Article

Machine Learning Assisted Discovery of Interactions between Pesticides, Phthalates, Phenols, and Trace Elements in Child Neurodevelopment

Vishal Midya,*,[⊥] Cecilia Sara Alcala,[⊥] Elza Rechtman, Jill K. Gregory, Kurunthachalam Kannan, Irva Hertz-Picciotto, Susan L. Teitelbaum, Chris Gennings, Maria J. Rosa,[¶] and Damaskini Valvi[¶]

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: A growing body of literature suggests that developmental exposure to individual or mixtures of environmental chemicals (ECs) is associated with autism spectrum disorder (ASD). However, investigating the effect of interactions among these ECs can be challenging. We introduced a combination of the classical exposure-mixture Weighted Quantile Sum (WQS) regression and a machine-learning method termed Signed iterative Random Forest (SiRF) to discover synergistic interactions between ECs that are (1) associated with higher odds of ASD diagnosis, (2) mimic toxicological interactions, and (3) are present only in a subset of the sample whose chemical concentrations are higher than certain thresholds. In a case-control Childhood Autism Risks from Genetics and Environment (CHARGE) study, we evaluated multiordered synergistic interactions among 62 ECs measured in the urine samples of 479 children in association with increased odds for ASD diagnosis (yes vs no). WQS-SiRF identified two synergistic two-ordered interactions between (1) trace-element cadmium (Cd) and the organophosphate pesticide metabolite diethyl-phosphate (DEP); and (2) 2,4,6-trichlorophenol (TCP-246) and DEP. Both interactions were suggestively associated with increased odds of ASD diagnosis in the subset of children with urinary concentrations of Cd, DEP, and TCP-246 above the 75th percentile. This study demonstrates a novel method that combines the inferential power of WQS and the predictive accuracy of machine-learning algorithms to discover potentially biologically relevant chemical−chemical interactions associated with ASD.

KEYWORDS: autism spectrum disorder, environmental chemical exposures, iterative random forests, random intersection tree, exposure mixture model, synergistic interactions
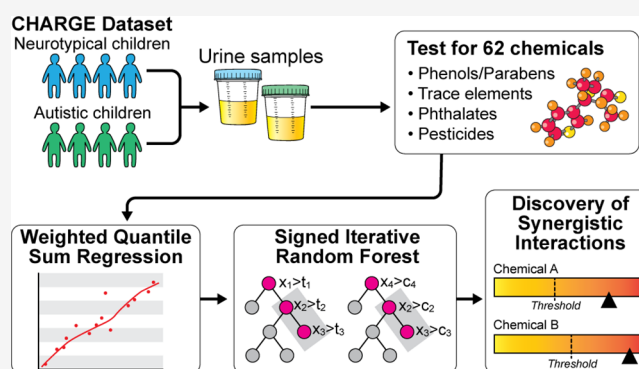
INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by deficits in social communication and interaction and repetitive and stereotyped interests and behaviors.[1] ASD prevalence has increased drastically in recent years and is a public health concern worldwide. According to the Centers for Disease Control program Autism and Developmental Disabilities Monitoring (ADDM) Network, approximately 1 in 44 children have been diagnosed with ASD.[2,3] In the past decade, a growing number of epidemiological studies have associated early life environmental exposures with ASD.[4] These environmental exposures include air pollution,[5−9] nutrition, and several endocrine-disrupting chemicals (EDCs). Among other EDCs, studies on certain metals have been associated with ASD,[10,11] with a compelling link between arsenic exposure and ASD in children.[12] Other EDCs, such as bisphenol A (BPA), and parabens have also been proposed as potential risk factors for

child behavioral outcomes,[14−16] though this evidence is less consistent across studies.[13]

Although the etiology of ASD remains unclear, an interplay of multiple genetic and early environmental contributions that differ between individuals likely underlies disease risk.[4,17,18] Genetic and environmental factors may impact typical brain development, including neuron formation and migration, synapse formation, or neural connectivity, ultimately leading to ASD.[4] Environmental chemical exposures may impact neurodevelopment through multiple mechanisms, including the direct disruption of cells and structures of the nervous

system, endocrine hormone- or immune system-mediated effects, and/or epigenetic changes, among others.[4] However, there is a lack of studies assessing potential chemical-chemical interactions in ASD. Among the very few studies, Curtin et al. examined the dynamic interaction of zinc−copper cycles, which regulate metal metabolism, is disrupted in ASD.[19] Findings showed that the interaction between cyclical co-occurrence between zinc and copper is disrupted in ASD.[19,20]

The concept of "interaction" has been construed in many ways through different scientific fields.[21] For example, in epidemiological studies, interactions are usually reported through association estimates of their effect sizes or inclusion probabilities.[22−33] Though estimating associations is essential, most methods do not provide any mechanistic or biological insight, possibly because the reported interactions are of particular functional forms (for example, multiplication of exposures) rather than representing their collective activities beyond certain concentration thresholds.[34] Further, after applying certain dimension reductions, most interactions are reported between sets of reduced exposures, limiting interpretability. In addition, such interactions provide a population-level estimate, with each sample providing some contribution to the overall estimate.

In contrast, the toxicological representation of the interactions is easier to comprehend. Through the collective activities of the chemicals, (1) one can identify the mechanism of synergistic or antagonistic behavior that might arise beyond the concentration thresholds (and not just the regression coefficient of multiplicative associations), and (2) the use of concentration thresholds reflect the toxicological underpinning of classical threshold based on chemical dose−response studies.[35−37] Moreover, as the number of chemical exposures increases, searching for multiordered interactions gets computationally intensive. Most current methods, therefore, "hard code" or prespecify interaction terms in models, but such strategies are limited due to restrictions on sample size and are usually underpowered.[38,39] In comparison, Kernel Machine Regression or Bayesian factorization-inspired methods discover interactions with certain functional forms that do not represent any collective activity or concentration threshold.[30,32,40] The lack of similarity with toxicological threshold-based dose−response studies makes it difficult to find biologically relevant interpretations of the recovered interactions. It is also possible that such interactions can only be present in a subset of the population since not every sample will have chemical concentrations beyond certain thresholds. Novel analytical approaches are required to account for these challenges and to move the field forward.

As a possible alternative to address this problem of interpretability of complex interactions among chemicals, tree-based machine learning (ML) models have been proposed that can offer a solution to represent collective activities of exposures as threshold-based interactions. Nevertheless, a related challenge is that most of the tree-based ML models are black-box, creating tension between prediction quality and meaningful biological insight. Moreover, a predictive ML model might not be the optimal model for inference.[41] However, in recent epidemiological studies, interpretable tree-based ML tools were used to discover simultaneously co-occurring chemicals, similar to classical Weighted Quantile Sum (WQS) Regression models.[42−46] Separately in computational biology, using a novel ML algorithm called random intersection trees,[47] Basu et al.[48,49] introduced the "signed

iterative random forest" (SiRF) algorithm to discover interactions through collective activities. SiRF can efficiently search for the few stable and highly occurring interactions instead of going through each possible interaction term. Since exposure to environmental chemicals occurs continuously, we aimed to use a combination of the WQS regression and the ML method Signed Iterative Random Forest to search for interactions that mimic toxicological interactions. Using data from the Childhood Autism Risks from the Genetics and Environment (CHARGE) study, we aimed to identify multiordered synergistic interactions between environmental chemicals at specific exposure thresholds associated with higher odds of ASD. We further examined whether the directionality of the interactions remained unaltered even after adjusting for potential effects of the overall chemical mixture.

■ METHODS

Study Design and Population. Details about the CHARGE study have been reported in Bennett et al.[13] Briefly, the Childhood Autism Risks from Genetics and Environment (CHARGE) is a case-control study that recruited three groups of children between 2006 and 2017: (1) children with ASD (2) children with developmental delay (DD) but not ASD, and (3) children with typical development (TD).[49] Children from the first two groups were mainly identified by the California Department of Developmental Services. The department coordinates services for individuals with developmental disabilities and is inclusive of all residents of California regardless of their place of birth, religion, or financial resources.[13] The third group (controls) was sampled from California birth files utilizing frequency matching of ASD cases comprising the following characteristics: age, sex, and broad geographic regions up to 10 counties. Children from all three groups were (a) aged 24−60 months at recruitment, (b) living with a biological parent who speaks English or Spanish, (c) born in California, and (d) residing in the study catchment area. CHARGE study included all children with at least 16 mL of urine collected at their assessment and available for chemical analysis. In addition, detailed demographic characteristics of the parents and children were collected during the study visit. However, in this present study, we included only children with either ASD (from group 1) or typical development (from group 3), totaling a sample size of 479. The CHARGE study protocol was approved by the institutional review boards of the State of California and the University of California at Davis. Before collection of any data, all participants provided written informed consent.

