METHODS IN ENVIRONMENTAL EPIDEMIOLOGY (AZ POLLACK AND NJ PERKINS, SECTION EDITORS)

Complex Mixtures, Complex Analyses: an Emphasis on Interpretable Results

Elizabeth A. Gibson[1] · Jeff Goldsmith[2] · Marianthi-Anna Kioumourtzoglou[1]

Abstract

Purpose of Review The purpose of this review is to outline the main questions in environmental mixtures research and provide a non-technical explanation of novel or advanced methods to answer these questions.

Recent Findings Machine learning techniques are now being incorporated into environmental mixture research to overcome issues with traditional methods. Though some methods perform well on specific tasks, no method consistently outperforms all others in complex mixture analyses, largely because different methods were developed to answer different research questions. We discuss four main questions in environmental mixtures research: (1) Are there specific exposure patterns in the study population? (2) Which are the toxic agents in the mixture? (3) Are mixture members acting synergistically? And, (4) what is the overall effect of the mixture?

Summary We emphasize the importance of robust methods and interpretable results over predictive accuracy. We encourage collaboration with computer scientists, data scientists, and biostatisticians in future mixture method development.

Keywords Environmental mixtures · Multi-pollutant · Dimension reduction · Variable selection · Bayesian statistics

Introduction

We are exposed daily to numerous environmental pollutants. Only a small proportion of these has been assessed for toxicity, with most studies conducted in experimental settings and not necessarily involving humans [1]. Furthermore, studies evaluating adverse health have traditionally conducted single-chemical analyses. This approach, however, does not represent reality; we are exposed to a mixture of chemicals at any given time, which can act synergistically or antagonistically. Furthermore, due to high correlations among many of these chemicals, we might detect associations between some of them and the outcome of interest due to their correlation with the actual "bad actor(s)," i.e., the actual toxic agent(s) in the mixture. Finally, testing a plethora of chemicals in single-pollutant models—i.e., multiple comparisons—dramatically increases the chances of spurious findings and, consequently, may increase disagreement across studies. For these reasons, the US Environmental Protection Agency, National Research Council (NRC), and National Institute of Environmental Health Sciences (NIEHS) have all recognized the necessity to assess exposure to mixtures [2–5].

Assessing exposures to mixtures, nonetheless, is especially challenging. First, the dimensionality of the data dramatically increases when one includes multiple chemicals in the statistical model. Many studies do not have the power to accommodate this need. Furthermore, high correlation among chemicals can lead to collinearity and subsequently inflated standard errors and unstable effect estimates. Two main issues stemming from current limitations in mixtures analyses have been identified: the need for (a) novel and robust statistical approaches to assess exposure to mixtures, and (b) appropriate use of available statistical methods in epidemiologic studies [5, 6].

Given the increasing need to incorporate complex high-dimensional data in environmental health studies, researchers have progressively turned toward machine learning (ML) methods. Adapting ML and data science methods can be especially advantageous, leading to more comprehensive studies of environmental exposure impacts on human health. Nonetheless, these methods were developed to serve a different purpose, mostly focusing on optimizing predictive accuracy, which is not necessarily well-aligned with Public and Environmental Health. Environmental health researchers, therefore, should be especially cautious when using such methods and preferably should work with computer and data scientists, in collaboration with biostatisticians, to best adapt and extend ML methods for appropriate use in environmental health.

The goal of this paper is not to give a comprehensive overview of all existing methods to analyze exposure to mixtures. Instead, we will discuss four types of scientific questions that are of interest in mixtures research. We will provide a conceptual description of analytic techniques appropriate to answer each question, along with examples in recent studies. No single method, to date, can adequately address all four types of mixtures-related scientific questions [5, 6]. Although other reviews exist on mixtures methods [7–10], here we emphasize the need for methods that ensure robust results while focusing on interpretability and inference. Although the specific research question(s) might differ across studies, the two aims of mixtures analyses are universal: to [1] better understand biological pathways of pathogenesis, and [2] informally efficient targeted interventions and policies to best protect the public and prevent disease. For both of these aims, it is of utmost importance to select robust methods that provide interpretable, and therefore actionable, results.

Complex Mixture Methods

To discuss complex mixture methods, one must first define what a mixture is. Although there is no strict definition, according to one NIEHS statement, "a mixture must have at least three independent chemicals or chemical groups" [11]. Generally, exposure to a mixture indicates exposure to multiple "stressors" simultaneously, which can include both chemical and non-chemical (e.g., socioeconomic status, diet, etc.) components. The question becomes how can we represent the complexity of reality in a statistical model?

The selected method(s) should be based on the primary research question. If the interest lies in identifying exposure patterns or groups of people with similar exposure profiles, some dimensionality reduction is required [12–15]. To identify the toxic agent(s) in a mixture, variable selection approaches may be more appropriate [16, 17]. If the aim is to evaluate synergistic or antagonistic effects, the main options

are to hard-code interactions into the health model or take advantage of more flexible semi- or non-parametric models [18••, 19, 20]. Finally, to observe the effect of the overall mixture, one may create a weighted index of exposure or compute the full posterior distribution using Bayesian methods [18••, 19, 21••].

