A Hierarchical Integrative Group LASSO (HiGLASSO) framework for analyzing environmental mixtures

Jonathan Boss[1,*], Alexander Rix[1], Yin-Hsiu Chen[2], Naveen N. Narisetty[3], Zhenke Wu[1],

Kelly K. Ferguson[4], Thomas F. McElrath[5], John D. Meeker[6], and Bhramar Mukherjee[1,**]

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

[2]Google Inc., Mountain View, California, U.S.A.

[3]Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, Illinois, U.S.A.

[4]Epidemiology Branch, National Institute of Environmental Health Sciences, Durham, North Carolina, U.S.A.

[5]Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Boston, Massachusetts, U.S.A.

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

*email: bossjona@umich.edu

**email: bhramar@umich.edu

Summary: Environmental health studies are increasingly measuring multiple pollutants to characterize the joint health effects attributable to exposure mixtures. However, the underlying dose-response relationship between toxicants and health outcomes of interest may be highly nonlinear, with possible nonlinear interaction effects. Existing penalized regression methods that account for exposure interactions either cannot accommodate nonlinear interactions while maintaining strong heredity or are computationally unstable in applications with limited sample size. In this paper, we propose a general shrinkage and selection framework to identify noteworthy nonlinear main and interaction effects among a set of exposures. We design hierarchical integrative group least absolute shrinkage and selection operator (HiGLASSO) to (a) impose strong heredity constraints on two-way interaction effects (hierarchical), (b) incorporate adaptive weights without necessitating initial coefficient estimates (integrative), and (c) induce sparsity for variable selection while respecting group structure (group LASSO). We prove sparsistency of the proposed method and apply HiGLASSO to an environmental toxicants dataset from the LIFECODES birth cohort, where the investigators are interested in understanding the joint effects of 21 urinary toxicant biomarkers on urinary 8-isoprostane, a measure of oxidative stress. An implementation of HiGLASSO is available in the higlasso R package, accessible through the Comprehensive R Archive Network.

Key words: Environmental exposures; Group LASSO; Interaction; Nonlinearity; Strong heredity.

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

1. Introduction

1.1 Background and Motivation

Studying the effects of chemical exposures and their interactions in relation to adverse health outcomes is an important topic in epidemiological and environmental research. Furthermore, exposure to endocrine disruptors, such as phthalates and phenols, is of particular interest due to the ubiquity of exposure in the U.S. general population (Crinnion, 2010). Phthalates are a group of chemicals that are widely used as plasticizers or solvents in products such as food packaging, cosmetics, and other industrial materials, which typically enter the human body through daily ingestion and inhalation (Schettler, 2006). Phthalates are known for anti-androgenic effects and reproductive toxicity and recent studies have reported that the modes of their action include mechanisms such as oxidative stress (Ferguson et al., 2011, 2012). Phenols are a class of chemical compounds used in the manufacture of polycarbonate plastics and epoxy resins. Applications of some phenols include use in pesticides and personal care products such as makeups and toothpastes (Darbre and Harvey, 2008). Phenols may possess estrogenic activity and are linked to higher levels of maternal oxidative stress, inflammation in pregnancy, and reduced fetal growth (Watkins et al., 2015; Ferguson et al., 2018).

Classical environmental epidemiology has focused on analyzing one toxicant at a time even though, in truth, subjects are simultaneously exposed to a mixture of compounds which may work in concert. Namely, potential synergistic and antagonistic effects of chemical mixtures have been minimally addressed in human studies. The primary reasons behind single toxicant analysis are the lack of studies with measures on multiple pollutants and a lack of a principled analytic strategy for understanding effects of multiple pollutants and their interactions with limited sample size. Modern assaying technology has made it possible to measure multiple pollutants on the same subjects and advances in statistical learning have enabled us to develop methods that capture nonlinearity and interactions in complex exposure-response

surfaces. Commonly used approaches that characterize the joint effects of mixtures on health outcomes in a flexible way include classification and regression tree (CART) (Loh, 2011) and Bayesian kernel machine regression (BKMR) (Bobb et al., 2015). However, the number of candidate effects, including main effects and interaction effects, may be much larger than the number of observations (i.e., $p >> n$). To address this issue, one common approach is to introduce sparsity during estimation to shrink coefficient estimates towards a subset of variables that have stronger effects (Roberts and Martin, 2005; Liu et al., 2018; Li and Ding, 2019; Ashrap et al., 2020), although targeted selection of interaction effects have received less attention in environmental applications. This paper proposes a variable selection framework to handle potential nonlinearity and interactions between a set of multiple exposures. To be clear, when we qualitatively describe a penalized regression method as nonlinear, we are referring to the fact that the method is designed to estimate nonlinear exposure-response surfaces, not that the method is nonlinear in the regression coefficients. We then apply this framework to data from the LIFECODES study, an ongoing prospective pregnancy/birth cohort at Brigham and Women's Hospital (BWH), to identify important exposures and two-way exposure interactions that are associated with 8-isoprostane, an oxidative stress biomarker (Ferguson et al., 2015).

1.2 *Overview of Interaction Selection Methods*

There are two major classes of methods for variable selection: penalty-based methods and forward/stepwise selection methods. The former adds a penalty term to an objective function which, upon optimization, induces sparsity. Some examples include the L_1 penalty in LASSO (Tibshirani, 1996), the L_0 penalty in nonnegative garrote (Breiman, 1995), the L_γ penalty with $\gamma \geqslant 1$ in bridge regression (Fu, 1998), the mixture of L_1 and L_2 penalties in elastic-net (Zou and Hastie, 2005), and the smoothly clipped absolute deviation (SCAD) penalty (Fan and Li, 2001). These methods can be used to incorporate interactions by treating interaction

terms as additional predictors. However, including interaction terms in the absence of at least one corresponding main effect deviates from a naturally interpretable hierarchical interaction structure. Nelder (1977) and McCullagh and Nelder (1989) introduced the concepts of weak/strong heredity and marginality respectively as conceptual constraints to simplify model interpretation (McCullagh, 1984) and improve statistical power (Cox, 1984). Recent penalty-based methods that respect these heredity principles include the strong heredity interaction model (SHIM) (Choi et al., 2010), the LASSO for hierarchical interactions (hierNet) (Bien et al., 2013), and the group-LASSO interaction network (GLinternet) (Lim and Hastie, 2015). In addition to penalization based methods, forward selection methods (Boos et al., 2009; Wasserman and Roeder, 2009; Luo and Ghosal, 2015) are also commonly used for variable selection in practice. Several forward selection methods which incorporate heredity constraints with linear and nonlinear interactions have been proposed (Wu et al., 2010; Crews et al., 2011; Hao and Zhang, 2014; Narisetty et al., 2018).

1.3 *Penalized Regression with Nonlinear Dose-Response Surfaces*

Nonlinear exposure-response relationships have also been explored in environmental studies. Failure to account for nonlinearity could result in important variables being left out. Moreover, not properly adjusting for nonlinear main effects might result in spurious detection of interaction effects (Cornelis et al., 2012; Mukherjee et al., 2012; He et al., 2017; Zhang et al., 2020). For example, the quadratic main effect terms of two predictors and interaction terms between the two predictors are not easily differentiable in practice, especially when the signal-to-noise level is low (He et al., 2017). Group LASSO (Yuan and Lin, 2006) can be adopted to model nonlinear effects where each group of variables represents the nonlinear expansion of a single predictor with respect to a chosen basis (Huang et al., 2010). Another work that considers modeling nonlinear main effects using penalization is the COmponent Selection and Smoothing Operator (COSSO) (Lin and Zhang, 2006). To our knowledge, Variable selection

using Adaptive Nonlinear Interaction Structures in High dimensions (VANISH) (Radchenko and James, 2010) is the only existing method that accounts for both nonlinear main and interaction effects with strong heredity enforced.

1.4 *Weighted Penalization and Selection Consistency*

Using the same tuning parameter λ (degree of penalization) for each predictor/group without assessing their relative importance may simultaneously reduce estimation efficiency and affect selection consistency (Leng et al., 2006). Adaptive shrinkage has been extensively discussed in previous literature (Wang et al., 2007; Zhang and Lu, 2007). For example, adaptive LASSO (Zou, 2006), adaptive elastic-net (Zou and Zhang, 2009), and adaptive Group LASSO (Wang and Leng, 2008) assign a separate penalty to each predictor/group, usually determined by the reciprocal of the absolute values of the corresponding coefficients. This ensures that smaller coefficients are shrunk to zero faster whereas larger coefficients are less penalized. Ordinary least squares (OLS) can be used to estimate the coefficients, however, when $p > n$, OLS cannot be implemented. Additionally, in high-dimensional scenarios, it might be difficult to supply a $\sqrt{n}$-consistent estimate of main and interaction effects to use as adaptive weights, implying that oracle properties are not maintained. In this paper, we bypass the need to specify a set of initial coefficient estimates by using integrative weighted group LASSO (Pan and Zhao, 2016) which jointly estimates weights and coefficients.

