RESEARCH REPORT

HEALTH EFFECTS INSTITUTE

Number 185
December 2015

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

Patrick H. Ryan, Cole Brokamp, Zhi-Hua (Tina) Fan, and M.B. Rao

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

Patrick H. Ryan, Cole Brokamp, Zhi-Hua (Tina) Fan, and M.B. Rao

with a Critique by the HEI Health Review Committee

Research Report 185
Health Effects Institute
Boston, Massachusetts

Trusted Science • Cleaner Air • Better Health

Publishing history: This document was posted at *www.healtheffects.org* in December 2015.

Citation for document:

Ryan PH, Brokamp C, Fan Z-H, Rao MB. 2015. 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. Research Report 185. Boston, MA:Health Effects Institute.

 Cameographics, Belfast, Me., Compositor. Printed by Recycled Paper Printing, Boston, Mass. Library of Congress Catalog Number for the HEI Report Series: WA 754 R432.

Cover paper: made with at least 55% recycled content, of which at least 30% is post-consumer waste; free of acid and elemental chlorine. Text paper: made with 100% post-consumer waste recycled content; acid free; no chlorine used in processing. The book is printed with soy-based inks and is of permanent archival quality.

CONTENTS

About HEI v

About This Report vii

HEI STATEMENT 1

INVESTIGATORS' REPORT *by Ryan et al.* 3

ABSTRACT 3

INTRODUCTION 4

SPECIFIC AIMS 5

METHODS AND STUDY DESIGN 5

STATISTICAL METHODS AND DATA ANALYSIS 6

- Overview of Statistical Approach 6
- Data Importation and Preparation 6
- Descriptive Analyses 6
- Principal Component Analyses 7
- Clustering Subjects 7
- Linear Regression 7
  - Regression Model 1: Outdoor Only 8
  - Regression Model 2: Outdoor + Indoor 8
  - Regression Model 3: Outdoor + Indoor + Personal and Home Characteristics 8
- Random Forest 8

RESULTS 9

- Descriptive Analyses 9
  - Correlation Between Indoor, Outdoor, and Personal Concentrations of Elements 15
  - Ratio of Elemental Concentrations in Indoor–Personal and Outdoor–Personal Samples 16
- Principal Component Analyses 17
  - Comparing Elemental Heterogeneity Across Sample Types by Study Site 17
  - Comparing Elemental Heterogeneity Across Study Sites by Sample Type 17
- Clustering Subjects 17
  - Personal and Home Characteristics Associated with Cluster Membership 20
- Linear Regression 24
  - Does the Inclusion of Indoor Elemental Concentrations Improve the Prediction of Personal Exposure Compared with Outdoor Elemental Concentrations Only? 24
  - Does the Inclusion of Personal and Home Characteristics Further Improve the Prediction of Personal Exposure Compared with Outdoor and Indoor Elemental Concentrations? 24

Random Forest Analysis 24
Does the Inclusion of Indoor Elemental Concentrations Improve the Prediction of Personal Exposure Compared with Outdoor Elemental Concentrations Only? 25
Does the Inclusion of Personal and Home Characteristics Further Improve the Prediction of Personal Exposure Compared with Outdoor and Indoor Elemental Concentrations? 25
DISCUSSION AND CONCLUSIONS 28
Correlations and Ratios of Elemental Concentrations in Indoor–Personal and Outdoor–Personal Samples 28
Comparing Elemental Heterogeneity Using Principal Component Analysis 29
Clustering Subjects Based on the Elemental Composition of Personal $PM_{2.5}$ 31
Personal and Home Characteristics Associated with Cluster Membership 32
Improving the Prediction of Personal Exposure Using Indoor $PM_{2.5}$ and Personal and Home Characteristics 33
IMPLICATIONS OF FINDINGS 34
SUMMARY OF KEY FINDINGS 36
REFERENCES 37
APPENDICES AVAILABLE ON THE WEB 39
ABOUT THE AUTHORS 39
OTHER PUBLICATION RESULTING FROM THIS RESEARCH 39
ABBREVIATIONS 40
ABBREVIATIONS OF CHEMICAL ELEMENTS 40

CRITIQUE *by the Health Review Committee* 41
INTRODUCTION 41
APPROACH 42
Summary of Results 43
HEALTH REVIEW COMMITTEE'S EVALUATION 43
SUMMARY AND CONCLUSIONS 45
ACKNOWLEDGMENTS 45
REFERENCES 45

Related HEI Publications 47

HEI Board, Committees, and Staff 49

ABOUT HEI

The Health Effects Institute is a nonprofit corporation chartered in 1980 as an independent research organization to provide high-quality, impartial, and relevant science on the effects of air pollution on health. To accomplish its mission, the institute

- Identifies the highest-priority areas for health effects research;
- Competitively funds and oversees research projects;
- Provides intensive independent review of HEI-supported studies and related research;
- Integrates HEI's research results with those of other institutions into broader evaluations; and
- Communicates the results of HEI's research and analyses to public and private decision makers.

HEI typically receives balanced funding from the U.S. Environmental Protection Agency and the worldwide motor vehicle industry. Frequently, other public and private organizations in the United States and around the world also support major projects or research programs. HEI has funded more than 330 research projects in North America, Europe, Asia, and Latin America, the results of which have informed decisions regarding carbon monoxide, air toxics, nitrogen oxides, diesel exhaust, ozone, particulate matter, and other pollutants. These results have appeared in more than 260 comprehensive reports published by HEI, as well as in more than 1000 articles in the peer-reviewed literature.

HEI's independent Board of Directors consists of leaders in science and policy who are committed to fostering the public–private partnership that is central to the organization. The Health Research Committee solicits input from HEI sponsors and other stakeholders and works with scientific staff to develop a Five-Year Strategic Plan, select research projects for funding, and oversee their conduct. The Health Review Committee, which has no role in selecting or overseeing studies, works with staff to evaluate and interpret the results of funded studies and related research.

All project results and accompanying comments by the Health Review Committee are widely disseminated through HEI's Web site (*www.healtheffects.org*), printed reports, newsletters and other publications, annual conferences, and presentations to legislative bodies and public agencies.

ABOUT THIS REPORT

Research Report 185, *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,* presents a research project funded by the Health Effects Institute and conducted by Dr. Patrick H. Ryan of the Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, and his colleagues. The report contains three main sections.

The HEI Statement, prepared by staff at HEI, is a brief, nontechnical summary of the study and its findings; it also briefly describes the Health Review Committee's comments on the study.

The Investigators' Report, prepared by Ryan and colleagues, describes the scientific background, aims, methods, results, and conclusions of the study.

The Critique is prepared by members of the Health Review Committee with the assistance of HEI staff; it places the study in a broader scientific context, points out its strengths and limitations, and discusses remaining uncertainties and implications of the study's findings for public health and future research.

This report has gone through HEI's rigorous review process. When an HEI-funded study is completed, the investigators submit a draft final report presenting the background and results of the study. This draft report is first examined by outside technical reviewers and a biostatistician. The report and the reviewers' comments are then evaluated by members of the Health Review Committee, an independent panel of distinguished scientists who have no involvement in selecting or overseeing HEI studies. During the review process, the investigators have an opportunity to exchange comments with the Review Committee and, as necessary, to revise their report. The Critique reflects the information provided in the final version of the report.

HEI 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 were conducted (using the R package mclust). Several linear and random-forest regression models were run, largely aiming to predict total personal

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. Research Report 185 contains both the detailed Investigators' Report and a Critique of the study prepared by the Institute's Health Review Committee.

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.

