HHS Public Access

Author manuscript

Ecotoxicol Environ Saf. Author manuscript; available in PMC 2023 March 14.

Published in final edited form as:

Ecotoxicol Environ Saf. 2022 December 15; 248: 114309. doi:10.1016/j.ecoenv.2022.114309.

Chemical mixture exposure patterns and obesity among U.S. adults in NHANES 2005–2012

Yuqing Zhang[a,1], **Xu Wang**[b,1], **Xu Yang**[c,d], **Qi Hu**[c,d], **Kuldeep Chawla**[e,f], **Bo Hang**[g], **Jian-Hua Mao**[f,g], **Antoine M. Snijders**[f,g], **Hang Chang**[f,g,*], **Yankai Xia**[c,d,**]

[a]Department of Obstetrics and Gynecology, Women ‘s Hospital of Nanjing Medical University, Nanjing Maternity and Child Health Care Hospital, Nanjing 210004, China

[b]Department of endocrinology, Children ‘s Hospital of Nanjing Medical University, Nanjing 210008, China

[c]State Key Laboratory of Reproductive Medicine, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing 211166, China

[d]Key Laboratory of Modern Toxicology of Ministry of Education, School of Public Health, Nanjing Medical University, Nanjing 211166, China

[e]Scientific Computing Group, Information Technology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

[f]Berkeley Biomedical Data Science Center, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

[g]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

Abstract

[*]Corresponding author at: Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [**]Correspondence to: State Key Laboratory of Reproductive Medicine, Center for Global Health, School of Public Health, Nanjing Medical University, No.101 Longmian Road, Nanjing 211166, China. hchang@lbl.gov (H. Chang), yankaixia@njmu.edu.cn (Y. Xia).

[1]These authors contributed equally to this work.

CRediT authorship contribution statement

Yuqing Zhang: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing - original draft. **Xu Wang:** Formal analysis, Validation, Writing - review & editing. **Xu Yang:** Writing - review & editing. **Qi Hu:** Writing - review & editing. **Kuldeep Chawla:** Resources, Software. **Bo Hang:** Supervision, Validation, Writing - review & editing. **Jian-Hua Maov:** Conceptualization, Methodology, Project administration, Supervision, Writing - review & editing. **Antoine M. Snijders:** Conceptualization, Methodology, Project administration, Supervision, Writing - review & editing. **Hang Chang:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing - review & editing. **Yankai Xia:** Funding acquisition, Project administration, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ecoenv.2022.114309.

Background: The effect of chemical exposure on obesity has raised great concerns. Real-world chemical exposure always imposes mixture impacts, however their exposure patterns and the corresponding associations with obesity have not been fully evaluated.

Objectives: To discover obesity-related mixed chemical exposure patterns in the general U.S. population.

Methods: Sparse Decompositional Regression (SDR), a model adapted from sparse representation learning technique, was developed to identify exposure patterns of chemical mixtures with exclusion (non-targeted model) and inclusion (targeted model) of health outcomes. We assessed the relationships between the identified chemical mixture patterns and obesity-related indexes. We also conducted a comprehensive evaluation of this SDR model by comparing to the existing models, including generalized linear regression model (GLM), principal component analysis (PCA), and Bayesian kernel machine regression (BKMR).

Results: Eight core exposure patterns were identified using the non-targeted SDR model. Patterns of high levels of MEP, high levels of naphthalene metabolites (EOH-Nap), and a pattern of high exposure levels of MCOP, MCNP, and MCPP were positively associated with obesity. Patterns of high levels of BP3, and a pattern of higher mixed levels of MPB, PPB, and MEP were found to have negative associations. Associations were strengthened using the targeted SDR model. In the single chemical analysis by GLM, BP3, MBP, PPB, MCOP, and MCNP showed significant associations with obesity or body indexes. The SDR model exceeded the performance of PCA in pattern identification. Both SDR and BKMR identified a positive contribution of EOH-Nap and MCOP, as well as a negative contribution of BP3 and PPB to obesity.

Conclusion: Our study identified five core exposure patterns of chemical mixtures significantly associated with obesity using the newly developed SDR model. The SDR model could open a new avenue for assessing health effects of environmental mixture contaminants.

Keywords

Sparse Decompositional Regression Model; Exposure mixtures; Exposure pattern; Obesity

1. Introduction

Obesity is a common public health concern that affects 13% of adults worldwide and imposes a major disease burden (Afshin et al., 2017). Except for genetic background and lifestyle changes (Lin et al., 2019; Ling and Ronn, 2019), the potential influence of environmental chemicals has raised great interest (Thayer et al., 2012). Fully evaluating the chemical exposure levels of obese people and identifying chemical mixture signals/patterns associated with obesity are critical for its prevention. Previous studies have demonstrated that certain chemicals disrupt endogenous hormonal homeostasis and are involved in the programming of adipogenesis, which may lead to weight gain (Heindel et al., 2017). Environmental phenols, parabens, polycyclic aromatic hydrocarbons (PAHs), and phthalates have potential endocrine disruption effects, and their associations with obesity have been reported in previous studies (Liu et al., 2017; Scinicariello and Buser, 2014a; Stahlhut et al., 2007). However, the classical methods to assess the health effect of chemicals mainly focus on individual chemical or limited types of environmental contaminants. An

environment-wide association study (EWAS) used data from NHANES to explore the associations between a panel of exposures and obesity in adolescents (Haddad et al., 2022), which successfully extended the scope of exposures without exploring the combined effect among chemicals. In summary, the real-world complex chemical mixture exposure and their synergistic health impact remain inadequately addressed. Such effects can result from the similarity of exposure sources and metabolite pathways (Kim et al., 2017; Shaheen et al., 2016). Although some advanced statistical approaches have been developed to explore the health effects of mixed exposure (Lazarevic et al., 2019; Taylor et al., 2016), identifying the core exposure patterns (fundamental chemical mixture signals/combination) from high-dimensional exposure data and thereafter estimating the health effect of these patterns haven't received enough attention.

Sparse coding technique, a newly developed representation learning technique initially applied in feature learning in the computer science domain, was adapted as Sparse Decompositional Regression (SDR) model to identify exposure patterns of study population. After learning sets of overcomplete features, SDR uses an L1 regularized optimization to obtain and highlight sparse coefficients to represent the original raw data. The sparse coding technique, with its capability of efficient and effective representation learning (Kavukcuoglu et al., 2009; Lee et al., 2006; Rozell et al., 2008), allows the capture of underlying signals/patterns from the raw data (Chang et al., 2018; Chang et al., 2021; Chang et al., 2015; Liu et al., 2022; Mao et al., 2022). In this study, obesity-related core exposure patterns composed of potential obesity-related chemicals, including phenols, parabens, PAHs, and phthalates (Hatch et al., 2008; Liu et al., 2017; Scinicariello and Buser, 2014a; Wu et al., 2019), were discovered by SDR model in a U.S. national representative population data from National Health and Nutrition Examination Survey (NHANES) 2005–2012. In addition, the SDR model allows the incorporation of targeted health outcomes (e.g., obesity) during representation learning, which helps enforce the relevance of derived core exposure patterns with health outcomes. We also employed evaluation of SDR model through comparing to the most commonly used model (generalized linear regression model (GLM)), another pattern identification model (principal components analysis (PCA)), and Bayesian kernel machine regression (BKMR) to determine the performance of SDR model. Furthermore, the association of chemical exposure patterns with obesity was evaluated. SDR model will provide a new avenue for the efficient and effective discovery and assessment of the impact of chemical exposure patterns on health.