Chemical Analysis. A spot urine sample was collected from each participant during their visit. All samples were frozen immediately at −20 °C and remained frozen until analysis. The samples were shipped on dry ice to Wadsworth Center's Human Health Exposure Analysis Resource (HHEAR) Laboratory (Albany, NY) for analysis. Enzymatic deconjugation and liquid−liquid extraction were used in the determination of environmental phenols (i.e., benzophenone, bisphenols, chlorophenols, parabens, and triclosan), as previously described.[50,51] Additional description of the target phenolic compounds can be found in Bennett et al.[13,50−52] Twenty urinary phthalate metabolites (PhMs) were analyzed using enzymatic deconjugation, solid-phase extraction (SPE), and an isotope dilution method of quantification.[53] Further details of the analysis of PhMs are described elsewhere.[13,53,54] Six dialkyl phosphate metabolites (DAPs) were as described else-
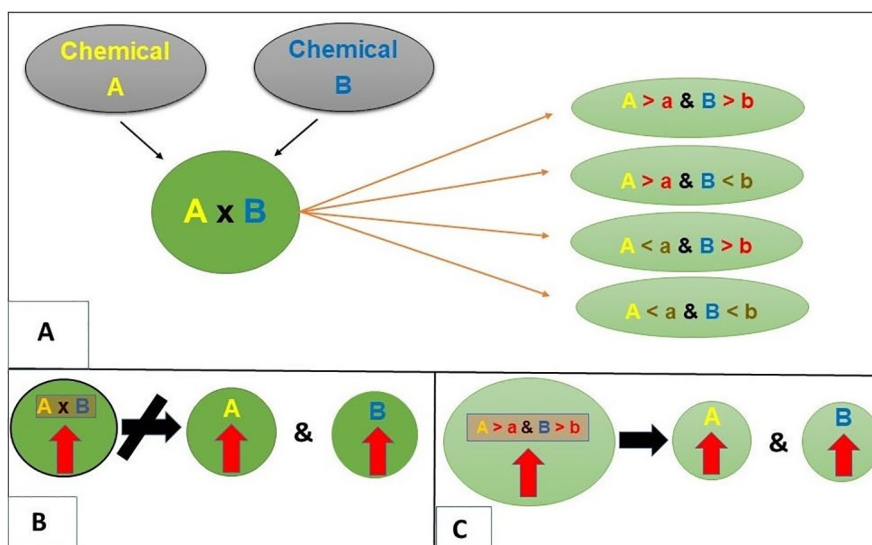
Figure 1. Conceptual illustration of multiplicative and toxicologically mimicked interactions. Figure 1A shows a usual multiplicative interaction between two chemicals and mapping to plausible threshold-based interactions. Figure 1B shows that a higher value of a multiplicative interaction does not imply a higher value for each of its constituents. On the other hand, Figure 1C denotes a higher value of a toxicologically mimicking interaction, implying a higher value for each of its constituents.

where.[13,55] High performance liquid chromatography–tandem mass spectrometry (HPLC–MS/MS) was used in the analysis of environmental phenols, PhMs, and DAPs. Trace elements were analyzed in urine specimens using inductively coupled plasma mass spectrometry (ICP-MS) at Wadsworth Center.[13,56] Quality assurance and harmonization for targeted biomonitoring of organic chemicals in the Human Health Exposure Analysis Resource (HHEAR) laboratory network has been detailed previously.[57] The method recoveries for analytes were within 80–120%, and the matrix effect was corrected using internal standards for each analyte.[57]

Urinary concentrations were corrected for specific gravity (SG) using the formula, $P_c = P \times [(SG_p - 1)/(SG - 1)]$.[58] P_c is the SG corrected metabolite concentration (ng/mL), and SG is the specific gravity of the urine sample. The median specific gravity of the CHARGE cohort participants was 1.0223 ng/mL (SG_p). In the event that the specific gravity correction factors were greater than 2, they were assigned a value of 2. For values below 0.5, they were assigned 0.5.[13]

Developmental Assessment. During the study visit, an assessment of ASD was conducted (to confirm the diagnosis of ASD indicated during the CHARGE enrollment process) using two gold standard psychometric instruments: the Autism Diagnostic Interview-Revised (ADI-R)[59−61] and the Autism Diagnostic Observation Schedules (ADOS).[62] The ADI-R is a semistructured interview administered by the primary caregiver to diagnose autism and to differentiate autism from other developmental disorders.[61] The ADOS is a semistructured, standardized assessment where the researcher observes the social interaction, communication, play, and imaginative use of materials by children suspected of having ASD.[13,62] We utilized the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) and followed standardized procedures from the ADOS and ADI-R to assign the final diagnosis of ASD.[63] All children were administered the Mullen Scales of Early Learning (MSEL) and the Vineland Adaptive Behaviors Scores (VABS).[13] To confirm that a child did not have ASD, we used the Social Communications Questionnaire to screen for ASD in children in both the developmental delay and general population groups.[64] If a child was positive, we administered the ADI-R and ADOS to determine if they had ASD. All other children enrolled because of a community diagnosis of ASD or DD, but were not confirmed for either of these two diagnoses, were grouped as Other Early Concerns (OEC).[13] Children were classified as TD and enrolled as general population controls who did not meet the criteria for either ASD or DD. All classifications are mutually exclusive. All clinicians participating in the study spoke English and/or Spanish. Additionally, they achieved research reliability on all of the instruments they administered.[13]

Statistical Analysis. We used the weighted quantile sum (WQS)[26] regression to model the adverse mixture effect of chemicals while simultaneously (1) accommodating the correlation structure of the chemicals and (2) controlling for covariates. Previous studies between chemical exposures and neurodevelopment in the CHARGE case-control study[13,65] found significant positive associations between the environmental chemical mixture and the outcomes. Given these results and the a priori hypothesis that environmental exposures increase the odds for ASD, we assumed a positive association (in an adverse direction) between the environmental chemical mixture and ASD diagnosis. To reduce spurious co-occurrences of chemicals, interactions were searched on top of the chemical-mixture effect. A conceptual schematic of different kinds of interactions has been shown in Figure 1. Briefly, these interactions mimic the classical toxicological paradigm in which an interaction occurs only if the concentration of certain chemicals is above a threshold. Conceptually, a usual multiplicative interaction between two chemicals (for instance, A and B) can be mapped to four toxicological interactions: (1) concentration of A is high, and concentration of B is high, (2) concentration of A high, and concentration of B is low, (3) concentration of A is low, and concentration of B is high, and (4) concentration of A is low, and concentration of B is low (see Figure 1A). Note that each of the four components is easier to interpret and could directly imply a plausible toxicological interpretation. Moreover, a positive association with multiplicative interaction does not

necessarily imply synergy since the higher value of multiplicative interaction does not imply that the concentrations of individual chemicals are also high. However, such a problem of interpretability does not arise for toxicologically mimicked interactions (Figures 1B and 1C). Lastly, multiplicative interactions provide a population-level interaction estimate—where all individuals contribute, whereas the mimicked toxicological interactions are present in only a subset of the population. In the following analysis, we searched for synergistic interactions in the adverse direction, i.e., chemical exposures higher than certain concentration thresholds, mimicking a toxicological interaction.

All models were controlled for the child's sex, year of birth, race/ethnicity, age at enrollment, maternal age at the time of childbirth, maternal metabolic conditions during pregnancy (any hypertensive disorder, including obesity or any diabetes), and parental homeowner status (as a proxy of socioeconomic status). These covariates were chosen *a priori* based on the previous analysis by Bennett et al.[13] To make the analysis robust; we implemented the random subset and repeated holdout[66,67] variants of WQS. Assuming the main chemical-mixture effect and the synergistic interactions are additive, we extracted the Pearson residuals from this model and treated the residual as the new outcome (the Pearson residual possesses asymptotic normality).[37,68] Therefore, we searched for synergistic interactions on the residuals after adjusting the first-order main mixture effect and the covariates.

We searched for interactions through signed-iterated Random Forest (SiRF), in which the Pearson residuals from WQS were the outcome and the chemicals were the exposures. The SiRF utilizes a combination of state-of-the-art machine-learning tools, iterative Random Forests (iRFs), and recently developed Random Intersection Trees (RITs) to search for interactions within a certain proportion of samples.[47,69−71] Instead of searching through all possible combinations, SiRF searches for combinations of exposures that sequentially occur within each tree's branches (or decision paths) in the RFs. Therefore, instead of looking for all possible combinations, the branches in the tree provide predetermined and possibly predictive combinations. SiRF searches for high-order chemical exposure interactions, as follows: First, the model begins with fitting the RF model and reweighting the important exposures. Using the reweighted exposures, multiple RF models are fitted iteratively to reduce the dimensionality of the exposure space without removing marginally unimportant exposures. Important exposures are denoted as higher and significant, while unimportant exposures are those with marginally low contributions to the prediction framework. Second, decision rules are extracted from the iterated RF and fed to a generalization of the RIT to efficiently discover high-order interactions from the decision paths. Last, a bagging step is introduced in the algorithm to assess the "stability" of the recovered interactions through a large number of bootstrapped iterations. Here stability implies the number of times an interaction is detected throughout the iterations; therefore, the higher the recovery rate, the better. Since SiRF searches through particular decision branches, it can incorporate meaningful directionality (in the current study, synergism) while recovering the interactions. The combination of WQS-SiRF can robustly search for interactions without the need to rely on p-values. It should be noted that the WQS-SiRF technique does not look through all possible combinations of exposures and selects only a few predictive combinations.

Therefore, the penalties for multiple comparison errors could be minimal, irrespective of sample size. The WQS technique combines all of the exposures to create an overall mixture index with a final one-degree-of-freedom hypothesis test. While the sheer number of combinatorial interactions can rapidly increase with an increase in the number of exposures, the SiRF algorithm only chooses the predictive ones. Finally, only in the next stage is the hypothesis testing on interactions carried out using a few selected combinations. Therefore, the combination of WQS-SiRF might help in diverting the loss in statistical power.

In the SiRF part, the model was trained on a subset of data and then bagging was introduced on the remaining held-out testing data. Therefore, to obtain robust results against the sensitivity of data partitioning, we chose three different data partitions: (1) 70% for training and 30% for testing, (2) 75% for training and 25% for testing, and (1) 80% for training and 20% for testing. Finally, we chose only those interactions with (1) more than 50% stability score and (2) common to all three data partitioning results. Since the discovered interactions were based on thresholds, they were present in only certain portions of the samples. However, SiRF does not directly estimate the thresholds by itself. Therefore, we created interaction indicators based on their joint concentrations to denote the presence or absence of interactions. For example, if the specific gravity-adjusted concentrations of the chemicals were more than the 75th percentile, then the interaction indicator would be nonzero; else, it would be zero. We created another set of indicators based on the 67th percentile threshold for sensitivity analysis. (1) For WQS analysis, we converted all chemical exposures to deciles and (2) to ensure that WQS-SiRF is not sensitive to adding exposures with low detection limits, we included all chemicals irrespective of their percentage detected above LOD. Note that the conversion in deciles for chemical exposures and the growing many decision trees through bootstraps protect against outlying and influential observations.