INVESTIGATORS' 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

Patrick H. Ryan,[1] Cole Brokamp,[2] Zhi-Hua (Tina) Fan,[3] and M.B. Rao[2]

[1]Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; [2]Department of Environmental Health, University of Cincinnati, Cincinnati, Ohio; [3]Department of Environmental and Occupational Medicine, Rutgers University, New Brunswick, New Jersey

ABSTRACT

The complex mixture of chemicals and elements that constitute particulate matter (PM*) varies by season and geographic location because source contributors differ over time and place. The composition of PM having an aerodynamic diameter < 2.5 μm ($PM_{2.5}$) is hypothesized to be responsible, in part, for its toxicity. Epidemiologic studies have identified specific components and sources of $PM_{2.5}$ that are associated with adverse health outcomes. The majority of these studies use measures of outdoor concentrations obtained from one or a few central monitoring sites as a surrogate for measures of personal exposure.

Personal $PM_{2.5}$ (and its elemental composition), however, may be different from the $PM_{2.5}$ measured at stationary outdoor sites. The objectives of this study were (1) to describe the relationships between the concentrations of various elements in indoor, outdoor, and personal $PM_{2.5}$ samples, (2) to identify groups of individuals with similar exposures to mixtures of elements in personal $PM_{2.5}$ and to examine personal and home characteristics of these groups, and (3) to evaluate whether concentrations of elements from outdoor $PM_{2.5}$ samples are appropriate surrogates for personal exposure to $PM_{2.5}$ and its elements and whether indoor $PM_{2.5}$ concentrations and information about home characteristics improve the prediction of personal exposure.

The objectives of the study were addressed using data collected as part of the Relationships of Indoor, Outdoor, and Personal Air (RIOPA) study. The RIOPA study has previously measured the mass concentrations of $PM_{2.5}$ and its elemental constituents during 48-hour concurrent indoor, outdoor (directly outside the home), and personal samplings in three urban areas (Los Angeles, California; Houston, Texas; and Elizabeth, New Jersey). The resulting data and information about personal and home characteristics (including air-conditioning use, nearby emission sources, time spent indoors, census-tract geography, air-exchange rates, and other information) for each RIOPA participant were downloaded from the RIOPA study database.

We performed three sets of analyses to address the study aims. First, we conducted descriptive analyses to describe the relationships between elemental concentrations in the concurrently gathered indoor, outdoor, and personal air samples. We assessed the correlation between personal exposure and indoor concentrations as well as personal exposure and outdoor concentrations of each element and calculated ratios between them. In addition, we performed principal component analysis (PCA) and calculated principal component scores (PCSs) to examine the heterogeneity of the elemental composition and then tested whether the mixture of elements in indoor, outdoor, and personal $PM_{2.5}$ was significantly different within each study site and across study sites.

Secondly, we performed model-based clustering analysis to group RIOPA participants with similar exposures to mixtures of elements in personal $PM_{2.5}$. We examined the association between cluster membership and the concentrations of elements in indoor and outdoor $PM_{2.5}$ samples and personal and home characteristics. Finally, we developed a

This Investigators' Report is one part of Health Effects Institute Research Report 185, which also includes a Critique by the Health Review Committee and an HEI Statement about the research project. Correspondence concerning the Investigators' Report may be addressed to Dr. Patrick H. Ryan, 3333 Burnet Avenue, ML 5041, Cincinnati, OH 45229; email: *patrick.ryan@cchmc.org*.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award CR–83467701 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and therefore may not necessarily reflect the views of the Agency, and no official endorsement by it should be inferred. The contents of this document also have not been reviewed by private party institutions, including those that support the Health Effects Institute; therefore, it may not reflect the views or policies of these parties, and no endorsement by them should be inferred.

* Lists of abbreviations appear at the end of the Investigators' Report.

series of linear regression models and random forest models to examine the association between personal exposure to elements in $PM_{2.5}$ and (1) outdoor measurements, (2) outdoor and indoor measurements, and (3) outdoor and indoor measurements and home characteristics. As we developed each model, the improvement in prediction of personal exposure when including additional information was assessed. Personal exposures to $PM_{2.5}$ and to most elements were significantly correlated with both indoor and outdoor concentrations, although concentrations in personal samples frequently exceeded those of indoor and outdoor samples. In general, for most $PM_{2.5}$ elements indoor concentrations were more highly correlated with personal exposure than were outdoor concentrations. PCA showed that the mixture of elements in indoor, outdoor, and personal $PM_{2.5}$ varied significantly across sample types within each study site and also across study sites within each sample type. Using model-based clustering, we identified seven clusters of RIOPA participants whose personal $PM_{2.5}$ samples had similar patterns of elemental composition. Using this approach, subsets of RIOPA participants were identified whose personal exposures to $PM_{2.5}$ (and its elements) were significantly higher than their indoor and outdoor concentrations (and vice versa). The results of linear and random forest regression models were consistent with our correlation analyses and demonstrated that (1) indoor concentrations were more significantly associated with personal exposure than were outdoor concentrations and (2) participant reports of time spent at their home significantly modified many of the associations between indoor and personal concentrations. In linear regression models, the inclusion of indoor concentrations significantly improved the prediction of personal exposures to Ba, Ca, Cl, Cu, K, Sn, Sr, V, and Zn compared with the use of outdoor elemental concentrations alone. Including additional information on personal and home characteristics improved the prediction for only one element, Pb.

Our results support the use of outdoor monitoring sites as surrogates of personal exposure for a limited number of individual elements associated with long-range transport and with a few local or indoor sources. Based on our PCA and clustering analyses, we concluded that the overall elemental composition of $PM_{2.5}$ obtained at outdoor monitoring sites may not accurately represent the elemental composition of personal $PM_{2.5}$. Although the data used in these analyses compared outdoor $PM_{2.5}$ composition collected at the home with indoor and personal samples, our results imply that studies examining the complete elemental composition of $PM_{2.5}$ should be cautious about using data from central outdoor monitoring sites because of the potential for exposure misclassification. The inclusion of personal and home characteristics only marginally improved the prediction of personal exposure for a small number of elements in $PM_{2.5}$. We concluded that the additional cost and burden of indoor and personal sampling may be justified for studies examining elements because neither outdoor monitoring nor questionnaire data on home and personal characteristics were able to represent adequately the overall elemental composition of personal $PM_{2.5}$.

INTRODUCTION

Epidemiologic studies have consistently demonstrated that PM air pollution is associated with cardiovascular and respiratory morbidity and mortality (Dockery 2009; Simkhovich et al. 2008; Zanobetti et al. 2009). These studies often used the mass concentration of $PM_{2.5}$ collected at ambient regulatory sampling sites as the primary exposure metric. However, $PM_{2.5}$ is a heterogeneous mix of solid and liquid particles of varying composition and from various sources. More recently, studies have demonstrated that the composition of $PM_{2.5}$ exhibits seasonal and geographic variations, which may in part explain the observed seasonal and geographic variability in $PM_{2.5}$-associated health effects (Bell et al. 2007, 2008, 2011; Dominici et al. 2006).

The effort to characterize the health effects of PM components has been identified as a research priority by the National Research Council of the National Academies (2004). Toxicological studies suggest that the organic compounds and transition metals found in $PM_{2.5}$ may be particularly important because they are able to elicit inflammation, with subsequent respiratory and cardiovascular effects (Schlesinger et al. 2006). Epidemiologic studies examining the sources and composition of PM have identified several specific components — including elemental carbon, organic carbon, and nitrates — associated with increased risk for cardiovascular and respiratory hospital admissions (Bell et al. 2009; Peng et al. 2009) and mortality (Ostro et al. 2007). Elemental components of $PM_{2.5}$, including As, Al, Br, Cr, Ni, Si, V, and Zn, have also been associated with increased cardiovascular and respiratory hospital admissions (Bell et al. 2009; Zanobetti et al. 2009), increased mortality (Franklin et al. 2008), and lower birth weight (Bell et al. 2010).

