Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

Bayesian Models for Multiple Outcomes in Domains With Application to the Seychelles Child Development Study

Luo Xiao, Sally W. Thurston, David Ruppert, Tanzy M. T. Love & Philip W. Davidson

To cite this article: Luo Xiao, Sally W. Thurston, David Ruppert, Tanzy M. T. Love & Philip W. Davidson (2014) Bayesian Models for Multiple Outcomes in Domains With Application to the Seychelles Child Development Study, Journal of the American Statistical Association, 109:505, 1-10, DOI: 10.1080/01621459.2013.830070

To link to this article: https://doi.org/10.1080/01621459.2013.830070

Bayesian Models for Multiple Outcomes in Domains With Application to the Seychelles Child Development Study

Luo XIAO, Sally W. THURSTON, David RUPPERT, Tanzy M. T. LOVE, and Philip W. DAVIDSON

The Seychelles Child Development Study (SCDS) examines the effects of prenatal exposure to methylmercury on the functioning of the central nervous system. The SCDS data include 20 outcomes measured on 9-year-old children that can be classified broadly in four outcome classes or "domains": cognition, memory, motor, and social behavior. Previous analyses and scientific theory suggest that these outcomes may belong to more than one of these domains, rather than only a single domain as is frequently assumed for modeling. We present a framework for examining the effects of exposure and other covariates when the outcomes may each belong to more than one domain and where we also want to learn about the assignment of outcomes to domains. Each domain is defined by a sentinel outcome, which is preassigned to that domain only. All other outcomes can belong to multiple domains and are not preassigned. Our model allows exposure and covariate effects to differ across domains and across outcomes within domains, and includes random subject-specific effects that model correlations between outcomes within and across domains. We take a Bayesian MCMC approach. Results from the Seychelles study and from extensive simulations show that our model can effectively determine sparse domain assignment, and at the same time give increased power to detect overall, domain-specific, and outcome-specific exposure and covariate effects relative to separate models for each endpoint. When fit to the Seychelles data, several outcomes were classified as partly belonging to domains other than their originally assigned domains. In retrospect, the new partial domain assignments are reasonable and, as we discuss, suggest important scientific insights about the nature of the outcomes. Checks of model misspecification were improved relative to a model that assumes each outcome is in a single domain. Supplementary materials for this article are available online.

KEY WORDS: Bayesian variable selection; Latent variable model; Markov chain Monte Carlo; Methylmercury; Sparsity.

1. INTRODUCTION

Many studies have examined the effects of prenatal exposure to methylmercury on multiple neurodevelopmental outcomes. Often, the multiple outcomes are manifestations of a smaller number of outcome classes, latent traits, or "domains" in which they are nested. Appropriate modeling of the nesting structure allows information from outcomes in different domains to be used in a single model. For example, Thurston, Ruppert, and Davidson (2009) used data from the Seychelles Child Development Study (SCDS) and fit a model to 20 outcomes measured on children nine years of age. The outcomes were in four domains: cognition, memory, motor, and social behavior. In Budtz-Jørgensen et al. (2002), 11 outcomes measured on 7-year-old children in the Faroes Islands cohort were classified into two domains: motor and verbal.

In many analyses of multiple outcomes data (e.g., Budtz-Jørgensen et al. 2002; Thurston, Ruppert, and Davidson 2009), the nesting of outcomes into domains is determined by expert knowledge and treated as known, with each outcome assigned to exactly one domain. Often, however, it makes sense scientifically and empirically to assume that at least some outcomes belong partly to multiple domains with domain membership not known a priori; for example, Thurston, Ruppert, and Davidson (2009) found evidence that some outcomes in the SCDS are related to more than one domain. In this article, we develop a framework for investigating the effects of covariates, including exposures to environmental toxins, on multiple correlated outcomes in which the data help determine the membership of outcomes in domains and in which outcomes may belong to more than one domain. Using this framework, we reanalyze the SCDS data and find strong support for allowing outcomes to belong to several domains and for the use of covariate information to assign outcomes to domains.

In Thurston, Ruppert, and Davidson (2009), each of the 20 SCDS outcomes was assumed to belong to a single known domain, and outcomes within a domain were assumed to be exchangeable. However, for two outcomes in the motor domain, Trailmaking A and Trailmaking B, girls did much better on average than boys, whereas the opposite was true for the other motor domain outcomes. This was addressed in Thurston, Ruppert, and Davidson (2009) by the expedient of allowing Trailmaking A and Trailmaking B to have a different slope for sex

Luo Xiao is Postdoctoral Fellow, Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA (E-mail: *lxiao@jhsph.edu*). Sally W. Thurston is Associate Professor and Tanzy M. T. Love is Assistant Professor, Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642, USA (E-mail: *thurston@lost.rochester.edu*, *tanzy_love@urmc.rochester.edu*). David Ruppert is Professor of Engineering, Department of Statistical Science and School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853, USA (E-mail: *dr24@cornell.edu*). Philip W. Davidson is Professor Emeritus, Departments of Pediatrics, Environmental Medicine, and Psychiatry, University of Rochester, Rochester, NY 14642, USA (E-mail: *phil_davidson@urmc.rochester.edu*). The authors thank the referees and Associate Editor for their very helpful suggestions, and Brittany Huber and Jean Reeves for their assistance in interpreting results from the Seychelles study. Luo Xiao's work was partly supported by National Institute of Biomedical Imaging And Bioengineering, NIH grant R01-EB012547 and National Institute of Neurological Disorders and Stroke, NIH grant R01-NS060910. Sally Thurston and Tanzy Love's work was partly supported by National Institute of Environmental Health Sciences (NIEHS), NIH grant P30-ES01247. Philip Davidson's work was partly supported by NIEHS, NIH grant R01-ES008442. This work represents the opinions of the researchers and not necessarily those of the granting organizations.

Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/jasa*.

Journal of the American Statistical Association
March 2014, Vol. 109, No. 505, Applications and Case Studies
DOI: 10.1080/01621459.2013.830070

than other motor domain outcomes. However, there is no good reason why some outcomes solely in the motor domain should have a slope for sex that differs substantially from that of other outcomes solely in that domain. The situation suggests instead that some outcomes have characteristics of two or more domains. A primary motivation for the current work is that, unlike the other motor domain outcomes, the two trailmaking tests require number and letter recognition and we hypothesized that these outcomes belong partly to the cognition domain.

As do some other authors such as Budtz-Jørgensen et al. (2002), we use a joint model that incorporates information from multiple correlated outcomes. Models of this type are more powerful than separate outcome-specific models to detect exposure effects that are small but nonetheless important when a large population is exposed. Our model framework differs from the structural equation model (SEM) used for the Faroes data (Budtz-Jørgensen et al. 2002). In SEMs such as used in Budtz-Jørgensen et al. (2002) and Sanchez et al. (2005), a factor loading matrix allows different loadings of outcomes in a domain; domains are treated as unobservable latent variables. The different factor loadings allow variability in how much an outcome belongs to a domain. Although the factor loading matrix could allow outcomes to belong to more than one domain, this additional flexibility is not typically used. In the SEM model used for the Faroes data (Budtz-Jørgensen et al. 2002; Sanchez et al. 2005), each outcome was assumed to belong to one domain. The key difference between our model and both the multiple outcomes model in Thurston, Ruppert, and Davidson (2009) and the SEM model in Budtz-Jørgensen et al. (2002) is that our model uses information in the data to determine domain assignment and allows outcomes to belong to multiple domains.

The model developed here fits into the general framework of mixed-membership models as defined by Erosheva, Fienberg, and Lafferty (2004) in which objects are clustered, but have membership probability vectors which assign them weighted membership in multiple groups. In our model, we group regression outcomes into classes or domains, instead of the usual situation in which the grouping pertains to subjects or variables. Erosheva, Fienberg, and Lafferty (2004) laid out a framework in which the variety of mixed-membership models are defined by assumptions at four levels: population, sampling scheme, subject, and latent variable. For our model, the population level assumptions are of multivariate linear relationships between the jth outcome, y_j, and the known exposure, x, and covariates, z. At the sampling scheme level, the n replications of each outcome are not independent, but are assumed to have a covariance as specified by the random effects (which can be domain specific), while the outcomes are conditionally independent given the random effects and the domain assignments (the standard subject level assumption). At the latent variable level, we use a sparse prior for group membership.

Our model is useful to investigators in three major ways. First, our model is the first to allow the investigator to learn more about outcomes by seeing how individual deviations and covariate relationships determine how outcomes are assigned to domains. Second, our model is more realistic than other multiple outcome models when some outcomes measure characteristics of more than one domain or latent trait, and accounting for the partial domain memberships in the model allows us to estimate exposure and covariate effects more accurately. Finally, like other multiple outcomes models, it allows estimation of exposure and covariate effects with more power than separate models for each outcome.

From an investigator's perspective, it is not particularly useful if an outcome has a very small membership in a particular domain. This motivates a need for sparsity of possible domains to which an outcome can belong. To accommodate this, we develop a sparsity-inducing prior for the domain membership.