For sensitivity analyses, (1) we repeated the WQS-SiRF algorithm with data partitioned in 75% for training and 25% for testing without chemicals whose % of detection above LOD was less than 60%, (2) we gradually increased the number of bootstraps, from 250, 500, to 1000, (3) we used the whole data set to test the model trained on the 75% data, (4) repeated WQS-SiRF to obtain interactions observed in the primary analysis after randomly permuting the ASD status, and (5) conducted a Bayesian Kernel Machine Regression (BKMR) analysis to compare and contrast the synergistic interactions discovered from WQS-SiRF.

For descriptive analysis, we calculated the Pearson correlation matrices of log-transformed and specific gravity-corrected chemicals exposures for ASD and TD children. All concentration values detected below the corresponding LODs were imputed by the value of LOD/2. Missing data in covariates were minimal (<5%) and were imputed using the R package "mice".[72] A two-tailed p-value less than alpha at 0.05 is considered statistically significant. All data were analyzed in R version 4.1.2. A detailed mathematical exposition of the algorithm was reported earlier.[69] In addition, the tuning parameters in WQS-SiRF and random seeds for training and testing data are provided in the Supporting Information. All the R codes are available online in Git Hub (https://github.com/vishalmidya/WQS_SiRF).
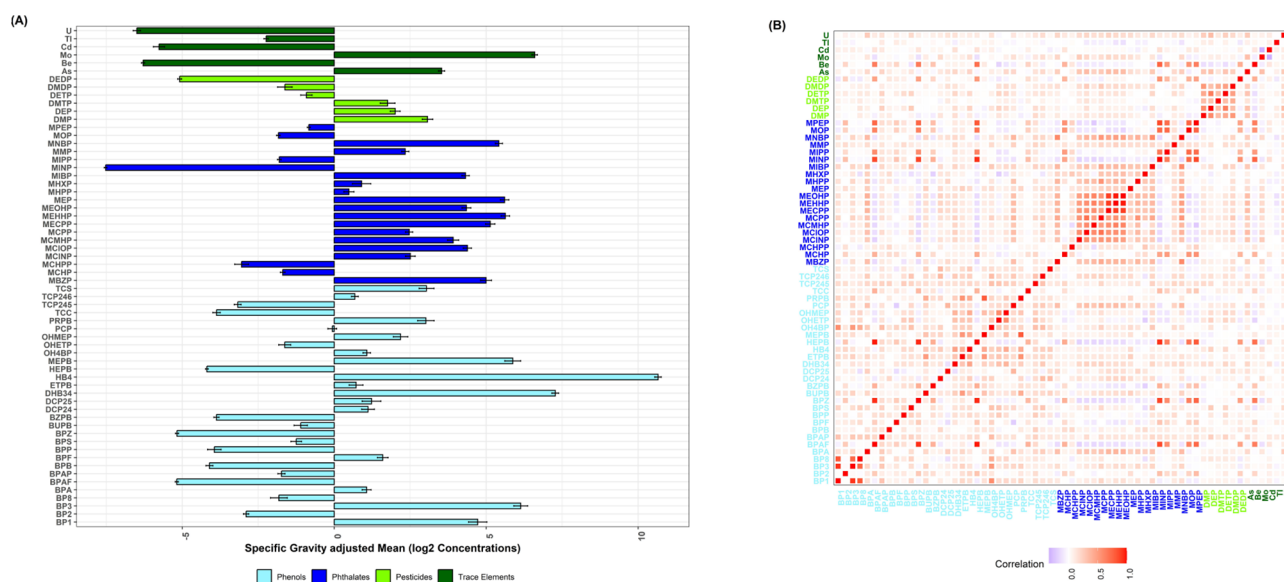
Figure 2. Specific gravity-adjusted and log-transformed (base 2) mean concentration and correlation plot of urinary environmental chemicals measured in children. Figure 2A denotes the log-transformed (base 2) mean concentration (standard error) of all 62 environmental chemicals included in this study. Figure 2B depicts the correlations among the 62 environmental chemicals. The chemical groups are color-coded.

Table 1. Characteristics of Mothers and Children Included in the Analysis from the CHARGE Cohort[a]

N = 479	All (Mean (Sd) or N (%))	TD	ASD	P-value
child sex				0.99
female	91 (19)	47	44	
male	388 (81)	201	187	
child race/ethnicity				0.25
white (non-hispanic)	246 (51.36)	135 (54.44)	111 (48.05)	
non-white (non-Hispanic)	102 (21.29)	46 (18.55)	56 (24.24)	
Hispanic any race	131 (27.35)	67 (27.02)	64 (27.71)	
child age at assessment (in years)	3.94 (0.75)	3.82 (0.75)	4.05 (0.73)	<0.01
child year of birth (baseline 2000)[c]	6.86 (3.08)	6.48 (2.91)	7.26 (3.21)	<0.01
parental homeowner status				0.09
no	137 (28.60)	62 (25.00)	75 (32.47)	
yes	342 (71.40)	186 (75.00)	156 (67.53)	
maternal age at child's birth	30.57 (5.56)	30.42 (5.43)	30.73 (5.71)	0.38
maternal metabolic condition[d]				0.01
healthy (BMI < 25) weight and no metabolic conditions	230 (48.02)	124 (50.00)	106 (45.89)	
overweight (BMI: 25−29.9) and no metabolic conditions	102 (21.29)	60 (24.19)	42 (18.18)	
obese (BMI > 30), no other metabolic conditions	68 (14.20)	36 (14.52)	32 (13.85)	
any hypertensive disorder (including obesity) or diabetes	79 (16.49)	28 (11.29)	51 (22.08)	
MEPB[b] (in ng/mL)	5.87 (2.87)	5.47 (2.81)	6.31 (2.87)	<0.01
DEP[b] (in ng/mL)	2.01 (1.78)	1.76 (1.64)	2.27 (1.89)	<0.01
PRPB[b] (in ng/mL)	3.02 (2.96)	2.66 (2.94)	3.39 (2.95)	<0.01

[a]All chemical concentrations were transformed to log (base 2) and corrected for specific gravity. [b]Top three chemicals in terms of weights from WQS regression. [c]All children were born after 2000. [d]The mutually exclusive covariate maternal metabolic condition was created in previous studies by merging BMI categories with any hypertensive disorder and obesity and was shown to be associated with neuro-developmental outcomes in children.[73] P-values for the difference between ASD and TD groups were calculated using the Fisher exact test for categorical variables and the Wilcoxon rank-sum test for continuous variables. ASD, Autism Spectrum Disorder; TD, typical development; BMI, body mass index; MEPB, methyl paraben; DEP, diethyl-phosphate; PRPB, propyl paraben.

■ RESULTS

There were 62 chemical exposures measured in the urine samples of children, which were included in this analysis. The list of all 62 target chemicals is presented in Supplemental Table S1, and their LODs (and % detected above LOD) are presented in Supplemental Table S2. Supplemental Tables S3−S5 show the log-transformed and specific-gravity-adjusted urinary concentrations of all 62 chemicals for all 479 children,

