Review

Statistical Methodology in Studies of Prenatal Exposure to Mixtures of Endocrine-Disrupting Chemicals: A Review of Existing Approaches and New Alternatives

Nina Lazarevic,[1] Adrian G. Barnett,[2] Peter D. Sly,[3] and Luke D. Knibbs[1,4]

[1]School of Public Health, Faculty of Medicine, The University of Queensland, Brisbane, Queensland, Australia
[2]School of Public Health and Social Work, Faculty of Health, Queensland University of Technology, Brisbane, Queensland, Australia
[3]Child Health Research Centre, The University of Queensland, Brisbane, Queensland, Australia
[4]Centre for Air Quality & Health Research and Evaluation, Glebe, New South Wales, Australia

BACKGROUND: Prenatal exposures to endocrine-disrupting chemicals (EDCs) during critical developmental windows have been implicated in the etiologies of a wide array of adverse perinatal and pediatric outcomes. Epidemiological studies have concentrated on the health effects of individual chemicals, despite the understanding that EDCs act together via common mechanisms, that pregnant women are exposed to multiple EDCs simultaneously, and that substantial toxicological evidence of adverse developmental effects has been documented. There is a move toward multipollutant models in environmental epidemiology; however, there is no current consensus on appropriate statistical methods.

OBJECTIVES: We aimed to review the statistical methods used in these studies, to identify additional applicable methods, and to determine the strengths and weaknesses of each method for addressing the salient statistical and epidemiological challenges.

METHODS: We searched Embase, MEDLINE, and Web of Science for epidemiological studies of endocrine-sensitive outcomes in the children of mothers exposed to EDC mixtures during pregnancy and identified alternative statistical methods from the wider literature.

DISCUSSION: We identified 74 studies and analyzed the methods used to estimate mixture health effects, identify important mixture components, account for nonmonotonicity in exposure–response relationships, assess interactions, and identify windows of exposure susceptibility. We identified both frequentist and Bayesian methods that are robust to multicollinearity, performing shrinkage, variable selection, dimension reduction, statistical learning, or smoothing, including methods that were not used by the studies included in our review.

CONCLUSIONS: Compelling motivation exists for analyzing EDCs as mixtures, yet many studies make simplifying assumptions about EDC additivity, relative potency, and linearity, or overlook the potential for bias due to asymmetries in chemical persistence. We discuss the potential impacts of these choices and suggest alternative methods to improve analyses of prenatal exposure to EDC mixtures. https://doi.org/10.1289/EHP2207

Introduction

A growing body of evidence suggests that prenatal exposures to environmental chemicals during critical windows of development may have lasting effects throughout the life course (Gillman 2005; Soto et al. 2008; Stroustrup and Swan 2013; Woodruff et al. 2009). Prenatal exposures to endocrine-disrupting chemicals (EDCs) are of particular concern because they can interfere with the synthesis, secretion, binding, transport, and metabolism of endogenous hormones that are involved in regulating developmental processes (Diamanti-Kandarakis et al. 2009; Meeker 2012). Exposure to EDCs during gestation can also induce epigenetic changes that have transgenerational effects, which have been observed in the children of women prenatally exposed to the miscarriage-prevention drug diethylstilbestrol (Diamanti-Kandarakis et al. 2009; Grandjean et al. 2015; Meeker 2012). Of concern is that the exposure levels required for endocrine disruption during gestation can be infinitesimally small (as low as picomolar serum concentrations) (Vandenberg et al. 2012); and, unexpectedly, lower doses may have more potent effects than higher doses (Diamanti-Kandarakis et al. 2009).

The developing fetus is uniquely sensitive to chemicals that mimic hormones because the protective mechanisms existing in adulthood are not completely functional *in utero* (Newbold et al. 2009; Romano et al. 2014). Hence, prenatal exposures to EDCs are thought to be implicated in the etiologies of several adverse perinatal and childhood outcomes, such as decreased fetal growth and length of gestation, reproductive tract defects, altered pubertal development, neurodevelopmental dysfunctions, obesity, and diabetes (Silver and Meeker 2015; Skakkebaek et al. 2011). This matter is of critical public health importance because the prevalence of many of these endocrine-related disorders and diseases has been increasing over time, a trend that cannot be explained solely by genetic factors (Darbre 2015; Skakkebaek et al. 2011).

A broad range of environmental chemicals are likely to be EDCs; many are high-production-volume chemicals: organochlorine pesticides, industrial chemicals and their byproducts, flame retardants, surfactants, plastics and plasticizers, and some metals (Diamanti-Kandarakis et al. 2009; Meeker 2012). Consequently, exposure to EDCs is widespread, multisource, and multiroute (Meeker 2012), and biomonitoring studies indicate that the pattern of exposure in pregnant women and neonates is chronic, low-dose, and involves multiple chemicals simultaneously rather than individual agents (Mitro et al. 2015; Stroustrup and Swan 2013; Woodruff et al. 2011). However, epidemiological studies of EDCs have concentrated on the health effects of individual chemicals, an approach that may not be suitable to the study of chemicals that behave like hormones (Gaudriault et al. 2017; Kortenkamp 2007).

Multiple EDCs can act via a common mechanism to produce an outcome (e.g., binding to a particular type of hormone receptor), which suggests that individual chemicals can act together at lower concentrations to achieve the same outcome than the concentration that would be required for each chemical on its own (Darbre 2015). A corollary of this postulation is that different mixtures of EDCs are also able to produce a common outcome via a common mechanism (Darbre 2015). This has been confirmed in experimental studies, *in vitro*, which have shown that combinations of xenoestrogens are able to produce significant effects even when each individual chemical is present at a dose below its no-observed-effects level (NOEL; Rajapakse et al. 2002; Silva et al. 2002). Additionally,

there is substantial toxicological evidence of adverse effects *in vivo* on both prenatal and postnatal development due to mixtures of EDCs, even with individual chemicals at levels below concentrations that cause observable effects (e.g., below the NOEL; Al-Gubory 2014; Kortenkamp 2008). Such a mixture effect was also recently observed in an epidemiological study of breast cancer using a novel biomarker of combined xenoestrogen exposure; a strong positive relationship was observed for the mixture though individual xenoestrogens showed no associations (Pastor-Barriuso et al. 2016), casting doubt on the null findings of meta-analyses of breast cancer risk and individual xenoestrogens such as p,p'-DDE (Ingber et al. 2013; Kortenkamp 2006, 2008). These findings suggest that epidemiological studies of individual chemicals with null findings may have substantially underestimated the risks of exposure to EDCs (Kortenkamp 2007).

Numerous authors have expressed a need for epidemiological studies to move beyond analyzing the health effects of individual chemicals toward the study of chemical mixtures (e.g., Braun et al. 2016; Carlin et al. 2013; Kortenkamp 2007). The study of chemical mixtures is one of the ongoing research priorities of the U.S. National Institute of Environmental Health Sciences (NIEHS) (Carlin et al. 2013) and is recognized by the U.S. National Academy of Sciences, which recommends that risk assessments consider the possibility of cumulative effects from multiple chemical exposures (Mitro et al. 2015; Woodruff et al. 2011). However, progress has been thwarted by the complexities of such studies: They place a high demand on the quantity and quality of data and require advanced statistical methods, the development of which is an active area of research (Billionnet et al. 2012; Braun et al. 2016; Sun et al. 2013; Taylor et al. 2016). The statistical challenges are considerable: High correlation between exposures can lead to inflated standard errors and instability in effect estimates; differing degrees of measurement error among exposures can bias effect estimates; and there may be insufficient power to estimate small effects in the presence of measurement error, multicollinearity, small sample size, interactions, and nonlinearity (Billionnet et al. 2012; Sun et al. 2013).

The study of chemical mixtures addresses some of the shortcomings of single-chemical epidemiology. Effect estimates may be biased in the presence of copollutant confounding (Braun et al. 2016), and the study of correlated exposures in separate models can lead to inflated false-positive discoveries. One exposure can also modify the effects of another [e.g., bisphenol A (BPA) can increase the expression of estrogen receptors and make cells more vulnerable to other EDCs; Brieño-Enríquez et al. 2012; Hayes et al. 2016], making it difficult to infer combined effects from knowledge of individual effects (Braun et al. 2016; Kortenkamp 2007). Moreover, choice of chemicals has typically been based on those of known concern or measurable with existing methods, meaning that studies may suffer from the streetlight effect (Braun et al. 2016). These issues have important regulatory implications if health effects have been incorrectly attributed to one exposure rather than to a correlated harmful exposure (Braun et al. 2016; Lenters et al. 2015a). Studies of chemical mixtures provide effects that more closely correspond to observed exposure patterns; they may enable better targeted interventions by identifying which exposures out of a set are the most detrimental to health; and they can reveal how exposures interact, whether their effects are additive, antagonistic, or synergistic (Braun et al. 2016; Sun et al. 2013).

There is no current consensus on the appropriate statistical methodology for use in epidemiological studies of chemical mixtures (Braun et al. 2016). The existing methodological reviews have concentrated on multipollutant methods in air pollution epidemiology and the analysis of exposome-health associations (Agier et al. 2016; Billionnet et al. 2012; Davalos et al. 2016; Patel 2017; Stafoggia et al. 2017; Sun et al. 2013), with the latter focusing on an exploratory and hypothesis-generating search for important exposures in higher dimensional datasets rather than etiologic research questions. Moreover, the analysis of EDCs requires methods that can account for nonmonotonicity in exposure–response relationships (i.e., changes in the sign of the slope are possible; Vandenberg et al. 2012). Statistical methods for multiple correlated exposures are also required for omics data; however, the focus there is on methods suitable for ill-conditioned problems where the number of parameters may exceed the sample size (Chadeau-Hyam et al. 2013).

A wide array of statistical methods were presented and compared in a 2015 workshop hosted by the NIEHS, where participants stressed the importance of choosing a method based on the specific research question and the need for improving existing methods with subject-specific knowledge (NIEHS 2015; Taylor et al. 2016). Subsequently, Braun et al. (2016) explicated three research questions that can be addressed by mixtures analyses, including the estimation of mixture health effects, and the identification of important mixture components and their interactions. Recent simulation studies have compared subsets of the available methods for identifying important exposures and interactions (Agier et al. 2016; Barrera-Gómez et al. 2017; Lenters et al. 2018; Sun et al. 2013). No single method has emerged as consistently superior, and choice of method may be context-dependent (NIEHS 2015; Sun et al. 2013).

We conducted a review to identify the statistical methods used in studies of prenatal coexposures to EDCs and identified additional applicable methods. We compared methods within the three research questions explicated by Braun et al. (2016), as well as in their ability to account for nonmonotonicity in exposure–response relationships and to assess windows of exposure susceptibility in a multipollutant context. We aimed to arrive at an improved understanding of the methods used for studying EDC mixtures, along with an analysis of issues affecting method choice that may help to inform methodological decisions in future studies of prenatal exposure to EDC mixtures.

Methods

We searched Embase, MEDLINE, and Web of Science for observational studies of prenatal exposure to EDC mixtures and pregnancy or childhood outcomes published between 1997 and September/October 2016 (see Tables S1 and S2 in the Supplemental Material for Embase/MEDLINE and Web of Science search queries, respectively). Search terms specified both known and suspected EDCs and included generic terms such as "endocrine disrupt*," "xenobiotic," and "pesticide." We also included a combination of general and specific search terms related to the prenatal and perinatal period and endocrine-sensitive outcomes. Both outcome and exposure search terms were chosen from reviews of EDCs, pregnancy outcomes, and child health (Meeker 2012; Silver and Meeker 2015; Slama and Cordier 2010). Search terms specifying the study of multiple chemicals included "multipollutant," "co-pollutant," "co-exposure," and terms such as "combin*," "mixture," or "multip*," in proximity to terms such as "exposure," "chemical," "pollutant," "toxicant," or "effects." Only English-language primary studies of human health were eligible. Web of Science results were additionally limited to public, environmental, and occupational health.

We screened abstracts for studies that measured prenatal (maternal) exposure to more than one individual EDC and assessed associations with outcomes in offspring. Studies that assessed early postnatal exposure through breast milk, as a proxy for prenatal exposure, were also examined. We then assessed full-text articles to identify studies in which exposures were analyzed in combination rather

than separately. We excluded studies that measured only exposure to classes of chemicals (e.g., through job-exposure matrices, questionnaires, or biomarkers of cumulative exposure) or considered only maternal outcomes. We supplemented database searches with studies identified through manual searching. Our review was conducted using a systematic approach consistent with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher et al. 2009).

For each eligible study, we recorded the statistical methods used and the goals of the analysis: only control for copollutant confounding (where one exposure was the focus), estimation of mixture health effects (i.e., joint or cumulative health effects from multiple chemical exposures), identification of important mixture components, assessment of nonmonotonicity in exposure–response relationships, assessment of statistical interactions (i.e., effect measure modification by coexposures), and identification of windows of exposure susceptibility. Although nonmonotonic exposure–response relationships are of interest in EDC studies (Vandenberg et al. 2012), we discuss statistical methods for nonlinearity that are able to address both nonlinear monotonic and nonmonotonic relationships. We recorded the overall study design reported by the authors of each study, which may not be specific to the mixture analyses conducted. We categorized eligible studies by dimensionality (one aggregate-, two-, or three or more exposure variables).

In addition to the review of studies that met our eligibility requirements, we provide a narrative review of methods that are available for mixtures analyses, including methods that were not used in the studies we reviewed. For each method, we recorded the outcome variable types that can be analyzed (continuous, dichotomous, categorical, count, survival, repeated measures) and recorded whether the method can address the analytical goals and challenges of mixtures analyses, including multicollinearity.

Results

Study Selection Process

We identified 942 unique records through database and manual searches, of which 112 full-text studies were assessed for eligibility, and 38 studies were excluded for the reasons outlined in Figure 1. The 74 studies that met our eligibility criteria are listed
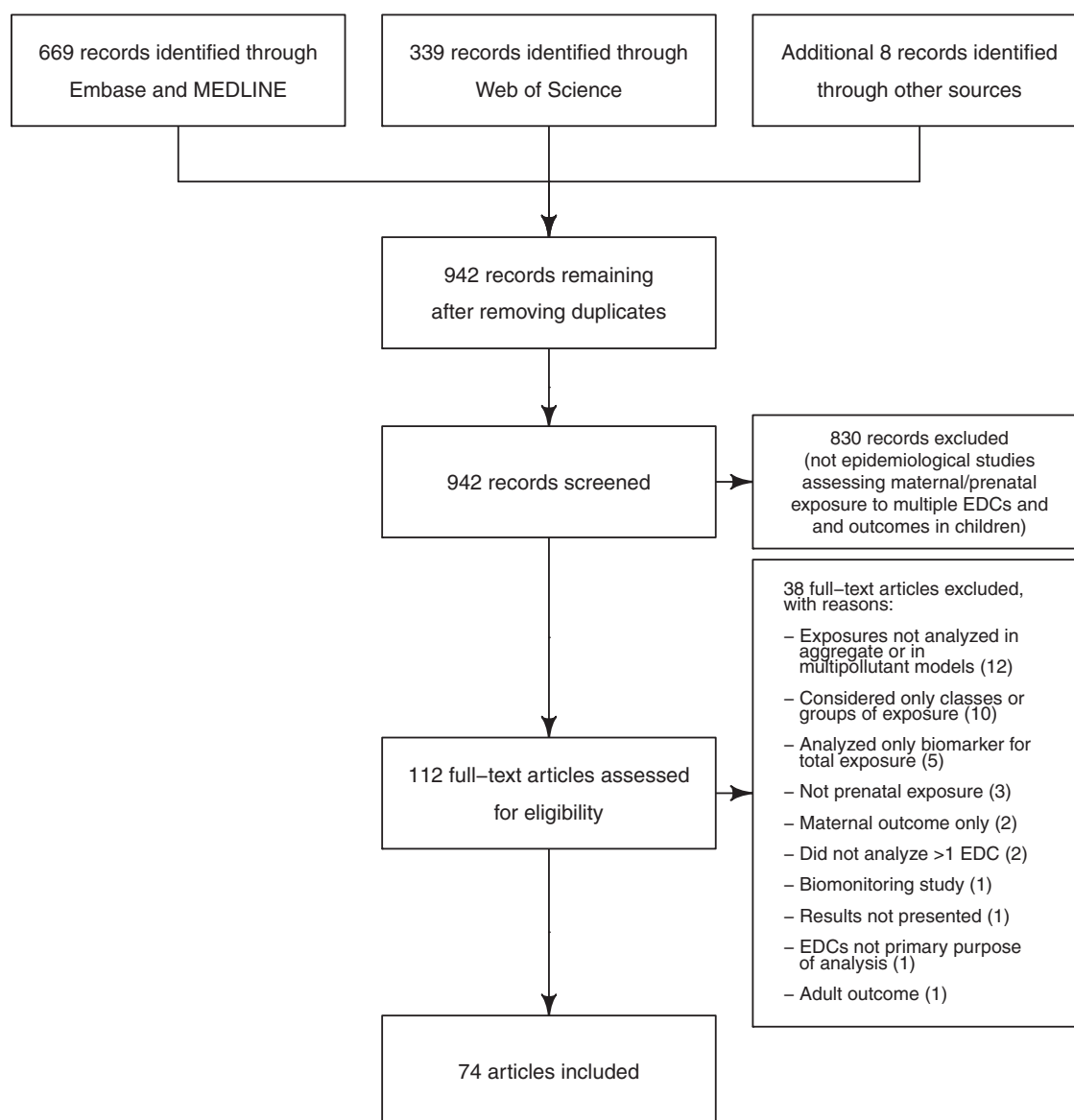

Figure 1. PRISMA flow diagram of the study selection process.

Table 1. Studies of prenatal coexposures to EDCs, categorized by dimensionality: one aggregate-, two-, three or more exposure variables presented in sections A, B, and C, respectively. A detailed summary of each study is available in Supplemental Material, Table S3. Number of studies is shown in parentheses.