When applied to the Seychelles data, several outcomes were found to have partial membership in several domains in which they were not originally thought to belong. The discovery of new partial membership of outcomes to domains can give important insights into the specific nature of these neurodevelopmental or other outcomes. Posterior predictive checks for the model in which each outcome is assumed to nest in a single domain (Thurston, Ruppert, and Davidson 2009) suggested some model misspecification of pairwise correlations between outcomes. The additional flexibility from our model resulted in substantial improvements in the posterior predictive checks when applied to the same data.

There is a large literature on Bayesian factor analysis related in various degrees to our work. Here we mention a few of these papers. Ghosh and Dunson (2009) proposed default prior distributions for factor loadings that lead to efficient computation of posterior distributions. The problems caused by normal priors that they mention do not apply here since in our model the factor loadings are on a compact set, the simplex. Also, the identifiability issues that they and other authors address do not arise in our model because of our use of prior information, in particular, sentinel outcomes. Ghosh and Dunson also developed methodology for the case where the number of factors is unknown. In their correlated topic model, Blei and Lafferty (2007) used a logistic-normal model (Aitchison and Shen 1980) to model the distribution of topics in a document. Their model is somewhat similar to model (4) in Section 3; it differs from our model in that it is not designed to induce sparsity. Additionally, it used a nondiagonal covariance matrix to accommodate correlated topics. The focused topic model of Williamson et al. (2010) used a compound Dirichlet process "to decouple across-data prevalence and within-data proportion." Motivated by gene expression analysis, West (2003) introduced sparsity into factor models using priors similar to model (4), although in his model the loadings were not restricted to a simplex.

The remainder of the article is organized as follows. Section 2 introduces our model, compares it with the traditional structural equation model (SEM; Skrondal and Rabe-Hesketh 2004), gives model assumptions and priors for model parameters, and describes our MCMC sampling algorithm. Section 3 develops a sparsity-inducing prior and proposes a sampling scheme for the domain membership. Results from fitting the model to the SCDS are presented in Section 4 and simulation results are given in Section 5. We conclude with some discussion in Section 6.

2. MODEL

In our notation, we use subscripts on covariates to indicate how they are treated in the model. Each covariate is specified as having fixed effects (no shrinkage between outcomes) and/or random effects (shrinkage between related outcomes) at different nesting levels based on expert knowledge. A particular covariate may appear several times with different subscripts; the

$\mathcal{F}$ subscript indicates a fixed effect. For the ith observation, we use the vectors $\mathbf{S}_{\mathcal{F},i}$ for covariates included in overall fixed effects, $\mathbf{S}_{\mathcal{D},\mathcal{F},i}$ for covariates for which we allow domain-specific fixed effects, $\mathbf{S}_{\mathcal{D},i}$ for covariates with domain-specific random effects, and $\mathbf{S}_{\mathcal{O},i}$ for covariates with outcome-specific random effects within domains. If a covariate was included in $\mathbf{S}_{\mathcal{D},\mathcal{F},i}$, then it would be excluded from $\mathbf{S}_{\mathcal{F},i}$ and $\mathbf{S}_{\mathcal{D},i}$ for identifiability.

To illustrate basic concepts, in this paragraph we will start with the simple case where each outcome is fixed in exactly one domain. We let $d(j)$ indicate the domain in which outcome j is nested. For notational consistency, we subscript $\boldsymbol{\beta}$ (which always represents a fixed effect) and $\mathbf{b}$ (which represents a random effect) in the same way. Then the model is

$$y^*_{i,j} = \mathbf{S}^T_{\mathcal{F},i}\,\boldsymbol{\beta}_{\mathcal{F}} + \mathbf{S}^T_{\mathcal{D},\mathcal{F},i}\,\boldsymbol{\beta}_{\mathcal{D},\mathcal{F},d(j)} + \mathbf{S}^T_{\mathcal{D},i}\mathbf{b}_{\mathcal{D},d(j)} + \mathbf{S}^T_{\mathcal{O},i}\mathbf{b}_{\mathcal{O},j} + r_i + r_{\mathcal{D},d(j),i} + \epsilon_{i,j}, \quad (1)$$

where $y^*_{i,j}$ is the scaled and centered jth outcome on the ith subject, $\dim(\boldsymbol{\beta}_{\mathcal{F}}) = p_{\mathcal{F}}$, $\dim(\boldsymbol{\beta}_{\mathcal{D},\mathcal{F},d(j)}) = p_{\mathcal{D},\mathcal{F}}$, $\dim(\mathbf{b}_{\mathcal{D},d(j)}) = p_{\mathcal{D}}$, $\dim(\mathbf{b}_{\mathcal{O},j}) = p_{\mathcal{O}}$, r_i is the overall random subject effect, $r_{\mathcal{D},d(j),i}$ is the domain-specific random subject effect, $\epsilon_{i,j}$ is an error term, and $1 \le i \le n$, $1 \le j \le J$. For covariates, which we expect to have similar effects across domains and between outcomes such as exposure in the model for the Seychelles example in Thurston, Ruppert, and Davidson (2009), we allow an overall fixed effect and random domain- and outcome-specific effects. This covariate would be included in $\mathbf{S}_{\mathcal{F},i}$, $\mathbf{S}_{\mathcal{D},i}$ and $\mathbf{S}_{\mathcal{O},i}$. The slope for that covariate would be the sum of the relevant terms in $\boldsymbol{\beta}_{\mathcal{F}}$, $\mathbf{b}_{\mathcal{D},d(j)}$ and $\mathbf{b}_{\mathcal{O},j}$ in Equation (1). Covariates, for which we allow domain-specific fixed effects (such as sex and mother's IQ in the Seychelles example), would be included in $\mathbf{S}_{\mathcal{D},\mathcal{F},i}$, but not in $\mathbf{S}_{\mathcal{F},i}$. Although it would be possible to treat outcome-specific slopes as fixed, we do not consider this case because our model is appropriate when the effects of exposure and covariates on outcomes in the same domain are assumed to be similar for all outcomes within the domain and a borrowing of information across outcomes is desired.

Model (1) includes terms specific to the domain in which outcome j is nested, given by $d(j)$. These terms are $\boldsymbol{\beta}_{\mathcal{D},\mathcal{F},d(j)}$, $\mathbf{b}_{\mathcal{D},d(j)}$, and $r_{\mathcal{D},d(j),i}$. Because we will be considering models in which the membership of outcome j in domain d is not fixed, it will be convenient to expand (1) with all specific references to domain $d(j)$ excluded. Let $\boldsymbol{\lambda}_j = (\lambda_{j,1}, \ldots, \lambda_{j,D})^T$, where $\lambda_{j,d} \ge 0$ and $\sum_d \lambda_{j,d} = 1$. Each $\lambda_{j,d}$ represents the proportion of outcome j that is in domain d, allowing for the situation that outcomes may belong partially to multiple domains. Then the expanded model can be written as

$$y^*_{i,j} = \mathbf{S}^T_{\mathcal{F},i}\,\boldsymbol{\beta}_{\mathcal{F}} + \left(\boldsymbol{\lambda}_j \otimes \mathbf{S}_{\mathcal{D},\mathcal{F},i}\right)^T \boldsymbol{\beta}_{\mathcal{D},\mathcal{F}} + \left(\boldsymbol{\lambda}_j \otimes \mathbf{S}_{\mathcal{D},i}\right)^T \mathbf{b}_{\mathcal{D}} + \mathbf{S}^T_{\mathcal{O},i}\mathbf{b}_{\mathcal{O},j} + r_i + \boldsymbol{\lambda}_j^T \mathbf{r}_{\mathcal{D},i} + \epsilon_{i,j}, \quad (2)$$

where $\boldsymbol{\beta}_{\mathcal{D},\mathcal{F}} = (\boldsymbol{\beta}^T_{\mathcal{D},\mathcal{F},1}, \ldots, \boldsymbol{\beta}^T_{\mathcal{D},\mathcal{F},D})^T$, $\mathbf{b}_{\mathcal{D}}$=$(\mathbf{b}^T_{\mathcal{D},1}, \ldots, \mathbf{b}^T_{\mathcal{D},D})^T$, and $\mathbf{r}_{\mathcal{D},i} = (r_{\mathcal{D},1,i}, \ldots, r_{\mathcal{D},D,i})^T$. Therefore, $\boldsymbol{\beta}_{\mathcal{D},\mathcal{F}}$ is the vector which stacks each of the D domain-specific terms together and $\mathbf{b}_{\mathcal{D}}$ and $\mathbf{r}_{\mathcal{D},i}$ are similarly defined.

Throughout this article, the number of domains, D, is fixed. The focus of this article is to allow the data to help determine the membership of outcomes in domains, represented by $\boldsymbol{\lambda}_j$. One important source of information about the domain to which an outcome belongs is the prior information based on expert opinion. For instance, Budtz-Jørgensen et al. (2002) examined the effect of MeHg on multiple outcomes within two neurobehavioral functions: motor and verbal; here the prior information for assigning outcomes into the two functions comes from social science. In our model, prior information is used to set some $\lambda_{j,d} = 1$, meaning that outcome j belongs only to domain d. We set $\lambda_{j,d} = 1$ for one outcome in each domain. These sentinel outcomes, chosen by the subject matter expert, help to define the domain and prevent label switching.

