

Latent Factor Regression Models for Grouped Outcomes

D. B. Woodard

School of Operations Research and Information Engineering, Cornell University
and

T. M. T. Love, S. W. Thurston

Department of Biostatistics and Computational Biology, University of Rochester
and

D. Ruppert

Department of Statistical Science and School of Operations Research
and Information Engineering, Cornell University
and

S. Sathyanarayana

Departments of Occupational and Environmental Health Science and Pediatrics,
University of Washington
and

S. H. Swan

Department of Preventive Medicine, Mount Sinai School of Medicine.

SUMMARY: We consider regression models for multiple correlated outcomes, where the outcomes are nested in domains. We show that random effect models for this nested situation fit into a standard factor model framework, which leads us to view the modeling options as a spectrum between parsimonious random effect multiple outcomes models and more general continuous latent factor models. We introduce a set of identifiable models along this spectrum that extend an existing random effect model for multiple outcomes nested in domains. We characterize the tradeoffs between parsimony and flexibility in this set of models, applying them to both simulated data and data relating sexually dimorphic traits in male infants to explanatory variables. Supplementary material is available in an online appendix.

KEY WORDS: epidemiology, factor analysis, multiple outcomes, regression.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

Multiple-outcome regression models pool information across related outcome variables; this can lead to higher power to detect a significant predictor effect than fitting separate regression models (Thurston et al., 2009). Such joint models are popular, for instance, in epidemiological studies, which often have multiple measures of physiological or psychological health and attempt to detect small but important effects of low-dose exposure on those outcomes. In such contexts it is crucial to use the available information as efficiently as possible. We use as a motivating example data from the Study for Future Families (Swan et al. 2003), relating sexually dimorphic traits in male infants to explanatory variables.

There are two general approaches to modeling the relationship of predictor variables to multiple correlated outcomes like these. One approach models the predictor effect on the outcomes directly (Sammel, Lin, and Ryan 1999; Lin et al. 2000; Coull, Hobert, Ryan, and Holmes 2001; Roy, Lin, and Ryan 2003) and induces correlations between outcomes with random effects. Another approach, called the continuous latent factor approach, introduces one or more continuous unobserved variables that are manifested by the multiple outcomes (Dunson, 2000; Muthén, 2002; Budtz-Jorgensen et al., 2003; Sanchez et al., 2005). The direct modeling approach includes the case where the outcomes are nested in domains (Thurston et al., 2009). This is a common situation in epidemiology studies, where one is interested in the relationship of the explanatory variables to a set of outcomes within domains such as motor function, intelligence, and attention. In the Study for Future Families the outcomes fall into natural domains including skinfold thickness measures, other “fatness” measures like weight and body mass index, and head circumference (which forms its own domain).

We show that some direct modeling approaches are special cases of the continuous latent factor model framework, even when the multiple outcomes are nested in domains. In other words, the two modeling frameworks above are closely related. This is not surprising

since continuous latent factor models are extremely general, and non-identifiable in the unrestricted case. However, expressing the direct modeling approach in this way suggests extensions and allows us to view the options for modeling grouped outcomes as a spectrum between parsimonious, but less flexible, random effect models and highly parameterized, but more flexible, latent factor models. We introduce a set of models along this spectrum, starting with a single restriction to the latent factor model framework. We show that this restriction is enough to ensure identifiability, while maintaining the ability to model complex dependencies (Section 2). So all our models are identifiable, while capturing desirable features like correlations between the latent variables and different correlations between different pairs of outcomes in the same domain, features that are more commonly obtained by designing very context-specific models and separately verifying identifiability.

Our models provide a set of general-purpose tools for the situation of multiple outcomes nested in domains, yielding estimates of the predictor effects both at the domain level and at the outcome level. They share information across outcomes, in part by shrinkage of the estimated outcome-specific predictor effect across outcomes in a domain. The models differ in the degree of shrinkage, in the most parsimonious case assuming that there is a common predictor effect for all outcomes in the same domain.

We characterize the tradeoffs between parsimony and flexibility in our models by applying them to both simulated data (Section 3) and to the Study for Future Families (Section 4). We demonstrate the estimation accuracy of the models, and show how to select from among the models using goodness-of-fit measures.

2. Modeling Grouped Outcomes

First we describe the random effect model of Thurston et al. (2009) for multiple outcomes nested in domains, and show that it is a type of continuous latent factor model in the sense of Sanchez et al. (2005), with one factor for each domain. The continuous latent factor model

is itself a special case of a structural equation model (SEM: cf. Sanchez et al. 2005). So we traverse a spectrum from parsimony to flexibility as we go from random effect models to latent variable models to SEMs, and the model-choice decision is not between a SEM and a random effect model, but rather about the appropriate amount of parsimony when considering model restrictions within the SEM framework.

Denote the outcome measurements by Y_{ij} for subjects $i = 1, \dots, n$ and outcomes $j = 1, \dots, p$. Although we focus on the case of continuous outcomes, one can handle the discrete case by use of the generalized linear model framework. The outcomes are grouped into domains $d(j) \in \{1, \dots, d\}$, which are defined to contain strongly positively correlated outcomes. Denote the covariates by the length- r vector \mathbf{Z}_i , and the (observed) exposure by η_i . In accordance with the epidemiology literature we distinguish notationally between these two sets of predictors, although they will be modeled in identical fashion; one can also drop η_i in the following models in order to obtain a single undifferentiated set of predictors.

The model of Thurston et al. (2009) extends the linear mixed model approach to borrow information across outcomes and domains while estimating the exposure effect. It provides shrinkage of this effect across domains and across outcomes within a domain, and has higher power to detect an effect than separate regression models. Their original model is as follows, where the outcome variables, exposure, and covariates are assumed to be standardized, and where the notation $\overset{\text{ind}}{\sim}$ indicates that the random effects are independently distributed.

$$Y_{ij} = (b_{o,\eta,j} + b_{D,\eta,d(j)} + b_\eta)\eta_i + (\mathbf{b}_{o,z,j} + \mathbf{b}_{D,z,d(j)} + \mathbf{b}_z)\mathbf{Z}_i + q_i + q_{i,d(j)} + e_{ij} \quad (1)$$

where b_η is a common exposure effect, $b_{D,\eta,k} \overset{\text{ind}}{\sim} N(0, \tau_D^2)$ for $k = 1, \dots, d$ is a domain-specific exposure effect, $b_{o,\eta,j} \overset{\text{ind}}{\sim} N(0, \tau_o^2)$ is an outcome-specific exposure effect, \mathbf{b}_z is a vector of overall covariate effects, $b_{D,z,k,\ell} \overset{\text{ind}}{\sim} N(0, \tau_{D,\ell}^2)$ is a domain-specific covariate effect for the ℓ th covariate, $b_{o,z,j,\ell} \overset{\text{ind}}{\sim} N(0, \tau_{o,\ell}^2)$ is an outcome-specific covariate effect for the ℓ th covariate, $q_i \overset{\text{ind}}{\sim} N(0, \tau_q^2)$ is a subject-specific random effect, $q_{i,k} \overset{\text{ind}}{\sim} N(0, \tau_{q,k}^2)$ is a subject-domain effect,

and $e_{ij} \stackrel{\text{ind}}{\sim} N(0, \sigma_j^2)$ is the residual error. The subject random effect q_i captures the situation where all outcome measures are positively correlated even after accounting for covariates and exposure. The subject-domain effect $q_{i,k}$ captures additional correlation between outcomes within a domain. No intercept parameters are included by Thurston et al. (2009) in model (1) due to centering of outcomes, exposure, and covariates. The class of models proposed by Thurston et al. (2009) allows the domain-specific exposure and covariate effects, $b_{\text{D},\eta,d(j)}$ and $\mathbf{b}_{\text{D},z,d(j)}$, to be treated either as random effects as in (1) or as fixed effects.

