

RESEARCH ARTICLE

Model choice for estimating the association between exposure to chemical mixtures and health outcomes: A simulation study

Lauren Hoskovec^{1*}, Wande Benka-Coker², Rachel Severson², Sheryl Magzamen², Ander Wilson¹

1 Department of Statistics, Colorado State University, Fort Collins, CO, United states of America,

2 Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins, CO, United states of America

* lvheck@rams.colostate.edu



OPEN ACCESS

Citation: Hoskovec L, Benka-Coker W, Severson R, Magzamen S, Wilson A (2021) Model choice for estimating the association between exposure to chemical mixtures and health outcomes: A simulation study. PLoS ONE 16(3): e0249236. <https://doi.org/10.1371/journal.pone.0249236>

Editor: Jaymie Meliker, Stony Brook University, Graduate Program in Public Health, UNITED STATES

Received: October 8, 2020

Accepted: March 13, 2021

Published: March 25, 2021

Copyright: © 2021 Hoskovec et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data for the study are protected under Colorado State University's Institutional Review Board (IRB). Data cannot be shared publicly due to concerns for protection of human subjects. The data contain protected health information and data that could potentially identify subjects. The data underlying the results presented in the study are available from Colorado State University's IRB. The IRB contact email is RICRO_IRB@mail.colostate.edu.

Abstract

Challenges arise in researching health effects associated with chemical mixtures. Several methods have recently been proposed for estimating the association between health outcomes and exposure to chemical mixtures, but a formal simulation study comparing broad-ranging methods is lacking. We select five recently developed methods and evaluate their performance in estimating the exposure-response function, identifying active mixture components, and identifying interactions in a simulation study. Bayesian kernel machine regression (BKMR) and nonparametric Bayes shrinkage (NPB) were top-performing methods in our simulation study. BKMR and NPB outperformed other contemporary methods and traditional linear models in estimating the exposure-response function and identifying active mixture components. BKMR and NPB produced similar results in a data analysis of the effects of multipollutant exposure on lung function in children with asthma.

Introduction

Individuals are continuously exposed to complex mixtures of environmental chemicals. Mounting evidence from epidemiological studies links environmental exposures to increased morbidity and mortality [1–5]. Traditional epidemiological studies have focused on a single pollutant and additive models with a small number of exposures; however, studying pollutants in isolation can lead to biased estimates [6, 7] and does not reflect the reality that people are jointly exposed to mixtures of pollutants. Hence, interest is rapidly growing in studying health outcomes associated with simultaneous exposure to mixtures of pollutants [8, 9]. The National Institute for Environmental Health Sciences (NIEHS) identified the study of mixtures as a goal in its 2012–2017 strategic plan while noting that this will require novel quantitative approaches [10]. As such, numerous statistical methods have been proposed. There is a need to identify the most appropriate statistical methods currently available for estimating health outcomes associated with exposure to mixtures [11, 12].

Funding: This work was supported by grants ES023815 awarded to A.W. and ES028811 awarded to S.M. from the National Institutes of Health (www.nih.gov). This work utilized the RMACC Summit supercomputer, which is supported by the National Science Foundation (www.nsf.gov, awards ACI-1532235 and ACI-1532236 to A.W. and L.H.), the University of Colorado Boulder and Colorado State University. The RMACC Summit supercomputer is a joint effort of the University of Colorado Boulder and Colorado State University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. This research was supported by USEPA grants RD-839278. Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the USEPA. Further, USEPA does not endorse the purchase of any commercial products or services mentioned in the publication.

Competing interests: The authors have declared that no competing interests exist.

Studying health outcomes associated with exposure to mixtures is complicated by small effect sizes, highly correlated exposures, possible nonlinear and interaction effects, and often small sample sizes. In this context, traditional regression methods are often inadequate as they may yield biased or unstable estimates [13] and have low power to detect effects, especially in the case of nonlinear associations and interactions. Common methods designed for variable selection tend to incorrectly select predictors when many predictors are highly correlated [14] and classical model selection techniques ignore uncertainty in both the selected model and selected mixture components when estimating the exposure-response function [15, 16].

In a broad literature review, Davalos et al. [17] identified five classes of methods currently used in mixtures analyses: additive main effects (AME), effect measure modification (EMM), unsupervised dimension reduction (UDR), supervised dimension reduction (SDR), and non-parametric (NP). AME and EMM methods are typically regression based. AME allows only additive effects, while EMM includes multiplicative interactions. Hierarchical and penalized regression methods have been applied to AME and EMM models to identify important mixture components and improve precision [18–22]. The next two groups are dimension reduction techniques (UDR and SDR) that transform exposure data to reduce the dimension of the predictor and, therefore, the required parameter space. UDR methods such as *k*-means [23, 24] transform exposure data without regard to the health outcome [25–28]. SDR methods, including supervised principle components analysis [29], let the outcome inform exposure data transformation [30–35]. Finally, NP methods like Bayesian kernel machine regression [36] are flexible data-driven techniques for estimating a complex exposure-response function that may include interactions and nonlinear effects [37, 38].

Choosing an appropriate statistical model depends on the research objectives [11, 39] and requires understanding the empirical performance of methods. Recent studies have compared several methods in subsets of the model classes proposed. Among those evaluated include linear regression AME [40] and EMM methods [14], principle components analysis [34], structural equation models [41], Bayesian kernel machine regression [41], and Bayesian semi-parametric regression [42]. These studies highlight challenges induced by highly correlated data in estimating complex exposure-response functions and characterizing uncertainty. To our knowledge, there has been no formal evaluation of methods from all five classes identified by Davalos et al. [17] in a single simulation study.

Evaluating the empirical performance of methods across a wide spectrum of model classes is important as it guides researchers in choosing across classes of models and aids in interpreting results and understanding the limitations of epidemiological studies using these methods. In addition, the existing literature is sparse with regards to a comparison among Bayesian methods, which are favorable in the multipollutant setting as they can incorporate prior information and fully characterize uncertainty [12, 34, 43]. To this end, we focus on a comparison of Bayesian methods across a variety of model classes in this paper. By comparing performance across classes of models, researchers can also gain insight into promising future directions for statistical methods development.

Motivated by research linking mixtures of air pollutant and pesticide exposures to child respiratory health, we conducted a simulation study to compare contemporary methods developed for estimating the association between health outcomes and exposure to mixtures. We considered one method from each of the five classes identified by Davalos et al. [17] and evaluated each method in three data-generating scenarios. The data-generating scenarios cover a range of linear to nonlinear functions of multiple pollutants with synergistic effects on the response in order to test each method in its ability to estimate both simple and complex exposure-response functions that may be encountered in practice.

In contrast to many recent studies that have compared methods from a conceptual standpoint or compared their performance in the analysis of a single data set, the primary contribution of our work is to compare diverse methods in a simulation study addressing a variety of research questions. Specifically, we quantified four aspects of model performance corresponding to previously identified epidemiological questions of interest: 1) how well does the model estimate the exposure-response function, 2) can the model identify important mixture components, 3) can the model identify components not associated with the outcome, and 4) can the model identify interactions among exposures [39].

A secondary contribution of our work is to provide software for the tested methods that currently lack software. Our simulation study describes the strengths and weaknesses of each method and available software encourages practitioners to use the most appropriate methods in a given application. Software is available in the form of the R package `mmpack` [44] to reproduce the simulation. Further, the software allows researchers to easily conduct a simulation study using the same methods and simulated exposure-response functions but substituting in their own exposure data which will have a different correlation structure and may result in different model performance. Hence, researchers can determine which methods are most appropriate for their own study. Finally, we applied each method to a data analysis of a cohort study investigating the relationship between air pollutant and pesticide exposures and lung function in children with asthma. We describe the differences in results among the methods, highlighting the importance of model choice.

Materials and methods

Data

Health data. This study was approved by the Institutional Review Board of Colorado State University, Protocol Number 19-9437H. This was a secondary data analysis from a closed cohort with all personal identifying information stripped from the database. We used data from Fresno Asthmatic Children's Environment Study (FACES). The study design, including recruitment, eligibility criteria, and measurement procedures, is described elsewhere [45–50]. FACES includes data for children aged 6–11 years with asthma symptoms at the time of enrollment and living within a 20 kilometer radius of one of Fresno's EPA air quality monitoring sites. The health outcome of interest was baseline forced expiratory volume in the first second (FEV_1) measured via spirometry. We regressed FEV_1 on age, sex, height and ethnicity and used the residuals as the outcome in our data analysis [51–53]. Age, sex, height, and ethnicity are well-known predictors of FEV_1 so we remove all variation from these predictors before looking into the effects of air pollution and pesticide exposure on FEV_1 . Other covariates have not been as well studied regarding their association with FEV_1 and are including the model as potential confounding variables. Complete exposure, health, and covariate data were available for 153 children.

The data contain information on covariates and potential confounding variables (S1 Table in [S1 File](#)). We included average temperature and precipitation over three months, the temporal scale of the pesticide exposure data, prior to baseline as covariates. Subject-specific covariates include body mass index (BMI, kg/m^2) and indicators for: self-reported residence within one block of a freeway, any smoking in the home, positive atopy or allergy test, modified Global Initiative for Asthma (GINA) score ≥ 3 at baseline, household income greater than \$30K/year, mother having post-secondary education, child not covered by insurance, and season of baseline spirometry test. Temperature, precipitation, and BMI were scaled to have mean 0 and variance 1. Approximately 1% of the covariate data was missing, including any smoking in the home (16%), household income (3%), and mother having post-secondary

Table 1. Pesticide and air pollutant exposure data summary statistics. Table shows mean, standard deviation (SD), minimum, 25th percentile, median, 75th percentile, and maximum concentration for each exposure.

	mean	SD	min	25 th	median	75 th	max
C × 10 ⁶ (kg/3km ²)	0.15	0.33	0.00	0.00	0.00	0.15	2.35
MeBr × 10 ⁶ (kg/3km ²)	3.88	9.90	0.00	0.00	0.00	0.00	48.92
OP × 10 ⁶ (kg/3km ²)	0.93	1.08	0.00	0.00	1.11	1.17	5.40
O ₃ (ppb)	0.04	0.01	0.01	0.03	0.04	0.04	0.06
NO ₂ (ppb)	15.48	3.26	9.49	12.64	14.42	17.96	23.07
PM _{2.5} (μg/m ³)	16.35	9.80	6.66	10.14	11.23	18.20	40.21
PM ₁₀ (μg/m ³)	37.89	10.68	19.55	30.30	32.49	47.23	65.94

<https://doi.org/10.1371/journal.pone.0249236.t001>

education (1%). As all covariates with missing data were binary variables, we singly imputed the missing values with 0 and then added a dummy variable for each covariate with any missing data that indicated which values of that covariate were missing.

