Assessing Disparities in Americans' Exposure to PCBs and PBDEs based on NHANES Pooled Biomonitoring Data

Yan Liu, Dewei Wang, Li Li & Dingsheng Li

Taylor & Francis
Taylor & Francis Group

Check for updates

Assessing Disparities in Americans' Exposure to PCBs and PBDEs based on NHANES Pooled Biomonitoring Data

Yan Liu[a] , Dewei Wang[b] , Li Li[a] , and Dingsheng Li[a]

[a]School of Public Health, University of Nevada, Reno, NV; [b]Department of Statistics, University of South Carolina, Columbia, SC

ABSTRACT

The National Health and Nutrition Examination Survey (NHANES) has been continuously biomonitoring Americans' exposure to two families of harmful environmental chemicals: polychlorinated biphenyls (PCBs) and polybrominated diphenyl ethers (PBDEs). However, biomonitoring these chemicals is expensive. To save cost, in 2005, NHANES resorted to pooled biomonitoring; that is, amalgamating individual specimens to form a pool and measuring chemical levels from pools. Despite being publicly available, these pooled data gain limited applications in health studies. Among the few studies using these data, racial/age disparities were detected, but there is no control for confounding effects. These disadvantages are due to the complexity of pooled measurements and a dearth of statistical tools. Herein, we developed a regression-based method to unzip pooled measurements, which facilitated a comprehensive assessment of disparities in exposure to these chemicals. We found increasing dependence of PCBs on age and income, whereas PBDEs were the highest among adolescents and seniors and were elevated among the low-income population. In addition, Hispanics had the lowest PCBs and PBDEs among all demographic groups after controlling for potential confounders. These findings can guide the development of population-specific interventions to promote environmental justice. Moreover, both chemical levels declined throughout the period, indicating the effectiveness of existing regulatory policies. Supplementary materials for this article are available online.

1. Introduction

The general population is exposed to a wide range of persistent organic pollutants. Notorious examples include two families of organochlorine contaminants called polychlorinated biphenyls (PCBs) and polybrominated diphenyl ethers (PBDEs). In history, PCBs were widely used as insulation fluids in capacitors, transformers, and other electrical and electronic equipment (Breivik et al. 2007), whereas PBDEs were widely used in indoor commodities such as furniture, electronics, and plastics for their fire-resistant properties (Abbasi, Li, and Breivik 2019). Once released into the environment, PCBs and PBDEs can accumulate in the human body through different routes. For instance, most PCBs can be bound strongly to animal lipids and thus ingested along with the consumption of lipid-rich foods like fish and meats (Li, Arnot, and Wania 2018). By contrast, most PBDEs are sorptive to indoor dust or hard surfaces and thus mainly ingested via hand-to-mouth contacts like palm mouthing and finger licking (Li et al. 2020). Once exposure occurs, the strong binding of these chemicals to human adipose tissues precludes their effective elimination from the human body. As such, they can stay in the human body for years and even decades (Li, Arnot, and Wania 2018; Li et al. 2020).

Since exposure to PCBs/PBDEs can cause adverse health effects such as endocrine disruption, diabetes, metabolic syndrome, and cancer (Fantke and Jolliet 2016), many countries have included these chemicals in nationwide biomonitoring

programs, such as Canada Health Measures Survey, German Environmental Survey, and French National Survey on Nutrition and Health. However, in the United States, the *only* nationwide biomonitoring of these chemicals is conducted as a part of the National Health and Nutrition Examination Survey (NHANES), a continuous program that has been releasing biannual data since 1999. In addition to serum concentration levels of PCBs/PBDEs, NHANES also provides participants' demographic information, medical examination results, and clinical/nutritional biomarkers measurements. By now, these data until 2016 have been made fully available on the NHANES website.

In this article, we leverage the NHANES biomonitoring data of PCBs and PDBEs to assess the disparities in Americans' exposure to these chemicals. Analysis of disparities in human chemical exposure is critical to making population-specific tailored environmental health policies, safeguarding minorities and people of lower socioeconomic status, and promoting environmental justice. Though NHANES has been continuously collecting and providing the measurements of PCBs and PBDEs, chemical exposure and health studies seemed reluctant to use them after 2004, not just in the analysis of disparities but also in other analyses. For instance, a search on PubMed using the keyword "PBDEs" on 11/15/2022 returned 6164 articles published between 2005 and 2022. However, the number dropped to 40 after adding the keyword "NHANES". Of these 40 articles, 10 analyzed the data collected from other sources, the majority

of the rest studies relied on data collected before 2005, and only four actually conducted analyses involving data since 2005 (Sjödin et al. 2014, 2019; Gaylord et al. 2020; Chang et al. 2020). Publications for PCBs follow a similar pattern: only a handful of articles analyzed data collected after 2004 (Caudill 2012, 2015; Bichteler et al. 2017).

The lack of usage actually results from the change in how NHANES measures PCBs and PBDEs. In 2005, NHANES switched from individual biomonitoring to pooled biomonitoring, which physically mixes several participants' serum samples together to form a pool and then measures chemical levels from pooled specimens. The most appealing advantage of pooling is cost-saving. As reported by Caudill (2012), for the measurements of PCBs and PBDEs in the 2005/06 cycle alone, pooled biomonitoring helped NHANES save approximately $2.76 million in total. However, pooling masks variability between individuals in pools, creates challenges for analysis, and reduces the usage of NHANES when information on individuals' chemical concentrations is required. To increase the potential for application of NHANES' long-term efforts, this article provides a novel statistical regression framework for pooled biomonitoring data. The main purpose herein is to assess the disparities, but our method can be easily adapted to studies of other purposes that need a regression analysis.

The handful of existing studies using NHANES pooled biomonitoring data have provided suggestive evidence of disparities in exposure to PCBs and PBDEs across subpopulations, such as people from different racial/ethnic groups and at different ages. For instance, Sjödin et al. (2014) found that the serum levels of PCBs tended to increase with age, whereas those of PBDEs did not differ significantly between different age groups. In addition, African Americans and "other groups" (e.g., Asian Americans) demonstrated significantly higher levels of PCBs in serum than non-Hispanic Whites, whereas Hispanics generally had the lowest PCBs among all racial/ethnic groups (Sjödin et al. 2014; Nguyen et al. 2020). However, disparities in chemical exposure likely result from complex and tangled relationships between different demographic, economic, lifestyle, and even cultural factors. While existing studies focused separately on evaluating exposure disparities in a single factor (age or race), none of them attempted to reveal how these factors jointly determine a subpopulation's exposure to PCBs and PBDEs, and whether the disparities still remain if certain key factors are controlled in statistical analysis. Inadequately addressing confounding effects may prevent policymakers and regulatory agencies from offering cost-effective measures for exposure mitigation and risk management.

The regression-based approach we present is the first attempt to address this issue. Our approach can facilitate revealing and understanding disparities in the exposure of subpopulations to PCBs and PBDEs based on pooled data. Specifically, the new approach supports unzipping pooled data, hence, allowing disentangling of various factors from each other in multivariate regression analysis. With this approach, we systematically investigate (i) the association of age and income with serum levels of PCBs and PBDEs, (ii) the racial/ethnic disparities in exposure to these compounds among Americans after controlling for confounding covariates, and (iii) the effects of national regulations on the reduction in human exposure to these compounds.

Such an investigation corroborates earlier findings based on small-scale biomonitoring studies (as opposed to the national NHANES); for example, lower serum levels of these chemicals among females compared to males, and the dependence of the serum levels on age. More importantly, it provides novel insights into these issues. For instance, when depicting racial/ethnic disparities in human exposure to these compounds, this approach for the first time allows adjusting for potential confounding factors, such as age and income, and capturing differentiated rates of decline in the serum level over time among racial/ethnic groups. Such findings can hardly be obtained, or are likely to be manipulated, by directly analyzing the pooled data due to the interaction between variables used for pooling (age group, gender, and race/ethnicity) and other variables (e.g., income and obesity).

Statistically, our goal is to estimate an individual-level regression model based on pooled biomonitoring data, and we would like to obtain an inference comparable with the one if individual biomonitoring is used. In practice, individual biomonitoring data are often right-skewed (see Figure S.1 in Section A of supplementary material for the histogram of the individual-level data obtained by NHANES before 2005). To account for the right skewness, the standard approach adopted by environmental health studies is log-transformation (see, e.g., Xu et al. 2019; Wang et al. 2019). Consequently, previous works performed analyses on pooled measurements (without regression) proceeded under the assumption that individual concentrations of PCBs and PBDEs follow a lognormal distribution (Caudill 2012, 2015; Sjödin et al. 2014, 2019). Extending this assumption to regression analysis for pooled data is very challenging, mainly because the probability density function of pooled measurements is now a convolution in the form of intractable integrals. Mitchell, Lyles, and Schisterman (2015) proposed an approach yielding analytical expression of integrals under a gamma regression model. However, this approach requires two stringent conditions on pooling: (a) every subject in the same pool must have identical covariate values; (b) the pooled chemical level is an arithmetic average of the individual levels. Unfortunately, NHANES pooled data do not satisfy either condition. To date, only two regression approaches have been developed under the skewed distributional assumption on chemical levels and no stringent conditions on pooling (Mitchell et al. 2014; Liu, McMahan, and Gallagher 2017). However, both methods approximate the intractable integrals using the Monte Carlo method, which usually requires a large Monte Carlo sample size, especially when the integral dimension is high (e.g., larger than 3). In NHANES, most pools are formed by combining eight individual serum specimens, yielding integrals in seven dimensions. More importantly, multiple application studies (Zota et al. 2008; Sjödin et al. 2014, 2019) have noticed nonlinear dependents of PBDEs exposure on age among the U.S. population, which the above two approaches cannot capture because they only considered linear covariate effects.

The remainder of this article is organized as follows. In Section 2, we introduce the motivating pooled biomonitoring data from NHANES. Section 3 presents the proposed regression framework, complete with a Markov chain Monte Carlo (MCMC) procedure. In Section 4, we investigate the finite sample performance of the proposed approach through extensive

simulation studies across various settings. Section 5 analyzes the pooled measurements of three PCB congeners (PCB-99, PCB-153, PCB-180) and three PBDE congeners (BDE-47, BDE-99, BDE-100) and presents some insightful results. In Section 6, we provide a concluding discussion and describe a possible future research topic.