1.5 *Structure of Article*

We propose *hierarchical integrative group LASSO (HiGLASSO)*, to deal with both nonlinear main and interaction effects under strong heredity while incorporating integrative weights for improving selection properties. The rest of the article is organized as follows. We briefly review the existing penalty-based interaction selection methods with heredity constraints in Section 2. In Section 3, we describe HiGLASSO, the optimization procedure, and prove spar-sistency for the resulting HiGLASSO estimator. We examine the performance of HiGLASSO

by comparing it to other procedures that address nonlinearity, interaction terms, and/or group structure in Section 4. In Section 5, we analyze data from the LIFECODES study to identify important phthalates, phenols, and their possible interactions that associate with the oxidative stress biomarker 8-isoprostane. We conclude with a discussion in Section 6.

2. Review of existing penalty-based interaction selection methods with heredity constraints

First we overview existing penalized regression methods that select interaction terms subject to heredity constraints. Consider the standard regression setting with p predictors and n observations where $\boldsymbol{x}_j$ denotes the $n \times 1$ predictor vector corresponding to the jth regression coefficient β_j, for $j = 1, \cdots, p$. Let γ_{kl} be the coefficient of interaction effect between $\boldsymbol{x}_k$ and $\boldsymbol{x}_l$. Strong and weak heredity principles for interaction effects are defined as follows.

- Strong heredity principle: If an interaction term is included in the model *both* of its corresponding main effects must be present in the model. That is, if $\gamma_{kl} \neq 0$, then $\beta_k \neq 0$ *and* $\beta_l \neq 0$.

- Weak heredity principle: If an interaction term is included in the model, *at least one* of the corresponding main effects must be present in the model. That is, if $\gamma_{kl} \neq 0$, then $\beta_k \neq 0$ *or* $\beta_l \neq 0$.

2.1 *Methods for linear interactions*

A generic second-order model accounting for pairwise interaction effects with linear predictors is given as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{X}_{(I)}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \tag{1}$$

where $\boldsymbol{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_p]$ denotes the $n \times p$ design matrix for main effects, $\boldsymbol{X}_{(I)} = [\boldsymbol{x}_1 \odot \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{p-1} \odot \boldsymbol{x}_p]$ denotes the $n \times [p(p-1)/2]$ design matrix for interactions where " $\odot$ " indicates the element-wise product, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^\top$, $\boldsymbol{\gamma} = (\gamma_{12}, \cdots, \gamma_{p-1,p})^\top$, and $\boldsymbol{\epsilon}$ is a

multivariate Gaussian error vector. Without loss of generality, we assume all variables are standardized and exclude the intercept from our regression model. We first review existing methods for selecting interaction effects which satisfy the strong heredity principle.

SHIM: SHIM reparametrizes the interaction coefficients as scaled products of component main effect terms, namely $\gamma_{ij} = \eta_{ij}\beta_i\beta_j$ for $1 \leqslant i < j \leqslant p$, $\eta_{ij} \in \mathbb{R}$. A penalty is imposed on $\boldsymbol{\eta} = \{\eta_{ij}\}$ rather than the interaction coefficients $\boldsymbol{\gamma}$ to preserve heredity of the interaction terms in the selected model. SHIM minimizes the objective function

$$\frac{1}{2}||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}_{(I)}\boldsymbol{\gamma}||_2^2 + \lambda_1||\boldsymbol{\beta}||_1 + \lambda_2||\boldsymbol{\eta}||_1$$

using an algorithm that iterates between LASSO and group LASSO.

hierNet: hierNet is a LASSO-based approach which minimizes

$$\frac{1}{2}\Big\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \sum_{k=1}^{p}\sum_{l=1}^{p}(\boldsymbol{x}_k \cdot \boldsymbol{x}_l)\gamma_{kl}\Big\|_2^2 + \lambda_1\sum_{j=1}^{p}|\beta_j| + \frac{1}{2}\lambda_2\sum_{k=1}^{p}\sum_{l=1}^{p}|\gamma_{kl}|,$$

subject to symmetry constraints $\gamma_{kl} = \gamma_{lk}, \ \forall \ 1 \leqslant k, l \leqslant p$, and heredity constraints $\sum_{l=1}^{p}|\gamma_{kl}| \leqslant |\beta_k|, \forall k = 1, \cdots, p$, which ensure that the interaction effects are zero given that any of the corresponding main effects are zero. Alternating Direction Method of Multipliers (ADMM) (Boyd et al., 2011) is used to solve the constrained optimization.

GLinternet: GLinternet uses an overlapping group LASSO penalty to enforce strong heredity. The objective function is given by

$$\frac{1}{2}\Big\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \sum_{k=2}^{p}\sum_{l=1}^{k-1}[\boldsymbol{x}_k, \boldsymbol{x}_l, \boldsymbol{x}_k \cdot \boldsymbol{x}_l]\boldsymbol{\gamma}_{kl}^*\Big\|_2^2 + \lambda_1||\boldsymbol{\beta}||_1 + \frac{1}{2}\lambda_2\sum_{k=2}^{p}\sum_{l=1}^{k-1}||\boldsymbol{\gamma}_{kl}^*||_2,$$

where each $\boldsymbol{\gamma}_{kl}^*$ is a three dimensional vector with the first two elements corresponding to main effects and the third element corresponding to the interaction effect. Note that the main effects appear multiple times inside the L_2-norm (parameterized by β_k, $\boldsymbol{\gamma}_{kl}^*$ for $l < k$, and $\boldsymbol{\gamma}_{lk}^*$ for $l > k$) and hence are multiply penalized. An iterative soft thresholding algorithm (Beck and Teboulle, 2009) can be used to solve the GLinternet optimization problem.

2.2 *Methods for nonlinear interactions*

Basis functions such as cubic splines are often used to incorporate nonlinear main effects and nonlinear interaction effects into regression models. Consider S groups of predictors each of which corresponds to a pre-specified nonlinear basis expansion. Let $\boldsymbol{X}_j$ and $\boldsymbol{\beta}_j$ denote the $n \times p_j$ design matrix and coefficient vector of length p_j corresponding to group j of basis size p_j, respectively, for $j = 1, \cdots, S$. Let $\boldsymbol{X}_{kl}$ be the $n \times (p_k p_l)$ design matrix for two-way interaction between group k and group l and $\boldsymbol{\gamma}_{kl}$ be the corresponding $(p_k p_l)-$vector of interaction coefficients for $1 \leqslant k < l \leqslant S$. Note that $\boldsymbol{X}_j$ and $\boldsymbol{X}_{kl}$ are distinct from their section 2.1 counterparts, $\boldsymbol{X}$ and $\boldsymbol{X}_{(I)}$, because we are now working with basis expansions of exposures rather than linear exposure terms. We focus on the second-order model with interaction effects for S groups of nonlinear predictors as

$$\boldsymbol{y} = \sum_{j=1}^{S} \boldsymbol{X}_j \boldsymbol{\beta}_j + \sum_{1 \leqslant k < l \leqslant S} \boldsymbol{X}_{kl} \boldsymbol{\gamma}_{kl} + \boldsymbol{\epsilon}. \tag{2}$$

VANISH is the only existing penalty-based method that imposes sparsity and strong heredity on model (2).

VANISH: VANISH optimizes penalized least squares as

$$\frac{1}{2} \Big\| \boldsymbol{y} - \sum_{j=1}^{S} \boldsymbol{X}_j \boldsymbol{\beta}_j - \sum_{1 \leqslant k < l \leqslant S} \boldsymbol{X}_{kl} \boldsymbol{\gamma}_{kl} \Big\|_2^2$$
$$+ \lambda_1 \sum_{j=1}^{S} \Big(||\boldsymbol{\beta}_j||_2^2 + \sum_{k<j} ||\boldsymbol{\gamma}_{kj}||_2^2 + \sum_{l>j} ||\boldsymbol{\gamma}_{jl}||_2^2 \Big)^{1/2} + \lambda_2 \sum_{1 \leqslant k < l \leqslant S} ||\boldsymbol{\gamma}_{kl}||_2.$$

By construction, $\boldsymbol{\beta}_j$'s and $\boldsymbol{\gamma}_{kl}$'s are folded together in the first penalty term so main effect coefficients and interaction coefficients are either all zero or all nonzero, based on the property of the group LASSO penalty. The same structure applies to all S groups of main effects so strong heredity is guaranteed. A block gradient descent algorithm involving a single sweep through all the variables is applied to obtain a solution to the VANISH objective function.