2. Methods

2.1. Study population

All of the study population was from NHANES, a cross-sectional study conducted every two years to assess the nutrition and health status of a representative U.S. population (Zipf et al., 2013). Subsamples of the participants were measured with different combinations of environmental chemicals. To extend the scope of exposure chemicals, we used data from NHANES 2005–2012 and all the participants provided informed consent. The survey protocol was approved by the National Center for Health Statistics Research Ethics Review Board. Information such as demographic characteristics, dietary intake, body examination,

and biospecimen measurement was recorded comprehensively. The detailed patient selection strategy was illustrated in Fig. S1. In total, 40790 participants took part in NHANES 2005–2012. Due to the study design, a representative subsample was available of the targeted chemicals (n = 10769 left). We excluded those with missing values of the chemical concentration or anthropometric data (including body weight, height, and waist circumference) (n = 1307 excluded). Those who were less than 20 years old, pregnant, previously diagnosed with cancer or malignancy, or underweight (BMI≤18.5) were excluded (n = 3985 excluded). Participants with missing data on covariates were also excluded from the final analysis (n = 664 excluded). The final sample size was 4813 participants.

2.2. Measurement of chemicals in urine

Spot urine samples were collected at mobile examination centers. The concentrations of environmental phenols, parabens, phthalates metabolites, and PAHs metabolites were analyzed using previously reported methods (CDC, 2018b). Chemicals with less than 90% detected were excluded from our analysis to reduce the evaluation bias derived from simple substitution of the values under the lower limit of detection (LOD) (Lubin et al., 2004). Specifically, triclosan, butyl paraben, ethyl paraben, mono-n-methyl phthalate, mono-isononyl phthalate, and mono-(2-ethyl)-hexyl phthalate, with detection frequencies ranging from 22.3% to 77.0% were excluded (Table S2). We further added up the molar concentration of metabolites from the same parent compound. More details are provided in the Supplementary Material. Finally, fifteen chemicals were retained in the analysis (including PAH: naphthalene metabolites (ΣOH-Nap), fluorine metabolites (ΣOH-Flu), phenanthrene metabolites (ΣOH-Phe), 1-hydroxypyrene (1-OH-Pyr); phenol: benzophenone-3 (BP3) and bisphenol A (BPA); paraben: methyl paraben (MPB) and propyl paraben (PPB), phthalate: mono(carboxynonyl) phthalate (MCNP), mono(carboxyoctyl) phthalate (MCOP), diethylhexyl phthalate metabolites (ΣDEHP), dibutyl phthalate metabolites (ΣDBP), mono-(3-carboxypropyl) phthalate (MCPP), mono-ethyl phthalate (MEP), and mono-benzyl phthalate (MBzP)).

2.3. Outcomes and co-variation assessment

Anthropometric indexes were measured at the mobile examination centers. Participants with a BMI of 30 kg/m^2 or higher were general obese according to WHO reference (WHO, 2017). Abdominal obesity was classified by absolute waist circumference (≥102 cm in men and ≥88 cm in women).

Information on demography, nutrition intake, habits, and medical status was collected by direct interview. We considered age, sex, race, educational level, family income to poverty ratio (PIR), smoking status, and total caloric intake status per day as potential covariates as previously described (Zhang et al., 2019). In addition, creatinine concentration was natural logarithm transformed and considered a covariate to account for urinary dilution (Barr et al., 2005).

2.4. Sparse decompositional regression (SDR) model

To explore the higher-level dependencies as well as the underlying regularities in the raw chemical exposure data, we combined the concept of sparse feature learning and feed-

forward regression into a unified framework, namely Sparse Decompositional Regression (SDR), to realize efficient and effective mining of combinatorial patterns related to the mixture of chemical exposures, as well as their association with health outcomes. Given $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_N] \in \mathbf{R}^{m \times N}$ as a set of samples (N) with a combination of chemical exposure levels (i.e., raw chemical exposure profile with m chemicals; e.g., m=15 in NHANES dataset) and **L** as the health outcome (e.g., BMI), the formulation of SDR model is defined as follows (Fig. 1B).

$$min_{A(*),\mathbf{D},\mathbf{Z},\mathbf{W},\mathbf{G}} \| \mathbf{X} - \mathbf{DZ} \|_F^2 + \| \mathbf{Z} - \mathbf{G}\sigma(\mathbf{WX}) \|_F^2 + \lambda \boldsymbol{I} \| \mathbf{Z} \|_1 + \| \mathbf{AZ} - \mathbf{L} \|_F^2$$
$$s.t. \| \boldsymbol{d}_i \|_2^2 = \boldsymbol{I}, \forall \boldsymbol{i} = \mathbf{1}, \ldots, \boldsymbol{h}; (*) target - only\ item$$

A detailed description of the items in the formulation can be found in Fig. 1B and Supplementary Material. To avoid or reduce potential false predictions where chemical concentrations are varied in exposure levels, feature transformation and scaling were conducted in data preprocessing. The concentration of the chemicals was right-skewed, and a lntransformation was used to improve the normality. Afterward, minmax normalization was applied for feature scaling, and **X**, the raw chemical exposure profile was a transformed and scaled dataset. The first constraint:$\| \mathbf{X} - \mathbf{DZ} \|_F^2$, penalizes the reconstruction error of raw chemical exposure with mixture chemical exposure patterns (**D**) and the corresponding sparse feature matrix (**Z**), which helps minimize the loss of information; the second constraint:$\| \mathbf{Z} - \mathbf{G}\sigma(\mathbf{WX}) \|_F^2$, penalizes the approximation error of sparse feature matrix (**Z**) with the auto-encoder, which helps improve the accuracy of sparse feature matrix approximation for new participants; the third constraint: $\| \mathbf{Z} \|_1$ penalizes the sparsity of the sparse feature matrix, which helps ensure the utilization/activation of dominant mixture chemical patterns during the learning process, and only sparse patterns were used for data reconstruction; and the last term (target-only): $\| \mathbf{AZ} - \mathbf{L} \|_F^2$ penalizes the approximation error of outcome (**L**) with sparse feature matrix (**Z**), which helps refine the learning process toward improved outcome association. A step-wised joint minimization of the above equation leads to a highly efficient and effective solution for mixture chemical exposure pattern discovery (Fig. 1B and Supplementary Material).

To optimize the number (***h***) of mixture chemical exposure patterns while stabilizing the reconstruction of the raw chemical exposure profile (**X**) based on mixture chemical exposure patterns (**D**), we first bootstrapped the SDR model 100 times with samples randomly selected per iteration. Consequently, it led to the discovery of 100 sets of chemical exposure patterns, $[\mathbf{D}_1, \ldots, \mathbf{D}_{100}]$, and the corresponding sparse feature matrix$[\mathbf{Z}_1, \ldots, \mathbf{Z}_{100}]$, where $\mathbf{D}_i = [\mathbf{d}_1^i, \ldots, \mathbf{d}_h^i]$ and $\mathbf{Z}_i = [\mathbf{Z}_1^i, \ldots, \mathbf{Z}_N^i]$. Then, we utilized the consensus clustering strategy with all preidentified chemical exposure patterns ($[\mathbf{D}_1,...,\mathbf{D}_{100}]$) to obtain the core chemical exposure patterns (abbreviated as core patterns) as $\widehat{\mathbf{D}} = \left[\widehat{\mathbf{d}}_1, \ldots, \widehat{\mathbf{d}}_k\right]$, where ***k*** is the number of clusters optimized by consensus clustering approach (helping reduce redundant exposure patterns, detailed information on consensus clustering is provided in Supplementary Material), and $\widehat{\mathbf{d}}_i$ is core chemical exposure pattern ***i*** defined as the median chemical exposure patterns within cluster ***i***. In the present study, the chemicals with higher

relative abundance in each core pattern were used to represent the corresponding core pattern. Last, we fixed the dictionary as the core patterns ($\widehat{\mathbf{D}}$), and updated <**Z, G, W**> (non-targeted) or <**Z, G, W, A**> (targeted) through our SDR model. The SDR parameters, including the number of core chemical exposure patterns (k), dictionary size (*h*, referring to the mixture chemical exposure pattern number in each bootstrap, as shown in Fig. 1A), and sparsity (referring to the number of mixture chemical exposure patterns to be used for raw chemical exposure profile reconstruction) of this study were optimized with a stepwise strategy (see Supplementary Material for details).