231 ASD children, and 248 TD children, respectively. Supplemental Tables S6, S7, and S8 show the log-transformed and specific-gravity-unadjusted urinary concentrations of all 62 chemicals for all 479 children, 231 ASD children, and 248 TD children, respectively. Supplemental Tables S9 and S10 present univariate associations between ASD diagnosis and log-transformed (base = 2), specific gravity-adjusted (and unadjusted) urinary biomarker concentrations, respectively.
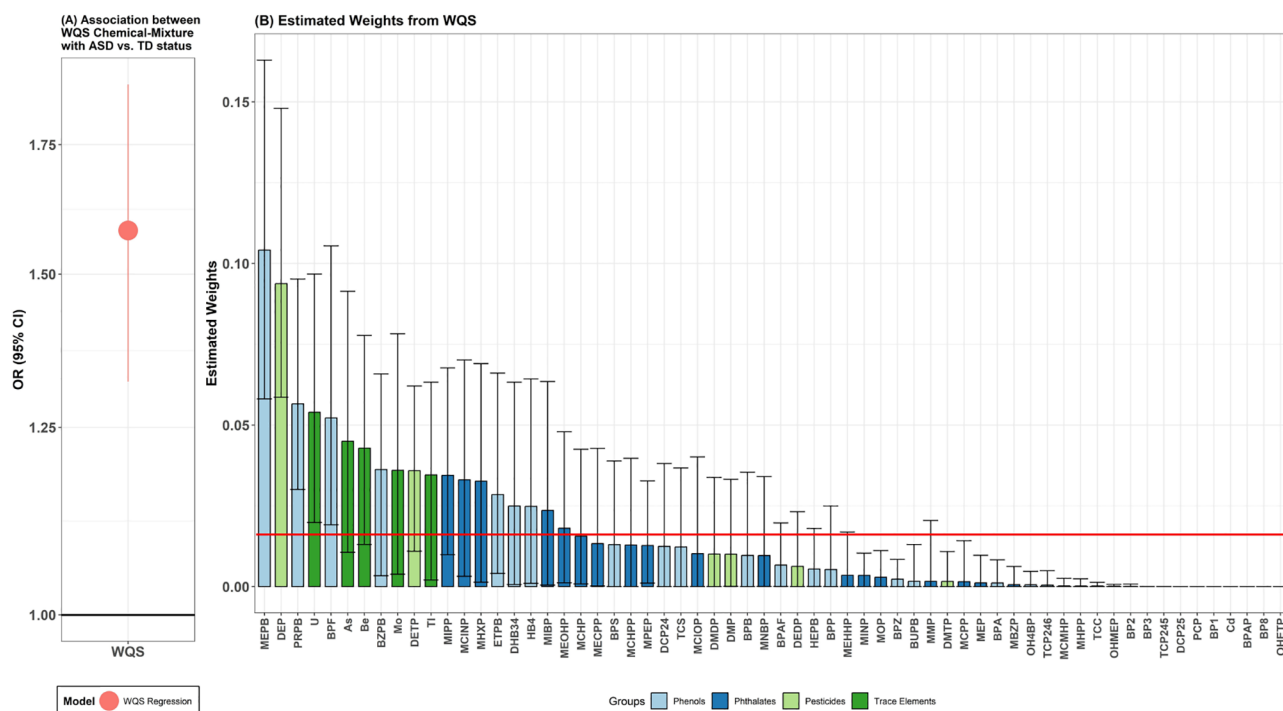
Figure 3. Results from WQS chemical mixture regression. (A) The overall odds ratio (OR) between the WQS chemical mixture with ASD vs TD status and (B) the corresponding weights contributed to the overall OR (note that the weights sum up to 1).
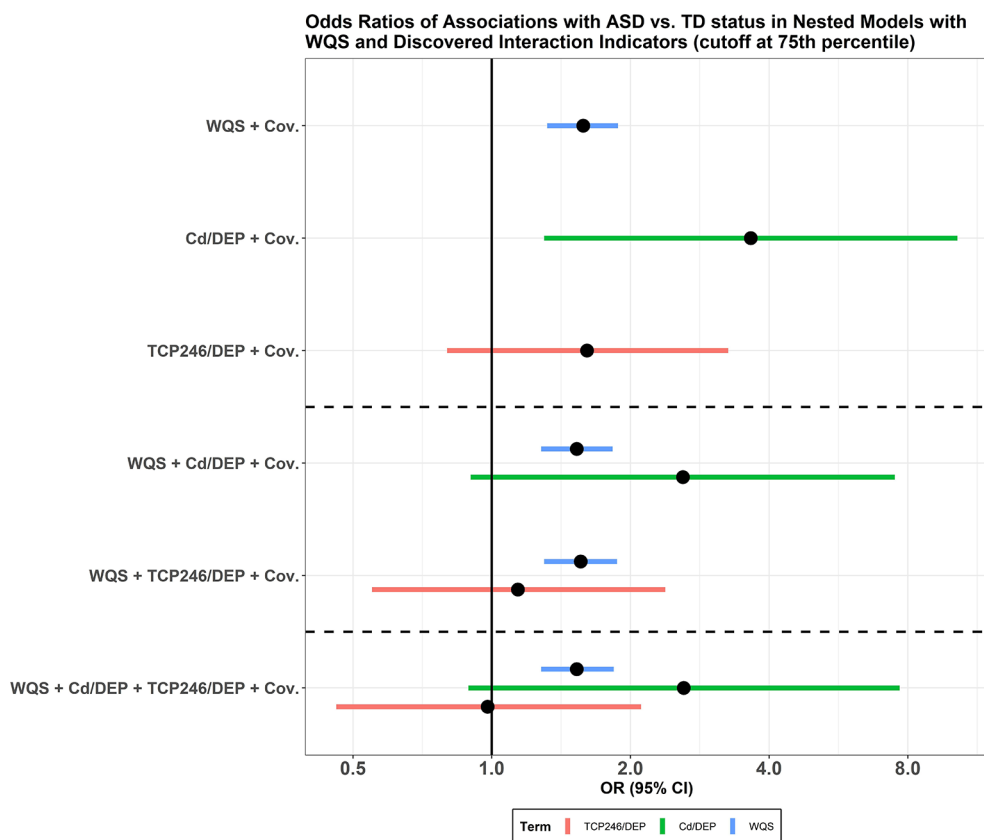

Figure 4. Results from nested linear models with WQS and discovered interaction indicators (cutoff set at 75th percentile) and the WQS chemical mixture. In the top three models, the WQS chemical mixture and the two interaction indicators were used in separate models. Both interaction indicators were adjusted with the WQS chemical mixture in the following two models. In the last model, interaction indicators and the WQS chemical mixture were put in the same model. All models were adjusted for the covariates listed above.

Among 62 chemicals, 42 had a more than 60% detection rate above the LOD (Supplemental Table S2). The specific gravity-adjusted concentrations and correlation matrices of the chemicals are presented in Figure 2.

There were moderate to strong (0.3 to 0.7) within-group correlations among pesticides and phenols. The distributions of the child's sex and race/ethnicity were not significantly different between ASD and TD children (Table 1). Furthermore, there was no significant difference in the parental homeowner status. However, children with ASD were more likely to be older at their age of assessment, and their mothers were more likely to have any hypertensive disorder or diabetes in any BMI category. The chemical concentrations of methyl paraben, DEP metabolite, and propyl paraben (the top three chemicals based on weights from WQS) were significantly higher in children with ASD. It is worth noting that several organophosphorus pesticides including chlorpyrifos, malathion, and diazinon can be metabolized to DEP.

■ WQS- SIRF RESULT

In the WQS model (with binary outcome ASD vs TD and without any interaction term), the mixture index was significantly associated with higher odds of ASD (OR[95% CI]: 1.58[1.32, 1.88]) after controlling for covariates. There were 20 chemicals with higher than chance contribution (weight $>1/62$) to the overall mixture effect. The top five chemicals were methyl paraben, diethyl-phosphate, propyl paraben, trace-metal uranium, and Bisphenol F (BPF). The estimated weights (and the corresponding 95% CIs) were presented in Figure 3.

WQS-SiRF searched for interactions of multiple orders (2 or more) and found two synergistic two-order interactions with more than 75% stability. The interactions were (1) urinary trace element cadmium (Cd) and DEP, denoted by Cd/DEP; and (2) environmental phenol 2,4,6-trichlorophenol (TCP-246) and DEP, denoted by TCP-246/DEP. However, both interactions were only observed in a subset of the sample whose urinary chemical concentrations of Cd, DEP, and TCP-246 were above certain thresholds. Therefore, based on a 75th percentile threshold cutoff, we created two separate interaction indicators to test these discovered interactions for association analysis. For example, if both the specific gravity-adjusted concentrations of Cd and DEP were more than the 75th percentile, then the interaction indicator Cd/DEP would be nonzero; else, it would be zero. In the sample, the calculated prevalences of these interactions were 5% and 8.4% for Cd/DEP and TCP-246/DEP, respectively. The results of SiRF from all three different data partitions are presented in Supplemental Table S11.

In two separate adjusted models (after controlling for the main WQS chemical mixture and covariates), each interaction indicator was associated with increased odds of ASD, 2.60 [0.90, 7.50] and 1.14 [0.55, 2.38] for Cd/DEP and TCP-246+/DEP, respectively. ORs and corresponding CIs in the forest plot are presented in Figure 4. Among the two interactions, Cd/DEP had the strongest association, and in all of the models, the WQS chemical mixture remained statistically significant, with just a slight change in the ORs.

In the sensitivity analyses, (1) the interactions Cd/DEP and TCP-246/DEP were replicated when the WQS-SiRF algorithm was refitted without chemicals whose percent of detection above LOD was less than 60% (Supplemental Table S12). (2) Furthermore, the gradual increase in the number of bootstraps,

from 250, 500, to 1000, did not alter the results. Both of the discovered interactions remained unaltered when the whole data set ($n = 479$) was used to test the model trained on 75% data (Supplemental Table S13). Moreover, (3) the directionality of the ORs did not alter even when the interaction threshold of the 75th percentile was changed to the 67th percentile (Supplemental Figure S1), and (4) the interactions Cd/DEP and TCP-246/DEP were not found in the permutation tests. We also compared the interactions through a BKMR analysis (Supplemental Figure S2). The interactions, DEP/Cd and DEP/TCP-246 from BKMR between were challenging to interpret, possibly due to the nature of how interactions were analyzed based on projections.

■ DISCUSSION

We leveraged data from the CHARGE study to assess the synergistic interactions among environmental chemicals, pesticides, phthalates, phenols, and trace elements and ASD. Utilizing WQS-SiRF, we found two suggestive synergistic interactions associated with increased odds of ASD diagnosis between (1) Cd and DEP and (2) 2,4,6-trichlorophenol and DEP among children with the urinary concentration of interacting chemicals over certain thresholds. When the main WQS mixture and the necessary covariates were controlled, cadmium/DEP and TCP-246/DEP were associated with increased odds of ASD, respectively. Between the two interactions, cadmium/DEP had the strongest association and was previously shown to form chemical complexes.[13] The identified interactions could be experimentally tested and are potentially biologically meaningful. This paper is a continuation of the study of the main effects by Bennett et al.,[13] which concluded that many urinary chemicals were associated with increased odds of ASD at 2−5 years of age. The present study adds value by examining multiordered synergistic interactions between exposures to pesticides, phthalates, phenols, and trace elements and ASD and providing evidence for suggestive two-order interactions between Cd/DEP and TCP-246/DEP.

A major aim of studying the effect of chemical mixtures is to determine whether there is any departure from the additive effect of the individual chemicals. Moreover, interactions among the mixture components may be dose-dependent compared with remaining constant over the entire dose range. Konemann and Pieters[74] and Gennings et al.[75] showed that interactions among environmental exposures may be dose-dependent. Moreover, the U.S. EPA[76] and Carpy et al.[77] suggested that lower exposure ranges of chemical mixtures might be associated with additivity, while synergistic (i.e., greater than additive) interactions might occur as the dose increases. This current method, WQS-SiRF, identifies toxicologically mimicking interactions detected only beyond certain thresholds, which denotes an essential difference from the multiplicative interaction. Therefore, such interactions can potentially identify suggestive interactions of potential biological relevance, which can later be validated or discarded in laboratory-based experimental studies.