Equation (2) shows that the data contain three sources of information on the membership of outcomes in domains: (i) random covariate effects ($\mathbf{b}_{\mathcal{D}}$); (ii) fixed covariate effects ($\boldsymbol{\beta}_{\mathcal{D},\mathcal{F}}$); and (iii) domain-specific subject random effects ($\mathbf{r}_{\mathcal{D},i}$). The $\mathbf{r}_{\mathcal{D},i}$ model the partial correlations between pairs of outcomes in the same domain, after adjusting for covariate effects. When these partial correlations are large for pairs of outcomes in the same domain relative to outcome pairs in different domains, the domain-specific subject effects will be relatively large. This was the case for the SCDS data in the model fit by Thurston, Ruppert, and Davidson (2009), suggesting that for these data the domain-specific subject effects carry substantial information about the membership of outcomes in domains.

2.1 Model Comparison

We first note that both models (1) and (2) can be regarded as structural equation models (SEM, Skrondal and Rabe-Hesketh 2004). To see this, let $\mathbf{x}_i = (\mathbf{S}^T_{\mathcal{F},i}, \mathbf{S}^T_{\mathcal{O},i})^T$ and $\mathbf{u}_i = (\mathbf{S}^T_{\mathcal{D},\mathcal{F},i}, \mathbf{S}^T_{\mathcal{D},i})^T$. Further, let $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_J)^T$, $\boldsymbol{\Gamma}$ be the $D \times (p_{\mathcal{D},\mathcal{F}} + p_{\mathcal{D}})$ matrix with dth row $(\boldsymbol{\beta}^T_{\mathcal{D},\mathcal{F},d}, \mathbf{b}^T_{\mathcal{D},d})$, and let $\mathbf{K}$ be the $J \times (p_{\mathcal{F}} + p_{\mathcal{O}})$ matrix with jth row $(\boldsymbol{\beta}^T_{\mathcal{F}}, \mathbf{b}^T_{\mathcal{O},j})$. Then model (2) can be written as

$$\begin{aligned} \mathbf{y}^*_i &= \boldsymbol{\nu} + \mathbf{K}\mathbf{x}_i + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\xi}_i \\ \boldsymbol{\eta}_i &= \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\mathbf{u}_i + \boldsymbol{\zeta}_i, \text{ so } \boldsymbol{\eta}_i = (\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{u}_i + \boldsymbol{\zeta}_i), \end{aligned} \quad (3)$$

where $\mathbf{y}^*_i$ is the J length vector of outcomes for the ith subject, $\boldsymbol{\eta}_i$ is a D length vector of latent factors for the ith subject, $\mathbf{B}$ is a $D \times D$ matrix with 0's on the diagonal, and $\mathbf{I} - \mathbf{B}$ is invertible. To get model (2), take $\boldsymbol{\nu} = \mathbf{0}$, $\boldsymbol{\alpha} = \mathbf{0}$, $\mathbf{B} = \mathbf{0}$, $\boldsymbol{\xi}_i = \boldsymbol{\epsilon}_i + r_i\mathbf{1}_J$, and $\boldsymbol{\zeta}_i = \mathbf{r}_{\mathcal{D},i}$, where $\boldsymbol{\epsilon}_i$ is the J length vector of $\epsilon_{i,j}$s for the ith subject.

Woodard et al. (2013) showed that if we take the covariance matrices of both $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ in (3) to be diagonal, then there exists a matrix $\mathbf{B}$ such that the SEM in (3) is equivalent to the model from Thurston, Ruppert, and Davidson (2009) when for each outcome j, $\lambda_{j,d}$ equals 1 for only one domain d and equals 0 for other domains. Model (3) is a continuous latent factor model, a special case of an SEM. In a conventional SEM, each outcome belongs to a single domain or factor, so each $\boldsymbol{\lambda}_j$ has only one nonzero element. The nonzero $\lambda_{j,d}$ quantifies the extent to which outcome j belongs to domain d, and unlike the Thurston, Ruppert, and Davidson (2009) model, these are not necessarily equal to 1.

When none of the latent factors are assumed to have a direct effect on any other latent factor (such as when the D latent factors are all outcome functions), then $\mathbf{B} = \mathbf{0}$ would be appropriate. In this case, the traditional SEM has the same limitations as Thurston et al.'s model when modeling correlations between outcomes. However, in our model, a key advantage of using this simple correlation structure is that the residual correlations between outcomes can help determine domain membership. Had

our model assumed a more general correlation structure, it would have been more difficult to identify partial domain membership.

When $\mathbf{B} = \mathbf{0}$, for a covariate that is included in $\mathbf{u}_i$ but not in $\mathbf{x}_i$, its slope on two outcomes in the same domain would differ only by the ratio of the corresponding $\lambda_{j,d}$'s. This ratio would apply not only to all covariate effects, but also to the variance of $\boldsymbol{\zeta}$. Also since all $\lambda_{j,d} \geq 0$, the sign of a covariate slope will be the same for all outcomes in a domain. In contrast, in our model (2), the slope for a covariate on two outcomes depends on the two $\boldsymbol{\lambda}_j$ vectors and the covariate slopes for two outcomes that share partial membership in the same domain may differ in sign.

When a latent factor may have a direct effect on another latent factor then $\mathbf{B} \neq \mathbf{0}$. For example, Budtz-Jørgensen et al. (2002) modeled latent mercury exposure, motor function and verbal function; the slopes of the relationships between these latent variables are the elements of $\mathbf{B}$. In this case since $(\mathbf{I} - \mathbf{B})^{-1}$ is not diagonal, the slope of a covariate on two outcomes entirely in the same domain may differ due to the nonzero domain-specific elements in $(\mathbf{I} - \mathbf{B})^{-1}$. This allows for a more general correlation structure between outcomes in the same domain, but assumes that the modifications to the correlation structure are adequately captured by the specific function of the slopes in $(\mathbf{I} - \mathbf{B})^{-1}$. In contrast, in our model the correlations are modified by partial domain memberships in $\boldsymbol{\Lambda}$. For example, if two outcomes are in separate domains except for common partial membership in one domain, then their domain-level correlation will depend on their loadings on the common domain, through $\boldsymbol{\Lambda}$.

In the Seychelles study, multiple domain memberships are needed for the trailmaking outcomes to properly capture their correlations with outcomes in the motor domain and cognition domains, as well as to capture the slopes for several covariates. In particular, the slope for sex was strongly positive for all motor outcomes except for the trailmaking outcomes, where it was strongly negative. Under model (2), the slope for sex will be a weighted average of domain-specific slopes, whereas in a traditional SEM the slope for sex could not be negative if Trailmaking A is in the motor domain, except through outcome-specific deviations in $\mathbf{b}_{\mathcal{O},j}$. Also, in model (2), Trailmaking A can be correlated with outcomes in the cognition domain, but there is no clear reason to assume that, conditional on r_i, the other outcomes in the motor domain should be correlated to those in the cognition domain.

2.2 Model Assumptions and Prior Distributions

Here we give our model assumptions and priors for model components except $\boldsymbol{\Lambda}$, which we will discuss in Section 3. We assume that $\mathbf{b}_{\mathcal{D},d} (1 \leq d \leq D)$, $\mathbf{b}_{\mathcal{O},j} (1 \leq j \leq J)$, $r_i (1 \leq i \leq n)$, $\mathbf{r}_{\mathcal{D},i} (1 \leq i \leq n)$, $\boldsymbol{\beta}_{\mathcal{F}}$, $\boldsymbol{\beta}_{\mathcal{D},\mathcal{F},d} (1 \leq d \leq D)$, and $\boldsymbol{\epsilon}_i (1 \leq i \leq n)$ are independent. We assume for $i = 1, \ldots, n$, $r_i \sim \mathrm{N}(0, \sigma_r^2)$, $\mathbf{r}_{\mathcal{D},i} \sim \mathrm{N}(0, \boldsymbol{\Sigma}_{r,\mathcal{D}})$, where $\boldsymbol{\Sigma}_{r,\mathcal{D}} = \mathrm{diag}(\sigma^2_{r,\mathcal{D},1}, \ldots, \sigma^2_{r,\mathcal{D},D})$ allows domain-specific variances. For the error terms, we assume $\boldsymbol{\epsilon}_i \sim \mathrm{N}(0, \boldsymbol{\Sigma}_\epsilon)$ and $\boldsymbol{\Sigma}_\epsilon = \mathrm{diag}(\sigma^2_{\epsilon,1}, \ldots, \sigma^2_{\epsilon,J})$.