Air pollution and pesticide data. We obtained air pollution data from the EPA Air Quality System Data Mart. Air pollutant concentrations were calculated as 24-hour averages for particles ≤ 2.5 μm in aerodynamic diameter (PM_{2.5}) and particles ≤ 10 μm in aerodynamic diameter (PM₁₀), 8-hour daily maximum levels for ozone (O₃) and one-hour daily maximum levels for nitrogen dioxide (NO₂) [47]. Concentrations were taken from the air monitoring site closest to each child's residence and exposure levels were summarized as averages over three months prior to baseline spirometry tests to be consistent with available pesticide exposure data. Due to right-skewed distributions, air pollutant exposures were square-root transformed and then scaled to have mean 0 and variance 1.

We obtained data on the date, location, and amount (kilograms) of applied agricultural pesticides from the California Pesticide Use Report (PUR) [54]. Based on previous evidence linking pesticide exposure to respiratory illness [55, 56], we considered three pesticide classes: carbamates (C), methyl bromide (MeBr), and organophosphates (OP). Pesticide exposures were estimated using the `purexposure` [57] package in R. We applied inverse distance weighting to the total reported pesticide amount over three months prior to baseline spirometry tests (as PUR reports are aggregated quarterly) to estimate exposures within a 3km buffer of each child's residence. Pesticide exposures were also highly skewed and so were square-root transformed and then scaled to have mean 0 and variance 1.

Exposure data summary statistics are shown in Table 1. Strong Spearman correlation existed between NO₂ and PM_{2.5} ($\rho = 0.88$) and between NO₂ and PM₁₀ ($\rho = 0.72$). Moderate Spearman correlation existed between PM_{2.5} and PM₁₀ ($\rho = 0.67$), O₃ and NO₂ ($\rho = -0.63$), O₃ and PM_{2.5} ($\rho = -0.54$), O₃ and OP ($\rho = 0.53$), and OP and NO₂ ($\rho = -0.53$) (Table 2).

Table 2. Spearman correlation among all pairs of air pollutant and pesticide exposures.

	MeBr	OP	O ₃	NO ₂	PM _{2.5}	PM ₁₀
C	0.27	0.12	0.09	0.08	0.06	0.01
MeBr		-0.08	0.02	0.07	-0.03	-0.13
OP			0.53	-0.53	-0.38	-0.24
O ₃				-0.63	-0.54	-0.22
NO ₂					0.88	0.72
PM _{2.5}						0.67

<https://doi.org/10.1371/journal.pone.0249236.t002>

Statistical methods

Our primary interest was to estimate the association between exposures to p pollutants $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ and a continuous outcome y_i , while controlling for q potential confounders $\mathbf{w}_i = (w_{i1}, \dots, w_{iq})^T$ in a sample $i = 1, \dots, n$. We considered five recently proposed methods. The first two are the AME model nonparametric Bayes shrinkage with main effects only (NPBr) and the EMM model nonparametric Bayes shrinkage with main effects and all pairwise multiplicative interactions (NPB) as proposed by Herring [19]. The next two models are unsupervised (UPR) and supervised Bayesian profile regression (SPR) as proposed by Molitor et al. [58]. The fifth is the NP model Bayesian kernel machine regression (BKMR) [36]. We chose these methods since they represent the five classes identified by Davalos et al. [17] and are recently developed Bayesian methods for estimating health outcomes associated with exposure to mixtures. These five methods cover a variety of exposure-response function shapes, handle multicollinearity in various ways, and include options for variable selection. BKMR is presented exactly as proposed by Bobb et al. [36]; NPB and SPR have been modified to accommodate the continuous outcome with normal residuals rather than the logistic model originally proposed by Herring [19] and Molitor et al. [58], respectively; and NPBr and UPR are further modifications of those previously introduced methods. For a baseline comparison, we also included a normal linear model with main effects only (LM) and with all pairwise interactions (LM-int), both estimated with least squares. All models considered in this paper have the form

$$y_i = h(\mathbf{x}_i) + \mathbf{w}_i^T \boldsymbol{\gamma} + \epsilon_i, \quad (1)$$

where ϵ_i are independent $N(0, \sigma^2)$ and $h(\mathbf{x}_i)$ represents the exposure-response function. All models were fit in R version 3.6.0 [59].

Nonparametric Bayes shrinkage. Nonparametric Bayes shrinkage [19] was originally introduced as a logistic regression EMM model and was adapted to the linear regression setting used here. We consider two variations. NPB, originally proposed by Herring [19] is an EMM model including main effects and all pairwise interactions, where

$$h(\mathbf{x}_i) = \sum_{j=1}^p x_{ij} \beta_j + \sum_{j=1}^{p-1} \sum_{k=j+1}^p x_{ij} x_{ik} \zeta_{jk}. \quad (2)$$

NPBr is a reduced AME model not originally proposed in Herring [19] that includes only main effects:

$$h(\mathbf{x}_i) = \sum_{j=1}^p x_{ij} \beta_j. \quad (3)$$

Both models place a Dirichlet Process (DP) prior on regression coefficients. The base distribution of the DP is a finite mixture of a normal distribution and a point mass at 0 to induce sparsity in the model. Hence, some coefficients are set exactly to 0, effectively selecting out variables that do not contribute to the health outcome. Correlated exposures can be clustered and assigned equal regression coefficients to reduce variance [19, 60]. This effectively reparameterizes the model to have a single effect for the sum of two correlated predictors and is particularly advantageous in situations where it is difficult to differentiate the effects of two highly

correlated predictors. The DP prior for main effects is constructed as:

$$\begin{aligned} \beta_j|D_1 &\sim D_1, \quad j = 1, \dots, p \tag{4} \\ D_1|\alpha_1, D_{01} &\sim \text{DP}(\alpha_1 D_{01}) \\ D_{01}|\pi_{01}, G_1 &= \pi_{01} \delta_0 + (1 - \pi_{01}) G_1 \\ G_1|\mu_1, \phi_1^2 &\equiv \text{N}(\mu_1, \phi_1^2), \end{aligned}$$

where δ_0 represents the Dirac delta function at 0. The model is completed with standard hyperpriors $\alpha_1 \sim \text{Gamma}(\alpha_{\alpha_1}, \beta_{\alpha_1})$, $\pi_{01} \sim \text{Beta}(\alpha_{\pi_1}, \beta_{\pi_1})$, $\mu_1 \sim \text{N}(0, \sigma_{\mu_1}^2)$, and $\phi_1^{-2} \sim \text{Gamma}(\alpha_{\phi_1}, \beta_{\phi_1})$. The DP prior for interactions is similarly constructed. Specifically,

$$\begin{aligned} \zeta_{jk}|D_2 &\sim D_2, \quad j = 1, \dots, p - 1 \ \& \ k = j + 1, \dots, p \tag{5} \\ D_2|\alpha_2, D_{02} &\sim \text{DP}(\alpha_2 D_{02}) \\ D_{02}|\pi_{02}, G_2 &= \pi_{02} \delta_0 + (1 - \pi_{02}) G_2 \\ G_2|\mu_2, \phi_2^2 &\equiv \text{N}(\mu_2, \phi_2^2). \end{aligned}$$

The hyperpriors on α_2 , π_{02} , μ_2 , and ϕ_2^{-2} come from the same families specified for the main effects. The distributions on the main effects and interactions are independent a priori.

Posterior inclusion probabilities (PIPs) are calculated for each mixture component as the posterior probability of the regression coefficient being assigned a non-zero value. Both NPBr and NPb were fit using the R package `mmpack` [44].

Bayesian profile regression. Bayesian profile regression is a dimension reduction approach that classifies pollutant exposure profiles, \mathbf{x}_i , into a parsimonious set of clusters using a DP mixture model (DPMM) [58, 61]. Each cluster represents a group of observations with similar exposure levels across the vector of pollutants. The health outcome is regressed on cluster indicators to estimate

$$h(\mathbf{x}_i) = \theta_c \tag{6}$$

if profile \mathbf{x}_i is assigned to cluster c . We introduce a latent variable $z_i = c$ if exposure profile i is assigned to cluster c . Conditional on cluster assignment, the model for an individual exposure profile is

$$\begin{aligned} \mathbf{x}_i|z_i = c, \mu_c, \Sigma_c &\sim \text{N}(\mu_c, \Sigma_c) \tag{7} \\ \mu_c &\sim \text{N}(v_0, \Lambda_0) \\ \Sigma_c^{-1} &\sim \text{Wish}_p(R, r). \end{aligned}$$

The DPMM for cluster assignment places a truncated stick-breaking prior on the assignment

probabilities to each cluster. The stick-breaking process and cluster assignment model are

$$V_1, \dots, V_{C-1} | \alpha \sim \text{Beta}(1, \alpha), \quad V_C = 1 \quad (8)$$

$$\alpha \sim \text{Gamma}(\alpha_x, \beta_x)$$

$$P(z_i = c) = \psi_c = V_c \prod_{h=1}^{c-1} (1 - V_h)$$

$$z_i \sim \text{Categorical}(\psi).$$

Subject to a maximum of C clusters, the DPMM allows the number of non-empty clusters to be estimated from the data. To identify the most optimal partitioning of the data, we follow the approach described in Dahl [62] and Molitor et al [58]. First, we construct an $n \times n$ score matrix at each iteration with a 1 in the i, j location if individuals i and j belong to the same cluster and a 0 otherwise. Then we calculate a probability matrix \mathbf{S} by averaging the score matrices. The most optimal clustering is the clustering from the MCMC iteration that has a score matrix with minimum least squared distance to the probability matrix \mathbf{S} . We calculate model averaged estimates of the cluster-specific parameters θ_c to incorporate the uncertainty present in the best clustering [58].