Studies	Study design[a]	Goals in mixtures analyses[b]	Statistical methods[b]
SECTION A. Studies that analyzed only one aggregate exposure variable at a time			
24 studies (Barrett et al. 2016; Buscail et al. 2015; Chevrier et al. 2012; Damgaard et al. 2006; Gascon et al. 2015b; Harley et al. 2011; Jusko et al. 2010; Konishi et al. 2009; Koskenniemi et al. 2015; Miller et al. 2004; Naksen et al. 2015; Neugebauer et al. 2015; Ochiai et al. 2014; Park et al. 2009; Philippat et al. 2012; Polanska et al. 2014; Rosa et al. 2011; Sonneborn et al. 2008; Sturza et al. 2016; Tang-Péronard et al. 2015; Tran et al. 2016; Wickerham et al. 2012; Wolff et al. 2008; Yolton et al. 2013)	Prospective birth cohort (20) Case–control (4)	Estimation of mixture health effects (24) Assessment of nonmonotonicity (6)[c] Identification of windows of exposure susceptibility (6)	Generalized linear models and their extensions (23) Factor analysis (1)[d] Analysis of variance (1)[d]
SECTION B. Studies that analyzed two individual or aggregate exposure variables at a time			
19 studies (Casas et al. 2015; Gray et al. 2005; Huel et al. 2004; Huen et al. 2014; Kim et al. 2013; Kippler et al. 2012; LaRocca et al. 2016; Lignell et al. 2013; Limousi et al. 2014; Migeot et al. 2013; Roen et al. 2015; Stroustrup et al. 2016; Suzuki et al. 2010; Symanski et al. 2016; Vejrup et al. 2016; Whyatt et al. 2004; Windham et al. 2006; Wojtyniak et al. 2010; Yorifuji et al. 2011)	Prospective birth cohort (14) Retrospective cohort (2) Case–control (2) Cross-sectional (1)	Identification of important mixture components (10)[e] Assessment of interactions between exposures (9) Estimation of mixture health effects (7) Assessment of nonmonotonicity (6)[c] Identification of windows of exposure susceptibility (4) Only control for copollutant confounding, where one exposure is the focus (3)[f]	Generalized linear models and their extensions (19)
SECTION C. Studies that analyzed three or more individual or aggregate exposure variables at a time			
31 studies (Agay-Shay et al. 2015; Berg et al. 2016; Braun et al. 2014; Buckley et al. 2016a, 2016b; Casas et al. 2016; Chevrier et al. 2011; Claus Henn et al. 2016; Erkin-Cakmak et al. 2015; Fernández et al. 2007; Forns et al. 2016; Gascon et al. 2015a; Govarts et al. 2016; Heilmann et al. 2006; Jacobson et al. 2015; Kobrosly et al. 2014; Krysiak-Baltyn et al. 2012; Lenters et al. 2015a; Liew et al. 2015; Maresca et al. 2016; Marques et al. 2014; Rodrigues et al. 2016; Rull et al. 2006; Stewart et al. 2008; Swartz et al. 2015; Talbott et al. 2015; Tatsuta et al. 2014; Vafeiadi et al. 2014; Vafeiadi et al. 2015; Valvi et al. 2012; Xu et al. 2015)	Prospective birth cohort (24) Case–control (7)	Identification of important mixture components (26)[e] Estimation of mixture health effects (23) Assessment of nonmonotonicity (17)[c] Identification of windows of exposure susceptibility (8) Assessment of interactions between exposures (4) Only control for copollutant confounding, where one exposure is the focus (3)[g]	Generalized linear models and their extensions (22) Principal components analysis/ regression (5) Bayesian hierarchical regression (5) Elastic net regression (2)[h] Partial least squares regression (2)[i] Bayesian model averaging (1)[h] Hierarchical clustering (1)[i] Machine learning classifiers (1)[i] Semiparametric regression (1)[j] Structural equation modeling (1)[j] Novel algorithm (1)[k]

[a]The reported study design may not be specific to the mixture analyses conducted but refers to the overall study design.
[b]The number of studies shown may exceed the totals in each section where studies had multiple goals or used multiple methods.
[c]Although nonmonotonic exposure-response relationships are of interest in EDC studies (Vandenberg et al. 2012), studies using statistical methods for nonlinearity that are able to address both nonlinear monotonic and nonmonotonic relationships were included in the tally for this goal.
[d]Factor analysis used by Ochiai et al. (2014); analysis of variance used by Tran et al. (2016).
[e]Studies that focused on the health effects of only one exposure while controlling for copollutant confounding were included in the tally for this goal if results were reported for multiple exposures.
[f]Gray et al. (2005); Roen et al. (2015); Vejrup et al. (2016).
[g]Claus Henn et al. (2016); Erkin-Cakmak et al. (2015); Kobrosly et al. (2014).
[h]Elastic net regression used by Forns et al. (2016) and Lenters et al. (2015a); Bayesian model averaging used by Forns et al. (2016).
[i]Berg et al. (2016) used partial least squares regression and hierarchical clustering; Krysiak-Baltyn et al. (2012) used partial least squares, support vector machine, and neural network classifiers.
[j]Semiparametric regression used by Claus Henn et al. (2016); structural equation model used by Heilmann et al. (2006).
[k]Novel algorithm for identifying important mixture components, based on ranking *p*-values and averaging Z-scores, used by Govarts et al. (2016).

in Table 1, and a detailed summary of each study is available in Supplemental Material, Table S3. These tables are divided into three sections according to dimensionality: studies of one aggregate-, two-, or three or more exposure variables are listed in sections A, B, and C, respectively.

Study Designs

The majority of studies were prospective birth cohort studies (58 studies, 78%), including two studies of pooled cohorts (Buckley et al. 2016a; Casas et al. 2015), and case–control studies (13 studies, 18%), including one case–cohort study (Chevrier et al. 2011) and one analysis of pooled data from two case–control studies (Rull et al. 2006) (Tables 1 and S3). There were also two retrospective cohort studies (Limousi et al. 2014; Migeot et al. 2013) and one cross-sectional study (Huel et al. 2004).

Outcomes

Perinatal outcomes included fetal growth (e.g., birth weight, length, and head circumference) and gestational age (22 studies, 30%), and birth defects, such as neural tube defects, cryptorchidism, and hypospadias (7 studies, 9%) (Table S3). Other perinatal outcomes included anogenital distance (Barrett et al. 2016), neonatal neurobehavior (Suzuki et al. 2010), enzyme activity (Huel et al. 2004), neonatal thyroid stimulating hormone concentrations

(Berg et al. 2016), and placental miRNA expression (LaRocca et al. 2016).

Neurodevelopmental outcomes were the most common pediatric outcomes (22 studies, 30%), including measures of intelligence, cognitive and psychomotor development, behavior, autism, and attention deficit hyperactivity disorder (Table S3). Eight studies (11%) assessed pediatric measures of body mass (e.g., body mass index, waist circumference) and cardiometabolic markers (e.g., leptin; Tang-Péronard et al. 2015; Vafeiadi et al. 2015). Four studies (5%) examined respiratory and atopic outcomes in infancy and childhood, including asthma and eczema (Gascon et al. 2015a; Miller et al. 2004; Ochiai et al. 2014; Rosa et al. 2011). Other pediatric outcomes were postvaccination antibody response (Heilmann et al. 2006; Jusko et al. 2010), auditory outcomes (Buscail et al. 2015; Sturza et al. 2016), and leukemia (Symanski et al. 2016). DNA methylation at birth and in childhood was considered in one study (Huen et al. 2014).

Most studies considered outcomes that were continuous (44 studies), dichotomous (19), or both (9) (Table S3). Few studies considered polytomous (Rull et al. 2006; Stroustrup et al. 2016) and count (Gascon et al. 2015b; Miller et al. 2004; Roen et al. 2015) outcomes. Nine studies analyzed an outcome measured at multiple time points in a longitudinal analysis (Braun et al. 2014; Buckley et al. 2016a, 2016b; Casas et al. 2016; Erkin-Cakmak et al. 2015; Gascon et al. 2015a; Huen et al. 2014; Kippler et al. 2012; Maresca et al. 2016) (Table S3).

Individual and Aggregate Exposures

Exposures were primarily measured using biomarkers (63 studies, 85%) and were analyzed individually or in aggregate to reduce dimensionality (Table S3). Exposures were primarily aggregated by raw or weighted sums, such as the toxic equivalence quotient (TEQ) or molar sum. Exposures were only summed if they shared similar physicochemical properties. Common aggregate variables were sums of polychlorinated biphenyl (PCB) congeners (in total or separated into dioxin-like and nondioxin-like congeners) (e.g., Jusko et al. 2010; Stewart et al. 2008; Vafeiadi et al. 2014), sums of phthalate metabolites (grouped into di(2-ethylhexyl) phthalate (DEHP) metabolites, low- and high-molecular-weight phthalate metabolites) (e.g., Chevrier et al. 2012; LaRocca et al. 2016; Wolff et al. 2008), TEQ of polychlorinated dibenzodioxins and dibenzofurans (PCDD/Fs) (e.g., Konishi et al. 2009; Koskenniemi et al. 2015; Neugebauer et al. 2015), and sums of polybrominated diphenyl ether (PBDE) congeners (e.g., Harley et al. 2011; Huen et al. 2014; Lignell et al. 2013). The health effects of sums of parabens (Chevrier et al. 2012; LaRocca et al. 2016; Philippat et al. 2012), polycyclic aromatic hydrocarbons (PAHs) (e.g., Miller et al. 2004; Polanska et al. 2014; Rosa et al. 2011), and dialkylphosphate (DAP) metabolites (e.g., Naksen et al. 2015; Yolton et al. 2013) were also analyzed. Exposures were also aggregated as counts of individual chemicals (e.g., above detection limits), particularly in studies of pesticides (e.g., Sturza et al. 2016; Wickerham et al. 2012).

Estimation of EDC-Mixture Health Effects

Of the 54 studies (73%) estimating mixture health effects, 51 studies estimated the effects of one or multiple aggregate exposure variables concurrently (24 studies in Section A, 7 studies in Section B, and 20 studies in Section C, Table S3). For example, summed concentrations of PCBs and summed PBDEs in breast milk were analyzed in a model of birth weight (Lignell et al. 2013). The majority of these studies used generalized linear models (GLMs) and their extensions, including conditional logistic regression for a nested case–control study (Chevrier et al. 2012). Studies performing longitudinal analyses used population-averaged models estimated by generalized estimating equations (e.g., Gascon et al. 2015a) and linear or logistic mixed effects regression models (e.g., Buckley et al. 2016a; Huen et al. 2014).

Studies that examined a large number of exposures used dimension reduction methods to assess mixture health effects (Section C, Tables 1 and S3). The most common were principal components analysis (PCA) (Agay-Shay et al. 2015; Forns et al. 2016; Govarts et al. 2016; Maresca et al. 2016; Talbott et al. 2015) and partial least squares regression (PLSR) (Berg et al. 2016; Krysiak-Baltyn et al. 2012). Methods used to partition exposures into groups included hierarchical clustering (Berg et al. 2016) and factor analysis (Ochiai et al. 2014). The health effects of predefined exposure groups were estimated using a structural equation model (SEM) (Heilmann et al. 2006).

Identification of Important Mixture Components

Almost half of the studies (36 studies, 49%) aimed to identify important individual exposures or subsets of exposures within a mixture (Tables 1 and S3). Within this group, GLMs were used by all ten of the studies that considered only two exposure variables at a time (Section B), and by 16 studies that analyzed more than two exposure variables at a time (Section C).

Studies analyzing problems of greater dimensionality used variable selection strategies (Section C, Tables 1 and S3). Elastic net regression (ENR), a frequentist shrinkage method, was used by Forns et al. (2016) and Lenters et al. (2015a). Both studies capitalized on the variable selection feature of ENR and then used the corresponding unpenalized regression methods to obtain unbiased effect estimates for the selected variables. Forns et al. (2016) compared their ENR estimates with those estimated by Bayesian model averaging (BMA), a method that accounts for uncertainty in model selection. In another approach, Govarts et al. (2016) presented a novel algorithm for identifying important subsets of exposures, based on ranking p-values and averaging z-scores from single pollutant models for all possible combinations of 16 exposures.

Shrinkage estimators in a semi-Bayesian hierarchical regression (semi-BHR) setting were used by Braun et al. (2014) and Rull et al. (2006), in which coefficients were shrunk toward the mean of their exchangeability group, representing chemicals with similar physicochemical properties (e.g., phthalate metabolites grouped by parent compounds). Bayesian hierarchical regression (BHR) models were also estimated by Buckley et al. (2016a, 2016b), relating multimetabolite phthalate exposures to measures of body mass. BHR with stochastic search variable selection (BHR-SSVS) was used to identify important air toxicants associated with spina bifida (Swartz et al. 2015).

Nonmonotonicity in Exposure–Response Relationships

Twenty-nine studies reported assessing whether exposure–response relationships were nonlinear (39%; Tables 1 and S3). Of these, 19 studies used splines and generalized additive models (GAMs) to assess nonlinearity. Other studies analyzed exposure quantiles or categories instead of continuous variables (15 studies; e.g., Buckley et al. 2016b; Erkin-Cakmak et al. 2015; Wolff et al. 2008) or used locally weighted scatterplot smoothers (LOWESS) (Harley et al. 2011; Kippler et al. 2012). A further nine studies used categorical exposure variables in their analyses without reporting that they were assessing nonlinearity. Some studies additionally performed hypothesis tests of trends across exposure categories (seven studies; e.g., Casas et al. 2015; Tang-Péronard et al. 2015).

Fourteen studies used a method that may address nonlinearity in a mixture (Table S3). Of these, 11 studies analyzed exposure quantiles or categories, rather than continuous variables, in a multipollutant model (e.g., Agay-Shay et al. 2015; Claus Henn et al. 2016; Valvi et al. 2012). Four studies modeled nonlinear additive relationships among exposures in a mixture using polynomials (Rodrigues et al. 2016) or splines (Buckley et al. 2016a; Claus Henn et al. 2016; Kippler et al. 2012). For example, Kippler et al. (2012) used a piecewise linear spline with one knot to model an inverse-V-shaped relationship between cadmium exposure and fetal size, controlling for arsenic exposure.

Interactions between Chemical Exposures

Interactions between chemical exposures were assessed in 13 studies (18%; Tables 1 and S3). These were predominantly studies of two or three exposure variables only, using GLMs with product terms between pairs of exposures to assess two-way interactions. Three studies considered interactions between pairs of aggregate exposure variables or between an aggregate exposure and individual chemicals; e.g., Huen et al. (2014) examined interactions between summed PBDE congeners and dichlorodiphenyl-trichloroethylene/dichloroethane (DDT/E) compounds in relation to DNA methylation.

Windows of Exposure Susceptibility

Eighteen studies (24%) analyzed exposure at multiple times to investigate etiological windows of exposure susceptibility (Tables 1 and S3). Most studies analyzed exposure windows in separate models (10 studies), predominantly using GLMs. For example, Jusko et al. (2010) estimated the health effects of PCBs measured in maternal serum at delivery, cord serum, and 6-month infant serum, on postvaccination antibody response, in separate linear regression models.

Eight studies analyzed exposure measured at multiple prenatal/postnatal times in the same model. Six of these studies used GLMs. For example, Roen et al. (2015) estimated the effects of prenatal and postnatal BPA exposure on child behavior, controlling for prenatal mono-*n*-butyl phthalate metabolite (MnBP) exposure. To address the possibility of multicollinearity when analyzing multiple exposures at multiple time points, Maresca et al. (2016) used PCA to obtain prenatal and postnatal phthalate concentration component scores that were then regressed on the outcome. In another approach, Heilmann et al. (2006) used a SEM with latent variables for both prenatal and postnatal exposure to assess cumulative health effects of PCBs on postvaccination antibody response, allowing prenatal exposure to be a determinant of postnatal exposure.

Discussion

We reviewed studies of prenatal exposure to EDC mixtures to determine the analytical strategies and statistical methods that have been employed and concurrently identified several novel and applicable statistical methods. We now compare the abilities of each method for addressing the salient epidemiological and statistical challenges. A detailed summary of each method is presented in Table 2, along with software suggestions.

Identification of Important Mixture Components

Correlation between chemicals with common sources, exposure pathways, or metabolic processes, can induce multicollinearity when analyzed simultaneously (Carrico et al. 2015; Vrijheid et al. 2016). This is arguably the paramount statistical challenge facing the study of chemical mixtures. In GLMs, multicollinearity can cause standard errors to become inflated, leading to unstable coefficient estimates that are sensitive to minor changes in model specification. Consequently, it may be difficult to assess the relative importance of individual exposures, and therefore to separate potential etiological agents from copollutant confounders and redundant variables (Woodruff et al. 2009). The problem is potentiated with the addition of interaction and nonlinear terms in the model (Schisterman et al. 2017), and at an extreme, maximum likelihood techniques may fail to converge to a solution (MacLehose et al. 2007).

Frequentist shrinkage methods, such as ridge regression, least absolute shrinkage and selection operator (LASSO) regression, and ENR, present an appealing solution (Table 2) (only ENR was used in the reviewed studies, by Forns et al. 2016; Lenters et al. 2015a). By imposing a penalty on the size of the coefficients, they ensure convergence and confer a degree of immunity to multicollinearity (Hastie et al. 2009). These methods shrink the coefficients of explanatory variables toward zero, with the degree of shrinkage determined by the strength of the association. LASSO and ENR also perform variable selection, whereby the coefficients of the least predictive variables are shrunk to exactly zero (Hastie et al. 2009), yielding a sparse model with greater interpretability and more-precise coefficient estimates. However, these estimators are biased and pose challenges for statistical inference because valid standard error estimates (and therefore confidence intervals) may be difficult to obtain and may depend on the choice of tuning parameters (i.e., parameters that are adjusted to find an optimal model) (Kyung et al. 2010; Pfeiffer et al. 2017). Confidence intervals may also lack proper coverage if the uncertainty introduced by variable selection is ignored in post-model-selection inference (Pfeiffer et al. 2017). Moreover, data-driven tuning parameter selection via *k*-fold cross-validation may lead to highly variable model selection results (Roberts and Nowak 2014); this instability may be especially apparent in small sample-size studies with weak signals and correlated exposures (Kirk et al. 2013; Lim and Yu 2016). Methods such as the percentile-LASSO (Roberts and Nowak 2014), estimation stability cross-validation (Lim and Yu 2016), and others (e.g., Frommlet et al. 2012; Kirk et al. 2013; Waldron et al. 2011), may address model selection instability.

ENR may be preferred to LASSO in mixtures analyses. Whereas LASSO may choose only one exposure among a set of highly correlated exposures and drop the others (Hastie et al. 2009), ENR possesses a grouping property, ensuring that highly correlated variables are retained or dropped together by being assigned coefficients of similar magnitude (Zou and Hastie 2005). However, not all correlated exposures selected by this property may be associated with the outcome, and those exposures dropped by either method may be causal agents, so it cannot be assumed that nonselected exposures are safe. In a recent simulation study, ENR was found to have slightly higher sensitivity but slightly higher false discovery proportion than LASSO, with comparable accuracy (mean-squared error) (Lenters et al. 2017).

Weighted quantile sum regression (WQSR) (not used in the reviewed studies) was developed as an alternative to LASSO and ENR for cross-sectional analyses and was reported to have greater accuracy (lower mean-squared error), comparable sensitivity, and improved specificity for variable selection among correlated exposures in simulation studies (Carrico et al. 2015) (Table 2). In WQSR, exposures are combined into an empirically weighted index, which provides both estimates of mixture health effects and indicators of exposure importance (i.e., weights); however, because the weights are nonzero, a threshold choice is required for variable selection (Carrico et al. 2015).

Table 2. Summary of statistical methods.

