HEALTH EFFECTS INSTITUTE

STATEMENT

Synopsis of Research Report 185

Elemental Composition of Indoor, Outdoor, and Personal $PM_{2.5}$ Samples Using RIOPA Data

INTRODUCTION

$PM_{2.5}$ has been associated with adverse health effects. It is, however, a complex mixture of many components that vary in composition and size and originate from a variety of outdoor sources. Assessments of exposure to $PM_{2.5}$ and its components and their associated health effects are further complicated by the fact that there are also indoor sources and that individual behaviors may influence exposure substantially. Patrick H. Ryan and his colleagues used data from the Relationships of Indoor, Outdoor, and Personal Air (RIOPA) study to explore relationships among the elemental compositions of indoor, outdoor, and personal $PM_{2.5}$ samples in greater detail than was done in the original RIOPA study.

The RIOPA study was conducted in Los Angeles, California; Houston, Texas; and Elizabeth, New Jersey. It included approximately 300 subjects who did not smoke and who lived at various distances from air pollution sources. In addition to indoor, outdoor (directly outside the home), and personal measurements, the original investigators collected data on factors that might affect exposures, such as personal and home characteristics and geographic information.

The aims of the current study were to explore the relationships among the elemental compositions of indoor, outdoor, and personal $PM_{2.5}$ samples; to identify clusters of individuals with similar exposures; and to investigate whether indoor, outdoor, and personal and home characteristics can be used to predict personal exposure to $PM_{2.5}$ elements.

APPROACH

Analyses were limited to 168 adults with complete data for at least one concurrently obtained set of sample types (indoor, outdoor, and personal). Twenty-four elements were analyzed that had detectable values in at least 70% of the personal samples. In pooled and city-specific analyses, relationships among the elemental compositions of the three sample types were explored using Spearman correlation coefficients, calculation of outdoor/personal and indoor/personal ratios, and principal component analysis. To identify clusters of individuals, model-based cluster analyses of personal samples

What This Study Adds

- Ryan and colleagues used RIOPA data to explore relationships among the elemental composition of indoor, outdoor, and personal $PM_{2.5}$ samples.
- Outdoor concentrations did not represent personal exposures well for elements other than those associated with long-range transport, such as sulfur and vanadium. For the other elements, outdoor concentrations did not predict personal exposure well and the addition of indoor concentrations and personal and home characteristics did not improve the prediction of personal exposure for most of them.
- The results should be interpreted with caution because important clustering in the data was not accounted for, the number of predictor variables in the models was large compared with the size of the data set, and the influence of outlier values was not tested. Performing studies such as RIOPA remains useful in order to quantify exposure measurement error and, ultimately, allow researchers to take this quantification into account in health analyses.

This Statement, prepared by the Health Effects Institute, summarizes a research project funded by HEI and conducted by Dr. Patrick H. Ryan at Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, and colleagues. The complete report, *Analysis of Personal and Home Characteristics Associated with the Elemental Composition of $PM_{2.5}$ in Indoor, Outdoor, and Personal Air in the RIOPA Study* (© 2015 Health Effects Institute), can be obtained from HEI or our Web site (see next page).

were conducted (using the R package mclust). Several linear and random-forest regression models were run, largely aiming to predict total personal exposure using all 24 elements measured indoors and outdoors and several home characteristics as predictors in the same models. The investigators used cross-validation methods to test the performance of their models. *P* values less than 0.05 were considered significant.

MAIN RESULTS AND INTERPRETATION

Outdoor concentrations did not represent personal exposures well for elements other than those associated with long-range transport (and with few known indoor sources), such as sulfur and vanadium. For the other elements, outdoor concentrations did not predict personal exposure well and the addition of indoor concentrations and personal and home characteristics did not improve the prediction of personal exposure for most of them. Only in the linear regression analyses did inclusion of indoor concentrations significantly improve the prediction of personal exposure (for nine elements — Ba, Ca, Cl, Cu, K, Sn, Sr, V, and Zn).

In its independent review of the study, the HEI Health Review Committee noted that the authors had conducted an extensive set of analyses on data from the 168 RIOPA participants for whom concurrent indoor, outdoor, and personal exposure concentration data for elements in $PM_{2.5}$ were available. The analyses included traditional approaches to comparing sample types, such as ratio and correlation measures; a traditional approach applied in a unique way (i.e., principal component analysis); and a novel approach (i.e., random forest analysis). However, the Committee identified several important issues with the analytic approaches summarized below that warrant caution in interpreting the results, in particular the linear regression analyses.

The analyses presented in the report were not adjusted for clustering or correlation within cities, among individuals, or by season. RIOPA was designed to capture data on various air pollution sources and weather conditions in the three cities. Principal component analysis results showed, as expected, notable differences in air pollution across the cities. Therefore, the analyses using pooled data across cities, such as the model-based cluster analyses, provided limited meaningful insights.

The Committee questioned in particular the linear regression analyses, because the number of predictor variables in the models was large (71 in the final model) compared with the number of observations ($N = 168$). Inclusion of a large number of predictors in a regression model intended for prediction and inference testing at the same time is problematic because unnecessary variables reduce the ability of the model to predict the outcome variable properly and may introduce multi-collinearity problems if they are correlated with other variables that destabilize the model. The results from the random forest analyses — which were notably different from the results from the linear regression analyses — were not affected by the issues described above. A possible explanation of the differences between the results could be the influence of outlier values, which were abundantly present in RIOPA data.

CONCLUSIONS

Ryan and colleagues used RIOPA data to explore relationships among elements found in indoor, outdoor, and personal samples of $PM_{2.5}$. Analyses included traditional approaches to comparing sample types, such as ratio and correlation measures; a traditional approach applied in a unique way (i.e., principal component analysis); and a novel approach (i.e., random forest analysis). In its independent review of the study, the HEI Health Review Committee noted that caution is warranted in interpreting the results, in particular the linear regression analyses, because important clustering in the data was not accounted for, the number of predictor variables in the models was large compared with the size of the data set, and the influence of outlier values was not tested. Conducting detailed exposure measurement studies such as RIOPA remains important in order to quantify exposure measurement error and, ultimately, allow researchers to take this quantification into account in health analyses.

HEALTH EFFECTS INSTITUTE

101 Federal Street, Suite 500
Boston, MA 02110, USA
+1-617-488-2300 phone
+1-617-488-2335 fax

pubs@healtheffects.org
www.healtheffects.org

Recycled Paper