The model has been extended to include variable selection to identify mixture components actively contributing to cluster assignment [63–65]. Briefly, binary random variables are introduced that indicate whether the mean for a mixture component within a cluster is unique to that cluster or common among all clusters. Hence, mixture components that are selected into the model are interpreted as being informative in partitioning the exposure data into clusters, but are not necessarily related to the health outcome.

We consider two variations of profile regression. The first, supervised profile regression (SPR), originally introduced by Molitor et al. [58] belongs to the SDR class of methods since cluster assignments are influenced by the health outcome. The second is an unsupervised adaptation (UPR) not originally proposed by Molitor et al. [58] that belongs to the UDR class. The difference between the two variations manifests when the latent cluster assignment variable z_i is updated. In the supervised case, we jointly model the response and estimate cluster assignments. Hence, there is feedback between the health outcome model and the profile assignment model where the health outcomes can influence cluster assignment. The full conditional for z_i depends on both the likelihood of exposures \mathbf{x}_i and the likelihood of the response y_i :

$$P(z_i = c | \mathbf{x}_i, y_i, \cdot) = \frac{\psi_c f(\mathbf{x}_i | z_i = c, \mu_c, \Sigma_c) f(y_i | z_i = c, \theta_c, \beta, \sigma^2)}{\sum_{c=1}^C \psi_c f(\mathbf{x}_i | z_i = c, \mu_c, \Sigma_c) f(y_i | z_i = c, \theta_c, \beta, \sigma^2)}. \quad (9)$$

Hence, in SPR, individuals with similar exposure profiles but different health outcomes may be assigned to different clusters depending on their responses.

The unsupervised case involves a two-step procedure where we first estimate cluster assignments independently of the response and then model the response conditional on cluster assignment. Here, z_i depends only on the exposure likelihood:

$$P(z_i = c | \mathbf{x}_i, \cdot) = \frac{\psi_c f(\mathbf{x}_i | z_i = c, \mu_c, \Sigma_c)}{\sum_{c=1}^C \psi_c f(\mathbf{x}_i | z_i = c, \mu_c, \Sigma_c)}. \quad (10)$$

Since the response does not inform cluster assignment in UPR, there may be high uncertainty

in the estimates of the cluster indicators θ_c if individuals with similar exposure profiles have very different health outcomes.

We fit SPR using the R package `PREMIUM` [65] and UPR using the R package `mmpack` developed for this paper [44].

Bayesian kernel machine regression. Bayesian kernel machine regression (BKMR) [36] belongs to the NP class of methods and flexibly models the exposure-response function to allow for nonlinear associations and higher order interactions. In BKMR, $h(\mathbf{x})$ is a smooth function represented using a Gaussian kernel. The response is modeled as

$$y_i \sim N(h_i + \mathbf{w}_i^T \boldsymbol{\gamma}, \sigma^2) \quad (11)$$

$$\mathbf{h} \equiv (h_1, \dots, h_n)^T \sim N(\mathbf{0}, \tau K),$$

where K is the kernel matrix with (i, i') element $K(\mathbf{x}_i, \mathbf{x}_{i'}) = \exp\{-\sum_{j=1}^p r_j(x_{ij} - x_{i'j})^2\}$, τ is a hyperparameter, and $\mathbf{r} = (r_1, \dots, r_p)^T$ are variable selection parameters. Estimated health outcomes for individuals with similar exposure levels across the p predictors are shrunk towards each other, resulting in a smooth but flexible exposure-response function.

BKMR allows for both component-wise and hierarchical variable selection (HVS) to identify important mixture components. In our simulation and data analysis, we implemented component-wise variable selection and calculated PIPs for each exposure. We also implemented HVS in our data analysis to address sensitivity of results. We partitioned the mixture components into groups of air pollutants (PM_{2.5}, PM₁₀, NO₂, and O₃) and pesticides (C, MeBr, and OP) and calculated PIPs for each group (group PIPs) and each component within a group, conditional on group inclusion (conditional PIPs). We fit BKMR using the R package `bkmr` [66].

Simulation study design

We evaluated the proposed methods in a simulation study. We ensure a realistic correlation structure among the pollutants by using the observed exposure data from 153 individuals in the FACES data set in our simulation study. We also use the observed covariate data in our simulation study. Health responses were simulated for three exposure-response scenarios, denoted h_k , $k = 1, 2, 3$, as $y_i = h_k(\mathbf{x}_i) + \mathbf{w}_i^T \boldsymbol{\gamma} + \varepsilon_i$, with $\varepsilon_i \sim N(0, 1)$. The covariate coefficients $\gamma_1, \dots, \gamma_q$ were simulated as independent $N(0, 1)$.

The first scenario, h_1 (linear), is an EMM model. For exposures x_j , $j = 1, \dots, 4$, the exposure-response function is

$$h_1(\mathbf{x}) = x_1 - x_2 + x_3 - x_4 + 0.7x_1x_2 - 0.5x_3x_4. \quad (12)$$

Second, h_2 (nonlinear) includes nonlinear sigmoidal functions of three pollutants and a multiplicative interaction between two of those pollutants:

$$h_2(\mathbf{x}) = \frac{2}{1 + \exp(-3x_1)} + \frac{2}{1 + \exp(-5x_2)} - \frac{2}{1 + \exp(-5x_3)} - 0.4x_1x_2. \quad (13)$$

Last, h_3 (fixed profiles) groups individuals into four distinct clusters based on dichotomous cut-offs for two pollutants. We assign a constant health effect to individuals in the same

cluster:

$$h_3(\mathbf{x}) = \begin{cases} -2, & x_1 \leq \text{median}(x_1) \text{ and } x_2 \leq \text{median}(x_2) \\ -1, & x_1 \leq \text{median}(x_1) \text{ and } x_2 > \text{median}(x_2) \\ 0, & x_1 > \text{median}(x_1) \text{ and } x_2 \leq \text{median}(x_2) \\ 2, & x_1 > \text{median}(x_1) \text{ and } x_2 > \text{median}(x_2). \end{cases} \quad (14)$$

We selected these three exposure-response scenarios to cater to different methods in our simulation study. The linear scenario plays to NPBr and NPB, the nonlinear scenario plays to BKMR, and the fixed profiles scenario plays to UPR and SPR. We hypothesize that the methods to which each scenario caters will perform best in that scenario. We are interested in evaluating how methods perform in exposure-response scenarios for which they were not explicitly developed.

We simulated 200 data sets for each scenario and fit all five Bayesian methods plus LM and LM-int. As results can be sensitive to which pollutants, x_j , $j = 1, \dots, 4$, are included in $h(\mathbf{x})$, we randomly selected pollutants to be the active components in each simulated data set. All pollutants, even those not selected as one of the active components, are included as inputs in the estimated models. By randomly selecting which exposures are the active components of the mixture, each simulated data set has a different correlation structure among the active exposures, which adds robustness to our simulation study results. We calculated the Calinski-Harabasz index [67], the silhouette statistic [68], and the number of clusters to maximize the gap width [69] to measure the grouping structure of the data generated by the fixed profiles scenario. Although the exposure data remains the same for each data set, the exposures used in the exposure-response function differ for each data set; hence the clustering in the fixed profiles scenario, which is based on the response, differs for each data set. Across the 200 data sets used in our simulation study, the median Calinski-Harabasz index was 22.54, the median silhouette was 0.15, and the median number of clusters to maximize the gap width was 6. The distribution of each of these statistics can be found in S2 Table in S1 File. In general, the fixed profiles scenario did not always generate a strong grouping structure with this data, but instead represents a wide variety of clustering schemes.

We evaluated exposure-response function estimation using root mean squared error (RMSE) and interval coverage (Cvg). RMSE was calculated as $\sqrt{\frac{1}{n} \sum_{i=1}^n [h(\mathbf{x}_i) - \hat{h}(\mathbf{x}_i)]^2}$ and coverage was calculated as the percent of $h(\mathbf{x}_i)$'s covered by 95% credible or confidence intervals. RMSE measures the variation between estimated and true values of the exposure-response function. Coverage measures how often the 95% credible or confidence intervals for the estimated exposure-response function capture the true mixture effect. A method with high RMSE and low coverage fails to capture the overall mixture effect. In this way, RMSE and coverage measure the ability of each method to capture the overall mixture effect.

We summarized variable selection through true and false selection rates. In the Bayesian methods, we consider a variable with a PIP above 0.5 as selected into the model [70]. In LM and LM-int, a variable is selected if the 95% confidence interval for the respective regression coefficient does not contain 0. We calculated true selection rate (TSR) as the proportion of mixture components active in the exposure-response function as main effects that were selected into the model as main effects, and false selection rate (FSR) as the proportion of mixture components not in the exposure-response function as main effects that were selected into the model as main effects. All seven exposures are included in the models as inputs, but the active mixture components are those that define the exposure-response function for each

simulated data set. For scenario 1, the active main effects are the randomly selected exposures denoted by x_1 , x_2 , x_3 , and x_4 ; for scenario 2, the active main effects are x_1 , x_2 , and x_3 ; and for scenario 3 the active main effects are x_1 and x_2 . In most methods (NPBr, UPR, SPR, BKMR, and LM), TSR and FSR are calculated only for main effects. In NPBr we can calculate PIPs for interactions and in LM-int we can calculate confidence intervals for the interaction effects. Hence, we also evaluate variable selection rates for interactions in NPBr and LM-int. We calculate true selection rate for interactions (TSR_{int}) as the proportion of the exact pairwise interactions active in the exposure response function that were selected into the model as interactions, and false selection rate for interactions (FSR_{int}) as the proportion of interactions that were not active in the exposure-response function that were selected into the model as interactions. In scenario 1, the true active interactions are $x_1 x_2$ and $x_3 x_4$ and in scenarios 2 and 3 the only active interaction is $x_1 x_2$.