Pairwise *Pearson* correlation was utilized to measure the similarity between core patterns learned from non-targeted and targeted SDR models, and for each pairwise comparison, the patterns were considered consistent when *Pearson's* correlation (r) > 0.9 and $p < 0.05$. The importance of individual chemicals in each core pattern was evaluated by a perturbation strategy to identify the driving components in the associations between core patterns and health as described in Supplementary Material.

Finally, we assessed the association of core mixture chemical exposure patterns ($\widehat{\mathbf{D}}$) with health outcomes through regression analysis with adjustment for other confounding factors (Fig. 1B and C).

2.5. Statistical analysis

Comparisons of continuous and categorical variables between groups were conducted by t-test and chi-square tests, respectively. Concentrations of chemicals were ln-transformed to improve the skewness. *Pearson* correlation analysis was carried out to estimate the relationship between chemicals.

As a further evaluation, we also compared our newly developed SDR model with GLM, PCA, and BKMR.

2.5.1. Generalized linear regression model—Generalized linear regression model was employed to assess the association between single chemical exposure and obesity-related outcomes (i.e., logistic regression for general obesity and abdominal obesity, multivariable linear regression for BMI and waist circumference). The participants were divided into tertile groups based on chemical concentration, and the association between individual chemicals and obesity was assessed by comparing the second and third tertiles to the first tertile of a chemical's concentration. All multivariable analyses were adjusted for gender, ethnicity, educational levels, age, family income-to-poverty ratio, smoking status, energy intake levels, and ln-transformed creatinine.

2.5.2. Principal components analysis (PCA)—PCA is a dimension reduction method in which uncorrelated principal components (PCs) are created as the linear combination of the highly correlated components Wold et al., 1987). The first few PCs, with a major contribution to the variance, were used in the regression analysis. Varimax rotation was used to make a smaller number of chemicals of high factor loadings and the rest of low factor loadings. Principal components with eigenvalues larger than one were retained in the following analysis.

2.5.3. Bayesian kernel machine regression (BKMR) model—BKMR model can fit the high-dimensional nonlinear and nonadditive exposure reaction function and obtain health effects of mixed exposures on outcomes by comparing all of the chemicals fixed at a certain percentile with all of them at median levels (Bobb et al., 2018, 2015). The posterior inclusion probabilities (PIPs), which represent the probability that a particular chemical was included in the model, were calculated with 50,000 iterations. Fixing all the other chemicals at median levels, the dose-response curve of chemicals with outcome was estimated.

Sensitivity analysis was conducted with further adjusting for physical activity. Since standard Global Physical Activity Questionnaire (GPAQ) was not available in all the NHANES 2005–2012 cycles, we defined physical active as doing any vigorous or moderate activities for at least 10 min in the past 30 days (NHANES 2005–2006) or in a typical week (NHANES 2007–2012).

Due to the complex sampling design of NHANES, the sample weights are usually applied in analysis using NHANES data (CDC, 2018a; Haddad et al., 2022; Patel et al., 2010). However, weighted estimation could introduce over-adjustment bias when variables used for calculating weights are already adjusted as covariates in regression analysis. Therefore, the unweighted estimation was used in all the models throughout this study (Blount et al., 2006; Gelman, 2007; Graubard and Korn, 1999). A P value of 0.05 was set as the significant criterion. SDR was conducted in Matlab (version R2018b), and all the other analyses were performed with R (3.5.1). BKMR was provided by the open-source R package “bkmr” (version 0.2.0).

3. Results

3.1. Population characteristics

This study consisted of 4813 participants, 37.8% of whom were classified as generally obese and 55.8% were classified as having abdominal obesity (Table S1). The average age was significantly higher in the obesity group and females were overrepresented. We also found the distributions of race, educational levels, family income to poverty ratio, and smoking status were significantly different between groups (Table S1).

3.2. Urinary chemical concentrations

Twenty-three out of 29 chemicals were detected in ≥ 90% of study participants (Table S2). The molar concentrations of chemicals derived from the same parent compound were summed (Table S3), resulting in 15 exposure biomarkers for downstream analysis. The concentrations of all 15 exposure biomarkers showed right-skewed distributions (Table S3). Urinary concentrations of BP3, MPB, metabolites of DEP (MEP), and the metabolites of naphthalene (ΣOH-Nap) were proportionally most abundant in phenols, parabens, phthalates, and PAHs, respectively. The *Pearson* correlation of these chemicals showed significant positive pairwise correlations (r ranging from 0.02 to 0.90) (Fig. S2). Chemicals with similar structures exhibited higher correlations, particularly for metabolites of PAHs and parabens.

3.3. Core pattern identification and its association with obesity

With preoptimized parameters of the SDR model (Fig. S3 and S4) and a predetermined number of core patterns, the core patterns were then discovered via the non-targeted SDR model followed by consensus clustering, where no health outcome was involved in the model construction. Specifically, 1600 individual patterns were obtained after bootstrapping 100 times in non-targeted SDR model with dictionary size and sparsity set to 16 and 4, respectively. Then, 8 clusters corresponding to 8 different core patterns were identified using consensus clustering based on all 1600 patterns. The composition and distribution of each chemical in the 8 core-pattern-related clusters are shown in Fig. S5. Afterwards, the expression level of each chemical in the core pattern is defined as the median expression level of the chemical in the corresponding cluster. Last, the SDR model is fine-tuned with fixed core patterns preobtained from previous steps. In addition, correlation analysis of the expression of these eight patterns (Fig. S6) revealed low and moderate correlations between patterns ($0.01 < |r| < 0.51$, $p < 0.05$).

The association of each pattern with body index-related outcomes was estimated by adjusting for covariates, and five patterns were significantly associated with obesity and body indexes. Core pattern 1 (high expression of ΣOH-Nap), core pattern 4 (high expression of MEP), and core pattern 7 (high expression of MCNP, MCOP, and MCPP) were positively associated with obesity and body indexes ($p < 0.05$). Core pattern 2 (high expression of BP3) and core pattern 5 (high expression of MPB, PPB, and MEP) were negatively associated with obesity and body indexes ($p < 0.05$) (Table 1).

To further improve the association between core patterns and health outcomes, we incorporate health outcomes (i.e., general obesity, BMI, abdominal obesity, and waist circumference) in pattern discovery using the targeted SDR model. The majority of core patterns in the non-targeted and targeted models were significantly correlated (*Pearson* correlation > 0.9 and p value<0.05) (Fig. 2), which means that the exposure data itself makes the greatest contribution to core pattern identification. Interestingly, we observed differences between the two models in association with health outcomes (Fig. 2, Fig. S7, Table S4, and Table S5). Most of the associations between core patterns and outcomes were strengthened. For example, five core patterns were found to be associated with general obesity in the non-targeted SDR model, while the associations of four of the core patterns were strengthened (larger OR in positive association and smaller OR in negative association) in targeted SDR model, and one additional association was found in the targeted SDR model. The explanation of these findings is due to that the targeted model was forced to identify core patterns more associated with outcomes.

3.4. Individual chemical importance score in core patterns

To identify the most influential chemicals in each core pattern, the percentage of the change in the regression coefficient (Δp) and importance scores (*IS*) were analyzed. In general, the chemicals with high relative abundance in pattern composition were those with a high percentage of the change in regression coefficient and importance scores (Fig. S8–S11). For example, the core pattern with high levels of MEP (Core pattern 4) was found to be positively associated with obesity, and MEP was assigned a high importance score,

indicating that MEP was the largest contributor to health outcomes in this exposure pattern (Fig. S8–S11). However, another core pattern characterized by high levels of MPB, PPB, and MEP (Core pattern 5) was negatively associated with obesity, and MPB and PPB, but not MEP, was assigned a large importance score, indicating that MPB and PPB made the largest contribution to the negative association while MEP made no or a neglectable contribution (Fig. S8–S11). The importance score was also refined in the targeted SDR model. The targeted SDR model of abdominal obesity didn't identify the core pattern with high levels ΣOH-Nap (Core pattern 1). However, ΣOH-Nap was assigned a relatively high importance score (*IS*=0.261) in the core pattern characterized by high levels of MEP (Core pattern 4) (Fig. S9).

