathrm{N}(\mu_{0k}, \sigma_{0k}^2)$. The form of $p(\boldsymbol{\gamma}|\boldsymbol{\theta}_0, \boldsymbol{\theta}_1)$ is given by Equation (3.1) as in the main text and the normalizing constants are canceled in the ratio.

Sampling β_{cw}

$$\log p(\beta_{cw}|.) = -\sum_{i=1}^{n} \{ \frac{(C_{iw}\beta_{cw})^2}{2\sigma_e^2} + \sigma_e^{-2} C_{iw}(Y_i - \boldsymbol{M_i}^\top \boldsymbol{\beta_m} - A_i\beta_a - \sum_{s \neq w} C_{iw}\beta_{cw})\beta_{cw} \}$$

$$p(\beta_{cw}|.) \sim N(\frac{\sum_{i=1}^n C_{iw}(Y_i - A_i\beta_a - \boldsymbol{M_i}^\top \boldsymbol{\beta_m} - \sum_{s \neq w} C_{iw}\beta_{cw})}{\sum_{i=1}^n C_{iw}^2}, \frac{\sigma_e^2}{\sum_{i=1}^n C_{iw}^2})$$

Sampling $(\boldsymbol{\alpha_{cw}})_j$

$$\log p((\boldsymbol{\alpha_{cw}})_j|.) = -\sum_{i=1}^{n} \{ \frac{(C_{iw}(\boldsymbol{\alpha_{cw}})_j)^2}{2\sigma_g^2} + \sigma_g^{-2} C_{iw}(M_i^{(j)} - A_i\alpha_{aj} - \sum_{s \neq w} C_{is}(\boldsymbol{\alpha_{cs}})_j)(\boldsymbol{\alpha_{cw}})_j \}$$

$$p((\boldsymbol{\alpha_{cw}})_j|.) \sim N(\frac{\sum_{i=1}^n C_{iw}(M_i^{(j)} - A_i\alpha_{aj} - \sum_{s \neq w} C_{is}(\boldsymbol{\alpha_{cs}})_j)}{\sum_{i=1}^n C_{iw}^2}, \frac{\sigma_g^2}{\sum_{i=1}^n C_{iw}^2})$$

Swendsen-Wang algorithm

We propose to use Swendsen-Wang algorithm [5] to update the Markov random field, $\boldsymbol{\gamma}$. It is a particular case of auxiliary variable methods. In applying SW, we introduce "bond variables", $\boldsymbol{u} = \{u_{ij}, i \sim j\}$, for each neighbor pair $i \sim j$. Given $\boldsymbol{\gamma}$, the non-negative random variable u_{ij} are assumed to be independent and uniformly distributed as below,

$$\begin{aligned} p(u_{ij}|\boldsymbol{\gamma}) &= \exp\{-\sum_{k=1}^{4} \theta_{1k} I[\gamma_i = \gamma_j = k]\} \times I[0 \leq u_{ij} \leq \exp\{\sum_{k=1}^{4} \theta_{1k} I[\gamma_i = \gamma_j = k]\}] \\ p(\boldsymbol{u}|\boldsymbol{\gamma}) &= \prod_{i \sim j} \exp\{-\sum_{k=1}^{4} \theta_{1k} I[\gamma_i = \gamma_j = k]\} \times I[0 \leq u_{ij} \leq \exp\{\sum_{k=1}^{4} \theta_{1k} I[\gamma_i = \gamma_j = k]\}] \end{aligned} \quad (1)$$

Furthermore,

$$p(\boldsymbol{\gamma}|\boldsymbol{u},\cdot) \propto p(\boldsymbol{\beta_m},\boldsymbol{\alpha_a}|\boldsymbol{\gamma})\exp\{\sum_{i=1}^{p}\sum_{k=1}^{4}\theta_{0k}I[\gamma_i=k]\}\times\prod_{i\sim j}I[0\le u_{ij}\le\exp\{\sum_{k=1}^{4}\theta_{1k}I[\gamma_i=\gamma_j=k]\}] \tag{2}$$

To sample from the joint posterior of $\boldsymbol{\gamma}$ and $\boldsymbol{u}=\{u_{ij}, i\sim j\}$, we can iteratively sample from Equation (1) and (2). To sample from (2), we note that $u_{ij}>1$ implies that $\gamma_i=\gamma_j$, so that the bond variable $\boldsymbol{u}$ partitions mediators into same-labeled clusters, and this happens with a probability of $1-\exp\{-\sum_{k=1}^{4}\theta_{1k}I[\gamma_i=\gamma_j=k]\}$. For a particular cluster, C, the probability of belonging to component k is $\propto\prod_{i\in C}p((\boldsymbol{\beta_m})_i,(\boldsymbol{\alpha_a})_i|\gamma_i)\exp\{\theta_{0k}\gamma_i\}$, and each cluster can be updated independently in turn according to its conditional distribution. The SW implementation can be described as below:

1. Update each bond variable according to a uniform distribution:

$$u_{ij}|\boldsymbol{\gamma}\sim U[0,\exp\{\sum_{k=1}^{4}\theta_{1k}I[\gamma_i=\gamma_j=k]\}]$$

   Bonds are forbidden from forming wherever the two neighbors are in different groups.

2. Form the same-labeled clusters (connected components) induced by u_{ij}

   i. The Union-Find algorithm

   ii. Simplifies in the 1-D case

3. For each cluster C, update its label according to its conditional distribution,

$$p(\gamma_C=k|\cdot)\propto\prod_{i\in C}p((\boldsymbol{\beta_m})_i,(\boldsymbol{\alpha_a})_i|\gamma_i)\exp\{\theta_{0k}\gamma_i\},k=1,2,3,4$$

We alternate between Swendsen-Wang updates of $\boldsymbol{\gamma}$ and single site Gibbs updates to ensure movement in large patches.