We assume for $1 \leq d \leq D$, $\mathbf{b}_{\mathcal{D},d} \sim \mathrm{N}(\mathbf{0}_{p_\mathcal{D}}, \boldsymbol{\Sigma}_{b,\mathcal{D}})$, and for $1 \leq j \leq J$, $\mathbf{b}_{\mathcal{O},j} \sim \mathrm{N}(\mathbf{0}_{p_\mathcal{O}}, \boldsymbol{\Sigma}_{b,\mathcal{O}})$. As for $\boldsymbol{\Sigma}_{b,\mathcal{D}}$ and $\boldsymbol{\Sigma}_{b,\mathcal{O}}$, we can assume that $\boldsymbol{\Sigma}_{b,\mathcal{D}} = \mathrm{diag}(\sigma^2_{b,\mathcal{D},1}, \ldots, \sigma^2_{b,\mathcal{D},p_\mathcal{D}})$ and $\boldsymbol{\Sigma}_{b,\mathcal{O}} = \mathrm{diag}(\sigma^2_{b,\mathcal{O},1}, \ldots, \sigma^2_{b,\mathcal{O},p_\mathcal{O}})$ which allows covariate-specific variances.

Furthermore, we assume $\boldsymbol{\beta}_{\mathcal{F}} \sim \mathrm{N}(\boldsymbol{\beta}_{\mathcal{F},0}, \boldsymbol{\Sigma}_0)$, $\boldsymbol{\beta}_{\mathcal{D},\mathcal{F},d} \sim \mathrm{N}(\boldsymbol{\beta}_{\mathcal{D},\mathcal{F},d,0}, \boldsymbol{\Sigma}_{\mathcal{D},\mathcal{F},0})$, $\sigma_r^2 \sim \mathrm{IG}(A_{0,r}, B_{0,r})$, $\sigma^2_{r,\mathcal{D},1}, \ldots, \sigma^2_{r,\mathcal{D},D} \sim \mathrm{IG}(A_{0,r,\mathcal{D}}, B_{0,r,\mathcal{D}})$, $\sigma^2_{b,\mathcal{D},1}, \ldots, \sigma^2_{b,\mathcal{D},p_\mathcal{D}} \sim \mathrm{IG}(A_{0,b,\mathcal{D}}, B_{0,b,\mathcal{D}})$, $\sigma^2_{b,\mathcal{O},1}, \ldots, \sigma^2_{b,\mathcal{O},p_\mathcal{O}} \sim \mathrm{IG}(A_{0,b,\mathcal{O}}, B_{0,b,\mathcal{O}})$ and $\sigma^2_{\epsilon,1}, \ldots, \sigma^2_{\epsilon,J} \sim \mathrm{IG}(A_{0,\epsilon}, B_{0,\epsilon})$, where $\mathrm{IG}(A, B)$ denotes the inverse gamma distribution with shape parameter A and scale parameter B. We use hyperparameter values that correspond to weakly informative priors. We take $\boldsymbol{\beta}_{\mathcal{F},0} = \mathbf{0}$, $\boldsymbol{\Sigma}_0^{-1} = \mathrm{diag}\{0.00001\}$, $\boldsymbol{\beta}_{\mathcal{D},\mathcal{F},d,0} = \mathbf{0}$, $\boldsymbol{\Sigma}^{-1}_{\mathcal{D},\mathcal{F},0} = \mathrm{diag}\{0.00001\}$, $A_{0,b,\mathcal{D}} = A_{0,b,\mathcal{O}} = A_{0,r} = A_{0,r,\mathcal{D}} = A_{0,\epsilon} = 0.5$, $B_{0,b,\mathcal{D}} = B_{0,b,\mathcal{O}} = B_{0,r} = B_{0,r,\mathcal{D}} = 0.00005$ and $B_{0,\epsilon} = 0.0005$. These prior values are the same as in the primary prior in Thurston, Ruppert, and Davidson (2009).

2.3 Sampling Scheme

We use Markov chain Monte Carlo (MCMC) to sample from the joint posterior distribution of all model parameters. For all model parameters except the domain membership vector $\boldsymbol{\lambda}_j$, we use Gibbs sampling and iteratively sample a parameter or parameter vector from its posterior full conditional. The full conditional posteriors of the model parameters are provided in Web Supplement Section S.2. We decompose $\boldsymbol{\lambda}_j$ into a product of a binary vector $\mathbf{z}_j$ and a log-transformed weight vector $\mathbf{v}_j$ (see Section 3 for details). We sample $\mathbf{z}_j$ like other model parameters using Gibbs sampling and provide a Metropolis-Hasting sampler for $\mathbf{v}_j$. The MCMC algorithm was coded in R.

3. MODELING Λ

We write the (j, d)th element of $\boldsymbol{\Lambda}$ as

$$\lambda_{j,d} = \frac{z_{j,d} \exp(v_{j,d})}{\sum_{d'=1}^{D} z_{j,d'} \exp(v_{j,d'})}, \quad j = 1, \ldots, J, \; d = 1, \ldots, D, \tag{4}$$

where $\mathbf{Z} = (z_{j,d})_{1 \leq j \leq J, 1 \leq d \leq D}$ and $\mathbf{V} = (v_{j,d})_{1 \leq j \leq J, 1 \leq d \leq D}$. As in Ghahramani, Griffiths, and Sollich (2006) and Griffiths and Ghahramani (2006), $\mathbf{Z}$ is a binary matrix specifying as to which elements of $\boldsymbol{\Lambda}$ are nonzero. What is different here is that the exponential of $\mathbf{V}$, rather than $\mathbf{V}$ itself, contains the relative membership weights. The expression (4) implies that all elements of $\boldsymbol{\Lambda}$ are nonnegative and the sum of each row of $\boldsymbol{\Lambda}$ is 1. Because of the binary $\mathbf{Z}$, the reparameterization in (4) is an extension of the multinomial logit transform. We use the $(\mathbf{Z}, \mathbf{V})$ parameterization because priors for $(\mathbf{Z}, \mathbf{V})$ can allow some elements in $\boldsymbol{\lambda}$ to be 0, while the commonly used Dirichlet prior does not provide such flexibility.

There are some technical issues with the logit transform in (4). First, the denominator in (4) cannot be zero and hence for each j, at least one $z_{j,d}$ has to be nonzero. One approach is to use prior information for each j to fix $z_{j,d}$ at 1 for some d. Doing this requires only weak prior information; for each outcome we need to know at least one domain where it has at least partial membership. We shall denote this d by $d^*(j)$. For the Seychelles data, we can use the outcome-to-domain assignments in Thurston, Ruppert, and Davidson (2009) to select $d^*(j)$. In the unlikely case where no prior information is available to select $d^*(j)$, a prior for $\mathbf{z}_j = (z_{j,1}, \ldots, z_{j,D})^T$ has to satisfy $p(\mathbf{z}_j = \mathbf{0}) = 0$.

Second, for each j, we need to fix one $v_{j,d}$ at 0 because $\boldsymbol{\lambda}_j$ in (4) is invariant to translations of $\mathbf{v}_j = (v_{j,1}, \ldots, v_{j,D})^T$. If $d^*(j)$ exists, then a convenient choice is to fix $v_{j,d^*(j)}$ at 0; otherwise a proper prior for $\mathbf{v}_j$ has to be carefully constructed. Section S.3 of the Web Supplement discusses the more complex case where no such $d^*(j)$ is available and provides a properly constructed prior for $(\mathbf{Z}, \mathbf{V})$ that was used in the simulation study.

3.1 Prior Specification

The remainder of this section focuses on the case where for each outcome a $d^*(j)$ has been selected and $v_{j,d^*(j)}$ is fixed at 0. The method below for sampling $\mathbf{Z}$ and $\mathbf{V}$ can be easily extended to the more complex case and, hence, is omitted.

Let $\mathcal{J}$ be the index set of nonsentinel outcomes and the prior is

$$p(\mathbf{Z}, \mathbf{V}) = \prod_{j \in \mathcal{J}} p(\mathbf{z}_j, \mathbf{v}_j) = \prod_{j \in \mathcal{J}} \left\{ I\left(z_{j,d^*(j)} = 1\right) I\left(v_{j,d^*(j)} = 0\right) \times \prod_{d \neq d^*(j)} p(z_{j,d}, v_{j,d}) \right\},$$

where $I(\cdot)$ is an indicator function that equals 0 if the statement inside the parenthesis is true and 0 otherwise. This prior forces the constraints $z_{j,d^*(j)} = 1$ and $v_{j,d^*(j)} = 0$ discussed above. We let $p(z_{j,d}, v_{j,d}) = p(z_{j,d})p(v_{j,d}|z_{j,d})$, where $p(z_{j,d})$ has a Bernoulli$(p_{j,d,0})$ distribution and $p(v_{j,d}|z_{j,d}) \sim z_{j,d} \cdot \mathrm{N}(0, \sigma_v^2) + (1 - z_{j,d}) \cdot \mathrm{N}^-(0, g\sigma_v^2)$ with g being fixed and large. Here $\mathrm{N}^-(0, g\sigma_v^2)$ denotes the distribution of $-|X|$ where X has a normal distribution $\mathrm{N}(0, g\sigma_v^2)$. Note that in Bayesian variable selection models (George and McCulloch 1993; Yi, George, and Allison 2003; O'Hara and Sillanpää 2009), mixture priors, for example, of two normals, are often used for selecting variables.

