Manuscript title

Author names

Affiliations

Contact information

**Conflicts of interest:** The authors declare they have no conflicts of interest related to this work to disclose.

**Abstract**

Background:

Objectives:

Methods:

Results:

Discussion:

**Introduction**

First citation test1.

Epidemiological studies on the health effects of exposure to chemical mixtures have gained increasing attention in recent years. The lived environment contains numerous chemicals, often in mixtures originating from common sources, which together contribute to the progression of adverse health outcomes. Exposure mixture studies aim to measure the joint health effect of these multiple exposures. Such studies can have more direct implications for environmental regulations than studies of single chemicals, as regulation often targets pollution sources responsible for emitting mixtures of chemicals. In recognition of this, the National Academies of Science, Engineering, and Medicine has advocated for a multipollutant regulatory approach, which is likely to be more protective of human health.

However, expanding the focus of analysis from one exposure to multiple exposures introduces unique statistical challenges. In addition to a common issue of small effect sizes and small sample sizes present in most exposure analyses, exposure mixture analyses must also contend with high-dimensionality, near-collinearity, non-linear effects, and non-additive interactions. In the past few years, a wide variety of statistical methods have been developed to overcome these challenges, which have been accompanied by a host of comparative simulation studies for general mixture scenarios. However, to our knowledge, there has yet to be a simulation study which provides guidance about using recently developed methods to conduct inference on non-additive interactions between exposures, particularly when the nature and effect sizes of these interactions vary. Moreover, there is no guidance in the literature on assessing interactions between covariates and exposures in exposure mixtures.

There is a need to better understand the capacity of current methods for detecting non-additive interactions, which have important implications for our understanding of exposure mixtures. From a risk assessment perspective, if non-additive interactions are not accounted for, they can lead to under- or over-estimation of the true effect of a mixture for certain groups in the study population. From a mechanistic perspective, a non-additive statistical interaction between two chemical exposures can suggest that these compounds may be functionally interacting with each other. This can occur through direct reactions between chemicals or through interactions with enzymes. The discovery of a statistical interaction can be followed up by a functional study to assess the underlying biology of the interaction. Finally, from a public health perspective, exposure mixtures may interact with other covariates, signaling that social and health factors might be mediating the relationship between the health outcome and components of the exposure mixture. In these cases, researchers should first consider the independent, additive association between the covariate and levels of exposure or rates of a health outcome, before determining the meaning of a potential interaction term. These conclusions can be relevant to policy makers, as the potential benefit of regulating an exposure might differ across groups.

In this paper, we conduct a simulation study comparing the performance of recently developed regression techniques -- Bayesian kernel machine regression (BKMR), Bayesian semiparametric regression (BSR), and quantile g-computation (QGC) -- for quantifying non-additive interactions between multiple environmental exposures and related covariates. We consider [xx]

We then apply these methods to a real-world data example. [xx]

Our aim is to benchmark the ability of current methods and provide guidance about analytical approaches for detecting non-additive interactions.

**Models**

**Bayesian kernel machine regression**

BKMR estimates the exposure-response relationship using a Gaussian kernel function in a Bayesian framework. As the kernel does not require assuming a functional form for the exposure-response relationship, BKMR can flexibly capture a wide range of relationships, including nonlinear exposure-response relationships and non-additive interactions between exposures. BKMR can perform either component-wise or hierarchical variable selection. Here, we employ component-wise selection, wherein each exposure is assigned a weight with a prior distribution. The weights control the degree to which each associated exposure contributes to the model. The posterior means of weights generated from the model fitting process represent each exposure’s posterior probability of inclusion (PIP).

The flexibility of BKMR’s model specification means that there is no direct measure for conducting inference on the presence of interactions. Currently, the most common approach for detecting interactions is through a qualitative assessment of visual diagnostic plots. A slightly more formal inferential approach for two-way interactions involves using summary statistics. Under this approach, researchers can generate a confidence interval of the difference in association between the response and one exposure at two quantiles, for example 0.25 and 0.75, of a second exposure. If the interval does not contain 0, then there is evidence of an interaction. The choice of quantiles is important, as the shape of the regression surface between the response and two exposures can influence the magnitude of the difference between the

Assessing interactions between exposures and covariates requires stratifying.

**Bayesian semiparametric regression**

BSR estimates the exposure-response relationship using natural spline regression in a Bayesian framework.

BSR explicitly incorporates interaction terms in its model formulation as a spline basis expansion of the product term. As a result, the model fitting process assigns PIPs for all combinations of interactions between exposures, which allows for a formal framework of inference for interactions.

**Quantile g-computation**

QGC

Package: GQCOMPINT

**Simulation example**

**Methods**

*NHANES data.* In order to make our simulations comparable to real-world exposure mixture studies, we based our simulation data on the XX. [Discuss the NHANES dataset.]

We followed the approach by XX for preparing the data for analysis.

*Simulating predictor data.* We simulated exposure and covariate data using a multivariate *t*-copula.

*Simulating predictor-response relationships.* Health outcome responses were simulated under several different scenarios, each of which included different effect sizes and functional forms for the interactions.

*Fitting models.* We ran four methods on our simulated datasets.

*Model assessment.* We assessed model performance based on detection of significant univariate chemicals as well as detection of interactions.

**Results**

X

**Application example**

**Methods**

X

**Results**

X

**Discussion**

X

**Conclusions**

X

**References**

1. National Academies of Sciences, Engineering, and Medicine, Division on Earth and Life Studies, Board on Environmental Studies and Toxicology, Committee on Incorporating 21st Century Science into Risk-Based Evaluations. *Using 21st Century Science to Improve Risk-Related Evaluations*. National Academies Press; 2017:24635. doi:10.17226/24635