3 Posterior Sampling Algorithm Details for GMM-CorrS

To sample from the posterior distribution using the Pólya-Gamma method, simply iterate two steps:

Sampling w_{jk} for each j and k

$$w_{jk}|\cdot \sim \text{Pólya-Gamma}(n_{jk}, b_{kj})$$

where $n_{jk} = 1 - \sum_{k'<k} I(\gamma_j = k')$, $n_{j1} = 1$. The samples from Pólya-Gamma distribution can be generated using the algorithm and software in [6].

Sampling $\boldsymbol{b_k}$

We can rewrite 4-dimensional multinomial in terms of 3 binomial densities $\widetilde{\pi}_{j1}$, $\widetilde{\pi}_{j2}$ and $\widetilde{\pi}_{j3}$. Specifically,

$$\begin{aligned} p(\boldsymbol{b_k}) \propto & \prod_j \widetilde{\pi}_{j1}^{I(\gamma_j=1)}((1-\widetilde{\pi}_{j1})\widetilde{\pi}_{j2})^{I(\gamma_j=2)}((1-\widetilde{\pi}_{j1})(1-\widetilde{\pi}_{j2})\widetilde{\pi}_{j3})^{I(\gamma_j=3)} \\ & ((1-\widetilde{\pi}_{j1})(1-\widetilde{\pi}_{j2})(1-\widetilde{\pi}_{j3})\widetilde{\pi}_{j4})^{I(\gamma_j=4)}\text{MVN}(\boldsymbol{a_k}, \sigma_{dk}^2\boldsymbol{D}) \\ \propto & \prod_j \widetilde{\pi}_{j1}^{I(\gamma_j=1)}(1-\widetilde{\pi}_{j1})^{1-I(\gamma_j=1)}\widetilde{\pi}_{j2}^{I(\gamma_j=2)}(1-\widetilde{\pi}_{j2})^{I(\gamma_j=3)+I(\gamma_j=4)}\widetilde{\pi}_{j3}^{I(\gamma_j=3)}(1-\widetilde{\pi}_{j3})^{I(\gamma_j=4)} \\ & \text{MVN}(\boldsymbol{a_k}, \sigma_{dk}^2\boldsymbol{D}) \end{aligned}$$

$$\begin{aligned} I(\gamma_j = 1) &\sim \text{Binom}(1, \text{expit}(b_{1j})), \\ I(\gamma_j = 2) &\sim \text{Binom}(n_{j2}, \text{expit}(b_{2j})), \\ I(\gamma_j = 3) &\sim \text{Binom}(n_{j3}, \text{expit}(b_{3j})) \end{aligned}$$

The multinomial distribution is now expressed with three binomial distributions involving b_{kj}, k = 1,2,3. Following the derivation in [6], we will have,

$$\boldsymbol{b_k}|\cdot \sim \text{MVN}(\boldsymbol{\mu_{bk}}, \boldsymbol{V_{bk}})$$

where

$$\begin{aligned} \boldsymbol{V_{bk}} &= (\boldsymbol{\Omega} + (\sigma_{dk}^2)^{-1}\boldsymbol{D}^{-1})^{-1} \\ \boldsymbol{\mu_{bk}} &= \boldsymbol{V_{bk}}(\boldsymbol{\kappa_k} + (\sigma_{dk}^2)^{-1}\boldsymbol{D}^{-1}\boldsymbol{a_k}) \end{aligned}$$

where $\boldsymbol{\Omega}$ is the diagnol matrix of w_{jk}'s, and $\boldsymbol{\kappa_k} = (I(\gamma_1 = k) - n_{1k}/2, I(\gamma_2 = k) - n_{2k}/2, ..., I(\gamma_p = k) - n_{pk}/2)$. Then we can update $\boldsymbol{\pi_j}$ accordingly.

Sampling σ^2_{dk}

$$\sigma^2_{dk}|\cdot \sim \text{IG}(u + \frac{p}{2}, v + \frac{(\boldsymbol{b_k} - \boldsymbol{a_k})^T \boldsymbol{D}^{-1} (\boldsymbol{b_k} - \boldsymbol{a_k})}{2})$$

The other parameters can be sampled in a similar way as in the GMM-Potts, with details described in the previous section.

4 Empirical FDR Results and Additional Simulations

To estimate the FDR and identify a significance threshold for declaring active mediators, we compute the local false discovery rate for each mediator following [7]. We define the local false discovery rate for the j-th mediator being in the active group as locfdr_{j1}, and it can be expressed as $1 - P(\gamma_j = 1|\text{ Data})$. We first sort locfdr_{j1} from the smallest to the largest, where the jth ordered value is $\text{locfdr}_1^{(j)}, j = 1, ..., p$. Then the cutoff value c_1 for locfdr_{j1} to guarantee a 10% FDR can be identified from,

$$\arg\max_{c_1} \frac{1}{\sum_{j=1}^{p} I(\text{locfdr}_1^{(j)} < c_1)} \sum_{j=1}^{p} I(\text{locfdr}_1^{(j)} < c_1) \text{locfdr}_1^{(j)} < 0.1$$

where I is an indicator function. Following [8], we declare mediators with an locfdr_{j1} smaller than the threshold c_1 as active mediators.

As another practical procedure, we also consider a cutoff on the posterior inclusion probabilities (PIP) to declare active mediators. To evaluate the performance of those significance rules, we report the empirical FDR and TPR in Table S1 and S2 under all the simulation scenarios.