We present the four main research questions most relevant for mixtures analyses in Table 1. In the next sections, we describe appropriate methods to address each of these questions and provide applied examples. Please note that many of the methods discussed may answer multiple questions and thus fall under multiple subsections. To avoid repetition, we present applications in detail under the research question to which they contribute most uniquely and mention them as appropriate when applicable to other sections.

Pattern or Profile Identification

Identification of exposure patterns in the population, e.g., due to common sources or behaviors, is highly desirable if the goal is to inform targeted interventions and regulations. Once common patterns are identified, they can be included as the exposures of interest in health models, resulting in subsequent identification of the most toxic sources/behaviors. Regulatory agencies, then, can act on certain sources, and interventions can be designed to target specific behaviors. Methods adopted from the pattern recognition field are powerful tools to help researchers identify these shared exposure patterns.

Questions about pattern or profile identification usually involve unsupervised techniques to describe the variability among correlated chemicals in fewer unobserved (i.e., latent) factors or to identify subgroups of individuals with similar exposure profiles (i.e., clusters). The solution of unsupervised approaches is obtained independently of any outcome(s) of interest. Both clustering and factor analysis involve dimensionality reduction of the original data. Clustering groups study analysis units (e.g., participants in a cohort study or days in a time-series), and factor analysis techniques group chemicals into factors using combinations of the mixture members within each factor, i.e., patterns. To be meaningful, the number of clusters or patterns should be substantially lower in dimension than the original data.

Table 1 The four main possible questions in mixture analyses

1. Are there specific patterns of exposure in the study population?
2. Which are the toxic agents in the mixture? Or, what are the independent effects of each mixture member on the health outcome of interest?
3. Are there synergistic effects or interactions among mixture members?
4. What is the overall effect of the mixture on the outcome of interest?

Clustering partitions observations (e.g., study participants) into distinct homogeneous groups so that observations within groups are similar and observations across groups are different. Clustering is often used in exploratory analyses, although the identified clusters can later be included in a health model as indicators. Though clustering is not particularly useful in estimating main effects, this approach can be advantageous when assessing effect modification by high-dimensional modifiers [22]. Although the results from clustering can be quite interpretable, there is no "golden rule" for choosing the number of clusters [23••], highlighting the importance of expert knowledge in interpretation.

It may be more appropriate in environmental mixtures analyses to identify exposure patterns as functions of all mixture members representing specific sources of exposure or common behaviors in the study population. Pattern identification requires expert knowledge to assign interpretable labels to the estimated patterns. Principal component analysis (PCA) [12] is the most commonly used dimension reduction technique employed in environmental epidemiology [24–27]. PCA aims to explain as much of the total variance in the data as possible using a smaller number of variables (called components), which are linear combinations of the original variables. The researcher must then decide the appropriate number of components to include in further analyses based on predefined criteria, by, e.g., having a priori defined a desired amount of the total variance explained. Although PCA is still widely used, its limitations include an orthogonal solution (which might be contrary to reality if the exposure patterns to be identified are not independent), no guarantee of an interpretable solution, and reliance on the researcher to decide on the number of components to retain for subsequent analyses.

While more advanced methods of matrix factorization exist [28], including positive matrix factorization (PMF) and sparse non-negative matrix underapproximation (SNMU), the structure of the results appears largely similar. PMF and SNMU are similar to traditional factor analysis in that the number of mixture components is designated by the researcher [15, 29, 30], but they both include constraints in the matrix factorization that enhance interpretability. First, the non-negativity constraint in both PMF and SNMU ensures that individual scores and variable loadings on factors are on the same range as the original variables [15, 31], as all environmental data are positive (e.g., chemical concentrations). The factors and individual exposures can be easily described—factors by the relative proportions of variables, and individual exposures by the relative proportions of factors. Second, both PMF and SNMU, unlike PCA, provide nonorthogonal results which can more realistically describe human exposure [15, 31]. Finally, SNMU adds a sparsity constraint on the solution by including a penalty term forcing the lowest contributing variables in the factor loadings to zero, ignoring chemicals that do not add to the mixture [30].

Traoré et al. implemented SNMU to identify mixtures of 210 environmental contaminants, including pesticide residues, trace elements, and minerals, in two cohorts of pregnant women in France [32]. The authors selected the optimal number of mixture components in terms of relevance and quality of interpretation, choosing eight [32]. They additionally applied hierarchical clustering to identify groups of women with similar co-exposure profiles [32], clustering participants based on the patterns identified by the SNMU.

Identification of Toxic Agents and Independent Effects

When interested in the identification of specific toxic agents within a mixture and the characterization of their exposure-response curves, the method of choice should help us estimate the independent effects of each mixture member. Any analysis, therefore, should incorporate information on the outcome of interest (i.e., supervised approaches).

Variable selection is one family of methods that may aid in identifying toxic agents by choosing a subset of relevant mixture members. The most traditional form is subset selection, including automated forward and backward selection and best subset selection [23••]. While these are easy to implement, they can be unstable, as small changes in the data can greatly affect variable inclusion in the model, and the uncertainty in the variable selection portion is ignored [33, 34], resulting in an increased type I error rate [35–37].