We assessed convergence for a few simulated data sets by visualizing trace plots and comparing results from multiple chains. We found evidence of convergence by 20,000 iterations for all methods. To ensure convergence across all simulated data sets, we based inference on 25,000 samples after a burn-in of 25,000 samples.

We conducted three additional simulation studies to further assess method performance. First, we considered a null scenario, $h_4(\mathbf{x})$, where none of the exposures are associated with the response. Second, we considered a complex mixture scenario, $h_5(\mathbf{x})$, where we simulated data for seven additional pollutants to have a total of 14 mixture components. Third, we applied our original simulation study design to a larger sample of size $n = 1000$ for each data set. Details on the additional simulations can be found in S1 Appendix in [S1 File](#).

Data analysis

We conducted a data analysis on 153 individuals with complete data in the FACES data set. We used regression-adjusted FEV_1 as the outcome. S1 Table in [S1 File](#) summarizes the characteristics of the sample. Pesticide and air pollutant exposures and covariate data were identical to that in our simulation study (Tables 1, 2, and S1 Table in [S1 File](#)). We fit the same models as in the simulation study. Prior specification for the Bayesian models is listed in S2 Appendix in [S1 File](#).

Results

Simulation study results

Simulation results are shown in [Table 3](#). Standard errors are shown in S3–S5 Tables in [S1 File](#). We show the computational time for each method to run for 5000 iterations in [Table 4](#).

Overall BKMR and NPBr were the best performing methods with BKMR performing slightly better in the nonlinear and fixed profiles scenarios. Regarding RMSE for the exposure-response function, BKMR (RMSE = 0.55) and NPBr (RMSE = 0.54) tied for lowest in the linear scenario. In the nonlinear scenario, BKMR (RMSE = 0.59) pulled slightly ahead of NPBr (RMSE = 0.69), while in the fixed profiles scenario, BKMR (RMSE = 0.69) outperformed all other methods by a substantial margin. UPR had the highest RMSE in all three scenarios with SPR having the second highest RMSE.

In addition to having the lowest RMSE in all three scenarios, BKMR consistently had interval coverage closest to the nominal level. LM-int also had interval coverage near the nominal level in all three scenarios and NPBr performed well in the linear scenario. BKMR (Cvg = 0.96), NPBr (Cvg = 0.95), and LM-int (Cvg = 0.95) all achieved the nominal coverage level (0.95) in the linear scenario. In the nonlinear scenario, BKMR (Cvg = 0.92) and LM-int (Cvg = 0.91) came closest to the nominal level, with NPBr next best but trailing behind (Cvg = 0.86). BKMR

Table 3. Summary of method performance in three data-generating scenarios. Table shows means across all data sets for: root mean squared error (RMSE), coverage (Cvg), true selection rate for main effects (TSR), false selection rate for main effects (FSR), true selection rate for interactions (TSR_{int}), and false selection rate for interactions (FSR_{int}). Top-performing methods will have low RSME, coverage near the nominal level (0.95), high TSR and low FSR. For each measure and exposure-response scenario, results from the top-performing method(s) are listed in bold.

Method	RMSE	Cvg	TSR	FSR	TSR _{int}	FSR _{int}
<i>h</i> ₁ (x): linear with multiplicative interactions						
NPBr	1.02	0.73	0.85	0.35	–	–
NPB	0.54	0.95	0.92	0.10	0.59	0.02
UPR	2.01	0.56	0.25	0.26	–	–
SPR	1.59	0.54	0.63	0.53	–	–
BKMR	0.55	0.96	1.00	0.39	–	–
LM	1.01	0.73	0.84	0.29	–	–
LM-int	0.73	0.95	0.68	0.04	0.32	0.04
<i>h</i> ₂ (x): nonlinear with multiplicative interactions						
NPBr	0.77	0.80	0.79	0.22	–	–
NPB	0.69	0.86	0.78	0.16	0.25	0.01
UPR	1.42	0.56	0.27	0.24	–	–
SPR	1.27	0.58	0.68	0.58	–	–
BKMR	0.59	0.92	0.96	0.48	–	–
LM	0.78	0.81	0.78	0.17	–	–
LM-int	0.89	0.91	0.54	0.08	0.20	0.07
<i>h</i> ₃ (x): constant function of fixed profiles						
NPBr	1.11	0.66	0.66	0.11	–	–
NPB	1.02	0.75	0.68	0.13	0.06	0.02
UPR	1.41	0.55	0.27	0.25	–	–
SPR	1.38	0.54	0.68	0.59	–	–
BKMR	0.69	0.91	0.97	0.64	–	–
LM	1.13	0.70	0.69	0.14	–	–
LM-int	0.99	0.91	0.56	0.14	0.12	0.11

<https://doi.org/10.1371/journal.pone.0249236.t003>

(Cvg = 0.91) and LM-int (Cvg = 0.91) had the highest coverage by far in the fixed profiles scenario. Again, UPR and SPR performed poorly with the lowest coverage in all three scenarios.

The story is more complex when it comes to variable selection. While BKMR had the highest TSR in all three scenarios, it also had the highest FSR. Again, NPB performed very well in the linear scenario but not as well in the other scenarios, while UPR and SPR had consistently poor selection rates. Regarding TSR, BKMR (TSR = 1.00) and NPB (TSR = 0.92) performed best in the linear scenario. BKMR had the highest TSR in the nonlinear scenario (TSR = 0.96), where the next best methods, NPBr, NPB, and LM, all had mean TSR just under 0.80. BKMR is singled out with the best TSR in the fixed profiles scenario (TSR = 0.97). UPR, SPR, and LM-int tended to have low TSR in all three scenarios.

Table 4. Computational time for each method to run 5000 iterations on MacBook Pro in R version 3.6.1. Time is reported in seconds. Results reflect 10 evaluations of each method.

method	minimum	mean	maximum
NPBr	6.90	7.03	7.17
NPB	24.73	24.95	25.23
BKMR	219.43	222.96	235.41
UPR	57.82	58.66	59.50
SPR	90.34	92.47	98.65

<https://doi.org/10.1371/journal.pone.0249236.t004>

A low false selection rate indicates a model does not erroneously classify exposures as associated with the outcome when they are not. Here, BKMR had some of the highest FSR across the three scenarios. In the linear scenario, LM-int (FSR = 0.04) and NPB (FSR = 0.10) had the lowest FSR. LM-int also had the lowest FSR in the nonlinear scenario (FSR = 0.08). In the fixed profiles scenario, NPBr, NPB, LM, LM-int all had similar FSR at or below 0.14. Along with BKMR, SPR had high FSR in all three scenarios.

When considering overall variable selection performance, NPB takes the top spot in the linear scenario, with high TSR and low FSR. No method was able to simultaneously achieve dominant TSR and FSR in the nonlinear or fixed profiles scenarios.

Only NPB and LM-int directly parameterized variable selection for interactions in an easily interpretable manner. Interpretable variable selection for interactions is itself an advantage of these approaches over the other methods. In the linear scenario, NPB ($TSR_{int} = 0.59$) had higher TSR_{int} than LM-int ($TSR_{int} = 0.32$). Both methods had poor TSR_{int} in the nonlinear and fixed profiles scenarios, with values at or below 0.25. Regarding FSR_{int} , both methods performed well in all three scenarios, with FSR_{int} consistently at or below 0.11.

The additional simulations produced similar results, with NPB and BKMR being consistently top-performing methods in terms of estimating the exposure-response function and identifying active mixture components. In the null scenario, NPBr and NPB had lowest FSR, meaning these methods were the best at not selecting any mixture components into the model when none were associated with the response (S6 Table in [S1 File](#)). Results from the complex mixture scenario largely mirrored those from the linear scenario (S7 Table in [S1 File](#)). BKMR and NPB remained top-performing in the larger sample size simulation and TSR improved for all methods. Here, UPR and SPR had high TSR and FSR, meaning they often selected all of the mixture components into the model (S8 Table in [S1 File](#)).

Data analysis results

The results from our analysis of the FACES data set varied across the methods. First we consider the traditional models LM and LM-int. LM showed evidence for main effects of NO_2 ($\hat{\beta}$: -0.32, CI: -0.54, -0.10) and PM_{10} ($\hat{\beta}$: 0.19, CI: 0.02, 0.35). LM-int showed evidence for main effects of MeBr ($\hat{\beta}$: 0.17, CI: 0.05, 0.29), NO_2 ($\hat{\beta}$: -0.68, CI: -1.10, -0.25), and PM_{10} ($\hat{\beta}$: 0.50, CI: 0.08, 0.93) and an interaction between C and $PM_{2.5}$ ($\hat{\beta}$: 0.28, CI: 0.01, 0.54) ([Table 5](#)). The

Table 5. Results from analysis of FACES data set using LM and LM-int. Table includes effect estimates ($\hat{\beta}$), 95% confidence intervals, and associated p -values for all main effects in LM and LM-int plus the interaction effects in LM-int with p -values ≤ 0.10 . The regression coefficient $\hat{\beta}$ is the expected change in FEV_1 for a 1 standard deviation increase in the square root transformed exposures.