None of the existing variable selection methods described so far account for both strong heredity in interaction selection and differential penalization via adaptive weighting. We

propose HiGLASSO as a novel approach to select two-way interaction effects under strong

heredity constraints using penalization with integrative weights, circumventing the need for

initial coefficient estimates.

3. Hierarchical integrative group LASSO (HiGLASSO)

3.1 *HiGLASSO formulation*

Consider the regression model in (2). To enforce heredity constraints, we rewrite (2) as

$$\boldsymbol{y} = \sum_{j=1}^{S} \boldsymbol{X}_j \boldsymbol{\beta}_j + \sum_{1 \leqslant j < j' \leqslant S} \boldsymbol{X}_{jj'} [\boldsymbol{\eta}_{jj'} \odot (\boldsymbol{\beta}_j \otimes \boldsymbol{\beta}_{j'})] + \boldsymbol{\epsilon} \tag{3}$$

by reparameterizing $\boldsymbol{\gamma}_{jj'} = \boldsymbol{\eta}_{jj'} \odot (\boldsymbol{\beta}_j \otimes \boldsymbol{\beta}_{j'})$ for $1 \leqslant j < j' \leqslant S$. Here "$\otimes$" denotes the

Kronecker product and $\boldsymbol{\eta}_{jj'}$ is a $(p_j p_{j'})-$vector of scalars for interactions between variables

in group j and group j' following SHIM (Choi et al., 2010). Note that strong heredity

constraints are satisfied because $\boldsymbol{\gamma}_{jj'}$ is non-zero only if both main effects are non-zero. To

see this, $\boldsymbol{\beta}_j = \boldsymbol{0}$ and/or $\boldsymbol{\beta}_{j'} = \boldsymbol{0}$ implies that $\boldsymbol{\gamma}_{jj'} = \boldsymbol{0}$. Similarly, $\boldsymbol{\gamma}_{jj'} \neq \boldsymbol{0}$ implies that

$\boldsymbol{\eta}_{jj'} \neq \boldsymbol{0}$, $\boldsymbol{\beta}_j \neq \boldsymbol{0}$, and $\boldsymbol{\beta}_{j'} \neq \boldsymbol{0}$.

Consider the penalized least squares criterion

$$\underset{\boldsymbol{\beta}_j, \boldsymbol{\eta}_{jj'}}{\arg \min} \frac{1}{2} \left\| \boldsymbol{y} - \sum_{j=1}^{S} \boldsymbol{X}_j \boldsymbol{\beta}_j - \sum_{1 \leqslant j < j' \leqslant S} \boldsymbol{X}_{jj'} [\boldsymbol{\eta}_{jj'} \odot (\boldsymbol{\beta}_j \otimes \boldsymbol{\beta}_{j'})] \right\|_2^2 \tag{4}$$

$$+ \lambda_1 \sum_{j=1}^{S} ||\boldsymbol{\beta}_j||_2 + \lambda_2 \sum_{1 \leqslant j < j' \leqslant S} ||\boldsymbol{\eta}_{jj'}||_2,$$

where λ_1 and λ_2 are tuning parameters that control the amount of main effect and interaction

effect shrinkage toward 0, respectively. To remedy potential estimation inefficiency and

selection inconsistency, we work with a modified version of (4) to differentially penalize

parameters in the spirit of adaptive group LASSO (Wang and Leng, 2008). We consider

$$\underset{\boldsymbol{\beta}_j, \boldsymbol{\eta}_{jj'}}{\arg \min} \frac{1}{2} \left\| \boldsymbol{y} - \sum_{j=1}^{S} \boldsymbol{X}_j \boldsymbol{\beta}_j - \sum_{1 \leqslant j < j' \leqslant S} \boldsymbol{X}_{jj'} [\boldsymbol{\eta}_{jj'} \odot (\boldsymbol{\beta}_j \otimes \boldsymbol{\beta}_{j'})] \right\|_2^2 \tag{5}$$

$$+ \lambda_1 \sum_{j=1}^{S} w_j ||\boldsymbol{\beta}_j||_2 + \lambda_2 \sum_{1 \leqslant j < j' \leqslant S} w_{jj'} ||\boldsymbol{\eta}_{jj'}||_2,$$

where w_j's and $w_{jj'}$'s are pre-specified weight functions of unknown coefficients $\{\boldsymbol{\beta}_j\}$ and $\{\boldsymbol{\eta}_{jj'}\}$.

To concurrently estimate weights and model parameters following (Pan and Zhao, 2016), we consider weight functions based on the extreme values of each group, namely,

$$w_j \equiv \exp\left\{ - \frac{||\boldsymbol{\beta}_j||_\infty}{\sigma} \right\} \text{ for } j = 1, \cdots, S, \tag{6}$$

$$w_{jj'} \equiv \exp\left\{ - \frac{||\boldsymbol{\eta}_{jj'}||_\infty}{\sigma} \right\} \text{ for } 1 \leqslant j < j' \leqslant S, \tag{7}$$

where $||\boldsymbol{\mu}||_\infty$ is the L_∞ norm of $\boldsymbol{\mu}$ and σ is a pre-determined scale parameter. That is, the weights decay exponentially with the extremum norm of the coefficients within a group. Figure 1 illustrates the weight function for a two-dimensional coefficient vector. We adopt the L_∞ norm, instead of the L_0, L_1, and L_2 norms, because the groups in our motivating example are basis expansions of each exposure. We do not want to impose sparsity within each group; therefore, to assess the effect size of the entire basis expansion, taking the extremum of the coefficients within a group is more meaningful than taking an "average" coefficient.

[Figure 1 about here.]

In summary, HiGLASSO has the following four features:

(1) Imposes strong heredity on two-way interaction (Hierarchical);

(2) Incorporates adaptive weights without requiring initial coefficient estimates (Integrative);

(3) Induces sparsity for variable selection (LASSO);

(4) Maintains group structure (Group LASSO). The HiGLASSO framework is general and the group structure can be defined based on the specific application. For example, the group structure could be:

- A set of basis functions representing nonlinear relationships,

- Dummy variables representing different levels of categorical variables,

- A natural grouping based on domain knowledge.

3.2 *Optimizing the HiGLASSO objective function*

The objective function in (5) is non-convex and is difficult to globally minimize, however (Pan and Zhao, 2016) proposed a generalized local quadratic approximation which we utilize to find a local minimum. The first term in (5) involves the product of $\boldsymbol{\beta}_j$'s and $\boldsymbol{\eta}_{jj'}$'s. We use an iterative approach to cycle through $\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_S$, and the $\boldsymbol{\eta}_{jj'}$'s until convergence using gradient descent. We first optimize over $\boldsymbol{\beta}_j$ given the current $\hat{\boldsymbol{\beta}}_{j'}$'s with $j' \neq j$ and $\hat{\boldsymbol{\eta}}_{jj'}$'s. Then we iteratively obtain $\hat{\boldsymbol{\eta}}_{jj'}$ estimates given current $\hat{\boldsymbol{\beta}}_j$'s. The optimization routine is summarized in Web Appendix A. The *higlasso* R package, available on the Comprehensive R Archive Network (CRAN), implements the proposed optimization routine.

3.3 *Sparsistency of HiGLASSO estimator*

We now establish sparsistency of the HiGLASSO estimator obtained as the minimizer of (5). Let $\boldsymbol{\theta}$ denote the vector of all coefficients, including main effect coefficients and interaction coefficients. Namely, $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \cdots, \boldsymbol{\beta}_S^\top)^\top$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_{12}^\top, \cdots, \boldsymbol{\gamma}_{S-1,S}^\top)^\top$, and $\boldsymbol{\gamma}_{jj'} = \boldsymbol{\eta}_{jj'} \odot (\boldsymbol{\beta}_j \otimes \boldsymbol{\beta}_{j'})$. Denote $\boldsymbol{\theta}_{\mathcal{P}} = (\boldsymbol{\beta}_{\mathcal{P}_1}{}^\top, \boldsymbol{\gamma}_{\mathcal{P}_2}{}^\top)^\top$ and $\boldsymbol{\theta}_{\mathcal{P}^c} = (\boldsymbol{\beta}_{\mathcal{P}_1^c}{}^\top, \boldsymbol{\gamma}_{\mathcal{P}_2^c}{}^\top)^\top$ where $\mathcal{P}_1$ is the true nonzero set for $\boldsymbol{\beta}$, $\mathcal{P}_1^c$ is the true zero set for $\boldsymbol{\beta}$, $\mathcal{P}_2$ is the true nonzero set for $\boldsymbol{\gamma}$, $\mathcal{P}_2^c$ is the true zero set for $\boldsymbol{\gamma}$, $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$, and $\mathcal{P}^c = \mathcal{P}_1^c \cup \mathcal{P}_2^c$. Let $a_n = \min(\lambda_1(n), \lambda_2(n))$ and $b_n = \sigma(n)$. That is, $\lambda_1(n)$, $\lambda_2(n)$, and $\sigma(n)$ depend on sample size.