3.5. Evaluation of SDR model by comparing to classic models

We also applied GLM, PCA, and BKMR to explore the association between the fifteen chemicals and obesity-related outcomes. GLM is the most classic model used in environmental epidemiology studies and mainly focuses on individual chemical effects. Similar to the SDR model, PCA and BKMR enable the assessment of the association of mixed exposure on health outcomes.

In multivariate logistic and linear regression analysis, the chemicals were grouped based on tertiles, and BP3, MBP, PPB, and MCOP showed significant associations (after adjusting for all the covariates) with obesity and body indexes in the highest group (Table S6–S7). Also, MCNP revealed a positive association with body indexes.

Using PCA, we identified four important principal components that explained 71.6% of the variance. We conducted logistic and linear regression to evaluate the association of PCs and obesity and body indexes, respectively. We found that PC 4, mainly composed of MPB, PPB, and BP3, was significantly associated with both obesity and body indexes; PC 3, mainly composed of MCOP, MCPP, and MCNP, was significantly associated with abdominal obesity and body indexes (Table 2).

According to the BKMR models, the joint effects of mixed chemical exposure on general obesity, BMI, abdominal obesity, and waist circumference showed negative trends (Fig. S12). The probabilities of inclusion (PIPs) derived from BKMR models are summarized in Table S8. The PIP is a ranking measurement to see to what extent the data favors the inclusion of a variable in the regression. PPB, BP3, MCOP, 1-OH-Pyr, and ΣOH-Nap contributed the most to the association of joint chemicals exposure with obesity. We also estimated the exposure-response functions of each chemical with outcomes (Fig. 3). We observed that ΣOH-Nap and MCOP had positive exposure-response associations with outcomes when all the other chemicals were fixed at their median levels, while BP3, PPB, and 1-OH-Pyr showed negative associations.

Table 3 summarizes the results obtained using the different models. Consistent with the *Pearson* correlation results (Supplementary material Fig. S2), both PCA and SDR are capable of capturing underlying patterns from the data. The SDR model not only provides consistent and robust discoveries but also identifies associations that are beyond the scope of

PCA (Table 3), which leads to improved predictive power toward health outcomes (i.e., AUC and RSME, Fig. S13, $p < 0.001$).

3.6. Sensitivity analysis with further adjusting for physical activity

After further adjusting for physical activity, the results of non-targeted and targeted SDR models remained stable (Table S9 and S10). For the comparison between two pattern identification models (SDR and PCA, results of PCA are shown in Table S11), SDR also showed improved predictive power toward health outcomes (i.e., AUC and RSME, Fig. S14, $p < 0.001$).

4. Discussion

We identified eight representative core exposure patterns in the general U.S. population by non-targeted SDR model. The exposure patterns of high expression of ΣOH-Nap, MEP, and a high coexpression pattern of MCNP, MCOP, and MCPP were positively associated with obesity, while the patterns of high expression of BP3 and coexpression of MPB, PPB, and MEP were negatively associated with obesity and body indexes. In the targeted SDR model, their associations were strengthened. We also explored three commonly used models, including the generalized linear regression model, PCA, and BKMR. The associations of BP3, MPB, PPB, MCOP, MCNP, and ΣOH-Nap with obesity were consistent in both SDR model and classic models, while the association of MEP with obesity was specifically identified by SDR model only.

Humans are exposed to various kinds of environmental chemicals; however, the identification of chemical exposure patterns and their association with human health remains largely unexplored and requires urgent investigation in terms of both technology development and deployment. Existing models used in multiagent studies mainly focus on the discovery of the most relevant chemicals and their combinatorial effects on health outcomes (Lazarevic et al., 2019; Stafoggia et al., 2017), ignoring potential contributions of coexposure patterns in the general population. Principal component analysis (PCA) is a classic model attempting to identify and estimate the effect of chemical patterns on outcomes (Wold et al., 1987). By reducing the dimension of chemical exposure space based on the first few principal components computed from chemical exposure data while preserving the majority of the chemical data variation, PCA leads to the most variation-preserving exposure patterns that are not necessarily interpretable or related to the outcomes. In this study, we applied the SDR model for identifying multivariate core exposure patterns by mining intrinsic patterns from raw data (non-target model) and enabling the outcome incorporation during pattern discovery (targeted model). As shown in Fig. S3 and S4, the number of core patterns is largely independent of parameter settings, which is due to the capability and robustness of SDR in discovering the underlying characteristics of the data. Our study suggests that the general population of NHANES 2005–2012 was exposed to eight core exposure patterns. In non-targeted and targeted SDR models, the core exposure patterns have similar profiles, which indicates that the exposure data itself makes the greatest contribution to core pattern identification. By penalizing the approximation error of outcome

(L) with sparse feature matrix (Z), the core patterns identified in targeted model are refined toward improved outcome association.

Different from dimension-reduction in PCA, SDR falls into the category of sparse representation learning, which learns sets of overcomplete features to reconstruct the original data (Mairal et al., 2008; Rubinstein et al., 2010). In detail, the SDR model first increases the dimensionality of the chemical exposure space to improve the representation power through sparse coding and to discover underlying exposure patterns, which contributes to the sparse reconstruction of the original exposure space without linear restrictions of relationships among chemical components. Next, consensus clustering on the underlying exposure patterns, learned with bootstrapping strategy, leads to the discovery of robust core patterns minimizing redundancy and the risk of missing core exposure patterns due to the random nature of sampling. In PCA, the exposure status of all the participants could be represented by the combination of the same patterns, and each pattern was a combination of all the chemicals. In SDR, we assume the exposure status of the study population is represented by sparse exposure patterns and each individual is a combination of a limited number of patterns. Under this condition, the patterns identified in SDR are more precise in representing the exposure status of each individual. Consistent with the *Pearson* correlation results (Fig. S2), both PCA and SDR are capable of capturing underlying patterns from the data. However, through extensive experimental evaluation, the SDR model exceeds the performance of PCA in the identification of core exposure patterns by improving specificity, especially among chemicals of highly correlated exposure levels. PCA only provided limited combinations of chemicals (the number of PCs ≤ the number of chemicals) and tended to group chemicals from the same class together, which potentially resulted in weakened/distorted associations. For example, in our study, the PAH metabolites were grouped in PC1, which consequently weakened the previously reported association of ΣOH-Nap with obesity (Bushnik et al., 2019; Scinicariello and Buser, 2014b; Zhang et al., 2019), but the significant contribution of ΣOH-Nap to the positive association between core pattern 1 and obesity was observed using the SDR model and BKMR. Similarly, the association between MEP and obesity in PC2 was also weakened in PCA but identified using the SDR model. The positive association of MEP with obesity has been reported (Hatch et al., 2008; Stahlhut et al., 2007). As a result, compared with the patterns derived from PCA, the core patterns learned from SDR demonstrated significantly improved predictive power toward health outcomes (i.e., AUC and RSME, Fig. S11, $p < 0.001$).

The SDR model provides a mechanism to evaluate the importance of individual chemicals in the association of core exposure patterns and outcomes. Chemicals with high importance scores were considered the driver components in the associations between core patterns and health. In general, the chemicals with a high importance score are those with high relative abundance within each core pattern composition, but the importance score is much more precise in evaluating their contributions. In our study, the core pattern highly expressed of MBP, PPB, and MEP was found to be negatively associated with obesity, and MEP was the third most highly expressed chemical in the core pattern. However, the importance score of MEP is much smaller than MBP and PPB, which indicates MEP had no or a neglectable contribution to the negative association. The BKMR model also performed well in evaluating the importance of chemicals and their dose-response association with health.

In this study, the positive contribution of ΣOH-Nap and MCOP, as well as the negative contribution of BP3 and PPB on obesity were identified in both SDR and BKMR models.