Method	Description	Purpose/properties	Goals	Outcome types	Strengths	Caveats or weaknesses	Studies included in review	Further information	Software (open source listed where possible)
Generalized linear models (GLMs)	Generalized linear models with one or a few single or aggregate exposures.	Regression	Identification of important mixture components; interactions; nonlinearity (e.g., using nonlinear additive terms such as polynomials or splines)	Any GLM link function and family; longitudinal/multilevel data	- Well known properties, accessible software. - Straightforward interpretation of coefficients. - Straightforward adjustment for confounding. - Recognized methods for adjusting for multiple testing.	- Multicollinearity issues if exposures are highly correlated. - High false positive rate for single exposure models with correlated exposures. - Not suitable for higher-dimensional data as does not perform variable selection.	See Tables 1 and S3	Dobson and Barnett (2008)	glm function in R (R Core Team 2017); lme4 in R (Bates et al. 2015)
Ridge regression, least absolute shrinkage and selection operator (LASSO), elastic net regression (ENR)	Shrinkage (penalized regression) methods. LASSO and ENR perform variable selection. ENR also performs grouped selection, where highly correlated exposures tend to be retained or dropped together, by being assigned coefficients of similar magnitude (Zou and Hastie 2005). Grouped LASSO penalty is available for pre-defined groups of exposures (Hastie et al. 2009). Sparse hierarchical interaction models can be built using hiernet in R (Bien et al. 2013) and glinternet in R (Lim and Hastie 2015).	Shrinkage; variable selection	Identification of important mixture components; pairwise interactions	Any	- Robust to multicollinearity. - Lower coefficient variance than ordinary least squares regression. - Greater prediction accuracy has been demonstrated in ENR than LASSO in simulation studies (Zou and Hastie 2005). - Ability to control for confounding.	- Biased toward zero; however, for LASSO and ENR, a subsequent unpenalized regression can be used to obtain unbiased effect estimates for the selected variables (Hastie et al. 2009). - In LASSO: among a set of highly correlated exposures, only one may be chosen and the others dropped, giving the impression that they are not associated with the outcome (Hastie et al. 2009; Zou and Hastie 2005). - Tuning parameter selection via k-fold cross-validation may lead to highly variable model selection; however, methods exist that address this instability (e.g., Lim and Yu 2016; Roberts and Nowak 2014). - Post-selection statistical inference tools are required for inference, including generating confidence intervals (e.g., Berk et al. 2013; Lockhart et al. 2014; Meinshausen and Bühlmann 2010; Tibshirani et al. 2016).	Lenters et al. (2015a); Forns et al. (2016)	Sun et al. (2013); Zou and Hastie (2005); Hastie et al. (2009); Bien et al. (2013); Lim and Hastie (2015)	glmnet in R (Friedman et al. 2010); hierNet in R (Bien et al. 2013); glinternet in R (Lim and Hastie 2015); selectiveInference in R (Tibshirani et al. 2016)
Weighted quantile sum regression (WQSR), lagged WQSR	Empirically weighted sums of exposure quantiles are used in a regression model, with weights and parameters estimated by nonlinear programming and bootstrap (Carrico et al. 2015). Inference is focused in a specific	Variable selection	Identification of important mixture components; estimation of mixture health effects; identification of windows of exposure susceptibility (lagged WQSR).	Any GLM link function and survival data	- Robust to multicollinearity (Gennings et al. 2013). - The use of quantiles reduces the impact of outlying observations (Gennings et al. 2013). - May have greater accuracy (lower mean-squared error), comparable sensitivity, and higher specificity than shrinkage methods	- Loss of information through the use of quantiles (Carrico et al. 2015). - Weights are non-zero so a threshold choice is required for identifying important mixture components (Carrico et al. 2015). - Interactions are not supported (Carrico et al. 2015).	None	Carrico et al. (2015); Czarnota et al. (2015); Gennings et al. (2013); Bello et al. (2017)	wqs in R (Czarnota and Wheeler 2015); gWQS in R (Renzetti et al. 2018)

Table 2. (*Continued.*)

Method	Description	Purpose/properties	Goals	Outcome types	Strengths	Caveats or weaknesses	Studies included in review	Further information	Software (open source listed where possible)
	direction, i.e., in the direction of increased risk or decreased risk (Carrico et al. 2015). Lagged WQSR can be used for mixtures analyses with numerous repeated exposure measures (Bello et al. 2017).				such as LASSO and ENR (Carrico et al. 2015; Czarnota et al. 2015). - Ability to control for confounding.				
Semi-Bayes hierarchical regression (semi-BHR)	Two-level semi-Bayes parametric model that can incorporate prior information on exposure categories and the variance of the effect estimates.	Shrinkage	Identification of important mixture components	Any	- Able to control for traditional and co-pollutant confounders (Braun et al. 2014). - Able to incorporate prior knowledge on exposure groupings, e.g. based on similarity of chemical properties or expected mechanisms. Exchangeability groups do not need to be mutually exclusive and can be implied from a set of covariates rather than directly specified (MacLehose et al. 2007; Thomas et al. 2007). - Coefficients are shrunk toward the mean of their exchangeability group to counteract multicollinearity (Greenland 1993; MacLehose et al. 2007). - Mitigation of multiple testing issues (Momoli et al. 2010).	- Quality of prior knowledge determines the validity and efficiency of the analysis (Momoli et al. 2010). - Coefficient variability needs to be set (fixed prior), so information about coefficient variability from the observed data is ignored (MacLehose et al. 2007).	Braun et al. (2014); Rull et al. (2006); Buckley et al. (2016a, 2016b)	Greenland (1993, 1994); MacLehose (2007); Momoli et al. (2010)	See MacLehose et al. (2007) for WinBUGS code.
Bayesian hierarchical regression (BHR), and BHR with stochastic search variable selection (BHR-SSVS)	A flexible class of models, which can be parametric (e.g., the Bayesian empirical-Bayes model, Greenland 1994) or semiparametric (e.g., using a Dirichlet process prior for the exposure coefficients, MacLehose et al. 2007). Variable	Shrinkage; variable selection; clustering; smoothing	Identification of important mixture components; nonlinearity; interaction	Any	- Able to control for traditional and co-pollutant confounders (Braun et al. 2014). - Able to incorporate prior information. - Mitigation of multiple testing issues (Momoli et al. 2010). - Shrinkage estimator, so robust to multicollinearity: effect estimates have	- Specification may require subject-matter knowledge that may not be available (Greenland 1994; Thomas et al. 2007). - May be computationally demanding.	Swartz et al. (2015)	Greenland (1993, 1994); MacLehose et al. (2007); Herring (2010)	See MacLehose et al. (2007) for WinBUGS code.

Table 2. (Continued.)

Method	Description	Purpose/properties	Goals	Outcome types	Strengths	Caveats or weaknesses	Studies included in review	Further information	Software (open source listed where possible)
	selection priors are also possible (e.g., a Bayesian hierarchical mixture model with stochastic search variable selection, Swartz et al. 2015).				greater accuracy than maximum likelihood estimates (MacLehose et al. 2007). - Adaptive rather than fixed shrinkage, allowing the data to inform the amount of shrinkage based on support for the prior distribution (MacLehose et al. 2007). - Inference less sensitive to specification of prior parameters than semi-Bayes model (MacLehose et al. 2007).				
Bayesian kernel machine regression (BKMR) with component-wise or hierarchical variable selection (BKMR-HVS)	Nonparametric method which estimates joint effects of multiple exposures, allows for nonlinearity and interactions, and performs component-wise or hierarchical (grouped) variable selection (Bobb et al. 2015; Coull et al. 2015).	Smoothing; variable selection; shrinkage	Identification of important mixture components; estimation of mixture health effects; nonlinearity; pairwise and higher-order interactions	Continuous; extension to binary and count outcomes planned for development (Claus Henn in NIEHS 2015; Coull et al. 2015); longitudinal outcomes	- Simultaneously estimates mixture health effects while identifying drivers of the association (Coull et al. 2015). - Reduces the dimensionality of the problem (Bobb et al. 2015). - Robust to multicollinearity (Bobb et al. 2015). - Increased power to detect associations due to grouping of highly correlated exposures, using BKMR-HVS (Bobb et al. 2015). - Estimation of posterior inclusion probabilities. - Able to model nonlinearity in a high-dimensional exposure-response surface (Bobb et al. 2015).	- Computation for very large datasets currently not feasible (see Coull et al. 2015). - Assumes exposures within a group do not interact (Bobb et al. 2015). - Currently does not allow overlapping groups, however this extension is planned for development (Bobb et al. 2015). - Only one exposure can be selected within a group in BKMR-HVS (Coull et al. 2015). - Magnitude, but not relative ranking, of the posterior inclusion probabilities is sensitive to the choice of tuning parameters (Coull et al. 2015).	None	Bobb et al. (2015); Coull et al. (2015)	*bkmr* in R (Bobb et al. 2015)
Structural equation modeling (SEM)	A path modeling approach where observed exposure variables are seen as manifestations of latent true exposure variables, which are then assumed to	Dimension reduction	Identification of important mixture components; estimation of mixture health effects	Continuous	- Can model multiple related outcomes and exposures simultaneously, for example allowing prenatal and postnatal exposures to be analyzed in the same model (Heilmann et al. 2006).	- Requires greater subject matter knowledge for correct specification than single outcome-exposure models (Sánchez et al. 2005). - Imposes multiple strong assumptions on the analysis (regarding the existence and	Heilmann et al. (2006)	Sánchez et al. (2005); VanderWeele (2012)	*sem* in R (Fox et al. 2017); *blavaan* in R (Merkle and Rosseel 2016)

Table 2. (Continued.)

Method	Description	Purpose/ properties	Goals	Outcome types	Strengths	Caveats or weaknesses	Studies included in review	Further information	Software (open source listed where possible)
	affect one or more outcome variables (which can themselves also be seen as manifestations of multiple observed outcome variables) (Sánchez et al. 2005). This is a frequentist method that can also be estimated in a Bayesian setting for more flexibility.				- May correct for measurement error when multiple measurements for each exposure are available, under certain assumptions (Budtz-Jørgensen et al. 2002; Sánchez et al. 2005). - Greater model parsimony than modeling each outcome-exposure pair separately (Sánchez et al. 2005). - Multiple testing issues are reduced through the use of latent variables that represent multiple observed exposure variables (Sánchez et al. 2005). - Variables that are related to exposure but not outcome can be included to increase power (Sánchez et al. 2005). - Budtz-Jørgensen et al. (2002) describe an approach for handling missing data in SEMs with continuous outcome variables. - Ability to control for confounding.	direction of causal relationships, and distributional and functional forms) that may be difficult to verify in practice and may lead to severe biases (VanderWeele 2012).			
Principal components analysis (PCA) and regression (PCR)	Unsupervised PCA reduces exposure data to several uncorrelated components (assuming orthogonal rotation), which are then regressed on the outcome. Supervised PCA uses both exposure and outcome data to obtain components, by first selecting variables that are most correlated with the outcome (Bair et al. 2006).	Dimension reduction; variable selection	Estimation of mixture health effects; identification of important mixture components	Any	- Eliminates multicollinearity by reducing data to several uncorrelated components (assuming orthogonal rotation). - Supervised PCA also performs variable selection and provides variable importance scores with fixed ordering (Bair et al. 2006). - Ability to control for confounding (in subsequent regression model using components from PCA).	- Components may not have biologically-relevant interpretations. - Can only be used with continuous exposures (correspondence analysis can be used for categorical exposures).	Forns et al. (2016); Agay-Shay et al. (2015); Govarts et al. (2016); Berg et al. (2016); Maresca et al. (2016); Talbott et al. (2015)	Hastie et al. (2009); Roberts and Martin (2006); Bair et al. (2006)	*prcomp* function in R (R Core Team 2017); *nsprcomp* in R (Sigg and Buhmann 2008)

Table 2. (*Continued.*)

Method	Description	Purpose/properties	Goals	Outcome types	Strengths	Caveats or weaknesses	Studies included in review	Further information	Software (open source listed where possible)
Partial least squares regression (PLSR)	A supervised dimension reduction method that constructs latent variables by forming linear combinations of exposures that maximize the covariance between the exposures and the outcome. Sparse PLSR additionally performs variable selection (Agier et al. 2016; Chun and Keleş 2010).	Dimension reduction; variable selection	Estimation of mixture health effects; identification of important mixture components	Continuous, categorical (using PLSR-DA, i.e., Discriminant Analysis)	- Addresses multicollinearity by constructing a set of orthogonal latent variables.	- Adversely affected by the inclusion of irrelevant variables (Bair et al. 2006). - In sparse PLS, one exposure may be chosen somewhat arbitrarily among a set of highly correlated exposure variables (Hastie et al 2009; Lenters et al. 2015b). - Adjusting for confounders is not straightforward in PLSR (Lenters et al. 2015b), however both the exposures and outcome may be preadjusted for confounders and their residuals used in the analysis.	Berg et al. (2016)	Lenters et al. (2015b); Agier et al. (2016); Chun and Keleş (2010); Hastie et al. (2009)	*pls* in R (Mevik et al. 2016); *spls* in R (Chung et al. 2018)
Generalized additive models (GAMs) and semiparametric models	Commonly used data-driven method for modeling nonlinearity in which the outcome is regressed on a sum of nonparametric smoothing functions (Billionnet et al. 2012).	Smoothing	Nonlinearity	Any GLM link function and family; survival data	- Flexibility in functional form, does not impose assumptions on the shape of the exposure-response curve (Billionnet et al. 2012; Eisen et al. 2004). - Provides a visual representation of exposure-response relationships. - Pairwise and higher-order interactions can be modeled using multivariate smoothers (Ruppert et al. 2003).	- May be affected by concurvity, the nonlinear analog to collinearity; however, this may be addressed by partial GAMs (Gu et al. 2010).	19 studies used splines and GAMs (e.g., Vejrup et al. 2016; Claus Henn et al. 2016) (Table S3).	Eisen et al. (2004); Ruppert et al. (2003, 2009); Gu et al. (2010)	*SemiPar* in R (Ruppert et al. 2003, 2009); *mgcv* in R (Wood 2017)
Multivariate adaptive regression splines (MARS)	A nonparametric method suitable for higher-dimensional problems that automatically selects variables, models nonlinearity, and identifies interactions (Hastie et al. 2009; Morlini 2006).	Smoothing; variable selection	Interactions; nonlinearity; identification of important mixture components	Continuous; categorical	- Flexibility in functional form.	- May select an exposure somewhat arbitrarily in the presence of high correlation between exposures (Morlini 2006). - Meaningful standard error estimates are not available; bootstrapping techniques are required for statistical inference (Friedman and Roosen 1995). - Confounder terms are also subject to selection.	None	Friedman and Roosen (1995); Hastie et al. (2009); Waters in NIEHS (2015)	*earth* in R (Milborrow 2017)
Classification and regression trees (CART) and ensemble methods: stochastic gradient boosting (SGB),	Statistical learning methods that partition data into a tree. Numerous regression trees can be combined to improve predictive performance.	Classification; variable selection	Identification of important mixture components; pairwise and higher-order interactions; nonlinearity; identification of windows of	Continuous; binary	- Does not assume additivity and linearity in exposure-response relationships (Coull et al. 2015). - Able to automatically deal with missing values through surrogate splits (Hastie et al. 2009).	- CART requires many splits to create a linear or nonlinear association and hence is better suited to categorical exposures (Hastie et al. 2009). - Variable importance scores (RF and SGB) reflect predictive performance, not direction or	None	Lampa et al. (2014); Hastie et al. (2009); Gass et al. (2014); BART: Dunson in NIEHS	*gbm* in R (Ridgeway 2017); *mboost* in R (Bühlmann and Hothorn 2007); *bartMachine*

Table 2. (Continued.)

Method	Description	Purpose/properties	Goals	Outcome types	Strengths	Caveats or weaknesses	Studies included in review	Further information	Software (open source listed where possible)
random forests (RF), and Bayesian additive regression trees (BART); tree-based distributed lag models	using ensemble methods.		exposure susceptibility (tree-based distributed lag models)		- Immune to outliers (Hastie et al. 2009). - BART provides exposure effect estimates with uncertainty intervals (Chipman et al. 2010; Hernández et al. 2015).	magnitude of exposure-response associations (Bobb et al. 2015; Coull et al. 2015); i.e., effect estimates are not available. - Search for interactions is automatic, so spurious interactions may be identified (Lampa et al. 2014). - Confidence intervals are not immediately provided for SGB and RF, however can be obtained by bootstrap (Lampa et al. 2014). - No straightforward way to control for confounding (Coull et al. 2015), however see Gass et al. (2014) for a CART algorithm that controls for confounding. - Choice of threshold strategy for variable selection in BART requires cross-validation or prior knowledge of the sparsity of the solution (Bleich et al. 2014).		(2015); Chipman et al. (2010); Hernández et al. (2015); Bleich et al. (2014); tree-based distributed lag models: Bello et al. (2017)	in R (Kapelner and Bleich 2016); see Gass et al. (2014) for a CART algorithm that controls for confounding.
Bayesian model averaging (BMA)	A Bayesian technique that provides weighted average estimates over a set of all possible models rather than requiring selection of a single best model (Sun et al. 2013).	Variable selection; model averaging	Identification of important mixture components	Any GLM link function and family; survival data	- Accounts for uncertainty in model selection. - Supports the inclusion of interactions and nonlinearity.	- Not robust to multicollinearity (Sun et al. 2013). - Computational inefficiency involved in considering numerous models with large numbers of exposure variables (Sun et al. 2013).	Forns et al. (2016)	Sun et al. (2013); Wang et al. (2004)	BMA in R (Raftery et al. 2018)

Performing Bayesian shrinkage may offer a number of advantages beyond robustness to multicollinearity. Bayesian methods offer flexibility of specification, can incorporate external information, can handle missing data, are robust to overfitting in small sample sizes, and offer probability-based uncertainty intervals (Hernández et al. 2015). BHR methods treat exposure coefficients as random variables from a prior distribution (Table 2). They deal with multicollinearity by shrinking coefficients toward the prior mean, with the amount of shrinkage depending on the magnitude of the prior variance (MacLehose et al. 2007). The prior variance can be treated as known (semi-Bayes model; used by Braun et al. 2014; Buckley et al. 2016a, 2016b; Rull et al. 2006), requiring the researcher to set the parameter to a plausible value. However, unlike tuning parameters in frequentist shrinkage methods that are set by cross-validation, this tuning parameter has an intuitive interpretation as the variance of the exposure coefficients (Greenland 1994). If there is uncertainty about the magnitude of the parameter, it can be allowed to vary, with the advantage of making estimates more robust to misspecification of the prior (Greenland 1994; MacLehose et al. 2007).

Coefficients can also be shrunk toward group-specific means using BHR; e.g., groups based on physicochemical similarity, as in the reviewed studies by Braun et al. (2014) and Rull et al. (2006). These groups may be overlapping and can be implied from a set of covariates rather than directly specified (MacLehose et al. 2007; Thomas et al. 2007). Alternatively, if groups cannot be chosen *a priori*, a Dirichlet process prior can be used to perform data-driven clustering (MacLehose et al. 2007).

Methods that perform variable selection in addition to shrinkage can realize a further gain in precision by allowing exposures that do not affect the outcome to be excluded, a feature that is especially important in small samples (MacLehose et al. 2007). Unlike frequentist shrinkage methods that perform variable selection (i.e., LASSO and ENR), mixture priors in the "spike and slab" form (such as BHR-SSVS, used by Swartz et al. 2015) can also provide a measure of uncertainty in variable selection (Herring 2010). These methods also have the ability to account for truncation of exposures due to detection limits (Herring 2010).

Like frequentist shrinkage estimators, BHR estimates are biased toward their prior mean. However, this bias may be more than compensated for by the reduction in variance (Greenland 1994). Because maximum likelihood estimates are only asymptotically unbiased, hierarchical estimates are likely to have a small-sample advantage in terms of accuracy when modeling multiple correlated exposures (Greenland 1994). Moreover, simulation studies of chemical mixtures have observed accuracy and precision advantages of BHR methods over their frequentist method counterparts (e.g., Roli and Monari 2015).

Despite the availability of methods that are robust to multicollinearity, no statistical method is able to differentiate between very highly correlated exposures in the absence of further information. Preprocessing steps may be required, such as choosing an individual chemical to represent a group of very highly correlated chemicals or dimension reduction using PCA (Coull et al. 2015). For example, Braun et al. (2014) analyzed only the chemical with the highest median concentration in pairs of chemicals with correlation coefficients greater than 0.95 in absolute value. However, estimated associations may lack specificity if the chosen chemical is not a suitable surrogate for the others (e.g., a common proxy choice for PCBs is nonplanar PCB-153, but it is not representative of coplanar PCBs) (Engel and Wolff 2013). Moreover, inclusion of exposures unrelated to the outcome but correlated with exposures that are associated with the outcome may enhance, diminish, or reverse associations between the outcome and the important exposures due to the reversal paradox (Tu et al. 2008).

One method that allows prior information on the exposure correlation structure to be incorporated is Bayesian kernel machine regression (BKMR) (Table 2). BKMR is a novel approach for mixtures analyses (not used in the reviewed studies), offering both component-wise and hierarchical (grouped) variable selection (BKMR-HVS) (Bobb et al. 2015; Coull et al. 2015). BKMR-HVS concurrently estimates the importance (i.e., posterior inclusion probability) of groups of highly correlated exposures as well as individual chemicals within the group, enabling selection of an exposure in a group of highly correlated exposures (Bobb et al. 2015; Coull et al. 2015). Because BKMR estimates the joint exposure–response function, it can also provide an estimate of an overall mixture effect, representing the risk associated with all exposures at a particular quantile vs. the median (Bobb et al. 2015).

Estimation of EDC-Mixture Effects

A robust body of evidence demonstrates mixture effects due to low-dose EDCs acting together via common mechanisms (Kortenkamp et al. 2007; Ribeiro et al. 2017). EDCs that act in combination may have only weak signals when analyzed individually, and these signals may not be detected in studies lacking sufficient power. Here we examine methods that may be used to estimate the health effects of EDC mixtures directly when the estimation of individual effects is not possible.