We now show why the mixture prior specified above is suitable and how to construct a proper hierarchical prior for σ_v^2. We use the following equality

$$\lambda_{j,d} = z_{j,d} \exp(v_{j,d}) \lambda_{j,d^*(j)}, \tag{5}$$

where $\lambda_{j,d^*(j)} > 0$ and the equality can be derived from the transform (4). Similar to the common practice in Bayesian variable selection models (George and McCulloch 1993; Yi, George, and Allison 2003; O'Hara and Sillanpää 2009), if $z_{j,d} = 0$, by Equation (5) it is desirable to specify a prior for $v_{j,d}$ to give a small value to $\exp(v_{j,d})$, or equivalently, a large magnitude and negative value to $v_{j,d}$. Hence, we let $p(v_{j,d}|z_{j,d} = 0) \sim \mathrm{N}^-(0, g\sigma_v^2)$ with a large value of g. We let $g = 10$ in the simulation study and in the Seychelles study. Other large values of g can also be used and the robustness of our model to the choice of g is studied with simulations.

Next we construct a proper prior for σ_v^2. Because the sum of all elements in $\boldsymbol{\lambda}_j$ is 1, $v_{j,d}$ is generally not large. For instance, if $\lambda_{j,d^*(j)} = 0.15$, then by Equation (5) $v_{j,d}$ has to be smaller than 2 because $0.15 \exp(2) > 1$. Since $d^*(j)$ is usually selected with $\lambda_{j,d^*(j)}$ being relatively large, it is sensible to set an upper bound, such as 4, for σ_v^2. Furthermore it is also useful to set a lower bound, such as 1, for σ_v^2 so that the samples of $\mathbf{v}_j$ move sufficiently and do not get stuck near $\mathbf{0}$. Therefore, we place a Uniform$(1, 4)$ prior for σ_v^2. The robustness of our model to the prior specification of σ_v^2 is also evaluated in the simulation study.

3.2 Sampling $(\mathbf{Z}, \mathbf{V})$

We briefly describe our method of sampling $(\mathbf{Z}, \mathbf{V})$ from the posterior and provide further details in Web Supplement Section S.4. First it is easy to show that the full conditional posterior of a $z_{j,d}$ is Bernoulli and, hence, can be sampled by a Gibbs sampler. Next for $\mathbf{v}_j$, the full conditional posterior of which does not have a standard distribution, we consider a random-walk Metropolis–Hastings sampler used in Staudenmayer, Ruppert, and Buonaccorsi (2007).

4. RESULTS FROM THE SEYCHELLES CHILD DEVELOPMENT STUDY

The SCDS (Marsh et al. 1995; Davidson et al. 1998; Myers et al. 2003), began enrollment in 1989 to examine the effects of prenatal methylmercury (MeHg) exposure on childhood outcomes. MeHg is a known neurotoxicant, and prenatal exposure to MeHg, such as from maternal fish consumption, is thought to be particularly detrimental (Clarkson 2002). The MeHg level in maternal hair grown during pregnancy is used as a biomarker for the fetal MeHg exposure level. Pregnant mothers in the SCDS cohort consumed an average of 12 fish meals per week (Myers et al. 2003), and had MeHg hair levels averaging 6.83 ppm, nearly nine times greater than the average among U.S. women of child-bearing age who ate fish at least three times per week (McDowell et al. 2004).

Thurston, Ruppert, and Davidson (2009) developed a model for multiple outcomes nested in domains which they applied to 20 SCDS outcomes in four domains. In Thurston, Ruppert, and Davidson (2009), each outcome was considered to be part of a single domain which was specified a priori. The model adjusted for prenatal MeHg exposure and six covariates: sex (1 = male, 0 = female), maternal age, the HOME score (a measure of the stimulation of the home environment), K-BIT (a measure of maternal IQ), the Hollingshead SES (a composite measure based on parental education and employment), and the child's age at testing (averaging 9 years). The model in Thurston, Ruppert, and Davidson (2009) was fit on the 533 eligible subjects with complete covariate data who had measures of at least two outcomes in each domain. Missing outcome data were handled by a data augmentation step (Schafer 1997).

In this article we fit our model to the same 20 SCDS outcomes in four domains, using data from the same 533 subjects and using data augmentation (Schafer 1997) for missing outcome data. Like Thurston, Ruppert, and Davidson (2009), we treat MeHg exposure as having an overall fixed effect with random domain- and outcome-specific deviations, and allow each of the six covariates to have fixed domain-specific slopes. We set the appropriate $\lambda_{j,d} = 1$ for one sentinel outcome in each domain: WISC Verbal IQ (cognition domain), CVLT recognition (memory domain), grooved pegboard dominant hand (motor domain), and CBCL internalizing T score (behavior domain). These tests were selected as sentinel outcomes because they were validated for assessing the domain and were well known and widely accepted as measures of the domain (Sattler 2008; Sattler and Hoge 2006). We list these sentinel outcomes first

for the readers' convenience. Setting these values of $\lambda_{j,d}$ to 1 (and the other $\lambda_{j,d}$ to 0 for these outcomes) defines the four domains and prevents label switching. This also ensures that the denominator for $\lambda_{j,d}$ is > 0 for all j, as discussed in Section 3.

We ran the MCMC for 35,000 iterations using a single chain with starting values known to be reasonable when the membership of outcomes into domains is known. We discarded the first 5,000 iterations as burn-in. We assessed convergence for $v_{j,d}$ rather than for $\lambda_{j,d}$ since the latter is often a mixture with a point mass at zero. The Raftery-Lewis (Raftery and Lewis 1996) diagnostic indicated that 30,000 draws were sufficient for all model parameters to estimate the $q = 0.025$th quantile of all relevant model parameters to within $r = \pm 0.0125$ with $s = 95\%$ probability. The effective sample size was more than 500 for all parameters.

Table 1 shows the posterior means of the 30,000 draws for each $\lambda_{j,d}$. As anticipated, results indicated that Trailmaking A and Trailmaking B belong partly to the motor domain and partly to the cognition domain. Many other outcomes were found to have substantial membership in more than one domain, suggesting that many of these tasks depend on multiple processing levels. Our model results suggest that two of the three outcomes classified originally as belonging to the cognition domain (WISC performance IQ and Boston Naming Test: no cues) do belong primarily to this domain, but also belong partly to the motor domain. This is quite reasonable for the WISC performance IQ, which includes subtests with a motor component.

Table 1. Posterior mean values of $\lambda_{j,d}$, which indicate the allocation of outcomes to domains. Horizontal lines delineate previously assigned domains, where domain 1 is cognition, domain 2 is memory, domain 3 is motor, and domain 4 is behavior

	$D = 1$	$D = 2$	$D = 3$	$D = 4$
y1 = WISC verbal IQ*	**1.00**	0.00	0.00	0.00
y2 = WISC performance IQ	**0.77**	0.00	**0.23**	0.00
y3 = Boston naming test: no cues	**0.60**	**0.08**	**0.31**	0.00
y4 = CVLT: recognition*	0.00	**1.00**	0.00	0.00
y5 = CVLT: immediate recall	0.02	**0.97**	0.00	0.01
y6 = CVLT short delay recall	0.00	**1.00**	0.00	0.00
y7 = CVLT long delay recall	0.00	**1.00**	0.00	0.00
y8 = WISC-R digit span	**0.59**	**0.21**	0.02	**0.17**
y9 = WRAML design memory	**0.56**	**0.18**	**0.23**	0.04
y10 = Grooved pegboard dominant hand*	0.00	0.00	**1.00**	0.00
y11 = Grooved pegboard nondominant hand	**0.24**	0.00	**0.76**	0.00
y12 = Finger tapping dominant hand	0.00	0.00	**1.00**	0.00
y13 = Finger tapping nondominant hand	0.00	0.00	**1.00**	0.00
y14 = Trailmaking A	**0.60**	0.01	**0.39**	0.00
y15 = Trailmaking B	**0.78**	0.02	**0.19**	0.01
y16 = Bruinincks-Oseretsky	0.08	**0.11**	**0.79**	0.03
y17 = Beery-Buktenica	**0.45**	**0.05**	**0.37**	**0.13**
y18 = CBCL: internaling T score*	0.00	0.00	0.00	**1.00**
y19 = CBCL: externaling T score	0.00	0.00	0.00	**1.00**
y20 = CTRS hyperactivity index	**0.18**	**0.18**	**0.33**	**0.31**

NOTE: Sentinel outcomes (assigned to a domain with probability 1) are indicated by an asterisk. Bold typeface indicates that the posterior median is nonzero.

The Boston Naming Test (BNT) includes images of vehicles and tools such as a comb and a toothbrush. Psycholinguistic theory and neuroimaging studies have suggested that knowledge of objects is organized by their sensory and motor properties. The BNT may have loaded on the motor domain because people think about what they do with these tools when finding the appropriate word to match the picture.

The WISC-R digit span, originally in the memory domain, is classified as mainly in the cognition domain and only partly in the memory domain, as well as partly in the behavior domain. Part of the WISC-R digit span requires the subject to reverse digits presented aurally, a task that requires cognition. The WRAML design memory, originally classified as belonging to the memory domain, was found to be primarily in the cognitive domain and to a lesser extent in both the motor and the memory domains. Partial membership of the WRAML design memory in the cognition and motor domains is reasonable since this test has a component of spatial reasoning (a cognitive task) and a perceptual-motor component (a motor task).