To address flaws in subset selection, penalized regression techniques can be used; these outperform traditional regression in their predictive capacity. Notably, penalized regression methods perform better in highly correlated settings, finding a unique solution even when the number of chemicals is larger than the number of observations [38]. These methods cannot, as no method can, determine causal agents in highly correlated mixtures, but they continue to predict well in these settings, where traditional regression would provide unstable effect estimates and inflated standard errors. By penalizing the magnitude of the coefficients, "unimportant" variables shrink toward zero, i.e., their estimated effects are restricted, allowing estimation of the coefficients that are more strongly associated with the outcome. This trades some bias in the estimated coefficients for lower variance and overall mean squared error (MSE) of the predicted outcome.

Multiple penalization forms exist. Ridge regression shrinks the sum of the squares of the coefficients, resulting in non-zero coefficients that are smaller than or equal to those that would have been obtained using traditional regression [39]. Lasso (least absolute shrinkage and selection operator) shrinks the sum of the absolute values of the coefficients, which pushes some coefficients to zero, yielding a sparse solution [16]. Elastic net includes both penalization terms [17].

The penalization term in each above-mentioned approach includes a tuning parameter (λ) between zero (making the model equivalent to traditional regression) and infinity (where all coefficients are shrunk toward zero) [23••, 40]. Usually, model fitting includes a training set and a validation set to choose λ, followed by a test set to estimate the true MSE of the model [41]. In environmental epidemiology, a test set may not be necessary and is often unavailable, but some form of hold-out or cross-validation analysis to justify the choice of λ is warranted.

Lasso has been more commonly used than ridge regression recently because it produces a sparse solution (i.e., the coefficients of some exposures will be estimated as exactly zero). One simulation study showed that lasso outperforms other penalized methods when there is a small to moderate number of moderate-sized true effects, while ridge regression performs better when there is a large number of small true effects [16]. However, this simulation was not performed in an environmental epidemiology setting and should be interpreted cautiously. When mixture members are highly correlated, ridge and elastic net will push coefficients toward each other [17, 39]; lasso will keep one of the correlated variables in the model and push the others to zero [16]. If multiple toxic agents in correlated mixtures are hypothesized, elastic net may provide the best balance of sparsity and inclusion of correlated variables that best predict the health outcome.

The coefficients for the selected variables are not necessarily the same as those that would have been obtained from traditional regression including only that subset. It is even possible for corresponding coefficients in the two models to be in different directions [16]. A large drawback for use of these penalization methods in environmental health is the difficulty in obtaining valid inferences, as the coefficients are non-linear and non-differentiable [16]. To overcome this, many researchers have first fit a penalized regression (e.g., Lasso) and subsequently included the selected variables in a traditional regression model. This practice is not well justified for inference, as it underestimates standard errors by ignoring uncertainty in the variable selection step.

Nwanji-Enwerem et al. used an adaptive lasso to select $PM_{2.5}$ constituents associated with DNA methylation age [42]. This approach incorporates user-specified weights to penalize individual coefficients differently so that constituents with larger effects are penalized less than those with smaller effects [43]. The mixture of interest in their analysis included five $PM_{2.5}$ constituents (organic and elemental carbon, sulfate, nitrate, and ammonium) that made up 89% of the total $PM_{2.5}$ mass concentration. With covariates fixed in the model so that only constituents could be penalized, sulfate and ammonium remained in the model, positively predicting Horvath DNA methylation age [42].

Additionally, Bayesian kernel machine regression (see "Interactions") models the independent exposure-response functions between all exposures and the outcome and can be used to identify independent effects and characterize the exposure-response relationship. Weighted quantile sum regression (see "Overall Mixture Effect") assigns weights to mixture components which are interpreted as variable importance factors; these can identify potentially toxic agents but fail to provide individual effect estimates.

Interactions

Identification of potentially synergistic effects among chemicals is essential if there is reason to believe that the combined health effect is greater (or less) than the sum of the independent effects. This is often hypothesized when studying chemicals that share stereochemical features or that target the same biological pathway. If regulatory action or interventions aim only to lower exposure to one chemical below a certain threshold, while this chemical works synergistically with another, then the necessary reduction will be underestimated among people exposed to both chemicals. Methods to assess interactions between chemicals can identify susceptible groups in those exposed to interacting chemicals simultaneously. Interactions can be hard-coded into models, including lasso and weighted quantile sum (WQS) regression (see "Overall Mixture Effect"). However, this practice requires a priori deciding which interaction terms to include and can only accommodate a small number of all potential high-order and non-linear interactions. To address this limitation, semi- or non-parametric methods are preferred.

Non-parametric methods make no assumptions about the functional form of the association, instead using tuning parameters to estimate a curve as closely as possible to each point without over-fitting [23••]. Such approaches can more accurately fit nonlinear exposure-response relationships and allow for non-additive interactions among all mixture members without explicitly including them in the model. Semi-parametric methods combine the flexibility of non-parametric models with a parametric portion which is computationally easier to estimate [40], allowing for the adjustment of potential confounders. However, such approaches often require a larger sample size than is typically needed for a parametric approach, since they do not reduce the problem of estimating the functional form of the data to a few parameters [23••].