A common approach for estimating EDC-mixture effects involves the use of aggregating variables, such as raw or weighted sums, or counts of exposures above detection limits (e.g., Koskenniemi et al. 2015; LaRocca et al. 2016; Table S3). These approaches are appealing because of their ease of implementation; however, they discard information and make strong assumptions that effects are additive (implying independence and no interaction) and that exposure–response relationships are linear. Raw sums and counts assume component chemicals are equipotent; these aggregate variables will be driven by the highest concentration pollutants, which may not necessarily be the most detrimental to health (Axelrad et al. 2009; Braun et al. 2016). Even when component chemicals have been transformed to a common scale, heterogeneous concentration ranges and detection limits among mixture components imply that these chemicals are measured with differing precision, which may induce noise in the aggregate variable (Engel and Wolff 2013). Relative potencies of individual chemicals within a mixture may be addressed by a weighted sum; however, the quality of the resulting aggregate variable depends on the availability of accurate outcome-specific potency measures as those derived in different biological contexts may be inappropriate (Gennings et al. 2010; Zoeller et al. 2014). Findings based on aggregate variables will be sensitive to these assumptions, as well as the particular choice of mixture components and approach chosen for aggregating exposures.

The analysis of multiple aggregate exposure variables in one model makes further assumptions regarding independence and additivity of groups of exposures. Exposure groups based on physicochemical similarity, a common grouping method, may not share similar biological mechanisms. In an extreme case, Suzuki et al. (2010) and Tatsuta et al. (2014) summed 209 PCB congeners that were above detection limits. However, individual congeners may have varying effects and opposing actions (e.g., some are estrogenic, and others are antiestrogenic; Cooke et al. 2001), and these classifications may be specific to the tissues and diseases under investigation (Gennings et al. 2010). To separate EDCs into distinct groups based on modes of action may not be possible as the mode may be dose-dependent for individual EDCs; for example, low doses of some xenoestrogens may act by binding to estrogen receptors, whereas, at high doses, these same chemicals may bind to the receptors of other hormones (e.g., androgen or thyroid hormone receptors), through receptor cross-talk (Myers et al. 2009).

Here, methods that allow estimation of health effects for overlapping groups of exposures, as well as careful consideration of group composition, may be more appropriate. For example, Varshavsky et al. (2016) recently presented a novel potency-weighted metric for cumulative exposure to antiandrogenic phthalates based on common adverse outcomes.

For predefined groups of exposures, health effects may be estimated using WQSR and grouped LASSO (Hastie et al. 2009). WQSR addresses some of the limitations of aggregate exposure variables discussed above; WQSR uses empirically determined weights that take into account the exposure correlation structure and, by focusing inference in the direction of increased risk, produces an index that is readily interpretable (Carrico et al. 2015). SEM may also be used to assess health effects of predefined exposure groups (Table 2, used by Heilmann et al. 2006). In this approach, chemicals sharing similar characteristics are assumed to reflect a latent true exposure, which are then used to estimate effects for one or more outcomes (Heilmann et al. 2006). This pooling of information from several exposures results in a more powerful analysis than using linear regression methods (Budtz-Jørgensen et al. 2002; Heilmann et al. 2006); however, the gain in power may come at a considerable cost. SEMs are susceptible to misspecification because they impose multiple strong assumptions (regarding the existence and direction of causal relationships specified in a path diagram, as well as their distributional and functional forms) that may be difficult to verify in practice and may lead to severe biases if incorrect (VanderWeele 2012). SEMs are therefore considered confirmatory methods, suitable in settings where sufficient knowledge exists for the specification of a path diagram and sensible parametric assumptions (Sánchez et al. 2005).

Dimension reduction methods present an alternative to the use of aggregating variables and are particularly useful when the number of exposures is large in comparison with the sample size. PCA reduces multidimensional exposure data to several orthogonal (uncorrelated) components that can be used in a regression model in place of the original exposure variables (Table 2) (e.g., Agay-Shay et al. 2015; Maresca et al. 2016), providing a measure of health effects for components that summarize the variability in the exposure data. Although the component compositions may be statistically informative, they may not necessarily possess biologically relevant interpretations or provide a good explanation of the outcome (Bair et al. 2006). Better explanatory power (but not necessarily interpretability) may be obtained using supervised PCA, which captures information in both the exposures and outcome by preselecting exposures most correlated with the outcome prior to constructing components (Bair et al. 2006). However, because this preselection is based on univariate exposure–outcome relationships, there may be some redundancy in the set of selected exposures because highly correlated exposures are likely to be chosen together (Hastie et al. 2009). This redundancy may be ameliorated through the use of preconditioned LASSO (i.e., applying the LASSO as a postprocessor using a predicted outcome from supervised PCA), a two-stage procedure that produces a sparse model (Hastie et al. 2009; Paul et al. 2007). An alternative supervised dimension reduction method is partial least squares regression (PLSR; Table 2) (e.g., Berg et al. 2016), which constructs latent variables by forming linear combinations of exposures that maximize the covariance between the exposures and outcome (Hastie et al. 2009). However, although both unsupervised and supervised dimension reduction methods address the problem of multicollinearity, they do so at the cost of interpretability. The interpretability of the latent variables can be improved through the use of sparse PLSR, which performs simultaneous dimension reduction and variable selection (Chun and Keleş 2010).

The use of biomarkers of cumulative EDC exposure, such as the total effective xenoestrogen burden (TEXB; Fernández et al. 2004) and total effective xenobiotic burden of antiandrogens (TEXB-AA; Arrebola et al. 2015), present a nonstatistical alternative for estimating mixture effects. For example, Fernández et al. (2007) analyzed both TEXB and 16 organochlorine pesticides in relation to the risk of cryptorchidism and hypospadias in male neonates. TEXB provides dual fractions composed primarily of organohalogenated lipophilic xenoestrogens (alpha fraction) and endogenous hormones with more polar xenoestrogens (beta fraction) (Fernández et al. 2004; Pastor-Barriuso et al. 2016). The beta fraction has been used to control for potential confounding by endogenous hormones (e.g., Vilahur et al. 2014); however, this adjustment is thought to induce a selection bias (Pastor-Barriuso et al. 2016).

Accounting for Nonmonotonicity in EDC Exposure–Response Relationships

Motivation for assessing nonmonotonicity. The breadth of evidence for nonmonotonic EDC dose–response relationships observed across cell culture and animal experimental studies as well as the epidemiological literature has led some authors to recommend that nonmonotonicity should be assumed by default for chemicals that behave like hormones (Myers et al. 2009; Vandenberg et al. 2012). Despite this recommendation, fewer than half of the reviewed studies reported assessing nonlinearity (Tables 1 and S3). The assumption of linearity may be made to simplify statistical analyses, with the view that nonmonotonic relationships are unlikely to be observed over the comparatively narrow exposure ranges common in epidemiological studies (the principle of low-dose linearity; Thomas 2009), or the possibility that sampling biases in observational studies may create the impression of nonmonotonicity when dose–response is truly linear (Engel and Wolff 2013). Data sparsity in regions of nonlinearity may also make these relationships difficult to detect; a large overall sample size and sufficiently large sample size in regions of nonlinearity are required to detect nonlinearity (May and Bigelow 2006). Small sample sizes may yield wide confidence intervals across the exposure–response curve and uncertainty in functional form. Nevertheless, assuming linearity when exposure–response relationships are nonmonotonic can lead to misleading conclusions; for example, a horizontal fitted line when there is a U-shaped relationship (May and Bigelow 2006). Such model misspecification may be more severe in an analysis of multiple correlated exposures, where bias due to omitting important variables (e.g., quadratic terms) may be exacerbated by multicollinearity (Ganzach 1997; Schisterman et al. 2017). Moreover, although a narrow exposure range is common in observational studies of one chemical, the cumulative exposure range of a mixture of similarly acting chemicals is likely to be wider, which increases the chances of observing a nonmonotonic association.

Methods for addressing nonmonotonicity. We identified several statistical approaches for both diagnosing and modeling nonmonotonicity. A common approach in the reviewed studies involved modeling categorical exposure variables in place of their continuous counterparts (e.g., using quantiles; Berg et al. 2016). Although the simplicity of this approach may be appealing, it has several limitations, including loss of information that may preclude the identification of more complex functional forms with few categories, sensitivity to the arbitrary choice of categorization (Greenland 1995), and bias away from the null if observations near category cutpoints are misclassified (Eisen et al. 2004). Moreover, the large number of additional parameters requiring estimation can lead to a loss of power, which may make this approach unfeasible in mixtures studies with many exposures and comparatively small sample size.

An alternate approach that avoids these pitfalls is the inclusion of polynomial exposure terms (e.g., quadratic or cubic) directly in a parametric model (e.g., Rodrigues et al. 2016). These terms enable assessment of U-shaped, inverted U-shaped, or more complex multiphasic exposure–response relationships. They are straightforward to implement and particularly useful when the functional form conforms to a known parametric type. An improvement in fit and more flexibility in functional form may be obtained by using fractional polynomials to automatically identify the optimal power transformations according to goodness-of-fit criteria (May and Bigelow 2006). However, including polynomial terms for multiple correlated exposures may exacerbate multicollinearity issues and reduce the interpretability of parameter estimates. Moreover, these approaches apply a global fit; i.e., they assume the same functional form at all levels of exposure, which may lead to poor fit in some exposure regions (Keele 2007).

In the absence of hypotheses regarding functional form, GAMs may be used to fit flexible nonparametric exposure–response functions based on splines (Table 2) (e.g., Heilmann et al. 2006). For diagnosing nonlinearity, GAMs are powerful tools that impose few assumptions on the data. One of the primary strengths of GAMs is that they are directly comparable with nested parametric GLM fits using likelihood-based statistics (Keele 2007). Semiparametric regression extends GAMs by allowing additional linear additive terms with the usual parametric specifications (e.g., linear confounder terms) (Keele 2007). This method was used by Claus Henn et al. (2016), who analyzed the effect of arsenic on fetal growth while controlling for smoothed lead and manganese terms and linear confounders. Bootstrapping may be used to compare the fits of parametric and semiparametric models (Keele 2007).

One drawback of nonlinear approaches is the lack of a single reportable summary statistic. However, plots of the mean nonlinear association with confidence intervals can provide a more informative summary of the strength of a statistical relationship than single statistics (Keele 2007). Nonetheless, a number of authors used GAMs only as exploratory tools for the diagnosis of nonlinearity (e.g., Agay-Shay et al. 2015 analyzed exposure tertiles after detecting nonlinearity through GAMs) and to inform a parametric functional form (e.g., Rodrigues et al. 2016).

GAMs may suffer from concurvity, the nonlinear analog to collinearity, yielding unstable estimates or misleading functional forms (Gu et al. 2010). Although shrinkage methods can improve numerical stability and predictive accuracy in the presence of collinearity, they do not address concurvity (Gu et al. 2010). Instead, the method of partial GAMs, which adjusts for concurvity through a variable selection procedure that sequentially eliminates functional dependencies between covariates, may be useful (Gu et al. 2010). This method yields a more parsimonious model that is easier to interpret and a graphical representation of concurvity between covariates (Gu et al. 2010).

BKMR is an alternative nonparametric approach, not used in the reviewed studies, which addresses the multifaceted goals of mixtures analyses (Table 2). BKMR offers flexible estimation of a multivariate exposure–response surface, allowing for nonlinearity and interactions while simultaneously performing componentwise or hierarchical variable selection, with the latter providing selection among a group of highly correlated exposures (Bobb et al. 2015; Coull et al. 2015). Other methods that can simultaneously and automatically select variables and functional forms include shrinkage approaches in GAMs (Marra and Wood 2011) and Bayesian approaches based on spike and slab priors (Scheipl et al. 2013).

For higher-dimensional problems, statistical learning methods may be more appropriate than GAMs and BKMR. MARS is a nonparametric method that proceeds in a manner similar to stepwise regression, building a model from piecewise linear basis functions and their products (Hastie et al. 2009; Morlini 2006) (Table 2). It is able to automatically select variables and identify interactions in large datasets, as well as modeling nonlinearity. However, it is not able to control for confounding as confounder terms are also subject to selection, and in the presence of high correlation between exposures, it may select exposures somewhat arbitrarily (Morlini 2006).

Interactions between EDCs

The assessment of interactions compounds the methodological complexity of mixtures analyses. Fitting a GLM with product terms between pairs of exposures quickly becomes intractable as the number of exposures increases; for k exposures, the assessment of all two-way interactions introduces $\binom{k}{2}$ interaction terms, which may overwhelm modest sample sizes and exacerbate multicollinearity. Even in cases where the parameter count-to-sample size ratio is small, correlated measurement error among biomarkers may diminish the power available to detect interactions (Pollack et al. 2013). The paucity of studies in this review examining interactions perhaps reflects these difficulties (Tables 1 and S3). Nevertheless, synergistic effects between EDCs are biologically plausible (Claus Henn et al. 2014; Diamanti-Kandarakis et al. 2009; Kortenkamp 2007; Rosa et al. 2011; Silins and Högberg 2011) and have been reported in studies of metal mixtures and child health (Claus Henn et al. 2014). Antagonistic and synergistic EDC-mixture effects have also been observed in human cell lines (Ribeiro et al. 2017) and toxicological studies (e.g., Christiansen et al. 2009); however, whereas interaction is defined as a departure from additivity, the definition of additivity can vary widely between (and among) toxicologists and epidemiologists (Howard and Webster 2013).

Assessment of interactions among EDCs needs to take into account the possibility of nonmonotonicity in exposure–response relationships. When both nonmonotonicity and interaction exist, the consequences of misspecifying either of these characteristics may be severe. In a linear regression model, given two correlated exposures that interact and have U- or inverted U-shaped relationships with the outcome, assuming linearity may indicate that the relationship is synergistic when it is truly antagonistic, or vice versa (Ganzach 1997). Moreover, failure to assess interaction may result in exposure–response relationships that appear flat or convex when they are truly concave, or vice versa (Ganzach 1997). To our knowledge, these issues have been insufficiently explored in analyses of environmental chemical mixtures.

We identified several methods for identifying and modeling interaction between chemical exposures (Table 2), none of which were used in the reviewed studies. To address the risk of multicollinearity when including main effects and pairwise product terms of correlated exposures, LASSO and grouped LASSO regression have been extended to fit sparse hierarchical interaction models, performing variable selection while honoring the hierarchy principle for interactions (that for nonzero interaction coefficients, one or both main effects should also be included in the model) (Bien et al. 2013; Lim and Hastie 2015). Sparse hierarchical interaction models can also be built using BHR-SSVS by introducing dependence into the usual independent Bernoulli prior for variable importance (Bien et al. 2013; Chipman 1996). Where nonlinearity is expected, pairwise and possibly higher-order interactions can be modeled using multivariate smoothers in GAMs, such as tensor products or thin plate splines (Ruppert et al. 2003). Nonlinear multivariate interactions can also be identified using BKMR, by plotting cross-sections of the estimated joint nonparametric exposure–response function (Bobb et al. 2015).

Tree-based statistical learning methods present an alternative for datasets with a large number of exposures. These methods do not

assume additivity or linearity in exposure–response relationships and are able to automatically identify and account for nonlinearity and pairwise or higher-order interaction effects (Lampa et al. 2014). Classification and regression trees (CART) is a nonparametric method that partitions data into a decision tree, and multiple trees can be combined to improve predictive performance using ensemble methods, such as stochastic gradient boosting (SGB) and random forests (RF) (Table 2). SGB combines trees by fitting a weighted additive model of trees based on random subsets of the data, whereas RF averages trees based on bootstrapped data and uses a random sample of exposures for determining splits in individual trees (Hastie et al. 2009). Although both methods provide a measure of variable importance, SGB may be preferred for identifying interactions in studies of chemical mixtures because it provides a measure of interaction strength (Lampa et al. 2014). Although the inability of these methods to impose a monotonicity constraint may be advantageous for modeling EDC exposure–response relationships, it is problematic when linearity needs to be enforced in the handling of confounders (Sun et al. 2013). This issue was addressed by Gass et al. (2014), who developed a CART algorithm that controls for confounding. A further drawback is that spurious interactions may be identified in the presence of nonlinearity and high correlation between exposures (Lampa et al. 2014). One solution may be to encourage repeated splits on the same variable in the CART algorithm (Friedman and Popescu 2008; Lampa et al. 2014); another option may be to use principal component scores in place of subsets of very highly correlated exposures (Lampa et al. 2014). Consequently, these methods need to be considered exploratory unless the results can be validated (e.g., using external data or split-sample validation; Lampa et al. 2014).

Some of the drawbacks of tree-based statistical learning methods are addressed by Bayesian additive regression trees (BART), a Bayesian ensemble method that fits a sum of trees model (Chipman et al. 2010) (Table 2). Although SGB and RF provide point estimates of variable importance, these scores reflect predictive performance, not direction or magnitude of exposure–response associations (Bobb et al. 2015), and there is no measure of uncertainty. BART shares the benefits of ensemble methods along with the benefits of a fully specified Bayesian model; i.e., full posterior inference capabilities, providing exposure-effect estimates and probability-based uncertainty intervals, and the ability to incorporate prior knowledge (Chipman et al. 2010; Hernández et al. 2015). Moreover, BART can be extended to include linear confounder terms and random effects (Chipman et al. 2010).

Simulation studies comparing the performance of methods for assessing interactions in analyses of environmental chemical mixtures are lacking. In a recent exposome simulation study, boosted regression trees had lower sensitivity but lower false discovery proportion than hierarchical grouped LASSO (Barrera-Gómez et al. 2017). In a simulation study of 4–20 exposures, LASSO, BMA, and supervised PCA all had high interaction detection rates, with BMA and LASSO possessing the lowest false positive rates (Sun et al. 2013). Using CART as a preprocessor to restrict the analysis to a subset of exposures improved interaction detection rates and false positive rates (Sun et al. 2013). The deletion/substitution/addition algorithm, not reviewed here, had poor performance in identifying interactions (Sun et al. 2013). Further simulation studies are warranted to characterize the performance of the available methods in identifying interactions between EDCs.

Exposure Measurement Error and Timing of Exposure Measurement

Careful timing of exposure measurement is crucial in studies of prenatal exposure to chemical mixtures, requiring consideration of both the etiologically relevant window of exposure and the persistence of chemical species within the mixture. EDCs exhibit a broad gamut of biological half-lives; at one extreme, the metabolism and excretion of some phenols and phthalates occur within hours, whereas the half-lives of lipophilic persistent organic pollutants are in the order of years. The levels of nonpersistent chemicals in pregnant women have been shown to exhibit high intraindividual variability (Braun et al. 2012), reflecting patterns in the use of personal care products (Braun et al. 2012), plastics (Braun et al. 2013), seasonality (e.g., in pesticide use; Whyatt et al. 2004), and changes in xenobiotic metabolism associated with pregnancy (Braun et al. 2012; Jusko et al. 2014). This variability suggests the need for repeated measurements, as spot measurements may not adequately capture long-term exposure levels or average exposure levels across a targeted window of susceptibility (Braun et al. 2012).

For persistent chemicals, the intraindividual variability is likely to be overshadowed by interindividual variability (Gennings et al. 2013; Longnecker et al. 1999), suggesting that timing of exposure assessment is more flexible for these chemicals (Romano et al. 2014). However, even within classes of persistent chemicals, the magnitude of intraindividual variability is likely to be chemical-specific (e.g., the nondioxin-like di-ortho PCB congeners have substantially longer half-lives than the dioxin-like mono-ortho PCBs; Heilmann et al. 2006). Increasing the complexity, further sources of measurement error include differential placental transfer rates of chemicals (e.g., for PBDE congeners, there is a lower rate of placental transfer with increasing degree of bromination; Frederiksen et al. 2010), differential rates of metabolism and bioaccumulation within the fetus across chemicals (e.g., fetal methylmercury levels are greater than maternal levels, whereas the converse is true for PBDEs; Frederiksen et al. 2010; Stern and Smith 2003), and the lower precision in measurement of exposures with concentration ranges near detection limits than higher-concentration-range compounds (Engel and Wolff 2013). These asymmetries between EDCs suggest that exposure may be misclassified to differing degrees for individual chemicals within a mixture, especially if exposure is measured at only a single point in pregnancy.