Our finding that the grooved pegboard nondominant hand has a substantial cognitive component, unlike the grooved pegboard dominant hand or either fingertapping endpoint, suggests that this task requires more cognitive resources than performing the same action with the dominant hand. The absence of a cognitive component when performing the grooved pegboard task with the dominant hand is due to the neuronal efficiency in executing a task with the dominant hand. With more frequency in use and practice of a motor skill, the more automatic the movement sequence becomes as demonstrated in motor learning studies (Arbib 2003; Steele and Penhune 2010). Performing the grooved pegboard task with the nondominant hand might require higher level processing and allocation of attentional resources to plan and carry out the correct motor sequence (Strenge and Niederberger 2008). The lack of a cognitive component in the nondominant fingertapping results could be due to the less demanding nature of the task.

Our estimate for the posterior probability that the overall MeHg effect is greater than zero is 79%, which is very similar to the case when each outcome is assumed to be nested in a single, fixed domain as in Thurston, Ruppert, and Davidson (2009). The length of the 95% posterior intervals for the outcome-specific exposure slopes were very similar to Thurston, Ruppert, and Davidson (2009). Like Thurston, Ruppert, and Davidson (2009), the posterior medians of the variance components for the domain- and outcome-specific exposure effects for MeHg were very small, also suggesting that a model with a common MeHg slope might be sufficient. The posterior medians of σ_r^2, and $\sigma_{\epsilon,j}^2$ for most but not all j, were smaller than in the model with fixed membership. Of note, the posterior medians for the domain-specific subject effects, $\sigma_{r,\mathcal{D},d}^2$, $d = 1, \ldots, 4$ were considerably larger than in the model fit in Thurston, Ruppert, and Davidson (2009), suggesting that the structure of $\mathbf{\Lambda}$ explains more of the correlation between outcomes after adjusting for exposure and covariates.

We carried out the same posterior predictive checks as Thurston, Ruppert, and Davidson (2009). Like Thurston, Ruppert, and Davidson (2009), we found that most of the summary statistics from the SCDS data were consistent with those from draws from the posterior predictive data. However,

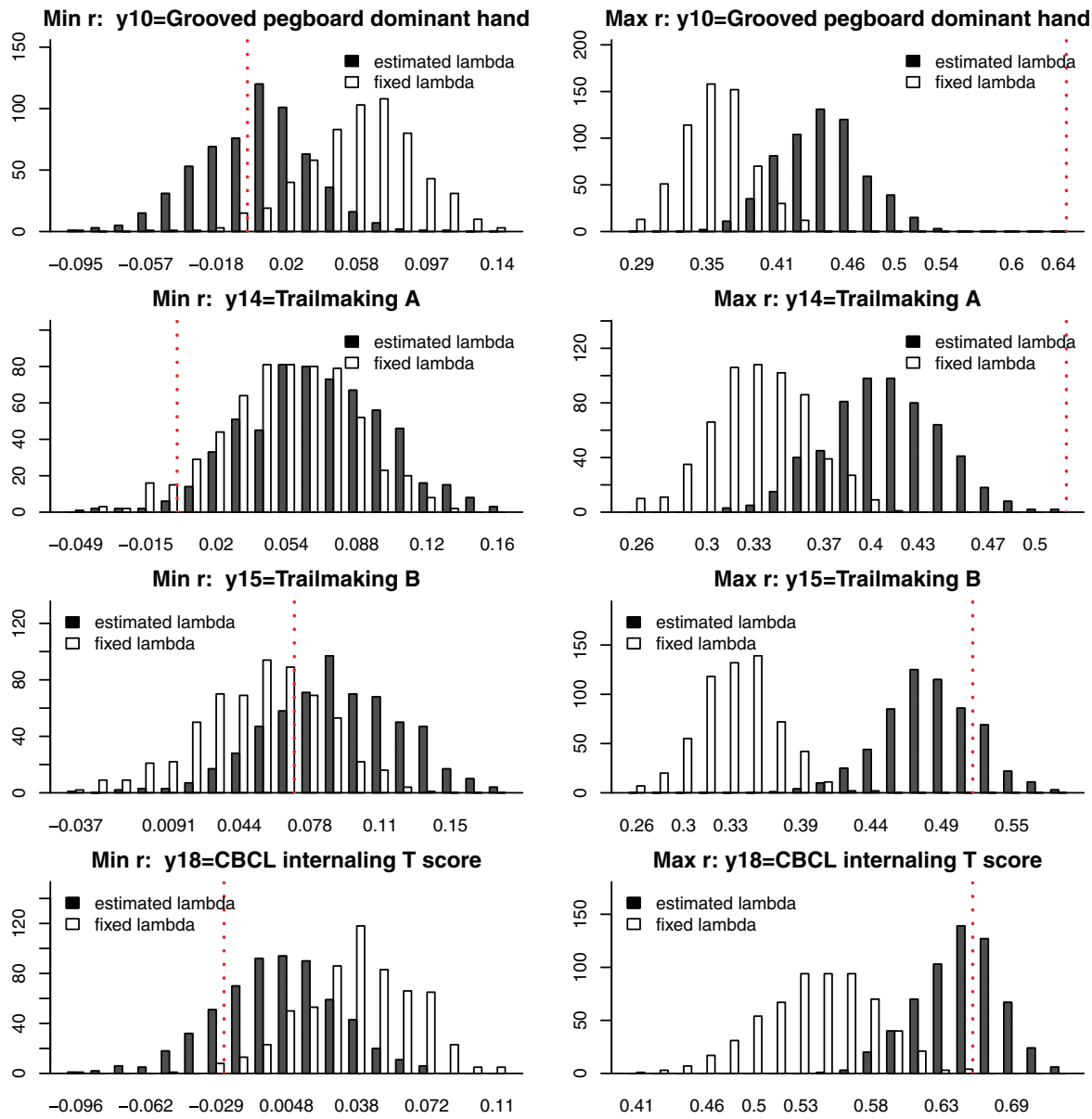

Figure 1. Histograms of the posterior predictive distribution of the minimum correlation (left column) and maximum correlation (right column) between outcome j and all other outcomes, under the model used, for four outcomes. Light bars use a model in which all $\lambda_{j,d}$ values are fixed at 0 or 1 as in Thurston, Ruppert, and Davidson (2009). Dark bars are from the model in this article. Vertical dashed lines indicate observed values in the Seychelles data.

Thurston et al. reported that for the four outcomes in which this was checked, both the maximum correlation and the minimum correlation between an outcome and the other 19 SCDS outcomes were generally very different from the corresponding values from the posterior predictive draws, suggesting a lack of model fit for these features. We examined these posterior predictive checks for all 20 outcomes under both our models and the model in Thurston, Ruppert, and Davidson (2009). The checks for the minimum correlations under our model were somewhat very reasonable in nearly all cases. We did find noticeable lack of fit for the maximum correlations for eight outcomes, of which five with most severe lack of fit (y_{10}, y_{11}, y_{12}, y_{13}, and y_{17}) were all originally classified as belonging to the motor domain. Six of the eight motor outcomes are part of a pair of closely related outcomes, such as fingertapping in the dominant and nondominant hands. The correlations between these outcome pairs are much larger than the other correlations within the "motor" domain, and our model failed to predict these very large correlations. These high correlations could be modeled by allowing $\epsilon_{i,j}$ and $\epsilon_{i,j'}$ to be correlated when outcomes j and j' form such a pair.

In Figure 1 we show the distribution of the minimum correlation and the maximum correlation for four outcomes under our model and under the model in Thurston, Ruppert, and Davidson (2009). On each plot we have superimposed a vertical dashed line showing the corresponding observed statistic in the Seychelles data. These distributions are shown for three motor outcomes, specifically the sentinel outcome, grooved pegboard dominant hand (y_{10}), Trailmaking A (y_{14}) and Trailmaking B (y_{15}), as well as for the sentinel outcome in the behavior domain, CBCL internalizing T score (y_{18}). The posterior

predictive checks for the maximum correlation for the two sentinel outcomes are poor for one outcome (y_{10}), but good for the other (y_{18}) under our model. Similarly, both the maximum and minimum correlations are reasonable for Trailmaking B, but not for Trailmaking A. These indicate that partial membership of an outcome in multiple domains is not the only factor determining fit or lack of fit. Nonetheless, the posterior predictive checks were improved relative to Thurston, Ruppert, and Davidson (2009) for nearly all outcomes. This improvement suggests that, for these data, allowing an outcome to belong partly to multiple domains better captures the correlation structure between outcomes than a similar model without this feature.

