HEI STATEMENT

Synopsis of Research Report 234

New Statistical Approaches for Addressing Challenging Questions About Air Pollution and Health

BACKGROUND

Recent decades have seen an increase in the development of statistical methods to answer questions about whether exposure to outdoor air pollution is causally linked to the risk of death. The need for investigators to satisfy specific data and analytic conditions to estimate causal effects from observational data is a key challenge to conducting such analyses. Dr. Joseph Antonelli of the University of Florida and colleagues sought to develop new statistical approaches for addressing challenging questions about potential causal links between air pollution mixtures and risk of death. Their study was funded through HEI's Request for Applications *19-2*: Walter A. Rosenblith New Investigator Award.

APPROACH

Antonelli and colleagues proposed developing approaches to address four specific challenges associated with conducting causal inference analyses related to air pollution mixtures and health:

1. Develop an approach to identify the effect of long-term exposure to a given air pollutant on health when multiple pollutants are present and when only limited data are available about their distributions.
2. Develop a method to estimate the effects of exposure to a mixture of pollutants on the risk of mortality and to identify subgroups of the population that are most susceptible to the effects of exposure to air pollution.
3. Evaluate the effect of accounting for daily mobility patterns on estimates of the health effects associated with long-term exposures to air pollution.
4. Develop a method to account for bias in cases for which potential confounding variables are not measured or accounted for perfectly in the statistical model.

These proposed aims were expected to advance the field by highlighting the challenges inherent in studying the health effects of multiple pollutant exposures and expanding the variety of research questions that can be answered regarding air pollution mixtures.

What This Study Adds

- This study is among the first to address a series of common challenges faced by researchers assessing the health effects of exposures to air pollution mixtures, from exposure assessment to causal inference.
- Antonelli and colleagues describe several new approaches to support causal inference, focusing on exposure to multiple pollutants, addressing bias resulting from confounding, and enhancing exposure assessment by incorporating people's mobility patterns.
- The approaches are applied to hypothetical scenarios and to data from a US Medicare cohort to demonstrate the effectiveness of the practice methods.
- Overall, the approaches presented here provide new solutions for overcoming challenges to assessing causal links between air pollution and health.

In some cases, the approaches and statistical models were described for theoretical applications, and in others, Antonelli and colleagues applied their methods to real-world data using a nationwide study of air pollution and health in the US Medicare population. Those analyses are based on the health information from over 30 million Medicare beneficiaries during the years 2000–2016 living in about 30,000 zip codes across the United States. For air pollution data, they used annual estimated concentrations of fine particles and several other pollutants from existing spatiotemporal models, all at a spatial resolution of about 1 km × 1 km, aggregated to zip codes.

KEY RESULTS

One key assumption for the validity of causal inference in environmental epidemiology is that every exposure value potentially of interest is, in fact, possible for all individuals in the dataset. That assumption would be violated if, for example, no

This Statement, prepared by the Health Effects Institute, summarizes a research project funded by HEI and conducted by Dr. Joseph Antonelli at the University of Florida and colleagues. Research Report 234 contains the detailed Investigators' Report and a Commentary on the study prepared by the HEI Review Committee.

men who are 40–45 years of age in the study dataset were exposed to the air pollution concentration of interest, as it would not be possible to learn about the risk of the health outcome associated with that exposure level among such men. As such, Antonelli and colleagues presented strategies to redefine causal models in a way that respects the limitations in the scope of the available exposure data.

The investigators also introduced a flexible statistical approach that allows one to better understand which population subgroups are most affected by exposure to air pollution or who might benefit most from an intervention to reduce exposures.

Antonelli and colleagues explored how estimates of health effects could be biased when mobility is ignored in estimating exposures to air pollution (e.g., by considering exposures only at study participants' residential addresses). Here, they presented a method that used location data derived from cell phones. They compared epidemiological models in which exposure was estimated for study participants, a) by using only the zip code in which each participant lived, and b) by modeling how much time each participant spent in their home and in other zip codes. Here, they found that although most participants spent about 22% of their time outside their home zip code, incorporating mobility information at that scale did not lead to appreciably different estimates of exposures, nor to appreciably different estimates of health effects. For example, the mean exposure to fine particles averaged across all study years using only the home zip code was 9.34 $\mu g/m^3$, and that which incorporated mobility was 9.64 $\mu g/m^3$.

INTERPRETATION AND CONCLUSIONS

In its independent evaluation of the Investigators' Report, the HEI Review Committee commended the investigators for developing original study aims and for tackling important issues for environmental epidemiology with which many statisticians continue to struggle. The Committee members agreed that the investigators did an excellent job setting the context and explaining the rationale for pursuing causal inference approaches to the study of air pollution exposures.

The innovative work presented by the investigators includes a rigorous assessment of the assumptions required to draw causal conclusions when studying complex mixtures of pollutants. They introduced several interesting theoretical approaches that advance the field. Their finding that incorporating mobility did not lead to appreciably different estimates of exposures corroborates recent findings from other studies.

Not surprisingly, the study team did not resolve all the important challenges described. For example, an overarching limitation of this study is that the various issues addressed by the investigators do not exist in isolation, yet the methods proposed here were presented independently. Their study is a valid starting point; however, addressing each challenge individually does not fully resolve the broader challenge of developing a comprehensive approach incorporating all of the important challenges in a single framework.

Another limitation of the study relates to the applied analyses with the Medicare data. The theoretical foundations presented here were evaluated in a very specific dataset with unique features (i.e., older adults and data available only at the zip code level) that might prevent valid inferences for reasons beyond the issues addressed by the investigators.

Although there is more work to be done to resolve these complex issues fully, this study provides a strong foundation for future work to extend the concepts described here to other settings.