Different approaches are best suited to answer different questions, and they all have limitations. Specifically, GLM ignores the fact that chemicals are exposed simultaneously and the joint effect is not simply equal to the sum of all the effects (Kim, 2019). Consequently, GLM could not evaluate the joint effect of chemical mixtures, and results obtained with GLM suffer from distortion introduced by chemical collinearity (Marill, 2004). The PCA model is limited by linear function assumptions to identify linear subspace feature extractors, which doesn't necessarily represent the real exposure status (Chin and Suter, 2007). BKMR estimates the whole effect of chemical mixtures on health outcomes with each chemical increased by one unit, where the whole effect of chemical mixtures with both high and low exposed chemicals could not be estimated (Bobb et al., 2015). The major limitation of SDR originates from its design/focus on the combinatorial impact evaluation of multiple chemical exposures; as a result, it leaves the detailed interaction information among these chemicals from the mixed patterns unassessed. In addition, the limitations of this study include (1) the chemicals explored in this study are limited, and the concept of exposome could be deployed to better understand the origin of diseases and extend the scope of exposures from environmental chemicals to all the nongenetic factors, including synthetic chemicals, dietary factors, physiology status, physical activity, and the corresponding biological responses (Haddad et al., 2019; Vermeulen et al., 2020; Wild, 2005, 2012); (2) we excluded chemicals with detection frequency less than 90% and simply substituted the values under LOD, which could leave out important chemicals; (3) other covariates, such as medicine usage and surgery, are important factors associated with obesity but were not included as the covariates since they are not available in all the NHANES 2005–2012 cycles, which could be considered in future studies with improved cohort; (4) the cross-sectional study design couldn' t assess the causal relationship between chemicals and exposure. To overcome these limitations, prospective cohort studies with extended exposure are warranted to explore the non-genetic origin of obesity.

In summary, we identified five exposure patterns/signals associated with obesity including the positive association of ΣOH-Nap, MEP, and the combination of MCNP, MCOP, and MCPP, as well as the negative association of BP3, and the combination of MPB, PPB, and MEP. Our newly developed SDR model undoubtedly offers an improved capability and novel solution, specifically in mixture chemical pattern discovery as well as the evaluation of their combinatorial impact on health outcomes. With a comprehensive evaluation of other models, we identified the contribution of ΣOH-Nap, MCOP, BP3, MPB, and PPB to the associations with obesity. Considering the limitation of the cross-sectional study design, prospective cohort studies are required to explore the risk of obesity after chemical pattern exposure. Nevertheless, we believe that SDR model could open a new avenue for assessing health effects of environmental mixture contaminants.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Cancer Institute (NCI) at the National Institutes of Health (NIH) (grant no. R01CA184476), China-U. S.Program for Biomedical Collaborative Research (NSFC-NIH) (grant no. 81961128022), U.S.-China Program for Biomedical Collaborative Research NIEHS (grant no. R01ES031322) and National Natural Science Foundation of China (grant no. 82103793).

List of abbreviations:

1-OH-Pyr	1-hydroxypyrene
BKMR	Bayesian kernel machine regression
BP3	benzophenone-3
BPA	bisphenol A
BUP	butyl paraben
DBP	dibutyl phthalate
DEHP	diethylhexyl phthalate
EPB	ethyl paraben
GLM	generalized linear regression model
LOD	lower limit of detection
MBP	Mono-n-butyl phthalate
MBzP	Mono-benzyl phthalate
MCNP	Mono(carboxynonyl) phthalate
MCOP	Mono(carboxyoctyl) phthalate
MCPP	Mono-(3-carboxypropyl) phthalate
MECPP	Mono-2-ethyl-5-carboxypentyl phthalate
MEHHP	Mono-(2-ethyl-5-hydroxyhexyl) phthalate
MEHP	Mono-(2-ethyl)-hexyl phthalate
MEOHP	Mono-(2-ethyl-5-oxohexyl) phthalate
MEP	Mono-ethyl phthalate
MiBP	Mono-isobutyl phthalate
MNMP	Mono-n-methyl phthalate
MNP	Mono-isononyl phthalate
MPB	methyl paraben

NHANES	National Health and Nutrition Examination Survey
PAHs	polycyclic aromatic hydrocarbons
PCA	principal components analysis
PPB	propyl paraben
SDR	Sparse Decompositional Regression

References

Afshin A, et al. , 2017. Health effects of overweight and obesity in 195 countries over 25 years. N. Engl. J. Med 377, 13–27. 10.1056/NEJMoa1614362. [PubMed: 28604169]

AnonWHO, 2017. Defining Adult Overweight and Obesity.

Barr DB, et al. , 2005. Urinary creatinine concentrations in the U.S. population: implications for urinary biologic monitoring measurements. Environ. Health Perspect. 113, 192–200. 10.1289/ehp.7337. [PubMed: 15687057]

Blount BC, et al. , 2006. Urinary perchlorate and thyroid hormone levels in adolescent and adult men and women living in the United States. Environ. Health Perspect. 114, 1865–1871. 10.1289/ehp.9466. [PubMed: 17185277]

Bobb JF, et al. , 2015. Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. Biostatistics 16, 493–508. 10.1093/biostatistics/kxu058. [PubMed: 25532525]

Bobb JF, et al. , 2018. Statistical software for analyzing the health effects of multiple concurrent exposures via Bayesian kernel machine regression. Environ. Health 17, 67. 10.1186/s12940-018-0413-y. [PubMed: 30126431]

Bushnik T, et al. , 2019. Association of urinary polycyclic aromatic hydrocarbons and obesity in children aged 3–18: Canadian Health Measures Survey 2009–2015. J. Dev. Orig. Health Dis. 1–9. 10.1017/S2040174419000825. [PubMed: 30919803]

CDC, National Health and Nutrition Examination Survey Analytic Guidelines. 2018a. (https://wwwn.cdc.gov/nchs/nhanes/analyticguidelines.aspx).

CDC, NHANES 2011–2012 Laboratory Methods. 2018b. (https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/labmethods.aspx?BeginYear=2011).

Chang H, et al. , 2015. Stacked Predictive Sparse Decomposition for Classification of Histology Sections. Int J Comput Vis. 113, 3–18. 10.1007/s11263-014-0790-9. [PubMed: 27721567]

Chang H, et al. , 2018. Unsupervised Transfer Learning via Multi-Scale Convolutional Sparse Coding for Biomedical Applications. IEEE Trans Pattern Anal Mach Intell. 40, 1182–1194. 10.1109/tpami.2017.2656884. [PubMed: 28129148]

Chang H, et al. , 2021. From Mouse to Human: Cellular Morphometric Subtype Learned From Mouse Mammary Tumors Provides Prognostic Value in Human Breast Cancer. Front Oncol. 11, 819565. 10.3389/fonc.2021.819565. [PubMed: 35242697]

Chin T-J, Suter D, 2007. Incremental kernel principal component analysis. IEEE Trans. Image Process. a Publ. IEEE Signal Process. Soc. 16, 1662–1674.

Gelman A, 2007. Struggles with survey weighting and regression modeling. Stat. Sci. 22 (153–164), 12. 10.1214/088342306000000691.

Graubard BI, Korn EL, 1999. Analyzing health surveys for cancer-related objectives. JNCI: J. Natl. Cancer Inst. 91, 1005–1016. 10.1093/jnci/91.12.1005. [PubMed: 10379963]

Haddad N, et al. , 2019. A scoping review on the characteristics of human exposome studies. Curr. Pollut. Rep. 5. 10.1007/s40726-019-00130-7.

Haddad N, et al. , 2022. An exposome-wide association study on body mass index in adolescents using the National Health and Nutrition Examination Survey (NHANES) 2003–2004 and 2013–2014 data. Sci. Rep. 12, 8856. 10.1038/s41598-022-12459-z. [PubMed: 35614137]

Hatch EE, et al. , 2008. Association of urinary phthalate metabolite concentrations with body mass index and waist circumference: a cross-sectional study of NHANES data, 1999–2002. Environ. Health 7, 27. 10.1186/1476-069x-7-27. [PubMed: 18522739]

Heindel JJ, et al. , 2017. Metabolism disrupting chemicals and metabolic disorders. Reprod. Toxicol. 68, 3–33. 10.1016/j.reprotox.2016.10.001. [PubMed: 27760374]

Kavukcuoglu K, et al., Learning invariant features through topographic filter maps. 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009, pp. 1605–1612.