Bayesian kernel machine regression (BKMR) is a semi-parametric technique that models the exposure-response relationship as a non-parametric kernel function of the mixture members, adjusting for covariates parametrically [18••, 19, 20]. The Gaussian kernel is commonly used for flexibly capturing a wide range of underlying functional forms, including non-additive interactions, without specifying the shape of the individual exposure-response curves or the existence of interactions among mixture members [18••, 44]. BKMR also

assesses independent effects, allows for component-wise or hierarchical variable selection, and estimates the overall effect of a mixture [18••, 19, 20], but we include it in this section due to its unique ability to detect nonlinear interactions.

Wasserman et al. used BKMR to estimate the joint effects of exposure to a mixture of five metals (arsenic, lead, manganese, cadmium, and selenium, measured cross-sectionally) and peri-natal arsenic on intellectual function in adolescents in Bangladesh [45]. While no interactions were observed, they found that increased arsenic and cadmium were associated with decreased raw full scale IQ, as was the overall mixture exposure [45].

Other methods to assess high-order and non-linear interactions include tree-based methods [40]. Regression and classification decision trees yield highly interpretable results, but they tend to be unstable, i.e., small changes in the data can cause large changes in the estimated trees. More complex tree-based methods, such as random forests, are more robust to variation and have improved prediction, but they lose the interpretability of the single tree [23••]. Several groups have begun to implement these methods in environmental mixtures [46–48].

Overall Mixture Effect

Characterizing the overall effect of combined chemical exposures is necessary to adequately define the total body burden of environmental mixtures. When exposure to individual compounds is below a set regulatory concentration or too low to show independent effects, an overall effect may still exist in combination with other exposures which target a common health endpoint. The NRC now recommends that risk assessment efforts account for cumulative risk associated with chemicals that affect the same health outcome [9, 49]. If no interaction is present, i.e., effects are believed to be additive, a composite of chemicals or a weighted index allows for the estimation of the combined effects of individual compounds without reducing the unique exposures to a simple sum.

Various methods exist to create a weighted score of exposure prior to the modeling step. Toxic equivalency factors (TEF), for example, are often used with dioxins and dioxin-like chemicals to weigh their toxicity in terms of the most toxic dioxin. Individual weights are determined by structural and binding similarities, ability to elicit a toxic response, persistence, and bio-magnification. A single number—a toxic equivalency (TEQ) score—is estimated as the sum of the products of each chemical's concentration and its individual TEF value, and can be used as a cumulative measure of exposure to these related chemicals [50, 51]. Use of TEQ, however, is limited to chemicals whose main mechanism of action is shared with dioxin. Creating such indices, therefore, for other mixtures can be challenging, especially if such prior knowledge is not available.

When less is known a priori about the individual toxicity of the mixture members, WQS regression creates an empirically weighted index which can be more widely implemented for any mixture. The estimated coefficient of this index is interpreted as the mixture effect [21••]. As the name implies, WQS categorizes the continuous exposures into quantiles to reduce the impact of outliers and ensure that all exposure variables are on the same scale [21••, 52], but this also reduces the amount of information in the data. WQS is analogous to the variable selection methods discussed in "Identification of Toxic Agents and Independent Effects", with each variable's penalization determined by its respective weight. WQS then assigns a single coefficient to the weighted index—the sum of the concentration quantiles of each member multiplied by its weight. The weights identify toxic agents and "zero out" chemicals with negligible associations [21••, 53]. If the index coefficient is statistically significant, important components of the index (i.e., toxic agents) can be identified as those with the highest weights [21••]. The weights provide information on the relative importance of individual mixture members but no corresponding effect estimates.

White et al. used WQS to estimate the overall effect of a mixture of ten metals (antimony, arsenic, cadmium, chromium, cobalt, lead, manganese, mercury, nickel, selenium) on breast cancer risk [54]. The WQS index was positively associated with postmenopausal breast cancer but not with overall or ER+ breast cancer. Cadmium, lead, and mercury had the largest weights in the post-menopausal breast cancer index [54].

Bayesian methods, such as BKMR (see "Interactions"), can also estimate the overall effect of the mixture by modeling the entire multi-dimensional posterior distribution.

Bayesian Methods

Though challenges still remain, recent advances in computational performance and scalability [55] have opened the door to Bayesian methods in environmental epidemiology. Bayesian methods explicitly use probability to quantify uncertainty in inference, i.e., there is (in principle) no impediment to fitting models with many parameters, correlated exposure variables, or complicated exposure-response specifications [56••], and these methods may be used to answer multiple mixtures questions in the same analysis. Given the flexibility of Bayesian methods, they are a promising direction for new development.

Bayesian methods estimate the full posterior distribution of the unobserved quantities [56••], meaning that all Bayesian models can estimate an overall effect. Additionally, inclusion of prior information—a hallmark of Bayesian data analysis—becomes a powerful tool in environmental mixture methods. Prior knowledge of effect estimates (magnitude or direction

taken from expert knowledge or previous research) or chemical groupings (by exposure source, biological pathway, or shared toxicity) can be explicitly incorporated in the model.

BKMR, for example, assesses potentially non-linear independent effects and the overall effect in addition to interactions among mixture members. It also allows for hierarchical grouping of mixture members [18••, 19, 20]. Other examples of Bayesian methods in environmental mixtures exist, as well. Bayesian hierarchical methods [57–59], Bayesian model averaging [60–64], Bayesian additive regression trees [65–67], Bayesian profile regression [10, 68, 69], and semi-Bayesian methods (which provide faster results) [70–72] have been implemented in environmental mixtures research, but they are not yet widely used. Computational advances in processing speed coupled with developments in ML and biostatistical modeling can make these methods accessible to environmental epidemiologists. There is space and need for more method development, in collaboration with data and computer scientists and biostatisticians, in our field.