Theorem (Sparsistency of HiGLASSO estimator): Suppose that the data are generated from the model given by (3) with the errors ϵ following an i.i.d. normal distribution with mean zero and variance $\tau^2 > 0$. Assume that the design matrix $\boldsymbol{X}$ is random such that $\frac{1}{n}\boldsymbol{X}^\intercal\boldsymbol{X} = \frac{1}{n}E(\boldsymbol{X}^\intercal\boldsymbol{X}) + O_p(n^{-1/2})$, $\frac{1}{n}E(\boldsymbol{X}^\intercal\boldsymbol{X})$ is invertible, all the eigenvalues of $\frac{1}{n}\boldsymbol{X}^\intercal\boldsymbol{X}$ are bounded away from 0 and ∞ with probability converging to one, and that there exists

some constant U that uniformly bounds the L_2-norm of the HiGLASSO estimator for all n. If $a_n/\sqrt{n} \to \infty$, $a_n/n \to 0$, and $b_n \to 0$ as $n \to \infty$, then we have $P\big(\big\|\hat{\boldsymbol{\beta}}_{\mathcal{P}_1^c}\big\|_2 = 0\big) \to 1$ and $P\big(\big\|\hat{\boldsymbol{\gamma}}_{\mathcal{P}_2^c}\big\|_2 = 0\big) \to 1$.

Proof: See Web Appendix B.

The theorem ensures that spurious covariates will be eliminated by the HiGLASSO procedure when the number of covariates is fixed as $n \to \infty$. However, the theorem assumes conditions on the design matrix $\boldsymbol{X}$ which do not allow the number of covariates to diverge. Generalizations of sparsistency of the HiGLASSO estimator in high dimensional settings, i.e. when $|\mathcal{P} \cup \mathcal{P}^c| = o(n)$, are not discussed here.

4. Simulation study

The goal of the simulation study is to compare the performance of HiGLASSO with alternative approaches for selecting main and pairwise interaction effects. The competing methods accounting for linear main effects and linear pairwise interaction terms include LASSO and hierNet. An alternative method accounting for nonlinear main effects and, potentially, nonlinear interaction terms is group LASSO. "Nonlinear" in this context refers to nonlinear basis expansions of the original exposure variables. For the present simulation study we use a cubic basis expansion, where each scalar exposure variable x_j is expanded to $(x_j, x_j^2, x_j^3)^\top$. Nonlinear interactions are therefore comprised of all pairwise multiples of individual terms in the corresponding basis expansions. For all methods with group structure, i.e. group LASSO and HiGLASSO, the full nonlinear basis expansions for each covariate define the groups ($p_s = 3$, $\forall s = 1, \cdots, S$). Similarly, for pairwise nonlinear interactions, all pairwise multiples of individual terms in the two basis expansions are considered a group. The R package glmnet was used to implement LASSO, the R package hierNet was used to implement hierNet, the R package gglasso was used to implement group LASSO, and the R package higlasso was used to

implement HiGLASSO. VANISH was not considered in this simulation study because there is no publicly available implementation on CRAN.

4.1 *Simulation setting*

For the present simulation study, we consider 9 different scenarios, each with 500 simulated datasets and a sample size of either $n = 1000$ or $n = 10000$. The data generation mechanism for the simulated datasets is to first generate covariate vectors from a $N(0, \Sigma)$ distribution where Σ is an compound symmetric matrix with unit variance and pairwise correlations equal to 0.3 and then draw $\boldsymbol{y}|\boldsymbol{x}_1, ..., \boldsymbol{x}_p$ from the regression model

$$\boldsymbol{y} = f(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_p) + \boldsymbol{\epsilon}, \ \ \boldsymbol{\epsilon} \sim N(0, 9\boldsymbol{I})$$

A list of the mean functions ($f(\cdot)$) and the number of predictors ($p = 10$, $p = 20$), across the six $n = 1000$ simulation scenarios are provided in Table 1. The $n = 10000$ simulation settings have the same mean functions as the $n = 1000$ simulation settings, but were only considered with $p = 10$ in order to assess the large sample behavior of each method. In the 'Scenario' column in Table 1, L refers to scenarios with true linear main and interaction effects, PL refers to scenarios with true piecewise linear main and interaction effects, and NL refers to scenarios with true nonlinear main and interaction effects.

[Table 1 about here.]

If we consider the cubic spline expansion with all possible two-way interactions, $p = 10$ corresponds to $435 = 10 * 3 + C_2^{10} * 3 * 3$ effective predictors in our design matrix and $p = 20$ corresponds to $1770 = 20 * 3 + C_2^{20} * 3 * 3$ effective predictors in our design matrix. All tuning parameters for each regularized regression method are selected via 10-fold cross-validation, with the exception of fixing $\sigma = 1$ for HiGLASSO. For LASSO, group LASSO, and hierNet, the largest tuning parameter value within one standard error of the minimum cross-validation error is selected. Since HiGLASSO is naturally conservative

with respect to interaction selection, the tuning parameter pair that results in the lowest cross-validation error is selected. With these tuning parameter values, the corresponding regularized regression methods are then re-fit on the full data.

4.2 *Performance metrics*

The simulation metrics that we will focus on are the following:

(1) **False negative interaction effects rate (FNI)**: The average number of times that a non-null interaction effect term is not selected by a model.

(2) **False negative main effects rate (FNM)**: The average number of times that a non-null main effect term is not selected by a model.

(3) **False positive interaction effects rate (FPI)**: The average number of times that a null interaction effect term is selected by a model.

(4) **False positive main effects rate (FPM)**: The average number of times that a null main effect term is selected by a model.

These four metrics are scaled to a range between 0 and 100, reflecting the average percent error rate per simulated data set and per important/unimportant term. Note that smaller values of all four metrics indicate better variable selection performance.

4.3 *Simulation results*

Simulation results for the $n = 1000$ and $p = 10$ simulation scenarios are presented in Figure 2. Panel (a) corresponds to case L10 with linear main and interaction effects, panel (b) corresponds to case PL10 with piecewise linear main and interaction effects, and panel (c) corresponds to case NL10 with nonlinear main and interaction effects (see Web Figure 1 for the $n = 10000$ simulation results). In L10, LASSO is correctly specified, and therefore leads to relatively low FNI, FNM, FPI, and FPM. LASSO's FNM, FPI, and FPM in PL10 are comparable to the respective metrics in L10, however the FNI is notably larger (FNI

= 37%). For NL10, some of the main effects contain absolute values and quadratic terms, which are more difficult for LASSO with only linear main and interaction terms to detect, hence the elevated FNI (FNI = 29%) and FNM (FNM = 26%). hierNet tends to do well with respect to FNI, FNM, and FPI, but on average has the highest FPM for L10 (FPM = 64%), PL10 (FPM = 27%), and NL10 (FPM = 35%). Conversely, HiGLASSO has the highest FNI rate for L10 (FNI = 16%), PL10 (FNI 55%), and NL10 (FNI = 45%), but has relatively low FNM, FPI, and FPM. That is, HiGLASSO is conservative for interaction selection, but when HiGLASSO selects interactions, they are almost always true interactions. Group LASSO's behavior is difficult to characterize across the three simulation scenarios. One general theme is that the FPM for group LASSO is above 20% for L10, PL10, and NL10. Group LASSO also has an FNI of 44% for PL10. The FNI, FNM, FPI, and FPM patterns for the $n = 10000$ and $p = 10$ simulation scenarios are similar to the $n = 1000$ and $p = 10$ simulation scenarios, however there is a general decrease in false negative and false positive rates across all methods.

[Figure 2 about here.]

Figure 3 summarizes the simulation results for $n = 1000$ and $p = 20$ simulation scenarios. Panel (a) corresponds to case L20 with linear main and interaction effects, panel (b) corresponds to case PL20 with piecewise linear main and interaction effects, and panel (c) corresponds to case NL20 with nonlinear main and interaction effects. Simulation results for L20 and PL20 are nearly identical to the simulation results for L10 and PL10, however the simulation results for NL20 are different from the simulation results for NL10. The notable difference in NL20 is that HiGLASSO now has the lowest FNI (FNI = 31%), FNM (FNM = 1%), and FPI (FPI = 0.2%), but also has very low FPM (FPM = 4%). For NL20, hierNet maintains an elevated FPM (FPM = 27%), LASSO has increased false negative rates (FNI

= 36%, FNM = 28%), and group LASSO has a large FNI (FNI = 55%) as opposed to the higher false positive rates from NL10.

[Figure 3 about here.]

