

► DIRECT FROM CDC ENVIRONMENTAL HEALTH SERVICES

Innovative Applied Epidemiology Methods for Analysis of Environmental Mixtures

Jennifer Bobb, PhD
Kaiser Permanente Washington
Health Research Institute

Chris Gennings, PhD
Icahn School of Medicine at Mount Sinai

Marianthi-Anna Kioumourtoglou, ScD
School of Public Health
at Brown University

Vishal Midya, MStat, PhD
Icahn School of Medicine at Mount Sinai

Editor's Note: The National Environmental Health Association strives to provide up-to-date and relevant information on environmental health and to build partnerships in the profession. In pursuit of these goals, we periodically feature columns on environmental health from the Centers for Disease Control and Prevention (CDC) in the *Journal*.

In these columns, authors from CDC's Water, Food, and Environmental Health Services Branch, as well as guest authors, will share tools, resources, and guidance for environmental health practitioners. The conclusions in this column are those of the author(s) and do not necessarily represent the views of CDC.

Dr. Jennifer Bobb is an associate investigator in the Biostatistics Division at Kaiser Permanente Washington Health Research Institute and an affiliate associate professor in biostatistics at the University of Washington. Dr. Chris Gennings is a professor and director of the Biostatistics Division within the Department of Environmental Medicine and Climate Science at the Icahn School of Medicine at Mount Sinai. Dr. Marianthi-Anna Kioumourtoglou is a professor of epidemiology and environment and society at the School of Public Health at Brown University. Dr. Vishal Midya is an assistant professor in the Department of Environmental Medicine and Climate Science at the Icahn School of Medicine at Mount Sinai.

Introduction

Environmental epidemiologic studies have traditionally focused on examining associations between health outcomes and a single stressor, whether chemical or non-chemical. Human exposure, however, rarely occurs in isolation as individuals are simultaneously subjected to complex mixtures of pollutants. For example, wildfires generate a heterogeneous mixture of hazardous pollutants, including fine particulate matter, carbon

monoxide, polycyclic aromatic hydrocarbons, and volatile organic compounds. These constituents can interact synergistically or antagonistically, complicating the attribution of health effects to individual components.

Consequently, advancing epidemiologic research requires methodologies that account for the combined and potentially interactive effects of pollutants in multipollutant exposures to better characterize their negative impact on human health. The Division of

Environmental Health Science and Practice within the Centers for Disease Control and Prevention (CDC) hosted its third annual environmental epidemiology methods workshop, Innovative Methods for Analysis of Environmental Mixtures in Epidemiology, on August 1, 2024, to help address this issue. The Environmental Epidemiology Methods Workshop Planning Committee included Johnni Daniel, Dana Flanders, Denise Gaughan, Sarah Segerlind, Andrea Winkvist, and Fuyuen Yip.

Building Capacity

This workshop series, in collaboration with academic partners, has provided information on innovative and applied methods in environmental epidemiology for public health practitioners. There are currently few scientific tools and resources for public health practitioners to analyze data to address important environmental epidemiology research questions and to translate study findings into actionable information for professionals. These workshops have served as an important scientific resource and forum for information exchange, supporting capacity building at the state and local levels to advance environmental public health practice in applied environmental epidemiology. According to the Epidemiology Capacity Assessment report from the Council of State and Territorial Epidemiologists (2024), over 2,500 additional epidemiologists are needed to reach full capacity to support the 10 Essential Public Health Services, and there is a particular need for data analytics training for the public health workforce. It is therefore important to identify approaches to continue building capacity in applied environmental epidemiology.

The 2024 workshop featured four presentations:

1. Overview of Issues in the Analysis of Mixtures by Marianthi-Anna Kioumourtzoglou, ScD
2. Weighted Quantile Sum Regression by Chris Gennings, PhD
3. Kernel Machine Regression by Jennifer Bobb, PhD
4. Other Machine Learning Applications by Vishal Midya, PhD

The recorded presentations are available on the CDC website (www.cdc.gov/eh-nexus/media/3rd-epi-workshop.html). They are a useful resource for researchers and public health professionals interested in environmental epidemiology and in evaluating exposure to mixtures.