5. A SIMULATION STUDY

We use simulations to investigate our method's ability to estimate $\mathbf{\Lambda}$, and, in particular, to detect possible sparse structure in $\mathbf{\Lambda}$. The robustness of the proposed prior specification of $p(v|z)$ is also investigated. Most importantly, we compare our method to separate regressions fit to each outcome and show that our method has greater ability to detect exposure effects. This significantly strengthens the result in Thurston, Ruppert, and Davidson (2009) that multiple outcome models have greater power to detect exposure effects *when* $\mathbf{\Lambda}$ *is known exactly.*

5.1 Settings

We consider a model with $J = 20$ outcomes in $D = 4$ domains and with $n = 500$ observations. The model is

$$y_{i,j} = x_i \left(\beta_{\mathcal{F}} + \sum_{1 \le d \le D} \lambda_{j,d} b_{\mathcal{D},d} + b_{\mathcal{O},j} \right) + \mathbf{z}_i^T \left(\sum_{1 \le d \le D} \lambda_{j,d} \boldsymbol{\beta}_{\mathcal{D},\mathcal{F},d} \right) + r_i + \sum_{1 \le d \le D} \lambda_{j,d} r_{\mathcal{D},d,i} + \epsilon_{i,j},$$

for $1 \le i \le n$, $1 \le j \le J$. The model has exposure, x, and a vector of six covariates, $\mathbf{z} = (z_1, \ldots, z_6)^T$.

We designed a full 2^3 factorial experiment with three factors each with two levels: the outcome-to-domain membership matrix ($\mathbf{\Lambda}$), the fixed covariate effects ($\boldsymbol{\beta}_{\mathcal{D},\mathcal{F}}$), and the variances of the domain-specific subject effects ($\sigma^2_{r,\mathcal{D},1}, \ldots, \sigma^2_{r,\mathcal{D},D}$). Of the two different membership matrices, $\mathbf{\Lambda}_1$ is very sparse with few nonzeros while $\mathbf{\Lambda}_2$ has more nonzeros. We used two levels of fixed covariate effects, $\boldsymbol{\beta}_{\mathcal{D},\mathcal{F}} = \boldsymbol{\beta}^a_{\mathcal{D},\mathcal{F}}$ or $\boldsymbol{\beta}^b_{\mathcal{D},\mathcal{F}}$. For $\boldsymbol{\beta}^a_{\mathcal{D},\mathcal{F}}$, each of the six pairs of domains is distinguished maximally by one covariate, while the overall covariate effect is zero. For instance, the covariate effect for z_1 is $(-0.05, -0.20, 0.20, 0.05)$ which maximally differentiates domain 2 and domain 3. For $\boldsymbol{\beta}^b_{\mathcal{D},\mathcal{F}}$, the difference across domains is much smaller, and the overall covariate effects are nonzero. For instance, the covariate effect for z_1 now becomes $(0.10, 0.15, 0.15, 0.10)$. The details of the membership matrices and the fixed covariate effects are provided in Web Supplement Section S.5.

The variances of the domain-specific subject effects are the same across the domains and they are either 0.45^2 or 0.8^2, indicating small or large domain-specific subject effects. Motivated by the Seychelles study, we assume a relatively small range of exposure effects across the outcomes, and choose $\beta_{\mathcal{F}} = -0.05$, $\mathbf{b}_{\mathcal{D}} = (-0.02, 0.02, -0.01, 0.01)^T$. For simplicity, we let $\mathbf{b}_{\mathcal{O}} = \mathbf{0}^T$, that is, the outcome-specific deviations are all 0. The variance of the overall random subject effect (σ_r^2) is fixed to 0.35^2, a moderate value.

The values of σ_ϵ^2's are determined by the other model parameters with the following data generating procedure. We let the variances of all outcomes, the exposure and covariates be 1. Conditional on the exposure and covariate effects, we derive the correlation matrix of the vector $(y_1, \ldots, y_J, x, z_1, \ldots, z_6, r, r_{\mathcal{D},1}, \ldots, r_{\mathcal{D},D})^T$ and draw samples of this vector from a multivariate normal distribution. Note that the pairwise correlation between outcomes in the same domain after adjusting for covariate effects is $\sigma_r^2 + \sigma_{r,\mathcal{D}}^2$.

Our experiment creates eight different sets of models parameters (model conditions) to examine; see Web Supplement Table S.3 for details. We drew 50 datasets under each model condition. Outcomes 1, 4, 10, and 18 were set to be sentinel outcomes for the four domains. The value of g in the mixture prior for $p(v|z)$ was 10 and σ_v^2 had a uniform distribution in $(1, 4)$, as described in Section 3. Moreover, we focused on the more complex case where there is no prior information as to which domain each outcome always has a membership, that is, no information is available for selecting d^* defined in Section 3.1. We used the prior in Web Supplement Section S.3 and randomly selected pairs of (j, d) for which the vs are fixed at 0 as starting values. We used a single MCMC chain with reasonable starting values. The MCMC chain was long enough so that it has an effective sample size of at least 400 and contains more draws than required by the Raftery-Lewis diagnostic using $q = 0.025$, $r = \pm 0.0125$, and $s = 95\%$ as in Section 4.

5.2 Results

Table 2 gives the results of estimating the outcome-specific slopes for exposure from both our model and separate regressions, under one model condition. Let β_j be the exposure effect for the jth outcome and $\hat{\beta}_j$ be an estimate. We report: the median relative absolute bias which is the median of $|\hat{\beta}_j/\beta_j - 1|$; the length of posterior (confidence) intervals; the relative mean squared errors (rMSE) which is the mean of $|\hat{\beta}_j/\beta_j - 1|^2$ over all datasets; and the coverage probability of the true exposure effects. From Table 2 we see that on average the outcome-specific slopes from the joint model have smaller median relative absolute bias than those from separate regressions. Compared to the confidence intervals of outcome-specific slopes given by separate regressions, posterior intervals under our model are shorter, have smaller rMSE, and generally have higher coverage probability. Similar results can be found for other model conditions and are shown in Web Supplement Section S.5.

It is of interest to examine whether the estimates (posterior means) of exposure effects are biased or the empirical biases are mainly due to the simulation variability, since we only did 50 simulations per model condition. We use results from model condition 1 as an example. For the estimate of overall exposure effects, the empirical bias is 0.004 while the standard error is 0.003 indicating no evidence of bias in the estimate. For the estimates of outcome-specific exposure effects, the standard errors of empirical bias are around 0.003 while the empirical biases in absolute value range from 0.003 to 0.018 showing that for

Table 2. Simulation results for the exposure effects under model condition 1. Rows labeled slope 1 to slope 20 refer to the estimated exposure effects on outcomes 1 to 20. Column headings under "Model results" and "Separate regressions" give the median relative absolute bias, length of 95% posterior (confidence) intervals, relative mean squared error (rMSE) and coverage probability

		Model results				Separate regressions			
	True	Relative bias	Length	rMSE	Coverage	Relative bias	Length	rMSE	Coverage
$\beta_{\mathcal{F}}$	−0.050	0.36	0.07	0.20	1.00				
slope 1	−0.070	0.35	0.09	0.19	0.94	0.37	0.15	0.33	0.94
slope 2	−0.068	0.36	0.09	0.18	0.94	0.48	0.15	0.42	0.90
slope 3	−0.070	0.35	0.09	0.19	0.94	0.32	0.14	0.31	0.96
slope 4	−0.030	0.54	0.09	0.89	0.98	0.90	0.19	2.23	0.84
slope 5	−0.030	0.54	0.09	0.88	0.98	0.80	0.20	2.40	0.86
slope 6	−0.030	0.54	0.09	0.88	0.98	0.55	0.18	1.96	0.88
slope 7	−0.030	0.54	0.09	0.89	0.98	0.77	0.17	2.11	0.90
slope 8	−0.048	0.41	0.08	0.25	0.98	0.76	0.16	0.85	0.96
slope 9	−0.042	0.43	0.08	0.31	0.98	0.65	0.17	1.24	0.90
slope 10	−0.060	0.42	0.08	0.21	1.00	0.62	0.15	0.54	0.98
slope 11	−0.060	0.42	0.08	0.21	1.00	0.55	0.13	0.55	0.96
slope 12	−0.060	0.42	0.08	0.21	1.00	0.47	0.15	0.46	0.98
slope 13	−0.060	0.42	0.08	0.21	1.00	0.64	0.16	0.57	0.96
slope 14	−0.060	0.42	0.08	0.21	1.00	0.59	0.15	0.52	0.96
slope 15	−0.060	0.42	0.08	0.21	1.00	0.63	0.18	0.66	0.90
slope 16	−0.041	0.48	0.08	0.35	0.98	0.82	0.16	1.41	0.94
slope 17	−0.060	0.42	0.08	0.21	1.00	0.60	0.15	0.54	0.98
slope 18	−0.040	0.48	0.08	0.35	1.00	0.84	0.14	1.01	0.94
slope 19	−0.040	0.48	0.08	0.35	1.00	0.74	0.14	0.96	0.94
slope 20	−0.050	0.38	0.08	0.22	1.00	0.54	0.16	0.80	0.94

some outcomes the estimates are quite biased. We found that outcomes 1 to 3 (slopes 1 to 3 in Table 2) have the most biased estimates, which is reasonable because these estimates are substantially shrunk to the overall exposure and hence biased.