2. NHANES Pooled Biomonitoring Data

NHANES started collecting measurements of PCBs and PBDEs, along with individual-level covariate information from the 1999/2000 and 2003/04 sampling cycles, respectively. Since the 2005/06 cycle and continued afterward, NHANES switched to pooled biomonitoring for both types of chemicals. In particular, during each cycle, NHANES randomly selects one-third of all survey participants over 12 years whose serum sample had a sufficient volume for laboratory testing and then physically combines these individual specimens using a homogeneous pooling strategy presented below. Currently, individual and pooled measurements are available on the NHANES website containing data collected from nine sampling cycles spanning from 1999 to 2016 for PCBs and seven sampling cycles spanning from 2003 to 2016 for PBDEs.

The homogeneous pooling strategy adopted by NHANES proceeds as follows. First, individuals are assigned into different groups defined by the stratification of (a) age group (12–19, 20–39, 40–59, and 60+ years), (b) gender, and (c) race/ethnicity. In each group, individuals are "homogeneous;" that is, having similar ages and the same gender and race/ethnicity. Once the homogeneous groups are determined, individual specimens in the same group are physically combined to form pooled specimens with sizes ranging from 2 to 8 (most are of size 8). It is important to note that in NHANES, each participant is assigned a sampling weight to account for under or oversampling certain population subgroups, and to produce descriptive statistics (e.g., a weighted sample mean) that can accurately depict population characteristics (Sjödin et al. 2019). In order to incorporate the sampling weights into pooled measurements, the volume of each individual specimen contributing to the pool is determined based on the ratio of its sampling weight to the sum of sampling weights of all individuals in the pool (i.e., the pooling weights are proportional to the sampling weights). As a result, the level of a pooled specimen is a weighted arithmetic average of individual chemical concentrations (see Section 3.1 for more details).

In this article, we build a regression model to assess the disparities in Americans' exposure to three PCB congeners (PCB-99, PCB-153, PCB-180) and three PBDE congeners (BDE-47, BDE-99, BDE-100) based on the pooled measurements collected from NHANES. Along with the pooled data, NHANES also provides individual-level information about the participants. Our regression analysis takes the response variable as the logarithm of a contaminant's latent individual-level concentration level and covariates as age, gender, race/ethnicity, obesity, and the ratio of family income to poverty guidelines. After removing subjects with missing values, for PCB-99 (PCB-153, PCB-180), 3104 (4152, 3856) specimens were tested individually and 6015 (6051, 6087) specimens were grouped into 784 (788, 793) pools where 708 (713, 717) pools are of size 8 and 76 (75, 76) pools are of sizes

from 2 and 7. For BDE-47, BDE-99, and BDE-100, respectively, 1842, 1299, and 1765 specimens were tested individually; for each chemical 6087 individuals were divided into 793 pools, including 717 pools of size 8 and 76 pools of sizes from 2 to 7.

3. Model and Method

We now introduce the regression model and present our estimation procedure in a general context. Suppose specimens (e.g., serum or urine) were collected from N individuals and divided into J pools, where the jth pool consists of c_j individual specimens, for $j = 1, \ldots, J$. Further for the ith individual in the jth pool, let $\widetilde{Y}_{ij}$ denote the associated chemical level that is right-skewed, and ω_{ij} be the sampling weight. Under the assumption that individual chemical levels are conditionally independent given the covariates, the relationship between $\widetilde{Y}_{ij}$ and associated covariates is described as follows

$$\log \widetilde{Y}_{ij} | \boldsymbol{Z}_{ij}, \boldsymbol{X}_{ij}$$
$$= \boldsymbol{Z}_{ij}^\top \boldsymbol{\beta} + \sum_{k=1}^{q} f_k(X_{ijk}) + \epsilon_{ij}, i = 1, \ldots, c_j, j = 1, \ldots, J, \quad (1)$$

where $\boldsymbol{Z}_{ij} = (1, Z_{ij1}, \ldots, Z_{ijp})^\top$, $\boldsymbol{X}_{ij} = (X_{ij1}, \ldots, X_{ijq})^\top$ are vectors of covariates for the ith individual in the jth pool, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^\top$ is a vector of regression coefficients, and $f_k(\cdot)$, $k = 1, \ldots, q$, are unknown smooth functions. To avoid the identifiability issue, we require f_k to be centered at 0 (see Hastie et al. (2009) and Section B of supplementary material for a more in-depth discussion). The error terms ϵ_{ij}'s are independent normal random variables with mean 0 and variance σ^2/ω_{ij}, a commonly adopted assumption to account for sampling weights in regression (see, e.g., Pfeffermann 1993; Li, Graubard, and Korn 2010). This assumption implicitly assumes that the sampling weights depend on the covariates but not the response variable. We view it fits our application because NHANES determined the sampling weights before shipping participants' specimens to labs for chemical analyses. We also want to note that if one ignores the sampling weights and fits the data by setting the variance of ϵ_{ij} to be σ^2, the resulting point estimators of the regression coefficients may not change much, however, the standard errors and the resulting inference will be heavily impacted. Desired estimates of the regression coefficients (i.e., $\boldsymbol{\beta}$ and $f_k(\cdot)$, $k = 1, \ldots, q$) in Model (1) can be obtained via off-the-shelf software packages (e.g., `gam` or `mgcv` in R), When individual chemical levels $\widetilde{Y}_{ij}$'s can be observed. In the pooled biomonitoring, since $\widetilde{Y}_{ij}$'s are never observed, estimating (1) becomes challenging.

3.1. Augmented Data Likelihood

The estimation of Model (1) in the context of pooling relies on how to relate the pooled measurements back to $\widetilde{Y}_{ij}$'s. A common assumption is that the chemical level of the jth pool (denoted by Y_j) is a weighted average of individual chemical levels in the same pool, that is, $Y_j = \sum_{i=1}^{c_j} w_{ij} \widetilde{Y}_{ij} / W_j$ where $W_j = \sum_{i=1}^{c_j} w_{ij}$ for some pooling weights w_{ij}'s. In Mitchell et al. (2014) and Liu, McMahan, and Gallagher (2017), $w_{ij} = 1$ since they assumed that each individual contributed the same amount of specimens to a pool. As introduced in Section 2, NHANES determines the

volume of each individual specimen contributing to the pool based on the ratio of its sampling weight to the sum of the sampling weights of all individuals in the pool. Therefore, in this article we focus on the case $w_{ij} = \omega_{ij}$ where ω_{ij} is the sampling weight of the corresponding individual; that is, $Y_j = \sum_{i=1}^{c_j} \omega_{ij} \widetilde{Y}_{ij}/W_j$ and $W_j = \sum_{i=1}^{c_j} \omega_{ij}$. This assumption is standard in the studies of the NHANES pooled-biomonitoring (see, e.g., Caudill 2012). In other practices, extending our method for different setups of w_{ij} is not difficult.

To obtain the observed data likelihood, it is essential to derive the probability density function (pdf) of pooled measurements. Under the aforementioned assumptions $Y_j = \sum_{j=1}^{c} \omega_{ij} \widetilde{Y}_{ij}/W_j$, the pdf of Y_j can be expressed as a convolution of the pdf's of $\widetilde{Y}_{1j}, \ldots, \widetilde{Y}_{c_j j}$. Unfortunately, the analytical expression of the convolution is not available, which hinders the use of observed data likelihood to conduct statistical inference. Existing approaches have used Monte Carlo method to approximate the integrals (Mitchell et al. 2014; Liu, McMahan, and Gallagher 2017). This technique usually requires a large size of random samples and thus has low computational efficiency. To circumvent this issue, we propose a data augmentation procedure; that is, introducing individual chemical level $\widetilde{Y}_{ij}$ as a latent random variable. Note, since Y_j is known, we only need to introduce $c_j - 1$ latent variables, which are $\widetilde{Y}_{2j}, \ldots, \widetilde{Y}_{c_j j}$. Let $\boldsymbol{Y}_p = (Y_1, \ldots, Y_J)^\top$, $\widetilde{\boldsymbol{Y}}_{(-1)j} = (\widetilde{Y}_{2j}, \ldots, \widetilde{Y}_{c_j j})^\top$, and $\widetilde{\boldsymbol{Y}}_{(-1)} = (\widetilde{\boldsymbol{Y}}_{(-1)1}, \ldots, \widetilde{\boldsymbol{Y}}_{(-1)J})$, then the following joint distribution is obtained,

$$\pi(\boldsymbol{Y}_p, \widetilde{\boldsymbol{Y}}_{(-1)} | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\eta})$$

$$= \prod_{j=1}^{J} \left[\frac{\sqrt{\omega_{1j}}}{Y_{1j}^* \sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{\omega_{1j}(\log Y_{1j}^* - \eta_{ij})^2}{2\sigma^2} \right\} I(Y_{1j}^* > 0) \right.$$

$$\left. \times \prod_{i=2}^{c_j} \frac{\sqrt{\omega_{ij}}}{\widetilde{Y}_{ij} \sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{\omega_{ij}(\log \widetilde{Y}_{ij} - \eta_{ij})^2}{2\sigma^2} \right\} I(\widetilde{Y}_{ij} > 0) \right], \quad (2)$$

where $Y_{1j}^* = \omega_{1j}^{-1}(W_j Y_j - \sum_{i=2}^{c_j} \omega_{ij} \widetilde{Y}_{ij})$, $\eta_{ij} = \boldsymbol{Z}_{ij}^\top \boldsymbol{\beta} + \sum_{k=1}^{q} f_k(X_{ijk})$, $\boldsymbol{\eta}_j = (\eta_{1j}, \ldots, \eta_{c_j j})^\top$, and $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_J)$.