Overview of Issues in the Analysis of Mixtures

Numerous methods have been developed for the evaluation of mixtures (Joubert et al., 2022). Some methods were specifically developed for environmental mixtures, while others have been adopted from other fields. A key step before any mixtures-related analysis is to develop a well-defined question, which will guide the choice of methods.

In 2015, the National Institute of Environmental Health Sciences held a workshop concerning methods to characterize mixture effects (Taylor et al., 2016). Two main conclusions were: 1) the choice of an analytic method should depend on the research question and 2) no method outperforms all other methods across all research questions. Several articles have identified mixtures-related research questions and provided examples of methods for each class of questions (Gibson, Goldsmith, & Kioumourtzoglou, 2019; Gibson, Nunez, et al., 2019).

Weighted Quantile Sum Regression

Weighted quantile sum (WQS) regression is a statistical method for estimating mixture effects on developmental or health outcomes and identifying key contributing components. It involves two main steps: 1) estimating a weighted index of standardized (i.e., quantiles) concentrations using a nonlinear ensemble model with bootstrap averaging (Bello, 2014; Carrico et al., 2015; Curtin et al., 2019) and 2) testing the significance of the regression coefficient for this index. Ran-

dom data splits for weight estimation and hypothesis testing, with multiple iterations, enhance result generalizability (Tanner et al., 2019, 2020). Extensions include interaction-specific weight estimation, stratified coefficients, and the 2iWQS model for simultaneous positive and negative direction estimation, all available in the gWQS R package (Bennett et al., 2022; Renzetti et al., 2023).

Bayesian Kernel Machine Regression

Bayesian kernel machine regression (BKMR) is a flexible statistical method for estimating the health effects of environmental mixtures, capturing complex, nonlinear, and interactive exposure-response relationships (Bobb et al., 2015). The method incorporates variable selection to discern important mixture components, including a hierarchical approach to address multicollinearity among highly correlated exposures (Bobb et al., 2018). The Bayesian framework enables the incorporation of uncertainty and prior information, enhancing the interpretability and flexibility of the analysis. The BKMR R package supports diverse model configurations, such as continuous or binary outcomes and random intercept models. Additionally, BKMR has been extended for applications involving time-varying exposures, longitudinal outcomes, and mediation analysis, broadening its applicability in epidemiologic research (Bobb et al., 2018; Devick et al., 2022; Liu et al., 2017, 2018).

Other Machine Learning Applications

With the advent of precision environmental health (Baccarelli et al., 2023), there is a need for a statistical modeling framework that identifies populations with exposure patterns that might confer differential susceptibility to diseases and disorders. As the number of chemical components and omics layers increases, however, identifying interpretable subgroups can become challenging due to computational complexity. Interpretable machine learning tools, such as tree-based models, can help identify such subgroups. The repeated-holdout signed iterative random forest (rh-SiRF) (Basu et al., 2018; Midya et al., 2023) and gradient boosting-based algorithm (MOXIE) (Midya & Gennings, 2024) are a few such recently used methods in envi-

ronmental health. Similar to the conceptual idea of genetic signature, these algorithms could identify chemical combinations (or chemical clique signatures) that characterize specific subgroups with altered susceptibility to disease. Additionally, such algorithms can be seamlessly used to 1) identify multi-ordered nonlinear interactions with chemical mixture models (Midya et al., 2023) and 2) discover multi-omic signatures (e.g., metallomics and gut microbiome) associated with health outcomes (Midya et al., 2024).

Summary

CDC's workshops on environmental epidemiology methods have demonstrated significant success, consistently attracting more than 500 participants each year. Participant feedback has underscored the value of these workshops as essential resources for accessing timely, practically meaningful epidemiological methods and information to bolster applied environmental epidemiology at the state and local levels. The workshops aimed to address a critical gap in applied epidemiology capacity within state, tribal, local, and territorial health departments, providing a pivotal opportunity for workforce upskilling. By enhancing applied epidemiology competencies, the educational workshops can help attendees more effectively leverage their data to inform evidence-based public health actions and advance environmental public health initiatives.