Discussion

Although multiple methods currently exist for environmental mixture research, no method can answer all mixture questions, highlighting the importance of a well-defined research question to guide method selection. The interpretability of results (over predictive accuracy) is critical in determining the usefulness of novel statistical, data science, or ML methods in environmental epidemiology. Despite statistical advances, all methods share certain limitations. Given high correlations across chemicals and varying measurement error in species-specific concentrations, any statistical method will pick the chemical with the least amount of measurement error that either is the toxic agent or is correlated with the toxic agent (but measured with less error) [73, 74]. Furthermore, if exposure biomarkers are used (e.g., chemicals or metabolites measured in biosamples), their half-lives and the timing of sample collection with respect to exposure matter. Depending on the chemicals' half-life and the critical window of exposure, all approaches are susceptible to selecting a chemical whose concentration was high during the critical exposure window and remained high during sampling; exposure to this selected chemical likely co-occurred with exposure to the actual toxic agent that—if it has a short half-life—might be undetected at sampling or measured with excess noise depending on the varying time between the critical exposure window and sampling across subjects. It is also conceivable that the actual toxic agent is not included in the mixture to be analyzed. Focusing, therefore, on identifying the toxic agent(s) might lead to the wrong conclusion under such scenarios, regardless of the choice (and performance) of method.

In mixture analyses, the researchers define the mixture to investigate, which is most often not the full true mixture to which the study population is exposed. The above-mentioned issues, then, may be amplified when the examined mixture is small (relative to the size of the true mixture), due to residual confounding from unmeasured chemicals or shared sources. Caution should also be applied when using the terms "overall" or "cumulative" for such small mixtures, as these are usually only a subset of the actual mixture of interest. The complexity of environmental mixtures—chemical and non-chemical—and analytical limitations for measurement of chemicals add to the difficulty of arriving at a perfectly specified model. Including correlated exposure variables in any model may amplify rather than reduce confounding bias [75]. These methods, additionally, present challenges in power estimation, but simulations can be used to calculate power. Simulations require certain assumptions about the data structure, as do power calculations for traditional regression. Finally, uncertainty propagation is an often-overlooked concern, mostly of unsupervised methods. Many researchers simply include PCA scores or cluster membership in health models ignoring the uncertainty inherent in the solution selection, often based on implicit assumptions. Propagation of uncertainty will lead to more valid inferences and can result in fewer spurious results and more consistent findings across methods and studies [25].

Although in this paper, we did not discuss study designs, most of the discussed methods can accommodate outcome distributions beyond normal, and can be used in multiple study designs, for example—but not limited to—longitudinal or time-to-event analyses. We note, however, that certain designs might introduce additional challenges. For instance, assessing exposure to a mixture that varies over time can be challenging, especially if the different chemicals in the mixture induce toxicity at different time points.

In future mixture analyses and method development, researchers should focus on robustness of findings. Different populations experience different exposure mixtures and different distributions of potential modifiers, so we should not expect to replicate results (patterns or effect estimates) across populations. Rather, unstable methods should be avoided, and multiple methods should be used, whenever possible, to address a research question. When investigating an overall effect using WQS, for example, BKMR may be used as sensitivity analysis. Care should be taken, however, when employing different methods—if a specific research question is not stated, different methods may provide results that appear conflicting. For methods that employ simulations or rely on user-specified prior information (i.e., Bayesian methods), internal assessment of reproducibility is also warranted.

These limitations and model-specific assumptions should be carefully considered when interpreting results of mixtures analyses. Additionally, groups developing mixtures methods should consider extensions that take this information into

account when estimating health effects. Furthermore, combining methods may be of interest, for example, coupling factor analysis with BKMR if one is interested in assessing the exposure-response of exposure patterns and their potentially non-linear interactions.

Bayesian approaches, furthermore, inherently accommodate supervised pattern recognition, fully propagating uncertainty in the health model, thus identifying patterns specific to each outcome and better characterizing biological pathways. New Bayesian (and semi-Bayesian) methods could further allow more flexible modeling, explicit incorporation of uncertainty, inclusion of prior knowledge, and the ability to answer multiple questions simultaneously. Method development should involve direct collaboration with computer scientists, data scientists, and biostatisticians to take advantage of computationally efficient ML algorithms and to obtain interpretable results from sophisticated models. Complex ML prediction methods generate enthusiasm across disciplines, but if their results are not directly interpretable in health effects analyses, they are unlikely to benefit the ultimate research goals of understanding biological pathways and informing regulatory action.

Conclusion

With careful incorporation of ML and data science methods, environmental epidemiologists are better able to explore complex relationships between environmental mixtures and adverse health. While each new prediction method appears to improve upon previous methods, effect estimation rather than outcome prediction should be the desired result. To this end, environmental epidemiologists must work with experts outside of our field to better adapt ML methods to our goals, instead of simply employing methods as they come. As method development for environmental mixtures continues, we recommend Bayesian methods for their flexibility and interpretability of their results. Although no single model to date can answer all mixtures questions, a well-defined research question will point toward the correct approach—whether identification of patterns or independent, synergistic, or overall effect(s). Results are only useful, no matter how sophisticated the method, if they are robust, reproducible, interpretable, and, finally, actionable.