Although acknowledging the degree of exposure misclassification is universally important in epidemiological studies, it is of particular importance in studies of chemical mixtures. Exposure measurement error in a single-pollutant model is expected to bias exposure–response associations toward the null, when the error is nondifferential with respect to the outcome (Zeka and Schwartz 2004). However, when exposures with differing degrees of (non-differential) measurement error are analyzed in a multipollutant model, the measurement error in one variable is able to contribute bias to the coefficients of other variables (Zeka and Schwartz 2004). Importantly, both bias toward the null and away from the null are possible, with the overall direction and magnitude of the bias dependent on the extent of correlation between the exposures, and the variances and correlation of the measurement errors (Zeger et al. 2000). Regardless of the true strength of an association, an exposure with lower measurement error may appear more strongly associated with an outcome than an exposure with higher measurement error may appear (Vedal and Kaufman 2011; Woodruff et al. 2009). If identifying causal agents within a mixture is the goal of the analysis, this issue may lead to misleading conclusions, unless the precision of exposure measurement is considered or explicitly incorporated into statistical models (Woodruff et al. 2009).

Not only is the use of single measurements anticipated to bias effect estimates, it also reduces the power to detect exposure–response associations (Jusko et al. 2012; Perrier et al. 2016). In the Generation R study, Jusko et al. (2012) demonstrated that the required sample size for detecting linear associations was almost

halved when using the mean across three BPA measurements vs. one measurement. A more powerful analysis can be realized not only from increasing the number of individuals but also from increasing the number of measurements per individual of short half-life chemicals (Jusko et al. 2012; Perrier et al. 2016), though this gain in power is attenuated with increasing intraindividual correlation. To reduce costs of repeated exposure measurements, pooling multiple biospecimens per individual before assaying chemicals may decrease bias and increase power, in comparison with the use of a single spot measurement (Perrier et al. 2016).

The majority of studies we reviewed analyzed nonpersistent chemicals on the basis of a single measurement (Table S3). The few studies that took repeated measurements to reduce measurement error used the average of two spot measurements across pregnancy (e.g., Casas et al. 2016; Gascon et al. 2015b). Although this approach is useful for summarizing longitudinal exposure data when measurements are unstable over time, a more powerful analysis may be obtained by taking into account the longitudinal nature of the exposure data, using methods such as a two-stage mixed effects model, generalized additive mixed model, and functional clustering model (Chen et al. 2015). These methods avoid loss of information about exposure temporal variation and may enable other time-varying covariates to be incorporated (Chen et al. 2015).

Only two reviewed studies took exposure measurement error into account in their statistical methodology (Claus Henn et al. 2016; Heilmann et al. 2006) (Table S3). For incorporating measurement error into statistical models, numerous applicable methods exist that form part of an extensive statistical literature dealing with this issue (e.g., Carroll et al. 2006; Gustafson et al. 2003; Thomas 2009). An exhaustive review is outside the scope of this article, but we present SEM and BHR as examples. SEM (used by Heilmann et al. 2006) enables the simultaneous modeling of a system of multiple outcomes and exposures that are measured with error (Table 2) (Grandjean et al. 2004; Sánchez et al. 2005). SEM can correct for classical measurement error when multiple exposure measurements are available (e.g., exposure measured across multiple biological matrices or multiple metabolites), under certain distributional and independence assumptions (e.g., independent and normally distributed measurement errors). This method was demonstrated by Budtz-Jørgensen et al. (2002), in which a latent methylmercury exposure variable was assumed to be measured with error by two methylmercury biomarkers in umbilical cord blood and maternal hair. This approach is purported to correct for bias due to both laboratory imprecision and biological variation (Grandjean et al. 2004); however, as discussed above, SEMs have been criticized for their susceptibility to misspecification and reliance on strong causal and parametric assumptions that may lead to severe biases (VanderWeele 2012). BHR provides a flexible framework for correcting for bias due to measurement error, by treating exposures as latent variables and specifying additional submodels for the measurement error and prior (Gustafson et al. 2003; Muff et al. 2015; Richardson et al. 2002). Correcting measurement error bias requires prior knowledge of the measurement error mechanism (e.g., classical or Berkson) and the measurement error variance from repeated measures, external studies, or expert knowledge (Gustafson et al. 2003; Muff et al. 2015). Where prior information is lacking, sensitivity analysis may allow examination of the extent to which inferences are affected by measurement error correction.

Identifying Windows of Exposure Susceptibility in Mixtures Analyses

The challenges in identifying vulnerable gestational windows in mixtures analyses are considerable. Teasing out the independent

effects of multiple correlated exposures at multiple times, which may have nonlinear relationships with the outcome, may be an intractable problem when the number of exposures is large relative to the sample size. Consequently, most of the reviewed studies that compared windows of exposure susceptibility analyzed time points separately in parallel cross-sectional models (e.g., Erkin-Cakmak et al. 2015; Gascon et al. 2015a) (Table S3). Although this approach is straightforward, it does not allow for formal tests of differences in effect estimates across windows and may result in bias if exposure is not uniform throughout the window (Sánchez et al. 2011). Missing data may further hamper comparisons, if effect estimates at separate time points are based on different subsets of the data (Chen et al. 2015; Sánchez et al. 2011).

Several of the reviewed studies analyzed multiple exposures at multiple times in one regression model (e.g., Jacobson et al. 2015; Roen et al. 2015) (Table S3). Although this approach may enable the estimation of independent effects, estimates may be based on a reduced sample if a complete-case analysis is conducted due to missing exposure data at varying time points across individuals (Chen et al. 2015; Sánchez et al. 2011). The interpretation of some effect estimates may be temporally invalid (Chen et al. 2015); e.g., the effect of prenatal exposure controlling for postnatal exposure. Additionally, prenatal effect estimates may be subject to overadjustment bias if postnatal exposure is on the causal pathway between prenatal exposure and the outcome (Schisterman et al. 2009), which may be the case for persistent chemicals. Estimates may also be affected by multicollinearity, although dimension reduction methods may be used to address this (e.g., Maresca et al. 2016). Allowing prenatal exposure to be a determinant of latent postnatal exposure in a SEM analysis, as performed by Heilmann et al. (2006), is appealing; however, marginal structural models (not reviewed here) may be better equipped to address time-varying exposures and confounders while making fewer assumptions than SEMs (VanderWeele 2012).

Several additional methods have been suggested in the single-exposure setting (Sánchez et al. 2011). Multiple informant models (MIM) may be used to jointly estimate separate regression models for each window, allowing for formal testing of differences in effect estimates across windows (Sánchez et al. 2011). Baek et al. (2014) extended MIM for hierarchical data with multiple correlated predictors, examining the school food environment and childhood obesity, an approach that may be applicable to the study of several correlated chemical exposures over time. Alternatively mixed models may be used with random intercepts and slopes if exposure is measured in a sufficient number of occasions to enable the estimation of exposure trajectories (Sánchez et al. 2011). However, we are not aware of any studies that have implemented these methods in the multipollutant setting, and they may be limited to models of only a few exposures.

Two novel methods have recently been proposed for mixtures analyses with numerous repeated exposure measurements. Lagged WQSR can be used for modeling high-resolution exposure trajectories where there is heterogeneity in the timing of exposure measurement across individuals (e.g., from baby tooth biomarkers) (Bello et al. 2017). Based on a mixed modeling framework, lagged WQSR quantifies the magnitude and significance of the mixture effect over time, enabling the identification of critical windows of exposure susceptibility as well as the drivers of the association at each time point (Bello et al. 2017). The method is also attractive because of its robustness to missing values in exposure measures over time; however, one caveat is that the direction of expected effects of individual exposures on the outcome are required to be homogenous (Bello et al. 2017). Tree-based distributed lag models do not require homogeneity in the direction of expected effects but can only quantify the combined explanatory power of exposures

within a time window, not the direction of the mixture effect, and are more suited to settings with uniformly timed exposure measurement across individuals and datasets with no missing exposure values (Bello et al. 2017). Neither method is able to quantify individual exposure-effect estimates or interactions between mixture components over time, which would require estimation of the complex multidimensional exposure–response surface (Bello et al. 2017).

Synthesis and Recommendations

We have examined a variety of frequentist and Bayesian methods performing dimension reduction, shrinkage, variable selection, statistical learning, or smoothing, which may be preferred to the use of more established methods when modeling EDC mixtures. We concentrated on methods that may be able to achieve the manifold goals of mixtures analyses; however, few methods are able to achieve multiple goals simultaneously. Two methods developed specifically for mixtures analyses in environmental epidemiology deserve special mention: Both WQSR (Carrico et al. 2015) and BKMR (Bobb et al. 2015) provide a measure of mixture health effects while identifying drivers of the association, with the latter also accounting for nonlinearity and interaction in multivariate exposure–response relationships. For multifaceted research questions assessing the health effects of numerous correlated exposures, we suggest exploring staged statistical methods (e.g., preconditioned LASSO or using CART as a preprocessor; Sun et al. 2013), and leveraging the flexibility offered by Bayesian methods such as BHM, BKMR, or BART. However, method selection should be made with cognizance of the dimensionality of the problem, the sample size, the ability to adjust for confounding, and the relative importance of each goal to the research question, so prescriptive recommendations are not possible.

Hypothesis tests should be conducted with awareness of multiple testing issues; methods for addressing multiple testing have been discussed in depth elsewhere (Braun et al. 2016; Chadeau-Hyam et al. 2013; Patel 2017). Multiple testing issues are inadequately addressed by many variable selection methods, so tools for valid selective inference may need to be explored (e.g., Berk et al. 2013; Lockhart et al. 2014; Meinshausen and Bühlmann 2010; Tibshirani et al. 2016). Moreover, the use of data-driven model optimization approaches, such as *k*-fold cross-validation for the selection of tuning parameters in penalized regression, may lead to model selection instability and concerns regarding the reproducibility of results (Lim and Yu 2016; Roberts and Nowak 2014). Cross-validation selects a solution that optimizes predictive performance but not necessarily performance metrics for variable selection and parameter estimation (Lim and Yu 2016). We recommend considering methods that address estimation stability (e.g., Lim and Yu 2016; Roberts and Nowak 2014); focusing on both estimation stability and predictive performance may improve false positive rates (Kirk et al. 2013; Lim and Yu 2016). Bayesian variable selection may also be highly sensitive to prior specification (Coull et al. 2015) and involve uncertainty in the choice of variable selection thresholds (Bleich et al. 2014); therefore, careful specification of priors and hyperparameters is needed.

One of our key conclusions is that the selection of chemicals to analyze as a mixture should be made with care. Too many chemicals in one model may increase the risk of multicollinearity, and no statistical method is able to separate the signals of very highly correlated exposures without further information. Conversely, too few chemicals in one model risks bias from omitted copollutant confounders, with the possibility of detecting an association that is due to a correlated copollutant (Braun et al. 2016). Collinearity can exacerbate bias due to model misspecification, so careful elucidation of a

causal framework prior to data analysis is essential (Schisterman et al. 2017). These considerations are important in the study of EDCs, where data on important confounders, such as endogenous hormone levels (Kortenkamp 2007), dietary exposures [e.g., omega-3 fatty acids (Engel and Wolff 2013), phytoestrogens, mycoestrogens], and the microbiome [which is involved in xenobiotic metabolism (Dietert and Silbergeld 2015)], may be unavailable. Moreover, bias due to the joint effect of multiple unmeasured (even weak) confounders can be substantial, especially if they are mutually correlated but independent of the included confounders (Groenwold et al. 2016), which may be the case when one class of EDCs is the focus of an analysis and other important classes of EDCs are excluded. For the reasons outlined, opportunistic selection of variables with the assumption that small data sets will necessarily reveal true patterns in exposure–health associations should be discouraged, in favor of careful consideration of biological mechanisms, modes of action, phenomenological similarity, and the potential for interaction, when combining EDCs (Kortenkamp 2007).

We stress the importance of considering exposure measurement error in studies of chemical mixtures. Failure to do so may bias effect estimates toward or away from the null, to an unpredictable extent (Zeka and Schwartz 2004). Exposure measurement error can be reduced for nonpersistent chemicals by taking multiple exposure measurements during a hypothesized window of susceptibility; these measurements can be pooled and assayed once per individual to reduce costs (Perrier et al. 2016). Nevertheless, mixture components are likely to be measured with differing degrees of measurement error, so we encourage selection of a statistical method that explicitly accounts for measurement error or exploring the impact of measurement error on effect estimates through sensitivity analysis, rather than rationalizing the sequelae of disregarding this important issue (Thomas et al. 2007; Woodruff et al. 2009).

Limitations

Our findings should be considered in the context of the following limitations. Our results are limited to studies that stated that they were analyzing multiple exposures in the abstract. Although this approach may not have identified all of the existing EDC-mixtures studies, we are likely to have captured the majority of studies that used nonconventional statistical methods, as these studies are more likely to have expressed their methodological novelty in the abstract. In addition, we performed a targeted review of the statistical literature that did not include all possible statistical methods for studying the health effects of prenatal EDC mixtures.

Conclusions

Biomonitoring studies suggest that pregnant women are exposed to multiple EDCs simultaneously. This finding, coupled with toxicological evidence of adverse developmental effects following prenatal exposure to mixtures of EDCs and the possibility of underestimating effects if exposures are studied one at a time, provides compelling motivation for the analysis of EDC exposures in combination rather than separately. Understanding the health risks of exposure to EDC mixtures requires taking into account the potential for copollutant confounding and interaction between exposures, prudent selection of mixture components, as well as acknowledging that analyses based on simplifying assumptions of the nature of EDC exposure–response relationships may have unintended consequences. To this end, we examined the impact of methodological choices and assumptions regarding additivity, linearity, and exposure measurement timing

on the appropriateness of causal inference in the study of prenatal exposure to EDC mixtures. We concurrently identified multiple applicable statistical methods and compared their strengths and weaknesses for modeling these exposures. Although several reviews have discussed multipollutant statistical methods in air pollution epidemiology and for the analysis of exposome–health associations (Billionnet et al. 2012; Davalos et al. 2016; Patel 2017; Stafoggia et al. 2017), none to our knowledge have placed them in the context of the specific statistical and epidemiological challenges facing the study of prenatal exposure to EDCs. Nevertheless, our recommendations may also apply to studies of postnatal exposures or exposures to environmental chemicals not acting through the endocrine system.

Epidemiological studies are uniquely placed to provide insight into the potential human health risks of complex EDC mixtures that are more representative of real-world exposure patterns than the analysis of single chemicals. Understanding the role of prenatal exposure to EDC mixtures on development may help to solidify evidence for the suspected developmental etiologies of many endocrine-sensitive outcomes in childhood and adulthood, the increasing incidence of which remains unexplained (Diamanti-Kandarakis et al. 2009). We hope that this resource will aid researchers in selecting the appropriate analytical strategies and statistical methodology in this vital research area.

Acknowledgments

References

Agay-Shay K, Martinez D, Valvi D, García-Esteban R, Basagaña X, Robinson O, et al. 2015. Exposure to endocrine-disrupting chemicals during pregnancy and weight at 7 years of age: a multi-pollutant approach. Environ Health Perspect 123(10):1030–1037, PMID: 25956007, https://doi.org/10.1289/ehp.1409049.

Agier L, Portengen L, Chadeau-Hyam M, Basagaña X, Giorgis-Allemand L, Siroux V, et al. 2016. A systematic comparison of linear regression–based statistical methods to assess exposome-health associations. Environ Health Perspect 124(12):1848–1856, PMID: 27219331, https://doi.org/10.1289/EHP172.

Al-Gubory KH. 2014. Environmental pollutants and lifestyle factors induce oxidative stress and poor prenatal development. Reprod BioMed Online 29(1):17–31, PMID: 24813750, https://doi.org/10.1016/j.rbmo.2014.03.002.

Arrebola JP, Molina-Molina JM, Fernández MF, Sáenz JM, Amaya E, Indiveri P, et al. 2015. A novel biomarker for anti-androgenic activity in placenta reveals risks of urogenital malformations. Reproduction 149(6):605–613, PMID: 25784770, https://doi.org/10.1530/REP-14-0525.

Axelrad DA, Goodman S, Woodruff TJ. 2009. PCB body burdens in US women of childbearing age 2001–2002: an evaluation of alternate summary metrics of NHANES data. Environ Res 109(4):368–378, PMID: 19251256, https://doi.org/10.1016/j.envres.2009.01.003.

Baek J, Sánchez BN, Sanchez-Vaznaugh EV. 2014. Hierarchical multiple informants models: examining food environment contributions to the childhood obesity epidemic. Stat Med 33(4):662–674, PMID: 24038440, https://doi.org/10.1002/sim.5967.

Bair E, Hastie T, Paul D, Tibshirani R. 2006. Prediction by supervised principal components. J Am Stat Assoc 101(473):119–137, https://doi.org/10.1198/016214505000000628.

Barrera-Gómez J, Agier L, Portengen L, Chadeau-Hyam M, Giorgis-Allemand L, Siroux V, et al. 2017. A systematic comparison of statistical methods to detect interactions in exposome-health associations. Environ Health 16(1):74, PMID: 28709428, https://doi.org/10.1186/s12940-017-0277-6.

Barrett ES, Parlett LE, Sathyanarayana S, Redmon JB, Nguyen RHN, Swan SH. 2016. Prenatal stress as a modifier of associations between phthalate exposure and reproductive development: results from a multicentre pregnancy cohort study. Paediatr Perinat Epidemiol 30(2):105–114, PMID: 26576028, https://doi.org/10.1111/ppe.12264.

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. J Stat Softw 67(1), https://doi.org/10.18637/jss.v067.i01.

Bello GA, Arora M, Austin C, Horton MK, Wright RO, Gennings C. 2017. Extending the Distributed Lag Model framework to handle chemical mixtures. Environ Res 156:253–264, PMID: 28371754, https://doi.org/10.1016/j.envres.2017.03.031.

Berg V, Nøst TH, Pettersen RD, Hansen S, Veyhe AS, Jorde R, et al. 2016. Persistent organic pollutants and the association with maternal and infant thyroid homeostasis: a multipollutant assessment. Environ Health Perspect 125(1):127–133, PMID: 27219111, https://doi.org/10.1289/EHP152.

Berk R, Brown L, Buja A, Zhang K, Zhao L. 2013. Valid post-selection inference. Ann Statist 41(2):802–837, https://doi.org/10.1214/12-AOS1077.

Bien J, Taylor J, Tibshirani R. 2013. A lasso for hierarchical interactions. Ann Stat 41(3):1111–1141, PMID: 26257447, https://doi.org/10.1214/13-AOS1096.

Billionnet C, Sherrill D, Annesi-Maesano I, GERIE Study. 2012. Estimating the health effects of exposure to multi-pollutant mixture. Ann Epidemiol 22(2):126–141, PMID: 22226033, https://doi.org/10.1016/j.annepidem.2011.11.004.

Bleich J, Kapelner A, George EI, Jensen ST. 2014. Variable selection for BART: an application to gene regulation. Ann Appl Stat 8(3):1750–1781, https://doi.org/10.1214/14-AOAS755.

Bobb JF, Valeri L, Claus Henn B, Christiani DC, Wright RO, Mazumdar M, et al. 2015. Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. Biostatistics 16(3):493–508, PMID: 25532525, https://doi.org/10.1093/biostatistics/kxu058.