There are two major obstacles in the development of an efficient Markov chain Monte Carlo (MCMC) procedure for posterior inference. First, the conditional distribution of latent individual measurements $\widetilde{Y}_{ij}$ is difficult to sample from directly. We developed an efficient Metropolis-Hastings (MH) algorithm in Section 3.2. Second, estimating nonparametric smooth functions is challenging as the parameter space is infinite-dimensional. Here we propose to use penalized splines because (a) it is easy to implement and spline coefficients can be sampled along with regression coefficients; (b) they will not suffer from the overfitting issue which is often experienced by regular regression splines. Details of the penalized spline regression are provided in Section 3.3.

3.2. Posterior Sampling of Latent Individual Chemical Levels

From the augmented data likelihood (2), the conditional density function of $\widetilde{Y}_{ij}$, for $i = 2, \ldots, c_j$ and $j = 1, \ldots, J$, satisfies

$$\pi(\widetilde{Y}_{ij} | \sigma^2, Y_j, \boldsymbol{\eta}_j, \widetilde{\boldsymbol{Y}}_{(-1,-i)j})$$

$$\propto \frac{1}{Y_{1j}^* \widetilde{Y}_{ij}} I(\widetilde{Y}_{ij} > 0) I(Y_{1j}^* > 0)$$

$$\times \exp\left\{ -\frac{\omega_{1j}(\log Y_{1j}^* - \eta_{1j})^2 + \omega_{ij}(\log \widetilde{Y}_{ij} - \eta_{ij})^2}{2\sigma^2} \right\},$$

where $\widetilde{\boldsymbol{Y}}_{(-1,-i)j}$ represents outcome values for all subjects in the jth pool except for subject 1 and i. Sampling from the above distribution is not straightforward, thus, we developed an MH algorithm. Given that $Y_{1j}^* = \omega_{1j}^{-1}(W_j Y_j - \sum_{i=2}^{c_j} \omega_{ij} \widetilde{Y}_{ij})$ must be positive, a natural upper bound for $\widetilde{Y}_{ij}$ is $B_{ij} = \omega_{ij}^{-1}(W_j Y_j - \sum_{i'=2,i'\neq i}^{c_j} \omega_{i'j} Y_{i'j})$ for $i = 2, \ldots, c_j$, therefore, a truncated normal distribution is employed as the proposal distribution for $\log \widetilde{Y}_{ij}$. In particular, let $\widetilde{Y}_{ij}^{(m)}$ and $\sigma^{(m)}$ denote the values of $\widetilde{Y}_{ij}$ and σ from the mth iteration of MCMC procedure, for $i = 2, \ldots, c_j$ and $j = 1, \ldots, J$. Then in the $(m+1)$th iteration, the proposal distribution for $\log \widetilde{Y}_{ij}$ is a truncated normal; that is,

$$\mathcal{TN}\left(\text{mean} = \log \widetilde{Y}_{ij}^{(m)}, \text{sd} = \sigma^{(m)} \omega_{ij}^{-1/2}; \text{lower} = -\infty, \right.$$

$$\left. \text{upper} = \log B_{ij}^{(m)}\right).$$

Formulas for $B_{ij}^{(m)}$ and the acceptance probability of the proposed sample are provided in Section C.1 of supplementary material.

3.3. Approximating Unknown Functions with Penalized Splines

B-spline functions are popular to approximate smooth functions; that is, $f_k(\cdot) \approx \sum_{l=1}^{p_k} B_{kl}(\cdot)\alpha_{kl}$, for $k = 1, \ldots, q$, where $B_{kl}(\cdot)$ is a B-spline basis function and α_{kl} is the corresponding coefficient. A routine procedure of using spline functions is to select the optimal number of knots to avoid under or overfitting by cross-validation. This process is time-consuming if the model includes multiple functions. A popular alternative approach is to specify a relatively large number of knots and then penalize the spline functions to avoid overfitting. Following Lang and Brezger (2004), we specify the following priors for $\boldsymbol{\alpha}_k = (\alpha_{k1}, \ldots, \alpha_{kp_k})^\top$; that is, $\boldsymbol{\alpha}_k | \tau_k \propto \exp\left\{-\boldsymbol{\alpha}_k^\top \boldsymbol{D}_k \boldsymbol{\alpha}_k/(2\tau_k)\right\}$, where $\boldsymbol{D}_k$ is a penalty matrix, and τ_k can be viewed as the inverse of the penalty parameter and controls the smoothness of unknown functions. Note, the matrix $\boldsymbol{D}_k$ is not necessarily full rank. Here we let $\boldsymbol{D}_k = \boldsymbol{\Delta}_k^{(r)\top} \boldsymbol{\Delta}_k^{(r)}$, where $\boldsymbol{\Delta}_k^{(r)}$ is the discrete difference operator matrix of order r and has been widely used because it is easy to specify and yields sparse penalty matrix (Eilers and Marx 1996). This operator matrix is best defined recursively as $\boldsymbol{\Delta}_k^{(r+1)} = \boldsymbol{\Delta}_k^{(1,r)} \boldsymbol{\Delta}_k^{(r)}$, where

$$\boldsymbol{\Delta}_k^{(1,r)} = \begin{bmatrix} -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(p_k - r - 1) \times (p_k - r)},$$

and $\boldsymbol{\Delta}_k^{(1)} = \boldsymbol{\Delta}_k^{(1,0)}$. Through extensive numerical studies, we have found that setting $r = 2$ is sufficient when implementing the proposed approach; that is, $\boldsymbol{\alpha}_k^\top \boldsymbol{\Delta}_k^{(2)\top} \boldsymbol{\Delta}_k^{(2)} \boldsymbol{\alpha}_k = \sum_{l=3}^{p_k}(\alpha_{kl} - $

$2\alpha_{k,l-1} + \alpha_{k,l-2})^2$. For τ_k, we specify a conjugate prior; that is, $\tau_k \sim \mathcal{IG}(a_k, b_k)$ meaning τ_k has an inverse gamma prior with predetermined parameters a_k and b_k. The values of hyperparameters a_k and b_k should lead to almost flat priors for τ_j (Lang and Brezger 2004). A common choice is $a_k = 1$ and a small value for b_k. We examined $b_k = 0.005$ and $b_k = 0.0005$ using simulation and did not find any noticeable difference in the estimation. Thus, we fix $a_k = 1$ and $b_k = 0.005$ throughout this article.

Proceeding with the aforementioned notations and assumptions, we derived the joint distribution of latent variables and parameters and obtained a full MCMC procedure. Technical details are provided in Section C.2 of supplementary material. In summary, our MCMC procedure iteratively samples $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_k$'s from normal distributions and σ^2 and τ_k's from inverse-gamma distributions. These explicit distributions facilitate a computationally efficient implementation of our methods.

4. Simulation Studies

In this section, we present numerical studies to illustrate the finite-sample performance of our proposed method. We considered two individual-level additive partially linear models, both of which take the form

$$\log \widetilde{Y} | Z_1, Z_2, Z_3, X_1, X_2$$
$$= \beta_0 + Z_1\beta_1 + Z_2\beta_2 + Z_3\beta_3 + f_1(X_1) + f_2(X_2) + \epsilon,$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3) = (1, 0.5, 0.5, -0.5)^\top$ and $\epsilon \sim \mathcal{N}(0, \sigma^2/\omega)$ in which $\sigma = 1$ and ω is the sampling weight and generated from $\mathcal{U}(1, 10)$ to ensure the ratio of the largest and smallest weight within the same pool is less than 10, sample as in NHANES. Moreover, we generate $Z_1 \sim$ Bernoulli(0.5), $Z_2 \sim$ Bernoulli(0.3), $Z_3 \sim$ Bernoulli(0.5), $X_1 \sim \mathcal{U}(-2, 2)$, and $X_2 \sim \mathcal{N}(0, 1)$ to imitate the covariates gender (1 for Male; 0 otherwise), obesity (BMI $\geq$ 30), race (1 for White; 0 otherwise), (scaled) age, and (standardized) ratio of family income to poverty in the NHANES data, respectively. To account for the correlation between gender and race, we let $X_1 = 4U_1 - 2$ and $X_3 = \Phi^{-1}(U_2)$, in which Φ^{-1} is the quantile function of $\mathcal{N}(0, 1)$, and $(U_1, U_2)'$ follows a bivariate Gaussian copula with a correlation coefficient $\rho = 0.2$, about the same as the observed correlation between the two covariates. The $f_1(\cdot)$ and $f_2(\cdot)$ take different forms in the two considered models to cover a broad range of nonlinear patterns. In the first model (M1), we have

$$f_1(X_1) = 0.8 \exp\left\{-0.98^{-1}(X_1 + 1)^2\right\}$$
$$+ 0.6 \exp\left\{-1.125^{-1}(X_1 - 1)^2\right\} - 0.58,$$
$$f_2(X_2) = 0.7 \exp\left[4^{-1}\left\{-I(X_2 > 0)1.2^2 - I(X_2 < 0)1.2^{-2}\right\}X_2^2\right]$$
$$- 0.568,$$

where $I(\cdot)$ is the usual indicator function; in the second model (M2),

$$f_1(X_1) = \frac{0.2 \sin\{\pi(X_1 - 0.5)/2.5\} + 0.4}{\exp[\{X_1 + (X_1 + 0.5)^2 I(X_1 > -0.5)\}/6]}$$
$$- 0.288, \quad f_2(X_2) = \frac{4 \exp(2X_2)}{5 + 5 \exp(2X_2)} - 0.4.$$

Note that we introduced the additive constants (e.g., -0.58) to ensure $E\{f_k(X_k)\} = 0$ such that f_k is identifiable.

We set $J = 250$ to mimic a fixed-budget design (i.e., only J assays are affordable within financial confines) and considered a common pool size $c_j = c \in \{4, 8\}$ for all j. For each c, we generated pooled biomonitoring data following both random and homogeneous pooling strategies. Under random pooling, we first generated $N = Jc$ independent and identically distributed (iid) copies of $(\widetilde{Y}, Z_1, Z_2, Z_3, X_1, X_2, \omega)$ from each model (M1 and M2), then randomly assigned the N observations to J pools, and labeled them by $(\widetilde{Y}_{ij}, Z_{ij1}, Z_{ij2}, Z_{ij3}, X_{ij1}, X_{ij2}, \omega_{ij})$ for $i = 1, \ldots, c$ and $j = 1, \ldots, J$. Finally, the jth pooled response was calculated by $Y_j = W_j^{-1} \sum_{i=1}^{c} \omega_{ij} \widetilde{Y}_{ij}$ with $W_j = \sum_{i=1}^{c_j} \omega_{ij}$, for $j = 1, \ldots, J$.