When there are nonlinear main and interaction effects in the true exposure-response model such that the nonlinear interactions obey the strong heredity principle, HiGLASSO has excellent performance with respect to FNM, FPI, and FPM. HiGLASSO can be conservative for interaction selection, as evidenced by elevated FNI in Figure 2 and Figure 3, for which there are several explanations. When the true outcome-exposure association involves sufficiently linear main and interaction effects, HiGLASSO overparameterizes the exposure-response model and therefore unnecessarily introduces additional parameters that require estimation. Estimating the additional parameters results in a loss of power to detect all of the true interactions (although the false discovery rate for main and interaction effects is very low). Another explanation is that using a cubic basis expansion to handle nonlinear main and interaction effects involves a certain level of approximation error. Nevertheless, HiGLASSO shows great promise in the NL20 setting, which is the scenario that it is specifically designed for.

5. Application to the LIFECODES study

5.1 *Data overview*

LIFECODES is a biobank that longitudinally collects biospecimens and medical data across pregnancy with the two-part goal of (a) understanding biophysiological processes underlying fetal development and (b) identifying environmental risk factors for adverse birth outcomes. A subset of pregnant women in the LIFECODES cohort ($n = 482$) had 21 phthalate, phenol, and paraben concentrations (see Table 2) measured longitudinally at approximately, 10 weeks, 18 weeks, 26 weeks, and 35 weeks gestation. Due to known temporal variability in the

analytes of interest, specific gravity adjusted geometric averages across the first three visits for each contaminant and each subject were used as covariates to minimize measurement error (Meeker et al., 2012). The fourth visit measurement was omitted because many women with preterm deliveries had already delivered by 35 weeks. Of those 482 women, our working dataset contains $n = 477$ women (128 preterm deliveries and 349 full-term deliveries) after removal of subjects with no phenol measurements. Study details including exclusion criteria, handling and storage of biological samples, assessment of contaminant concentrations, and institutional review board approval can all be found in (Ferguson et al., 2015).

In this section, we apply LASSO, group LASSO, hierNet, and HiGLASSO to the data collected as a part of LIFECODES where the covariates are the 21 phthalate, paraben, and phenol geometric averages (log-transformed and standardized) and the outcome is specific gravity corrected 8-isoprostane, a biomarker that is indicative of oxidative stress, averaged over the first three visits (log-transformed and centered) (Montuschi et al., 1999). For the nonlinear methods we expand each of the 21 exposure variables into a group of two variables using a quadratic basis expansion. Tuning parameters for all methods are determined by 10-fold cross-validation.

[Table 2 about here.]

5.2 *Initial analyses*

To perform an interaction search, many analysts will proceed by adding linear pairwise interaction terms one-at-a-time and then subsequently assess the statistical significance of each interaction. Therefore, as a cursory analytical step, we will regress log-transformed 8-isoprostane on every possible linear pairwise interaction term one at a time, keeping the 21 linear main effects for each exposure in the model throughout. Web Figure 2 provides a visualization of the resulting p-values for each pairwise interaction (diagonal entries of the heatmap are p-values for the addition of a squared term in the linear regression model). We

observe that there are several interactions that fall below the $p < 0.05$ threshold, including MBzP×MCPP ($p = 0.026$), BPS×2,4-DCP ($p = 0.025$), and BPS×2,5-DCP ($p = 0.016$). Moreover, the Wald tests for inclusion of a 2,5-DCP squared term ($p = 0.015$) and MePB squared term ($p = 0.033$) are significant at the $\alpha = 0.05$ level. Lastly, looking at the unadjusted, marginal exposure-response associations we can clearly identify several nonlinear relationships (see Web Figure 3). These exploratory steps affirm that a model accounting for nonlinearity and interaction structure in the exposure-response surface may be desired.

5.3 *Variable selection results*

The selected main effects and interaction effects for each method are enumerated in Table 3. The two methods that only account for linear pairwise interaction effects, LASSO and hierNet, have very similar results. Namely, all 9 main effects and 2 out of 3 interaction effects selected by LASSO are also selected by hierNet. The one interaction that is selected by LASSO but not hierNet is MEP×TCS, which violates strong heredity. There are more main effects selected by group LASSO than any other method. Moreover, the set of main effects selected by all other methods is a proper subset of the main effects selected by group LASSO. However, 4 of the 6 interactions selected by group LASSO violate strong heredity, the only exceptions being MBzP×MCPP and BPS×2,5-DCP. HiGLASSO selects fewer main effects and interaction effects than group LASSO, but the interactions both satisfy strong heredity. In fact, the interactions selected by HiGLASSO are MBzP×MCPP and BPS×2,5-DCP, which is consistent with group LASSO. One other interesting observation is that group LASSO and HiGLASSO both select MePB, while LASSO and hierNet do not. Referring to Web Figure 3, we can visually identify a marginal quadratic relationship between MePB (log-transformed) and 8-isoprostane (log-transformed), which when modeled by a linear term would be relatively flat. The quadratic term in the basis expansion facilitates detection of an association between MePB and 8-isoprostane that would have been missed otherwise.

[Table 3 about here.]

6. Discussion

This paper presents a new penalized variable selection algorithm to handle groups or sets of correlated predictors and their possibly nonlinear interactions. HiGLASSO imposes strong heredity, induces sparsity as in group LASSO, and maintains efficiency and sparsistency through the use of integrative weights. The integrative weights in HiGLASSO also help select a more parsimonious model compared to other penalized regression strategies, as seen in the LIFECODES data example. By defining groups through basis expansions, the method can handle nonlinear main effects and nonlinear pairwise interactions. Our simulation results indicate that HiGLASSO controls false discovery rates while having competitive true discovery rates for both main effects and interactions, particularly when there is true nonlinearity in the exposure-response surface. Further extension of HiGLASSO to an elastic-net framework is needed in order to handle highly collinear groups of environmental exposures. Additionally, following the application of an initial interaction screening algorithm, principled post model selection inference and robust replication strategies are other ares of statistical research that require rapid development.

Although quadratic and cubic B-splines were used throughout the simulation and data example, there are many other types of basis functions that could be considered. One such type of basis functions are penalized smoothing splines, which are frequently used in generalized additive models (Eilers and Marx, 1996). In the $n = 1000$ and $p = 10$ simulation settings we added a HiGLASSO implementation with penalized smoothing splines and compared the performance against HiGLASSO with cubic B-splines (Web Figure 4). We observe that the patterns and relative rankings across methods are largely similar, however penalized smoothing splines tend to result in lower FNI and lower FPI compared to cubic B-splines. FNM and FPI were small regardless of which basis functions were used. This

brief comparison provides some initial evidence that penalized smoothing splines might be a slightly more appropriate default choice, although a more comprehensive investigation of different classes of basis functions is needed.

To choose the degree of nonlinearity, several tools can be adopted as a part of an exploratory analysis. In particular, fitting a generalized additive model and examining the predictive residual plots - analogous to partial residual plots in linear regression - the estimated degrees of freedom - a higher departure from unity implying stronger evidence in favor of nonlinearity - and model fit statistics, such as the Akaike information criterion and the genearlized cross-validation score, can help gauge the need for nonlinearity. Formal tests comparing a series of nested models with different smoothing terms using ANOVA type techniques can also aid with choosing the degree of nonlinearity. An integrated one-step method with penalization for variable selection as well as smoothness parameter selection is incredibly challenging and beyond the scope of this paper. Thus our approach is to start with a rich set of basis functions based on our exploratory analysis and primarily focus on variable selection given this chosen set of basis functions.

In our simulations and data example, we set the integrative weight parameter to $\sigma = 1$. This is an ad hoc choice and there are likely better approaches for determining σ. One option is to treat σ as a tuning parameter, however tuning over a 3-dimensional grid of candidate tuning parameter values is computationally cumbersome. In our view, a more promising direction is to fix $\sigma = o(n)$ in accordance with the sparsistency result.

A reviewer pointed out that the HiGLASSO may naturally extend to spatio-temporal modeling, where the distinction between separable and non-separable spatio-temporal auto-correlation can be mathematically framed as interactions between spatial and temporal main effects. Parallels could be drawn between LASSO-style approaches that are already being utilized in this area (Hefley et al., 2017) and a re-branding of HiGLASSO may have some

additional applications in this context. That being said, a more computationally efficient algorithm for optimizing the HiGLASSO objective function would be necessary to practically extend HiGLASSO to spatio-temporal applications.

Because exposures never occur in isolation, identifying exposure by exposure and exposure by covariate interactions is crucial to advancing the understanding of how the environment holistically influences health. We show that nonlinearity in exposure-response associations and interactions, a common feature in epidemiologic studies, can make these effects difficult to quantify. HiGLASSO is useful in this space as a pairwise interaction detection tool that can help identify possibly nonlinear interaction effects and, ultimately, advance research on environmental chemical mixtures beyond models that strictly assume additive exposure effects or linear interaction effects.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to the sensitive nature of biological samples and demographic information in human subjects datasets.