The novelty of this work is the demonstration of the utility of integrating exposure mixture model analytical methods widely used in environmental health research with a machine learning tool to identify synergistic interactions among multiple environmental chemicals in ASD. Biological confirmation of the discovered interactions was beyond the goal of this study, which used observational case-control data. However, the proposed methodology provides a way to

discover possible multiordered interactions within environmental exposure mixtures on a health outcome, which could later be validated in experimental studies. Other exposure mixture analytical approaches, such as BKMR or g-computation, can also be coupled with SiRF for the investigation of potential interaction effects. When the directionality of the association between exposures and the outcome is hypothesized beforehand and interest lies in the joint mixture effect in a certain direction, a WQS-SiRF framework can be utilized. Alternatively, when the directionality of the association is determined in a data-driven way, and the interest lies in the overall effect of the exposure mixtures, a BKMR-SiRF framework can be implemented instead.[78] BKMR or related models estimate interactions based on mathematical projections or multiplications, and therefore future work comparing and contrasting interaction results across different exposure mixture methods coupled with SiRF would be informative. Lastly, if the interest lies in the overall mixture effect, irrespective of the hypothesized directionality, a quantile g-computation-SiRF algorithm can be implemented.[79] Another interesting area for future investigations in this field would be the determination of threshold cutoffs. Similar to the Extreme Gradient Boosting algorithm,[80] sparsity-aware algorithms for sparse data and novel quantile sketches for approximate tree learning algorithms can be implemented in the iterative Random Forests.

There are few studies on interactions associated with ASD, including gene-environment,[81] social,[82] and chemical[4] factors. Moreover, there is a lack of studies demonstrating chemical–chemical interactions in this context. Previous studies have shown an association between heavy metals, like cadmium, and ASD.[83,84] Kern et al. discovered that cadmium and other trace elements were significantly lower in the hair of children with autism than others.[85] This supports the concept that children with autism may have issues excreting cadmium, resulting in a higher body burden that could contribute to symptoms of autism.[85,86] Children could be exposed to cadmium through inhalation and ingestion. It is commonly found in the food chain, soil, cigarette smoke, and manufactured products.[84] Research on pesticide exposure during childhood, specifically glyphosate,[87,88] chlorpyrifos,[88] diazinon,[88] and the development of ASD continues to emerge.[89–91] Potential routes of pesticide exposure in children include food contaminated with pesticides (ingestion), in utero or through breastmilk, and household exposures via dermal contact.[92,93] However, there is a lack of studies showing any associations between the interaction of DEP and TCP-246 with ASD. Regarding possible biochemical significance, the cation, Cd^{2+} forms a complex with phosphate ester, particularly with DEP ($C_4H_{10}O_4P^-$), forming cadmium diethyl phosphate, $C_4H_{10}CdO_4P^-$.[94,95] Although for the TCP-246+/DEP+ interaction, many details are not known, a chemical complex "2,4,6-trichlorophenyl dialkyl phosphate" was patented (in 1952) for use as parasiticides and control of agricultural and household pests through aqueous suspensions employed as sprays.[96] However, the activities of both chemical complexes in biological media are not known in detail.

Our study limitations include the following: (1) The urine samples were collected postdiagnosis, i.e., months and sometimes years after the symptoms emerged, with only few urine samples collected at the time of diagnosis. Therefore, we cannot rule out reverse causation and that the disease or associated lifestyle changes due to diagnosis may affect chemical concentrations measured in this study and not vice versa. (2) Urinary measurements of the environmental chemicals assessed in this study represent recent exposures due to their short half-lives in the human body. In the absence of repeated urine samples collected at various time points,[13,99–101] we cannot rule out the possibility of exposure misclassification influencing results. (3) Because of the limited sample size, we did not study potential sex-specific associations with ASD diagnosis, although sexually dimorphic effects have been previously documented.[3] (4) Additionally, we used the same confounders used in the original analysis by Bennett et al.[13] However, these confounders were selected based on methyl paraben exposure because it has one of the strongest associations in the unadjusted model. (4) Similar to large case-control studies, residual confounding is possible. However, our results remained unaltered after adjusting for multiple confounders and covariates, negating residual confounding as the sole explanation. (5) The choice of cutoffs at the 75th or 67th percentile is ad-hoc and sample-specific and therefore needs to be replicated in a separate independent study population. Further, using random intersection trees within the SiRF algorithm makes it difficult to extract the absolute threshold cutoffs directly. Future methodological studies are required to address this limitation. (6) In the present analysis, the same chemicals were used in the WQS and then again in the SiRF, raising the possibility of overfitting. A training, testing, and validation data split in an ideal large sample scenario would potentially guard against overfitting. However, in this moderate sample-sized study, the use of random subsets and repeated holdouts in training and testing samples of WQS and the drawing of a large number of bootstrapped samples with different training and testing splits in the SiRF could potentially induce a robust guard against overfitting. (7) It should be noted that organophosphate insecticides are metabolized in the body, forming dialkyl phosphate metabolites that are exerted through the urine, such as DEP. DEP is a common biomarker of exposure to organophosphate insecticides detected in urine,[97,98] and can indicate exposure to organophosphate insecticides, as well as their metabolites. Therefore, the reported complexes of Cd-DEP and DEP-TCP246 may indicate potential interactions with the parent organophosphate insecticides or other metabolites beyond DEP and require further investigation, which needs to be corroborated in other human and experimental studies to elucidate their potential effects on ASD diagnosis.

Our study also had several strengths: (1) CHARGE is a well-established case-control study with extensive demographic and covariate data, which allowed us to assess a wide range of real-world environmental chemical exposures in children, along with available data on ASD in a moderate sample size. (2) This is the first study to combine exposure mixture methods and machine learning tools to discover interactions that mimic classical threshold-based toxicological dose–response interactions, providing a meaningful way to extract potentially plausible mechanistic insights that are worthy of further investigation. (3) Even though the main effects may not be apparent or reach statistical for some chemicals, nonlinear interactions between chemicals may still exist and be of considerable importance. Therefore, a strength of the WQS-SiRF algorithm is that it can efficiently accommodate many chemical exposures without needing a prior step of variable selection based on individual chemical associations. (4) These toxicologically mimicking interactions are only present in a

subset of the sample, and therefore can be thought of as "personalized and precision" interactions. (5) WQS-SiRF can efficiently search for high-order interactions; therefore, the intended order should not be specified beforehand. (6) Regarding practical implementation, the WQS-SiRF algorithm is relatively fast and user-friendly, with both having robust R packages available for future studies.

In conclusion, we introduced a novel way of discovering threshold-based chemical interactions among urinary biomonitoring data from a case-control study. To the best of our knowledge, this is the first paper that combines the inferential power of WQS and the predictive accuracy of a machine-learning algorithm to discover threshold-based, personalized biologically suggestive interactions among environmental chemical exposures associated with ASD.

■ ASSOCIATED CONTENT

Data Availability Statement

The data set is freely available at the Human Health Exposure Analysis Resource (HHEAR) Data Center (https://hheardatacenter.mssm.edu/PublicFile/ViewPublicFile?projectId=17). In particular, we have used the following files for the analysis: (1) *Chemical concentrations data*: 1461_TARGETED_DATA.csv (DOI: 10.36043/1461_222). (2) *Epidemiologic data*: 1461_EPI_DATA.csv (DOI: 10.36043/1461_219). (3) *Semantic Data Dictionary (SDD)*: SDD-2016-1461.xlsx (DOI: 10.36043/1461_630_2022.2).

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.3c00848.

> Figure showing results from nested linear models with WQS and discovered interaction indicators (cutoff set at 67th percentile) and WQS chemical mixture; figure showing sensitivity analysis using Bayesian Kernel Machine Regression (BKMR) model; list of all chemical names, abbreviations, limit of detection (LOD) for individual chemicals, and percent detection above the LOD by chemical classes; distribution of log-transformed (base = 2) specific gravity-adjusted and unadjusted urinary phenol, phthalate, and trace element biomarker concentrations (ng/mL) among all 479, 231 ASD, and 248 TD participants; univariate associations between ASD diagnosis and log-transformed (base = 2), specific gravity-adjusted and unadjusted urinary phenol, phthalate, and trace element biomarker concentrations; results of SiRF from the three different data partitions and tuning parameters for WQS-SiRF (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Vishal Midya − *Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, New York 10029, United States;* ⓞ orcid.org/0000-0002-6643-5176; Email: vishal.midya@mssm.edu

Authors

Cecilia Sara Alcala − *Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, New York 10029, United States*

Elza Rechtman − *Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, New York 10029, United States*

Jill K. Gregory − *Instructional Technology Group, Icahn School of Medicine at Mount Sinai, New York, New York 10029, United States*

Kurunthachalam Kannan − *Department of Pediatrics and Department of Environmental Medicine, New York University School of Medicine, New York, New York 10016, United States;* ⓞ orcid.org/0000-0002-1926-7456

Irva Hertz-Picciotto − *Department of Public Health Sciences, School of Medicine, University of California at Davis, Davis, California 95616, United States; UC Davis MIND (Medical Investigations of Neurodevelopmental Disorders) Institute, University of California at Davis, Sacramento, California 95817, United States*

Susan L. Teitelbaum − *Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, New York 10029, United States*

Chris Gennings − *Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, New York 10029, United States*

Maria J. Rosa − *Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, New York 10029, United States;* ⓞ orcid.org/0000-0002-1509-7664

Damaskini Valvi − *Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, New York 10029, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.est.3c00848

Author Contributions

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) *Diagnostic and statistical manual of mental disorders*, 5th ed.; American Psychiatric Association, 2013.

