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 limited 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 non-additive interactions with different effect sizes and functional forms, two- and three-way interactions, and scenarios with differing sample sizes. 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 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. We use a slightly more formal inferential approach for two-way interactions involves using summary statistics. Under this approach, the association between the response and an exposure is estimated as the difference in estimated response values at two quantiles of the exposure. Then, the interaction is assessed by generating a confidence interval of the difference in the estimated association between the response and one exposure at two quantiles 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 estimated exposure-response association.

In order to assess interactions between exposures and covariates in this paper, we take the approach of stratifying BKMR models based on values of the covariate. Confidence intervals for the estimated exposure-response association are constructed for each stratified model, and then compared to determine the significance of an interaction between an exposure and a covariate.

**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 between exposures. This spline regression framework makes BSR slightly more parameterized than BKMR, while also maintaining flexibility to capture nonlinear exposure-response relationships and non-additive interactions between exposures. Each term is assigned an indicator that determines whether the term is included in the model. The posterior means of these indicators generated from the model fitting process represent each term’s PIP. As interactions are incorporated as explicit terms in the model formulation, the model fitting process assigns PIPs for all combinations of interactions between exposures, which allows for a formal framework of inference.

We assess interactions between exposures and covariates in a similar manner as described for BKMR, using BSR models stratified based on values of the covariate.

**Quantile g-computation**

QGC

As QGC does not provide a framework for inference on interactions between exposures, we do not assess interactions between exposures in this paper. However, QGC can assess interactions between exposures and covariates.

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 2001-2002 National Health and Nutrition Examination Survey (NHANES) data, originally used in a study by Mitro et al which assessed the association between leukocyte telomere length (LTL) and exposure to a mixture of persistent organic chemicals (POP).

We generally followed the approach used by Gibson et al., who used this dataset for a comparison of methods for assessing general mixture scenarios, to prepare the data for analysis. Briefly, we retained a total of 18 congeners in our final dataset, including eight non-dioxin-like poly-chlorinated biphenyls (PCBs), two non-ortho-substituted PCBs, one mono-ortho-substituted PCB, three chlorinated dibenzo-p-dioxins, and four dibenzo-furans, all of which were lipid-adjusted. Observations of congeners with values below the limits of detection (LOD) were adjusted using the sample-specific LOD divided by the square root of two. A total of X covariates were included in analyses: age, sex, race/ethnicity (non-Hispanic white, non-Hispanic black, Mexican American, other Hispanic, and other non-Hispanic), education attainment (less than high school, high school graduate, some college, college or more), BMI, serum cotinine, and blood cell count and distribution (white blood cell count, percent lymphocytes, percent monocytes, percent neutrophils, percent eosinophils, and percent basophils). After excluding observations with missing data, a total of 1,003 observations were retained in the final dataset.

*Simulating predictor data.*

We simulated exposure and covariate data using a multivariate copula fit on the 1,003 observations in the NHANES dataset. We used copulas as they can preserve both the correlation structure and marginal distributions form the observed data, allowing us to replicate conditions in a real-world scenario. Copulas require continuous variables with uniform distributions. We transformed continuous exposures and covariates to uniform distributions based on their empirical marginal cumulative distribution functions (CDFs). For sex, the only binary covariate, we used the checkerboard copula approach, which involved “jittering” the binary values with uniform random noise. There is currently no widely accepted approach for fitting copulas from unordered categorical variables with more than two levels. Thus, we excluded race and ethnicity and educational attainment from the copula model.

We performed model selection to identify the copula that best approximates the dependence structure of the observe data. We fit the set of multivariate copulas used by Lazarevic et al. in their simulation study, which included the Gaussian, *t*, Gumbel, Frank, Clayton, and Joe copulas. We fit two *t* copulas with 4 and 10 degrees of freedom, a *t* copula where the degrees of freedom were determined during the fitting process, and two versions each of the Gumbel, Frank, Clayton, and Joes copulas with θ = {2, 4}. Among these, we chose the *t* copula with degrees of freedom determined during the fitting process, as it maximized the likelihood of the copula. We simulated predictor data by randomly sampling from the fitted multivariate *t* copula distributions and transforming the resulting uniform variables to their original distributions using empirical marginal CDFs. We simulated the race and ethnicity and educational attainment variables by randomly assigning values to observations based on proportions in the observed dataset.

We generated one set of simulated datasets with the original sample size of 1,003, and another set of datasets with a smaller sample size of 250, which is typical in many cohort studies. We verified that the original structure of the observed dataset was preserved by visually comparing marginal distributions of exposures and covariates, as well as the correlation structure using Spearman’s rho.

*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. All scenarios were run for both the smaller (n=250) and larger (n=1003) sample sizes.