Web Appendices, Tables, and Figures referenced in the main text are available with this paper at the Environmetrics website on Wiley Online Library. The *higlasso* R package is currently available on CRAN https://cran.r-project.org/web/packages/higlasso/index.html.

References

Ashrap, P., Watkins, D. J., Mukherjee, B., Boss, J., Richards, M. J., Rosario, Z., Vélez-Vega, C. M., Alshawabkeh, A., Cordero, J. F., and Meeker, J. D. (2020). Predictors of urinary and blood metal(loid) concentrations among pregnant women in northern puerto rico. *Environmental Research* **183,** 109178.

Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences* **2,** 183–202.

Bien, J., Taylor, J., and Tibshirani, R. (2013). A lasso for hierarchical interactions. *The Annals of Statistics* **41,** 1111–1141.

Bobb, J. F., Valeri, L., Claus Henn, B., Christiani, D. C., Wright, R. O., Mazumdar, M., Godleski, J. J., and Coull, B. A. (2015). Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. *Biostatistics* **16,** 493–508.

Boos, D. D., Stefanski, L. A., and Wu, Y. (2009). Fast fsr variable selection with applications to clinical trials. *Biometrics* **65,** 692–700.

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning* **3,** 1–122.

Breiman, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics* **37,** 373–384.

Choi, N. H., Li, W., and Zhu, J. (2010). Variable selection with the strong heredity

constraint and its oracle property. *Journal of the American Statistical Association* **105,** 354–364.

Cornelis, M. C., Tchetgen Tchetgen, E. J., Liang, L., Qi, L., Chatterjee, N., Hu, F. B., and Kraft, P. (2012). Gene-environment interactions in genome-wide association studies: a comparative study of tests applied to empirical studies of type 2 diabetes. *American Journal of Epidemiology* **175,** 191–202.

Cox, D. R. (1984). Interaction. *International Statistical Review / Revue Internationale de Statistique* **52,** 1–24.

Crews, H. B., Boos, D. D., and Stefanski, L. A. (2011). Fsr methods for second-order regression models. *Computational Statistics & Data Analysis* **55,** 2026–2037.

Crinnion, W. J. (2010). The CDC fourth national report on human exposure to environmental chemicals: what it tells us about our toxic burden and how it assist environmental medicine physicians. *Alternative Medicine Review* **15,** 101–109.

Darbre, P. D. and Harvey, P. W. (2008). Paraben esters: review of recent studies of endocrine toxicity, absorption, esterase and human exposure, and discussion of potential human health risks. *Journal of Applied Toxicology* **28,** 561–578.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science* **11,** 89–121.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96,** 1348–1360.

Ferguson, K. K., Loch-Caruso, R., and Meeker, J. D. (2011). Urinary phthalate metabolites in relation to biomarkers of inflammation and oxidative stress: Nhanes 1999-2006. *Environmental Research* **111,** 718–726.

Ferguson, K. K., Loch-Caruso, R., and Meeker, J. D. (2012). Exploration of oxidative stress and inflammatory markers in relation to urinary phthalate metabolites: Nhanes 1999-2006.

Environmental Science and Technology **46,** 477–485.

Ferguson, K. K., McElrath, T. F., Chen, Y.-H., Loch-Caruso, R., Mukherjee, B., and Meeker, J. D. (2015). Repeated measures of urinary oxidative stress biomarkers during pregnancy and preterm birth. *American Journal of Obstetrics and Gynecology* **212,** 208.e1– 208.e8.

Ferguson, K. K., McElrath, T. F., Chen, Y.-H., Mukherjee, B., and Meeker, J. D. (2015). Urinary phthalate metabolites and biomarkers of oxidative stress in pregnant women: A repeated measures analysis. *Environmental Health Perspectives* **123,** 210–216.

Ferguson, K. K., Meeker, J. D., Cantonwine, D. E., Mukherjee, B., G.Pace, G., Weller, D., and McElrath, T. F. (2018). Environmental phenol associations with ultrasound and delivery measures of fetal growth. *Environment International* **112,** 243–250.

Fu, W. J. (1998). Penalized regressions: The bridge versus the lasso. *Journal of Computational and Graphical Statistics* **7,** 397–416.

Hao, N. and Zhang, H. H. (2014). Interaction screening for ultra-high dimensional data. *Journal of the American Statistical Association* **109,** 1285–1301.

He, Z., Zhang, M., Lee, S., Smith, J. A., Kardia, S. L. R., Diez-Roux, A. V., and Mukherjee, B. (2017). Set-based tests for the gene-environment interaction in longitudinal studies. *Journal of the American Statistical Association* **112,** 966–978.

Hefley, T. J., Hooten, M. B., Hanks, E. M., Russell, R. E., and Walsh, D. P. (2017). The bayesian group lasso for confounded spatial data. *Journal of Agricultural, Biological and Environmental Statistics* **22,** 42–59.

Huang, J., Horowitz, J. L., and Wei, F. (2010). Variable selection in nonparametric additive models. *The Annals of Statistics* **38,** 2282–2313.

Leng, C., Lin, Y., and Wahba, G. (2006). A note on the lasso and related procedures in model selection. *Statistica Sinica* **16,** 1273–1284.

Li, Y. and Ding, A. A. (2019). Double-structured sparse multitask regression with application of statistical downscaling. *Environmetrics* **30,** e2534.

Lim, M. and Hastie, T. (2015). Learning interactions via hierarchical group-lasso regularization. *Journal of Computational and Graphical Statistics* **24,** 627–654.

Lin, Y. and Zhang, H. H. (2006). Component selection and smoothing in multivariate nonparametric regression. *The Annals of Statistics* **34,** 2272–2297.

Liu, S. H., Bobb, J. F., Henn, B. C., Schnaas, L., Tellez-Rojo, M. M., Gennings, C., Arora, M., Wright, R. O., Coull, B. A., and Wand, M. P. (2018). Modeling the health effects of time-varying complex environmental mixtures: Mean field variational bayes for lagged kernel machine regression. *Environmetrics* **29,** e2504.

Loh, W. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **1,** 14–23.

Luo, S. and Ghosal, S. (2015). Prediction consistency of forward iterated regression and selection technique. *Statistics & Probability Letters* **107,** 79–83.

McCullagh, P. (1984). Generalized linear models. *European Journal of Operational Research* **16,** 285–292.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models, Second Edition.* Chapman and Hall/CRC Monographs on Statistics and Applied Probability Series. Chapman & Hall.

Meeker, J. D., Calafat, A. M., and Hauser, R. (2012). Urinary phthalate metabolites and their biotransformation products: predictors and temporal variability among men and women. *Journal of Exposure Science & Environmental Epidemiology* **22,** 376–385.

Montuschi, P., Corradi, M., Ciabattoni, G., Nightingale, J., Kharitonov, S. A., and Barnes, P. J. (1999). Increased 8-isoprostane, a marker of oxidative stress, in exhaled condensate of asthma patients. *American Journal of Respiratory and Critical Care Medicine* **160,**

216–220.

Mukherjee, B., Ahn, J., Gruber, S. B., and Chatterjee, N. (2012). Testing gene-environment interaction in large-scale case-control association studies: possible choices and comparisons. *American Journal of Epidemiology* **175,** 177–190.

Narisetty, N. N., Mukherjee, B., Chen, Y., Gonzalez, R., and Meeker, J. D. (2018). Selection of nonlinear interactions by a forward stepwise algorithm: Application to identifying environmental chemical mixtures affecting health outcomes. *Statistics in Medicine* **38,** 1582–1600.

Nelder, J. A. (1977). A reformulation of linear models. *Journal of the Royal Statistical Society. Series A (General)* **140,** 48–77.

Pan, Q. and Zhao, Y. (2016). Integrative weighted group lasso and generalized local quadratic approximation. *Computational Statistics & Data Analysis* **104,** 66–78.

Radchenko, P. and James, G. M. (2010). Variable selection using adaptive nonlinear interaction structures in high dimensions. *Journal of the American Statistical Association* **105,** 1541–1553.

Roberts, S. and Martin, M. (2005). A critical assessment of shrinkage-based regression approaches for estimating the adverse health effects of multiple air pollutants. *Atmospheric Environment* **39,** 6223–6230.

Schettler, T. (2006). Human exposure to phthalates via consumer products. *International Journal of Andrology* **29,** 134–139.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **58,** 267–288.

Wang, H. and Leng, C. (2008). A note on adaptive group lasso. *Computational Statistics & Data Analysis* **52,** 5277–5286.

Wang, H., Li, G., and Jiang, G. (2007). Robust regression shrinkage and consistent variable

selection through the lad-lasso. *Journal of Business & Economic Statistics* **25,** 347–355.

Wasserman, L. and Roeder, K. (2009). High-dimensional variable selection. *The Annals of Statistics* **37,** 2178–2201.