Under the homogeneous pooling strategy, instead of randomly, observations with similar values of (Z_1, Z_3, X_1) were pooled together. This strategy emulates the NHANES study which pooled individuals of the same gender and race and similar ages together. Specifically, N observations were first partitioned into four parts, each with a distinct value of the discrete vector $(Z_1, Z_3)'$, respectively. Then, each part was sorted into an increasing order with respect to X_1. Following this order, we sequentially assigned individual observations (i.e., $\widetilde{Y}_{ij}$) to pools of size c to generate the Y_j's. The data-generating process under both strategies was repeated $B = 500$ times for each regression model (M1 and M2) and each $c \in \{4, 8\}$.

For each dataset, we let regression coefficients $\boldsymbol{\beta}$ have a flat prior and σ^2 follows $\mathcal{IG}(a = 0.5, b = 0.5)$. Priors for τ_k were specified in Section 3.3. To ensure enough flexibility, Eilers and Marx (1996) suggested specifying a relatively large number of knots, usually between 20 and 40. In practice, we used 30 inner knots being placed at equally-spaced quantiles of the observed covariate values. Simulation results showed that it was enough to guarantee sufficient accuracy for the unknown functions. In each estimation, we discarded the first 1000 samples and retained every 10th iterate from the next 20,000 samples. Trace plots demonstrated our choice of the burn-in period and the thinning factor is sufficient for excellent mixing.

Table 1 summarizes the $B = 500$ posterior median estimates of β_1, β_2, and β_3 in both models (M1 and M2) under both pooling strategies when $c \in \{4, 8\}$. For comparison, we also included the results under the setting $c = 1$; that is, the regular individual biomonitoring. The findings from Table 1 are threefold. First, the proposed approach can obtain accurate parameter estimates and reliable inference results, across all considered settings. In particular, the bias (Bias) in the estimates of β_1, β_2, and β_3 is close to zero, and the average of estimated posterior standard deviation (ESE) is in close agreement with the sample standard deviation (SSD) of the parameter estimates. The empirical coverage probabilities of 95% credible intervals (CP95) are all at the nominal level. Second, homogeneous pooling can improve the accuracy and efficiency of estimates when compared to random pooling. For example with homogeneous pools, the square root of mean squared error (RMSE) for the estimates of β_1 and β_3 are approximately $1/\sqrt{c}$ of the estimates obtained using random pools. Note, the homogeneous pools are formed by grouping individuals of the same value of (Z_1, Z_3), but not Z_2, thus, estimates of β_2 behave similarly under both pooling strategies. Third, as sample size c increases, homogeneous pooling can lead to more accurate and efficient estimates. For instance, the RMSE

Table 1. Simulation results for Models (M1) and (M2) when $c \in \{1, 4, 8\}$ under both random and homogeneous pooling.

Model	Parameter	Summary	$c = 1$ Individual	$c = 4$ Random	Homogeneous	$c = 8$ Random	Homogeneous
M1	$\beta_1 = 0.5$	Bias×100	0.106	−0.062	0.204	−0.347	−0.072
		RMSE(CP95)	0.055(0.951)	0.054(0.941)	0.029(0.948)	0.057(0.932)	0.022(0.934)
		SSD(ESE)	0.055(0.056)	0.054(0.055)	0.029(0.029)	0.057(0.055)	0.022(0.021)
	$\beta_2 = 0.5$	Bias×100	−0.030	−0.088	0.052	−0.677	−0.427
		RMSE(CP95)	0.060(0.949)	0.053(0.943)	0.051(0.955)	0.050(0.956)	0.054(0.949)
		SSD(ESE)	0.060(0.061)	0.053(0.054)	0.052(0.053)	0.049(0.052)	0.053(0.052)
	$\beta_3 = -0.5$	Bias×100	0.172	0.167	0.023	−0.066	−0.112
		RMSE(CP95)	0.059(0.936)	0.055(0.935)	0.028(0.959)	0.057(0.943)	0.021(0.932)
		SSD(ESE)	0.059(0.056)	0.055(0.055)	0.028(0.029)	0.057(0.056)	0.021(0.021)
M2	$\beta_1 = 0.5$	Bias×100	0.206	−0.283	−0.005	−0.061	−0.044
		RMSE(CP95)	0.056(0.946)	0.055(0.944)	0.028(0.958)	0.055(0.946)	0.021(0.948)
		SSD(ESE)	0.056(0.056)	0.055(0.055)	0.028(0.029)	0.055(0.055)	0.021(0.021)
	$\beta_2 = 0.5$	Bias×100	−0.066	−0.281	−0.229	−0.374	−0.387
		RMSE(CP95)	0.059(0.964)	0.056(0.930)	0.054(0.941)	0.054(0.931)	0.052(0.943)
		SSD(ESE)	0.059(0.061)	0.056(0.054)	0.054(0.053)	0.054(0.052)	0.052(0.052)
	$\beta_3 = -0.5$	Bias×100	0.293	0.412	0.030	0.127	0.007
		RMSE(CP95)	0.053(0.956)	0.055(0.939)	0.029(0.946)	0.055(0.946)	0.021(0.946)
		SSD(ESE)	0.053(0.056)	0.055(0.055)	0.029(0.029)	0.055(0.055)	0.021(0.021)

NOTE: The summary statistics for β_1, β_2, β_3 include average bias (Bias), square root of mean squared error (RMSE), sample standard deviation (SSD) of the 500 posterior median estimates, the average of the 500 estimates of the posterior standard deviation (ESE), and empirical coverage probability of 95% credible intervals (CP95).

for the estimates of β_1 and β_3 when $c = 8$ are about 40% of the ones when $c = 1$. However, the same conclusion cannot be made when using random pooling as the RMSE stay almost the same with increasing pool sizes.

Figure 1 summarizes the estimates of the function $f_1(\cdot)$ in Model (M1) under both pooling strategies when $c \in \{1, 4, 8\}$. In all subfigures, we display the pointwise 2.5th, 50th (median), and 97.5th percentiles of the $B = 500$ estimated functions based on the posterior median. First, we note that all the pointwise median curves are in nearly perfect agreement with the true functions, meaning the proposed approach can return estimates of $f_1(\cdot)$ with little bias. Second, estimates of $f_1(\cdot)$ obtained using homogeneous pools show less variability (i.e., smaller shaded area) when compared to the results obtained using random pools. Third, under homogeneous pooling, larger pool size c will help reduce the variability of estimates of $f_1(\cdot)$. Estimates for the function $f_2(\cdot)$ in Model (M1) are summarized in Section D.1 of supplementary material. Recall that under the homogeneous pooling strategy, homogeneity was not based on X_2. Therefore, estimates of $f_2(\cdot)$ have similar performance under both pooling strategies. The same patterns can be concluded for Model (M2) by figures in Section D.1 of supplementary material.

Figure 2 visualizes a comprehensive comparison between estimates under the two pooling strategies when $c \in \{1, 4, 8\}$. For regression coefficients β_1, β_2, and β_3, we display the boxplots of the $B = 500$ posterior median estimates. For unknown functions $f_1(\cdot)$ and $f_2(\cdot)$, we calculated the square root of integrated squared error (RISE) of each posterior median estimate; that is, $\text{RISE}(\widehat{f_k}) = \sqrt{\ell^{-1} \sum_{l=1}^{\ell} \{\widehat{f_k}(u_l) - f_k(u_l)\}^2}$, where $\{u_l\}_{l=1}^{\ell}$ are evenly spaced grid points, and present the boxplot of the 500 RISEs for each function in the bottom two rows of Figure 2. These comparisons reinforce the significant gain in precision and efficiency via the use of homogeneous pooling. The same figure for Model (M2) is in Section D.1 of supplementary material.

In addition, we considered a fixed-number-of-subjects design, where the total number of subjects, N, is fixed to be $cJ = 2000$. We let the pool size c increase from $c = 1$, to $c = 4$ (i.e., $J = 500$) and $c = 8$ (i.e., $J = 250$). Results are summarized in Section D.2 of supplementary material. Under random pooling, we see that as c increases, the variability of estimates increases, indicating a loss of estimation efficiency, which is expected since using a larger pool size naturally aggregates more information in pooling. However, under homogeneous pooling, because the homogeneity holds for Z_1, Z_3, and X_1, we see that the variability of estimates of β_1, β_2, and f_1 retain nearly the same for $c \in \{1, 4, 8\}$. This observation further demonstrates the advantages of using homogeneous pooling in practice.