An important caveat to these studies is the assumption that the elemental composition of $PM_{2.5}$ obtained at stationary outdoor monitoring sites adequately represented that of personal exposure for a population in a given geographic region. Previous studies have found, however, that ambient PM concentrations did not represent personal PM

exposure, because of the time–activity patterns of individuals, which included time spent indoors and other personal activities (Hsu et al. 2012; Nerriere et al. 2005; Ozkaynak et al. 1995). The difference between outdoor and personal PM may be more pronounced when examining specific elements in $PM_{2.5}$ because the sources of these elements vary indoors, in vehicles, and in other locations. The difference between outdoor and personal PM is also affected by other factors, including air-conditioning use, air-exchange rates, sources of PM indoors, and other home characteristics.

Fewer studies have examined the composition of elements in indoor and outdoor $PM_{2.5}$ at an individual or household level in relationship to personal exposure to $PM_{2.5}$. In a study of elderly subjects in two European cities, investigators found that concentrations of elements in $PM_{2.5}$ primarily of outdoor origin were lower in personal and indoor samples than in outdoor samples, although they were significantly correlated (Brunekreef et al. 2005; Janssen et al. 2005). Elements with indoor sources, including Ca, Cu, and Si, exhibited less correlation between personal and outdoor samples. In a study of elderly patients with preexisting respiratory disease, investigators showed that spatial variability (reflecting regional versus local sources of PM) was an important factor affecting the correlation between concentrations of elements in outdoor and personal $PM_{2.5}$ samples. In the study, central monitoring site concentrations of S (a marker of PM of outdoor origin with low spatial variability) reflected personal exposure well, whereas concentrations of Ni (with high spatial variability) at ambient monitoring sites were not well correlated with personal concentrations (Hsu et al. 2012).

The RIOPA study presented a unique opportunity to examine, at an individual or household level, concurrent indoor, outdoor, and personal exposure to elements in $PM_{2.5}$. The RIOPA study was designed to collect data to evaluate the contribution of outdoor sources of air toxics and PM to personal exposure (Weisel 2005a). The overall design of the RIOPA study included homes from three study sites in Los Angeles, Houston, and Elizabeth. In each of two measurement seasons, continuous 48-hour sampling was conducted on a subset of the homes of participants to concurrently measure indoor, outdoor, and personal $PM_{2.5}$ concentrations (Turpin et al. 2007; Weisel 2005a, 2005b). In addition, data on home and personal characteristics were collected for all study participants. It should be noted that the RIOPA study collected outdoor measurements directly outside the home of the study participants rather than at central monitoring sites. Measurements of outdoor concentrations obtained in the RIOPA study may therefore be expected to exhibit a higher degree of correlation with indoor and personal concentrations than the concentrations obtained from measurements at central monitoring sites. In our analyses we use the RIOPA dataset to explore the association between the elemental composition of $PM_{2.5}$ in indoor, outdoor, and personal air samples collected at the level of individuals or households.

SPECIFIC AIMS

The objective of this study was to examine the relationships between the elemental compositions of personal, indoor, and outdoor $PM_{2.5}$ samples. The overarching purpose of the study was to assess how well concentrations of the mixtures of elements in outdoor $PM_{2.5}$ reflected those of personal $PM_{2.5}$ and to determine if indoor $PM_{2.5}$ measurements and other home characteristics were significantly associated with personal exposure. Using data obtained from the three RIOPA study sites, the following specific aims were addressed in the present study:

1. Explore the relationship between the concentrations of $PM_{2.5}$ elements sampled in indoor, outdoor, and personal air samples;
2. Identify groups (clusters) of individuals with similar exposures to mixtures of elements in personal $PM_{2.5}$ and ascertain the elemental composition of indoor and outdoor $PM_{2.5}$ as well as personal and home characteristics associated with each identified group; and
3. Investigate whether personal exposure to $PM_{2.5}$ elements was predicted by outdoor and indoor concentrations and by personal and home characteristics.

METHODS AND STUDY DESIGN

Our study included statistical analyses of data collected as part of the RIOPA study. A complete description of the RIOPA study, including its design, methods, and quality control procedures, is available elsewhere (Turpin et al. 2007; Weisel 2005a, 2005b). The mass concentrations of $PM_{2.5}$ and its elemental constituents, which were measured during 48-hour concurrent indoor, outdoor, and personal samplings in the RIOPA study, were downloaded from the RIOPA study database (*https://riopa.aer.com*). Elemental concentrations measured by X-ray fluorescence (XRF) were included in our study rather than those obtained by means of inductively coupled plasma mass spectrometry (ICP–MS) because 36 elements were analyzed by XRF and only to 22 were analyzed by ICP–MS. Earlier studies found good agreement between XRF and

ICP–MS results for most elements in the RIOPA dataset (Turpin et al. 2007). Although the mass concentration of $PM_{2.5}$, elemental carbon, and organic carbon were also included in the initial dataset, concentrations for elemental carbon and organic carbon concentrations are not available for personal samples and therefore were not included in our analyses of personal exposure.

Personal and home characteristics of RIOPA participants were also downloaded and included in our analyses. These personal and home characteristics are from the RIOPA Baseline Questionnaire and Technician Walk-Through databases and included the presence of a basement, unvented appliances, nearby industrial emission sources, nearby gas stations and restaurants, the presence and type of air conditioning, heating source, flooring material, and the type of transportation used (car or bus). Time–activity questionnaires were also completed by the study participants; the proportion of the sampling period spent at home was included our analyses. Air-exchange rates measured at the homes were also included. We examined census-tract-level information for the participating homes, including the elevation of the centroid of the census tract in which a participating home was located, the population and housing density of the census tract, and a measure of the percentage of the census tract classified as "highly developed."

STATISTICAL METHODS AND DATA ANALYSIS

OVERVIEW OF STATISTICAL APPROACH

In order to describe the elemental composition of indoor, outdoor, and personal $PM_{2.5}$ samples and the relationship between the various compositions of the three types of samples (Aim 1), we performed descriptive analyses, correlation analyses, and principal component analyses and examined the ratios of elemental concentrations in indoor-to-personal and outdoor-to-personal samples. Model-based clustering was conducted to identify groups of individuals across all study sites and within each study location whose personal $PM_{2.5}$ samples had similar elemental compositions (Aim 2). After identifying these clusters, we described each cluster in terms of the average concentrations of elements in indoor and outdoor $PM_{2.5}$. Personal and home characteristics of individuals within each cluster were also examined for their association with the clusters of individuals having similar personal $PM_{2.5}$ composition. Finally, linear regression and random forest analyses were performed to determine the contribution of outdoor $PM_{2.5}$ elements to personal $PM_{2.5}$ exposure (Aim 3). Improvement in the prediction of personal $PM_{2.5}$ exposure was assessed in each model (linear and random forest) after the inclusion of elemental concentrations from indoor sampling and personal and home characteristics. Additional details about each approach are provided below.

All statistical analyses were conducted using R software (version 3.0.2). Packages implemented within R included: mclust (version 4.2) (*http://cran.r-project.org/web/packages/mclust/index.html*) and randomForest (version 4.6-7) (*http://cran.r-project.org/web/packages/randomForest/index.html*) (Fraley et al. 2012; Liaw and Wiener 2002).

DATA IMPORTATION AND PREPARATION

Data were downloaded directly from the RIOPA database website. Because the overall focus of our study was on the relationship between elements in indoor and outdoor $PM_{2.5}$ with respect to the elemental composition of personal $PM_{2.5}$, elements with nondetectable concentrations in 30% or more of personal samples were excluded from the analyses. Samples with concentrations that were detectable but below the analytical limit of detection were included as reported in the RIOPA database, and nondetectable elemental concentrations were replaced by zero. We excluded personal $PM_{2.5}$ samples from children because the sample size (n = 23) was limited. We limited the analyses to 168 households with available data on concurrently sampled indoor, outdoor, and personal $PM_{2.5}$ (68 from California, 52 from Texas, and 48 from New Jersey). For 29 of the 68 California households, 23 of the 52 Texas households, and 16 of the 48 New Jersey households, complete data were obtained for both visits. Therefore, only the data from the first complete visit were retained for analysis. All analyses, with the exception of the descriptive analyses, were conducted after applying a $\log(1 + x)$ transformation to all elemental concentrations (and other continuous home and personal characteristics, as needed) in order to make the data normal and the variance relatively constant.