Watkins, D. J., Ferguson, K. K., Anzalota Del Toro, L. V., Alshawabkeh, A. N., Cordero, J., and Meeker, J. D. (2015). Associations between urinary phenol and paraben concentrations and markers of oxidative stress and inflammation among pregnant women in puerto rico. *International Journal of Hygiene and Environmental Health* **218,** 212–219.

Wu, J., Devlin, B., Ringquist, S., Trucco, M., and Roeder, K. (2010). Screen and clean: a tool for identifying interactions in genome-wide association studies. *Genetic Epidemiology* **34,** 275–285.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B* **68,** 49–67.

Zhang, H. H. and Lu, W. (2007). Adaptive lasso for cox's proportional hazards model. *Biometrika* **94,** 691–703.

Zhang, M., Yu, Y., Wang, S., Salvatore, M., Fritsche, L. G., He, Z., and Mukherjee, B. (2020). Interaction analysis under misspecification of main effects: Some common mistakes and simple solutions. *Statistics in Medicine* **39,** 1675–1694.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* **101,** 1418–1429.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **67,** 301–320.

Zou, H. and Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *The Annals of Statistics* **37,** 1733–1751.
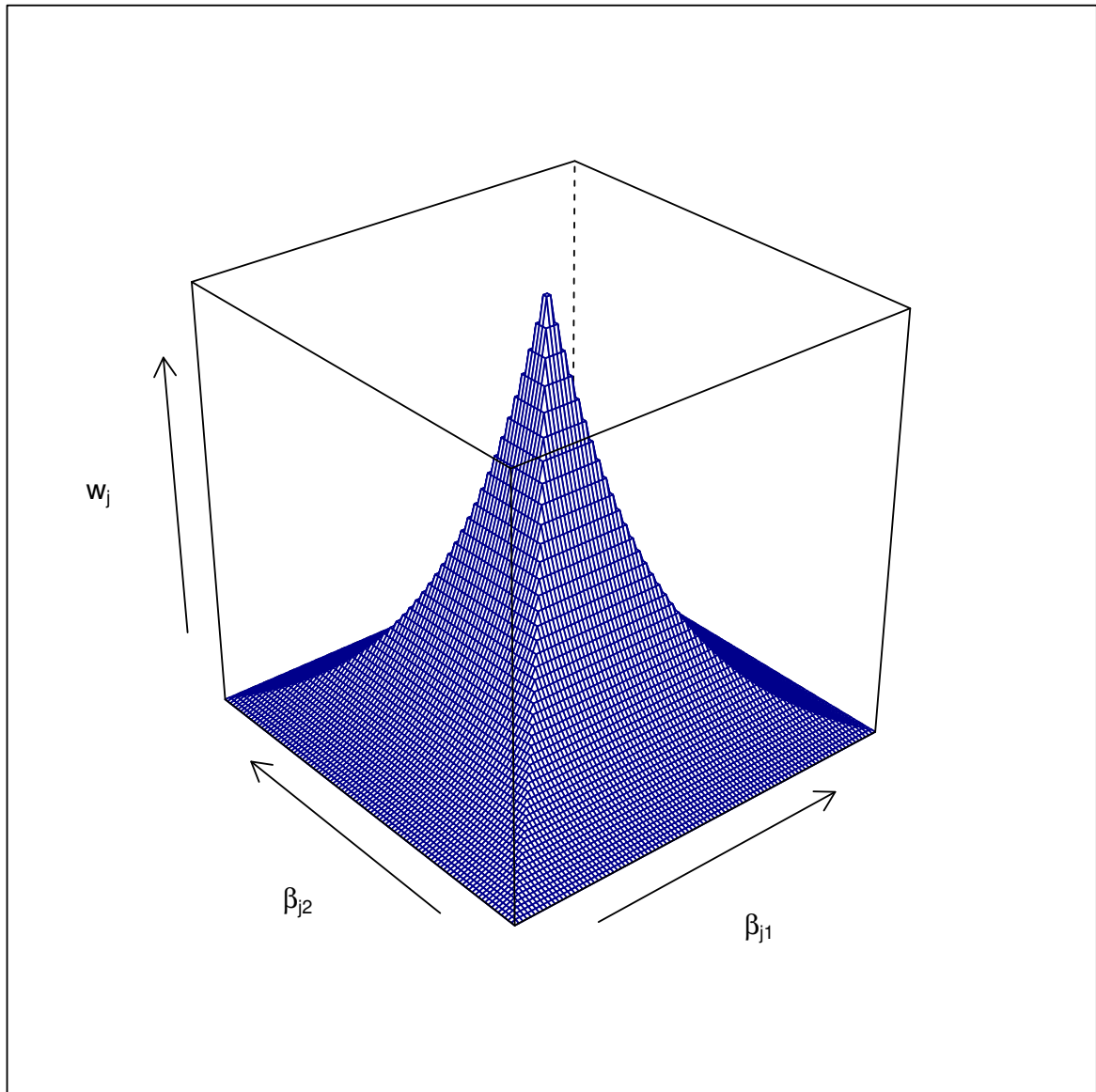
Figure 1. HiGLASSO weight function evaluated for a two-dimensional vector in $[-3, 3] \times [-3, 3]$ with $\sigma = 1$.
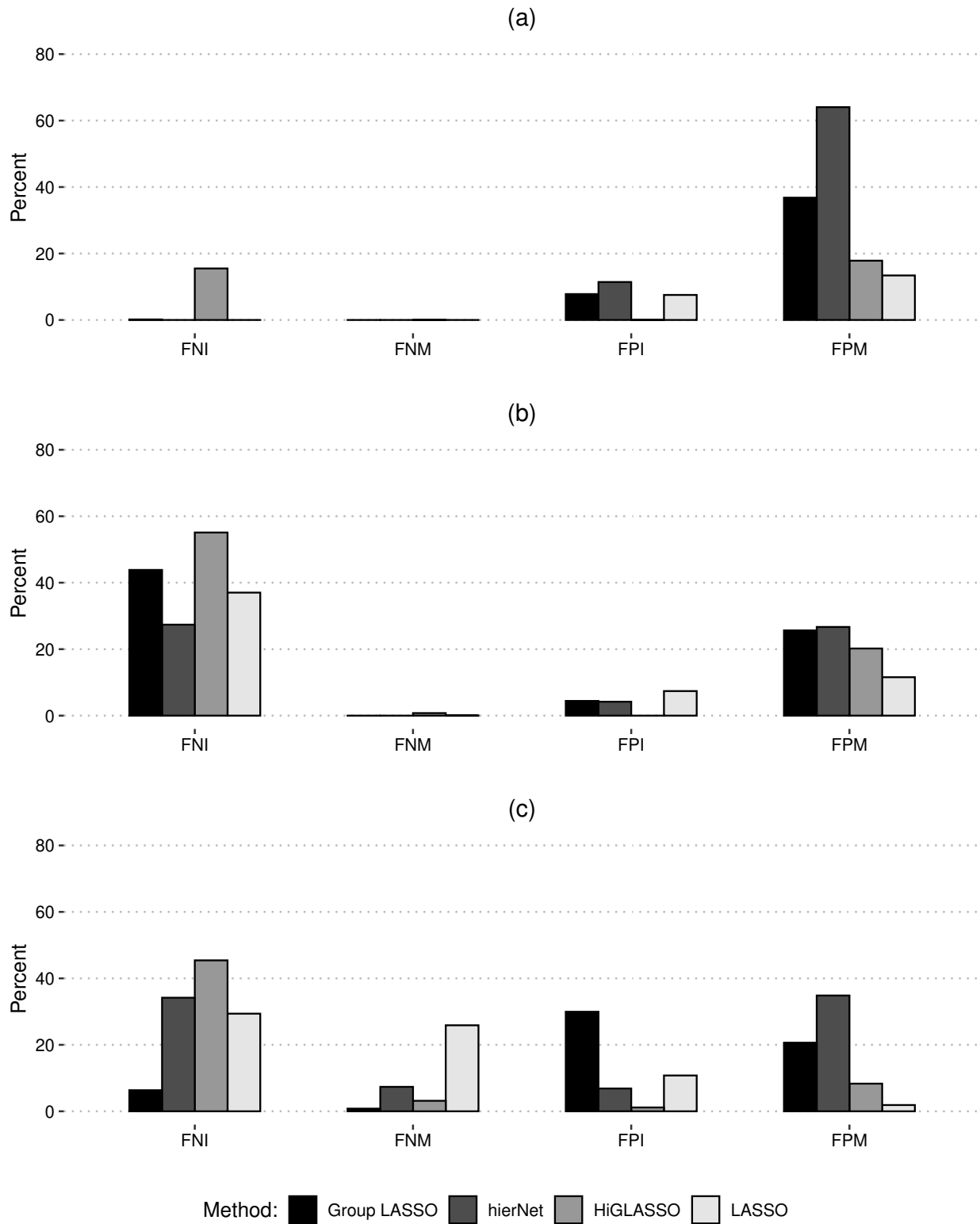
Figure 2. Simulation Results for the $n = 1000$ and $p = 10$ cases: (a) L10 (b) PL10 (c) NL10. FNI, FNM, FPI, and FPM are defined in Section 4.2.
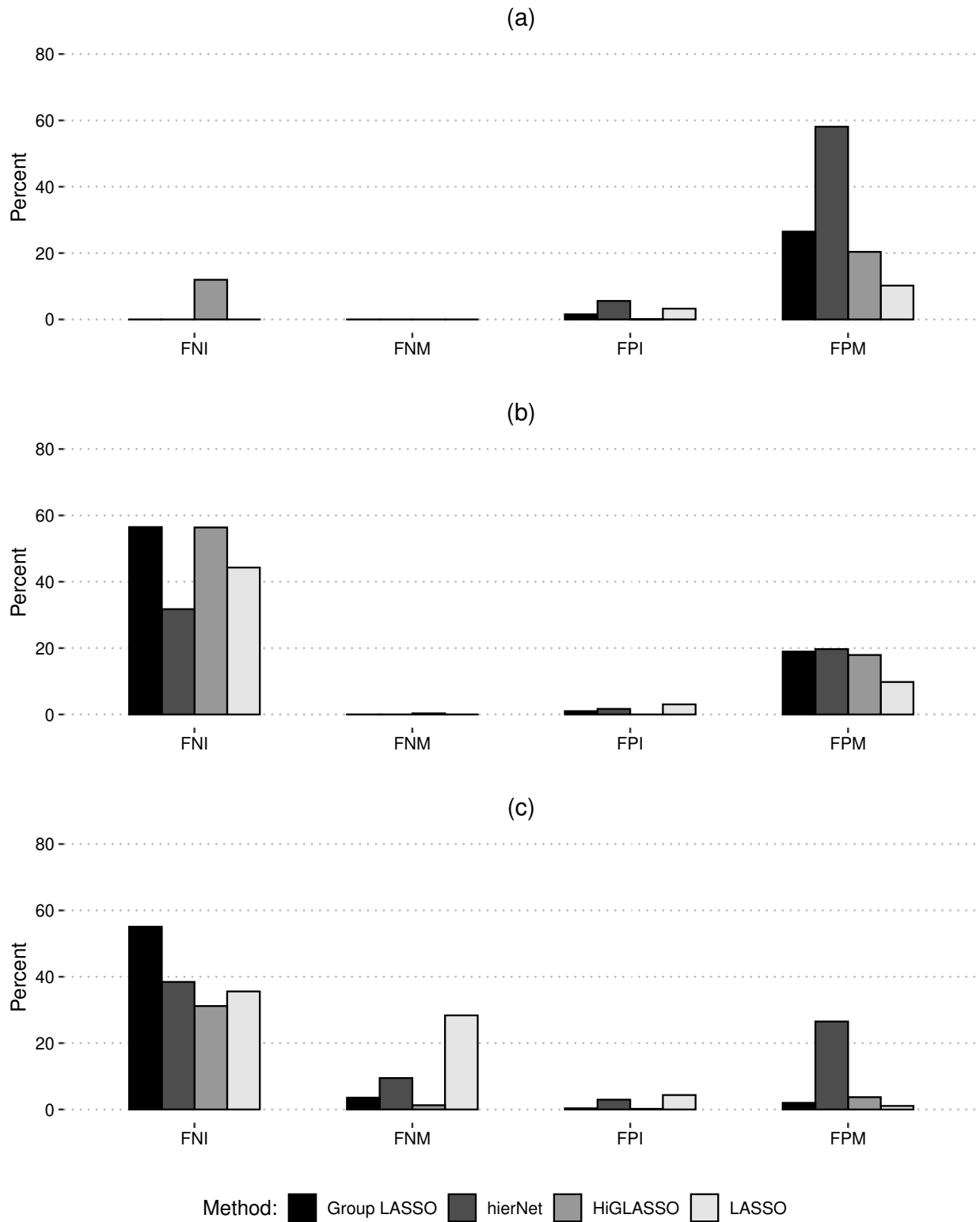
Figure 3. Simulation Results for the $n = 1000$ and $p = 20$ cases: (a) L20 (b) PL20 (c) NL20. FNI, FNM, FPI, and FPM are defined in Section 4.2.