5. Assessing Exposures to PCBs and PBDEs

We now analyze the pooled NHANES measurements of three PCB congeners (PCB-99, PCB-153, PCB-180) and three PBDE congeners (BDE-47, BDE-99, BDE-100). Because these chemicals were collected over a span of more than a decade, we included interactions between sampling cycles and covariates to study the change of covariates' effects over time. We chose the following full model to describe the individual serum chemical levels:

$$\log \widetilde{Y} | \mathbf{T}, \mathbf{E}, \mathbf{X} = \beta_0 + \sum_{p=1}^{S} \alpha_p T_p + \sum_{q=1}^{5} \zeta_q E_q + \sum_{p=1}^{S} \sum_{q=1}^{5} \gamma_{pq} T_p E_q$$

$$+ f_1(X_1) + f_2(X_2) + \sum_{p=1}^{S} \sum_{k=1}^{2} \delta_{pk} T_p X_k + \epsilon. \tag{3}$$

Herein, the $\mathbf{T} = (T_1, \ldots, T_S)^\top$ in which T_1 to T_S are binary indicator variables for the S previous sampling cycles before the 2015/16 cycle (this is the reference sampling period: $S = 8$ for PCB, and $S = 6$ for PBDE) to control for the time
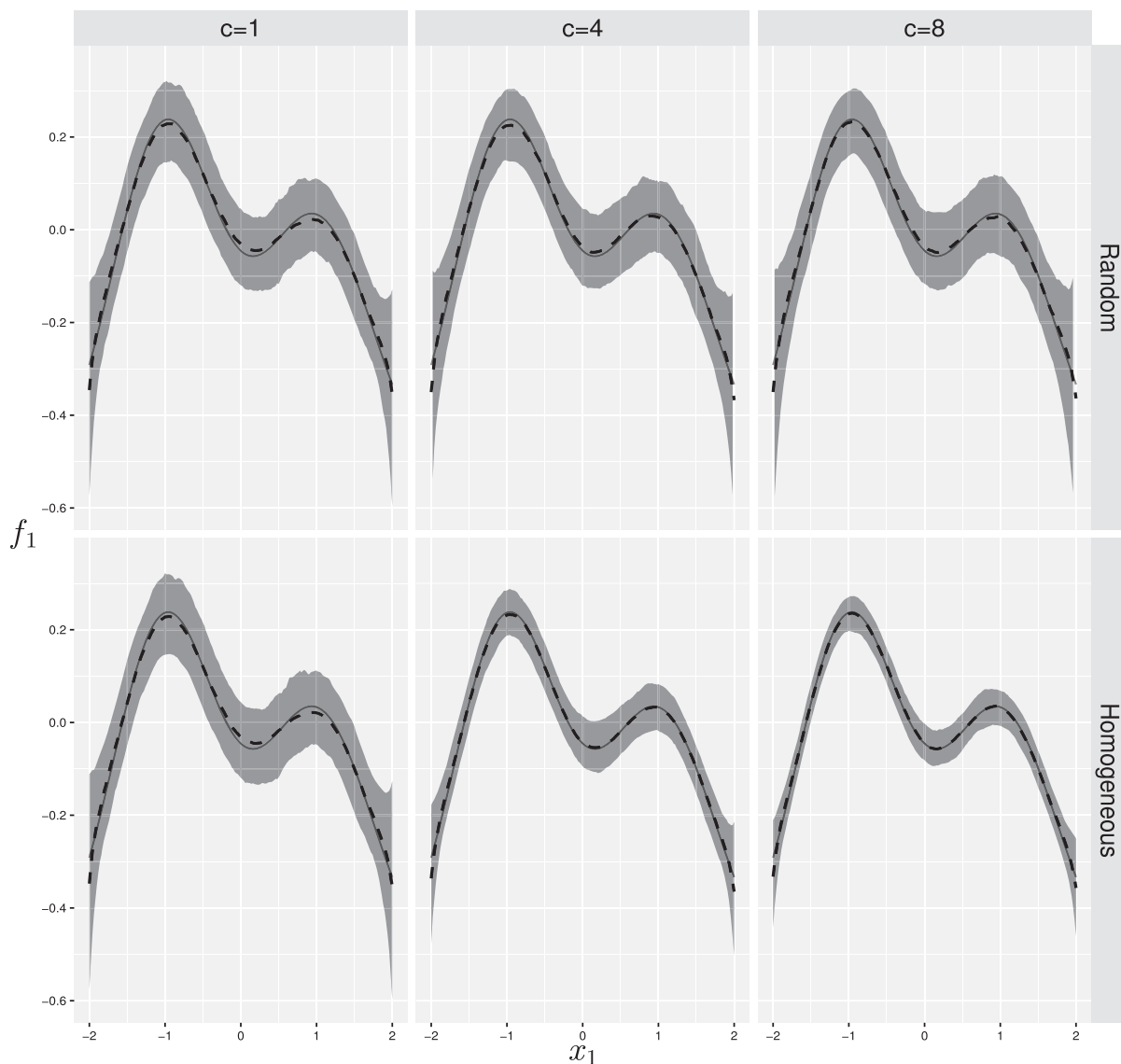
Figure 1. Summary of estimates of $f_1(\cdot)$ in Model (M1) when $c \in \{1, 4, 8\}$ under both random and homogeneous pooling. In each figure, gray solid and dashed lines depict the true function and pointwise median of 500 estimates of function, respectively. The shaded area is bound by the 2.5th and 97.5th percentiles of the 500 estimates of function.

effect; the $\boldsymbol{E} = (E_1, \ldots, E_5)^\top$, in which E_1 is the male status indicator ($= 1$ if Male; $= 0$ if Female), E_2 to E_4 are binary indicator variables for Hispanic (including Mexican American and Other Hispanics), Non-Hispanic Black, and Other Race (Non-Hispanic White is the reference group), respectively, and E_5 indicates obesity (BMI ≥ 30, where BMI stands for body mass index); the $\boldsymbol{X} = (X_1, X_2)^\top$ in which X_1 denotes age at examination (in years; ranges from 12 to 80), and X_2 denotes the ratio of family income to poverty guidelines (ranges from 0 to 5); all the $T_p E_q$ and $T_p X_k$ terms are interactions; and ϵ is a normal random variable with mean 0 and variance σ^2/ω where ω is the sampling weight.

It is easy to see that Model (3) is a special case of Model (1) as we can collect the T_p's, E_q's, $T_p E_q$'s, and $T_p X_k$'s into the covariate $\boldsymbol{Z}$ and the β_0, α_p's, ζ_q's, γ_{pq}'s, and δ_{pk}'s into the parameter $\boldsymbol{\beta}$. In our estimation, prior specifications for $\boldsymbol{\beta}$, σ^2 and τ_k are the same as in Section 4. The $\mathcal{N}(0, 10)$ is used as priors for the coefficients of interaction terms γ_{pq}'s and δ_{pk}'s. We discarded the

first 10,000 samples to establish the convergence of MCMC, and then retained every 10th sample from the next 100,000 iterations. For each considered dataset, we ran three chains, and for each chain, the initial values of regression and spline coefficients are randomly generated from a uniform distribution between -1 and 1. All 3 chains yield almost identical results, thus, we summarize the posterior inference results based on the first chain. Figure 3 presents the 95% equal tail credible intervals for the effect of binary covariates E_1 to E_5 across different sampling cycles for PCB-153 and BDE-47. Figures 4 displays the pointwise 2.5th, 50th (median), and 97.5th percentiles of the posterior samples for unknown smooth functions across different sampling cycles for PCB-153 and BDE-47, respectively. Section E of supplementary material presents the same figures for PCB-99, PCB-180, BDE-99, and BDE-100. As one can compare to see that the patterns of estimates for congeners in the same chemical family are very similar. Therefore, in the following, we mainly discuss the results for PCB-153 and BDE-47.
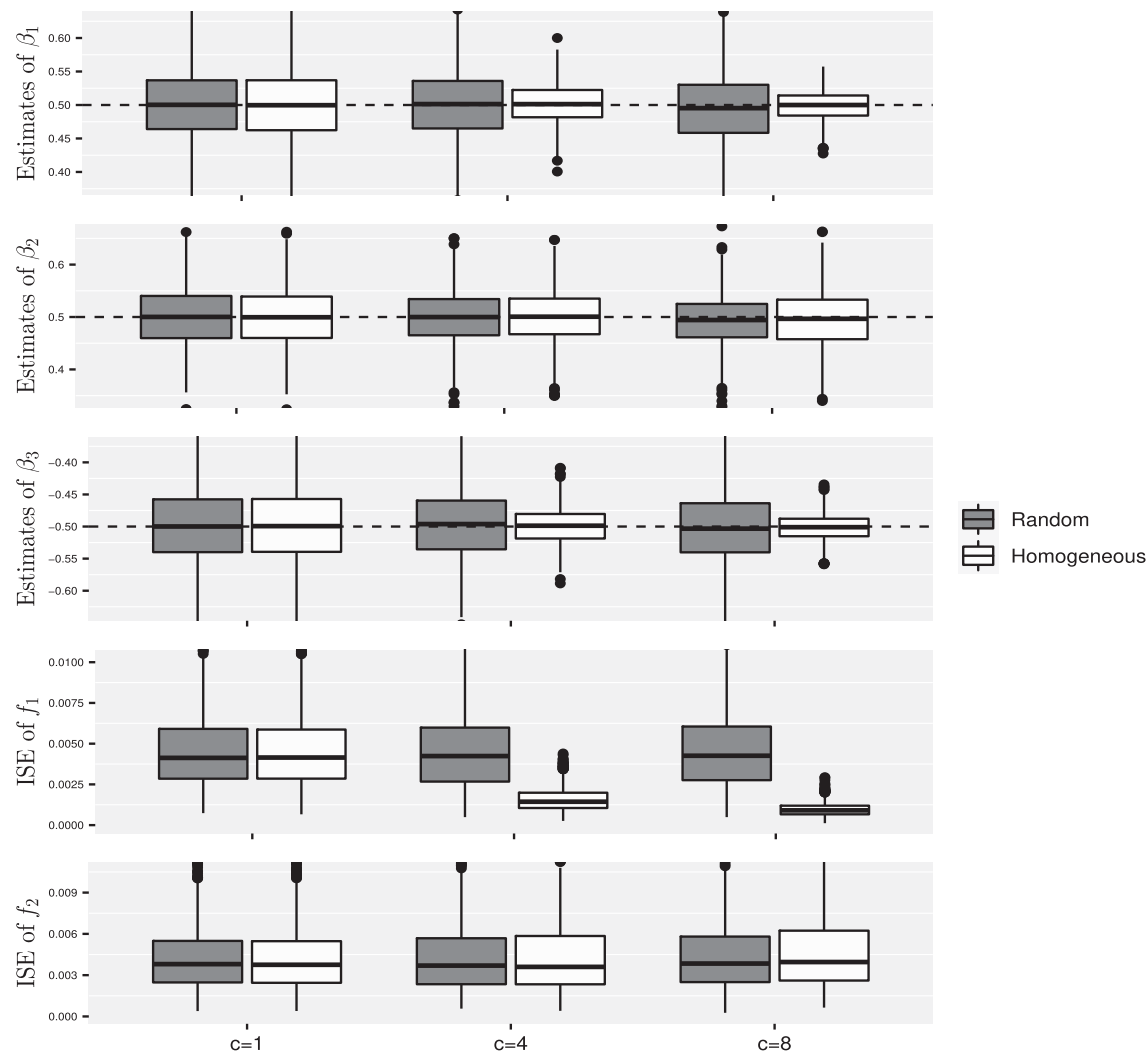
Figure 2. Simulation results for Model (M1) when $c \in \{1, 4, 8\}$ under both random (gray box) and homogeneous (white box) pooling strategies. Each boxplot is based on the 500 summary statistics (Mean estimates or RISE). For each row, two boxplots corresponding to $c = 1$ are the same as individual measurements are available. Dashed lines in the top three rows represent the true value.