Compliance with Ethical Standards

Conflict of Interest Elizabeth A. Gibson, Jeff Goldsmith, and Marianthi-Anna Kioumourtzoglou declare that they have received grants from the NIEHS during this study.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
•• Of major importance

1. Grandjean P, Landrigan PJ. Developmental neurotoxicity of industrial chemicals. Lancet. 2006;368(9553):2167–78.
2. U.S. EPA. Air, climate, and energy: strategic research action plan 2012–2016. Office of Research and Development, Available at: https://www.epa.gov/sites/production/files/2014-06/documents/strap-ace2012.pdf, 2012.
3. NRC (National Research Council). Air Quality Management in the United States. Washington, DC: National Academies Press; 2004.
4. NIEHS. Strategic plan 2012–2017—advancing science, Improving Health: A Plan for Envi-ronmental Health Research. US Department of Health and Human Services, National Insti-tutes of Health, Available at: https://www.niehs.nih.gov/health/materials/niehs_strategic_plan_20122017_frontiers_in_environmental_health_sciences_booklet_508.pdf, 2012.
5. NIEHS Statistical Approaches for Assessing Health Effects of Environmental Chemical Mixtures in Epidemiology Studies, Available at: http://www.niehs.nih.gov/about/events/pastmtg/2015/statistical/, July 2015.
6. Taylor KW, Joubert BR, Braun JM, Dilworth C, Gennings C, Hauser R, et al. Statistical approaches for assessing health effects of environmental chemical mixtures in epidemiology: lessons from an innovative workshop. Environ Health Perspect. 2016;124(12):A227.
7. Hamra GB, Buckley JP. Environmental exposure mixtures: questions and methods to address them. Curr Epidemiol Rep. 2018;5(2):160–5.
8. Stafoggia M, Breitner S, Hampel R, Basagaña X. Statistical approaches to address multi-pollutant mixtures and multiple exposures: the state of the science. Curr Environ Health Rep. 2017;4(4):481–90.
9. Huang H, Wang A, Morello-Frosch R, Lam J, Sirota M, Padula A, et al. Cumulative risk and impact modeling on environmental chemical and social stressors. Curr Environ Health Rep. 2018;5(1):88–99.
10. Coker E, Liverani S, Su JG, Molitor J. Multi-pollutant modeling through examination of susceptible subpopulations using profile regression. Curr Environ Health Rep. 2018;5(1):59–69.
11. NIEHS. Powering Research through Innovative Methods for mixtures in Epidemiology (PRIME) (R01), RFA-ES-17-001, Available at: https://grants.nih.gov/grants/guide/rfa-files/RFA-ES-17-001.html, 2017.
12. Jolliffe I. Principal component analysis. Springer-Verlag, New York, 2002.
13. Jolliffe IT. Principal component analysis and factor analysis. Principal component analysis, pages 150–166, Springer-Verlag, New York, 2002.
14. Thompson B. Exploratory and confirmatory factor analysis: understanding concepts and applications. Applied Psychological Measurement. 2007;31(3):245-8.
15. Paatero P, Tapper U. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. Environmetrics. 1994;5(2):111–26.

16. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B Methodol 1996; 58(1)267–288.

17. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B Stat Methodol. 2005;67(2):301–20.

18.•• Bobb JF, Valeri L, Henn BC, Christiani DC, Wright RO, Mazumdar M, et al. Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. Biostatistics. 16(3): 493–508, 2014 **BKMR was developed specifically for environmental mixtures by including kernel machine regression, a machine learning technique, in a Bayesian model.**

19. Coull BA, Bobb JF, Wellenius GA, Kioumourtzoglou M-A, Mittleman MA, Koutrakis P, et al. Development of Statistical Methods for Multipollutant Research; Part 1. Statistical Learning Methods for the Effects of Multiple Air Pollution Con-stituents, volume 183. Boston: Health Effects Institute; 2015.

20. Bobb JF, Henn BC, Valeri L, Coull BA. Statistical software for analyzing the health effects of multiple concurrent exposures via bayesian kernel machine regression. Environ Health. 2018;17(1): 67.

21.•• Carrico C, Gennings C, Wheeler DC, Factor-Litvak P. Characteriza-tion of weighted quantile sum regression for highly correlated data in a risk analysis setting. J Agric Biol Environ Stat. 2015;20(1):100–20 **WQS was developed specifically for environmental mixtures using a machine learning optimization algorithm.**

22. Kioumourtzoglou M-A, Austin E, Koutrakis P, Dominici F, Schwartz J, Zanobetti A. $PM_{2.5}$ and survival among older adults: effect modifica-tion by particulate composition. Epidemiology. 2015;26(3):321–7.

23.•• James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. New York: Springer; 2013. **This book provides guidance on how to implement statistical and machine learning methods without requiring a background in statistics or computer science. The authors give practical explanations of available methods and when to use them, including R code.**

24. Pang Y, Peng RD, Jones MR, Francesconi KA, Goessler W, Howard B-b V, et al. Metal mixtures in urban and rural populations in the US: the multi-ethnic study of atherosclerosis and the strong heart study. Environ Res. 2016;147:356–64.