	LM			LM-int		
	$\hat{\beta}$	95% CI	p -value	$\hat{\beta}$	95% CI	p -value
C	0.04	(-0.03, 0.11)	0.24	0.05	(-0.08, 0.19)	0.44
MeBr	0.00	(-0.06, 0.07)	0.96	0.17	(0.05, 0.29)	0.01
OP	0.05	(-0.03, 0.13)	0.24	0.02	(-0.17, 0.22)	0.80
O_3	-0.06	(-0.20, 0.07)	0.36	-0.13	(-0.32, 0.06)	0.17
NO_2	-0.32	(-0.54, -0.10)	0.01	-0.68	(-1.10, -0.25)	0.00
$PM_{2.5}$	-0.01	(-0.20, 0.17)	0.90	-0.11	(-0.48, 0.26)	0.55
PM_{10}	0.19	(0.02, 0.35)	0.03	0.50	(0.08, 0.93)	0.02
C: $PM_{2.5}$	-	-	-	0.28	(0.01, 0.54)	0.04
OP: PM_{10}	-	-	-	0.31	(-0.01, 0.62)	0.05
NO_2 : PM_{10}	-	-	-	0.33	(-0.05, 0.72)	0.09

<https://doi.org/10.1371/journal.pone.0249236.t005>

Table 6. Results from analysis of FACES data set using NPBr and NPB. Table shows estimates ($\hat{\beta}$), 95% credible intervals, and posterior inclusion probabilities (PIP) for main effect exposures in NPB and NPBr. The regression coefficient $\hat{\beta}$ is the expected change in FEV₁ for a 1 standard deviation increase in the square root transformed exposures. All interaction effects in NPB had posterior inclusion probabilities below 0.12.

	NPBr			NPB		
	$\hat{\beta}$	95% CI	PIP	$\hat{\beta}$	95% CI	PIP
C	0.00	(0.00, 0.04)	0.07	0.00	(0.00, 0.03)	0.07
MeBr	0.00	(-0.02, 0.00)	0.06	0.00	(-0.01, 0.00)	0.06
OP	0.02	(0.00, 0.12)	0.21	0.01	(0.00, 0.11)	0.16
O ₃	0.00	(-0.08, 0.02)	0.11	-0.01	(-0.12, 0.01)	0.11
NO ₂	-0.08	(-0.35, 0.00)	0.47	-0.12	(-0.36, 0.00)	0.60
PM _{2.5}	0.00	(-0.08, 0.06)	0.13	0.00	(-0.09, 0.05)	0.12
PM ₁₀	0.02	(0.00, 0.21)	0.21	0.02	(-0.01, 0.20)	0.19

<https://doi.org/10.1371/journal.pone.0249236.t006>

results from the linear models indicating a protective effect of PM₁₀ are counter-intuitive as there is an extensive literature on the deleterious health effects of PM on lung function. None of the other methods found evidence of protective effects for any of the exposures.

Next we consider the five contemporary methods. NPBr did not identify any exposures with PIPs above 0.5. The exposure with the highest PIP was NO₂ (PIP = 0.47), which was estimated to be negatively associated with FEV₁ ($\hat{\beta}$: -.08, CI: -0.35, 0.00). In NPB, NO₂ was selected (PIP = 0.60) and was also negatively associated with FEV₁ ($\hat{\beta}$: -0.12, CI: -0.36, 0.00) (Table 6). No other main effects or interactions were selected by either method (S10 Table in S1 File).

In BKMR, NO₂ was selected as an important mixture component with a PIP of 0.96 (S11 Table in S1 File). No other exposures had PIPs above 0.5. Results were similar using the HVS formulation (S12 Table in S1 File). NO₂ had a negative and nonlinear association with FEV₁ (Fig 1). To identify interactions, we plot the posterior distribution of the exposure-response function for each pair of exposures, holding all other exposures constant at their median values, and visually inspect changes in the response as both exposures change. In doing so we found no notable interactions among exposures (S1 Fig in S1 File).

As clustering algorithms, UPR and SPR reveal a different kind of story. UPR revealed four clusters as the best partitioning of the data. Each cluster had similar estimated health effects (Fig 2a); hence, despite partitioning the exposure space there was no meaningful association between the exposure profiles and the health outcome. Fig 2b–2e shows the empirical exposure means for individuals assigned to each cluster. The first cluster of $n = 25$ individuals was distinguished by higher than average exposure to MeBr. Cluster 2 ($n = 33$) had low exposure to OP and O₃ and high exposure to NO₂ and PM_{2.5} relative to the average. The third cluster ($n = 9$) was characterized by relatively high exposure to OP and low exposure to O₃. Individuals in cluster 4 ($n = 86$) had nearly average exposure to most pollutants except MeBr, which was notably below average; in addition, O₃ exposure was slightly above and PM_{2.5} exposure was slightly below average. UPR selected OP (PIP = 0.57), O₃ (PIP = 0.54), NO₂ (PIP = 0.61), and PM_{2.5} (PIP = 0.56) as important mixture components (S13 Table in S1 File).

SPR also revealed four clusters as the best partitioning of the data. The estimated exposure-response function for cluster 3, the smallest cluster ($n = 9$), had a 0.97 posterior probability of being greater than the overall mean estimated exposure-response function (Fig 3a). The cluster sample sizes and associated empirical exposure means were very similar to those in UPR (Fig 3b–3e), with the labels switched for clusters 1 and 4. In both UPR and SPR, cluster 3 was the smallest cluster and had an estimated mean health effect higher than average, but there was more uncertainty around the health effect in UPR likely due to the two-stage approach for

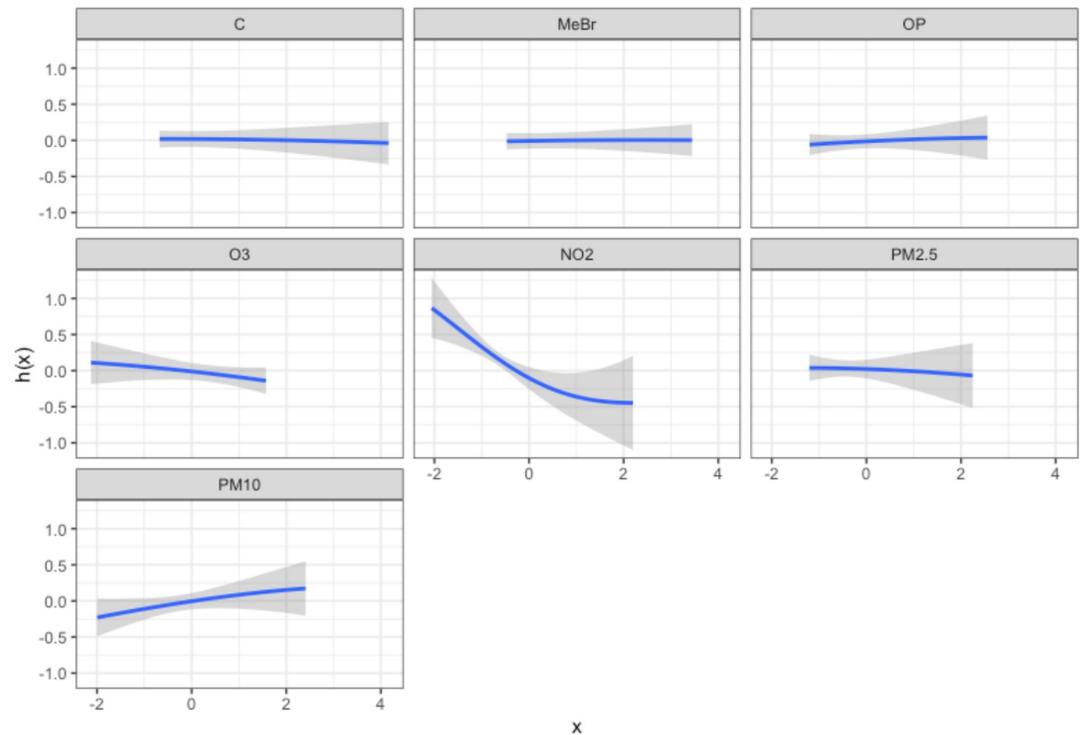


Fig 1. Results from analysis of FACES data set using BKMR. Figure shows the univariate relationship between each exposure and FEV₁, holding all other exposures at their median value.

<https://doi.org/10.1371/journal.pone.0249236.g001>

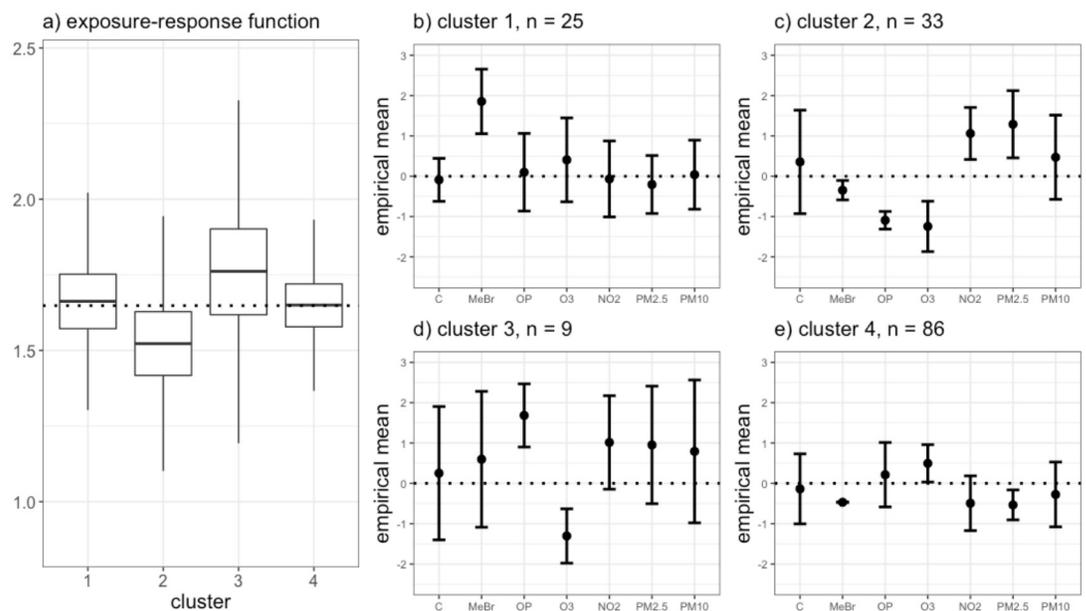


Fig 2. Results from analysis of FACES data set using UPR. Panel (a) shows the distribution of the model averaged estimated exposure-response function (θ_c) for each cluster identified in the best clustering by UPR. The dotted line represents the overall mean estimated exposure-response function across all clusters. Panels (b-e) show the empirical exposure means of the individuals assigned to each cluster in the best clustering, with 1 standard deviation error bars. The dotted lines are drawn at 0, the mean of the standardized exposure data.