In addition to the simulation scenarios included in the main text, we also perform additional simulations for a more comprehensive evaluation. First, we consider scenarios where there is a single active mediator in each high-correlated block. In these simulations, we follow the same simulation procedure and correlation structure as in the main text but each highly-correlated block only contains one active mediator. The simulation results are presented in Table S3. When the correlation structure underlying mediators provides little help in infer-

Method	TPR	TPR(locfdr)	FDR(locfdr)	TPR(PIP>0.5)	FDR(PIP>0.5)	TPR(PIP>0.9)	FDR(PIP>0.9)
$\rho_1 = 0.5 - 0.03\|i-j\|, \rho_2 = 0$, Signals in one block							
GMM-CorrS	0.78	0.69(0.021)	0.04(0.008)	0.82(0.019)	0.12(0.012)	0.49(0.016)	0.02(0.006)
GMM-Potts	0.93	0.79(0.019)	0.05(0.007)	0.86(0.014)	0.07(0.010)	0.61(0.017)	0.01(0.002)
$\rho_1 = 0.5 - 0.03\|i-j\|, \rho_2 = 0$, Signals in two blocks							
GMM-CorrS	0.62	0.52(0.018)	0.07(0.010)	0.67(0.021)	0.14(0.012)	0.40(0.015)	0.01(0.005)
GMM-Potts	0.49	0.34(0.041)	0.06(0.025)	0.66(0.023)	0.22(0.022)	0.24(0.032)	0.02(0.017)
$\rho_1 = 0.9 - 0.05\|i-j\|, \rho_2 = 0.1$, Signals in one block							
GMM-CorrS	0.81	0.49(0.020)	0.06(0.013)	0.83(0.014)	0.17(0.007)	0.36(0.018)	0.02(0.015)
GMM-Potts	0.92	0.51(0.043)	0.05(0.015)	0.83(0.049)	0.08(0.014)	0.23(0.014)	0.01(0.012)
$\rho_1 = 0.9 - 0.05\|i-j\|, \rho_2 = 0.1$, Signals in two blocks							
GMM-CorrS	0.49	0.31(0.032)	0.09(0.023)	0.55(0.032)	0.23(0.023)	0.22(0.021)	0.05(0.020)
GMM-Potts	0.40	0.27(0.006)	0.06(0.005)	0.43(0.038)	0.17(0.022)	0.17(0.006)	0.03(0.008)
$\rho_1 = 0$, Signals in two blocks							
GMM-CorrS	0.52	0.44(0.015)	0.03(0.008)	0.50(0.015)	0.07(0.012)	0.35(0.012)	0.01(0.004)
GMM-Potts	0.46	0.42(0.022)	0.06(0.016)	0.50(0.016)	0.19(0.019)	0.33(0.016)	0.02(0.011)
Weak correlation from MESA, Signals in two blocks							
GMM-CorrS	0.44	0.32(0.009)	0.03(0.009)	0.39(0.011)	0.08(0.013)	0.27(0.007)	0.01(0.006)
GMM-Potts	0.40	0.35(0.014)	0.07(0.015)	0.45(0.017)	0.23(0.022)	0.27(0.010)	0.03(0.011)

Table S1: Empirical estimates of TPR and FDR in simulations of $n = 100, p = 200$. The results are based on 200 replicates for each setting, and the standard errors are shown within parentheses. TPR is the true positive rate controlled at a fixed FDR of 10%; TPR(locfdr) and FDR(locfdr) are the empirical estimates based on the local FDR approach; TPR(PIP>0.9) and FDR(PIP>0.9) are the empirical estimates when the PIP threshold for identifying active mediators is 0.9; TPR(PIP>0.5) and FDR(PIP>0.5) are the empirical estimates when the PIP threshold for identifying active mediators is 0.5.

Method	TPR	TPR(locfdr)	FDR(locfdr)	TPR(PIP>0.5)	FDR(PIP>0.5)	TPR(PIP>0.9)	FDR(PIP>0.9)
$\rho_1 = 0.5 - 0.02\|i-j\|, p_{11} = 100$, Signals in five block							
GMM-CorrS	0.92	0.90(0.001)	0.08(0.002)	0.88(0.002)	0.02(0.012)	0.80(0.003)	0.00(0.001)
GMM-Potts	0.97	0.96(0.002)	0.09(0.002)	0.96(0.002)	0.01(0.002)	0.93(0.002)	0.00(0.002)
Weak correlation from MESA, $p_{11} = 100$, Signals in five blocks							
GMM-CorrS	0.83	0.83(0.002)	0.12(0.003)	0.81(0.003)	0.04(0.003)	0.77(0.003)	0.00(0.001)
GMM-Potts	0.76	0.86(0.017)	0.33(0.022)	0.88(0.010)	0.35(0.024)	0.82(0.011)	0.16(0.017)
$\rho_1 = 0.5 - 0.02\|i-j\|, p_{11} = 10$, Signals in two blocks							
GMM-CorrS	0.83	0.82(0.007)	0.09(0.006)	0.81(0.007)	0.05(0.007)	0.74(0.009)	0.01(0.003)
GMM-Potts	0.85	0.88(0.007)	0.34(0.012)	0.95(0.007)	0.65(0.008)	0.83(0.008)	0.04(0.012)
$\rho_1 = 0.25, p_{11} = 10$, Signals in two blocks							
GMM-CorrS	0.82	0.80(0.006)	0.09(0.007)	0.80(0.006)	0.06(0.007)	0.74(0.007)	0.01(0.003)
GMM-Potts	0.61	0.79(0.018)	0.35(0.043)	0.81(0.011)	0.56(0.037)	0.75(0.010)	0.15(0.047)

Table S2: Empirical estimates of TPR and FDR in simulations of $n = 1000, p = 2000$, p_{11} is the number of true active mediators. The results are based on 200 replicates for each setting, and the standard errors are shown within parentheses. TPR is the true positive rate controlled at a fixed FDR of 10%; TPR(locfdr) and FDR(locfdr) are the empirical estimates based on our PIP approach; TPR(PIP>0.9) and FDR(PIP>0.9) are the empirical estimates when the PIP threshold for identifying active mediators is 0.9; TPR(PIP>0.5) and FDR(PIP>0.5) are the empirical estimates when the PIP threshold for identifying active mediators is 0.5.

ring active mediators, we find that the three GMM-based methods behave quite similarly to each other, and outperform the frequentist methods. GMM-Potts tends to be less robust

Table S3: Simulation results of $(n, p) = (100, 200)/(1000, 2000)$ under different correlation structures, p_{11} is the number of true active mediators. TPR: true positive rate at false discovery rate (FDR) = 0.10. $\text{MSE}_{\text{non-null}}$: mean squared error for the indirect effects of active mediators. MSE_{null}: mean squared error for the indirect effects of inactive mediators. The results are based on 200 replicates for each setting. Bolded TPRs indicate the top two performers.