DESCRIPTIVE ANALYSES

We assessed the correlation between each measured element in indoor and personal (I–P) air samples and outdoor and personal (O–P) air samples by calculating the Spearman correlation coefficients. In addition, elemental concentrations in outdoor to personal (O/P) ratios and indoor to personal (I/P) ratios were determined for each RIOPA participant, and the distribution of O/P and I/P ratios was summarized in box-and-whisker plots.

PRINCIPAL COMPONENT ANALYSES

PCAs were conducted to examine the heterogeneity of elemental constituents in samples of indoor, outdoor, and personal $PM_{2.5}$. PCA is a data reduction technique in which the original data are transformed into a set of uncorrelated new variables, called principal components (PCs). The PCs are weighted combinations of the original data, and each component is chosen so that it has the largest possible variance while under the constraint that it must be orthogonal with all of the preceding components. The weighting factors, or loadings, can be examined to determine which constituents are the greatest contributors to each PC. It is important to note that PCA is used to reduce data dimensionality without losing any of the original information.

In order to quantitatively compare the elemental heterogeneity of $PM_{2.5}$ across the three types of samples (i.e., indoor, outdoor, and personal air samples) and the three study sites (in California, Texas, and New Jersey), two analyses were conducted.

In the first analysis, to examine elemental heterogeneity across sample types by study site, the elemental concentrations for each sample type (indoor, outdoor, and personal) were combined, and separate PCAs were conducted for each study site. The principal component scores (PCS) for each sample type within study site were derived as follows (Han et al. 2012):

$$PCS_p = \beta_{1p}(X_1) + \beta_{2p}(X_2) + \ldots + \beta_{ip}X_i \quad (1)$$

where PCS_p is the PCS on principal component p, β_{ip} is the loading for each measured element i on principal component p, and X_i is the elemental concentration scaled to have a one-unit variance. The PCSs were calculated using only PCs with eigenvalues greater than one. The mean PCS (and 95% confidence intervals) for the first three PCSs for each sample type by study site were plotted, and ANOVA was conducted to test if the mean PCS differed by sample type within each study site.

A similar approach was taken in the second analysis to examine elemental heterogeneity across study sites by sample type. Elemental concentrations for each study site (California, Texas, and New Jersey) were combined, and separate PCAs were conducted for each sample type. PCSs for each study site were derived as described in Equation 1, and analysis of variance (ANOVA) was used to test if the mean PCS differed by study site within each sample type.

CLUSTERING SUBJECTS

We performed model-based cluster analyses to group RIOPA participants with similar mixtures of elemental concentrations in personal $PM_{2.5}$ samples. Normal mixture modeling (using the R package mclust) (Fraley and Raftery 2002; Fraley et al. 2012) was used to identify groups (clusters) of participants based on the elemental concentrations in personal $PM_{2.5}$ samples. The mclust package uses an expectation maximization algorithm with a mixture of parameterized normal models. A Bayesian information criterion is used to identify both the optimal number of clusters and the appropriate mixture of Gaussian models. Clustering was performed for the pooled RIOPA dataset and also stratified by study site given the difference sources of $PM_{2.5}$ exposure and varying subject and home characteristics by study site. According to the Bayesian information criterion, the best model was found to be an equal-covariance mixture model with seven clusters for the pooled dataset. Clustering performed by study site identified five clusters for California and three each for New Jersey and Texas.

A heatmap was used to visualize individuals' personal, indoor, and outdoor $PM_{2.5}$ within each cluster by plotting the scaled mean concentration of each element (calculated for each element by subtracting the mean and dividing by the standard deviation to create a Z-score). We also examined the mean $PM_{2.5}$ in the indoor, outdoor, and personal samples and the personal and home characteristics of the RIOPA participants in each cluster. Differences in the participants' personal and home characteristics between clusters were tested using either a Pearson chi-squared test (with P values obtained by way of Monte Carlo simulation) or a Kruskal–Wallis rank sum test, as appropriate.

LINEAR REGRESSION

In order to examine the associations between personal exposure to $PM_{2.5}$ elements and indoor and outdoor concentrations, we developed a series of linear regression models. In our initial approach, for each element in personal $PM_{2.5}$, the indoor and outdoor concentration measurements were used to fit a linear regression model predicting the personal concentration of that same element:

$$personal = \beta_0 + \beta_{outdoor} \times X_{outdoor} + \beta_{indoor} \times X_{indoor} \quad (2)$$

In addition, we investigated the change in prediction of personal exposure when including participant reports of time spent indoors by including this information in the linear regression model. Specifically, we included the fraction of the total sampling time that the participant reported

spending indoors at home and its interaction with the indoor elemental concentration:

$$\begin{aligned} personal &= \beta_0 + \beta_{outdoor} \times X_{outdoor} \\ &+ \beta_{indoor} \times X_{indoor} + \beta_{time\ frac} \times X_{time\ frac} \\ &+ \beta_{indoor \times time\ frac} \times X_{indoor} \times X_{time\ frac} \end{aligned} \quad (3)$$

Both of these regression models were implemented for each of the 24 $PM_{2.5}$ elements separately.

Although regression models were developed using both indoor and outdoor concentrations in this study, indoor $PM_{2.5}$ data are not typically available in epidemiologic studies. Indeed, most epidemiologic studies use the elemental composition of outdoor $PM_{2.5}$ obtained at central monitoring sites as a surrogate for personal exposure. Therefore, to determine the association between outdoor $PM_{2.5}$ elemental concentrations and personal exposure and to assess the improvements in model prediction when indoor exposure data and participant home and personal characteristic information are available, three additional linear regression models were developed.

Regression Model 1: Outdoor Only

The first regression model, Outdoor Only, was derived to predict personal concentrations for each $PM_{2.5}$ element based on the concentrations of all elements in outdoor $PM_{2.5}$ only:

$$Y_{personal} = \beta_0 + \sum_{i=Ag}^{Zr} \beta_{i(outdoor)} \times X_{i(outdoor)} \quad (4)$$

where i is Ag, Al, ..., Zr for all 24 elements, and $\beta_{i(outdoor)} \times X_{i(outdoor)}$ represents the regression coefficients multiplied by the respective concentrations for every outdoor element. An intercept and 24 outdoor element measurements resulted in a regression model containing 25 predictive terms.

Regression Model 2: Outdoor + Indoor

The second regression model, Outdoor + Indoor, included the concentrations of all elements in both outdoor and indoor $PM_{2.5}$:

$$Y_{personal} = \beta_0 + \sum_{i=Ag}^{Zr} \begin{pmatrix} \beta_{i(outdoor)} \times X_{i(outdoor)} \\ + \beta_{i(indoor)} \times X_{i(indoor)} \end{pmatrix} \quad (5)$$

where i is Ag, Al, ..., Zr for all 24 elements, $\beta_{i(outdoor)} \times X_{i(outdoor)}$ represents the regression coefficients multiplied by the respective concentrations for every outdoor element, and $\beta_{i(outdoor)} \times X_{i(outdoor)}$ represents the regression coefficients multiplied by the respective concentrations for every indoor element. This model contains 49 predictive terms (the intercept, 24 outdoor element measurements, and 24 indoor element measurements).