Braun JM, Gennings C, Hauser R, Webster TF. 2016. What can epidemiological studies tell us about the impact of chemical mixtures on human health? Environ Health Perspect 124(1):A6–A9, PMID: 26720830, https://doi.org/10.1289/ehp.1510569.

Braun JM, Kalkbrenner AE, Just AC, Yolton K, Calafat AM, Sjödin A, et al. 2014. Gestational exposure to endocrine-disrupting chemicals and reciprocal social, repetitive, and stereotypic behaviors in 4- and 5-year-old children: the HOME study. Environ Health Perspect 122(5):513–520, PMID: 24622245, https://doi.org/10.1289/ehp.1307261.

Braun JM, Sathyanarayana S, Hauser R. 2013. Phthalate exposure and children's health. Curr Opin Pediatr 25(2):247–254, PMID: 23429708, https://doi.org/10.1097/MOP.0b013e32835e1eb6.

Braun JM, Smith KW, Williams PL, Calafat AM, Berry K, Ehrlich S, et al. 2012. Variability of urinary phthalate metabolite and bisphenol A concentrations before and during pregnancy. Environ Health Perspect 120(5):739–745, PMID: 22262702, https://doi.org/10.1289/ehp.1104139.

Brieño-Enríquez MA, Reig-Viader R, Cabero L, Toran N, Martinez F, Roig I, et al. 2012. Gene expression is altered after bisphenol A exposure in human fetal oocytes in vitro. Mol Hum Reprod 18(4):171–183, PMID: 22121209, https://doi.org/10.1093/molehr/gar074.

Buckley JP, Engel SM, Braun JM, Whyatt RM, Daniels JL, Mendez MA, et al. 2016a. Prenatal phthalate exposures and body mass index among 4- to 7-year-old children: a pooled analysis. Epidemiology 27(3):449–458, PMID: 26745610, https://doi.org/10.1097/EDE.0000000000000436.

Buckley JP, Engel SM, Mendez MA, Richardson DB, Daniels JL, Calafat AM, et al. 2016b. Prenatal phthalate exposures and childhood fat mass in a New York City cohort. Environ Health Perspect 124(4):507–513, PMID: 26308089, https://doi.org/10.1289/ehp.1509788.

Budtz-Jørgensen E, Keiding N, Grandjean P, Weihe P. 2002. Estimation of health effects of prenatal methylmercury exposure using structural equation models. Environ Health 1(1):2, PMID: 12513702, https://doi.org/10.1186/1476-069X-1-2.

Bühlmann P, Hothorn T. 2007. Boosting algorithms: regularization, prediction and model fitting. Statist Sci 22(4):477–505, https://doi.org/10.1214/07-STS242.

Buscail C, Chevrier C, Serrano T, Pelé F, Monfort C, Cordier S, et al. 2015. Prenatal pesticide exposure and otitis media during early childhood in the PELAGIE mother-child cohort. Occup Environ Med 72(12):837–844, PMID: 26347056, https://doi.org/10.1136/oemed-2015-103039.

Carlin DJ, Rider CV, Woychik R, Birnbaum LS. 2013. Unraveling the health effects of environmental mixtures: an NIEHS priority. Environ Health Perspect 121(1):a6–a8, PMID: 23409283, https://doi.org/10.1289/ehp.1206182.

Carrico C, Gennings C, Wheeler DC, Factor-Litvak P. 2015. Characterization of weighted quantile sum regression for highly correlated data in a risk analysis setting. J Agric Biol Environ Stat 20(1):100–120, PMID: 30505142, https://doi.org/10.1007/s13253-014-0180-3.

Carroll R, Ruppert D, Stefanski L, Crainiceanu C. 2006. *Measurement Error in Nonlinear Models*. 2nd ed. Boca Raton, FL:Chapman & Hall/CRC.

Casas M, Nieuwenhuijsen M, Martinez D, Ballester F, Basagaña X, Basterrechea M, et al. 2015. Prenatal exposure to PCB-153, p,p'-DDE and birth outcomes in 9000

mother–child pairs: Exposure–response relationship and effect modifiers. Environ Int 74:23–31, PMID: 25314142, https://doi.org/10.1016/j.envint.2014.09.013.

Casas M, Valvi D, Ballesteros-Gómez A, Gascon M, Fernández MF, García-Esteban R, et al. 2016. Exposure to bisphenol A and phthalates during pregnancy and ultrasound measures of fetal growth in the INMA-Sabadell cohort. Environ Health Perspect 124(4):521–528, PMID: 26196298, https://doi.org/10.1289/ehp.1409190.

Chadeau-Hyam M, Campanella G, Jombart T, Bottolo L, Portengen L, Vineis P, et al. 2013. Deciphering the complex: Methodological overview of statistical models to derive OMICS-based biomarkers. Environ Mol Mutagen 54(7):542–557, PMID: 23918146, https://doi.org/10.1002/em.21797.

Chen Y-H, Ferguson KK, Meeker JD, McElrath TF, Mukherjee B. 2015. Statistical methods for modeling repeated measures of maternal environmental exposure biomarkers during pregnancy in association with preterm birth. Environ Health 14:9, PMID: 25619201, https://doi.org/10.1186/1476-069X-14-9.

Chevrier C, Limon G, Monfort C, Rouget F, Garlantézec R, Petit C, et al. 2011. Urinary biomarkers of prenatal atrazine exposure and adverse birth outcomes in the PELAGIE birth cohort. Environ Health Perspect 119(7):1034–1041, PMID: 21367690, https://doi.org/10.1289/ehp.1002775.

Chevrier C, Petit C, Philippat C, Mortamais M, Slama R, Rouget F, et al. 2012. Maternal urinary phthalates and phenols and male genital anomalies. Epidemiology 23(2):353–356, PMID: 22317818, https://doi.org/10.1097/EDE.0b013e318246073e.

Chipman H. 1996. Bayesian variable selection with related predictors. Can J Statistics 24(1):17–36, https://doi.org/10.2307/3315687.

Chipman HA, George EI, McCulloch RE. 2010. BART: Bayesian additive regression trees. Ann Appl Stat 4(1):266–298, https://doi.org/10.1214/09-AOAS285.

Christiansen S, Scholze M, Dalgaard M, Vinggaard AM, Axelstad M, Kortenkamp A, et al. 2009. Synergistic disruption of external male sex organ development by a mixture of four antiandrogens. Environ Health Perspect 117(12):1839–1846, PMID: 20049201, https://doi.org/10.1289/ehp.0900689.

Chun H, Keleş S. 2010. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. J R Stat Soc Series B Stat Methodol 72(1):3–25, PMID: 20107611, https://doi.org/10.1111/j.1467-9868.2009.00723.x.

Chung D, Chun H, Keleş S. 2018. spls: Sparse partial least squares (SPLS) regression and classification. R package version 2.2-2. Available: https://CRAN.R-project.org/package=spls [accessed 7 June 2018].

Claus Henn B, Coull BA, Wright RO. 2014. Chemical mixtures and children's health. Curr Opin Pediatr 26(2):223–229, PMID: 24535499, https://doi.org/10.1097/MOP.0000000000000067.

Claus Henn B, Ettinger AS, Hopkins MR, Jim R, Amarasiriwardena C, Christiani DC, et al. 2016. Prenatal arsenic exposure and birth outcomes among a population residing near a mining-related superfund site. Environ Health Perspect 124(8):1308–1315, PMID: 26859631, https://doi.org/10.1289/ehp.1510070.

Cooke PS, Sato T, Buchanan DL. 2001. Disruption of steroid hormone signaling by PCBs. In: *PCBs: Recent advances in environmental toxicology and health effects*. Lexington, KY: University Press of Kentucky, 257–263.

Coull BA, Bobb JF, Wellenius GA, Kioumourtzoglou M-A, Mittleman MA, Koutrakis P, et al. 2015. Part 1. Statistical learning methods for the effects of multiple air pollution constituents. In: Development of statistical methods for multipollutant research. Research Report 183. Boston, MA:Health Effects Institute, 5–50.

Czarnota J, Gennings C, Wheeler DC. 2015. Assessment of weighted quantile sum regression for modeling chemical mixtures and cancer risk. Cancer Inform 14(Suppl 2):159–171, PMID: 26005323, https://doi.org/10.4137/CIN.S17295.

Czarnota J, Wheeler D. 2015. wqs: Weighted quantile sum regression. R package version 0.0.1. https://CRAN.R-project.org/package=wqs [accessed 7 June 2018].

Damgaard IN, Skakkebaek NE, Toppari J, Virtanen HE, Shen H, Schramm K-W, et al. 2006. Persistent pesticides in human breast milk and cryptorchidism. Environ Health Perspect 114(7):1133–1138, PMID: 16835070, https://doi.org/10.1289/ehp.8741.

Darbre PD. 2015. Chapter 2 - How could endocrine disrupters affect human health? In: *Endocrine Disruption and Human Health* (Darbre PD, ed). Boston:Academic Press, 27–45, https://doi.org/10.1016/B978-0-12-801139-3.00002-8.

Davalos AD, Luben TJ, Herring AH, Sacks JD. 2016. Current approaches used in epidemiologic studies to examine short-term multipollutant air pollution exposures. Ann Epidemiol, PMID: 28040377, https://doi.org/10.1016/j.annepidem.2016.11.016.

Diamanti-Kandarakis E, Bourguignon J-P, Giudice LC, Hauser R, Prins GS, Soto AM, et al. 2009. Endocrine-disrupting chemicals: an Endocrine Society scientific statement. Endocr Endocr Rev 30(4):293–342, PMID: 19502515, https://doi.org/10.1210/er.2009-0002.

Dietert RR, Silbergeld EK. 2015. Biomarkers for the 21st century: listening to the microbiome. Toxicol Sci 144(2):208–216, PMID: 25795652, https://doi.org/10.1093/toxsci/kfv013.

Dobson AJ, Barnett AG. 2008. *An Introduction to Generalized Linear Models*. 3rd ed. Boca Raton, FL:Chapman & Hall/CRC Press.

Eisen E, Agalliu I, Thurston S, Coull B, Checkoway H. 2004. Smoothing in occupational cohort studies: an illustration based on penalised splines. Occup

Environ Med 61(10):854–860, PMID: 15377772, https://doi.org/10.1136/oem.2004.013136.

Engel SM, Wolff MS. 2013. Causal inference considerations for endocrine disruptor Research in children's health. Annu Rev Public Health 34:139–158, PMID: 23514318, https://doi.org/10.1146/annurev-publhealth-031811-124556.

Erkin-Cakmak A, Harley KG, Chevrier J, Bradman A, Kogut K, Huen K, et al. 2015. In utero and childhood polybrominated diphenyl ether exposures and body mass at age 7 years: the CHAMACOS study. Environ Health Perspect 123(6):636–642, PMID: 25738596, https://doi.org/10.1289/ehp.1408417.

Fernández MF, Olmos B, Granada A, Lopez-Espinosa M-J, Molina-Molina J-M, Fernandez JM, et al. 2007. Human exposure to endocrine-disrupting chemicals and prenatal risk factors for cryptorchidism and hypospadias: a nested case-control study. Environ Health Perspect 115(Suppl 1):8–14, PMID: 18174944, https://doi.org/10.1289/ehp.9351.

Fernández MF, Rivas A, Olea-Serrano F, Cerrillo I, Molina-Molina JM, Araque P, et al. 2004. Assessment of total effective xenoestrogen burden in adipose tissue and identification of chemicals responsible for the combined estrogenic effect. Anal Bioanal Chem 379(1):163–170, PMID: 15024544, https://doi.org/10.1007/s00216-004-2558-5.

Forns J, Mandal S, Iszatt N, Polder A, Thomsen C, Lyche JL, et al. 2016. Novel application of statistical methods for analysis of multiple toxicants identifies DDT as a risk factor for early child behavioral problems. Environ Res 151:91–100, PMID: 27466755, https://doi.org/10.1016/j.envres.2016.07.014.

Fox J, Nie Z, Byrnes J. 2017. sem: Structural equation models. R package version 3.1-9. https://CRAN.R-project.org/package=sem [accessed 7 June 2018].

Frederiksen M, Thomsen C, Frøshaug M, Vorkamp K, Thomsen M, Becher G, et al. 2010. Polybrominated diphenyl ethers in paired samples of maternal and umbilical cord blood plasma and associations with house dust in a Danish cohort. Int J Hyg Environ Health 213(4):233–242, PMID: 20471317, https://doi.org/10.1016/j.ijheh.2010.04.008.

Friedman J, Hastie T, Tibshirani R. 2010. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 33(1):1–22, PMID: 20808728, https://doi.org/10.1109/TPAMI.2005.127.

Friedman JH, Popescu BE. 2008. Predictive learning via rule ensembles. Ann Appl Stat 2(3):916–954, https://doi.org/10.1214/07-AOAS148.

Friedman JH, Roosen CB. 1995. An introduction to multivariate adaptive regression splines. Stat Methods Med Res 4(3):197–217, PMID: 8548103, https://doi.org/10.1177/096228029500400303.

Frommlet F, Ruhaltinger F, Twaróg P, Bogdan MG. 2012. Modified versions of Bayesian information criterion for genome-wide association studies. Comput Stat Data Anal 56(5):1038–1051, https://doi.org/10.1016/j.csda.2011.05.005.

Ganzach Y. 1997. Misleading interaction and curvilinear terms. Psychol Methods 2(3):235–247, https://doi.org/10.1037/1082-989X.2.3.235.

Gascon M, Casas M, Morales E, Valvi D, Ballesteros-Gómez A, Luque N, et al. 2015a. Prenatal exposure to bisphenol A phthalates and childhood respiratory tract infections and allergy. J Allergy Clin Immunol 135(2):370–378.e7, PMID: 25445825, https://doi.org/10.1016/j.jaci.2014.09.030.

Gascon M, Valvi D, Forns J, Casas M, Martinez D, Júlvez J, et al. 2015b. Prenatal exposure to phthalates and neuropsychological development during childhood. Int J Hyg Environ Health 218(6):550–558, PMID: 26095249, https://doi.org/10.1016/j.ijheh.2015.05.006.

Gass K, Klein M, Chang HH, Flanders WD, Strickland MJ. 2014. Classification and regression trees for epidemiologic research: an air pollution example. Environ Health 13(1):17, PMID: 24625053, https://doi.org/10.1186/1476-069X-13-17.

Gaudriault P, Mazaud-Guittot S, Lavoué V, Coiffec I, Lesné L, Dejucq-Rainsford N, et al. 2017. Endocrine disruption in human fetal testis explants by individual and combined exposures to selected pharmaceuticals, pesticides, and environmental pollutants. Environ Health Perspect 125(8):087004, PMID: 28796631, https://doi.org/10.1289/EHP1014.

Gennings C, Carrico C, Factor-Litvak P, Krigbaum N, Cirillo PM, Cohn BA. 2013. A cohort study evaluation of maternal PCB exposure related to time to pregnancy in daughters. Environ Health 12(1):66, PMID: 23962309, https://doi.org/10.1186/1476-069X-12-66.

Gennings C, Sabo R, Carney E. 2010. Identifying subsets of complex mixtures most associated with complex diseases: polychlorinated biphenyls and endometriosis as a case study. Epidemiology 21(Suppl 4):S77–S84, PMID: 21422968, https://doi.org/10.1097/EDE.0b013e3181ce946c.

Gillman MW. 2005. Developmental origins of health and disease. N Engl J Med 353(17):1848–1850, https://doi.org/10.1056/NEJMe058187.

Govarts E, Remy S, Bruckers L, Hond Den E, Sioen I, Nelen V, et al. 2016. Combined effects of prenatal exposures to environmental chemicals on birth weight. Int J Environ Res Public Health 13(5):495, PMID: 27187434, https://doi.org/10.3390/ijerph13050495.

Grandjean P, Barouki R, Bellinger DC, Casteleyn L, Chadwick LH, Cordier S, et al. 2015. Life-long implications of developmental exposure to environmental stressors: new perspectives. Endocrinology 156(10):3408–3415, PMID: 26241067, https://doi.org/10.1210/EN.2015-1350.

Grandjean P, Budtz-Jørgensen E, Keiding N, Weihe P. 2004. Underestimation of risk due to exposure misclassification. Int J Occup Med Environ Health 17(1):131–136, PMID: 15212216.

Gray KA, Klebanoff MA, Brock JW, Zhou H, Darden R, Needham L, et al. 2005. In utero exposure to background levels of polychlorinated biphenyls and cognitive functioning among school-age children. Am J Epidemiol 162(1):17–26, PMID: 15961582, https://doi.org/10.1093/aje/kwi158.

Greenland S. 1993. Methods for epidemiologic analyses of multiple exposures: a review and comparative study of maximum-likelihood, preliminary-testing, and empirical-Bayes regression. Stat Med 12(8):717–736, PMID: 8516590, https://doi.org/10.1002/sim.4780120802.

Greenland S. 1994. Hierarchical regression for epidemiologic analyses of multiple exposures. Environ Health Perspect 102(Suppl 8):33–39, PMID: 7851328, https://doi.org/10.1289/ehp.94102s833.

Greenland S. 1995. Avoiding power loss associated with categorization and ordinal scores in dose-response and trend analysis. Epidemiology 6(4):450–454, PMID: 7548361.

Groenwold RHH, Sterne JAC, Lawlor DA, Moons KGM, Hoes AW, Tilling K. 2016. Sensitivity analysis for the effects of multiple unmeasured confounders. Ann Epidemiol 26(9):605–611, PMID: 27576907, https://doi.org/10.1016/j.annepidem.2016.07.009.

Gu H, Kenney T, Zhu M. 2010. Partial generalized additive models: an information-theoretic approach for dealing with concurvity and selecting variables. J Comput Graph Stat 19(3):531–551, https://doi.org/10.1198/jcgs.2010.07139.

Gustafson P. 2003. *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments*. Boca Raton, FL: Chapman & Hall/CRC.

Harley KG, Chevrier J, Aguilar Schall R, Sjödin A, Bradman A, Eskenazi B. 2011. Association of prenatal exposure to polybrominated diphenyl ethers and infant birth weight. Am J Epidemiol 174(8):885–892, PMID: 21878423, https://doi.org/10.1093/aje/kwr212.

Hastie T, Tibshirani R, Friedman J. 2009. *The Elements of Statistical Learning*. 2nd ed. New York: Springer-Verlag.

Hayes L, Weening A, Morey LM. 2016. Differential effects of estradiol and bisphenol A on SET8 and SIRT1 expression in ovarian cancer cells. Dose Response 14(2):1559325816640682, PMID: 27114721, https://doi.org/10.1177/1559325816640682.

Heilmann C, Grandjean P, Weihe P, Nielsen F, Budtz-Jørgensen E. 2006. Reduced antibody responses to vaccinations in children exposed to polychlorinated biphenyls. PLOS Med 3(8):e311, PMID: 16942395, https://doi.org/10.1371/journal.pmed.0030311.

Hernández B, Raftery AE, Pennington SR, Parnell AC. 2015. Bayesian additive regression trees using Bayesian model averaging. https://arxiv.org/abs/1507.00181v2 [accessed 9 May 2017].

Herring AH. 2010. Nonparametric Bayes shrinkage for assessing exposures to mixtures subject to limits of detection. Epidemiology 21(Suppl 4):S71–S76, PMID: 20526202, https://doi.org/10.1097/EDE.0b013e3181cf0058.

Howard GJ, Webster TF. 2013. Contrasting theories of interaction in epidemiology and toxicology. Environ Health Perspect 121(1):1–6, PMID: 23014866, https://doi.org/10.1289/ehp.1205889.

Huel G, Yazbeck C, Burnel D, Missy P, Kloppmann W. 2004. Environmental boron exposure and activity of delta-aminolevulinic acid dehydratase (ALA-D) in a newborn population. Toxicol Sci 80(2):304–309, PMID: 15141099, https://doi.org/10.1093/toxsci/kfh165.