<https://doi.org/10.1371/journal.pone.0249236.g002>

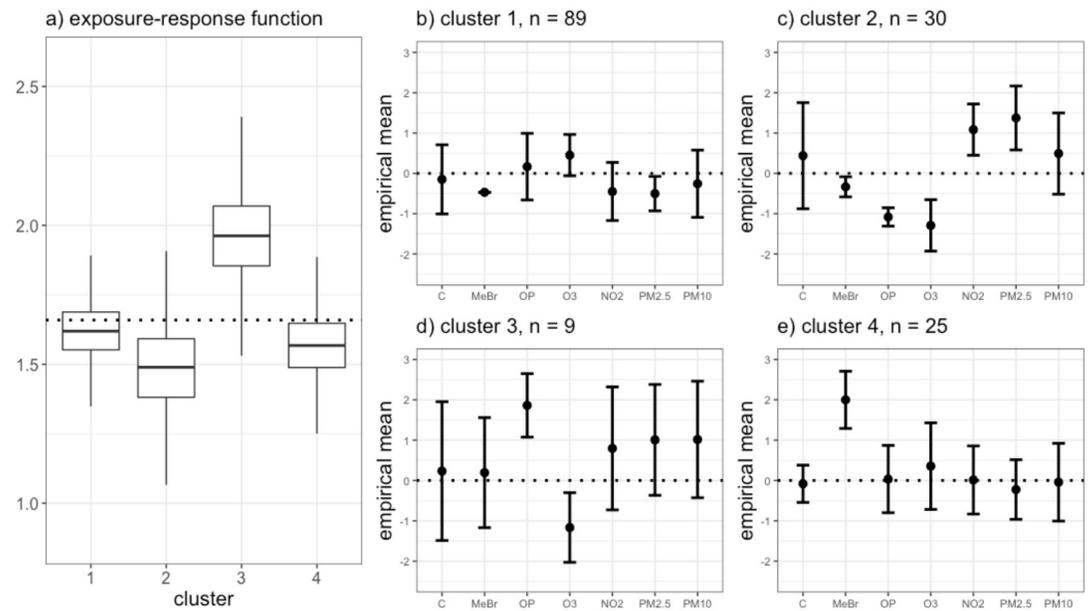


Fig 3. Results from analysis of FACES data set using SPR. Panel (a) shows the distribution of the model averaged estimated exposure-response function (θ_i) for each cluster identified in the best clustering by SPR. The dotted line represents the overall mean estimated exposure-response function across all clusters. Panels (b-e) show the empirical exposure means of the individuals assigned to each cluster in the best clustering, with 1 standard deviation error bars. The dotted lines are drawn at 0, the mean of the standardized exposure data.

<https://doi.org/10.1371/journal.pone.0249236.g003>

estimation. SPR selected five important mixture components: MeBr (PIP = 0.71), OP (PIP = 0.51), O₃ (PIP = 0.75), NO₂ (PIP = 0.67), and PM_{2.5} (PIP = 0.63) (S13 Table in S1 File). We found the clustering and PIPs in UPR and SPR to be sensitive to prior choice particularly for the cluster-specific precision matrix and error precision.

Discussion

Interest is rapidly growing in estimating the association between exposure to mixtures of environmental chemicals and health outcomes. As a result, new statistical approaches have been developed for studying health outcomes associated with exposure to mixtures. The purpose of this paper was to evaluate and compare recently developed methods for mixtures and determine which research questions they answer well and in which scenarios. We limited our study to contemporary Bayesian methods since they are under-studied, under-utilized, and may have the ability to answer multiple research questions. Our results highlight the advantages of the flexible modeling and Bayesian framework of BKMR and NPB in estimating the exposure-response function precisely and identifying mixture components most strongly associated with the health outcome.

Overall, BKMR was a top-performing method. In each of the scenarios, BKMR estimated the exposure-response function with coverage closest to the nominal level (0.95) and lowest RMSE. Despite being a more flexible approach based on Gaussian processes, BKMR had lower RMSE in the linear scenario than NPBr, LM, and LM-int, all of which assume linearity. This is likely because NPBr and LM do not account for interactions and LM-int can result in inflated standard errors in the presence of correlated data. BKMR identified active mixture components with the greatest frequency, but also included inactive components more often than other methods. Although we did not evaluate variable selection rates for interactions in BKMR

in our simulation, BKMR can identify linear or nonlinear interactions among exposures through visualization or summarizing the posterior distribution of the exposure-response function. A drawback to BKMR is that results are not as easily interpreted as in NPB or the linear models, though there are currently efforts to enhance interpretation and a suite of visualization approaches that aid in interpretation. BKMR is an appealing choice for mixtures because it makes minimal assumptions on the shape of the exposure-response function and includes a sophisticated variable selection algorithm for identifying important mixture components.

NPB was top-performing in the linear scenario regarding estimating the exposure-response function, identifying both active and inactive mixture components, and identifying interactions. NPB performed well even when the exposure-response function was mildly nonlinear, but lacks the flexibility of BKMR for the fixed profiles scenario, which is highly nonlinear. The AME method NPB poorly estimated the exposure-response function in the linear scenario, likely from not accounting for interactions. An advantage of NPB is its ease of interpretation, which is similar to interpreting a linear regression model. NPB estimates PIPs and effect sizes for all main effect and interaction terms, providing precise information regarding the contribution of each exposure to the mixture and its effect on the health outcome.

The profile regression methods, UPR and SPR, poorly addressed the research questions of interest in all three scenarios. Two explanations for this include lack of a clustering structure in the exposure data and a weak signal, both of which inhibit these methods from accurately estimating the multipollutant exposure-response function. Further, UPR and SPR do not have the ability to identify or estimate interactions or tease out individual effects of the pollutants within a mixture. These methods may not be appropriate for studies in which the primary objectives are to estimate the multipollutant exposure-response function and identify driving mixture components. As clustering methods, UPR and SPR are likely to perform better on data that has a strong grouping structure. Since we used a single data set in our simulation study, the results of our simulation should not be interpreted as representative of performance on all data structures. A particular advantage of UPR and SPR is that the number of clusters need not be pre-specified.

The linear model with interactions, LM-int, had coverage above 0.91 in all three scenarios, but had higher RMSE and lower TSR than BKMR and NPB. LM-int and NPB are both EMM methods, and NPB outperformed LM-int in the linear EMM scenario. LM and LM-int have the advantage of being easy to implement and interpret, but these methods estimated the exposure-response function with more uncertainty than the top-performing methods and generally lacked the ability to select truly active mixture components, likely due to high correlation among exposures.

In our application to the FACES data set, four methods (LM, LM-int, NPB, and BKMR) identified NO₂ as an important mixture component negatively associated with the health outcome. In addition, LM and LM-int estimated PM₁₀ to have a positive association with FEV₁, and PM₁₀ was positively correlated with NO₂. Further, the magnitude of the effect estimate for NO₂ in LM and LM-int was several times larger than that estimated in NPB, and the confidence intervals were also larger, reflecting more uncertainty. UPR and SPR also identified NO₂ as an important mixture component, but we cannot determine the sign of effect using these methods. Instead, UPR and SPR have the ability to estimate how the overall mixture is associated with the health outcome. UPR revealed four clusters with similar estimated health effects; hence, patterns in the exposure data were not associated with FEV₁. In SPR, the smallest cluster was associated with higher average FEV₁ than the other clusters, suggesting an association between a relatively rare mixture of exposures and the health outcome. Alternatively, this small cluster may reflect a strong influence from the health outcome in the clustering

using a supervised learner. Meanwhile, BKMR was able to describe a nonlinear association between NO_2 and FEV_1 .

Using missing indicators may have introduced some bias in the effect estimates. Additionally, all Bayesian methods are sensitive to prior specification and results may vary with more or less informative priors. PIPs are particularly sensitive to prior specification in all methods, so changing prior hyperparameters may lead to changes in TSR and FSR. We implemented all models using the default priors as specified by the authors to obtain an objective comparison of these methods.

Along with the primary research question, the best performing method is likely to depend on the exposure data. We used observed exposure data so our results are highly relevant to realistic settings. Our simulation results can be generalized to small data sets with a limited number of localized exposures, which is a frequent scenario in epidemiological studies.

In analyses of environmental mixtures and human health, model choice depends on the assumed exposure-response relationship and the primary questions of interest. NPB and BKMR are recently proposed methods that outperformed traditional regression models and offer promising tools for mixtures analyses. We recommend NPB when the exposure-response function is assumed to be approximately linear and a primary goal is accurately identifying which are the active and inactive components of the mixture. NPB is also highly interpretable and explicitly tests for interactions. We recommend BKMR if the exposure-response function is assumed to take on a complex form and the primary goal is estimating the form of the exposure-response function while at the same time identifying important mixture components. Our results suggest that UPR and SPR do not reliably answer our specified research questions, but may be applicable for different research questions such as pattern recognition. We further encourage users to take advantage of our R package `mmpack` [44] to replicate the simulation and determine how each method performs on their own data. Results will likely be different on different data sets. In particular, the profile regression methods may perform better on data that exhibits a stronger clustering structure in the fixed profiles scenario. We include the clustering statistics as part of the summary of the fixed profile scenario output so users can see how much grouping structure is in their own data. Replicating the simulation on their own data will enable users to choose the best method for their data and specific research question.