	Each block contains only one signal, $\boldsymbol{n = 100, p = 200, p_{11} = 10}$					
	(A) $\boldsymbol{\rho_1 = 0.5 - 0.03\|i-j\|, \rho_2 = 0}$			**(B)** $\boldsymbol{\rho_1 = 0.9 - 0.05\|i-j\|, \rho_2 = 0.1}$		
Method	**TPR**	$\textbf{MSE}_{\text{non-null}}$	$\textbf{MSE}_{\text{null}}$ $\times 10^{-4}$	**TPR**	$\textbf{MSE}_{\text{non-null}}$	$\textbf{MSE}_{\text{null}}$ $\times 10^{-4}$
GMM-CorrS	0.46	0.022	1.488	**0.39**	0.025	2.284
GMM-Potts	**0.48**	0.027	2.336	**0.39**	0.028	3.035
GMM	**0.47**	0.021	1.291	**0.41**	0.025	2.089
Bi-Lasso	0.36	0.101	0.753	0.26	0.195	0.763
Bi-Ridge	0.27	0.251	3.298	0.22	0.277	2.363
Pathway Lasso	0.29	0.160	0.344	0.27	0.194	0.286
	Each block contains only one signal, $\boldsymbol{n = 1000, p = 2000, p_{11} = 10}$					
	(A) $\boldsymbol{\rho_1 = 0.5 - 0.02\|i-j\|, \rho_2 = 0}$			**(B)** $\boldsymbol{\rho_1 = 0.9 - 0.03\|i-j\|, \rho_2 = 0.1}$		
Method	**TPR**	$\textbf{MSE}_{\text{non-null}}$ $\times 10^{-3}$	$\textbf{MSE}_{\text{null}}$ $\times 10^{-6}$	**TPR**	$\textbf{MSE}_{\text{non-null}}$ $\times 10^{-3}$	$\textbf{MSE}_{\text{null}}$ $\times 10^{-6}$
GMM-CorrS	**0.81**	2.317	1.484	**0.78**	4.501	4.920
GMM-Potts	**0.81**	2.892	0.724	0.73	7.257	4.076
GMM	**0.81**	2.396	1.256	**0.78**	4.511	4.970
Bi-Lasso	0.72	10.425	3.761	0.64	17.676	5.053
Bi-Ridge	0.50	14.084	15.273	0.41	28.471	14.049
Pathway Lasso	0.56	13.507	14.001	0.50	25.940	14.609

due to the inclusion of irrelevant neighbors, and therefore has slightly larger MSEs than GMM-CorrS and GMM.

Next, we consider a different ratio of n/p from the main text. Specifically, we set $n = 100, p = 500$ to examine the performance of the proposed methods. In these simulations, the effects are generated from the mixture of bivariate normals, $[\beta_{mj}, \alpha_{aj}]^\top$ from

$$[\beta_{mj}, \alpha_{aj}]^\top \sim \pi_1 \text{MVN}(\mathbf{0}, \begin{bmatrix} 1.0 & 0.2 \\ 0.2 & 1.0 \end{bmatrix}) + \pi_2 \text{MVN}(\mathbf{0}, \begin{bmatrix} 1.0 & 0 \\ 0 & 0 \end{bmatrix}) + \pi_3 \text{MVN}(\mathbf{0}, \begin{bmatrix} 0 & 0 \\ 0 & 1.0 \end{bmatrix}) + \pi_4 \boldsymbol{\delta}_0$$

where we set $\pi_1 = 10/500$, $\pi_2 = 20/500$, $\pi_3 = 40/500$, and $\pi_4 = 430/500$. For the correlation structure, we assume 25 highly-correlated blocks of size 10×10, within which the pairwise correlation of mediators is ρ_1, e.g. $\rho_1 = 0.5 - 0.03|i-j|$ or $0.9 - 0.05|i-j|$, and the correlation between blocks (ρ_2) is relatively weak (e.g. $\rho_2 = 0$ or 0.1). The 10 active mediators are either clustered within one block or scattered over a few blocks, while the other blocks contain no

active mediators. We also consider settings where there is either no correlation (i.e. using identical matrix to simulate mediators) or a lack of such structural information underlying active mediators (i.e. the estimated covariance was based on a random subset of DNAm from MESA to simulate mediators).

The results are summarized in Table S4. In general, we observe similar performance patterns with $p = 500$ as in $p = 200$: the proposed methods are better at identifying groups of correlated mediators than GMM and the frequentist methods in the presence of such correlation, and behave similarly to GMM in the absence of correlation or when the correlation information is not related. The TPRs get reduced with increased number of mediators, and GMM-Potts suffers more loss of accuracy when ρ_1 is small, e.g. $0.5 - 0.03|i - j|$, and signals are in two blocks. This is possibly due to the fact that the mis-classification in 0-1 neighborhood matrix increases more with the increased noise under larger number of mediators.