Also in Web Supplement Section S.5 are further simulation results including the posterior means of $\mathbf{\Lambda}$, and these results show that our model is capable of using information in the data to identify the sparse structure of $\mathbf{\Lambda}$. Moreover, the posterior means of $\mathbf{\Lambda}$ show that the nonzero elements in $\mathbf{\Lambda}$ were also accurately estimated.

To summarize, our simulations demonstrate that our method is capable of using the information in the data to accurately find the outcome-specific slopes by identifying the sparse structure of the outcome-to-domain membership. This can be done without pre-assigning outcomes to domains except for the sentinel outcomes which are required for defining the domains.

5.3 Sensitivity Analysis

We examined the robustness of our method for identifying the sparse structure of $\mathbf{\Lambda}$ to the prior for σ_v^2 and the value of g in the mixture prior for $p(v|z) = z \cdot \mathrm{N}(0, \sigma_v^2) + (1-z)\mathrm{N}^{-}(0, g\sigma_v^2)$. The simulations demonstrated that our method is not very sensitive to the specific value of g and the prior for σ_v^2. The details are in Web Supplement Section S.6.

6. DISCUSSION

A number of related models have examined the effect of exposure and covariates on multiple outcomes within a single model to increase power. In some cases, the outcomes belong to more than one class or "domain." Previous work assumes that each outcome belongs to a single domain known a priori (Budtz-Jørgensen et al. 2002; Thurston, Ruppert, and Davidson 2009). In contrast, we allow outcomes to belong to multiple domains determined by expert knowledge, exposure, and covariate effects, and partial correlations between outcomes.

We applied our model to data on 20 outcomes within four domains from the SCDS. Previous results suggested that at least two outcomes belonged partly to more than one domain. Our results suggest that mixed-membership of outcomes in the four domains is considerably more extensive than this.

Results from our model may be sensitive to the choice of sentinel outcomes. We recommend that sentinel outcomes always be chosen by the subject-matter expert so that domains are established sensibly. An alternative method would not use sentinel outcomes, but this would require handling label switching and the results might be less interpretable since there would not necessarily be any outcome belonging to only one domain.

Although the membership of outcomes in domains will differ with different choices for sentinel outcomes, in this application these differences are not likely to have much effect on the overall or outcome-specific exposure effects. Our model gives substantially more flexibility than a model in which each outcome is assigned to a single domain a priori. When applied to the SCDS data, our model found an improvement in model fit as compared to the model that assigns each outcome to a single domain. Our model is particularly useful in elucidating the membership of specific outcomes in domains. When sentinel domains cannot be identified a priori, further work to determine how sensitive results are to different choices of sentinel outcomes would be of interest.

SUPPLEMENTARY MATERIALS

Section S.1 discusses further model formulation. Section S.2 provides full conditional posteriors for model components. Section S.3 constructs a general prior for the outcome-domain membership. Section S.4 gives details on sampling the outcome-domain membership. Sections S.5 and S.6 provide additional simulation results and a sensitivity analysis.

[Received January 2012. Revised July 2013.]

REFERENCES

Aitchison, J., and Shen, S. (1980), "Logistic-Normal Distributions: Some Properties and Uses," *Biometrika*, 67, 261–272. [2]

Arbib, M. A. (2003), *The Handbook of Brain Theory and Neural Networks*, Cambridge, MA: MIT Press. [6]

Blei, D., and Lafferty, J. (2007), "A Correlated Topic Model of Science," *The Annals of Applied Statistics*, 1, 17–35. [2]

Budtz-Jørgensen, E., Keiding, N., Grandjean, P., and Weihe, P. (2002), "Estimation of Health Effects of Prenatal Methylmercury Exposure Using Structural Equation Models," *Environmental Health*, 1, 2. [1,2,3,4,9]

Clarkson, T. (2002), "The Three Modern Faces of Mercury," *Environmental Health Perspectives*, 110, 11–23. [5]

Davidson, P., Myers, G., Cox, C., Axtell, C., Shamlaye, C., Sloane-Reeves, J., Cernichiari, E., Needham, L., Choi, A., Wang, Y., et al. (1998), "Effects of Prenatal and Postnatal Methylmercury Exposure From Fish Consumption on Neurodevelopment Outcomes at 66 Months of Age in the Seychelles Child Development Study," *Journal of the American Medical Association*, 280, 701–707. [5]

Erosheva, E., Fienberg, S., and Lafferty, J. (2004), "Mixed-Membership Models of Scientific Publications," *Proceedings of the National Academy of Sciences of the United States of America*, 101, 5220–5227. [2]

George, E., and McCulloch, R. (1993), "Variable Selection Via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881–889. [5]

Ghahramani, Z., Griffiths, T., and Sollich, P. (2006), "Bayesian Nonparametric Latent Feature Models," *Bayesian Statistics*, 8, 201–226. [4]

Ghosh, J., and Dunson, D. (2009), "Default Prior Distributions and Efficient Posterior Computations in Bayesian Factor Analysis," *Journal of Computational and Graphical Statistics*, 18, 306–282. [2]

Griffiths, T., and Ghahramani, Z. (2006), "Infinite Latent Feature Models and the Indian Buffet Process," *Advances in Neural Information Processing Systems*, 18, 475–499. [4]

Marsh, D., Clarkson, T., Myers, G., Davidson, P., Cox, C., Cernichiari, E., Tanner, M., Lednar, W., Shamlaye, C., Choisy, O., et al. (1995), "The Seychelles Study of Fetal Methylmercury Exposure and Child Development: Introduction," *Neurotoxicology*, 16, 583–596. [5]

McDowell, M., Dillon, C., Osterloh, J., Bolger, P., Pellizzari, E., Fernando, R., De Oca, R., Schober, S., Sinks, T., Jones, R., et al. (2004), "Hair Mercury Levels in U.S. Children and Women of Childbearing Age: Reference Range Data From Nhanes 1999–2000," *Environmental Health Perspectives*, 112, 1165. [5]

Myers, G., Davidson, P., Cox, C., Shamlaye, C., Palumbo, D., Cernichiari, E., Sloane-Reeves, J., Wilding, G., Kost, J., Huang, L., et al. (2003), "Prenatal Methylmercury Exposure From Ocean Fish Consumption in the Seychelles Child Development Study," *The Lancet*, 361, 1686–1692. [5]

O'Hara, R., and Sillanpää, M. J. (2009), "A Review of Bayesian Variable Selection Methods: What, How and Which," *Bayesian Analysis*, 4, 85–118. [5]

Raftery, A., and Lewis, S. (1996), "Implementing MCMC," *Markov Chain Monte Carlo in Practice*, 115–130. [6]

Sanchez, B., Budtz-Jørgensen, E., Ryan, L., and Hu, H. (2005), "Structural Equation Models," *Journal of the American Statistical Association*, 100, 1443–1455. [2]

Sattler, J. (2008), *Assessment of Children: Cognitive Foundations*, La Mesa, CA: J. M. Sattler, Publisher Inc. [5]

Sattler, J., and Hoge, R. (2006), *Assessment of Children: Behavioral, Social, and Clinical Foundations*, La Mesa, CA: J. M. Sattler, Publisher Inc. [5]

Schafer, J. (1997), *Analysis of Incomplete Multivariate Data*, New York: Chapman and Hall. [5]

Skrondal, A., and Rabe-Hesketh, S. (2004), *Generalized Latent Variable Modeling*, Boca Raton, FL: Chapman Hall/CRC. [2,3]

Staudenmayer, J., Ruppert, D., and Buonaccorsi, J. (2007), "Density Estimation in the Presence of Heteroscedastic Measurement Error," *Journal of the American Statistical Association*, 103, 726–736. [5]

Steele, C. J., and Penhune, V. B. (2010), "Specific Increases Within Global Decreases: A Functional Magnetic Resonance Imaging Investigation of Five Days of Motor Sequence Learning," *The Journal of Neuroscience*, 30, 8332–8341. [6]

Strenge, H., and Niederberger, U. (2008), "Unidirectional Interference in Use of Nondominant Hand During Concurrent Grooved Pegboard and Random Number Generation Tasks," *Perceptual and Motor Skills*, 106, 763–774. [6]

Thurston, S., Ruppert, D., and Davidson, P. (2009), "Bayesian Models for Multiple Outcomes Nested in Domains," *Biometrics*, 65, 1078–1086. [1,2,3,4,5,7,8,9]

West, M. (2003), "Bayesian Factor Regression Models in the 'Large p, Small n' Paradigm," *Bayesian Analysis*, 7, 723–732. [2]

Williamson, S., Wang, C., Heller, K. A., and Blei, D. M. (2010), "The IBP Compound Dirichlet Process and Its Application to Focused Topic Modeling," in *Proceedings of the 27th International Conference on Machine Learning*, pp. 1151–1158. [2]

Woodard, D., Love, T., Thurston, S., and Ruppert, D. (2013), "Latent Factor Regression Models for Grouped Outcomes," *Biometrics*, 69, 785–794. [3]

Yi, N., George, V., and Allison, D. (2003), "Stochastic Search Variable Selection for Identifying Multiple Quantitative Trait Loci," *Genetics*, 164, 1129–1138. [5]