Supporting information

S1 File.
(PDF)

Author Contributions

Data curation: Wande Benka-Coker, Rachel Severson, Sheryl Magzamen.

Formal analysis: Lauren Hoskovec.

Funding acquisition: Sheryl Magzamen.

Methodology: Lauren Hoskovec, Sheryl Magzamen, Ander Wilson.

Resources: Ander Wilson.

Software: Lauren Hoskovec.

Supervision: Sheryl Magzamen, Ander Wilson.

Visualization: Lauren Hoskovec.

Writing – original draft: Lauren Hoskovec.

Writing – review & editing: Wande Benka-Coker, Sheryl Magzamen, Ander Wilson.

References

1. Dockery DW, Pope CA, Xu X, Spengler JD, Ware JH, Fay ME, et al. An Association between Air Pollution and Mortality in Six U.S. Cities. *New England Journal of Medicine*. 1993; 329(24):1753–1759. <https://doi.org/10.1056/NEJM199312093292401>
2. Dockery DW, Pope CA. Acute Respiratory Effects of Particulate Air Pollution. *Annual Review of Public Health*. 1994; 15(1):107–132. <https://doi.org/10.1146/annurev.pu.15.050194.000543>
3. Di Q, Dai L, Wang Y, Zanobetti A, Choirat C, Schwartz JD, et al. Association of Short-term Exposure to Air Pollution With Mortality in Older Adults. *JAMA*. 2017; 318(24):2446–2456. <https://doi.org/10.1001/jama.2017.17923>
4. Di Q, Wang Y, Zanobetti A, Wang Y, Koutrakis P, Choirat C, et al. Air Pollution and Mortality in the Medicare Population. *New England Journal of Medicine*. 2017; 376(26):2513–2522. <https://doi.org/10.1056/NEJMoa1702747>
5. Pan L, Wu S, Li H, Xu J, Dong W, Shan J, et al. The short-term effects of indoor size-fractioned particulate matter and black carbon on cardiac autonomic function in COPD patients. *Environment International*. 2018; 112(38):261–268. <https://doi.org/10.1016/j.envint.2017.12.037>
6. Slama R, Vrijheid M. Some challenges of studies aiming to relate the Exposome to human health. *Occupational and Environmental Medicine*. 2015; 72(6):383–384. <https://doi.org/10.1136/oemed-2014-102546>
7. Weisskopf MG, Seals RM, Webster TF. Bias Amplification in Epidemiologic Analysis of Exposure to Mixtures. *Environmental Health Perspectives*. 2018; 126(4):047003. <https://doi.org/10.1289/EHP2450>
8. Samet JM. The perspective of the National Research Council's Committee on Research Priorities for Airborne Particulate Matter. *Journal of Toxicology and Environmental Health—Part A*. 2005; 68(13–14):1063–1067. <https://doi.org/10.1080/15287390590935905> PMID: 16024488
9. Dominici F, Peng RD, Barr CD, Bell ML. Protecting Human Health From Air Pollution. *Epidemiology*. 2010; 21(2):187–194. <https://doi.org/10.1097/EDE.0b013e3181cc86e8>
10. NIEHS. 2012-2017 Strategic Plan: Advancing Science, Improving Health: A plan for environmental health research; 2012.
11. Taylor KW, Joubert BR, Braun JM, Dilworth C, Gennings C, Hauser R, et al. Statistical Approaches for Assessing Health Effects of Environmental Chemical Mixtures in Epidemiology: Lessons from an Innovative Workshop. *Environmental Health Perspectives*. 2016; 124(12):A227–A229. <https://doi.org/10.1289/EHP547>
12. Hamra GB, Buckley JP. Environmental Exposure Mixtures: Questions and Methods to Address Them. *Current Epidemiology Reports*. 2018; 5:160–165. <https://doi.org/10.1007/s40471-018-0145-0>
13. Witte JS, Greenland S. Simulation Study of Hierarchical Regression. *Statistics in Medicine*. 1996; 15(11):1161–1170. [https://doi.org/10.1002/\(SICI\)1097-0258\(19960615\)15:11%3C1161::AID-SIM221%3E3.0.CO;2-7](https://doi.org/10.1002/(SICI)1097-0258(19960615)15:11%3C1161::AID-SIM221%3E3.0.CO;2-7)
14. Barrera-Gómez J, Agier L, Portengen L, Chadeau-Hyam M, Giorgis-Allemand L, Siroux V, et al. A systematic comparison of statistical methods to detect interactions in exposome-health associations. *Environmental Health: A Global Access Science Source*. 2017; 16(74). <https://doi.org/10.1186/s12940-017-0277-6>
15. Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: A tutorial. *Statistical Science*. 1999; 14(4):382–401. <https://doi.org/10.2307/2676803>
16. Clyde M. Model uncertainty and health effect studies for particulate matter. *Environmetrics*. 2000; 11(6):745–763. [https://doi.org/10.1002/1099-095X\(200011/12\)11:6%3C745::AID-ENV431%3E3.0.CO;2-N](https://doi.org/10.1002/1099-095X(200011/12)11:6%3C745::AID-ENV431%3E3.0.CO;2-N)
17. Davalos AD, Luben TJ, Herring AH, Sacks JD. Current approaches used in epidemiologic studies to examine short-term multipollutant air pollution exposures. *Annals of Epidemiology*. 2017; 27(2):145–153. <https://doi.org/10.1016/j.annepidem.2016.11.016>
18. Roberts S, Martin M. A critical assessment of shrinkage-based regression approaches for estimating the adverse health effects of multiple air pollutants. *Atmospheric Environment*. 2005; 39:6223–6230. <https://doi.org/10.1016/j.atmosenv.2005.07.004>
19. Herring AH. Nonparametric bayes shrinkage for assessing exposures to mixtures subject to limits of detection. *Epidemiology*. 2010; 21:S71–S76. <https://doi.org/10.1097/EDE.0b013e3181cf0058>

20. Carbajal-Arroyo L, Miranda-Soberanis V, Medina-Ramón M, Rojas-Bracho L, Tzintzun G, Solís-Gutiérrez P, et al. Effect of PM10 and O3 on infant mortality among residents in the Mexico City Metropolitan Area: A case-crossover analysis, 1997-2005. *Journal of Epidemiology and Community Health*. 2011; 65(8):715–721. <https://doi.org/10.1136/jech.2009.101212>
21. Winquist A, Kirrane E, Klein M, Strickland M, Darrow LA, Sarnat SE, et al. Joint effects of ambient air pollutants on pediatric asthma emergency department visits in atlanta, 1998-2004. *Epidemiology*. 2014; 25(5):666–673. <https://doi.org/10.1097/EDE.0000000000000146>
22. Lenters V, Portengen L, Rignell-Hydbom A, Jönsson BAG, Lindh CH, Piersma AH, et al. Prenatal Phthalate, Perfluoroalkyl Acid, and Organochlorine Exposures and Term Birth Weight in Three Birth Cohorts: Multi-Pollutant Models Based on Elastic Net Regression. *Environmental Health Perspectives*. 2016; 124(3):365–372. <https://doi.org/10.1289/ehp.1408933>
23. Austin E, Coull B, Thomas D, Koutrakis P. A framework for identifying distinct multipollutant profiles in air pollution data. *Environment International*. 2012; 45:112–121. <https://doi.org/10.1016/j.envint.2012.04.003>
24. Zanobetti A, Austin E, Coull BA, Schwartz J, Koutrakis P. Health effects of multi-pollutant profiles. *Environment International*. 2014; 71:13–19. <https://doi.org/10.1016/j.envint.2014.05.023>
25. Sacks JD, Ito K, Wilson WE, Neas LM. Impact of covariate models on the assessment of the air pollution-mortality association in a single-and multipollutant context. *American Journal of Epidemiology*. 2012; 176(7):622–634. <https://doi.org/10.1093/aje/kws135>
26. Pearce JL, Waller LA, Chang HH, Klein M, Mulholland JA, Sarnat JA, et al. Using self-organizing maps to develop ambient air quality classifications: a time series example. *Environmental Health: A Global Access Science Source*. 2014; 13(1):56. <https://doi.org/10.1186/1476-069X-13-56>
27. Pearce JL, Waller LA, Mulholland JA, Sarnat SE, Strickland MJ, Chang HH, et al. Exploring associations between multipollutant day types and asthma morbidity: Epidemiologic applications of self-organizing map ambient air quality classifications. *Environmental Health: A Global Access Science Source*. 2015; 14(55). <https://doi.org/10.1186/s12940-015-0041-8>
28. Pearce JL, Waller LA, Sarnat SE, Chang HH, Klein M, Mulholland JA, et al. Characterizing the spatial distribution of multiple pollutants and populations at risk in Atlanta, Georgia. *Spatial and Spatio-temporal Epidemiology*. 2016; 18:13–23. <https://doi.org/10.1016/j.sste.2016.02.002>
29. Roberts S, Martin MA. Using Supervised Principal Components Analysis to Assess Multiple Pollutant Effects. *Environmental Health Perspectives*. 2006; 114(12):1877–1882. <https://doi.org/10.1289/ehp.9226>
30. Wold S, Ruhe A, Wold H, Dunn WJ III. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *Society for Industrial and Applied Mathematics*. 1984; 5(3):735–743.
31. Roberts S, Martin M. Investigating the mixture of air pollutants associated with adverse health outcomes. *Atmospheric Environment*. 2006; 40(5):984–991. <https://doi.org/10.1016/j.atmosenv.2005.10.022>
32. Nikolov MC, Coull BA, Catalano PJ, Godleski JJ. An informative Bayesian structural equation model to assess source-specific health effects of air pollution. *Biostatistics*. 2007; 8(3):609–624. <https://doi.org/10.1093/biostatistics/kxl032>
33. Pachon JE, Balachandran S, Hu Y, Mulholland JA, Darrow LA, Sarnat JA, et al. Development of outcome-based, multipollutant mobile source indicators. *Journal of the Air and Waste Management Association*. 2012; 62(4):431–442. <https://doi.org/10.1080/10473289.2012.656218>
34. Sun Z, Tao Y, Li S, Ferguson KK, Meeker JD, Park SK, et al. Statistical strategies for constructing health risk models with multiple pollutants and their interactions: possible choices and comparisons. *Environmental Health: A Global Access Science Source*. 2013; 12(1):85. <https://doi.org/10.1186/1476-069X-12-85>
35. Carrico C, Gennings C, Wheeler DC, Factor-Litvak P. Characterization of Weighted Quantile Sum Regression for Highly Correlated Data in a Risk Analysis Setting. *Journal of Agricultural, Biological, and Environmental Statistics*. 2015; 20(1):100–120. <https://doi.org/10.1007/s13253-014-0180-3>
36. Bobb JF, Valeri L, Claus Henn B, Christiani DC, Wright RO, Mazumdar M, et al. Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. *Biostatistics*. 2015; 16(3):493–508. <https://doi.org/10.1093/biostatistics/kxu058>
37. Gass K, Klein M, Chang HH, Dana Flanders W, Strickland MJ. Classification and regression trees for epidemiologic research: An air pollution example. *Environmental Health: A Global Access Science Source*. 2014; 13(1):17. <https://doi.org/10.1186/1476-069X-13-17>
38. Gass K, Klein M, Sarnat SE, Winquist A, Darrow LA, Flanders WD, et al. Associations between ambient air pollutant mixtures and pediatric asthma emergency department visits in three cities: a classification