(2) Maenner, M. J.; Shaw, K. A.; Baio, J.; Washington, A.; Patrick, M.; DiRienzo, M.; Christensen, D. L.; Wiggins, L. D.; Pettygrove, S.;

Andrews, J. G.; Lopez, M.; Hudson, A.; Baroud, T.; Schwenk, Y.; White, T.; Rosenberg, C. R.; Lee, L.-C.; Harrington, R. A; Huston, M.; Hewitt, A.; Esler, A.; Hall-Lande, J.; Poynter, J. N.; Hallas-Muchow, L.; Constantino, J. N.; Fitzgerald, R. T.; Zahorodny, W.; Shenouda, J.; Daniels, J. L.; Warren, Z.; Vehorn, A.; Salinas, A.; Durkin, M. S.; Dietz, P. M. Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2016. *MMWR Surveillance summaries.* **2020**, *69*, 1.

(3) Maenner, M. J.; Shaw, K. A.; Bakian, A. V.; Bilder, D. A.; Durkin, M. S.; Esler, A.; Furnier, S. M.; Hallas, L.; Hall-Lande, J.; Hudson, A.; Hughes, M. M.; Patrick, M.; Pierce, K.; Poynter, J. N.; Salinas, A.; Shenouda, J.; Vehorn, A.; Warren, Z.; Constantino, J. N.; DiRienzo, M.; Fitzgerald, R. T.; Grzybowski, A.; Spivey, M. H.; Pettygrove, S.; Zahorodny, W.; Ali, A.; Andrews, J. G.; Baroud, T.; Gutierrez, J.; Hewitt, A.; Lee, L.-C.; Lopez, M.; Mancilla, K. C.; McArthur, D.; Schwenk, Y. D.; Washington, A.; Williams, S.; Cogswell, M. E. Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2018. *MMWR Surveillance Summaries.* **2021**, *70*, 1.

(4) Kalkbrenner, A. E.; Schmidt, R. J.; Penlesky, A. C. Environmental chemical exposures and autism spectrum disorders: a review of the epidemiological evidence. *Curr. Probl Pediatr Adolesc Health Care.* **2014**, *44*, 277−318.

(5) Lam, J.; Sutton, P.; Kalkbrenner, A.; Windham, G.; Halladay, A.; Koustas, E.; Lawler, C.; Davidson, L.; Daniels, N.; Newschaffer, C.; Woodruff, T. A systematic review and meta-analysis of multiple airborne pollutants and autism spectrum disorder. *PloS one.* **2016**, *11*, No. e0161851.

(6) Gong, T.; Dalman, C.; Wicks, S.; Dal, H.; Magnusson, C.; Lundholm, C.; Almqvist, C.; Pershagen, G. Perinatal Exposure to Traffic-Related Air Pollution and Autism Spectrum Disorders. *Environ. Health Perspect.* **2017**, *125*, 119−126.

(7) Raz, R.; Levine, H.; Pinto, O.; Broday, D. M.; Yuval; Weisskopf, M. G. Traffic-Related Air Pollution and Autism Spectrum Disorder: A Population-Based Nested Case-Control Study in Israel. *Am. J. Epidemiol.* **2018**, *187*, 717−725.

(8) Ritz, B.; Liew, Z.; Yan, Q.; Cuia, X.; Virk, J.; Ketzel, M.; Raaschou-Nielsen, O. Air pollution and Autism in Denmark. *Environ. Epidemiol.* **2018**, *2*, No. e028.

(9) Pagalan, L.; Bickford, C.; Weikum, W.; Lanphear, B.; Brauer, M.; Lanphear, N.; Hanley, G. E.; Oberlander, T. F.; Winters, M. Association of Prenatal Exposure to Air Pollution With Autism Spectrum Disorder. *JAMA Pediatr.* **2019**, *173*, 86−92.

(10) Rossignol, D. A.; Genuis, S. J.; Frye, R. E. Environmental toxicants and autism spectrum disorders: a systematic review. *Transl Psychiatry.* **2014**, *4*, No. e360.

(11) Grandjean, P.; Landrigan, P. J. Developmental neurotoxicity of industrial chemicals. *Lancet.* **2006**, *368*, 2167−78.

(12) Wang, M.; Hossain, F.; Sulaiman, R.; Ren, X. Exposure to Inorganic Arsenic and Lead and Autism Spectrum Disorder in Children: A Systematic Review and Meta-Analysis. *Chem. Res. Toxicol.* **2019**, *32*, 1904−1919.

(13) Bennett, D. H.; Busgang, S. A.; Kannan, K.; Parsons, P. J.; Takazawa, M.; Palmer, C. D.; Schmidt, R. J.; Doucette, J. T.; Schweitzer, J. B.; Gennings, C.; Hertz-Picciotto, I. Environmental exposures to pesticides, phthalates, phenols and trace elements are associated with neurodevelopment in the CHARGE study. *Environ. Int.* **2022**, *161*, 107075.

(14) Harley, K. G.; Gunier, R. B.; Kogut, K.; Johnson, C.; Bradman, A.; Calafat, A. M.; Eskenazi, B. Prenatal and early childhood bisphenol A concentrations and behavior in school-aged children. *Environ. Res.* **2013**, *126*, 43−50.

(15) Braun, J. M. Early-life exposure to EDCs: role in childhood obesity and neurodevelopment. *Nat. Rev. Endocrinol.* **2017**, *13*, 161−173.

(16) Philippat, C.; Nakiwala, D.; Calafat, A. M.; Botton, J.; De Agostini, M.; Heude, B.; Slama, R. Prenatal Exposure to Non-persistent Endocrine Disruptors and Behavior in Boys at 3 and 5 Years. *Environ. Health Perspect.* **2017**, *125*, 097014.

(17) Engel, S. M.; Daniels, J. L. On the complex relationship between genes and environment in the etiology of autism. *Epidemiology.* **2011**, *22*, 486−488.

(18) Landrigan, P. J. What causes autism? Exploring the environmental contribution. *Current opinion in pediatrics.* **2010**, *22*, 219−225.

(19) Curtin, P.; Austin, C.; Curtin, A.; Gennings, C.; Arora, M.; Tammimies, K.; Willfors, C.; Berggren, S.; Siper, P.; Rai, D.; Meyering, K.; Kolevzon, A.; Mollon, J.; David, A. S.; Lewis, G.; Zammit, S.; Heilbrun, L.; Palmer, R. F.; Wright, R. O.; Bolte, S.; Reichenberg, A. Dynamical features in fetal and postnatal zinc-copper metabolic cycles predict the emergence of autism spectrum disorder. *Sci. Adv.* **2018**, *4*, No. eaat1293.

(20) Austin, C.; Curtin, P.; Arora, M.; Reichenberg, A.; Curtin, A.; Iwai-Shimada, M.; Wright, R. O.; Wright, R. J.; Remnelius, K. L.; Isaksson, J.; Bolte, S.; Nakayama, S. F. Elemental Dynamics in Hair Accurately Predict Future Autism Spectrum Disorder Diagnosis: An International Multi-Center Study. *J. Clin Med.* **2022**, *11*, 7154.

(21) Gennings, C. On testing for drug/chemical interactions: definitions and inference. *J. Biopharm Stat.* **2000**, *10*, 457−67.

(22) Lee, M.; Rahbar, M. H.; Samms-Vaughan, M.; Bressler, J.; Bach, M. A.; Hessabi, M.; Grove, M. L.; Shakespeare-Pellington, S.; Coore Desai, C.; Reece, J.-A.; Loveland, K. A.; Boerwinkle, E. A generalized weighted quantile sum approach for analyzing correlated data in the presence of interactions. *Biometrical Journal.* **2019**, *61*, 934−954.

(23) Rahbar, M. H.; Samms-Vaughan, M.; Kim, S.; Saroukhani, S.; Bressler, J.; Hessabi, M.; Grove, M. L.; Shakspeare-Pellington, S.; Loveland, K. A. Detoxification Role of Metabolic Glutathione S-Transferase (GST) Genes in Blood Lead Concentrations of Jamaican Children with and without Autism Spectrum Disorder. *Genes.* **2022**, *13*, 975.

(24) Colicino, E.; Pedretti, N. F.; Busgang, S. A.; Gennings, C. Per- and poly-fluoroalkyl substances and bone mineral density: Results from the Bayesian weighted quantile sum regression. *Environ. Epidemiol.* **2020**, *4*, No. e092.

(25) Kowal, D. R.; Bravo, M.; Leong, H.; Bui, A.; Griffin, R. J.; Ensor, K. B.; Miranda, M. L. Bayesian variable selection for understanding mixtures in environmental exposures. *Statistics in Medicine.* **2021**, *40*, 4850−4871.

(26) Carrico, C.; Gennings, C.; Wheeler, D. C.; Factor-Litvak, P. Characterization of Weighted Quantile Sum Regression for Highly Correlated Data in a Risk Analysis Setting. *J. Agric Biol. Environ. Stat.* **2015**, *20*, 100−120.

(27) Keil, A. P.; Buckley, J. P.; O'Brien, K. M.; Ferguson, K. K.; Zhao, S.; White, A. J. A Quantile-Based g-Computation Approach to Addressing the Effects of Exposure Mixtures. *Environ. Health Perspect.* **2020**, *128*, 47004.

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

(29) Bellavia, A.; Dickerson, A. S.; Rotem, R. S.; Hansen, J.; Gredal, O.; Weisskopf, M. G. Joint and interactive effects between health comorbidities and environmental exposures in predicting amyotrophic lateral sclerosis. *International Journal of Hygiene and Environmental Health.* **2021**, *231*, 113655.