The workshop recordings from 2022 and 2023 are available on the following CDC websites: www.cdc.gov/eh-nexus/media/1st-epi-workshop.html and www.cdc.gov/eh-nexus/media/2nd-epi-workshop.html. ✨

Corresponding Author: Sarah Segerlind, Health Communications Specialist, Centers for Disease Control and Prevention, 4770 Buford Highway, Atlanta, GA 30341. Email: htz7@cdc.gov

References

- Baccarelli, A., Dolinoy, D.C., & Walker, C.L. (2023). A precision environmental health approach to prevention of human disease. *Nature Communications*, 14, Article 2449. <https://doi.org/10.1038/s41467-023-37626-2>
- Basu, S., Kumbier, K., Brown, J.B., & Yu, B. (2018). Iterative random forests to dis-

- cover predictive and stable high-order interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 115(8), 1943–1948. <https://doi.org/10.1073/pnas.1711236115>
- Bello, G. (2014). *Application and extension of weighted quantile sum regression for the development of a clinical risk prediction tool* [Doctoral dissertation, Virginia Commonwealth University]. VCU Archives. <https://doi.org/10.25772/3HVVW-ZQ89>
- Bennett, D.H., Busgang, S.A., Kannan, K., Parsons, P.J., Takazawa, M., Palmer, C.D., Schmidt, R.J., Doucette, J.T., Schweitzer, J.B., Gennings, C., & Hertz-Picciotto, I. (2022). Environmental exposures to pesticides, phthalates, phenols and trace elements are associated with neurodevelopment in the CHARGE study. *Environment International*, 161, Article 107075. <https://doi.org/10.1016/j.envint.2021.107075>
- Bobb, J.F., Claus Henn, B., Valeri, L., & Coull, B.A. (2018). Statistical software for analyzing the health effects of multiple concurrent exposures via Bayesian kernel machine regression. *Environmental Health*, 17, Article 67. <https://doi.org/10.1186/s12940-018-0413-y>
- Bobb, J.F., Valeri, L., Claus Henn, B., Christiani, D.C., Wright, R.O., Mazumdar, M., Godleski, J.J., & Coull, B.A. (2015). Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. *Biostatistics*, 16(3), 493–508. <https://doi.org/10.1093/biostatistics/kxu058>
- Carrico, C., Gennings, C., Wheeler, D.C., & Factor-Litvak, P. (2015). Characterization of weighted quantile sum regression for highly correlated data in a risk analysis setting. *Journal of Agricultural, Biological, and Environmental Statistics*, 20, 100–120. <https://doi.org/10.1007/s13253-014-0180-3>
- Council of State and Territorial Epidemiologists. (2024). *2024 epidemiology capacity assessment report*. https://eca.cste.org/wp-content/uploads/2024/09/2024-ECA-Report_FINAL-1.pdf
- Curtin, P., Kellogg, J., Cech, N., & Gennings, C. (2019). A random subset implementation of weighted quantile sum (WQS_{RS}) regression for analysis of high-dimensional mixtures. *Communications in Statistics—Simulation and Computation*, 50(4), 1119–1134. <https://doi.org/10.1080/03610918.2019.1577971>
- Devick, K.L., Bobb, J.F., Mazumdar, M., Claus Henn, B., Bellinger, D.C., Christiani, D.C., Wright, R.O., Williams, P.L., Coull, B.A., & Valeri, L. (2022). Bayesian kernel machine regression-causal mediation analysis. *Statistics in Medicine*, 41(5), 860–876. <https://doi.org/10.1002/sim.9255>
- Gibson, E.A., Goldsmith, J., & Kioumourtoglou, M.-A. (2019). Complex mixtures, complex analyses: An emphasis on interpretable results. *Current Environmental Health Reports*, 6, 53–61. <https://doi.org/10.1007/s40572-019-00229-5>