5.1. Results for PCB-153

Figure 3 displays the estimated regression coefficients for PCB-153 in different sampling cycles. The most striking result is the statistically significant decreasing trend in PCB-153 concentration among the U.S. population throughout 2005–2016 period. This result indicates the effectiveness of PCB regulation and remediation in the United States over decades.

At the significant level of 0.05, PCB-153 concentrations among males are similar to, or higher than, females for most sampling cycles. The lower chemical levels among females have previously been hypothesized to be a result of reproduction because reproduction is a remarkable mechanism for eliminating "lipophilic" chemicals that are strongly bound to human adipose tissues from mothers (Moysich et al. 2002; Sarcinelli et al. 2003; Bergonzi et al. 2009). Our result coincides with this hypothesis. In addition, Figure 3 shows that obesity is negatively correlated to PCB-153 concentration. This is likely because obese individuals' larger volume of adipose tissue apparently "dilutes" PCB concentration when they take in a similar level of PCBs as lean individuals do (Li and Wania 2017). As

Figure 3 displays, compared to Non-Hispanic Whites, Hispanic Americans have consistently lower PCB-153 concentrations, whereas Non-Hispanic Blacks and Other Race have higher ones than Non-Hispanic Whites for most sampling cycles. On the other hand, Figure 4 suggests that on average individuals with higher family income have a similar or even higher PCB-153 concentration, compared to those with lower family income.

Since diet is the main pathway responsible for human intake of PCBs, different dietary patterns between sub-populations result in the observed disparities in chemical concentration. An earlier comprehensive analysis (Undeman et al. 2010) has confirmed that consumption of seafood is the most predominant contributor, among all food items, to dietary intake of PCBs. Therefore, individuals consuming more seafood are anticipated to possess higher PCB concentrations. For instance, U.S. adult men (aged 20 and over; the same hereafter) consume more seafood than U.S. adult women, evident by the most recent nationally representative average rates of 0.69 and 0.56 ounces equivalent, respectively, documented in the U.S. Food Patterns
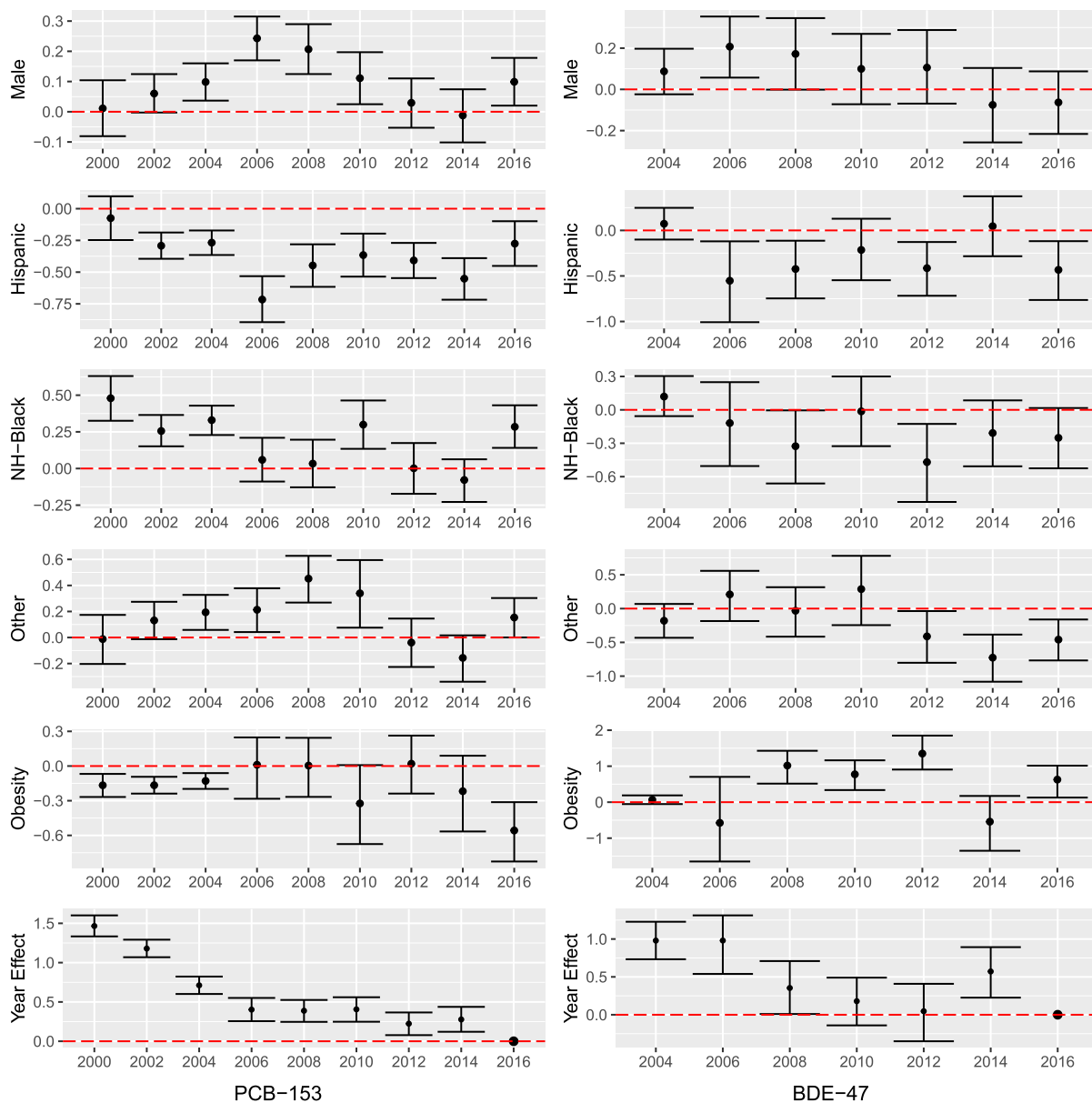
Figure 3. Estimates of covariate effects for PCB-153 and BDE-47 across different sampling cycles. In each figure, pointwise posterior median (dot in the middle), 2.5th (bottom bar), and 97.5th (top bar) percentiles are presented.

Equivalent Database (combining seafood low and high in *n*-3 fatty acids in the database; the same hereafter) (Agricultural Research Service 2018). This comparison agrees with our finding of higher PCB-153 concentrations observed among male participants in NHANES. Likewise, in a nationwide perspective, Hispanic adults consume less seafood (on average 0.63 ounces equivalent daily) than other racial groups (e.g., on average 0.73 and 1.06 ounces equivalent for African and Asian Americans, respectively) (Agricultural Research Service 2018). An earlier survey in an ethnically diverse location (Sacramento-San Joaquin Delta in California) shows that Hispanics consume 3% to 71% less fish than Whites, African Americans, and Asians, probably due to cultural differences (Silver et al. 2007). In addition, individuals with high income consume more seafood than those with low income; for instance, the U.S. Food Patterns Equivalent Database shows that individuals aged 20 and over with family income greater than $75,000 consume on average

0.67 ounces equivalent of seafood each day, higher than the daily rate for those with family income less than $25,000 of on average 0.52 ounces equivalent (Agricultural Research Service 2018). This disparity is consistent with the finding from a comprehensive and critical review that "consumption of [...] fish [..] was consistently associated with higher socioeconomic status groups" due to the differences in the energy cost between food items and in the physical accessibility to low-cost energy-dense foods between individuals with different income levels (Darmon and Drewnowski 2008). Interestingly, PCB-153 is not the only case that increased contaminant concentrations are found among individuals with higher income: since consumption of seafood is also a dominant contributor to human exposure to mercury, individuals with higher income eat more fish and have higher blood mercury as well (Mahaffey, Clickner, and Jeffries 2009).
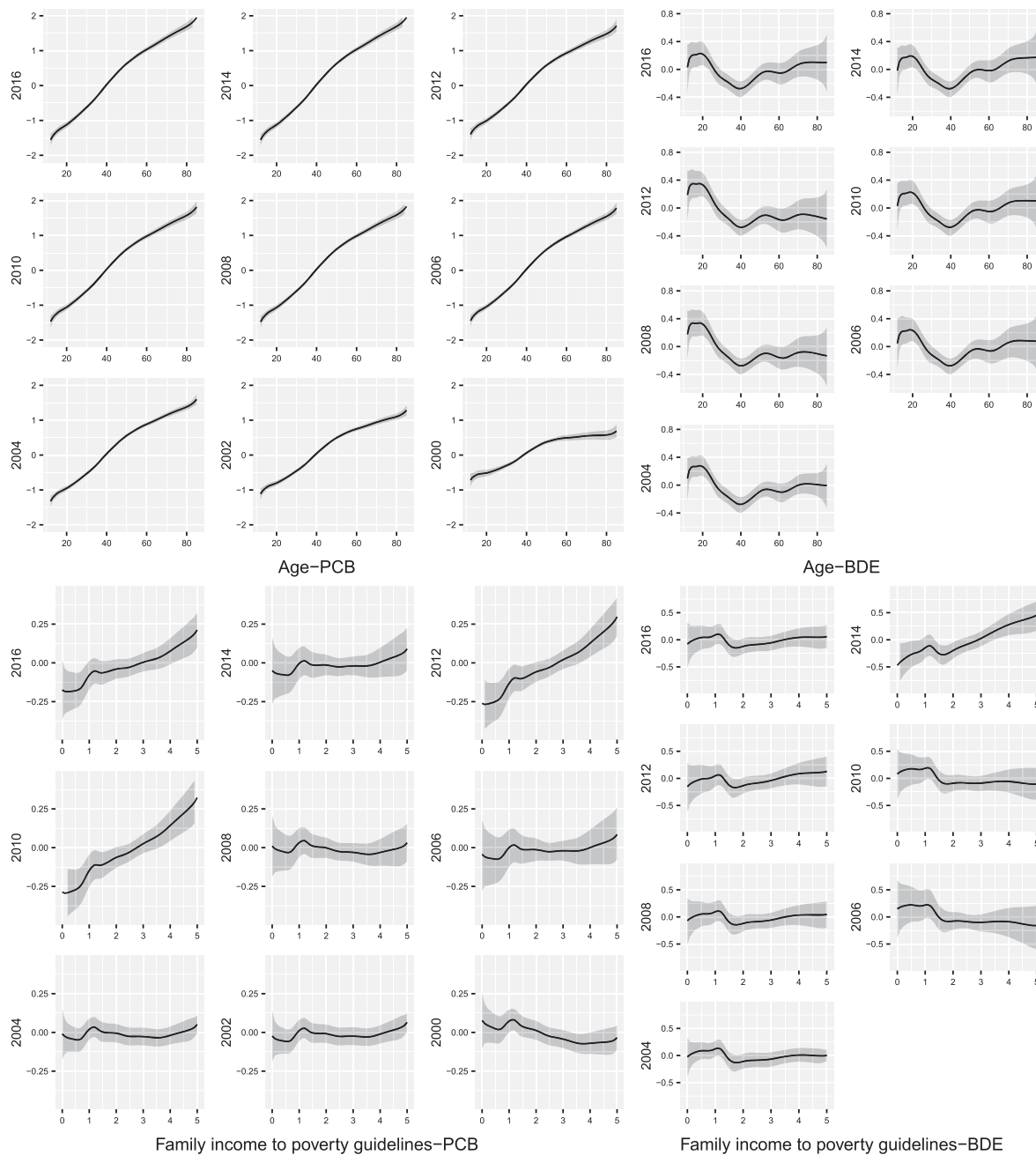
Figure 4. Estimates of covariate effects for PCB-153 and BDE-47 across different sampling cycles. In each figure, the solid curve is the pointwise posterior median. The shaded areas are bounded by the 2.5th and 97.5th percentiles.