Kim JH, 2019. Multicollinearity and misleading statistical results. Korean J. Anesth. 72, 558–569. 10.4097/kja.19087.

Kim S, et al. , 2017. Considering common sources of exposure in association studies - urinary benzophenone-3 and DEHP metabolites are associated with altered thyroid hormone balance in the NHANES 2007–2008. Environ. Int 107, 25–32. 10.1016/j.envint.2017.06.013. [PubMed: 28651165]

Lazarevic N, et al. , 2019. Statistical methodology in studies of prenatal exposure to mixtures of endocrine-disrupting chemicals: a review of existing approaches and new alternatives. Environ. Health Perspect. 127, 26001. 10.1289/ehp2207. [PubMed: 30720337]

Lee H, et al. , 2006. Efficient sparse coding algorithms. Adv. Neural Inf. Process. Syst. 19.

Lin WY, et al. , 2019. Performing different kinds of physical exercise differentially attenuates the genetic effects on obesity measures: Evidence from 18,424 Taiwan Biobank participants. PLoS Genet 15, e1008277. 10.1371/journal.pgen.1008277. [PubMed: 31369549]

Ling C, Ronn T, 2019. Epigenetics in human obesity and type 2 diabetes. Cell Metab. 29, 1028–1044. 10.1016/j.cmet.2019.03.009. [PubMed: 30982733]

Liu B, et al. , 2017. Bisphenol A substitutes and obesity in US adults: analysis of a population-based, cross-sectional study. Lancet Planet Health 1, e114–e122. 10.1016/s2542-5196(17)30049-9. [PubMed: 29308453]

Liu XP, et al. , 2022. Clinical Significance and Molecular Annotation of Cellular Morphometric Subtypes in Lower Grade Gliomas discovered by Machine Learning. Neuro Oncol. 10.1093/neuonc/noac154.

Lubin JH, et al. , 2004. Epidemiologic evaluation of measurement data in the presence of detection limits. Environ. Health Perspect. 112, 1691–1696. 10.1289/ehp.7199. [PubMed: 15579415]

Mairal J, et al. , 2008. Learning multiscale sparse representations for image and video restoration. Multiscale Model. Simul. 7, 214–241. 10.1137/070697653.

Mao XY, 2022. iCEMIGE: Integration of CEll-morphometrics, MIcrobiome, and GEne biomarker signatures for risk stratification in breast cancers. World J Clin Oncol 13, 616–629. 10.5306/wjco.v13.i7.616.

Marill KA, 2004. Advanced statistics: linear regression, part II: multiple linear regression. Acad. Emerg. Med. 11, 94–102. 10.1197/j.aem.2003.09.006. [PubMed: 14709437]

Patel CJ, et al. , 2010. An environment-wide association study (EWAS) on type 2 diabetes mellitus. PLoS One 5, e10746. 10.1371/journal.pone.0010746. [PubMed: 20505766]

Rozell CJ, et al. , 2008. Sparse coding via thresholding and local competition in neural circuits. Neural Comput. 20, 2526–2563. 10.1162/neco.2008.03-07-486. [PubMed: 18439138]

Rubinstein RB, A. M, Elad M, 2010. Dictionaries for sparse representation modeling. Proc. IEEE 98, 1045–1057. 10.1109/JPR0C.2010.2040551.

Scinicariello F, Buser MC, 2014a. Urinary polycyclic aromatic hydrocarbons and childhood obesity: NHANES (2001–2006). Environ. Health Perspect. 122, 299–303. 10.1289/ehp.1307234. [PubMed: 24380973]

Scinicariello F, Buser MC, 2014b. Urinary polycyclic aromatic hydrocarbons and childhood obesity: NHANES (2001–2006). Environ. Health Perspect. 122, 299–303. 10.1289/ehp.1307234. [PubMed: 24380973]

Shaheen N, et al. , 2016. Presence of heavy metals in fruits and vegetables: Health risk implications in Bangladesh. Chemosphere 152, 431–438. 10.1016/).chemosphere.2016.02.060. [PubMed: 27003365]

Stafoggia M, et al. , 2017. Statistical approaches to address multi-pollutant mixtures and multiple exposures: the State of the Science. Curr. Environ. Health Rep. 4, 481–490. 10.1007/s40572-017-0162-z. [PubMed: 28988291]

Stahlhut RW, et al. , 2007. Concentrations of urinary phthalate metabolites are associated with increased waist circumference and insulin resistance in adult U.S. males. Environ. Health Perspect. 115, 876–882. 10.1289/ehp.9882. [PubMed: 17589594]

Taylor KW, et al. , 2016. Statistical approaches for assessing health effects of environmental chemical mixtures in epidemiology: lessons from an innovative workshop. Environ. Health Perspect. 124, A227–a229. 10.1289/ehp547. [PubMed: 27905274]

Thayer KA, et al. , 2012. Role of environmental chemicals in diabetes and obesity: a National Toxicology Program workshop review. Environ. Health Perspect. 120, 779–789. 10.1289/ehp.1104597. [PubMed: 22296744]

Vermeulen R, et al. , 2020. The exposome and health: where chemistry meets biology. Science 367, 392–396. 10.1126/science.aay3164. [PubMed: 31974245]

Wild CP, 2005. Complementing the genome with an “exposome”: the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol. Biomark. Prev. 14, 1847–1850. 10.1158/1055-9965.Epi-05-0456.

Wold S, Kim Esbensen, Paul Geladi, 1987. Principal component analysis. Chemometrics and Intelligent Laboratory Systems. 2, 37–52. 10.1016/0169-7439(87)80084-9.

Wild CP, 2012. The exposome: from concept to utility. Int J. Epidemiol. 41, 24–32. 10.1093/ije/dyr236. [PubMed: 22296988]

Wu C, et al. , 2019. Repeated measurements of paraben exposure during pregnancy in relation to fetal and early childhood growth. Environ. Sci. Technol. 53, 422–433. 10.1021/acs.est.8b01857. [PubMed: 30427191]

Zhang Y, et al. , 2019. Association between exposure to a mixture of phenols, pesticides, and phthalates and obesity: comparison of three statistical models. Environ. Int 123, 325–336. 10.1016/j.envint.2018.11.076. [PubMed: 30557812]

Zipf G, et al. , 2013. National health and nutrition examination survey: plan and operations, 1999–2010. Vital. Health Stat. 1, 1–37.