Regression Model 3: Outdoor + Indoor + Personal and Home Characteristics

The third regression model, Outdoor + Indoor + Personal and Home Characteristics, included the concentrations of all elements in both outdoor and indoor $PM_{2.5}$ as well as personal and home characteristics:

$$\begin{aligned} Y_{personal} = \beta_0 &+ \sum_{i=Ag}^{Zr} \begin{pmatrix} \beta_{i(outdoor)} \times X_{i(outdoor)} \\ + \beta_{i(indoor)} \times X_{i(indoor)} \end{pmatrix} \\ &+ \sum_{j=1}^{23} \beta_{j(pers)} \times X_{j(pers)} \end{aligned} \quad (6)$$

where $\beta_{j(pers)} \times X_{j(pers)}$ represents the j^{th} of p total home and personal characteristics selected for inclusion in the predictive models. The personal and home characteristics available for at least 90% of the RIOPA participants were considered for inclusion in this model. An initial screening approach was conducted to determine which personal and home characteristics would be included in the final model. Screening was conducted by determining the association between each personal and home characteristic and the concentration of elements in personal $PM_{2.5}$. Variables associated at an alpha = 0.1 level calculated by Spearman correlation test or Mann–Whitney test, as appropriate, for 25% of the elements in personal $PM_{2.5}$ were included in the final model. Regression Model 3 contains an intercept, 24 indoor element measurements, 24 outdoor measurements, and the 23 personal and home characteristics that met the screening criteria, for a total of 72 predictive terms.

The squared Pearson correlation coefficient between the observed and predicted values (cross-validated [CV] R^2) was used to quantitatively compare the accuracy of each model in order to assess the improvement in modeling of personal exposure as additional information (indoor measurements and personal and home characteristics) was included. CV R^2 was estimated using 10-fold cross-validation repeated 10 times. Furthermore, a t test was used to see if the CV R^2 of Regression Model 2 was different than that of Regression Model 1 and if the CV R^2 of Regression Model 3 was different than that of Regression Model 2 for every element.

RANDOM FOREST

In addition to the linear regression approach, we also pursued a machine-learning technique known as random

forest analysis. The random forest technique is an ensemble learning method that uses bagging (i.e., bootstrapped aggregation of several regression trees) to predict an outcome. A regression tree employs a decision tree that seeks to make the most accurate prediction of a continuous outcome based on multiple predictor variables. At every split point in the tree, all possible values of the randomly selected descriptive variables are attempted and whichever split point results in the greatest reduction of entropy is used. An ensemble of many regression trees is known as a random forest and generates a better predictive performance than any single regression tree alone. Several hundred trees are generated, each using a different bootstrapped sample of the data and the prediction of each tree is aggregated to give one prediction for the entire ensemble of trees. This results in reduced variance of the prediction while maintaining the unbiased nature of the prediction. In our analyses described below, we used random forests to predict elemental concentrations in personal $PM_{2.5}$ based on both indoor and outdoor elemental concentrations as well as personal and home characteristics.

We assessed the prediction of personal exposure of each $PM_{2.5}$ element based on information about all elements by developing three random forest models, much like the linear regression modeling approach described above: (1) an Outdoor Only, (2) an Outdoor + Indoor, and (3) an Outdoor + Indoor + Personal and Home Characteristics model.

CV R^2 was calculated as described above for the linear models (10-fold cross-validation repeated 10 times) with the folds maintained so that comparisons between models were possible as well as comparisons for each element and model type between the linear regression model and the random forest model.

RESULTS

DESCRIPTIVE ANALYSES

We restricted our analyses to a subset of 24 elements (Ag, Al, As, Ba, Br, Ca, Cl, Cr, Cu, Fe, K, Mn, Ni, Pb, S, Sb, Se, Si, Sn, Sr, Ti, V, Zn, and Zr), with greater than 80% of the samples having detectable concentrations. A total of 208 RIOPA participants completed at least one portion of the $PM_{2.5}$ sampling (indoor, outdoor, or personal) (Turpin et al. 2007). Of these, 81% (n = 168) of the participating RIOPA households had available elemental concentrations for at least one visit for all sample types (68 from California, 52 from Texas, and 48 from New Jersey). A summary of the elemental concentrations obtained from the complete RIOPA dataset by study site is available elsewhere (Turpin et al. 2007). Tables 1–3 show the concentrations of $PM_{2.5}$ and the elemental constituents for personal, indoor, and outdoor samples, respectively, for the subset of RIOPA participants used in the current study. Data from the three locations were combined (pooled) in those analyses.

Table 4 presents a summary of home and personal characteristics for the overall cohort and for each location individually. Overall, the cohort was largely female (73%) and about 44% of the participants self-identified as Caucasian. These percentages varied widely by location: the cohort in California was a more even mix of sexes (59% female) than were the cohorts of Texas and New Jersey (88% and 80% female, respectively). The cohort in New Jersey was markedly different from the overall cohort in terms of race, with only 9% self-reporting as Caucasian. Overall, 82% of participants had education at or above completion of high school, and 54% had a household yearly income greater than $25,000.

On average, concentrations of $PM_{2.5}$ in personal samples were twice as high as either indoor or outdoor concentrations. The presence of extreme values was also observed, with maximum concentrations of many elements exceeding three standard deviations from the mean. Much as for $PM_{2.5}$, all of the concentrations of elements in the personal samples (except As [outdoor], Br [outdoor], S, Se, and V) were higher than the concentrations in the indoor and outdoor samples. These results suggest that there are other sources of personal exposure besides those in the home and in the neighborhood air. People may have spent time in other microenvironments, such as on roadways, where higher concentrations of elements are expected than in the home.

Elemental concentrations varied largely by species. The highest concentrations among all of the elements in each sample type were documented for S, followed by Si, Ca, Fe, K, Cl, and Al. For these elements, natural sources, including crustal minerals and soil, are the dominant contributors. In addition, large variations in concentrations were observed for Ca, S, Si, Cl, and Zn. These variations were primarily driven by a small number of samples with extremely high concentrations. Overall, the outdoor elemental concentrations found in the RIOPA data were similar to those found in data from other urban areas (Han et al. 2012; Yu et al. 2011). The personal and indoor elemental concentrations found in the RIOPA data were similar to those found in data from urban areas in the United States (Hsu et al. 2012) but higher than those found in data from studies in European countries (Janssen et al. 2005).

A summary of personal and home characteristics hypothesized a priori to be associated with personal exposure to elemental concentrations, along with personal

Table 1. Summary of Personal Concentrations of $PM_{2.5}$ and Elemental Constituents for all RIOPA Participants (Pooled, n = 168)[a]

	Mean	Median	SD	Minimum	Maximum	IQR	% ND
$PM_{2.5}$ (μg/m³)							
	38.4	31.7	22.8	7.6	151.5	21.6	0.0
Element (ng/m³)							
Ag	1.2	1.0	1.1	0.0	4.5	1.9	26.8
Al	119.4	67.1	135.0	0.0	671.2	140.4	4.8
As	0.8	0.6	0.8	0.0	4.1	1.1	17.9
Ba	27.1	22.7	35.8	1.7	467.7	12.6	0.0
Br	5.2	3.5	11.3	0.3	146.1	3.1	0.0
Ca	461.1	246.8	617.3	24.2	5047.0	395.8	0.0
Cl	180.1	109.6	262.6	5.3	2303.3	158.4	0.0
Cr	3.6	1.6	8.8	0.0	84.8	2.1	2.4
Cu	18.2	9.7	63.2	0.0	776.5	10.6	1.2
Fe	202.4	153.6	162.9	40.0	998.2	134.9	0.0
K	167.0	118.9	149.7	32.4	1028.0	106.0	0.0
Mn	4.0	3.3	2.8	0.4	20.1	2.6	0.0
Ni	5.4	3.8	7.9	0.4	85.3	3.2	0.0
Pb	7.1	4.5	18.2	0.0	232.3	4.1	1.2
S	914.9	766.3	615.4	131.1	3577.4	681.2	0.0
Sb	1.8	1.6	1.7	0.0	8.2	2.3	19.6
Se	0.7	0.5	0.7	0.0	4.0	0.7	11.9
Si	446.0	298.6	480.5	37.0	4492.1	367.7	0.0
Sn	6.3	6.0	3.4	0.5	22.7	4.2	0.0
Sr	2.9	2.0	3.8	0.0	37.3	2.2	1.8
Ti	25.9	19.9	22.4	2.7	211.3	19.2	0.0
V	3.5	2.5	3.3	0.0	13.9	4.2	10.1
Zn	103.4	36.0	366.0	10.1	3361.7	30.4	0.0
Zr	12.1	4.7	23.0	0.0	137.4	9.6	1.8