In contrast with (1), a continuous latent factor model induces correlation between related outcomes by assuming that they are all manifestations of a set of common unmeasurable variables. The general form of a continuous latent factor model is the following (Sammel and Ryan, 1996; Muthén, 2002; Sanchez et al., 2005), where we take the number of factors equal to the number of domains:

$$\begin{aligned} \mathbf{Y}_i &= \boldsymbol{\alpha} + \boldsymbol{\beta}_{o,\eta}\eta_i + \boldsymbol{\beta}_{o,z}\mathbf{Z}_i + \boldsymbol{\Lambda}\boldsymbol{\xi}_i + \boldsymbol{\epsilon}_i \\ \boldsymbol{\xi}_i &= \boldsymbol{\beta}_{\text{D},\eta}\eta_i + \boldsymbol{\beta}_{\text{D},z}\mathbf{Z}_i + \mathbf{B}\boldsymbol{\xi}_i + \boldsymbol{\zeta}_i. \end{aligned} \quad (2)$$

Here, \mathbf{Y}_i is the length- p vector of outcomes for the i th subject, $\boldsymbol{\alpha}$ is a length- p vector of intercepts, $\boldsymbol{\beta}_{o,\eta}$ and $\boldsymbol{\beta}_{o,z}$ are $p \times 1$ and $p \times r$ matrices of regression coefficients, $\boldsymbol{\Lambda}$ is a $p \times d$ matrix of factor loadings, $\boldsymbol{\xi}_i$ is a length- d vector of latent factors, $\boldsymbol{\epsilon}_i$ is a length- p vector of independent residuals such that $\epsilon_{ij} \stackrel{\text{ind}}{\sim} N(0, \sigma_j^2)$, $\boldsymbol{\beta}_{\text{D},\eta}$ and $\boldsymbol{\beta}_{\text{D},z}$ are $d \times 1$ and $d \times r$ matrices of regression coefficients, $\boldsymbol{\zeta}_i$ is a length- d vector such that $\zeta_{i,k} \stackrel{\text{ind}}{\sim} N(0, \tau_{\zeta,k}^2)$, and \mathbf{B} is a $d \times d$ matrix with zero diagonal elements and $(\mathbf{I} - \mathbf{B})$ invertible. Without further restrictions this model is non-identifiable.

To see that (1) is a special case of (2), specify the factor loadings matrix $\boldsymbol{\Lambda}$ so that the latent traits $\boldsymbol{\xi}_i$ correspond to the outcome domains. To do this, the nonzero elements of $\boldsymbol{\Lambda}$ should be the elements $(j, d(j))$ for each j , which we call λ_j . For example, if there are two domains and

four outcomes with $d(1) = d(2) = 1$ and $d(3) = d(4) = 2$ then $\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & \lambda_4 \end{pmatrix}^T$.

For identifiability, it is common practice in factor analysis to set $\lambda_j = 1$ for the first outcome measurement j in each domain (Sanchez et al., 2005).

Next, a parsimonious way to induce \mathbf{B} is to take the second line of (2) to be $\boldsymbol{\xi}_i = \boldsymbol{\beta}_{\mathbf{D},\eta}\eta_i + \boldsymbol{\beta}_{\mathbf{D},z}\mathbf{Z}_i + \phi_i + \boldsymbol{\psi}_i$, where $\phi_i \stackrel{\text{ind}}{\sim} N(0, \tau_\phi^2)$ is a scalar and $\boldsymbol{\psi}_i$ is a length- d vector such that $\psi_{ik} \stackrel{\text{ind}}{\sim} N(0, \tau_{\psi,k}^2)$. This is shown in Appendix A to be a special case of (2). Using these specifications,

$$\begin{aligned} Y_{ij} &= \alpha_j + \beta_{o,\eta,j}\eta_i + \boldsymbol{\beta}_{o,z,j}\mathbf{Z}_i + \lambda_j\xi_{i,d(j)} + \epsilon_{ij} & i = 1, \dots, n; j = 1, \dots, p \\ \xi_{ik} &= \beta_{\mathbf{D},\eta,k}\eta_i + \boldsymbol{\beta}_{\mathbf{D},z,k}\mathbf{Z}_i + \phi_i + \psi_{ik} & k = 1, \dots, d \end{aligned} \quad (3)$$

where $\boldsymbol{\beta}_{o,z,j}$ and $\boldsymbol{\beta}_{\mathbf{D},z,k}$ are the j th and k th rows of the matrices $\boldsymbol{\beta}_{o,z}$ and $\boldsymbol{\beta}_{\mathbf{D},z}$, respectively. With these choices of $\mathbf{\Lambda}$ and \mathbf{B} , a rich covariance structure can be captured, since $\text{Cov}(Y_{ij}, Y_{i\ell} \mid \eta_i, \mathbf{Z}_i) = \lambda_j\lambda_\ell \left(\tau_\phi^2 + \tau_{\psi,d(j)}^2 \mathbf{1}_{\{d(j)=d(\ell)\}} \right)$ for $j \neq \ell$; i.e., the covariance/correlation can be different for different pairs of outcomes, even within a single domain.

One could simplify (3) by setting $\lambda_j = 1$ for every j , so that each outcome variable puts equal weight on the associated latent factor $\xi_{i,d(j)}$:

$$Y_{ij} = \alpha_j + (\beta_{o,\eta,j} + \beta_{\mathbf{D},\eta,d(j)})\eta_i + (\boldsymbol{\beta}_{o,z,j} + \boldsymbol{\beta}_{\mathbf{D},z,d(j)})\mathbf{Z}_i + \phi_i + \psi_{i,d(j)} + \epsilon_{ij}. \quad (4)$$

The outcome-specific exposure and covariate effects for outcome j in (3) are $\text{OS}_{\eta,j} \triangleq \beta_{o,\eta,j} + \lambda_j\beta_{\mathbf{D},\eta,d(j)}$ and $\text{OS}_{z,j,\ell} \triangleq \beta_{o,z,j,\ell} + \lambda_j\beta_{\mathbf{D},z,d(j),\ell}$, respectively, which simplify in (4) to $\beta_{o,\eta,j} + \beta_{\mathbf{D},\eta,d(j)}$ and $\beta_{o,z,j,\ell} + \beta_{\mathbf{D},z,d(j),\ell}$, i.e. an outcome-level effect plus a domain-level effect. To obtain (1) from (4), drop the intercept term α_j and use the random effect specification $\beta_{\mathbf{D},\eta,k} \stackrel{\text{ind}}{\sim} N(b_\eta, \tau_{\mathbf{D}}^2)$, $\beta_{\mathbf{D},z,k,\ell} \stackrel{\text{ind}}{\sim} N(b_{z,\ell}, \tau_{\mathbf{D},\ell}^2)$, $\beta_{o,\eta,j} \stackrel{\text{ind}}{\sim} N(0, \tau_o^2)$, and $\beta_{o,z,j,\ell} \stackrel{\text{ind}}{\sim} N(0, \tau_{o,\ell}^2)$ for each k , j and ℓ , after standardizing the outcome variables, covariates, and exposure.

We will investigate models that fit into the framework (3), using the random effect assumption that $\beta_{o,\eta,j} \stackrel{\text{ind}}{\sim} N(0, \tau_o^2)$ and $\beta_{o,z,j,\ell} \stackrel{\text{ind}}{\sim} N(0, \tau_{o,\ell}^2)$ and taking $\lambda_j = 1$ for the first

outcome in each domain. We will see that this framework is identifiable without further restrictions. The random effect specification of $\beta_{o,\eta,j}$ and $\beta_{o,z,j,\ell}$ also has the advantage of leading to shrinkage estimation of the exposure/covariate effects for outcomes within each domain. When shrinkage of these effects across domains is desirable we will also specify $\beta_{D,\eta,k}$ and $\beta_{D,z,k,\ell}$ as random effects. Unlike (1) we will include the intercept terms α_j in our models; dropping the intercept term ignores the uncertainty in that intercept when estimating the parameters of interest, potentially affecting the quality of interval estimates.

In deriving model (3) we have made assumptions regarding the structure of the matrices \mathbf{A} and \mathbf{B} ; contrast these choices with those standard in the SEM literature, where it is conventional to assign the latent factors particular interpretations, such as “motor function” and “verbally mediated function.” Having done that, one manually selects a small number of nonzero elements in the matrices \mathbf{B} and \mathbf{A} corresponding to hypothesized associations among the interpreted factors and between the interpreted factors and the outcome variables (Sanchez et al., 2005; Palomo et al., 2007). Like these authors, we associate each latent factor with a domain, and assign each of the outcomes to a single domain. However, we avoid manually specifying the relationships between the latent factors, instead assuming that the latent factors are related to each other by inclusion of the subject random effect ϕ_i , and potentially by random effect modeling of the domain-specific coefficients $\beta_{D,\eta,k}$ and $\beta_{D,z,k,\ell}$.