25. Kioumourtzoglou M-A, Coull BA, Dominici F, Koutrakis P, Schwartz J, Suh H. The impact of source contribution uncertainty on the effects of source-specific $PM_{2.5}$ on hospital admissions: a case study in Boston, MA. Journal of expo-sure science and environmental. Epidemiology. 2014;24(4):365–71.

26. Robinson O, Tamayo I, De Castro M, Valentin A, Giorgis-Allemand L, Krog NH, et al. The urban exposome during pregnancy and its socioeconomic determi-nants. Environ Health Perspect. 2018;126(7):077005.

27. Manzano-León N, Serrano-Lomelin J, Sánchez BN, Quintana-Belmares R, Vega E, Vázquez-López I, et al. Tnf α and il-6 responses to particulate matter in vitro: variation according to PM size, season, and polycyclic aromatic hydrocarbon and soil content. *Environ Health Perspect*. 2015;124(4):406–12.

28. Emmanuel J, Candès XL, Ma Y, Wright J. Robust principal component analysis? *J ACM*. 2011;58(3):11.

29. Gillis N, Glineur F. Using underapproximations for sparse nonnegative matrix factorization. Pattern Recogn. 2010;43(4):1676–87.

30. Gillis N, Plemmons RJ. Sparse nonnegative matrix underapproximation and its application to hyperspectral image analysis. Linear Algebra Appl. 2013;438(10):3991–4007.

31. Lee DD, Seung HS. Algorithms for non-negative matrix factorization. In Advances in neural information processing systems, 2001;556–562.

32. Traoré T, Forhan A, Sirot V, Kadawathagedara M, Heude B, Hulin M, et al. To which mixtures are french pregnant women mainly exposed? a combination of the second french total diet study with the eden and elfe cohort studies. Food Chem Toxicol. 2018;111: 310–28.

33. Shen X, Ye J. Adaptive model selection. J Am Stat Assoc. 2002;97(457):210–21.

34. Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Stat Assoc. 2001;96(456):1348–60.

35. Leamer EE. Specification searches: ad hoc inference with nonexperimental data, John Wiley & Sons Incorporated, 1978.

36. Raftery AE. Approximate bayes factors and accounting for model uncertainty in gener-alised linear models. Biometrika. 1996;83(2): 251–66.

37. Draper D. Assessment and propagation of model uncertainty. J R Stat Soc Ser B Methodol 1995;45–97.

38. Fan J, Li R. Statistical challenges with high dimensionality: feature selection in knowledge discovery. arXiv preprint math/0602133 2006

39. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogo-nal problems. Technometrics. 1970;12(1):55–67.

40. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning, volume 1. New York: Springer series in Statistics; 2001.

41. Hal Daumé III. A course in machine learning. 2012;5–73.

42. Nwanaji-Enwerem JC, Dai L, Colicino E, Oulhote Y, Di Q, Kloog I, et al. Associations between long-term exposure to $PM_{2.5}$ component species and blood DNA methylation age in the elderly: the VA normative aging study. Environ Int. 2017;102:57–65.

43. Zou H. The adaptive lasso and its oracle properties. J Am Stat Assoc. 2006;101(476):1418–29.

44. Liu D, Lin X, Ghosh D. Semiparametric regression of multidimensional genetic pathway data: least-squares kernel machines and linear mixed models. Biometrics. 2007;63(4):1079–88.

45. Wasserman GA, Liu X, Parvez F, Chen Y, Factor-Litvak P, LoIacono NJ, et al. A cross-sectional study of water arsenic exposure and intellectual function in adolescence in Arai-Hazar, Bangladesh. Environ Int. 2018;118:304–13.

46. Stingone JA, Pandey OP, Claudio L, Pandey G. Using machine learning to identify air pollution exposure profiles associated with early cognitive skills among us children. Environ Pollut. 2017;230: 730–40.

47. Ouidir M, Lepeule J, Siroux V, Malherbe L, Meleux F, Rivière E, Launay L, Zaros C, Cheminat M, Charles M-A, et al. Is atmospheric pollution exposure during pregnancy associated with individual and contex-tual characteristics? a nationwide study in FranceJ Epidemiol Community Health, pages jech–2016, 2017.

48. Gass K, Klein M, Chang HH, Flanders WD, Strickland MJ. Classification and regression trees for epidemiologic research: an air pollution example. Environ Health. 13(1):17, 2014.

49. National Research Council. Phthalates and cumulative risk assessment: the tasks ahead. National Academies Press, 2009.

50. Van den Berg M, Birnbaum LS, Denison M, De Vito M, Farland W, Feeley M, et al. The 2005 world health organization reevaluation of human and mammalian toxic equivalency fac-tors for dioxins and dioxin-like compounds. Toxicol Sci. 2006;93(2):223–41.

51. Mitro SD, Birnbaum LS, Needham BL, Zota AR. Cross-sectional associations between exposure to persistent organic pollutants and leukocyte telomere length among us adults in nhanes, 2001–2002. Environ Health Perspect. 2015;124(5):651–8.