In the first scenario, we specified a “base case” model with no interactions:

where *ε* ~ *N*(0, 5), i.i.d. This model includes a linear term for XX, two S-shaped logistic terms for X and X with varying effect sizes, and a symmetric inverse U-shaped quadratic term for X. Moreover, we included covariate terms as linear effects in the model. We chose the standard deviation on the normal random error term in order to achieve an R^2 of around 0.1-0.3 in a multiple linear regression that included only the true functional form of the significant chemicals. This R^2 range approximates realistic signal-to-noise ratios in exposure mixture studies.

In subsequent scenarios, we added an additional interaction term to the base case model. First, we considered four cases of interactions between two or three exposures: a two-way interaction between exposures that are univariately significant, a two-way interaction between exposures that are univariately insignificant, a two-way interaction between exposures that are moderately correlated, and a three-way interaction. For each case, we considered two functional forms – multiplicative and polynomial – and a lower and higher effect size, which we set by defining the weight on the interaction term in the model. The higher effect sizes were selected in order to achieve a power of approximately 0.5 at an alpha level of 0.05 in the smaller sample size (n=250) case, using a multiple linear regression with the true functional form of the chemicals specified and the covariate terms included. The weights on the lower effect sizes were set equal to half of the higher effect sizes. Table XX shows the specification of interaction terms.

Next, we considered interactions between the race an ethnicity covariate and an exposures. We increased the coefficient of XX in XX individuals (n=XX in the original NHANES dataset) for the first scenario, and in XX individuals (n=XX in the original NHANES dataset) for the second scenario. For each scenario, we specified a lower effect size by increasing the coefficient on XX by 1.5 (i.e., from 1\*XX to 1.5\*XX) in the target group, and a higher effect size by increasing the coefficient on XX by 2.

This resulted in a total of 42 scenarios. For each scenario, we generated 100 simulated datasets, resulting in a total of 4,200 datasets. Hereafter, we refer to the two different effect sizes of interactions using “lower” and “higher,” and the two different sample sizes as “smaller” and “larger.”

*Fitting models.*

We ran four methods on our simulated datasets. All exposures and continuous covariates were standardized in analysis. To obtain a baseline, we ran a multiple linear regression, including all exposures and covariates as linear, additive terms in the model. We refer to this model as the naive MLR. Then, we ran a multiple linear regression with the true model explicitly specified by excluding non-significant exposures and specifying the known form of non-linear terms and non-additive interactions. We refer to this model as the oracle MLR.

Next, we ran BKMR using the `bkmr` package in R with the default priors. We ran the Markov chain Monte Carlo (MCMC) sampler for 50,000 iterations, as recommended by Bobb et al., and discarded the first 25,000 iterations for burn-in. For larger size datasets, we sped up computations by employing a Gaussian predictive process on 10 knots specified evenly across the predictor space. We ran BSR using the `NLinteraction` package in R with the default priors and with the default Gibbs sampler. We fit BSR on a grid of four degrees of freedom values, {1, 2, 3, 4}, using 5,000 MCMC iterations, discarding the first 2,5000 iterations for burn-in each time. We selected the degrees of freedom using the WAIC criterion recommended by Antonelli et al. Then, we fit the full BSR model using 50,000 MCMC iterations, discarding the first 25,000 iterations for burn-in and thinning each chain based on default settings.

In scenarios where we simulated an interaction between race and XX, we ran stratified BKMR and BSR models. This involved running XX separate models for each category, each with the same settings specified above. We also ran QGC models using the `qgcompint` package in R. [SPECIFY SETTINGS]

*Model assessment.*

We assessed model performance based on detection of significant univariate chemicals as well as detection of interactions. For the naive and oracle MLRs, we considered a p-value less than 0.05 to indicate detection of a significant term. For BKMR and BSR, we considered a PIP greater than or equal to 0.5 to indicate detection of a significant term. For two-way interactions in BKMR, we considered formal detection of an interaction based on 95% confidence intervals constructed around the estimated response at the 0.25 and 0.75 quantiles of exposures while holding all other exposures at their 0.5 quantiles, as described above. We considered a three-way interaction to be detected if at least one of the three confidence intervals of two-way interactions between exposures participating in the three-way interactions did not include 0. We used the Bonferroni correction to adjust for multiple comparisons.

For both BKMR, we also visually assessed detection of interactions by plotting the estimated exposure-response surface for one chemical while fixing one (or two, for three-way interactions) other exposure(s) at their 0.1, 0.5, and 0.9 quantiles. In all scenarios, we calculated the sensitivity as the proportion of times a significant term was correctly detected. False discovery rates were calculated as the proportion of times a significant term was incorrectly detected.

For stratified models, we compared the estimated response across each separate model. Specifically, for BKMR, we computed a confidence interval for the difference in estimated response at the 0.25 and 0.75 quantiles of XX on each category of race. We adjusted for multiple comparisons based on the Bonferroni procedure. We considered an interaction as correctly detected if there was at least one non-overlap between the target group’s confidence interval and all other confidence intervals, and all other confidence intervals overlapped. For BSR, [XX what did we do?]. For QGC, [XX what did we do?].

**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