**A Simulation Study for Evaluating Model Performance in the Estimation of Health Effects From Complex Environmental Mixtures**

**Abstract**

*V2*

Bayesian Kernel Machine Regression (BKMR) and other machine learning (ML) methods are commonly deployed to identify individual causal effects on health outcomes for elements within high-dimensional environmental mixtures. However, there remains a need for direct head-to-head comparison in such settings with complex structure, including non-linearities and interactions, as well as moderate multicollinearity, between exposures and confounders in the functional form of the outcome. Here, simulation studies are conducted on various ML methods — particularly those commonly used in Environmental Health Sciences (EHS) research and that allow for inference and flexible functional form specification — to better understand comparative performance. Data is generated for 10 exposures (metals) and five confounders. Evaluation occurs initially in two scenarios: one with a linear relationship between exposures and outcome, the other with a more complex relationship between exposures and outcome, but both with complex confounding and similar correlation structure. Model performance explores (relative) bias of estimates vs. true effects, estimate variance, computational run-time, and other measures that will enable scientists to make better decisions about which model(s) to use in their mixture studies. Ultimately, no single method consistently and accurately controls for confounding in these settings, indicating a continued need for model and ensemble optimization in EHS applications.

*V1*

Bayesian Kernel Machine Regression (BKMR) and other machine learning (ML) methods are commonly deployed to isolate and identify individual causal effects on health outcomes for elements within high-dimensional environmental mixtures. Although prior work has compared BKMR to other ML techniques in a variety of simulation settings, there remains a need for direct head-to-head comparison in settings with complex structure, including non-linearities and interaction effects with confounders in the functional form of the outcome, as well as moderate or high multicollinearity between exposures and/or confounders. Here, simulation studies are performed on a variety of ML methods — particularly those that are commonly used in Environmental Health Sciences (EHS) research and allow for multiple continuous exposures, inference, and flexible yet relatively automated functional form specification — to better understand how different algorithms perform against each other. Data is generated using parameters (mean, variance/covariance structure) from the Strong Heart Study for 10 exposures (metals) and five confounders. Evaluation of ML models occurs initially in two scenarios: one with a linear relationship between exposures and outcome, the other with a more complex relationship between exposures and outcome, but both with complex confounding and similar correlation structure. Model performance explores bias and relative bias of estimates compared to true effects, variance of estimates, computational run-time, and other measures that will enable environmental scientists and statisticians to make better-informed decisions about which model or set thereof to apply to their mixture studies. Ultimately, we find that no single ML method consistently and accurately controls for confounding in such settings, indicating a sustained need for continued model optimization and improved ensemble techniques in EHS applications.