The subject random effect ϕ_i captures positive correlation between all of the outcomes, conditional on exposure and covariates; to enforce this we restrict $\lambda_j > 0$ for each j . This is appropriate in many contexts, for instance the SFF context of Section 4 and the democratization example of Palomo et al. (2007). When not all of the outcomes are positively correlated it may be possible to multiply some of the outcomes by -1 so that our models can be applied; for instance, in the methylmercury analysis of Budtz-Jorgensen et al. (2003), where for most outcome variables a higher value indicates better neurological development,

but in a few it indicates worse development. In our models one could potentially allow $\lambda_j = 0$ in addition to $\lambda_j > 0$. However, allowing $\lambda_j = 0$ simultaneously for all j leads to non-identifiability, and it is reasonable to require the latent factor associated with a domain to influence all of the outcomes within that domain.

2.1 Model Formulation

Next we define a set of models for grouped outcomes, from flexible latent factor models to parsimonious random effect models. All variables are standardized before model-fitting.

Model A: Given in (3), treating $\beta_{\mathbf{D},\eta,k}$, $\beta_{\mathbf{D},z,k}$, α_j , and λ_j as fixed effects restricting to $\lambda_j > 0$, and recalling that $\lambda_j = 1$ for the first outcome in each domain and that $\beta_{o,\eta,j} \sim N(0, \tau_o^2)$, $\beta_{o,z,j,\ell} \sim N(0, \tau_{o,\ell}^2)$, $\phi_i \sim N(0, \tau_\phi^2)$, and $\psi_{ik} \sim N(0, \tau_{\psi,k}^2)$ are random effects. Recall that (3) is obtained from the latent factor model (2) by specifying \mathbf{A} , \mathbf{B} as described.

Model B: Identical to Model A except that it models $\beta_{\mathbf{D},\eta,k}$ as a random effect, $\beta_{\mathbf{D},\eta,k} \sim N(b_\eta, \tau_{\mathbf{D}}^2)$ (reducing the effective number of free parameters in the model). Since this induces shrinkage of $\beta_{\mathbf{D},\eta,k}$ across domains k , it is only reasonable if the effect of η_i is believed to be similar for the outcomes in all domains.

Model C: Identical to Model A except that $\beta_{o,\eta} = \beta_{o,z} = 0$. This gives a more parsimonious model while allowing the outcome-specific predictor effects (now $\text{OS}_{\eta,j} = \lambda_j \beta_{\mathbf{D},\eta,d(j)}$ and $\text{OS}_{z,j,\ell} = \lambda_j \beta_{\mathbf{D},z,d(j),\ell}$) to be different for outcomes in the same domain.

Model D: Identical to Model A except that $\lambda_j = 1$ for all j , so each outcome puts the same weight on the latent factors. Close to the model (1) of Thurston et al. (2009).

Model E: Identical to Model D except that $\beta_{o,\eta} = \beta_{o,z} = 0$. Here the outcome-specific effects $\text{OS}_{\eta,j} = \beta_{\mathbf{D},\eta,d(j)}$ and $\text{OS}_{z,j,\ell} = \beta_{\mathbf{D},z,d(j),\ell}$ are the same for all outcomes in a domain.

All our models capture positive correlation between the outcomes, conditional on the predictors. Models A-D allow the predictor effects $\text{OS}_{\eta,j}$ and $\text{OS}_{z,j,\ell}$ to be different for

outcomes within a domain. Model B leads to shrinkage of the exposure effect across domains. In Appendix B we prove that our most general model (Model A) is identifiable so long as there is more than one domain and more than one outcome in each domain. If a particular domain has only one outcome, we show that identifiability can be achieved by setting $\beta_{o,\eta,j} = \tau_{\psi,d(j)}^2 = 0 = \beta_{o,z,j}$ for that outcome j . When applying Models A-E in practice small modifications may be needed to accommodate features seen in the data; in the Study for Future Families, for instance, we allow nonzero correlation between the residuals ϵ_{ij} and $\epsilon_{ij'}$ of two particular outcomes j, j' .

2.2 Estimation

We use Bayesian inference in the above models. All prior distributions not given in the model descriptions are specified as follows. The parameters $\log(\lambda_j)$, α_j , $\beta_{D,\eta,k}$, and the elements of the vector $\beta_{D,z,k}$, are given prior distributions that are uniform on the real line. For the variance parameters σ_j^2 , τ_o^2 , $\tau_{o,\ell}^2$, τ_D^2 , τ_ϕ^2 , and $\tau_{\psi,k}^2$, we use a uniform prior on the associated standard deviation (Gelman et al., 2004), with support on the interval from zero to two. This upper bound is reasonable: none of the variance parameters is expected to be greater than one due to standardization of the outcomes, but we use a slightly higher upper bound since in some cases the likelihood can be high near and just above one. To understand why variance parameter values with high likelihood typically are ≤ 1 , take the example of Model A, and consider an arbitrary outcome j . Then $\text{Var}(Y_{ij}|\eta_i, \mathbf{Z}_i) = \lambda_j^2 \tau_\phi^2 + \lambda_j^2 \tau_{\psi,d(j)}^2 + \sigma_j^2$. For parameter vectors with high likelihood, the model typically explains some of the variability in the outcome Y_{ij} in the sense that $\text{Var}(Y_{ij}|\eta_i, \mathbf{Z}_i) \leq \text{Var}(Y_{ij})$. Due to standardization of Y_{ij} , we then have $\lambda_j^2 \tau_\phi^2 + \lambda_j^2 \tau_{\psi,d(j)}^2 + \sigma_j^2 \leq 1$ for each j . This implies that $\sigma_j^2 \leq 1$, and taking j to be the first outcome in an arbitrary domain k yields $\lambda_j = 1$ and thus $\tau_\phi^2, \tau_{\psi,k}^2 \leq 1$. Similar arguments show that the standard deviations τ_o and $\tau_{o,\ell}$ of $\beta_{o,\eta,j}$ and $\beta_{o,z,j,\ell}$ should be ≤ 1 .

While the prior distributions for λ_j , α_j , $\beta_{D,\eta,k}$, and $\beta_{D,z,k}$ are nonintegrable, the posterior distributions for Models A-E are integrable (well-defined). We assessed prior sensitivity for the variance parameters by changing the upper bounds on the standard deviations; the results in Sections 3-4 were insensitive to increases in the upper bound and to moderate decreases in the upper bound. Not surprisingly, decreasing the upper bound so far as to truncate the region of high likelihood changed the parameter estimates in an undesirable way.

Computation is performed by Markov chain Monte Carlo. Parameter point estimates are taken to be the posterior mean, and $(1 - a)$ interval estimates for $a \in (0, 1)$ are given by the $a/2$ and $(1 - a/2)$ posterior quantiles. For the variance parameters and λ_j we find the posterior mean on the log scale and then exponentiate to obtain the point estimate, since the posterior distributions of these parameters are right-skewed. We verify convergence of the Markov chain by ensuring that the estimated Monte Carlo standard error is less than 1% of each parameter's posterior standard deviation (equivalently, the effective sample sizes are > 10000 ; Flegal, Haran, and Jones 2008). In the simulation study we allow slightly higher Monte Carlo standard error (5% of each the posterior standard deviation; equivalently, the effective sample sizes are > 400) to reduce computation time since we analyze a large number of scenarios. Using stricter convergence criteria can only be expected to improve our simulation results.

3. Simulation Study

3.1 Simulation Conditions

We performed a simulation study to compare the performance of the five models, evaluating the accuracy of point and interval estimation and assessing goodness-of-fit. We used one exposure η_i and one covariate Z_i , which were generated from a bivariate normal distribu-

tion with mean zero, variance one and covariance 0.2. The outcomes and predictors were standardized as described in Section 2.