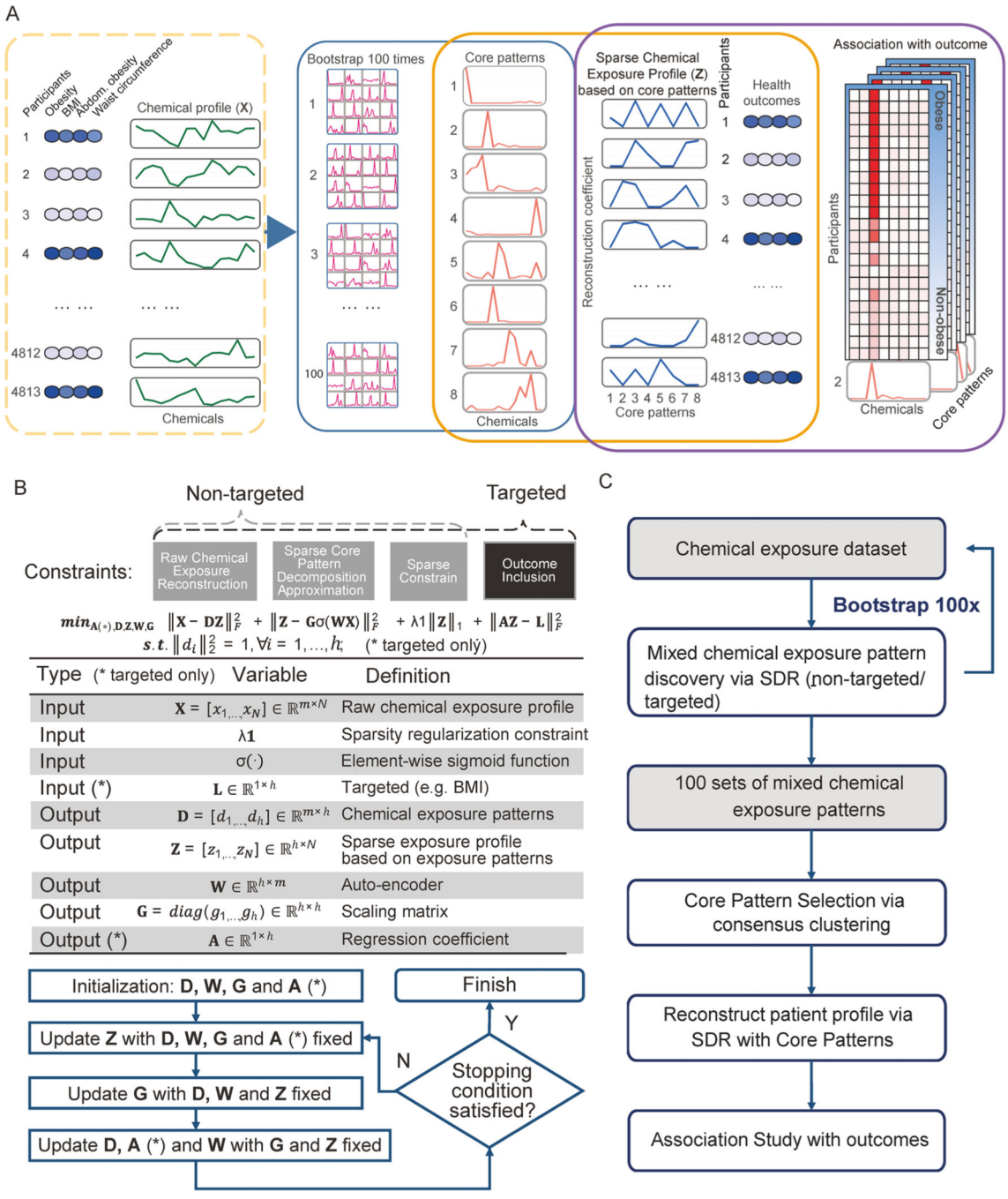

Fig. 1.

The main structure of the SDR model and its application in our study. A, Application of the SDR model to NHANES study data to identify core exposure patterns associated with obesity and other body index-related outcomes. Each participant had their exposure status (chemical profile). The SDR model was employed to identify exposure patterns (the chemical mixture signals), and limited exposure patterns were combined to reconstruct the original exposure profile of each participant. After bootstrapping 100 times, we obtained 100 sets of exposure patterns. Consensus clustering was used to reduce the redundancy of

exposure patterns to get the core patterns, and thereafter, the sparse combination of core patterns was used to reconstruct the original exposure status. The association of core patterns with outcomes was further evaluated. B, The mathematical formulation and optimization scheme of the SDR model in non-targeted and targeted modes; A detailed description of each item can be found in the Methods. C, The computational pipeline of the SDR model.

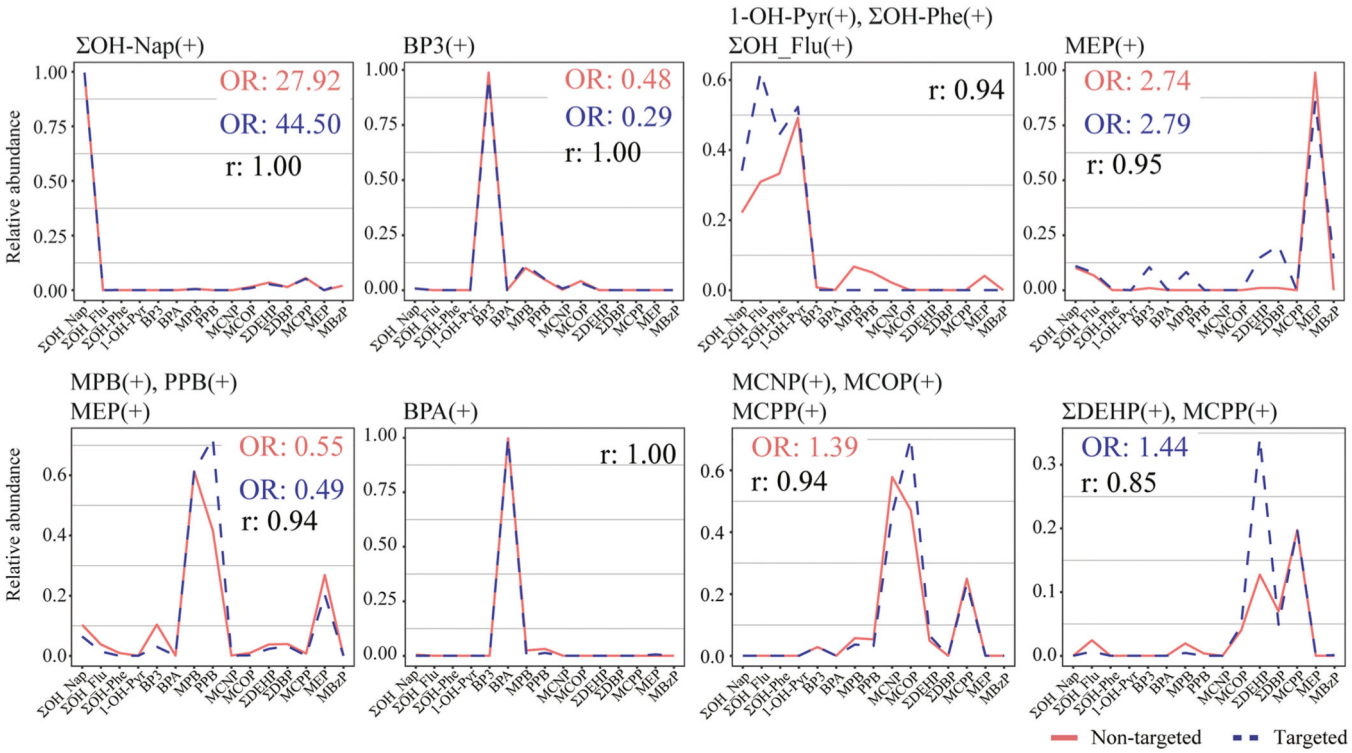

Fig. 2.

Association of core exposure patterns with general obesity using the non-targeted (pink color) and targeted (blue color) SDR models. The numbers in each panel represent the odds ratio (OR) for general obesity. All regressions were adjusted for sex, race, educational levels, age, family income–to-poverty ratio, smoking status, energy intake levels, and urinary creatinine. OR are only indicated for significant associations ($p < 0.05$).