Huen K, Yousefi P, Bradman A, Yan L, Harley KG, Kogut K, et al. 2014. Effects of age, sex, and persistent organic pollutants on DNA methylation in children. Environ Mol Mutagen 55(3):209–222, PMID: 24375655, https://doi.org/10.1002/em.21845.

Ingber SZ, Buser MC, Pohl HR, Abadin HG, Murray HE, Scinicariello F. 2013. DDT/DDE and breast cancer: a meta-analysis. Regul Toxicol Pharmacol 67(3):421–433, PMID: 24021539, https://doi.org/10.1016/j.yrtph.2013.08.021.

Jacobson JL, Muckle G, Ayotte P, Dewailly E, Jacobson SW. 2015. Relation of prenatal methylmercury exposure from environmental sources to childhood IQ. Environ Health Perspect 123(8):827–833, PMID: 25757069, https://doi.org/10.1289/ehp.1408554.

Jusko TA, De Roos AJ, Schwartz SM, Lawrence BP, Palkovicova L, Nemessanyi T, et al. 2010. A cohort study of developmental polychlorinated biphenyl (PCB) exposure in relation to post-vaccination antibody response at 6-months of age. Environ Res 110(4):388–395, PMID: 20378105, https://doi.org/10.1016/j.envres.2010.02.010.

Jusko T, Shaw P, Snijder C, Pierik F, Koch H, Hauser R, et al. 2012. O-051: Reliability of urinary bisphenol A concentrations measured during pregnancy and implications for epidemiologic study design [Abstract]. Epidemiology 23(5S):1, https://doi.org/10.1097/01.ede.0000416709.49302.6e.

Jusko TA, Shaw PA, Snijder CA, Pierik FH, Koch HM, Hauser R, et al. 2014. Reproducibility of urinary bisphenol A concentrations measured during pregnancy in the Generation R Study. J Expo Sci Environ Epidemiol 24(5):532–536, PMID: 24736100, https://doi.org/10.1038/jes.2014.23.

Kapelner A, Bleich J. 2016. bartMachine: machine learning with Bayesian additive regression trees. J Stat Softw 70(4):1–40, https://doi.org/10.18637/jss.v070.i04.

Keele L. 2007. *Semiparametric Regression for the Social Sciences*. Chichester, UK: John Wiley & Sons.

Kim Y, Ha E-H, Park H, Ha M, Kim Y, Hong Y-C, et al. 2013. Prenatal lead and cadmium co-exposure and infant neurodevelopment at 6 months of age: the Mothers and Children's Environmental Health (MOCEH) study. Neurotoxicology 35:15–22, PMID: 23220728, https://doi.org/10.1016/j.neuro.2012.11.006.

Kippler M, Wagatsuma Y, Rahman A, Nermell B, Persson L-Å, Raqib R, et al. 2012. Environmental exposure to arsenic and cadmium during pregnancy and fetal size: a longitudinal study in rural Bangladesh. Reprod Toxicol 34(4):504–511, PMID: 22985739, https://doi.org/10.1016/j.reprotox.2012.08.002.

Kirk P, Witkover A, Bangham CRM, Richardson S, Lewin AM, Stumpf MPH. 2013. Balancing the robustness and predictive performance of biomarkers. J Comput Biol 20(12):979–989, PMID: 23909374, https://doi.org/10.1089/cmb.2013.0018.

Kobrosly RW, Evans S, Miodovnik A, Barrett ES, Thurston SW, Calafat AM, et al. 2014. Prenatal phthalate exposures and neurobehavioral development scores in boys and girls at 6-10 years of age. Environ Health Perspect 122(5):521–528, PMID: 24577876, https://doi.org/10.1289/ehp.1307063.

Konishi K, Sasaki S, Kato S, Ban S, Washino N, Kajiwara J, et al. 2009. Prenatal exposure to PCDDs/PCDFs and dioxin-like PCBs in relation to birth weight. Environ Res 109(7):906–913, PMID: 19683226, https://doi.org/10.1016/j.envres.2009.07.010.

Kortenkamp A. 2006. Breast cancer, oestrogens and environmental pollutants: a re-evaluation from a mixture perspective. Int J Androl 29(1):193–198, PMID: 16466540, https://doi.org/10.1111/j.1365-2605.2005.00613.x.

Kortenkamp A. 2007. Ten years of mixing cocktails: a review of combination effects of endocrine-disrupting chemicals. Environ Health Perspect 115(Suppl 1):98–105, PMID: 18174957, https://doi.org/10.1289/ehp.9357.

Kortenkamp A. 2008. Low dose mixture effects of endocrine disrupters: implications for risk assessment and epidemiology. Int J Androl 31(2):233–240, PMID: 18248400, https://doi.org/10.1111/j.1365-2605.2007.00862.x.

Kortenkamp A, Faust M, Scholze M, Backhaus T. 2007. Low-level exposure to multiple chemicals: reason for human health concerns? Environ Health Perspect 115(Suppl 1):106–114, PMID: 18174958, https://doi.org/10.1289/ehp.9358.

Koskenniemi JJ, Virtanen HE, Kiviranta H, Damgaard IN, Matomäki J, Thorup JM, et al. 2015. Association between levels of persistent organic pollutants in adipose tissue and cryptorchidism in early childhood: a case–control study. Environ Health 14:78, PMID: 26403566, https://doi.org/10.1186/s12940-015-0065-0.

Krysiak-Baltyn K, Toppari J, Skakkebaek NE, Jensen TS, Virtanen HE, Schramm K-W, et al. 2012. Association between chemical pattern in breast milk and congenital cryptorchidism: modelling of complex human exposures. Int J Androl 35(3):294–302, PMID: 22519522, https://doi.org/10.1111/j.1365-2605.2012.01268.x.

Kyung M, Gill J, Ghosh M, Casella G. 2010. Penalized regression, standard errors, and Bayesian lassos. Bayesian Anal 5(2):369–411, https://doi.org/10.1214/10-BA607.

Lampa E, Lind L, Lind PM, Bornefalk-Hermansson A. 2014. The identification of complex interactions in epidemiology and toxicology: a simulation study of boosted regression trees. Environ Health 13:57, PMID: 24993424, https://doi.org/10.1186/1476-069X-13-57.

LaRocca J, Binder AM, McElrath TF, Michels KB. 2016. First-trimester urine concentrations of phthalate metabolites and phenols and placenta miRNA expression in a cohort of U.S. women. Environ Health Perspect 124(3):380–387, PMID: 26090578, https://doi.org/10.1289/ehp.1408409.

Lenters V, Portengen L, Rignell-Hydbom A, Jönsson BA, Lindh CH, Piersma AH, et al. 2015a. Prenatal phthalate, perfluoroalkyl acid, and organochlorine exposures and term birth weight in three birth cohorts: multi-pollutant models based on elastic net regression. Environ Health Perspect 124(3):365–372, PMID: 26115335, https://doi.org/10.1289/ehp.1408933.

Lenters V, Portengen L, Smit LAM, Jönsson BAG, Giwercman A, Rylander L, et al. 2015b. Phthalates, perfluoroalkyl acids, metals and organochlorines and reproductive function: a multipollutant assessment in Greenlandic, Polish and Ukrainian men. Occup Environ Med 72(6):385–393, PMID: 25209848, https://doi.org/10.1136/oemed-2014-102264.

Lenters V, Vermeulen R, Portengen L. 2018. Performance of variable selection methods for assessing the health effects of correlated exposures in case-control studies. Occup Environ Med 2018 Jul; 75(7):522–529, PMID: 28947495, https://doi.org/10.1136/oemed-2016-104231.

Liew Z, Ritz B, Ehrenstein von OS, Bech BH, Nohr EA, Fei C, et al. 2015. Attention deficit/hyperactivity disorder and childhood autism in association with prenatal exposure to perfluoroalkyl substances: a nested case-control study in the Danish National Birth Cohort. Environ Health Perspect 123(4):367–373, PMID: 25616253, https://doi.org/10.1289/ehp.1408412.

Lignell S, Aune M, Darnerud PO, Hanberg A, Larsson SC, Glynn A. 2013. Prenatal exposure to polychlorinated biphenyls (PCBs) and polybrominated diphenyl ethers (PBDEs) may influence birth weight among infants in a Swedish cohort

with background exposure: a cross-sectional study. Environ Health 12:44, PMID: 23724965, https://doi.org/10.1186/1476-069X-12-44.

Lim M, Hastie T. 2015. Learning interactions via hierarchical group-lasso regularization. J Comput Graph Stat 24(3):627–654, PMID: 26759522, https://doi.org/10.1080/10618600.2014.938812.

Lim C, Yu B. 2016. Estimation stability with cross-validation (ESCV). J Comput Graph Stat 25(2):464–492, https://doi.org/10.1080/10618600.2015.1020159.

Limousi F, Albouy-Llaty M, Carles C, Dupuis A, Rabouan S, Migeot V. 2014. Does area deprivation modify the association between exposure to a nitrate and low-dose atrazine metabolite mixture in drinking water and small for gestational age? A historic cohort study. Environ Sci Pollut Res Int 21(7):4964–4973, PMID: 23771441, https://doi.org/10.1007/s11356-013-1893-5.

Lockhart R, Taylor J, Tibshirani RJ, Tibshirani R. 2014. A significance test for the lasso. Ann Stat 42(2):413–468, PMID: 25574062, https://doi.org/10.1214/13-AOS1175.

Longnecker MP, Klebanoff MA, Gladen BC, Berendes HW. 1999. Serial levels of serum organochlorines during pregnancy and postpartum. Arch Environ Health 54(2):110–114, PMID: 10094288, https://doi.org/10.1080/00039899909602244.

MacLehose RF, Dunson DB, Herring AH, Hoppin JA. 2007. Bayesian methods for highly correlated exposure data. Epidemiology 18(2):199–207, PMID: 17272963, https://doi.org/10.1097/01.ede.0000256320.30737.c0.

Maresca MM, Hoepner LA, Hassoun A, Oberfield SE, Mooney SJ, Calafat AM, et al. 2016. Prenatal exposure to phthalates and childhood body size in an urban cohort. Environ Health Perspect 124(4):514–520, PMID: 26069025, https://doi.org/10.1289/ehp.1408750.

Marques RC, Bernardi JVE, Dórea JG, de Fatima R Moreira M, Malm O. 2014. Perinatal multiple exposure to neurotoxic (lead, methylmercury, ethylmercury, and aluminum) substances and neurodevelopment at six and 24 months of age. Environ Pollut 187:130–135, PMID: 24486466, https://doi.org/10.1016/j.envpol.2014.01.004.

Marra G, Wood SN. 2011. Practical variable selection for generalized additive models. Comput Stat Data Anal 55(7):2372–2387, https://doi.org/10.1016/j.csda.2011.02.004.

May S, Bigelow C. 2006. Modeling nonlinear dose-response relationships in epidemiologic studies: statistical approaches and practical challenges. Dose Response 3(4):474–490, PMID: 18648629, https://doi.org/10.2203/dose-response.003.04.004.

Meeker JD. 2012. Exposure to environmental endocrine disruptors and child development. Arch Pediatr Adolesc Med 166(6):E1–E7, PMID: 22664748, https://doi.org/10.1001/archpediatrics.2012.241.

Meinshausen N, Bühlmann P. 2010. Stability selection. J R Stat Soc: Series B Stat Methodol 72(4):417–473, https://doi.org/10.1111/j.1467-9868.2010.00740.x.

Merkle EC, Rosseel Y. 2016. blavaan: Bayesian structural equation models via parameter expansion. http://arxiv.org/abs/1511.05604v2 [accessed 7 June 2018].

Mevik B-H, Wehrens R, Liland KH. 2016. pls: Partial least squares and principal component regression. R package version 2.6-0. https://CRAN.R-project.org/package=pls [accessed 7 June 2018].

Migeot V, Albouy-Llaty M, Carles C, Limousi F, Strezlec S, Dupuis A, et al. 2013. Drinking-water exposure to a mixture of nitrate and low-dose atrazine metabolites and small-for-gestational age (SGA) babies: a historic cohort study. Environ Res 122:58–64, PMID: 23340115, https://doi.org/10.1016/j.envres.2012.12.007.

Milborrow S. 2017. earth: Multivariate Adaptive Regression Splines. R package version 4.6.3. https://CRAN.R-project.org/package=earth [accessed 7 June 2018].

Miller RL, Garfinkel R, Horton M, Camann D, Perera FP, Whyatt RM, et al. 2004. Polycyclic aromatic hydrocarbons, environmental tobacco smoke, and respiratory symptoms in an inner-city birth cohort. Chest 126(4):1071–1078, PMID: 15486366, https://doi.org/10.1378/chest.126.4.1071.

Mitro SD, Johnson T, Zota AR. 2015. Cumulative chemical exposures during pregnancy and early development. Curr Environ Health Rep 2(4):367–378, PMID: 26341623, https://doi.org/10.1007/s40572-015-0064-x.

Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 339:b2535, PMID: 19622551, https://doi.org/10.1136/bmj.b2535.

Momoli F, Abrahamowicz M, Parent M-É, Krewski D, Siemiatycki J. 2010. Analysis of multiple exposures: an empirical comparison of results from conventional and semi-Bayes modeling strategies. Epidemiology 21(1):144–151, PMID: 20010218, https://doi.org/10.1097/EDE.0b013e3181c297c7.

Morlini I. 2006. On multicollinearity and concurvity in some nonlinear multivariate models. JISS 15(1):3–26, https://doi.org/10.1007/s10260-006-0005-9.

Muff S, Riebler A, Held L, Rue H, Saner P. 2015. Bayesian analysis of measurement error models using integrated nested Laplace approximations. J R Stat Soc C 64(2):231–252, PMID: 25810131, https://doi.org/10.1111/rssc.12069.

Myers JP, Zoeller RT, Saal vom FS. 2009. A clash of old and new scientific concepts in toxicity, with important implications for public health. Environ Health Perspect 117(11):1652–1655, PMID: 20049113, https://doi.org/10.1289/ehp.0900887.

Naksen W, Prapamontol T, Mangklabruks A, Chantara S, Thavornyutikarn P, Srinual N, et al. 2015. Associations of maternal organophosphate pesticide exposure and PON1 activity with birth outcomes in SAWASDEE birth cohort,

Thailand. Environ Res 142:288–296, PMID: 26186137, https://doi.org/10.1016/j.envres.2015.06.035.

Neugebauer J, Wittsiepe J, Kasper-Sonnenberg M, Schöneck N, Schölmerich A, Wilhelm M. 2015. The influence of low level pre- and perinatal exposure to PCDD/Fs, PCBs, and lead on attention performance and attention-related behavior among German school-aged children: results from the Duisburg Birth Cohort Study. Int J Hyg Environ Health 218(1):153–162, PMID: 25456149, https://doi.org/10.1016/j.ijheh.2014.09.005.

Newbold RR, Padilla-Banks E, Jefferson WN. 2009. Environmental estrogens and obesity. Mol Cell Endocrinol 304(1–2):84–89, PMID: 19433252, https://doi.org/10.1016/j.mce.2009.02.024.

NIEHS (National Institute of Environmental Health Sciences). 2015. Statistical approaches for assessing health effects of environmental chemical mixtures in epidemiology studies [Workshop], 13–14 July 2015, National Research of Environmental Health Sciences, Research Triangle Park, North Carolina, USA. www.niehs.nih.gov/about/events/pastmtg/2015/statistical/ [accessed 5 April 2017].

Ochiai S, Shimojo N, Yuka I, Watanabe M, Matsuno Y, Suzuki S, et al. 2014. A pilot study for foetal exposure to multiple persistent organic pollutants and the development of infant atopic dermatitis in modern Japanese society. Chemosphere 94:48–52, PMID: 24080003, https://doi.org/10.1016/j.chemosphere.2013.09.009.

Park H-Y, Park J-S, Sovcikova E, Kocan A, Linderholm L, Bergman A, et al. 2009. Exposure to hydroxylated polychlorinated biphenyls (OH-PCBs) in the prenatal period and subsequent neurodevelopment in eastern Slovakia. Environ Health Perspect 117(10):1600–1606, PMID: 20019912, https://doi.org/10.1289/ehp.0900611.

Pastor-Barriuso R, Fernández MF, Castaño-Vinyals G, Whelan D, Pérez-Gómez B, Llorca J, et al. 2016. Total effective xenoestrogen burden in serum samples and risk for breast cancer in a population-based multicase-control study in Spain. Environ Health Perspect 124(10):1575–1582, PMID: 27203080, https://doi.org/10.1289/EHP157.

Patel CJ. 2017. Analytic complexity and challenges in identifying mixtures of exposures associated with phenotypes in the exposome era. Curr Epidemiol Rep 4(1):22–30, PMID: 28251040, https://doi.org/10.1007/s40471-017-0100-5.

Paul D, Bair E, Hastie T, Tibshirani R. 2007. "Pre-conditioning" for feature selection and regression in high-dimensional problems. Ann Statist 36(4):1595–1618, https://doi.org/10.1214/009053607000000578.

Perrier F, Giorgis-Allemand L, Slama R, Philippat C. 2016. Within-subject pooling of biological samples to reduce exposure misclassification in biomarker-based studies. Epidemiology 27(3):378–388, PMID: 27035688, https://doi.org/10.1097/EDE.0000000000000460.

Pfeiffer RM, Redd A, Carroll RJ. 2017. On the impact of model selection on predictor identification and parameter inference. Comput Stat 32(2):667–690, PMID: 28690368, https://doi.org/10.1007/s00180-016-0690-2.

Philippat C, Mortamais M, Chevrier C, Petit C, Calafat AM, Ye X, et al. 2012. Exposure to phthalates and phenols during pregnancy and offspring size at birth. Environ Health Perspect 120(3):464–470, PMID: 21900077, https://doi.org/10.1289/ehp.1103634.

Polanska K, Dettbarn G, Jurewicz J, Sobala W, Magnus P, Seidel A, et al. 2014. Effect of prenatal polycyclic aromatic hydrocarbons exposure on birth outcomes: the Polish mother and child cohort study. Biomed Res Int 2014:408939, PMID: 25140312, https://doi.org/10.1155/2014/408939.

Pollack AZ, Perkins NJ, Mumford SL, Ye A, Schisterman EF. 2013. Correlated biomarker measurement error: an important threat to inference in environmental epidemiology. Am J Epidemiol 177(1):84–92, PMID: 23221725, https://doi.org/10.1093/aje/kws209.

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing. www.R-project.org/ [accessed 7 June 2018].

Raftery A, Hoeting J, Volinsky C, Painter I, Yeung KY. 2018. BMA: Bayesian Model Averaging. R package version 3.18.8. https://CRAN.R-project.org/package=BMA [accessed 7 June 2018].

Rajapakse N, Silva E, Kortenkamp A. 2002. Combining xenoestrogens at levels below individual no-observed-effect concentrations dramatically enhances steroid hormone action. Environ Health Perspect 110(9):917–921, PMID: 12204827, https://doi.org/10.1289/ehp.02110917.

Renzetti S, Curtin P, Just AC, Bello G, Gennings C. 2018. gWQS: Generalized weighted quantile sum regression. R package version 1.1.0. https://CRAN.R-project.org/package=gWQS [accessed 7 June 2018].

Ribeiro E, Ladeira C, Viegas S. 2017. EDCs mixtures: a stealthy hazard for human health? Toxics 5(1):5, PMID: 29051438, https://doi.org/10.3390/toxics5010005.

Richardson S, Leblond L, Jaussent I, Green PJ. 2002. Mixture models in measurement error problems, with reference to epidemiological studies. J Royal Statistical Soc A 165(3):549–566, https://doi.org/10.1111/1467-985X.00252.

Ridgeway G. 2017. gbm: Generalized boosted regression models. R package version 2.1.3. https://CRAN.R-project.org/package=gbm [accessed 7 June 2018].

Roberts S, Martin MA. 2006. Using supervised principal components analysis to assess multiple pollutant effects. Environ Health Perspect 114(12):1877–1882, PMID: 17185279, https://doi.org/10.1289/ehp.9226.