Here we report results for Models A, B, D, and E; in additional results the performance of Model C was generally in between those of Models E and A, and was similar to A when the parameters $\beta_{o,\eta,j}$ and $\beta_{o,z,j}$ were not too large. We cannot compare the results of our models to those of a generic continuous latent factor model, or a generic SEM, because these frameworks are non-identified without further restrictions. In addition to varying the models used to simulate and fit the data, we also varied six factors each with two levels: (a) sample size ($n=100$ or 500); (b) number of outcomes in each domain, either $(4, 2, 1)$ or $(4, 6, 3)$; (c) λ_j values (all equal to 1, or some = 0.5); (d) $\bar{\beta}_{D,\eta} \triangleq \frac{1}{3} \sum_{k=1}^3 \beta_{D,\eta,k}$, either 0.05 or 0.15; (e) $\beta_{D,\eta,k} - \bar{\beta}_{D,\eta}$, either $(-0.03, 0, 0.03)$ or $(-0.1, 0, 0.1)$ (unless simulating from Model B, in which case either $\beta_{D,\eta,k} \sim N(\bar{\beta}_{D,\eta}, .03^2)$ or $\beta_{D,\eta,k} \sim N(\bar{\beta}_{D,\eta}, .1^2)$); and (f) $\beta_{D,z,k} - \bar{\beta}_{D,z}$, either $(-0.05, 0, 0.05)$ or $(-0.2, 0, 0.2)$ where $\bar{\beta}_{D,z} \triangleq \frac{1}{3} \sum_{k=1}^3 \beta_{D,z,k}$. For simulations in which some $\lambda_j = 0.5$, for the model with $(4,2,1)$ outcomes we used λ_j values of $(1, 1, 0.5, 0.5, 1, 0.5, 1)$ and for the model with $(4,6,3)$ outcomes, we used λ_j values of $(1, 1, 0.5, 0.5, 1, 1, 1, 0.5, 0.5, 0.5, 1, 0.5, 0.5)$. For all models we used $\bar{\beta}_{D,z} = 0.2$, $\tau_\phi = 0.2$, $\tau_{\psi,k} = 0.05$ for $k = 1, 2, 3$, and $\sigma_j = 1$ for all j . For Models A, B, D the standard deviations τ_o and $\tau_{o,1}$ of $\beta_{o,\eta,j}$ and $\beta_{o,z,j}$ were 0.05.

We generated 25 datasets for each of the following scenarios. For experiment one we both simulated from and fit Model A, using several combinations of factors (a)-(f) as shown in Table 1. For experiment two we tried several combinations of simulation model, estimation model, and λ_j values as shown in Table 2, while fixing: $n = 500$; $(4, 6, 3)$ outcomes in the domains; $\bar{\beta}_{D,\eta} = 0.15$; $\beta_{D,\eta,k} - \bar{\beta}_{D,\eta}$ equal to $(-0.1, 0, 0.1)$; and $\beta_{D,z,k} - \bar{\beta}_{D,z}$ equal to $(-0.2, 0, 0.2)$.

3.2 Evaluation

We evaluated the bias, relative bias (bias / true parameter value) and root mean squared error (RMSE) of point estimates, as well as coverage and width of 95% interval estimates

for the parameters $\beta_{D,\eta,k}$, $\beta_{D,z,k}$, $OS_{\eta,j}$, $OS_{z,j}$, and λ_j . Since each model has multiple $\beta_{D,\eta,k}$ parameters, we averaged the performance measures across the multiple parameters in such cases. For relative bias, since some of the parameters can take values arbitrarily close to zero we divide the average of the bias by the average of the parameter true value, which is strictly positive in all cases considered. We also calculated four goodness-of-fit measures: the estimated root mean squared error of approximation (RMSEA; Browne & Cudeck 1992), the p-value for a test of whether population RMSEA is less than .05 (a recommended threshold for close fit), the p-value for a test of whether population RMSEA = 0 (exact fit), and the average residual correlations of the outcomes. RMSEA is a measure of how well the model covariance matrix approximates the empirical covariance matrix.

As shown in Tables 1 and 2, the absolute bias and RMSE are small for the parameters of interest $\beta_{D,\eta,k}$, $\beta_{D,z,k}$, $OS_{\eta,j}$ and $OS_{z,j}$, under most scenarios. The true values of the outcome-specific effects $OS_{\eta,j}$ in this simulation study typically range from $-.15$ to $.35$, while the values of the domain-specific effects $\beta_{D,\eta,k}$ range from $-.05$ to $.25$. By contrast the absolute bias is below .03 for $OS_{\eta,j}$ and below .025 for $\beta_{D,\eta,k}$ (ten times smaller than the parameter range) under all scenarios in which the true model is nested within the fitted model and Model B is not fit (eliminating Columns 3, 4, 6, 8, and 10 of Table 2). The RMSE is $\leq .11$, which is roughly five times smaller than the parameter range for $OS_{\eta,j}$ and three times smaller than that for $\beta_{D,\eta,k}$. The bias and RMSE are even smaller if we additionally restrict to scenarios having the larger sample size $n = 500$. In these cases the absolute bias is less than .004 for $OS_{\eta,j}$ and .014 for $\beta_{D,\eta,k}$, and the RMSE is less than .06 for both parameters.

The relative bias of $\beta_{D,\eta,k}$, $\beta_{D,z,k}$, $OS_{\eta,j}$ and $OS_{z,j}$ is typically below 20% in absolute value (again ignoring Columns 3, 4, 6, 8, and 10 of Table 2); the only exceptions are in cases where both $n = 100$ and the average true parameter value happens to be very small ($\leq .07$). We have found the relative bias to be sensitive to what we choose the true values to be, tending

to reflect the particular simulation study choices more than the estimation accuracy (in order to even report these measures the average true values had to be chosen to be nonzero), so comparing the bias to the spread of the true values as above provides more insight.

The estimates of the less well-identified parameter λ_j are less accurate, with absolute bias up to .75 and RMSE from .40 to .78 (compare to true values in the range 0.5 to 1.0). This is lessened somewhat by increasing the sample size; compare Columns 2-5 of Table 1. So caution should be taken in interpreting the λ_j estimates, but the estimates of the parameters of interest $\beta_{D,\eta,k}$, $\beta_{D,z,k}$, $OS_{\eta,j}$ and $OS_{z,j}$ are reliable when the model is chosen carefully.

The bias and RMSE for all parameters are higher, and the interval coverage is lower, when the true model is A and we fit the more parsimonious models D or E; compare Column 2 to Columns 3-4 in Table 2. On the other hand, when the simulated model is D the bias and RMSE are smaller when fitting Model D than when fitting Model A; compare Columns 5 and 7 in Table 2. This demonstrates the advantage of fitting a parsimonious model when that model is correct. We do not see this advantage when the simulated model is B; compare Columns 9 & 10 in Table 2. In general our results from fitting Model B are worse than those from fitting Model A (see also Columns 5-6 of Table 2); this could be due to the fact that Table 2 uses large true deviations $\beta_{D,\eta,k} - \bar{\beta}_{D,\eta}$, so that Model B is not very appropriate.

The confidence interval coverage is excellent ($> 90\%$) in all scenarios shown in Tables 1-2, except when either the true model is not nested in the fitted model, or when Model B is fit. As expected, the confidence intervals are substantially narrower when fitting Models B, D, and E than when fitting Model A.

Several of the goodness-of-fit measures, namely residual correlations, estimated RMSEA, and p-values for the test of whether population $RMSEA \leq .05$, give evidence of good fit in all cases. The average residual correlations are less than 8%, the average RMSEA values are always less than .05 (the cutoff for “good fit”), and the p-values for the test of whether

population RMSEA is $\leq .05$ are typically very close to one. This holds even when the true model is not nested in the fitted model, meaning that the lack of fit in these cases was not detected using these measures. However, p-values for the test of whether population RMSEA = 0 are close to zero (indicating lack of fit) when the data are drawn from Models A or D and Model E is fit (Columns 4 & 8 of Table 2). This does not occur when the true model is A and Model D is fit (Column 3 of Table 2), perhaps because the true values of λ_j are not very far from one, making the lack of fit hard to detect. These results suggest that testing population RMSEA is a reasonable approach to distinguish between models. While in the somewhat idealized setting of simulated data we need the more sensitive test of whether RMSEA = 0, in real-data situations we believe that the more conservative measure of whether RMSEA $\leq .05$ is more appropriate, for reasons described in Browne and Cudeck (1992).