(30) Antonelli, J.; Mazumdar, M.; Bellinger, D.; Christiani, D.; Wright, R.; Coull, B. Estimating the health effects of environmental mixtures using Bayesian semiparametric regression and sparsity inducing priors. *Annals of Applied Statistics.* **2020**, *14*, 257−275.

(31) McGee, G.; Wilson, A.; Webster, T. F.; Coull, B. A. Bayesian multiple index models for environmental mixtures. *Biometrics* **2023**, *79*, 462.

(32) Liu, J. Z.; Deng, W.; Lee, J.; Lin, P-iD; Valeri, L.; Christiani, D. C.; Bellinger, D. C.; Wright, R. O.; Mazumdar, M. M.; Coull, B. A. A Cross-Validated Ensemble Approach to Robust Hypothesis Testing of Continuous Nonlinear Interactions: Application to Nutrition-

Environment Studies. *Journal of the American Statistical Association.* **2022**, *117*, 561−573.

(33) Ferrari, F.; Dunson, D. B. Bayesian Factor Analysis for Inference on Interactions. *Journal of the American Statistical Association.* **2021**, *116*, 1521−1532.

(34) Kumbier, K.; Basu, S.; Frise, E.; Celniker, S. E.; Brown, J. B.; Yu, B. Signed iterative random forests to identify enhancer-associated transcription factor binding. *arXiv* **2018**, No. 1810.07287, DOI: 10.48550/arXiv.1810.07287.

(35) Hamm, A. K.; Hans Carter, W., Jr; Gennings, C. Analysis of an interaction threshold in a mixture of drugs and/or chemicals. *Statistics in Medicine.* **2005**, *24*, 2493−2507.

(36) Yeatts, S. D.; Gennings, C.; Wagner, E. D.; Simmons, J. E.; Plewa, M. J. Detecting Departure From Additivity Along a Fixed-Ratio Mixture Ray With a Piecewise Model for Dose and Interaction Thresholds. *J. Agric Biol. Environ. Stat.* **2010**, *15*, 510−522.

(37) Gennings, C.; Schwartz, P.; Carter, W. H.; Simmons, J. E. Detection of Departures from Additivity in Mixtures of Many Chemicals with a Threshold Model. *Journal of Agricultural, Biological, and Environmental Statistics.* **1997**, *2*, 198−211.

(38) Gibson, E. A. *Statistical and Machine Learning Methods for Pattern Identification in Environmental Mixtures*; Columbia University, 2021.

(39) Joubert, B. R.; Kioumourtzoglou, M. A.; Chamberlain, T.; Chen, H. Y.; Gennings, C.; Turyk, M. E.; Miranda, M. L.; Webster, T. F.; Ensor, K. B.; Dunson, D. B.; Coull, B. A. Powering Research through Innovative Methods for Mixtures in Epidemiology (PRIME) Program: Novel and Expanded Statistical Methods. *Int. J. Environ. Res. Public Health.* **2022**, *19*, 19.

(40) Colicino, E.; Ferrari, F.; Cowell, W.; Niedzwiecki, M. M.; Foppa Pedretti, N.; Joshi, A.; Wright, R. O.; Wright, R. J. Non-linear and non-additive associations between the pregnancy metabolome and birthweight. *Environ. Int.* **2021**, *156*, 106750.

(41) Shmueli, G. To Explain or to Predict? *Statistical Science.* **2010**, *25*, 289−310.

(42) Lampa, E.; Lind, L.; Lind, P. M.; Bornefalk-Hermansson, A. The identification of complex interactions in epidemiology and toxicology: a simulation study of boosted regression trees. *Environmental Health.* **2014**, *13*, 57.

(43) Stingone, J. A.; Pandey, O. P.; Claudio, L.; Pandey, G. Using machine learning to identify air pollution exposure profiles associated with early cognitive skills among U.S. children. *Environ. Pollut.* **2017**, *230*, 730−740.

(44) Gass, K.; Klein, M.; Chang, H. H.; Flanders, W. D.; Strickland, M. J. Classification and regression trees for epidemiologic research: an air pollution example. *Environmental Health.* **2014**, *13*, 17.

(45) Ouidir, M.; Lepeule, J.; Siroux, V.; Malherbe, L.; Meleux, F.; Rivière, E.; Launay, L.; Zaros, C.; Cheminat, M.; Charles, M.-A.; Slama, R. Is atmospheric pollution exposure during pregnancy associated with individual and contextual characteristics? A nationwide study in France. *Journal of Epidemiology and Community Health.* **2017**, *71*, 1026.

(46) Li, Y.-C.; Hsu, H-HL; Chun, Y.; Chiu, P.-H.; Arditi, Z.; Claudio, L.; Pandey, G.; Bunyavanich, S. Machine learning−driven identification of early-life air toxic combinations associated with childhood asthma outcomes. *Journal of Clinical Investigation.* **2021**, *131*, 131.

(47) Shah, R. D.; Meinshausen, N. Random intersection trees. *Journal of Machine Learning Research.* **2014**, *15*, 629−654.

(48) Basu, S.; Kumbier, K.; Brown, J. B.; Yu, B. Iterative random forests to discover predictive and stable high-order interactions. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 1943−1948.

(49) Hertz-Picciotto, I.; Croen, L. A.; Hansen, R.; Jones, C. R.; van de Water, J.; Pessah, I. N. The CHARGE study: an epidemiologic investigation of genetic and environmental factors contributing to autism. *Environ. Health Perspect.* **2006**, *114*, 1119−25.

(50) Asimakopoulos, A. G.; Thomaidis, N. S.; Kannan, K. Widespread occurrence of bisphenol A diglycidyl ethers, p-hydroxybenzoic acid esters (parabens), benzophenone type-UV filters,

triclosan, and triclocarban in human urine from Athens, Greece. *Sci. Total Environ.* **2014**, *470−471*, 1243−9.

(51) Li, A. J.; Xue, J.; Lin, S.; Al-Malki, A. L.; Al-Ghamdi, M. A.; Kumosani, T. A.; Kannan, K. Urinary concentrations of environmental phenols and their association with type 2 diabetes in a population in Jeddah, Saudi Arabia. *Environ. Res.* **2018**, *166*, 544−552.

(52) Rocha, B. A.; Asimakopoulos, A. G.; Honda, M.; da Costa, N. L.; Barbosa, R. M.; Barbosa, F., Jr; Kannan, K. Advanced data mining approaches in the assessment of urinary concentrations of bisphenols, chlorophenols, parabens and benzophenones in Brazilian children and their association to DNA damage. *Environ. Int.* **2018**, *116*, 269−277.

(53) Li, A. J.; Martinez-Moral, M.-P.; Al-Malki, A. L.; Al-Ghamdi, M. A.; Al-Bazi, M. M.; Kumosani, T. A.; Kannan, K. Mediation analysis for the relationship between urinary phthalate metabolites and type 2 diabetes via oxidative stress in a population in Jeddah, Saudi Arabia. *Environ. Int.* **2019**, *126*, 153−161.

(54) Rocha, B. A.; Asimakopoulos, A. G.; Barbosa, F., Jr; Kannan, K. Urinary concentrations of 25 phthalate metabolites in Brazilian children and their association with oxidative DNA damage. *Sci. Total Environ.* **2017**, *586*, 152−162.

(55) Li, A. J.; Banjabi, A. A.; Takazawa, M.; Kumosani, T. A.; Yousef, J. M.; Kannan, K. Serum concentrations of pesticides including organophosphates, pyrethroids and neonicotinoids in a population with osteoarthritis in Saudi Arabia. *Sci. Total Environ.* **2020**, *737*, 139706.

(56) Minnich, M. G.; Miller, D. C.; Parsons, P. J. Determination of As, Cd, Pb, and Hg in urine using inductively coupled plasma mass spectrometry with the direct injection high efficiency nebulizer. *Spectrochimica Acta Part B: Atomic Spectroscopy.* **2008**, *63*, 389−395.

(57) Kannan, K.; Stathis, A.; Mazzella, M. J.; Andra, S. S.; Barr, D. B.; Hecht, S. S.; Merrill, L. S.; Galusha, A. L.; Parsons, P. J. Quality assurance and harmonization for targeted biomonitoring measurements of environmental organic chemicals across the Children's Health Exposure Analysis Resource laboratory network. *Int. J. Hyg Environ. Health.* **2021**, *234*, 113741.

(58) Hauser, R.; Meeker, J. D.; Park, S.; Silva, M. J.; Calafat, A. M. Temporal variability of urinary phthalate metabolite levels in men of reproductive age. *Environmental health perspectives.* **2004**, *112*, 1734−1740.

(59) Lord, C.; Rutter, M.; Le Couteur, A. The autism diagnostic interview-revised (ADI-R). *J. Autism Dev. Disord.* **1994**, *24*, 659−685.

(60) Lord, C.; Rutter, M.; Le Couteur, A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of autism and developmental disorders.* **1994**, *24*, 659−685.

(61) Lord, C.; Pickles, A.; McLennan, J.; Rutter, M.; Bregman, J.; Folstein, S.; Fombonne, E.; Leboyer, M.; Minshew, N. Diagnosing autism: analyses of data from the Autism Diagnostic Interview. *Journal of autism and developmental disorders.* **1997**, *27*, 501−517.

(62) Lord, C.; Risi, S.; Lambrecht, L.; Cook, E. H.; Leventhal, B. L.; DiLavore, P. C.; Pickles, A.; Rutter, M. The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of autism and developmental disorders.* **2000**, *30*, 205−223.

(63) Risi, S.; Lord, C.; Gotham, K.; Corsello, C.; Chrysler, C.; Szatmari, P.; Cook, E. H., Jr; Leventhal, B. L.; Pickles, A. Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child & Adolescent Psychiatry.* **2006**, *45*, 1094−1103.