Roberts S, Nowak G. 2014. Stabilizing the lasso against cross-validation variability. Comput Stat Data Anal 70:198–211, https://doi.org/10.1016/j.csda.2013.09.008.

Rodrigues EG, Bellinger DC, Valeri L, Hasan MOSI, Quamruzzaman Q, Golam M, et al. 2016. Neurodevelopmental outcomes among 2- to 3-year-old children in Bangladesh with elevated blood lead and exposure to arsenic and manganese in drinking water. Environ Health 15:44, PMID: 26968381, https://doi.org/10.1186/s12940-016-0127-y.

Roen EL, Wang Y, Calafat AM, Wang S, Margolis A, Herbstman J, et al. 2015. Bisphenol A exposure and behavioral problems among inner city children at 7-9 years of age. Environ Res 142:739–745, PMID: 25724466, https://doi.org/10.1016/j.envres.2015.01.014.

Roli G, Monari P. 2015. Hierarchical Bayesian models for the estimation of correlated effects in multilevel data: a simulation study to assess model performance. Commun Stat Theory and Methods 44(12):2644–2653, https://doi.org/10.1080/03610926.2013.806662.

Romano ME, Savitz DA, Braun JM. 2014. Challenges and future directions to evaluating the association between prenatal exposure to endocrine disrupting chemicals and childhood obesity. Curr Epidemiol Rep 1(2):57–66, PMID: 25328860, https://doi.org/10.1007/s40471-014-0007-3.

Rosa MJ, Jung KH, Perzanowski MS, Kelvin EA, Darling KW, Camann DE, et al. 2011. Prenatal exposure to polycyclic aromatic hydrocarbons, environmental tobacco smoke and asthma. Respir Med 105(6):869–876, PMID: 21163637, https://doi.org/10.1016/j.rmed.2010.11.022.

Rull RP, Ritz B, Shaw GM. 2006. Neural tube defects and maternal residential proximity to agricultural pesticide applications. Am J Epidemiol 163(8):743–753, PMID: 16495467, https://doi.org/10.1093/aje/kwj101.

Ruppert D, Wand MP, Carroll RJ. 2003. *Semiparametric Regression*. Cambridge, UK:Cambridge University Press.

Ruppert D, Wand MP, Carroll RJ. 2009. Semiparametric regression during 2003–2007. Electron J Stat 3:1193–1256, PMID: 20305800, https://doi.org/10.1214/09-EJS525.

Sánchez BN, Budtz-Jørgensen E, Ryan LM, Hu H. 2005. Structural equation models. J Am Stat Assoc 100(472):1443–1455, https://doi.org/10.1198/016214505000001005.

Sánchez BN, Hu H, Litman HJ, Téllez-Rojo MM. 2011. Statistical methods to study timing of vulnerability with sparsely sampled data on environmental toxicants. Environ Health Perspect 119(3):409–415, PMID: 21362588, https://doi.org/10.1289/ehp.1102453.

Scheipl F, Kneib T, Fahrmeir L. 2013. Penalized likelihood and Bayesian function selection in regression models. AStA Adv Stat Anal 97(4):349–385, https://doi.org/10.1007/s10182-013-0211-3.

Schisterman EF, Cole SR, Platt RW. 2009. Overadjustment bias and unnecessary adjustment in epidemiologic studies. Epidemiology 20(4):488–495, PMID: 19525685, https://doi.org/10.1097/EDE.0b013e3181a819a1.

Schisterman EF, Perkins NJ, Mumford SL, Ahrens KA, Mitchell EM. 2017. Collinearity and causal diagrams - a lesson on the importance of model specification. Epidemiology 28(1):47–53, PMID: 27676260, https://doi.org/10.1097/EDE.0000000000000554.

Sigg CD, Buhmann JM. 2008. Expectation-maximization for sparse and non-negative PCA. In: *Proceedings of ICML '08, The 25th International Conference on Machine Learning*, 05–09 July 2008, Helsinki, Finland 960–967, https://doi.org/10.1145/1390156.1390277.

Silins I, Högberg J. 2011. Combined toxic exposures and human health: biomarkers of exposure and effect. Int J Environ Res Public Health 8(3):629–647, PMID: 21556171, https://doi.org/10.3390/ijerph8030629.

Silva E, Rajapakse N, Kortenkamp A. 2002. Something from "nothing" — eight weak estrogenic chemicals combined at concentrations below NOECs produce significant mixture effects. Environ Sci Technol 36(8):1751–1756, PMID: 11993873, https://doi.org/10.1021/es0101227.

Silver MK, Meeker JD. 2015. Chapter 13 - Endocrine disruption of developmental pathways and children's health. In: *Endocrine Disruption and Human Health*. Darbre PD, ed. Boston:Academic Press, 237–255.

Skakkebaek NE, Toppari J, Söder O, Gordon CM, Divall S, Draznin M. 2011. The exposure of fetuses and children to endocrine disrupting chemicals: a European Society for Paediatric Endocrinology (ESPE) and Pediatric Endocrine Society (PES) call to action statement. J Clin Endocrinol Metab 96(10):3056–3058, PMID: 21832106, https://doi.org/10.1210/jc.2011-1269.

Slama R, Cordier S. 2010. Environmental contaminants and impacts on healthy and successful pregnancies. In: *Environmental Impacts on Reproductive Health and Fertility*. Woodruff TJ, Janssen SJ, Guillette JL Jr, Giudiceeds LC, eds). Cambridge, UK:Cambridge University Press, 125–144.

Sonneborn D, Park H-Y, Petrik J, Kocan A, Palkovicova L, Trnovec T, et al. 2008. Prenatal polychlorinated biphenyl exposures in eastern Slovakia modify effects of social factors on birthweight. Paediatr Perinat Epidemiol 22(3):202–213, PMID: 18426515, https://doi.org/10.1111/j.1365-3016.2008.00929.x.

Soto AM, Vandenberg LN, Maffini MV, Sonnenschein C. 2008. Does breast cancer start in the womb?. Basic Clin Pharmacol Toxicol 102(2):125–133, PMID: 18226065, https://doi.org/10.1111/j.1742-7843.2007.00165.x.

Stafoggia M, Breitner S, Hampel R, Basagaña X. 2017. Statistical approaches to address multi-pollutant mixtures and multiple exposures: the state of the science. Curr Environ Health Rep 4(4):481–490, PMID: 28988291, https://doi.org/10.1007/s40572-017-0162-z.

Stern AH, Smith AE. 2003. An assessment of the cord blood: maternal blood methylmercury ratio: implications for risk assessment. Environ Health Perspect 111(12):1465–1470, PMID: 12948885, https://doi.org/10.1289/ehp.6187.

Stewart PW, Lonky E, Reihman J, Pagano J, Gump BB, Darvill T. 2008. The relationship between prenatal PCB exposure and intelligence (IQ) in 9-year-old children. Environ Health Perspect 116(10):1416–1422, PMID: 18941588, https://doi.org/10.1289/ehp.11058.

Stroustrup A, Hsu H-H, Svensson K, Schnaas L, Cantoral A, Solano González M, et al. 2016. Toddler temperament and prenatal exposure to lead and maternal depression. Environ Health 15(1):71, PMID: 27312840, https://doi.org/10.1186/s12940-016-0147-7.

Stroustrup A, Swan SH. 2013. Endocrine disruptors. In: *Textbook of Children's Environmental Health*. Landrigan PJ, Etzel RA, eds. New York:Oxford University Press, 325–332.

Sturza J, Silver MK, Xu L, Li M, Mai X, Xia Y, et al. 2016. Prenatal exposure to multiple pesticides is associated with auditory brainstem response at 9 months in a cohort study of Chinese infants. Environ Int 92–93:478–485, https://doi.org/10.1016/j.envint.2016.04.035.

Sun Z, Tao Y, Li S, Ferguson KK, Meeker JD, Park SK, et al. 2013. Statistical strategies for constructing health risk models with multiple pollutants and their interactions: possible choices and comparisons. Environ Health 12(1):85, PMID: 24093917, https://doi.org/10.1186/1476-069X-12-85.

Suzuki K, Nakai K, Sugawara T, Nakamura T, Ohba T, Shimada M, et al. 2010. Neurobehavioral effects of prenatal exposure to methylmercury and PCBs, and seafood intake: neonatal behavioral assessment scale results of Tohoku study of child development. Environ Res 110(7):699–704, PMID: 20673887, https://doi.org/10.1016/j.envres.2010.07.001.

Swartz MD, Cai Y, Chan W, Symanski E, Mitchell LE, Danysh HE, et al. 2015. Air toxics and birth defects: a Bayesian hierarchical approach to evaluate multiple pollutants and spina bifida. Environ Health 14:16, PMID: 25971584, https://doi.org/10.1186/1476-069X-14-16.

Symanski E, Tee Lewis PG, Chen T-Y, Chan W, Lai D, Ma X. 2016. Air toxics and early childhood acute lymphocytic leukemia in Texas, a population based case control study. Environ Health 15(1):70, PMID: 27301866, https://doi.org/10.1186/s12940-016-0154-8.

Talbott EO, Marshall LP, Rager JR, Arena VC, Sharma RK, Stacy SL. 2015. Air toxics and the risk of autism spectrum disorder: the results of a population based case-control study in southwestern Pennsylvania. Environ Health 14:80, PMID: 26444407, https://doi.org/10.1186/s12940-015-0064-1.

Tang-Péronard JL, Heitmann BL, Jensen TK, Vinggaard AM, Madsbad S, Steuerwald U, et al. 2015. Prenatal exposure to persistent organochlorine pollutants is associated with high insulin levels in 5-year-old girls. Environ Res 142:407–413, PMID: 26232659, https://doi.org/10.1016/j.envres.2015.07.009.

Tatsuta N, Nakai K, Murata K, Suzuki K, Iwai-Shimada M, Kurokawa N, et al. 2014. Impacts of prenatal exposures to polychlorinated biphenyls, methylmercury, and lead on intellectual ability of 42-month-old children in Japan. Environ Res 133:321–326, PMID: 24998460, https://doi.org/10.1016/j.envres.2014.05.024.

Taylor KW, Joubert BR, Braun JM, Dilworth C, Gennings C, Hauser R, et al. 2016. Statistical approaches for assessing health effects of environmental chemical mixtures in epidemiology: lessons from an innovative workshop. Environ Health Perspect 124(12):A227–A229, PMID: 27905274, https://doi.org/10.1289/EHP547.

Thomas DC. 2009. *Statistical Methods in Environmental Epidemiology*. New York: Oxford University Press.

Thomas DC, Witte JS, Greenland S. 2007. Dissecting effects of complex mixtures: who's afraid of informative priors? Epidemiology 18(2):186–190, PMID: 17301703, https://doi.org/10.1097/01.ede.0000254682.47697.70.

Tibshirani RJ, Taylor J, Lockhart R, Tibshirani R. 2016. Exact post-selection inference for sequential regression procedures. J Am Stat Assoc 111(514):600–620, https://doi.org/10.1080/01621459.2015.1108848.

Tran NN, Pham TT, Ozawa K, Nishijo M, Nguyen ATN, Tran TQ, et al. 2016. Impacts of perinatal dioxin exposure on motor coordination and higher cognitive development in Vietnamese preschool children: a five-year follow-up. PLoS One 11(1):e0147655, PMID: 26824471, https://doi.org/10.1371/journal.pone.0147655.

Tu Y-K, Gunnell D, Gilthorpe MS. 2008. Simpson's Paradox, Lord's Paradox, and Suppression Effects are the same phenomenon – the reversal paradox. Emerg Themes Epidemiol 5:2, PMID: 18211676, https://doi.org/10.1186/1742-7622-5-2.

Vafeiadi M, Georgiou V, Chalkiadaki G, Rantakokko P, Kiviranta H, Karachaliou M, et al. 2015. Association of prenatal exposure to persistent organic pollutants

with obesity and cardiometabolic traits in early childhood: the Rhea mother-child cohort (Crete, Greece). Environ Health Perspect 123(10):1015–1021, PMID: 25910281, https://doi.org/10.1289/ehp.1409062.

Vafeiadi M, Vrijheid M, Fthenou E, Chalkiadaki G, Rantakokko P, Kiviranta H, et al. 2014. Persistent organic pollutants exposure during pregnancy, maternal gestational weight gain, and birth outcomes in the mother-child cohort in Crete, Greece (RHEA study). Environ Int 64:116–123, PMID: 24389008, https://doi.org/10.1016/j.envint.2013.12.015.

Valvi D, Mendez MA, Martinez D, Grimalt JO, Torrent M, Sunyer J, et al. 2012. Prenatal concentrations of polychlorinated biphenyls, DDE, and DDT and overweight in children: a prospective birth cohort study. Environ Health Perspect 120(3):451–457, PMID: 22027556, https://doi.org/10.1289/ehp.1103862.

Vandenberg LN, Colborn T, Hayes TB, Heindel JJ, Jacobs DR Jr, Lee D-H, et al. 2012. Hormones and endocrine-disrupting chemicals: low-dose effects and nonmonotonic dose responses. Endocr Rev 33(3):378–455, PMID: 22419778, https://doi.org/10.1210/er.2011-1050.

VanderWeele TJ. 2012. Invited commentary: structural equation models and epidemiologic analysis. Am J Epidemiol 176(7):608–612, PMID: 22956513, https://doi.org/10.1093/aje/kws213.

Varshavsky JR, Zota AR, Woodruff TJ. 2016. A novel method for calculating potency-weighted cumulative phthalates exposure with implications for identifying racial/ethnic disparities among U.S. reproductive-aged women in NHANES 2001–2012. Environ Sci Technol 50(19):10616–10624, PMID: 27579903, https://doi.org/10.1021/acs.est.6b00522.

Vedal S, Kaufman JD. 2011. What does multi-pollutant air pollution research mean? Am J Respir Crit Care Med 183(1):4–6, PMID: 21193783, https://doi.org/10.1164/rccm.201009-1520ED.

Vejrup K, Schjølberg S, Knutsen HK, Kvalem HE, Brantsæter AL, Meltzer HM, et al. 2016. Prenatal methylmercury exposure and language delay at three years of age in the Norwegian Mother and Child Cohort Study. Environ Int 92–93:63–69, PMID: 27058928, https://doi.org/10.1016/j.envint.2016.03.029.

Vilahur N, Fernández MF, Bustamante M, Ramos R, Forns J, Ballester F, et al. 2014. In utero exposure to mixtures of xenoestrogens and child neuropsychological development. Environ Res 134:98–104, PMID: 25086706, https://doi.org/10.1016/j.envres.2014.07.002.

Vrijheid M, Casas M, Gascon M, Valvi D, Nieuwenhuijsen M. 2016. Environmental pollutants and child health - a review of recent concerns. Int J Hyg Environ Health 219(4-5):331–342, PMID: 27216159, https://doi.org/10.1016/j.ijheh.2016.05.001.

Waldron L, Pintilie M, Tsao M-S, Shepherd FA, Huttenhower C, Jurisica I. 2011. Optimized application of penalized regression methods to diverse genomic data. Bioinformatics 27(24):3399–3406, PMID: 22156367, https://doi.org/10.1093/bioinformatics/btr591.

Wang D, Zhang W, Bakhai A. 2004. Comparison of Bayesian model averaging and stepwise methods for model selection in logistic regression. Stat Med 23(22):3451–3467, PMID: 15505893, https://doi.org/10.1002/sim.1930.

Whyatt RM, Rauh V, Barr DB, Camann DE, Andrews HF, Garfinkel R, et al. 2004. Prenatal insecticide exposures and birth weight and length among an urban minority cohort. Environ Health Perspect 112(10):1125–1132, PMID: 15238288, https://doi.org/10.1289/ehp.6641.

Wickerham EL, Lozoff B, Shao J, Kaciroti N, Xia Y, Meeker JD. 2012. Reduced birth weight in relation to pesticide mixtures detected in cord blood of full-term infants. Environ Int 47:80–85, PMID: 22796478, https://doi.org/10.1016/j.envint.2012.06.007.

Windham GC, Zhang L, Gunier R, Croen LA, Grether JK. 2006. Autism spectrum disorders in relation to distribution of hazardous air pollutants in the San Francisco Bay area. Environ Health Perspect 114(9):1438–1444, PMID: 16966102, https://doi.org/10.1289/ehp.9120.

Wojtyniak BJ, Rabczenko D, Jönsson BAG, Zvezday V, Pedersen HS, Rylander L, et al. 2010. Association of maternal serum concentrations of 2,2', 4,4'5,5'-hexachlorobiphenyl (CB-153) and 1,1-dichloro-2,2-bis (p-chlorophenyl)-ethylene (p,p-'DDE) levels with birth weight, gestational age and preterm births in Inuit and European populations. Environ Health 9:56, PMID: 20819217, https://doi.org/10.1186/1476-069X-9-56.

Wolff MS, Engel SM, Berkowitz GS, Ye X, Silva MJ, Zhu C, et al. 2008. Prenatal phenol and phthalate exposures and birth outcomes. Environ Health Perspect 116(8):1092–1097, PMID: 18709157, https://doi.org/10.1289/ehp.11007.

Wood SN. 2017. *Generalized Additive Models: An Introduction with R*. 2nd ed. Boca Raton, FL:Chapman & Hall/CRC Press.

Woodruff TJ, Parker JD, Darrow LA, Slama R, Bell ML, Choi H, et al. 2009. Methodological issues in studies of air pollution and reproductive health. Environ Res 109(3):311–320, PMID: 19215915, https://doi.org/10.1016/j.envres.2008.12.012.

Woodruff TJ, Zota AR, Schwartz JM. 2011. Environmental chemicals in pregnant women in the United States: NHANES 2003–2004. Environ Health Perspect 119(6):878–885, PMID: 21233055, https://doi.org/10.1289/ehp.1002727.

Xu X, Chiung YM, Lu F, Qiu S, Ji M, Huo X. 2015. Associations of cadmium, bisphenol A and polychlorinated biphenyl co-exposure in utero with placental gene expression and neonatal outcomes. Reprod Toxicol 52:62–70, PMID: 25687722, https://doi.org/10.1016/j.reprotox.2015.02.004.

Yolton K, Xu Y, Sucharew H, Succop P, Altaye M, Popelar A, et al. 2013. Impact of low-level gestational exposure to organophosphate pesticides on neurobehavior in early infancy: a prospective study. Environ Health 12(1):79, PMID: 24034442, https://doi.org/10.1186/1476-069X-12-79.

Yorifuji T, Debes F, Weihe P, Grandjean P. 2011. Prenatal exposure to lead and cognitive deficit in 7- and 14-year-old children in the presence of concomitant exposure to similar molar concentration of methylmercury. Neurotoxicol Teratol 33(2):205–211, PMID: 20868742, https://doi.org/10.1016/j.ntt.2010.09.004.

Zeger SL, Thomas D, Dominici F, Samet JM, Schwartz J, Dockery D, et al. 2000. Exposure measurement error in time-series studies of air pollution: concepts and consequences. Environ Health Perspect 108(5):419–426, PMID: 10811568, https://doi.org/10.1289/ehp.00108419.

Zeka A, Schwartz J. 2004. Estimating the independent effects of multiple pollutants in the presence of measurement error: an application of a measurement-error-resistant technique. Environ Health Perspect 112(17):1686–1690, PMID: 15579414, https://doi.org/10.1289/ehp.7286.

Zoeller RT, Bergman A, Becher G, Bjerregaard P, Bornman R, Brandt I, et al. 2014. A path forward in the debate over health impacts of endocrine disrupting chemicals. Environ Health 13:442, PMID: 25533907, https://doi.org/10.1186/1476-069X-13-118.

Zou H, Hastie T. 2005. Regularization and variable selection via the elastic net. J Royal Statistical Soc B 67(2):301–320, https://doi.org/10.1111/j.1467-9868.2005.00503.x.