Figure 4 also demonstrates that the PCB-153 concentration increases with age. Since PCBs resist elimination from the human body, their concentrations in the blood reflect the "memory" of the cumulative ongoing exposure over years and even more than a decade (Li, Arnot, and Wania 2018). Compared to younger individuals, senior individuals have a longer history of exposure in their lives. Most of their past exposure occurred in a time more contaminated than today, as PCB concentrations in the U.S. environment and foods have kept declining since the phase-out of PCBs in the late 1970s (evident by the declining temporal trend in Figure 3). For these reasons, it is not surprising to observe an increasing trend of chemical exposure with age, which coincides with the findings in Quinn and Wania (2012).

5.2. Results for BDE-47

Figure 3 displays the estimated regression coefficients for BDE-47 in different sampling cycles. The BDE-47 concentration started declining since the 2006/07 sampling cycle, with the only exception of a slight rebound in the 2013/14 sampling cycle, indicating a transition from the pre-ban to post-ban periods. This transition coincides with the decadal effort in banning PBDEs in the United States, which started early in 2006 when the U.S. Environmental Protection Agency proposed to end their use and eventually succeeded in 2013 when California's Smolder Resistance Technical Bulletin 117 was amended (Cordner et al. 2013). Figure 3 shows that obesity plays an inconsistent role in determining BDE-47 concentration: BDE-47 concentration is higher among obese individuals than lean individuals for

2007/08, 2009/10, 2011/12, and 2015/16 sampling cycles but not significantly different from the rest of sampling cycles.

Overall, males show slightly a higher BDE-47 concentration than females (Figure 3). Hispanics, Non-Hispanic Blacks, and Other Race all have significantly lower concentrations than Non-Hispanic White for most sampling cycles (Figure 3). The country of origin may contribute to the lower chemical concentrations in Hispanics and Non-Hispanic Blacks, given that serum PBDE levels have been observed to be significantly lower in foreign-born Mexican Americans than their U.S.-born counterparts because the U.S. environment is more contaminated than elsewhere (Zota et al. 2008).

Figure 4 demonstrates the dependence of BDE-47 concentration on age and income among individuals sampled. It is first noted that the three chemicals' concentrations peak in adolescence, decline after age 20, rebound after age 40, and maintain stable throughout the rest of adulthood. Such a non-monotonic trend reflects the interaction between the divergent age dependence of ingestion and elimination of this chemical. On the one hand, ingestion of PBDE-contaminated indoor dust via hand-to-mouth contact is the predominant route for human exposure to PBDEs (Li et al. 2020). Since such hand-to-mouth contact is more prominent and frequent for children and adolescents, ingestion of PBDEs is most remarkable before age 20 (Li, Hughes, and Arnot 2021). On the other hand, adults have a longer history of exposure; continuous exposure and accumulation will also lead to high levels of PBDEs in their bodies, which causes the rebound after age 40. In addition, adults older than 60 may have a reduced hepatic biotransformation capacity, which means they are less efficient at eliminating BDE-47 from their bodies. However, unlike the PCBs case where contamination maintained a continuous descending trend over the investigated period, PBDE contamination was in a transition from pre-ban to post-ban periods. Articles with PBDEs can still be used by households for many years and circulate in the secondhand market after the ban (Stapleton et al. 2011) and continue to be a source of exposure for years. In addition, the half-life of PBDEs is shorter than that of PCBs (Grandjean et al. 2008), which makes it less cumulative in the body. As such, we observe a relatively stable level of PBDEs for the rest of adulthood, instead of the constant increase in PCB concentrations with age seen in Section 5.1. The rich information on such age effects would have been masked by a simple ascending linear pattern if linear regression was used.

Figure 4 also shows that BDE-47 concentration maintains a relatively high level for individuals with a family income to poverty guidelines lower than 2, with the only exception for the 2013/14 sampling cycle probably due to random errors in sampling. A plausible explanation may be that individuals with low income may have increased accessibility to PBDE-treated consumer goods such as foam furniture: the cheapest furniture may be flame retarded with PBDEs, while the newer and more expensive furniture may embed alternative flame retardants which are environmentally friendly (Zota, Adamkiewicz, and Morello-Frosch 2010). Our statistical analysis reveals this intriguing question and warrants in-depth investigation in the future, which may not be possible to reveal by studying the pooled biomonitoring data without regression.

6. Discussion

In this article, we developed a flexible regression-based framework to assess disparities in the U.S. population's exposure to two families of notorious environmental chemicals (i.e., PCBs and PBDEs) using the pooled biomonitoring measurements collected by NHANES from 2005 to 2016. A data augmentation procedure and penalized spline basis are used to strike a balance between model flexibility and complexity. As pooled biomonitoring data for a large population is becoming increasingly popular due to its cost-effectiveness, our method could provide practitioners with a valuable tool to explore the underlying associations when outcomes are subject to pooling, and increase the depth of analysis in related environmental health research.

Using pooled biomonitoring data from NHANES, we, for the first time, systematically investigate the disparities in exposure to PCBs and PBDEs between subpopulations belonging to different racial/ethnic groups, at different ages, with or without obesity, and with different family income levels. In particular, our findings show males have higher levels of exposure to both chemicals than females. For PCBs, Hispanic have the lowest exposure levels among all demographic groups, and both age and income are positively associated with the serum concentrations of PCBs. For PBDEs, Non-Hispanic White have the highest exposure levels among all race/ethnicity groups, and obese people have higher serum levels than non-obese people. In addition, both age and family income exhibit nearly monotonic associations with PCB concentrations but non-monotonic associations with PBDE concentrations. Since multiple demographic and socioeconomic factors are integrated into the same statistical analysis, this approach ensures adequate control for potential confounding factors and thus prevents possible spurious statistical associations. We further demonstrate the need for control of confounding by a comparison between our results and the ones obtained by ignoring potential confounding factors. We refer the readers to Section F of supplementary material for this comparison.

Notably, our derived trends match well with the existing mechanistic understandings of the main pathways responsible for human exposure to these compounds. As such, this approach is mechanistically sound from the perspective of environmental health. Results of our case study help understand the difference in exposure to these chemicals among various subpopulations, which facilitates the formulation and implementation of population-specific tailored, cost-effective policies to promote environmental justice. Finally, our results also indicate descending trends in exposures to both PCBs and PBDEs among the U.S. population, which demonstrates the effectiveness of regulations on the production and use of these chemicals and the responsiveness of human exposure to environmental contamination. In summary, our proposed approach enables the use of pooled biomonitoring data for environmentally relevant and meaningful findings.

Lastly, we conclude this article with a possible direction for future research. In this article, we have the response variables observed in pools. An interesting topic could be to investigate how adverse health outcomes (e.g., diabetes or organ dysfunction) are associated with exposure to PCBs and PBDEs by leveraging the NHANES pooled biomonitoring data. But we would

like to point out that, with covariates' measurements subject to pooling, regression analysis could be much more methodologically challenging.

Supplementary Materials

The supplementary materials contain a discussion of model identifiability and technical details for the posterior sampling. Additional figures and tables pertaining to simulation studies and real data analysis are also presented in the supplementary materials. Furthermore, all the codes and data required to replicate our research are included in the supplementary materials.