[a] IQR = interquartile range; ND = nondetectable; SD = standard deviation.

characteristics like sex, race, education, and income, is presented in Table 4. Of the 168 RIOPA participants included in our analyses, 41% reported having industrial emission sources in their neighborhood, 64% reported having air conditioning, and 81% reported using a car for transportation during the sampling period. The majority of the personal and home characteristics significantly varied by study location, including the presence of nearby industrial emission sources (9% in California, 81% in Texas, and 37% in New Jersey) and central air conditioning (16% in California, 44% in Texas, and 0% in New Jersey). Overall, the average air-exchange rate was 1.4; this rate did not differ significantly across study sites. The average reported time spent indoors at home was 66%; this value varied significantly across the study sites (lowest in California [55%] and highest in Texas [82%]).

Table 2. Summary of Indoor Concentrations of $PM_{2.5}$ and Elemental Constituents for all RIOPA Participants (Pooled, n = 168)[a]

	Mean	Median	SD	Minimum	Maximum	IQR	% ND
$PM_{2.5}$ (µg/m³)							
	18.5	14.8	13.3	3.8	97.0	9.5	0.0
Element (ng/m³)							
Ag	0.8	0.7	0.7	0.0	3.0	1.2	23.2
Al	24.1	14.6	32.7	0.0	225.5	32.3	22.0
As	0.6	0.5	0.7	0.0	5.0	0.7	8.9
Ba	14.0	13.0	6.6	1.2	40.0	8.6	0.0
Br	3.8	3.4	3.2	0.4	28.7	2.9	0.0
Ca	107.1	72.9	152.2	0.0	1532.2	65.2	0.6
Cl	78.7	24.5	247.5	0.0	2713.0	47.1	6.5
Cr	1.9	0.5	6.1	0.0	54.2	0.8	7.1
Cu	7.8	4.0	33.7	0.0	436.8	3.8	1.8
Fe	76.4	61.9	48.5	12.1	254.6	59.1	0.0
K	93.9	64.9	129.7	8.3	1036.0	46.6	0.0
Mn	2.0	1.8	1.2	0.0	6.4	1.6	0.6
Ni	2.0	1.5	1.8	0.0	12.1	1.9	4.2
Pb	3.4	3.1	2.2	0.0	15.5	2.9	0.6
S	920.0	740.6	652.7	107.3	3655.6	775.7	0.0
Sb	1.2	0.7	1.3	0.0	8.1	1.9	20.8
Se	0.7	0.5	0.7	0.0	3.6	0.6	9.5
Si	131.0	99.2	93.9	18.1	632.0	103.4	0.0
Sn	4.6	4.3	2.8	0.0	18.7	2.9	0.6
Sr	0.9	0.6	1.3	0.0	12.2	0.9	6.5
Ti	7.8	5.3	12.2	0.7	149.1	5.7	0.0
V	4.0	2.9	3.8	0.0	20.2	4.3	4.8
Zn	67.6	12.2	428.3	1.9	4734.9	13.6	0.0
Zr	0.9	0.7	1.1	0.0	8.9	0.8	9.5

[a] IQR = interquartile range; ND = nondetectable; SD = standard deviation.

Table 3. Summary of Outdoor Concentrations of $PM_{2.5}$ and Elemental Constituents for all RIOPA Participants (Pooled, n = 168)[a]

	Mean	Median	SD	Min	Max	IQR	% ND
$PM_{2.5}$ (μg/m³)							
	17.9	14.2	12.4	4.5	94.9	8.8	0.0
Element (ng/m³)							
Ag	0.6	0.5	0.6	0.0	3.4	0.9	30.4
Al	54.6	11.8	137.7	0.0	1165.0	36.6	25.0
As	0.8	0.7	0.7	0.0	3.9	0.9	10.7
Ba	19.7	16.4	12.0	0.0	84.1	10.9	0.6
Br	4.5	4.2	2.8	0.0	17.6	3.1	0.6
Ca	119.5	76.3	175.3	0.9	1735.0	79.0	0.0
Cl	67.2	10.8	220.3	0.0	2004.3	33.3	19.0
Cr	2.6	0.6	15.0	0.0	188.9	0.8	7.7
Cu	5.9	3.3	8.7	0.3	62.5	3.8	0.0
Fe	173.8	108.5	231.0	1.0	1988.0	118.7	0.0
K	79.4	64.6	61.7	0.0	420.6	45.9	0.6
Mn	3.9	2.9	3.9	0.5	31.4	2.4	0.0
Ni	3.1	2.3	3.6	0.0	24.6	2.4	2.4
Pb	4.7	4.0	3.8	0.4	29.1	3.6	0.0
S	1209.4	1140.3	706.3	0.0	4512.9	906.8	0.6
Sb	1.8	1.4	2.0	0.0	17.6	2.3	18.5
Se	1.2	0.8	1.3	0.0	6.8	0.9	2.4
Si	201.1	106.4	324.6	1.2	2363.0	107.6	0.0
Sn	5.1	4.0	5.3	0.0	55.4	3.2	1.2
Sr	1.3	0.7	1.9	0.0	14.8	1.1	3.6
Ti	10.3	6.8	13.2	0.0	123.0	8.7	1.8
V	6.1	4.3	6.2	0.0	56.8	6.3	3.0
Zn	17.6	13.5	15.4	0.0	118.1	14.4	0.6
Zr	0.7	0.4	1.1	0.0	8.0	0.7	17.3

[a] IQR = interquartile range; ND = nondetectable; SD = standard deviation.

Table 4. Summary of Personal and Home Characteristics by Study Site

Characteristic	Pooled (N = 168)	Study Site		
		California (n = 68)	Texas (n = 53)	New Jersey (n = 47)
Female (%)	73.0	58.8	87.8	80.0
Self-identified as Caucasian (%)	43.9	57.4	50.0	8.6
Education above high school (%)	82.1	95.6	70.8	71.4
Household yearly income greater than $25K (%)	53.6	61.2	48.9	42.9
Census-tract elevation (meters)	73.1	167.0	8.0	9.2
Census-tract population density (per sq. km land)	3972.5	4104.6	1105.4	6831.7
Census-tract housing density (per sq. km land)	1517.2	1663.4	396.4	2500.8
Census-tract % high density developed	17.9	11.3	13.0	32.4
Study time spent indoors at home (%)	66	55	82	65
Wood flooring (%)	23.2	26.8	12.8	31.4
Area rugs (%)	18.1	23.2	6.4	25.7
Wall-to-wall carpet (%)	68.1	80.4	70.2	45.7
Cement floors (%)	3.6	0.0	8.5	2.9
Presence of basement (%)	26.8	8.9	0.0	91.4
Nearby industrial emissions (%)	40.6	8.9	80.9	37.1
Gas station < 200 ft (%)	15.2	3.6	6.4	45.7
Gas station < 1/2 mile (%)	67.4	66.1	72.3	62.9
Restaurant < 200 ft (%)	15.9	16.1	2.1	34.3
Restaurant < 1/2 mile (%)	58.7	60.7	53.2	62.9
Kerosene heater (%)	1.9	0.0	3.8	2.9
Gas heater (%)	3.2	0.0	7.7	2.9
Wood stove (%)	2.6	1.5	3.8	2.9
Fireplace (%)	27.1	45.6	19.2	2.9
Car for transportation (%)	81.3	83.8	86.5	68.6
Bus for transportation (%)	6.5	11.8	1.9	2.9
Pumped gasoline (%)	13.8	17.9	17.3	4.2
Use of deodorant, perfume, or hair spray (%)	75.4	70.1	80.8	77.1
Use of air freshener (%)	34.7	26.9	42.3	37.5
Traveled on roads (%)	76.6	91.0	75.0	58.3