4. Data Analysis

4.1 *Study for Future Families*

The Study for Future Families (SFF) is a pregnancy cohort study measuring a variety of infant anthropometric characteristics (Swan et al. 2003), including four skinfold thickness metrics, body mass index, weight percentile-for-age, and head circumference percentile-for-age. Most of these traits are known to be strongly sexually dimorphic; in particular, skinfold thickness measures tend to be larger in females at all ages up to three years, while head circumference and weight are larger in males than females of the same age (Rodriguez et al. 2004; U.S. Centers for Disease Control and Prevention 2000). These metrics fall into three natural domains: (1) skinfold thickness metrics, (2) weight percentile and BMI, which are closely related, and (3) head circumference percentile. We investigate explanatory variables that may be related to these sexually dimorphic traits in male infants, including exposure to phthalates (toxic chemicals that are believed to have an anti-androgenic effect) as measured

using the “phthalate score” (Swan, 2008). Other variables considered include infant’s age and gestational age, mother’s age at time of birth, mother’s race, mother’s educational level, mother’s smoking status, and the creatinine concentration for the urine sample from which the phthalate measurements were taken. We performed a preliminary analysis by regressing each of the anthropometry measurements on the covariates and phthalate score. After appropriate transformations all variables are standardized before fitting all regression models. This analysis gave evidence of a relationship between infant’s age, gestational age, mother’s age, and mother’s race (Caucasian/non-) and some of the anthropometry measures. We include these four predictors in our models, as well as the creatinine concentration (although it was not a significant predictor for the outcomes, it is a significant predictor for the phthalate concentrations and is included in order to adjust for this effect; Barr et al. 2005). We restrict to infants with complete data, leaving 118 male infants out of 172.

Table 3 gives summary statistics for the outcomes and covariates. It also shows the regression coefficient for phthalate score obtained from the separate regressions, with 95% confidence intervals. We do not see a significant relationship between phthalate score and the outcomes using the separate regressions; however, by fitting the multiple-outcomes models from Section 2.1 one would have higher power to detect a potential effect.

4.2 Results

We apply Models A, C, D, and E, omitting Model B because we do not hypothesize that the phthalate effect is similar across domains. The third domain defined in Section 4.1 has only one outcome variable (head circumference), so for identifiability we set $\beta_{o,\eta,j} = \tau_{\psi,d(j)}^2 = 0 = \beta_{o,z,j}$ for this outcome j . Models A, C, D, and E were appropriate since residuals from the separate regressions done in Section 4.1 were positively correlated for the different outcomes (having correlations .08 – .65).

The residual correlations of skinfold thickness quadriceps and skinfold thickness triceps

were large in all of the models (e.g., .32 in Model A and .40 in Model C), indicating that these two outcomes are more strongly correlated than was captured by the models. We added a random effect to the four models to account for this; with the random effect, the residual correlations for all pairs of outcomes in all models were $\leq .20$. These residual correlations are small and do not indicate lack of fit of any of the models (for variables X and Y with correlation .20, a regression of Y on X has $r^2 = .20^2$, meaning that the fraction of variance in Y explained by X is .20²; then the standard error of prediction of Y by X is only $(1 - \sqrt{1 - .20^2}) = 2.0\%$ less than the standard deviation of Y). However, the point estimates of the goodness-of-fit measure RMSEA (see Section 3) are .080, .120, .077, and .129 for Models A, C, D, and E respectively. Also, the p-values for a test of whether population RMSEA is $\leq .05$ are .089, 8.0×10^{-5} , .096, and 2.2×10^{-6} , respectively. Both of these measures indicate (Browne and Cudeck, 1992) that Models A and D have reasonable fit, while Models C and E have poor fit to these data. One should use the most parsimonious model that shows good fit, namely Model D for these data.

Although Model D is more parsimonious than separate regression models, we do not find a significant relationship between phthalate exposure and the anthropometry outcomes (the 95% interval estimates of the outcome-specific and domain-specific exposure effects $OS_{\eta,j}$ and $\beta_{D,\eta,j}$ all contain zero). If a link exists between phthalates and the anthropometric measures, we are unable to detect it due to the size of our dataset and the need to adjust for covariates.

Selected outcome-specific covariate effects $OS_{z,j}$ for Model D are shown in Table 4. Results for Model A are similar; the set of significant outcome-specific covariate effects is the same, the point estimates of $OS_{z,j}$ are different by 11% on average between the two models, and the left and right interval estimate endpoints differ by 5% and 10%, respectively. So there is not much sensitivity of the results to the choice of model, among those models for which the RMSEA hypothesis test indicates good fit.

We find a positive relationship between gestational age and weight percentile, which is in accordance with previous findings (U.S. Centers for Disease Control and Prevention, National Center for Health Statistics, 2000). We also find a negative relationship between infant age and several skinfold thickness measures as well as weight percentile. Such relationships are surprising, in part since weight percentile is already adjusted for infant age. These negative correlations also exist in the raw data; however, if we restrict to infants in the most rapid phase of growth (younger than 9 months) these become positive correlations, which are in accordance with previous findings. Regarding the domain-specific covariate effects $\beta_{D,z,k,\ell}$, in Model D none is significant, and in Model A only the coefficient for age in the skinfold thickness domain is significant (having 95% interval estimate $(-.71,-.08)$).

Figure 1 illustrates the varying degrees of shrinkage of the estimated effects $OS_{z,j}$ in Models A, C, D, E. Estimates from Model A show shrinkage towards the domain average, relative to estimates from the separate regressions described in Section 4.1. Model D shows a slight amount of shrinkage relative to Model A. Model E restricts to a single coefficient estimate per domain, and this estimate is close to the average of the outcome-specific estimates from either Model A or separate regressions. By contrast, no shrinkage between domains is visible; this is due to the fact that the domain-specific coefficients $\beta_{D,z,k}$ are fixed effects in these models. The coefficient estimates from Model C, at the bottom of Figure 1, also show shrinkage but are somewhat different than those of the other models, perhaps due to the poor fit of Model C for these data. Although the shrinkage of coefficients in Model A relative to separate regressions appears moderate in Figure 1, it can be dramatic when the signal-to-noise ratio is low. Compare Table 3, Column 3 to the phthalate coefficient estimates ($\times 100$) from Model A ($-3.3, -0.3, -2.4, 1.0, 9.5, 9.5$ and -1.4 , respectively); the coefficient estimates are strongly pulled together.

5. Conclusions

We introduced identifiable models for regression with multiple outcomes nested in domains, from very general continuous latent factor models to very parsimonious random effect models. Our models extend existing models for this context, in the sense that they introduce outcome-specific loadings λ_j for latent factors associated with the domains. Our methods are appropriate in situations where all outcomes are positively correlated conditional on the predictors; this can be checked empirically by verifying that the residuals from separate regressions are positively correlated. It is appropriate in some situations to multiply a subset of outcomes by -1 to get this property.

We applied our models to simulated data and the SFF data. In simulations we evaluated estimation accuracy as well as goodness-of-fit approaches to model selection. All models except Model B had high accuracy in estimating the outcome-specific and domain-specific predictor effects, and had excellent coverage of interval estimates, when the true model was nested in the fitted model. When the latter does not hold, the point estimation accuracy and coverage of interval estimates are lower. Conversely, there is a beneficial effect of fitting the more parsimonious models when the data come from those models. We did not find evidence supporting the use of Model B, perhaps because we used true domain effects that differed widely and Model B is intended for situations where these effects are similar. The parameters λ_j are less well-identified so they are estimated with more bias and higher error than the parameters of interest; care should be taken when interpreting the λ_j point estimates for small to moderate sample sizes.

Regarding goodness-of-fit measures in the simulation study, a test of population RMSEA found good fit whenever the true model was nested within the fitted model, and often detected lack of fit when this was not the case. So we recommend using this hypothesis test to find the subset of appropriate models; among these models the most parsimonious model

should be chosen, but all models with good fit can be used in a sensitivity analysis. Although in the somewhat idealized simulation setting the more sensitive test of whether population $RMSEA = 0$ was needed to accurately detect lack of fit, we argued that for real-data settings the more conservative test of whether $RMSEA \leq .05$ is more appropriate.

In the SFF analysis the RMSEA hypothesis test indicated that Models A and D had good fit. So we chose the more parsimonious Model D, obtaining estimates of the outcome-specific and domain-specific covariate effects, and noted that Model A gave very similar results. We also used the SFF data to illustrate the differing degrees of shrinkage of the outcome-specific effects induced by the various models. Models A, D, and E gave very similar estimates apart from the shrinkage. Models A, C, D, and E led to shrinkage of the outcome-specific effects within domains but not across domains, because they treat the domain-level effects as fixed and the outcome-level effects (present in A & D) as random. Desirably, the shrinkage is more dramatic in cases where the signal-to-noise ratio appears to be low (as evidenced by wide coefficient interval estimates that contain zero).