Acknowledgments

Funding

ORCID

Yan Liu ⬤ http://orcid.org/0000-0002-3787-5906
Dewei Wang ⬤ http://orcid.org/0000-0003-0822-8563
Li Li ⬤ http://orcid.org/0000-0002-5157-7366
Dingsheng Li ⬤ http://orcid.org/0000-0002-8432-4023

References

Abbasi, G., Li, L., and Breivik, K. (2019), "Global Historical Stocks and Emissions of PBDEs," *Environmental Science & Technology*, 53, 6330–6340. [1538]

Agricultural Research Service. (2018), "Food Patterns Equivalents Intakes from Food: Mean Amounts Consumed per Individual, by Gender and Age, in the United States, 2017–2018," U.S. Department of Agriculture. Available at *www.ars.usda.gov/nea/bhnrc/fsrg*. [1546]

Bergonzi, R., Specchia, C., Dinolfo, M., Tomasi, C., De Palma, G., Frusca, T., and Apostoli, P. (2009), "Distribution of Persistent Organochlorine Pollutants in Maternal and Foetal Tissues: Data from an Italian Polluted Urban Area," *Chemosphere*, 76, 747–754. [1545]

Bichteler, A., Wikoff, D. S., Loko, F., and Harris, M. A. (2017), "Estimating Serum Concentrations of Dioxin-like Compounds in the US Population Effective 2005–2006 and 2007–2008: A Multiple Imputation and Trending Approach Incorporating NHANES Pooled Sample Data," *Environment International*, 105, 112–125. [1539]

Breivik, K., Sweetman, A., Pacyna, J. M., and Jones, K. C. (2007), "Towards a Global Historical Emission Inventory for Selected pcb Congeners—A Mass Balance Approach: 3. An Update," *Science of the Total Environment*, 377, 296–307. [1538]

Caudill, S. P. (2012), "Use of Pooled Samples from the National Health and Nutrition Examination Survey," *Statistics in Medicine*, 31, 3269–3277. [1539,1541]

——— (2015), "Confidence Interval Estimation for Pooled-Sample Biomonitoring from a Complex Survey Design," *Environment International*, 85, 40–45. [1539]

Chang, C.-J., Terrell, M. L., Marcus, M., Marder, M. E., Panuwet, P., Ryan, P. B., Pearson, M., Barton, H., and Barr, D. B. (2020), "Serum Concentrations of Polybrominated Biphenyls (PBBs), Polychlorinated Biphenyls (PCBs) and Polybrominated Diphenyl Ethers (PBDEs) in the Michigan PBB Registry 40 Years after the PBB Contamination Incident," *Environment International*, 137, 105526. [1539]

Cordner, A., Mulcahy, M., and Brown, P. (2013), "Chemical Regulation on Fire: Rapid Policy Advances on Flame Retardants," *Environmental Science & Technology*, 47, 7067–7076. [1547]

Darmon, N., and Drewnowski, A. (2008), "Does Social Class Predict Diet Quality?" *The American Journal of Clinical Nutrition*, 87, 1107–1117. [1546]

Eilers, P. H., and Marx, B. D. (1996), "Flexible Smoothing with b-splines and Penalties," *Statistical Science*, 11, 89–102. [1541,1542]

Fantke, P., and Jolliet, O. (2016), "Life Cycle Human Health Impacts of 875 Pesticides," *The International Journal of Life Cycle Assessment*, 21, 722–733. [1538]

Gaylord, A., Osborne, G., Ghassabian, A., Malits, J., Attina, T., and Trasande, L. (2020), "Trends in Neurodevelopmental Disability Burden Due to Early Life Chemical Exposure in the USA from 2001 to 2016: A Population-based Disease Burden and Cost Analysis," *Molecular and Cellular Endocrinology*, 502, 110666. [1539]

Grandjean, P., Budtz-Jørgensen, E., Barr, D. B., Needham, L. L., Weihe, P., and Heinzow, B. (2008), "Elimination Half-Lives of Polychlorinated Biphenyl Congeners in Children," *Environmental Science & Technology*, 42, 6991–6996. [1548]

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Vol. 2), New York: Springer. [1540]

Lang, S., and Brezger, A. (2004), "Bayesian p-splines," *Journal of Computational and Graphical Statistics*, 13, 183–212. [1541,1542]

Li, L., Arnot, J. A., and Wania, F. (2018), "Revisiting the Contributions of Far-and Near-Field Routes to Aggregate Human Exposure to Polychlorinated Biphenyls (PCBs)," *Environmental Science & Technology*, 52, 6974–6984. [1538,1547]

Li, L., Hoang, C., Arnot, J. A., and Wania, F. (2020), "Clarifying Temporal Trend Variability in Human Biomonitoring of Polybrominated Diphenyl Ethers through Mechanistic Modeling," *Environmental Science & Technology*, 54, 166–175. [1538,1548]

Li, L., Hughes, L., and Arnot, J. A. (2021), "Addressing Uncertainty in Mouthing-Mediated Ingestion of Chemicals on Indoor Surfaces, Objects, and Dust," *Environment International*, 146, 106266. [1548]

Li, L., and Wania, F. (2017), "Mechanistic Pharmacokinetic Modeling of the Bioamplification of Persistent Lipophilic Organic Pollutants in Humans During Weight Loss," *Environmental Science & Technology*, 51, 5563–5571. [1545]

Li, Y., Graubard, B. I., and Korn, E. L. (2010), "Application of Nonparametric Quantile Regression to Body Mass Index Percentile Curves from Survey Data," *Statistics in Medicine*, 29, 558–572. [1540]

Liu, Y., McMahan, C., and Gallagher, C. (2017), "A General Framework for the Regression Analysis of Pooled Biomarker Assessments," *Statistics in Medicine*, 36, 2363–2377. [1539,1540,1541]

Mahaffey, K. R., Clickner, R. P., and Jeffries, R. A. (2009), "Adult Women's Blood Mercury Concentrations Vary Regionally in the United States: Association with Patterns of Fish Consumption (NHANES 1999–2004)," *Environmental Health Perspectives*, 117, 47–53. [1546]

Mitchell, E. M., Lyles, R. H., Manatunga, A. K., Danaher, M., Perkins, N. J., and Schisterman, E. F. (2014), "Regression for Skewed Biomarker Outcomes Subject to Pooling," *Biometrics*, 70, 202–211. [1539,1540,1541]

Mitchell, E. M., Lyles, R. H., and Schisterman, E. F. (2015), "Positing, Fitting, and Selecting Regression Models for Pooled Biomarker Data," *Statistics in Medicine*, 34, 2544–2558. [1539]

Moysich, K. B., Ambrosone, C. B., Mendola, P., Kostyniak, P. J., Greizerstein, H. B., Vena, J. E., Menezes, R. J., Swede, H., Shields, P. G., and Freudenheim, J. L. (2002), "Exposures Associated with Serum Organochlorine Levels among Postmenopausal Women from Western New York State," *American Journal of Industrial Medicine*, 41, 102–110. [1545]

Nguyen, V. K., Kahana, A., Heidt, J., Polemi, K., Kvasnicka, J., Jolliet, O., and Colacino, J. A. (2020), "A Comprehensive Analysis of Racial Disparities in Chemical Biomarker Concentrations in United States Women, 1999–2014," *Environment International*, 137, 105496. [1539]

Pfeffermann, D. (1993), "The Role of Sampling Weights When Modeling Survey Data," *International Statistical Review/Revue Internationale de Statistique*, 61, 317–337. [1540]

Quinn, C. L., and Wania, F. (2012), "Understanding Differences in the Body Burden–Age Relationships of Bioaccumulating Contaminants based on Population Cross Sections versus Individuals," *Environmental Health Perspectives*, 120, 554–559. [1547]

Sarcinelli, P. N., Pereira, A. C. S., Mesquita, S. A., Oliveira-Silva, J. J., Meyer, A., Menezes, M. A., Alves, S. R., Mattos, R. C., Moreira, J. C., and Wolff, M. (2003), "Dietary and Reproductive Determinants of Plasma Organochlorine Levels in Pregnant Women in Rio de Janeiro," *Environmental Research*, 91, 143–150. [1545]

Silver, E., Kaslow, J., Lee, D., Lee, S., Tan, M. L., Weis, E., and Ujihara, A. (2007), "Fish Consumption and Advisory Awareness among Low-Income Women in California's Sacramento–San Joaquin Delta," *Environmental Research*, 104, 410–419. [1546]

Sjödin, A., Jones, R. S., Caudill, S. P., Wong, L.-Y., Turner, W. E., and Calafat, A. M. (2014), "Polybrominated Diphenyl Ethers, Polychlorinated Biphenyls, and Persistent Pesticides in Serum from the National Health and Nutrition Examination Survey: 2003–2008," *Environmental Science & Technology*, 48, 753–760. [1539]

Sjödin, A., Jones, R. S., Wong, L.-Y., Caudill, S. P., and Calafat, A. M. (2019), "Polybrominated Diphenyl Ethers and Biphenyl in Serum: Time Trend Study from the National Health and Nutrition Examination Survey for Years 2005/06 through 2013/14," *Environmental Science & Technology*, 53, 6018–6024. [1539,1540]

Stapleton, H., Klosterhaus, S., Keller, A., Ferguson, P., VanBergen, S., Cooper, E., Webster, T., and Blum, A. (2011), "Identification of Flame Retardants in Polyurethane Foam Collected from Baby Products," *Environmental Science & Technology*, 45, 5323–5331. [1548]

Undeman, E., Brown, T. N., Wania, F., and McLachlan, M. S. (2010), "Susceptibility of Human Populations to Environmental Exposure to Organic Contaminants," *Environmental Science & Technology*, 44, 6249–6255. [1545]

Wang, S., Romanak, K. A., Hendryx, M., Salamova, A., and Venier, M. (2019), "Association between Thyroid Function and Exposures to Brominated and Organophosphate Flame Retardants in Rural Central Appalachia," *Environmental Science & Technology*, 54, 325–334. [1539]

Xu, C., Su, X., Xu, Y., Ma, S., Duan, W., and Mo, X. (2019), "Exploring the Associations of Serum Concentrations of PCBs, PCDDs, and PCDFs with Walking Speed in the US General Population: Beyond Standard Linear Models," *Environmental Research*, 178, 108666. [1539]

Zota, A. R., Adamkiewicz, G., and Morello-Frosch, R. A. (2010), "Are PBDEs an Environmental Equity Concern? Exposure Ddisparities by Socioeconomic Status," *Environmental Science & Technology*, 44, 5691–5692. [1548]

Zota, A. R., Rudel, R., Morello-Frosch, R. A., and Brody, J. G. (2008), "Elevated House Dust and Serum Concentrations of PBDEs in California: Unintended Consequences of Furniture Flammability Standards? *Environmental Science & Technology*, 42, 8158–8164. [1539,1548]