- Gibson, E.A., Nunez, Y., Abuawad, A., Zota, A.R., Renzetti, S., Devick, K.L., Gennings, C., Goldsmith, J., Coull, B.A., & Kioumourtoglou, M.-A. (2019). An overview of methods to address distinct research questions on environmental mixtures: An application to persistent organic pollutants and leukocyte telomere length. *Environmental Health*, 18, Article 76. <https://doi.org/10.1186/s12940-019-0515-1>
- Joubert, B.R., Kioumourtoglou, M.-A., Chamberlain, T., Chen, H.Y., Gennings, C., Turyk, M.E., Miranda, M.L., Webster, T.F., Ensor, K.B., Dunson, D.B., & Coull, B.A. (2022). Powering Research Through Innovative Methods for Mixtures in Epidemiology (PRIME) program: Novel and expanded statistical methods. *International Journal of Environmental Research and Public Health*, 19(3), Article 1378. <https://doi.org/10.3390/ijerph19031378>
- Liu, S.H., Bobb, J.F., Claus Henn, B., Gennings, C., Schnaas, L., Téllez-Rojo, M., Bellinger, D., Arora, M., Wright, R.O., & Coull, B.A. (2017). Bayesian varying coefficient kernel machine regression to assess cognitive trajectories associated with exposure to complex mixtures. *Statistics in Medicine*, 37(30), 4680–4694. <https://doi.org/10.1002/sim.7947>
- Liu, S.H., Bobb, J.F., Lee, K.H., Gennings, C., Claus Henn, B., Bellinger, D., Austin, C., Schnaas, L., Téllez-Rojo, M.M., Hu, H., Wright, R.O., Arora, M., & Coull, B.A. (2018). Lagged kernel machine regression for identifying time windows of susceptibility to exposures of complex mixtures. *Biostatistics*, 19(3), 325–341. <https://doi.org/10.1093/biostatistics/kxx036>
- Midya, V., Alcala, C.S., Rechtman, E., Gregory, J.K., Kannan, K., Hertz-Picciotto, I., Teitelbaum, S.L., Gennings, C., Rosa, M.J., & Valvi, D. (2023). Machine learning assisted discovery of interactions between pesticides, phthalates, phenols, and trace elements in child neurodevelopment. *Environmental Science & Technology*, 57(46), 18139–18150. <https://doi.org/10.1021/acs.est.3c00848>
- Midya, V., & Gennings, C. (2024). Detecting shape-based interactions among environmental chemicals using an ensemble of exposure-mixture regression and interpretable machine learning tools. *Statistics in Biosciences*, 16, 395–415. <https://doi.org/10.1007/s12561-023-09405-6>
- Midya, V., Nagdeo, K., Lane, J.M., Torres-Olascoaga, L.A., Torres-Calapiz, M., Gennings, C., Horton, M.K., Téllez-Rojo, M.M., Wright, R.O., Arora, M., & Eggers, S. (2024). Prenatal metal exposures and childhood gut microbial signatures are associated with depression score in late childhood. *Science of the Total Environment*, 916, Article 170361. <https://doi.org/10.1016/j.scitotenv.2024.170361>
- Renzetti, S., Gennings, C., & Calzo, S. (2023). A weighted quantile sum regression with penalized weights and two indices. *Frontiers in Public Health*, 11, Article 1151821. <https://doi.org/10.3389/fpubh.2023.1151821>
- Tanner, E.M., Bornehag, C.-G., & Gennings, C. (2019). Repeated holdout validation for weighted quantile sum regression. *MethodsX*, 6, 2855–2860. <https://doi.org/10.1016/j.mex.2019.11.008>
- Tanner, E.M., Hallerback, M.U., Wikström, S., Lindh, C., Kiviranta, H., Gennings, C., & Bornehag, C.-G. (2020). Early prenatal exposure to suspected endocrine disruptor mixtures is associated with lower IQ at age seven. *Environment International*, 134, Article 105185. <https://doi.org/10.1016/j.envint.2019.105185>
- Taylor, K.W., Joubert, B.R., Braun, J.M., Dilworth, C., Gennings, C., Hauser, R., Heindel, J.J., Rider, C.V., Webster, T.F., & Carlin, D.J. (2016). Statistical approaches for assessing health effects of environmental chemical mixtures in epidemiology: Lessons from an innovative workshop. *Environmental Health Perspectives*, 124(12), A227–A229.