52. Gennings C, Carrico C, Factor-Litvak P, Krigbaum N, Cirillo PM, Cohn BA. A cohort study evaluation of maternal pcb exposure related to time to pregnancy in daughters. Environ Health. 2013;12(1):66.

53. Yorita Christensen KL, Carrico CK, Sanyal AJ, Gennings C. Multiple classes of environmental chemicals are associated with liver disease: Nhanes 2003–2004. Int J Hyg Environ Health. 2013;216(6):703–9.

54. White AJ, O'Brien KM, Niehoff NM, Carroll R, Sandler DP. Metallic air pollutants and breast cancer risk in a nationwide cohort study. Epidemiology 20181;30(1):20-8..

55. Hoffman MD, Blei DM, Wang C, Paisley J. Stochastic variational inference. J Mach Learn Res. 2013;14(1):1303–47.

56.•• Gelman A, Stern HS, Carlin JB, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. Chapman and Hall/CRC, 2013. **This book is widely considered the leading text on Bayesian methods, with an accessible, applied approach to data analysis. The authors introduce basic concepts from a data-analytic perspective before presenting advanced methods.**

57. MacLehose RF, Dunson DB, Herring AH, Hoppin JA. Bayesian methods for highly correlated exposure data. Epidemiology 2007;18(2):199–207.

58. MacLehose RF, Hamra GB. Applications of bayesian methods to epidemi-ologic research. Curr Epidemiol Rep. 2014;1(3):103–9.

59. Furlong MA, Herring A, Buckley JP, Goldman BD, Daniels JL, Engel LS, et al. Prenatal exposure to organophosphorus pesticides and childhood neurodevelopmental phenotypes. Environ Res. 2017;158:737–47.

60. Fragoso TM, Bertoli W, Louzada F. Bayesian model averaging: a systematic review and conceptual classification. Int Stat Rev. 2018;86(1):1–28.

61. Wilson A, Zigler CM, Patel CJ, Dominici F. Model-averaged confounder adjustment for estimating multivariate exposure effects with linear regression. Biometrics, 2018;74(3):1034-44.

62. Berger K, Eskenazi B, Balmes J, Holland N, Calafat AM, Harley KG. Associations between prenatal maternal urinary concentrations of personal care product chemical biomarkers and childhood respiratory and allergic outcomes in the CHAMACOS study. Environ Int. 2018;121:538–49.

63. Berger K, Eskenazi B, Balmes J, Kogut K, Holland N, Calafat AM, Harley KG. Prenatal high molecular weight phthalates and bisphenol a, and childhood respiratory and allergic outcomes. Pediatr Allergy Immunol, 2019;30(1):36-46.

64. Berger K, Gunier RB, Chevrier J, Calafat AM, Ye X, Eskenazi B, et al. Associations of maternal exposure to triclosan, parabens, and other phenols with prenatal maternal and neonatal thyroid hormone levels. Environ Res. 2018;165:379–86.

65. Park SK, Tao Y, Meeker JD, Harlow SD, Mukherjee B. Envirsonmental risk score as a new tool to examine multi-pollutants in epidemiologic research: an example from the nhanes study using serum lipid levels. PLoS One. 2014;9(6):e98632.

66. Chipman HA, George EI, McCulloch RE, et al. Bart: Bayesian additive regression trees. Ann Appl Stat. 2010;4(1):266–98.

67. Ko Y-A, Mukherjee B, Smith JA, Kardia SL, Allison M, Diez AVR. Classification and clustering methods for multiple environmental factors in gene-environment interaction: application to the multi-ethnic study of atherosclerosis. Epidemiology. 2016;27(6):870–8.

68. Coker E, Gunier R, Bradman A, Harley K, Kogut K, Molitor J, et al. Association between pesticide profiles used on agricultural fields near maternal residences during pregnancy and iq at age 7 years. Int J Environ Res Public Health. 2017;14(5):506.

69. Molitor J, Papathomas M, Jerrett M, Richardson S. Bayesian profile regression with an application to the national survey of children's health. Biostatistics. 2010;11(3):484–98.

70. Kioumourtzoglou M-A, Zanobetti A, Schwartz JD, Coull BA, Dominici F, Suh HH. The effect of primary organic particles on emer-gency hospital admissions among the elderly in 3 us cities. Environ Health. 2013;12(1):68.

71. Kalkbrenner AE, Daniels JL, Chen J-C, Poole C, Emch M, Morrissey J. Perinatal exposure to hazardous air pollutants and autism spectrum dis-orders at age 8. Epidemiology. 2010;21(5):631–41.

72. Momoli F, Abrahamowicz M, Parent M-E, Krewski D, Siemiaty-cki J. Analysis of multiple exposures: an empirical comparison of results from conventional and semi-bayes modeling strategies. Epidemiology. 2010;21(1):144–51.

73. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement error in nonlinear models: a modern perspective. Chapman and Hall/CRC, 2006.

74. Pollack AZ, Perkins NJ, Mumford SL, Ye A, Schisterman EF. Correlated biomarker measurement error: an important threat to inference in environmental epidemiology. Am J Epidemiol. 2012;177(1):84–92.

75. Weisskopf MG, Seals RM, Webster TF. Bias amplification in epidemiologic analysis of exposure to mixtures. Environ Health Perspect. 2018;47003:1.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.