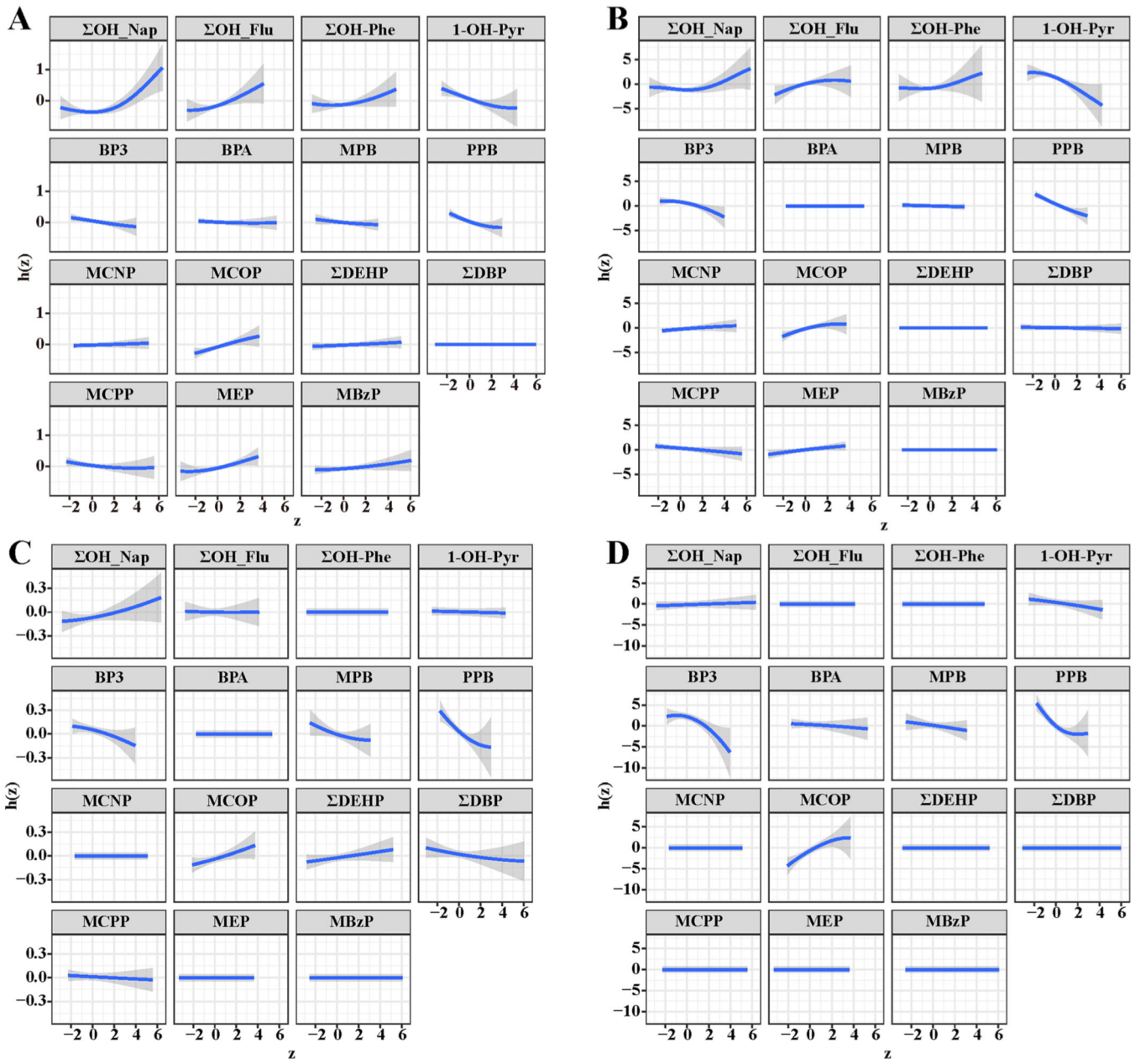

Fig. 3.

Univariate exposure-response function (95% CI) between selected chemical concentrations and obesity-related outcomes (A: general obesity, B: BMI, C: abdominal obesity, D: waist circumference) while fixing the concentrations of other chemicals at median values (N = 4813), NHANES 2005–2012. Models were adjusted for sex, race, educational levels, age, family income-to-poverty ratio, smoking status, energy intake levels, and urinary creatinine.

Table 1

Association of non-targeted SDR core patterns and obesity and body indexes (N = 4813), NHANES 2005–2012

Chemicals	General obesity		BMI		Abdominal obesity		Waist circumference	
	OR (95% CI)	*p* value	β (95% CI)	*p* value	OR (95% CI)	*p* value	β (95% CI)	*p* value
Core pattern1	27.92 (5.78, 153.84)	< 0.001	6.46 (1.90, 11.01)	0.006	12.19 (2.06, 91.55)	0.009	14.80 (3.86, 25.75)	0.008
Core pattern2	0.48 (0.34, 0.68)	< 0.001	−2.47 (−3.47, −1.48)	< 0.001	0.54 (0.38, 0.76)	< 0.001	−5.63 (−8.02, −3.25)	< 0.001
Core pattern3	0.98 (0.75, 1.30)	0.906	−0.12 (−0.95, 0.71)	0.777	1.03 (0.78, 1.36)	0.846	0.59 (−1.40, 2.58)	0.560
Core pattern4	2.74 (2.00, 3.76)	< 0.001	3.37 (2.43, 4.32)	< 0.001	2.90 (2.09, 4.03)	< 0.001	8.24 (5.97, 10.50)	< 0.001
Core pattern5	0.55 (0.46, 0.66)	< 0.001	−2.35 (−2.87, −1.83)	< 0.001	0.50 (0.41, 0.59)	< 0.001	−6.41 (−7.66, − 5.16)	< 0.001
Core pattern6	0.75 (0.51, 1.09)	0.134	−0.15 (−1.29, 1.00)	0.800	0.78 (0.53, 1.16)	0.217	−0.45 (−3.20, 2.30)	0.747
Core pattern7	1.39 (1.07, 1.79)	0.012	1.67 (0.91, 2.43)	< 0.001	1.43 (1.10, 1.86)	0.007	4.43 (2.60, 6.25)	< 0.001
Core pattern8	0.83 (0.41, 1.41)	0.542	−0.25 (−1.65, 1.14)	0.720	0.68 (0.35, 1.15)	0.186	−0.96 (−4.30, 2.39)	0.574

OR: odds ratio; CI: confidence interval

Models were adjusted for gender, ethnicity, educational levels, age, family income-to-poverty ratio, smoking status, energy intake levels, and ln-transformed creatinine.

Table 2

Principal components analysis of exposure biomarker concentration and obesity or BMI (N = 4813), NHANES 2005–2012.

Dominant Chemicals	PC1		PC2		PC3		PC4	
	ΣOH-Nap (+), ΣOH-Flu (+), ΣOH-Phe (+), 1-OH-Pyr (+)		BPA (+), ΣDEHP (+), ΣDBP (+), MEP (+), MBZP (+)		MCNP (+), MCOP (+), MCPP (+)		BP3 (+), MPB (+), PPB (+)	
	Effect (95% CI)	*p* value	Effect (95% CI)	*p* value	Effect (95% CI)	*p* value	Effect (95% CI)	*p* value
General obesity	0.99 (0.91, 1.08)	0.797	1.07 (0.99, 1.15)	0.091	1.04 (0.98, 1.12)	0.217	0.77 (0.72, 0.83)	< 0.001
BMI	−0.12 (−0.38, 0.13)	0.334	0.19 (T0.04, 0.42)	0.107	0.32 (0.12, 0.53)	0.002	−1.01 (− 1.22, − 0.80)	< 0.001
Abdominal obesity	1.00 (0.92, 1.09)	1.000	1.03 (0.95, 1.11)	0.458	1.07 (1.00, 1.15)	0.049	0.75 (0.69, 0.80)	< 0.001
Waist circumference	−0.12 (−0.38, 0.13)	0.334	0.19 (T0.04, 0.42)	0.107	0.32 (0.12, 0.53)	0.002	−1.01 (− 1.22, − 0.80)	< 0.001

Effect represents OR (odds ratio) in association with obesity; Effect represents β in association with BMI; CI: confidence interval.

Note: Adjusted for sex, race, educational levels, age, family income-to-poverty ratio, smoking status, energy intake levels, and urinary creatinine.

Table 3

Comparison of SDR with classic methods in the assessment of the association between chemical mixture exposure and obesity.

Methods	Chemicals/Chemical mixtures associated with obesity and body indexes
SDR	ΣOH-Nap↑, MEP↑, (MCNP, MCOP, and MCPP)↑, BP3↓, (MPB, PPB, and MEP) ↓
GLM	BP3↓, MPB↓, PPB↓, and MCOPf
PCA	(MCNP, MCOP, and MCPP)↑, (BP3, MPB, PPB)↓
BKMR	ΣOH-Nap↑, MCOP↑, 1-OH-Pyr↓, BP3↓, PPB↓

Note: ↑indicates the positive association with outcomes and ↓ indicates the negative associations.