5 Detailed Description of MESA Data

MESA is a population-based longitudinal study designed to identify risk factors for the progression of subclinical cardiovascular disease (CVD) [9]. A total of 6,814 non-Hispanic white, African-American, Hispanic, and Chinese-American women and men aged 45−84 without clinically apparent CVD were recruited between July 2000 and August 2002 from the following 6 regions in the US: Forsyth County, NC; Northern Manhattan and the Bronx, NY; Baltimore City and Baltimore County, MD; St. Paul, MN; Chicago, IL; and Los Angeles County, CA. Each field center recruited from locally available sources, which included lists of residents, lists of dwellings, and telephone exchanges. Neighborhood socioeconomic disadvantage scores for each neighborhood were created based on a principal components analysis of 16 census-tract level variables from the 2000 US Census. These variables reflect dimensions of education, occupation, income and wealth, poverty, employment, and housing. For the neighborhood measures, we use the cumulative average of the measure across all available MESA examinations. The descriptive statistics for the exposure and outcome

Table S4: Simulation results of $n = 100, p = 500$ under different correlation structures. TPR: true positive rate at false discovery rate (FDR) = 0.10. $\text{MSE}_{\text{non-null}}$: mean squared error for the indirect effects of active mediators. MSE_{null}: mean squared error for the indirect effects of inactive mediators. The results are based on 200 replicates for each setting. Bolded TPRs indicate the top two performers.

	$\boldsymbol{\rho_1 = 0.5 - 0.03\lvert i-j\rvert, \rho_2 = 0}$					
	(A) Signals in one block			**(B) Signals in two blocks**		
Method	**TPR**	$\textbf{MSE}_{\text{non-null}}$	$\textbf{MSE}_{\text{null}}$ $\times 10^{-4}$	**TPR**	$\textbf{MSE}_{\text{non-null}}$	$\textbf{MSE}_{\text{null}}$ $\times 10^{-4}$
GMM-CorrS	**0.54**	0.056	2.530	**0.46**	0.076	2.869
GMM-Potts	**0.63**	0.074	2.801	0.34	0.138	2.761
GMM	0.42	0.070	1.804	**0.42**	0.081	2.170
Bi-Lasso	0.18	0.256	0.414	0.20	0.248	0.398
Bi-Ridge	0.15	0.280	2.413	0.16	0.281	2.364
Pathway Lasso	0.24	0.233	2.598	0.23	0.180	6.405
	$\boldsymbol{\rho_1 = 0.9 - 0.05\lvert i-j\rvert, \rho_2 = 0.1}$					
	(A) Signals in one block			**(B) Signals in two blocks**		
Method	**TPR**	$\textbf{MSE}_{\text{non-null}}$	$\textbf{MSE}_{\text{null}}$ $\times 10^{-4}$	**TPR**	$\textbf{MSE}_{\text{non-null}}$	$\textbf{MSE}_{\text{null}}$ $\times 10^{-4}$
GMM-CorrS	**0.53**	0.140	4.166	**0.37**	0.152	3.903
GMM-Potts	**0.71**	0.124	1.803	**0.38**	0.162	4.955
GMM	0.30	0.172	5.157	0.32	0.151	2.882
Bi-Lasso	0.12	0.301	0.283	0.11	0.297	0.336
Bi-Ridge	0.15	0.296	2.249	0.16	0.298	1.929
Pathway Lasso	0.21	0.237	5.495	0.19	0.264	3.457
	No systematic correlation structure (signals in two blocks)					
	(A) $\boldsymbol{\rho_1 = 0}$			**(B) Weak correlation from MESA**		
Method	**TPR**	$\textbf{MSE}_{\text{non-null}}$	$\textbf{MSE}_{\text{null}}$ $\times 10^{-4}$	**TPR**	$\textbf{MSE}_{\text{non-null}}$	$\textbf{MSE}_{\text{null}}$ $\times 10^{-4}$
GMM-CorrS	**0.44**	0.060	2.470	**0.27**	0.126	1.757
GMM-Potts	0.40	0.106	1.986	0.24	0.144	2.582
GMM	**0.45**	0.057	1.466	**0.28**	0.131	1.458
Bi-Lasso	0.30	0.191	0.669	0.21	0.186	0.353
Bi-Ridge	0.25	0.260	2.726	0.17	0.201	2.141
Pathway Lasso	0.35	0.164	0.314	0.32	0.177	0.400

can be found in Table S5.

In the MESA data, between April 2010 and February 2012 (corresponding to MESA Exam 5), DNAm were assessed on a random subsample of 1,264 non-Hispanic white, African-American, and Hispanic MESA participants aged 55−94 from the Baltimore, Forsyth County, New York, and St. Paul field centers. After excluding respondents with missing data on one or more variables, we had phenotype and DNAm data from purified monocytes on a total of 1,225 individuals and we focused on this set of individuals for analysis. The detailed description of DNAm data collection, quantitation and data processing procedures can be found in Liu et al [10]. Briefly, the Illumina HumanMethylation450 BeadChip was used to measure DNAm, and bead-level data were summarized in GenomeStudio. Quantile normalization was performed using the *lumi* package with default settings [11]. Quality control (QC) measures included checks for sex and race/ethnicity mismatches and outlier identification by multidimensional scaling plots. Further probe filtering criteria included: "detected" DNAm levels in <90% of MESA samples (detection p-value cut-off = 0.05), existence of a SNP within 10 base pairs of the target CpG site, overlap with a non-unique region, and suggestions by DMRcate [12] (mostly cross-reactive probes). Those procedures leave us 403,713 autosomal probes for analysis.

For computational reasons, we first selected a subset of CpG sites to be used in the final multivariate mediation analysis model. In particular, for each single CpG site in turn, we fit the following linear mixed model to test the marginal association between the CpG site and the exposure variable:

$$M_i = A_i\psi_a + \boldsymbol{C_{1i}}^\top \boldsymbol{\psi_c} + \boldsymbol{Z_i}^\top \boldsymbol{\psi_u} + \epsilon_i, i = 1, ..., n \tag{3}$$

where A_i represents neighborhood SES value for the i'th individual and ψ_a is its coefficient; $\boldsymbol{C_{1i}}$ is a vector of covariates that include age, gender, race/ethnicity, childhood socioeconomic status, adult socioeconomic status and enrichment scores for each of 4 major blood cell types (neutrophils, B cells, T cells and natural killer cells) to account for potential contamination by non-monocyte cell types; $\boldsymbol{Z_i}^\top \boldsymbol{\psi_u}$ represent methylation chip and position random

effects and are used to control for possible batch effects. The error term $\epsilon_i \sim \text{MVN}(0, \sigma^2 I_n)$ and is independent of the random effects. We obtained p-values for testing the null hypothesis $\psi_a = 0$ from the above model. We further applied the R/Bioconductor package BACON [13] to these p-values to further adjust for possible inflation using an empirical null distribution. Based on these marginal p-values, we selected top 2,000 CpG sites with the smallest p-values for our Bayesian multivariate analysis.