Our methods extend the multiple-outcomes random effect models of Sammel et al. (1999), Lin et al. (2000), Roy et al. (2003), Thurston et al. (2009), and others. Our Model D is similar to the model of Thurston et al. (2009), while our Models A, B and C are extensions via inclusion of the factor loadings λ_j and our Model E is a parsimonious special case of Model D. The factor loadings allow for complex dependencies, like different correlations for different pairs of outcomes in the same domain, that cannot be captured by Model D. However, our work lends additional support to Model D, showing for instance that it is appropriate for the SFF data.

Like in the scaled linear mixed and scaled linear marginal models of Lin et al. (2000) and Roy et al. (2003), we use rescaled versions of the outcome variables, so that the predictor effects can be compared across outcomes (the scaling in the three cases is done differently).

Also like these authors, we include random effects to account for correlation of the outcomes conditional on the predictors. However, unlike these authors we handle the situation of outcomes nested in domains, obtaining estimates of both the outcome-specific and the domain-specific predictor effects, and capturing higher correlation of outcomes in the same domain than outcomes in different domains.

Our models also relate to some latent variable models for multivariate longitudinal data (Oort 2001, Sivo 2001, and others; see Verbeke, Fieuws, Molenberghs and Davidian 2012 for an overview), in the sense that both approaches define a set of continuous latent factors that are modeled as a function of an exogenous variable (exposure or time). The fundamental difference is that in the longitudinal context the outcomes are measured at multiple time points for each subject, while in our context outcomes are measured for a single value of the exposure, for each subject. This affects what types of models can be used, due to identifiability and practical considerations; for instance, in the some longitudinal contexts, unlike ours, one can use autoregressive models for the latent factors (Oort, 2001).

Acknowledgements. The authors thank the referees and Associate Editor for their suggestions. Research partly supported by: National Science Foundation grants CMMI-0926814 and DMS-1209103; National Institute of Environmental Health Sciences, NIH grant P30-ES01247; and STAR grant RD832515 from the U.S. Environmental Protection Agency.

References

- Barr, D. B., Wilder, L. C., Caudill, S. P., Gonzalez, A. J., Needham, L. L., and Pirkle, J. L. (2005). Urinary creatinine concentrations in the U.S. population: Implications for urinary biologic monitoring measurements. *Environmental Health Perspectives* **113**, 192–200.
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. Wiley & Sons, New York.
- Browne, M. W. and Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research* **21**, 230–258.
- Budtz-Jorgensen, E., Keiding, N., Grandjean, P., Weihe, P., and White, R. F. (2003). Statistical methods for the evaluation of health effects of prenatal exposure. *Environmetrics* **14**, 105–120.

- Coull, B. A., Hobert, J. P., Ryan, L. M., and Holmes, L. B. (2001). Crossed random effect models for multiple outcomes in a study of teratogenesis. *Journal of the American Statistical Association* **96**, 1194–1204.
- Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society, Series B.* **62**, 355–366.
- Flegal, J. M., Haran, M., and Jones, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statistical Science* **23**, 250–260.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*. Chapman and Hall, Boca Raton, FL, 2nd edition.
- Lin, X., Ryan, L., Sammel, M., Zhang, D., Padungtod, C., and Xu, X. (2000). A scaled linear mixed model for multiple outcomes. *Biometrics* **56**, 593–601.
- Muthén, B. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika* **29**, 81–117.
- Oort, F. J. (2001). Three-mode models for multivariate longitudinal data. *British Journal of Mathematical and Statistical Psychology* **54**, 49–78.
- Palomo, J., Dunson, D. B., and Bollen, K. (2007). Bayesian structural equation modeling. In Lee, S., editor, *Handbook of Latent Variable and Related Models*, pages 163–188, Amsterdam. Elsevier.
- Rodriguez, G., Samper, M. P., Ventura, P., Moreno, L. A., Olivares, J. L., and Perez-Gonzalez, J. M. (2004). Gender differences in newborn subcutaneous fat distribution. *European Journal of Pediatrics* **163**, 457–461.
- Roy, J., Lin, X., and Ryan, L. M. (2003). Scaled marginal models for multiple continuous outcomes. *Biostatistics* **4**, 371–383.
- Sammel, M., Lin, X., and Ryan, L. (1999). Multivariate linear mixed models for multiple outcomes. *Statistics in Medicine* **18**, 2479–2492.
- Sammel, M. D. and Ryan, L. M. (1996). Latent variable models with fixed effects. *Biometrics* **52**, 650–663.
- Sanchez, B. N., Budtz-Jorgensen, E., Ryan, L. M., and Hu, H. (2005). Structural equation models: A review with applications to environmental epidemiology. *Journal of the American Statistical Association* **100**, 1443–1455.
- Sivo, S. A. (2001). Multiple indicator stationary time series models. *Structural Equation Modeling: A Multidisciplinary Journal* **8**, 599–612.
- Swan, S. H. (2008). Environmental phthalate exposure in relation to reproductive outcomes and other health endpoints in humans. *Environmental Research* **108**, 177–184.
- Swan, S. H., Brazil, C., Drobnis, E. Z., Liu, F., Kruse, R. L., Hatch, M., Redmon, J. B., Wang, C., and Overstreet, J. W. (2003). Geographic differences in semen quality of fertile U. S. males. *Environmental Health Perspectives* **111**, 414–420.
- Thurston, S. W., Ruppert, D., and Davidson, P. W. (2009). Bayesian models for multiple outcomes nested in domains. *Biometrics* **65**, 1078–1086.
- U.S. Centers for Disease Control and Prevention, National Center for Health Statistics (2000). CDC Growth Charts. URL: <http://www.cdc.gov/growthcharts>.
- Verbeke, G., Fieuws, S., Molenberghs, G., and Davidian, M. (2012). The analysis of multivariate longitudinal data: A review. *Statistical Methods in Medical Research* **In press**.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

[Figure 1 about here.]

Web Appendix for “Latent Factor Regression Models for Grouped Outcomes” by D. B. Woodard, T. Love, S. W. Thurston, D. Ruppert, S. Sathyanarayana and S. H. Swan

Appendix : Derivation of the Matrix \mathbf{B}

We wish to specify $\tau_{\zeta,k}^2$ and \mathbf{B} in such a way that the factor vector $\boldsymbol{\xi}_i$, defined by

$$\boldsymbol{\xi}_i = \boldsymbol{\beta}_{\mathbf{D},\eta}\eta_i + \boldsymbol{\beta}_{\mathbf{D},z}\mathbf{Z}_i + \mathbf{B}\boldsymbol{\xi}_i + \boldsymbol{\zeta}_i \quad (\text{A.1})$$

where $\zeta_{i,k} \stackrel{\text{ind}}{\sim} N(0, \tau_{\zeta,k}^2)$, can be rewritten as follows.

$$\xi_{i,k} = \beta_{\mathbf{D},\eta,k}\eta_i + \beta_{\mathbf{D},z,k}\mathbf{Z}_i + \phi_i + \psi_{i,k} \quad k = 1, \dots, d$$

where $\phi_i \stackrel{\text{ind}}{\sim} N(0, \tau_{\phi}^2)$, $\psi_{i,k} \stackrel{\text{ind}}{\sim} N(0, \tau_{\psi,k}^2)$, and $\beta_{\mathbf{D},z,k}$ is the k th row of the matrix $\boldsymbol{\beta}_{\mathbf{D},z}$. First, let's rewrite (A.1) as

$$\boldsymbol{\xi}_i = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\beta}_{\mathbf{D},\eta}\eta_i + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\beta}_{\mathbf{D},z}\mathbf{Z}_i + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\zeta}_i$$

Since $(\mathbf{I} - \mathbf{B})^{-1}$ is a one-to-one transformation, we can reparametrize using $\boldsymbol{\beta}_{\mathbf{D},\eta}^* = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\beta}_{\mathbf{D},\eta}$ and $\boldsymbol{\beta}_{\mathbf{D},z}^* = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\beta}_{\mathbf{D},z}$. Suppressing the *s, we now have

$$\boldsymbol{\xi}_i = \boldsymbol{\beta}_{\mathbf{D},\eta}\eta_i + \boldsymbol{\beta}_{\mathbf{D},z}\mathbf{Z}_i + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\zeta}_i.$$