Scenario	p	Mean Function
L10, L20	10, 20	$x_1 + x_2 + x_3 + x_4 + x_5 + x_1 x_2 + x_1 x_3 + x_1 x_4 +$ $x_1 x_5 + x_2 x_3 + x_2 x_4 + x_2 x_5 + x_3 x_4 + x_3 x_5 + x_4 x_5$
PL10, PL20	10, 20	$x_1 I(x_1 > 0) + x_2 I(x_2 < 0) + x_3 I(x_3 > 0.5) + x_4 I(x_4 > 0) + x_5 I(x_5 < -0.5) +$ $x_1 x_2 I(x_1 > 0) I(x_2 < 0) + x_1 x_3 I(x_1 > 0) I(x_3 > 0.5) + x_1 x_4 I(x_1 > 0) I(x_4 > 0) +$ $x_1 x_5 I(x_1 > 0) I(x_5 < -0.5) + x_2 x_3 I(x_2 < 0) I(x_3 > 0.5) + x_2 x_4 I(x_2 < 0) I(x_4 > 0) +$ $x_2 x_5 I(x_2 < 0) I(x_5 < -0.5) + x_3 x_4 I(x_3 > 0.5) I(x_4 > 0) +$ $x_3 x_5 I(x_3 > 0.5) I(x_5 < -0.5) + x_4 x_5 I(x_4 > 0) I(x_5 < -0.5)$
NL10, NL20	10, 20	$x_1 I(x_1 > 0) + \exp(x_2) + |x_3| + x_4^2 + (x_5 + 1)^2 + x_1 \exp(x_2) I(x_1 > 0)$ $x_1 |x_3| I(x_1 > 0) + x_1 x_4^2 I(x_1 > 0) + x_1 (x_5 + 1)^2 I(x_1 > 0) + \exp(x_2) |x_3| +$ $\exp(x_2) x_4^2 + \exp(x_2)(x_5 + 1)^2 + |x_3| x_4^2 + |x_3|(x_5 + 1)^2 + x_4^2 (x_5 + 1)^2$

Table 1

Mean specifications for all simulation scenarios. In the scenario column, "L" indicates linear main and pairwise interaction effects, "PL" indicates piecewise linear main and interaction effects, and "NL" indicates nonlinear main and interaction effects. p represents the number of predictors.

Exposure Class	Full Name	Acronym
Phthalates		
	mono-n-butyl	MBP
	monobenzyl	MBzP
	mono(3-carboxypropyl)	MCPP
	mono(2-ethyl-5-carboxypentyl)	MECPP
	mono(2-ethyl-5-hydroxyhexyl)	MEHHP
	mono(2-ethylhexyl)	MEHP
	mono(2-ethyl-5-oxohexyl)	MEOHP
	monoethyl	MEP
	monoisobutyl	MiBP
	Summed di(2-ethylhexyl)	DEHP
Phenols		
	2,4-Dichlorophenol	2,4-DCP
	2,5-Dichlorophenol	2,5-DCP
	benzophenone-3	BP3
	Bisphenol A	BPA
	Bisphenol S	BPS
	butyl paraben	BuPB
	ethyl paraben	EtPB
	methyl paraben	MePB
	propyl paraben	PrPB
	triclocarban	TCC
	triclosan	TCS

Table 2

List of 21 exposure measurements including 10 phthalates and 11 phenols in the LIFECODES dataset.

Selected Term	LASSO	hierNet	Group LASSO	HiGLASSO
MBP	✓	✓	✓	
MBzP	✓	✓	✓	✓
MCPP		✓	✓	✓
MECPP	✓	✓	✓	✓
MEP	✓	✓	✓	✓
MiBP	✓	✓	✓	
BuBP	✓	✓	✓	✓
BPS	✓	✓	✓	✓
2,5-DCP	✓	✓	✓	✓
EtPB			✓	
MePB			✓	✓
TCC	✓	✓	✓	✓
MBP×BPA			✓	
MBP×MBzP		✓		
MBP×MCPP	✓	✓		
MBzP×MCPP	✓	✓	✓	✓
MECPP×BP3			✓	
MECPP×BPA			✓	
MEP×TCS	✓			
MiBP×MBzP		✓		
BP3×BPA			✓	
BPS×2,5-DCP			✓	✓

Table 3
Selected main effects and interaction effects from the LIFECODES study. Candidate main and interaction effects that were not selected are omitted for brevity.