(64) Rutter, M; Bailey, A; Lord, C. *The social communication questionnaire: Manual*; Western Psychological Services, 2003.

(65) Oh, J.; Shin, H. M.; Kannan, K.; Busgang, S. A.; Schmidt, R. J.; Schweitzer, J. B.; Hertz-Picciotto, I.; Bennett, D. H. Childhood exposure to per- and polyfluoroalkyl substances and neurodevelopment in the CHARGE case-control study. *Environ. Res.* **2022**, *215*, 114322.

(66) Tanner, E. M.; Bornehag, C.-G.; Gennings, C. Repeated holdout validation for weighted quantile sum regression. *MethodsX.* **2019**, *6*, 2855−2860.

(67) Curtin, P.; Kellogg, J.; Cech, N.; Gennings, C. A random subset implementation of weighted quantile sum (WQSRS) regression for analysis of high-dimensional mixtures. *Communications in Statistics - Simulation and Computation.* **2021**, *50*, 1119−1134.

(68) Agresti, A. *Categorical data analysis*; John Wiley & Sons, 2003.

(69) Basu, S.; Kumbier, K.; Brown, J. B.; Yu, B. Iterative random forests to discover predictive and stable high-order interactions. *Proceedings of the National Academy of Sciences.* **2018**, *115*, 1943−1948.

(70) Kumbier, K.; Basu, S.; Brown, J. B.; Celniker, S.; Yu, B. Refining interaction search through signed iterative Random Forests. *bioRxiv* **2018**, No. 467498.

(71) Breiman, L. Random forests. *Machine learning.* **2001**, *45*, 5−32.

(72) van Buuren, S.; Groothuis-Oudshoorn, K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software.* **2011**, *45*, 1−67.

(73) Krakowiak, P.; Walker, C. K.; Bremer, A. A.; Baker, A. S.; Ozonoff, S.; Hansen, R. L.; Hertz-Picciotto, I. Maternal Metabolic Conditions and Risk for Autism and Other Neurodevelopmental Disorders. *Pediatrics.* **2012**, *129*, No. e1121-e1128.

(74) Konemann, W. H.; Pieters, M. N. Confusion of concepts in mixture toxicology. *Food Chem. Toxicol.* **1996**, *34*, 1025−31.

(75) Gennings, C.; Carter, W. H.; Campain, J. A.; Bae, D-s; Yang, R. S. Statistical analysis of interactive cytotoxicity in human epidermal keratinocytes following exposure to a mixture of four metals. *Journal of Agricultural, Biological, and Environmental Statistics.* **2002**, *7*, 58−73.

(76) USEPA. *Guidance on cumulative risk assessment of pesticide chemicals that have a common mechanism of toxicity*; USEPA Office of Pesticide Programs, 2002.

(77) Carpy, S. A.; Kobel, W.; Doe, J. Health risk of low-dose pesticides mixtures: a review of the 1985−1998 literature on combination toxicology and health risk assessment. *J. Toxicol Environ. Health B Crit Rev.* **2000**, *3*, 1−25.

(78) Bobb, J. F.; Claus Henn, B.; Valeri, L.; Coull, B. A. Statistical software for analyzing the health effects of multiple concurrent exposures via Bayesian kernel machine regression. *Environ. Health.* **2018**, *17*, 67.

(79) Day, D. B.; Sathyanarayana, S.; LeWinn, K. Z.; Karr, C. J.; Mason, W. A.; Szpiro, A. A. A Permutation Test-Based Approach to Strengthening Inference on the Effects of Environmental Mixtures: Comparison between Single-Index Analytic Methods. *Environ. Health Perspect.* **2022**, *130*, 87010.

(80) Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13, 2016; San Francisco, California, USA, pp 785−794.

(81) Tordjman, S.; Somogyi, E.; Coulon, N.; Kermarrec, S.; Cohen, D.; Bronsard, G.; Bonnot, O.; Weismann-Arcache, C.; Botbol, M.; Lauth, B.; Ginchat, V.; Roubertoux, P.; Barburoth, M.; Kovess, V.; Geoffray, M.-M.; Xavier, J. Gene× Environment interactions in autism spectrum disorders: role of epigenetic mechanisms. *Frontiers in psychiatry.* **2014**, *5*, 53.

(82) Assaf, M.; Hyatt, C. J.; Wong, C. G.; Johnson, M. R.; Schultz, R. T.; Hendler, T.; Pearlson, G. D. Mentalizing and motivation neural function during social interactions in autism spectrum disorders. *Neuroimage Clin.* **2013**, *3*, 321−31.

(83) Shiani, A.; Sharafi, K.; Omer, A. K.; Kiani, A.; Karamimatin, B.; Massahi, T.; Ebrahimzadeh, G. A systematic literature review on the association between exposures to toxic elements and an autism spectrum disorder. *Sci. Total Environ.* **2023**, *857*, 159246.

(84) Sulaiman, R.; Wang, M.; Ren, X. Exposure to Aluminum, Cadmium, and Mercury and Autism Spectrum Disorder in Children: A Systematic Review and Meta-Analysis. *Chem. Res. Toxicol.* **2020**, *33*, 2699−2718.

(85) Kern, J. K.; Grannemann, B. D.; Trivedi, M. H.; Adams, J. B. Sulfhydryl-reactive metals in autism. *J. Toxicol Environ. Health A* **2007**, *70*, 715−21.

(86) Yorbik, O.; Kurt, I.; Haşimi, A.; Oztürk, O. Chromium, cadmium, and lead levels in urine of children with autism and typically developing controls. *Biol. Trace Elem Res.* **2010**, *135*, 10−5.

(87) Ongono, J. S.; Beranger, R.; Baghdadli, A.; Mortamais, M. Pesticides used in Europe and autism spectrum disorder risk: can novel exposure hypotheses be formulated beyond organophosphates, organochlorines, pyrethroids and carbamates?-A systematic review. *Environ. Res.* **2020**, *187*, 109646.

(88) von Ehrenstein, O. S.; Ling, C.; Cui, X.; Cockburn, M.; Park, A. S.; Yu, F.; Wu, J.; Ritz, B. Prenatal and infant exposure to ambient pesticides and autism spectrum disorder in children: population based case-control study. *BMJ.* **2019**, *364*, l962.

(89) Shelton, J. F.; Hertz-Picciotto, I.; Pessah, I. N. Tipping the balance of autism risk: potential mechanisms linking pesticides and autism. *Environ. Health Perspect.* **2012**, *120*, 944−51.

(90) Miani, A.; Imbriani, G.; De Filippis, G.; De Giorgi, D.; Peccarisi, L.; Colangelo, M.; Pulimeno, M.; Castellone, M. D.; Nicolardi, G.; Logroscino, G.; Piscitelli, P. Autism Spectrum Disorder and Prenatal or Early Life Exposure to Pesticides: A Short Review. *Int. J. Environ. Res. Public Health* **2021**, *18*, 18.

(91) Biosca-Brull, J.; Pérez-Fernández, C.; Mora, S.; Carrillo, B.; Pinos, H.; Conejo, N. M.; Collado, P.; Arias, J. L.; Martín-Sánchez, F.; Sánchez-Santed, F.; Colomina, M. T. Relationship between Autism Spectrum Disorder and Pesticides: A Systematic Review of Human and Preclinical Models. *Int. J. Environ. Res. Public Health* **2021**, *18*, 18.

(92) Liu, J.; Schelar, E. Pesticide exposure and child neuro-development: summary and implications. *Workplace Health Saf.* **2012**, *60*, 235−42.

(93) Chalupka, S.; Chalupka, A. N. The impact of environmental and occupational exposures on reproductive health. *J. Obstet Gynecol Neonatal Nurs.* **2010**, *39*, 84−102.

(94) *Compound Summary for CID 129652268, Cadmium diethylphosphate*; NIH National Library of Medicine, 2022.

(95) Miner, V. W.; Prestegard, J. H.; Faller, J. W. Cadmium diethyl phosphate: structure determination and comparison to cation phospholipid complexes. *Inorganic Chemistry.* **1983**, *22*, 1862−1865.

(96) Drake, L. R.; Erbel, A. J. 2,4,6-trichlorophenyl dialkylphosphates. US Patent 2599375A, 1952.

(97) Sagiv, S. K.; Harris, M. H.; Gunier, R. B.; Kogut, K. R.; Harley, K. G.; Deardorff, J.; Bradman, A.; Holland, N.; Eskenazi, B. Prenatal Organophosphate Pesticide Exposure and Traits Related to Autism Spectrum Disorders in a Population Living in Proximity to Agriculture. *Environ. Health Perspect.* **2018**, *126*, 047012.

(98) Barkoski, J. M.; Busgang, S. A.; Bixby, M.; Bennett, D.; Schmidt, R. J.; Barr, D. B.; Panuwet, P.; Gennings, C.; Hertz-Picciotto, I. Prenatal phenol and paraben exposures in relation to child neurodevelopment including autism spectrum disorders in the MARBLES study. *Environ. Res.* **2019**, *179*, 108719.

(99) Hoppin, J. A.; Brock, J. W.; Davis, B. J.; Baird, D. D. Reproducibility of urinary phthalate metabolites in first morning urine samples. *Environ. Health Perspect.* **2002**, *110*, 515−8.

(100) Barr, D. B.; Wang, R. Y.; Needham, L. L. Biologic monitoring of exposure to environmental chemicals throughout the life stages: requirements and issues for consideration for the National Children's Study. *Environ. Health Perspect.* **2005**, *113*, 1083−91.

(101) Perrier, F.; Giorgis-Allemand, L.; Slama, R.; Philippat, C. Within-subject Pooling of Biological Samples to Reduce Exposure Misclassification in Biomarker-based Studies. *Epidemiology.* **2016**, *27*, 378−88.