Our problem now has been simplified to that of specifying $\tau_{\zeta,k}^2$ and \mathbf{B} in such a way that $(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\zeta}_i \stackrel{\text{d}}{=} \boldsymbol{\psi}_i^*$, where $\boldsymbol{\psi}_i^*$ is the vector with elements $\phi_i + \psi_{i,k}$ for $k = 1, \dots, d$. Let's equivalently find an invertible matrix $\mathbf{A} = (\mathbf{I} - \mathbf{B})^{-1}$ such that $\mathbf{A}\boldsymbol{\zeta}_i \stackrel{\text{d}}{=} \boldsymbol{\psi}_i^*$ and such that the diagonal entries of \mathbf{A}^{-1} are equal to one. Since the vector $\boldsymbol{\psi}_i$ has a multivariate normal distribution, and since ϕ_i is normally distributed independent of $\boldsymbol{\psi}_i$, the vector $\boldsymbol{\psi}_i^*$ has a multivariate normal distribution. So we can write $\boldsymbol{\psi}_i^* \stackrel{\text{d}}{=} \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{X}_i$ where \mathbf{X}_i is a vector of i.i.d. standard normal random variables, $\boldsymbol{\mu} = E(\boldsymbol{\psi}_i^*) = 0$, and $\boldsymbol{\Sigma} = \text{Var}(\boldsymbol{\psi}_i^*)$ is the invertible variance-covariance matrix of $\boldsymbol{\psi}_i^*$. So for any values of the $\tau_{\zeta,k}^2 > 0$, taking $\mathbf{A} = \boldsymbol{\Sigma}^{1/2}\mathbf{C}$ where \mathbf{C} is a diagonal matrix with entries $1/\tau_{\zeta,k}$, we have that $\mathbf{A}\boldsymbol{\zeta}_i \stackrel{\text{d}}{=} \boldsymbol{\psi}_i^*$. The inverse $\mathbf{A}^{-1} = \mathbf{C}^{-1}\boldsymbol{\Sigma}^{-1/2}$ exists; furthermore, by taking $1/\tau_{\zeta,k}$ equal to the k th diagonal entry of $\boldsymbol{\Sigma}^{-1/2}$ ensures that \mathbf{A}^{-1} has diagonal entries equal to one.

Appendix : Identifiability

We show the identifiability of our most general model (A), for the case where η_i is included in the model but \mathbf{Z}_i is not; \mathbf{Z}_i can be handled analogously to η_i . We assume that there is more than one domain, and that the values η_i are i.i.d. from some unknown distribution with mean zero and variance one (the case of any finite mean and variance proceeds in the same fashion but is notationally cumbersome). We also assume that each domain either has more than one outcome, or if it has only one outcome j then we assume that $\beta_{o,\eta,j} = \tau_{\psi,d(j)}^2 = 0$.

Proof of identifiability can proceed by setting the empirical moments of \mathbf{Y}_i and η_i equal to the moments implied by the model (3), and showing that the solution for the parameter values is unique (Bollen, 1989). For model (3)

such uniqueness can be shown by looking at the first and second moments; in particular we will use $E(\mathbf{Y}_i)$, the first-moment vector for the outcomes; $\text{Var}(\mathbf{Y}_i)$, the variance-covariance matrix of the outcomes; and $\text{Cov}(\mathbf{Y}_i, \eta_i)$, the vector of covariances between η_i and the elements of the vector \mathbf{Y}_i . First, $E(Y_{ij}) = \alpha_j$, so α_j is identified for each outcome j . Next,

$$\begin{aligned} \text{Cov}(Y_{ij}, \eta_i) &= \text{Cov}(Y_{ij} - \alpha_j, \eta_i) = E[(Y_{ij} - \alpha_j)\eta_i] \\ &= [\beta_{o,\eta,j} + \lambda_j \beta_{D,\eta,d(j)}] E(\eta_i^2) = \beta_{o,\eta,j} + \lambda_j \beta_{D,\eta,d(j)} \end{aligned}$$

for each outcome j , so $[\beta_{o,\eta,j} + \lambda_j \beta_{D,\eta,d(j)}]$ is identified. The elements of the variance-covariance matrix $\text{Var}(\mathbf{Y}_i)$ are

$$\begin{aligned} \text{Cov}(Y_{ij}, Y_{i\ell}) &= E[(Y_{ij} - \alpha_j)(Y_{i\ell} - \alpha_\ell)] \\ &= (\beta_{o,\eta,j} + \lambda_j \beta_{D,\eta,d(j)})(\beta_{o,\eta,\ell} + \lambda_\ell \beta_{D,\eta,d(\ell)}) + \lambda_j \lambda_\ell \tau_\phi^2 + \mathbf{1}_{\{d(j)=d(\ell)\}} \lambda_j \lambda_\ell \tau_\psi^2 + \mathbf{1}_{\{j=\ell\}} \sigma_j^2. \end{aligned}$$

Using this expression, since $[\beta_{o,\eta,j} + \lambda_j \beta_{D,\eta,d(j)}]$ is identified $\forall j$, since $\lambda_j = 1$ for the first outcome in each domain, and since there is more than one domain, τ_ϕ^2 is identified. Since τ_ϕ^2 is identified, λ_j is identified for each j (by taking ℓ to be the first outcome in another domain). For any domain $k \in \{1, \dots, d\}$ that has more than one outcome, $\tau_{\psi,k}^2$ is then identified. For a domain with a single outcome, $\tau_{\psi,k}^2 = 0$ so we don't have to be concerned about its identifiability. Since everything else is identified in the expression for $\text{Cov}(Y_{ij}, Y_{i\ell})$, σ_j^2 is also identified for each j .

Next we address the question of whether $\beta_{o,\eta,j}$ and $\beta_{D,\eta,d(j)}$ are separately identified for each j . Since we have assigned $\beta_{o,\eta,j}$ a random effect prior distribution centered at zero, and since $[\beta_{o,\eta,j} + \lambda_j \beta_{D,\eta,d(j)}]$ and λ_j are identified for each j , $\beta_{o,\eta,j}$ and $\beta_{D,\eta,d(j)}$ are separately identified. So Model A is fully identified.

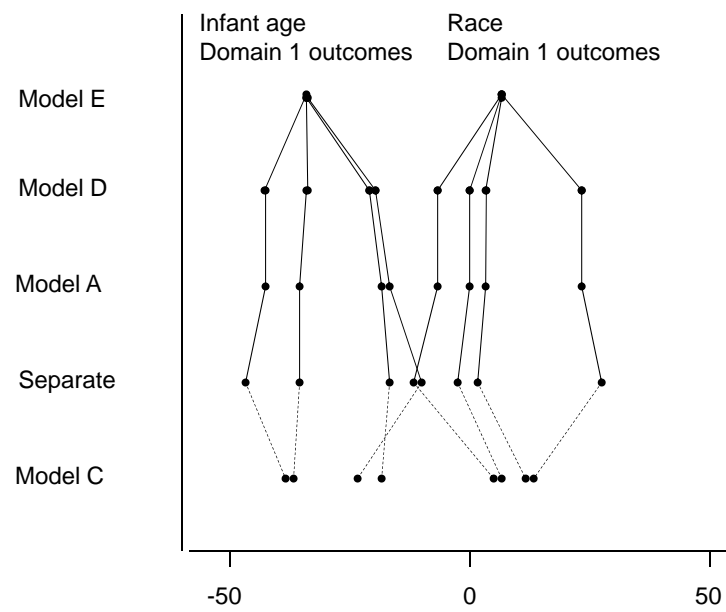


Figure 1. Outcome-specific covariate effects ($\times 100$) for Models A, C, D, and E and separate regressions on the SFF data, for two representative covariates and the Domain 1 outcomes.