- and regression tree approach. *Environmental Health: A Global Access Science Source*. 2015; 14(1):58. <https://doi.org/10.1186/s12940-015-0044-5>
39. Braun JM, Gennings C, Hauser R, Webster TF. What can epidemiological studies tell us about the impact of chemical mixtures on human health? *Environmental Health Perspectives*. 2016; 124(1):A6–A9. <https://doi.org/10.1289/ehp.1510569>
 40. Agier L, Portengen L, Chadeau-hyam M, Basagaña X, Giorgis-allemand L, Siroux V, et al. A Systematic Comparison of Linear Regression—Based Statistical Methods to Assess Exposome–Health Associations. *Environmental Health Perspectives*. 2016; 124(12):1848–1856. <https://doi.org/10.1289/EHP172>
 41. Chiu YH, Bellavia A, James-Todd T, Correia KF, Valeri L, Messerlian C, et al. Evaluating effects of prenatal exposure to phthalate mixtures on birth weight: A comparison of three statistical approaches. *Environment International*. 2018; 113(November 2017):231–239. <https://doi.org/10.1016/j.envint.2018.02.005>
 42. Antonelli J, Mazumdar M, Bellinger D, Christiani DC, Wright R, Coull BA. Estimating the health effects of environmental mixtures using Bayesian semiparametric regression and sparsity inducing priors. *Annals of Applied Statistics*. 2020; 14(1):257–275. <https://doi.org/10.1214/19-AOAS1307>
 43. Gibson EA, Goldsmith J, Kioumourtzoglou MA. Complex Mixtures, Complex Analyses: an Emphasis on Interpretable Results. *Current Environmental Health Reports*. 2019; 6(2):53–61. <https://doi.org/10.1007/s40572-019-00229-5>
 44. Hoskovec L. mmpack: Implement methods for multipollutant mixtures analyses. R package version 0.1.0.; 2019. Available from: <https://github.com/lvhoskovec/mmpack>.
 45. Mortimer K, Neugebauer R, Lurmann F, Alcorn S, Balmes J, Tager I. Early-lifetime exposure to air pollution and allergic sensitization in children with asthma. *Journal of Asthma*. 2008; 45(10):874–881. <https://doi.org/10.1080/02770900802195722>
 46. Margolis HG, Mann JK, Lurmann FW, Mortimer KM, Balmes JR, Hammond SK, et al. Altered pulmonary function in children with asthma associated with highway traffic near residence. *International Journal of Environmental Health Research*. 2009; 19(2):139–155. <https://doi.org/10.1080/09603120802415792>
 47. Mann JK, Balmes JR, Bruckner TA, Mortimer KM, Margolis HG, Pratt B, et al. Short-term effects of air pollution on wheeze in asthmatic children in Fresno, California. *Environmental Health Perspectives*. 2010; 118(10):1497–1502. <https://doi.org/10.1289/ehp.0901292>
 48. Noth EM, Hammond SK, Biging GS, Tager IB. A spatial-temporal regression model to predict daily outdoor residential PAH concentrations in an epidemiologic study in Fresno, CA. *Atmospheric Environment*. 2011; 45(14):2394–2403. <https://doi.org/10.1016/j.atmosenv.2011.02.014>
 49. Gale SL, Noth EM, Mann J, Balmes J, Hammond SK, Tager IB. Polycyclic aromatic hydrocarbon exposure and wheeze in a cohort of children with asthma in Fresno, CA. *Journal of Exposure Science and Environmental Epidemiology*. 2012; 22(4):386–392. <https://doi.org/10.1038/jes.2012.29>
 50. Padula AM, Balmes JR, Eisen EA, Mann J, Noth EM, Lurmann FW, et al. Ambient polycyclic aromatic hydrocarbons and pulmonary function in children. *Journal of Exposure Science and Environmental Epidemiology*. 2015; 25(3):295–302. <https://doi.org/10.1038/jes.2014.42>
 51. Van Sickle D, Magzamen S, Mullahy J. Understanding socioeconomic and racial differences in adult lung function. *American Journal of Respiratory and Critical Care Medicine*. 2011; 184(5):521–527. <https://doi.org/10.1164/rccm.201012-2095OC>
 52. Raanan R, Balmes JR, Harley KG, Gunier RB, Magzamen S, Bradman A, et al. Decreased lung function in 7-year-old children with early-life organophosphate exposure. *Thorax*. 2016; 71(2):148–153. <https://doi.org/10.1136/thoraxjnl-2014-206622>
 53. Benka-Coker W, Hoskovec L, Severson R, Balmes J, Wilson A, Magzamen S. The joint effect of ambient air pollution and agricultural pesticide exposures on lung function among children with asthma. *Environmental Research*. 2020; 190(February):109903. <https://doi.org/10.1016/j.envres.2020.109903>
 54. California Department of Pesticide Regulation. California Pesticide Use Reporting Data; 2015. Available from: <http://www.cdpr.ca.gov/docs/pur/purmain.htm>.
 55. Colovic MB, Krsti DZ, Lazarevi-Pati TD, Bondi AM, Vasi VM. Acetylcholinesterase Inhibitors: Pharmacology and Toxicology. *Current Neuropharmacology*. 2013; 11:315–335. <https://doi.org/10.2174/1570159X11311030006>
 56. Bulathsinghala AT, Shaw IC. The toxic chemistry of methyl bromide. *Human and Experimental Toxicology*. 2014; 33(1):81–91. <https://doi.org/10.1177/0960327113493299>
 57. Severson R. purexposure: Pull and Calculate Exposure to CA Pesticide Use Registry Records. R package version 0.1.0.; 2019. Available from: <https://github.com/leighseverson/purexposure>.
 58. Molitor J, Papathomas M, Jerrett M, Richardson S. Bayesian profile regression with an application to the National Survey of Children’s Health. *Biostatistics*. 2010; 11(3):484–498. <https://doi.org/10.1093/biostatistics/kxq013>

59. R Core Team. R: A language and environment for statistical computing; 2018.
60. Dunson DB, Herring AH, Engel SM. Bayesian Selection and Clustering of Polymorphisms in Functionally Related Genes. *Journal of the American Statistical Association*. 2008; 103(482):534–546. <https://doi.org/10.1198/016214507000000554>
61. Molitor J, Su JG, Molitor NT, Rubio VG, Richardson S, Hastie D, et al. Identifying vulnerable populations through an examination of the association between multipollutant profiles and poverty. *Environmental Science and Technology*. 2011; 45(18):7754–7760. <https://doi.org/10.1021/es104017x>
62. Dahl DB. Model-Based Clustering for Expression Data via a Dirichlet Process Mixture Model. *Bayesian Inference for Gene Expression and Proteomics*. 2006; p. 201–218. <https://doi.org/10.1017/CBO9780511584589.011>
63. Chung Y, Dunson DB. Nonparametric Bayes Conditional Distribution Modeling With Variable Selection. *Journal of the American Statistical Association*. 2009; 104(488):1646–1660. <https://doi.org/10.1198/jasa.2009.tm08302> PMID: 23580793
64. Papathomas M, Molitor J, Hoggart C, Hastie D, Richardson S. Exploring Data From Genetic Association Studies Using Bayesian Variable Selection and the Dirichlet Process: Application to Searching for Gene \times Gene Patterns. *Genetic Epidemiology*. 2012; 36(6):663–674. <https://doi.org/10.1002/gepi.21661>
65. Liverani S, Hastie DI, Azizi L, Papathomas M, Richardson S. PReMiuM: An R Package for Profile Regression Mixture Models Using Dirichlet Processes. *Journal of Statistical Software*. 2015; 64(7):1–30. <https://doi.org/10.18637/jss.v064.i07>
66. Bobb JF. bkmr: Bayesian Kernel Machine Regression. R package version 0.2.0; 2017. Available from: <https://cran.r-project.org/package=bkmr>.
67. Caliński T, Harabasz J. Communications in Statistics—Theory and Methods. *Communications in Statistics*. 1974; 3(1):1–27.
68. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*. 1987; 20(C):53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
69. Tibshirani R, Walther G, Hastie T. Estimating the number of data clusters via the gap statistic; 2001.
70. Barbieri MM, Berger JO. Optimal predictive model selection. *Annals of Statistics*. 2004; 32(3):870–897. <https://doi.org/10.1214/009053604000000238>