Table continues next page

Table 4 (*continued*). Summary of Personal and Home Characteristics by Study Site

Characteristic	Pooled (N = 168)	Study Site		
		California (n = 68)	Texas (n = 53)	New Jersey (n = 47)
Inside with a smoker (%)	6.0	3.0	0.0	16.7
In a vehicle with a smoker (%)	1.8	4.5	0.0	0.0
Swimming (%)	1.2	3.0	0.0	0.0
Used cleaning supplies (%)	34.7	28.4	36.5	41.7
Took a bath (%)	15.0	16.4	13.5	14.6
Took a shower (%)	88.0	88.1	90.4	85.4
Used felt-tip markers (%)	28.7	40.3	23.1	18.8
Burned candles, oil, or incense in house (%)	19.8	11.9	26.9	22.9
Air-exchange rate	1.4	1.6	1.2	1.2
Air conditioning (%)	63.9	63.2	51.9	82.9
Use of central air conditioning (%)	20.4	16.4	44.2	0.0
Use of window-unit air conditioner (%)	14.4	3.0	19.2	25.0
Use of ceiling or portable fan (%)	39.5	32.8	65.4	20.8
Use of an exhaust fan (%)	44.3	44.8	53.8	33.3
Open windows (%)	74.8	98.5	44.2	74.3
Open doors or windows at home (%)	63.5	77.6	44.2	64.6
Use of central heating (%)	16.2	22.4	9.6	14.6
Use of dishwasher (%)	25.1	38.8	26.9	4.2
Use of a clothes washer (%)	52.1	52.2	65.4	37.5
Use of a clothes dryer (%)	46.1	52.2	53.8	29.2
Use of nail polish remover (%)	10.2	10.4	9.6	10.4
Sprayed pesticides (%)	8.4	7.5	11.5	6.3
Vacuuming the house (%)	42.5	41.8	51.9	33.3
Sweeping the house (%)	53.3	37.3	63.5	64.6
Dusting the house (%)	34.1	22.4	36.5	47.9
Use of cleaning solutions (%)	46.7	34.3	57.7	52.1
Gardening (%)	16.8	17.9	21.2	10.4
Grilling or frying inside the house (%)	44.9	46.3	40.4	47.9
Motor vehicle parked in attached garage (%)	16.2	23.9	17.3	4.2
Motor vehicle started in attached garage (%)	16.2	23.9	17.3	4.2
Diesel vehicle parked around house (%)	43.4	32.3	57.7	42.2
Nearby or operated a diesel vehicle (%)	30.2	47.8	8.2	28.3

Correlation Between Indoor, Outdoor, and Personal Concentrations of Elements

For the pooled RIOPA data, the concentration of $PM_{2.5}$ in personal samples was significantly correlated with indoor (r = 0.46) and outdoor (r = 0.18) concentrations (Figure 1). All elemental concentrations measured in personal $PM_{2.5}$ were significantly correlated with concentrations measured in indoor $PM_{2.5}$ and ranged from 0.16 (Sb) to 0.90 (S) (Figure 1). The highest indoor-to-personal (I–P) correlations were observed for S (r = 0.90), V (r = 0.87), Br (r = 0.76), and Se (r = 0.69). Only Sb had I–P correlation coefficients less than 0.2. The correlation between elements in outdoor $PM_{2.5}$ and personal $PM_{2.5}$ samples (O–P) ranged from −0.13 (Ti) to 0.84 (S) and, with the exception of Ag, were lower than those for I–P correlations (Figure 1). O–P correlations were highest for S (r = 0.84), V (r = 0.82), and Se (r = 0.65) and lowest for Ti (r = −0.13), Zr (r = 0.01), Sb (r = 0.06), Cl (r = 0.08), and Si (r = 0.07) (Figure 1).

The correlation between elemental concentrations in I–P and O–P samples was also examined by study site; the results are shown in Figures A.1–A.3, in Appendix A, available on the Web). In the study sites in California and New Jersey, $PM_{2.5}$ mass concentration was significantly correlated between both I–P and O–P samples. I–P correlations, however, were greater than O–P correlations (r = 0.43 and 0.29 in California and New Jersey, respectively) (Figures A.1–A3 in Appendix A). In contrast, although I–P concentrations of $PM_{2.5}$ were significantly correlated in

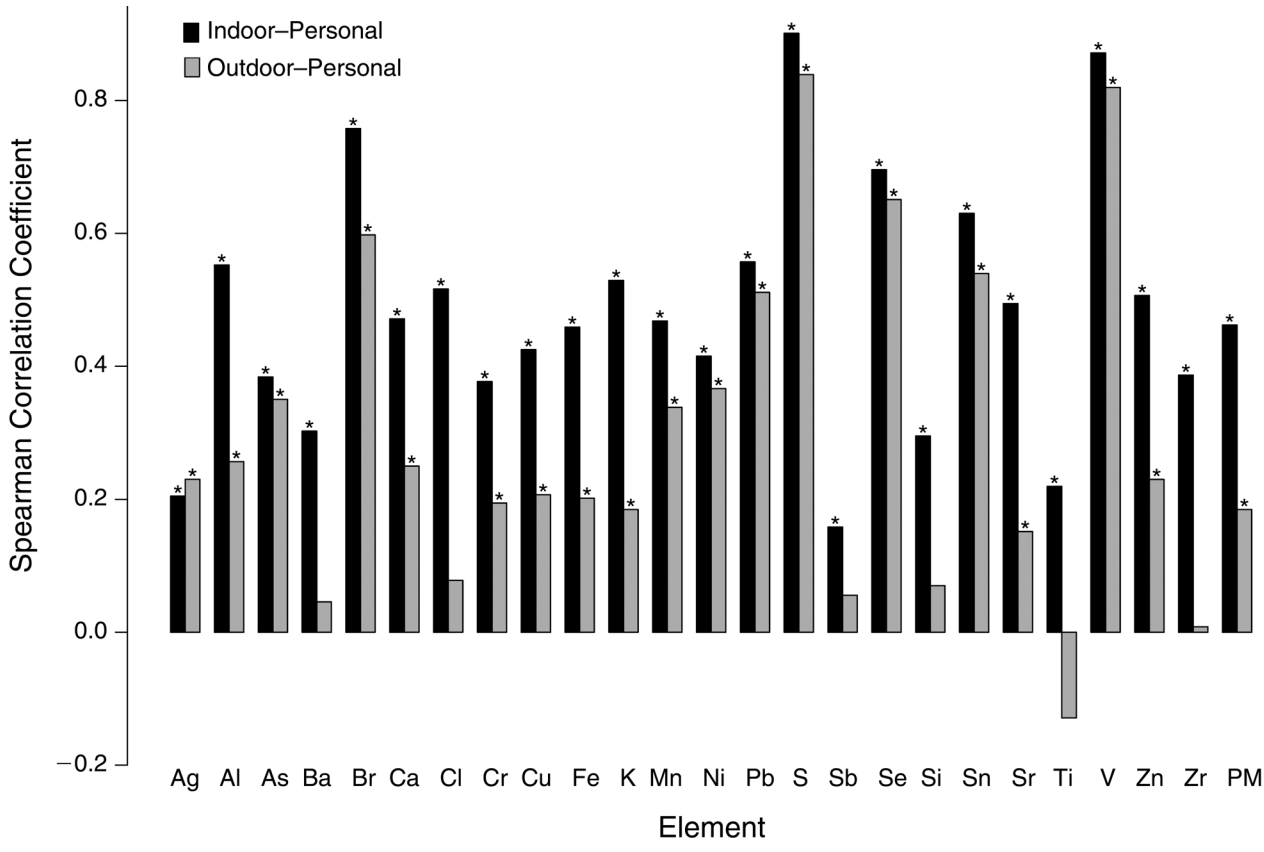

Figure 1. Spearman correlation coefficients between elemental concentrations in indoor and personal (I–P) $PM_{2.5}$ samples and between elemental concentrations in outdoor and personal (O–P) $PM_{2.5}$ (pooled dataset).