Num of Outcomes	4,2,1				4,6,3				0.15 Large
	0.05 Small*				0.15				
$\bar{\beta}_{D,\eta}$	All $\lambda_j = 1$		Some $\lambda_j = 0.5$		Small		Large		Some $\lambda_j = 0.5$
$\beta_{D,\cdot,k} - \bar{\beta}_{D,\cdot}$	100	500	100	500	0.05 Small	0.05 Large	0.15 Small	0.15 Large	
λ_j					All $\lambda_j = 1$				Some $\lambda_j = 0.5$
n					100				
bias $\beta_{D,\eta,k}$	-.019	.006	.022	.008	.025	.002	.016	.009	.015
bias $OS_{\eta,j}$	-.015	-.002	.020	.000	-.002	-.007	-.027	-.022	-.015
bias λ_j	-.748	-.603	-.455	-.388	-.641	-.640	-.584	-.539	-.381
relbias $\beta_{D,\eta,k}$ (%)	-39.1	12.7	47.2	17.2	50.9	4.6	11.4	6.6	11.3
relbias $\beta_{D,z,k}$	4.6	4.5	3.2	0.0	17.9	7.1	13.5	18.7	2.8
relbias $OS_{\eta,j}$	-40.5	-4.6	73.2	0.0	-4.1	-18.6	-19.6	-17.2	-14.9
relbias $OS_{z,j}$	-11.2	-2.1	-15.3	-6.6	-12.7	-14.4	-10.7	-6.0	-16.7
relbias λ_j	-72.6	-60.9	-67.4	-60.8	-64.2	-57.8	-56.3	-50.3	-60.8
CI cov $\beta_{D,\eta,k}$ (%)	97.3	98.7	98.7	97.2	97.3	100.0	100.0	98.7	100.0
CI width $\beta_{D,\eta,k}$.592	.322	.621	.345	.469	.476	.547	.517	.542
CI cov $OS_{\eta,j}$ (%)	98.3	94.9	98.3	92.3	94.8	99.1	93.8	96.9	96.6
CI width $OS_{\eta,j}$.344	.161	.348	.161	.301	.310	.338	.330	.319
RMSE $\beta_{D,\eta,k}$.102	.056	.086	.054	.100	.082	.085	.088	.085
RMSE $\beta_{D,z,k}$.108	.055	.091	.056	.110	.090	.084	.088	.103
RMSE $OS_{\eta,j}$.077	.042	.074	.041	.072	.063	.081	.080	.074
RMSE $OS_{z,j}$.102	.038	.082	.046	.097	.095	.086	.084	.087
RMSE λ_j	.776	.663	.498	.428	.714	.703	.660	.660	.479
RMSEA	.035	.010	.033	.011	.027	.031	.026	.025	.023
p-val RMSEA $\leq .05$.553	.981	.590	.983	.757	.686	.762	.759	.778
p-val RMSEA = 0	.349	.449	.392	.456	.338	.285	.356	.360	.376
residl corr (%)	6.3	3.0	6.2	3.0	7.0	7.2	7.1	7.0	6.9

* Small: $\beta_{D,\eta,k} - \bar{\beta}_{D,\eta} = (-0.03, 0, 0.03)$ and $\beta_{D,z,k} - \bar{\beta}_{D,z} = (-0.05, 0, 0.05)$; Large: $\beta_{D,\eta,k} - \bar{\beta}_{D,\eta} = (-0.1, 0, 0.1)$ and $\beta_{D,z,k} - \bar{\beta}_{D,z} = (-0.2, 0, 0.2)$.

Table 1

Bias, relative bias, RMSE, 95% interval coverage and width, and goodness-of-fit measures, for some scenarios in the simulation study. Here we use data simulated from Model A and report the results for fitting Model A. Also recall the true values of $\tau_o = \tau_{o,1} = 0.05$, $\sigma_j = 1$, $\tau_\phi = 0.2$, and $\tau_{\psi,k} = 0.05$.

λ_j Simulated Estimation	Some $\lambda_j = 0.5$			All $\lambda_j = 1$					
	Model A			Model D				Model B	
	A	D	E	A	B	D	E	A	B
bias $\beta_{D,\eta,k}$.009	-.038	-.036	.013	-.032	.003	.003	.010	-.033
bias $OS_{\eta,j}$	-.002	.003	.004	-.004	-.002	.000	.000	-.004	-.002
bias λ_j	-.254	.329	.329	-.234	.504	.000	.000	-.225	.481
relbias $\beta_{D,\eta,k}$ (%)	6.3	-27.8	-26.5	9.5	-23.2	2.0	2.3	7.5	-25.0
relbias $\beta_{D,z,k}$	1.3	-30.7	-29.4	5.2	-21.5	0.5	0.9	2.1	-21.2
relbias $OS_{\eta,j}$	-2.3	3.0	4.6	-3.0	-1.3	-0.1	0.4	-2.9	-1.2
relbias $OS_{z,j}$	-5.9	-2.4	-0.7	-1.2	-0.6	0.4	0.9	-2.0	-1.7
relbias λ_j	-39.4	51.0	51.0	-23.4	50.4	0.0	0.0	-23.0	49.3
CI cov $\beta_{D,\eta,k}$ (%)	98.6	74.7	48.0	98.7	69.3	96.0	76.0	100.0	69.3
CI width $\beta_{D,\eta,k}$.297	.181	.090	.238	.130	.150	.090	.266	.129
CI cov $OS_{\eta,j}$ (%)	93.8	91.7	48.9	94.2	92.6	90.5	54.8	93.5	91.7
CI width $OS_{\eta,j}$.155	.153	.087	.153	.147	.141	.088	.156	.148
RMSE $\beta_{D,\eta,k}$.055	.056	.054	.045	.047	.034	.035	.047	.050
RMSE $\beta_{D,z,k}$.053	.074	.071	.048	.053	.033	.033	.047	.062
RMSE $OS_{\eta,j}$.040	.042	.061	.038	.037	.038	.050	.038	.037
RMSE $OS_{z,j}$.039	.041	.065	.039	.038	.036	.047	.040	.039
RMSE λ_j	.392	.398	.398	.415	.582	.000	.000	.407	.568
RMSEA	.005	.004	.036	.006	.008	.005	.024	.003	.006
p-val RMSEA $\leq .05$	1.00	1.00	.928	1.00	1.00	1.00	.996	1.00	1.00
p-val RMSEA = 0	.523	.571	.017	.557	.510	.617	.103	.680	.555
residl corr (%)	3.3	3.5	3.6	3.3	3.3	3.4	3.4	3.2	3.3

Table 2

Bias, relative bias, RMSE, 95% interval coverage and width, and goodness-of-fit measures, for some scenarios in the simulation study. Here we use (4,6,3) outcomes in the three domains, a large sample size ($n = 500$), a large exposure effect ($\beta_{D,\eta} = 0.15$), and large deviations between domains ($\beta_{D,\eta,k} - \beta_{D,\eta} = (-0.1, 0, 0.1)$ and $\beta_{D,z,k} - \beta_{D,z} = (-0.2, 0, 0.2)$). Also recall the true values of $\tau_o = \tau_{o,1} = 0.05$, $\sigma_j = 1$, $\tau_\phi = 0.2$, and $\tau_{\psi,k} = 0.05$.

Variable	Counts or Mean \pm SD	Regression Coefficient for Phthalate Score ($\times 100$)
Mother's Race (Cauc. / Non-)	89 / 29	–
Mother's Age	30.1 \pm 5.08	–
Infant Age (mos.)	10.3 \pm 7.30	–
Gestational Age (wks.)	39.0 \pm 2.18	–
Creatinine	88.5 \pm 62.1	–
Skinfold Thickness Flank	5.56 \pm 1.86	–13.5 (–40.3, 13.3)
Skinfold Thickness Quadriceps	14.82 \pm 5.42	5.38 (–22.3, 33.1)
Skinfold Thickness Subscapular	6.96 \pm 1.98	–3.29 (–28.2, 21.6)
Skinfold Thickness Triceps	9.74 \pm 2.42	13.1 (–13.5, 39.8)
Body Mass Index	16.8 \pm 1.49	7.76 (–20.7, 36.2)
Weight Percentile	49.1 \pm 31.5	11.1 (–15.6, 37.7)
Head Circumference Percentile	56.3 \pm 30.1	–1.26 (–29.4, 26.9)

Table 3

Column 2: Summaries of the covariates and outcomes for the SFF data; Column 3: regression coefficient for phthalate score ($\times 100$) from separate regressions, with 95% confidence interval.

Outcome	Infant Age		Gestational Age	
S.T. Flank	-32.6	(-49.9, -15.0)	-8.7	(-27.0, 9.1)
S.T. Quadriceps	-16.7	(-35.9, 2.5)	8.6	(-9.5, 26.7)
S.T. Subscapular	-42.2	(-58.8, -25.4)	8.7	(-7.0, 24.8)
S.T. Triceps	-20.0	(-38.8, -0.8)	23.8	(4.5, 43.1)
BMI	-10.0	(-27.8, 7.9)	7.4	(-10.1, 25.0)
Weight Pct.	-27.2	(-44.6, -9.4)	26.6	(9.8, 43.7)
Head Circ. Pct.	16.2	(-3.4, 35.8)	10.4	(-9.2, 29.9)

Table 4

Selected outcome-specific covariate effects $OS_{z,j}$ times 100 in Model D for the SFF data (95% interval estimates shown in parentheses).