	Full Sample (n, %)	**Neighborhood Socioeconomic Disadvantage Mean (SD)**	**Glucose Mean (SD)**
Full sample	1225 (100)	-0.32 (1.11)	29.5 (5.49)
Age			
55−65 years	462 (38)	-0.18 (0.96)	30.3 (6.02)
66−75 years	397 (32)	-0.30 (1.16)	30.1 (5.21)
76−85 years	300 (24)	-0.47 (1.15)	28.2 (4.65)
86−95 years	66 (5)	-0.67 (1.46)	26.6 (4.66)
Race/ethnic group			
Non-Hispanic white	580 (47)	-0.56 (1.18)	28.7 (5.40)
African-American	263 (22)	-0.16 (0.98)	30.5 (5.69)
Hispanic	382 (31)	-0.05 (1.00)	30.0 (5.32)
Gender			
Female	633 (52)	-0.24 (1.09)	30.1 (6.20)
Male	592 (48)	-0.40 (1.12)	28.9 (4.54)

Table S5: Characteristics of 1225 participants from MESA. %: proportion in the corresponding category. SD: standard deviation.

6 Detailed Description of LIFECODES Data

The LIFECODES prospective birth cohort enrolled approximately 1,600 pregnant women between 2006 and 2008 at the Brigham and Women's Hospital in Boston, MA. Participants between 20 and 46 years of age were all at less than 15 weeks gestation at the initial study visit, and followed up to four visits (targeted at median 10, 18, 26, and 35 weeks gestation). At the initial study visit, questionnaires were administered to collect demographic and health-related information. Subjects' urine and plasma samples were collected at each study visit. Among participants recruited in the LIFECODES cohort, 1,181 participants were followed until delivery and had live singleton infants. The birth outcome, gestational age, was also recorded at delivery, and preterm birth was defined as delivery prior to 37 weeks gestation. This study received institutional review board (IRB) approval from the Brigham

and Women's Hospital and all participants provided written informed consent. All of the methods within this study were performed in accordance with the relevant guidelines and regulations approved by the IRB. Additional details regarding recruitment and study design can be found in [14, 15].

In this study, we focused on a subset of $n = 161$ individuals with their urine and plasma samples collected at one study visit occurring between 23.1 and 28.9 weeks gestation (median = 26 weeks). Characteristics of the subset sample is described in Table S6. Subjects' urine samples were refrigerated (4°C) for a maximum of 2 hours before being processed and stored at −80°C. Approximately 10mL of blood was collected using ethylenediaminetetraacetic acid plasma tubes and temporarily stored at 4°C for less than 4 hours. Afterwards, blood was centrifuged for 20 minutes and stored at −80°C. Environmental exposure analytes were measured from urine samples by NSF International in Ann Arbor, MI, following the methods developed by the Centers for Disease Control (CDC) [16]. Those exposure analytes include phthalates, phenols and parabens, trace metals and polycyclic aromatic hydrocarbons. To adjust for urinary dilution, specific gravity (SG) levels were measured in each urine sample using a digital handheld refractometer (ATAGO Company Ltd., Tokyo, Japan), and was included as a covariate in regression models. Urine and plasma were subsequently analyzed for endogenous biomarkers, including 51 eicosanoids, five oxidative stress biomarkers and five immunological biomarkers in the present study. For a detailed description of the biomarkers that we analyzed and the media (urine or plasma) in which they were measured, please refer to [17].

We also utilize the biological pathway information to construct neighbors in GMM-Potts and correlation matrix in GMM-CorrS. The results are shown in Table S7.

7 Choices of $\psi's$

Following the reviewer's comment, we performed sensitivity analysis on the ψ parameters (ψ_{01} and ψ_{02}) in the covariances of the mixture model. Specifically, we added different level of changes to the Bi-Lasso variance estimations and re-fitted both models. The results are shown in Table S8. We found that the posterior inference is robust to mild changes in $\psi's$,

	Full Sample (n = 161)	**Preterm (<37 weeks gestation, n = 52)**	**Control (n = 109)**
Age[a]	32.7 (4.4)	32.1 (5.0)	33.0 (4.2)
BMI at Initial Visit[a]	26.7 (6.4)	28.5 (7.6)	25.8 (5.6)
Race/ethnic group[b]			
White	102 (63%)	29 (56%)	73 (67%)
African-American	18 (11%)	7 (13%)	11 (10%)
Other	41 (26%)	16 (31%)	25 (23%)
Gestational weeks[a]	37.5 (3.1)	34.1 (3.2)	39.1 (1.1)

Table S6: Characteristics of all participants in the subset sample from the LIFECODES prospective birth cohort (n = 161). [a]Continuous variables presented as: mean (standard deviation). [b]Categorical variables presented as: count (proportion).

Method	Selected Mediators	PIP	$\hat{\beta}_{mj}\hat{\alpha}_{aj}$ (95% CI)
Polycyclic aromatic hydrocarbons $\rightarrow$ *Biomarkers* $\rightarrow$ *Gestational Age*			
GMM-Potts	8(9)-EET	0.99	0.698(0.391, 1.005)
	9,10-DiHOME	0.91	-0.359(-0.671, 0.000)
GMM-CorrS	12(13)-EpoME	1.00	1.132(0.867, 1.426)
	9-oxoODE	0.99	-0.834(-1.119, -0.535)
GMM	8(9)-EET	1.00	0.698(0.407, 0.990)
	9,10-DiHOME	0.99	-0.394(-0.693, -0.091)

Table S7: Summary of the identified active mediators from the data application on LIFECODES study based on 10% FDR with the local FDR approach. Both GMM-Potts and GMM-CorrS use the neighborhood structure based on biological pathways. Besides the PIP, we also report the effect estimation $\hat{\beta}_{mj}\hat{\alpha}_{aj}$ and its 95% credible interval.

especially as we increase the values of $\psi's$. The results also indicate that values smaller than the variances estimated by Bi-Lasso may not be good choices for $\psi's$.

In addition, we explored the use of the deviance information criterion (DIC) proposed in [18] as a measure of model fit to select ψ. The results are shown in Figure S1. We found that, while the DICs are relatively stable across a range of ψ parameters, the smallest DICs do correspond to the best/close-to-best models in terms of TPR in Table S8 for both GMM-CorrS and GMM-Potts. Therefore, when presented with a set of potential ψ parameters, users can use DIC as one criterion for model selection.

Combining the above results and given that the choice of $\psi's$ is not of the main focus of the present study, we believe our current approach of using Bi-Lasso can provide good choices

Table S8: Sensitivity analysis of GMM-CorrS and GMM-Potts regarding ψ parameters under $n = 100, p = 200/500$. We add various percentages of change to the estimated ψ_{01} and ψ_{02} from Bi-Lasso and re-evaluate the model performance.

	$\boldsymbol{p = 200, \rho_1 = 0.5 - 0.03\|i-j\|, \rho_2 = 0}$, **signals in two blocks**					
	GMM-CorrS			**GMM-Potts**		
% of change	**TPR**	$\mathbf{MSE}_{\text{non-null}}$	$\mathbf{MSE}_{\text{null}}$ $\times 10^{-4}$	**TPR**	$\mathbf{MSE}_{\text{non-null}}$	$\mathbf{MSE}_{\text{null}}$ $\times 10^{-4}$
-90	0.49	0.035	2.701	0.38	0.124	2.106
-50	0.56	0.031	2.771	0.41	0.063	2.881
0	0.62	0.039	1.919	0.49	0.040	2.112
50	0.62	0.039	1.919	0.48	0.050	3.086
10^2	0.62	0.028	2.771	0.50	0.036	3.239
2×10^2	0.62	0.026	2.687	0.49	0.034	3.384
5×10^2	0.63	0.025	2.857	0.51	0.034	3.105
10^3	0.64	0.024	2.948	0.51	0.029	3.163
10^4	0.62	0.030	2.374	0.47	0.037	1.910
	$\boldsymbol{p = 500, \rho_1 = 0.5 - 0.03\|i-j\|, \rho_2 = 0}$, **signals in two blocks**					
	GMM-CorrS			**GMM-Potts**		
% of change	**TPR**	$\mathbf{MSE}_{\text{non-null}}$	$\mathbf{MSE}_{\text{null}}$ $\times 10^{-4}$	**TPR**	$\mathbf{MSE}_{\text{non-null}}$	$\mathbf{MSE}_{\text{null}}$ $\times 10^{-4}$
-90	0.41	0.088	1.900	0.25	0.192	3.141
-50	0.42	0.076	2.732	0.27	0.162	3.037
0	0.46	0.076	2.869	0.34	0.138	2.761
50	0.44	0.076	2.316	0.32	0.144	2.789
10^2	0.44	0.075	2.548	0.32	0.138	3.153
2×10^2	0.45	0.066	2.563	0.32	0.133	2.888
5×10^2	0.45	0.067	3.578	0.32	0.128	2.748
10^3	0.46	0.059	3.560	0.35	0.112	1.812
10^4	0.47	0.059	3.897	0.32	0.104	2.511

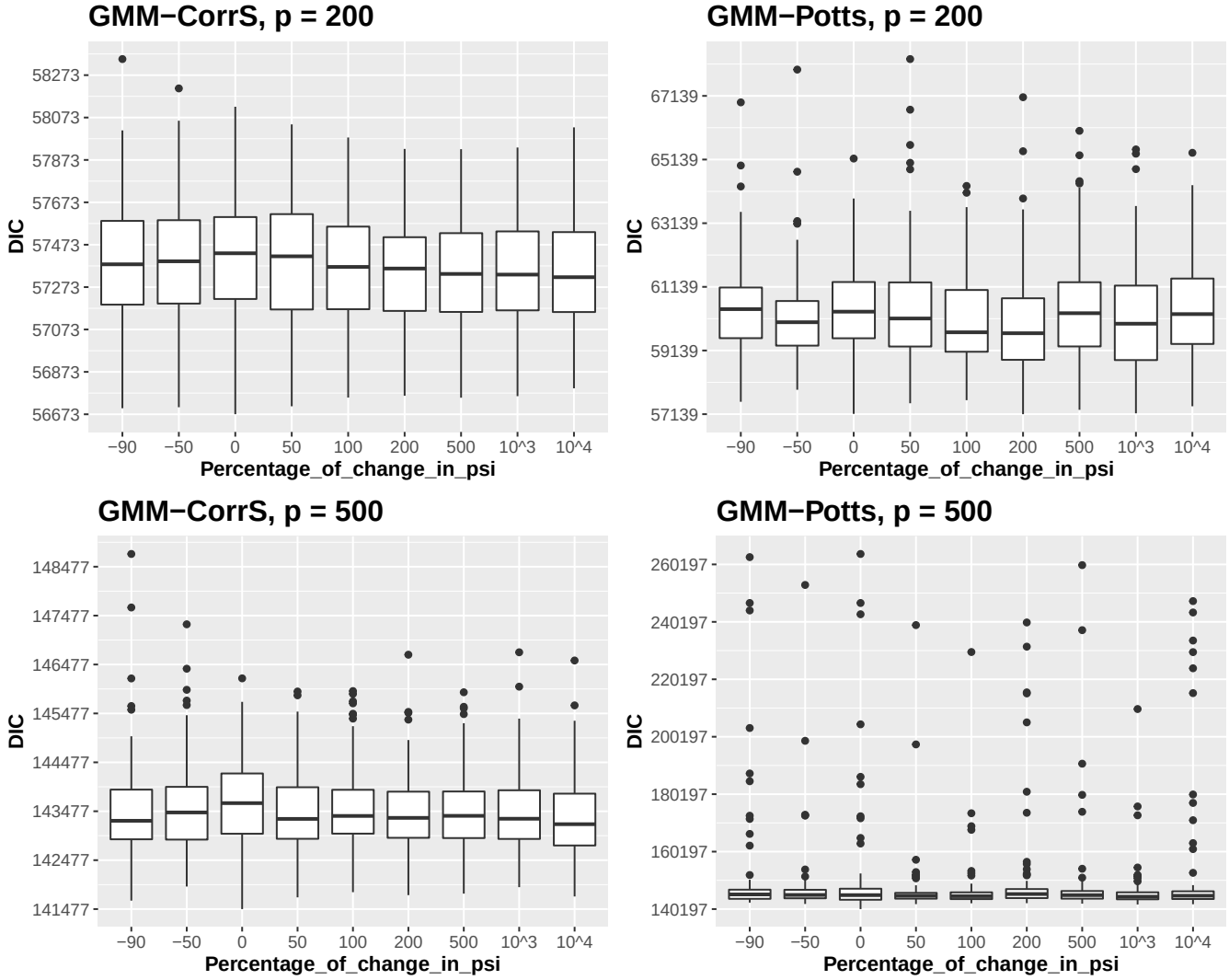

Figure S1: The deviance information criterion (DIC) vs the percentage of changes in $\psi's$ for both models: GMM-CorrS and GMM-Potts. Top: when $p = 200, n = 100$; Bottome: when $p = 500, n = 100$. A smaller value of DIC indicates a better model fitting.

for $\psi's$. In addition, model fitting criteria, such as DIC, can be used to select the optimal $\psi's$. For future work, we could also consider specifying priors on $\psi's$, e.g. a discrete prior over a pre-determined set of values, and conduct fully Bayesian inference.

References

[1] Donald B Rubin. "Randomization analysis of experimental data: The Fisher randomization test comment". In: *Journal of the American Statistical Association* 75.371 (1980), pp. 591–593.

[2] Donald B Rubin. "Comment: Which ifs have causal answers". In: *Journal of the American Statistical Association* 81.396 (1986), pp. 961–962.

[3] Tyler VanderWeele and Stijn Vansteelandt. "Mediation analysis with multiple mediators". In: *Epidemiologic methods* 2.1 (2014), pp. 95–115.

[4] Faming Liang. "A double Metropolis–Hastings sampler for spatial models with intractable normalizing constants". In: *Journal of Statistical Computation and Simulation* 80.9 (2010), pp. 1007–1022.

[5] David M Higdon. "Auxiliary variable methods for Markov chain Monte Carlo with applications". In: *Journal of the American statistical Association* 93.442 (1998), pp. 585–595.

[6] Nicholas G Polson, James G Scott, and Jesse Windle. "Bayesian inference for logistic models using Pólya–Gamma latent variables". In: *Journal of the American statistical Association* 108.504 (2013), pp. 1339–1349.

[7] Bradley Efron et al. "Size, power and false discovery rates". In: *The Annals of Statistics* 35.4 (2007), pp. 1351–1377.

[8] Michael A Newton et al. "Detecting differential gene expression with a semiparametric hierarchical mixture method". In: *Biostatistics* 5.2 (2004), pp. 155–176.

[9] Diane E Bild et al. "Multi-ethnic study of atherosclerosis: objectives and design". In: *American journal of epidemiology* 156.9 (2002), pp. 871–881.

[10] Yongmei Liu et al. "Methylomics of gene expression in human monocytes". In: *Human molecular genetics* 22.24 (2013), pp. 5065–5074.

[11] Pan Du, Warren A Kibbe, and Simon M Lin. "lumi: a pipeline for processing Illumina microarray". In: *Bioinformatics* 24.13 (2008), pp. 1547–1548.

[12] Yi-an Chen et al. “Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray”. In: *Epigenetics* 8.2 (2013), pp. 203–209.

[13] Maarten van Iterson, Erik W van Zwet, and Bastiaan T Heijmans. “Controlling bias and inflation in epigenome-and transcriptome-wide association studies using the empirical null distribution”. In: *Genome biology* 18.1 (2017), p. 19.

[14] Thomas F McElrath et al. “Longitudinal evaluation of predictive value for preeclampsia of circulating angiogenic factors through pregnancy”. In: *American journal of obstetrics and gynecology* 207.5 (2012), 407–e1.

[15] Kelly K Ferguson et al. “Variability in urinary phthalate metabolite levels across pregnancy and sensitive windows of exposure for the risk of preterm birth”. In: *Environment international* 70 (2014), pp. 118–124.

[16] Manori J Silva et al. “Quantification of 22 phthalate metabolites in human urine”. In: *Journal of Chromatography B* 860.1 (2007), pp. 106–112.

[17] Max T Aung et al. “Prediction and associations of preterm birth and its subtypes with eicosanoid enzymatic pathways and inflammatory markers”. In: *Scientific reports* 9.1 (2019), pp. 1–17.

[18] Andrew Gelman, Xiao-Li Meng, and Hal Stern. “Posterior predictive assessment of model fitness via realized discrepancies”. In: *Statistica sinica* (1996), pp. 733–760.