data from the study site in Texas (r = 0.65), the O–P concentrations of $PM_{2.5}$ were not correlated. Overall, I–P and O–P correlation coefficients in California were similar to those observed in the pooled dataset (Figure 1); I–P and O–P correlation coefficients were highest for Br, S, Se, and V (Spearman $r > 0.75$) and lowest for Cr and Ti ($r < 0.2$). In Texas, the observed pattern of correlations between elemental concentrations in I–P samples was similar to the pooled dataset and California. However, concentrations in outdoor and personal samples in Texas were not well correlated, with the exception of Ag, Br, S, Se, Sn, and V. In New Jersey, the correlation between I–P $PM_{2.5}$ concentrations exceeded O–P concentrations (r = 0.47 and 0.3, respectively); both I–P and O–P correlations were greatest for S and V ($r > 0.8$).

Ratio of Elemental Concentrations in Indoor–Personal and Outdoor–Personal Samples

The distribution of the ratios of elemental concentrations in indoor-to-personal (I/P) and outdoor-to-personal (O/P) samples for the pooled dataset are shown in Figure 2. I/P and O/P ratios for $PM_{2.5}$ mass concentrations are less than 1, indicating that personal $PM_{2.5}$ exposure exceeded both indoor and outdoor concentrations of $PM_{2.5}$. Similarly, the median I/P ratio for most elements was also less than 1, indicating that personal exposure for more than half of RIOPA participants exceeded indoor concentrations. There were some exceptions: the median I/P ratio for S and Se was approximately 1, suggesting comparable concentrations of these elements in indoor and personal $PM_{2.5}$, and the median I/P ratio for V exceeded 1. Overall, the variability of

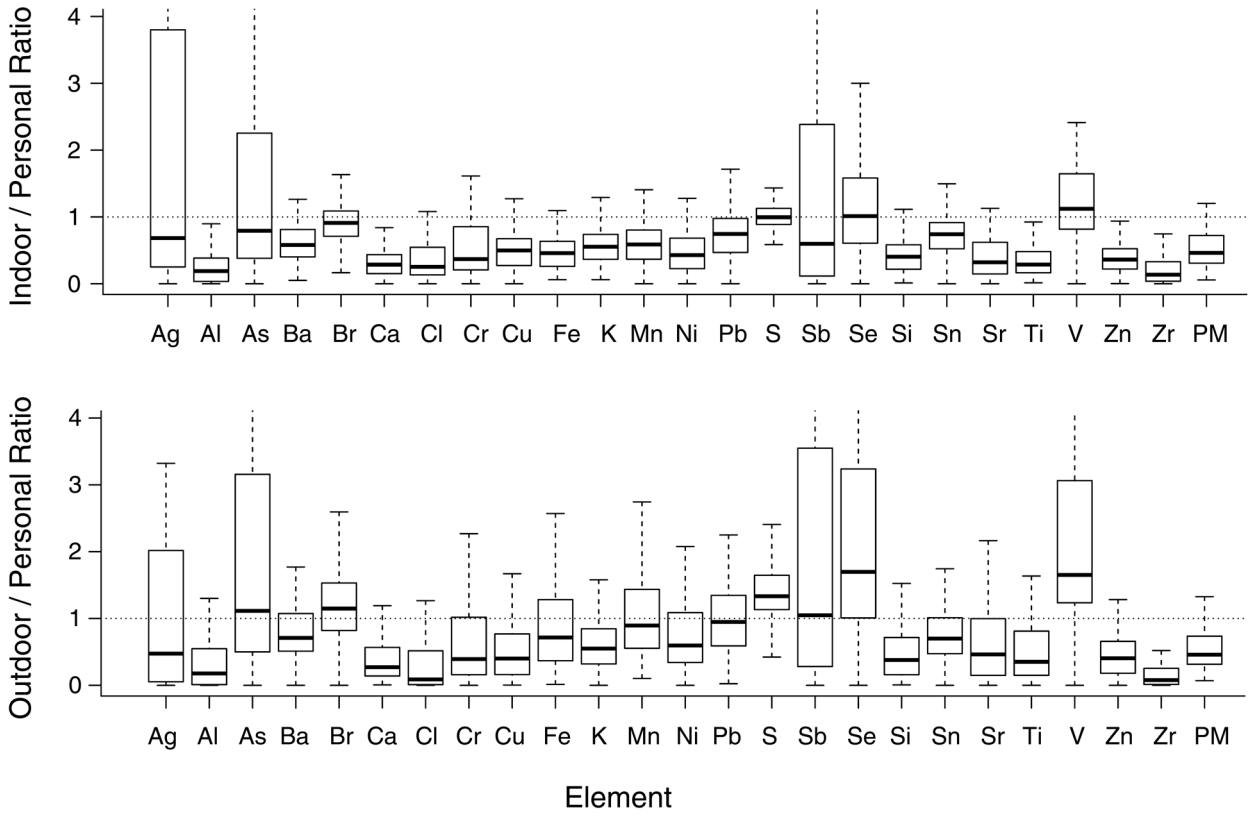

Figure 2. Box plot of indoor/personal (I/P) and outdoor/personal (O/P) ratios (pooled dataset).

I/P ratios was greater for Ag, As, Sb, Se, and V compared with the other elements.

Median O/P ratios less than 1 were observed for Ag, Al, Ba, Ca, Cl, Cr, Cu, Fe, K, Ni, Si, Sn, Sr, Ti, Zn, and Zr, suggesting that personal exposure to these elements exceeded outdoor concentrations for at least half of the RIOPA study participants. Outdoor concentrations were similar to or greater than personal exposures, however, for the majority of RIOPA participants for As, Br, S, Sb, Se, and V.

I/P and O/P ratios were also calculated separately for each study site and are shown in Figures A.4–A.6 in Appendix A (available on the Web). Overall, results for I/P and O/P ratios for both $PM_{2.5}$ and most elements observed for the individual study sites were similar to those in the pooled data. For all sites, personal concentrations of $PM_{2.5}$ exceeded, for the majority of RIOPA participants, both indoor and outdoor concentrations. Similarly, personal concentrations of most elements were also higher than both indoor and outdoor concentrations with the exception of S, Se, and V, whose median I/P and O/P ratios were close to or greater than 1 in all three study locations. The variability of I/P and O/P ratios was similar across the three study sites, although the variability of O/P ratios in New Jersey was much greater.

PRINCIPAL COMPONENT ANALYSES

Comparing Elemental Heterogeneity Across Sample Types by Study Site

Five PCs with an eigenvalue greater than 1 were extracted from the California study site's elemental data, and six PCs with eigenvalues greater than 1 were identified in the elemental data for each of the Texas and New Jersey study sites. The total variance of the dataset explained by the PCs with eigenvalues greater than 1 was similar across all study sites (69.5% in California, 69.7% in Texas, and 71% in New Jersey).

The loading factors for all elements for PC1–PC3 by sample location are shown in Table B.1 (in Appendix B, available on the Web). Elements whose absolute loading factors exceeded 0.3 were considered highly loaded (and are boldfaced in the table). The first PC for all three locations, PC1, was not highly loaded by any elements, as is common in PCA. PC2 was primarily explained by S, Se, and V for all three study site locations and also by Br for California and New Jersey but not for Texas. PC3 in California was largely explained by Cl, Cr, and Ni; in Texas by As, Br, Cl, and Sn, and in New Jersey by As, Cl, K, and Zn.

Principal component scores (PCSs) were derived for each sample type (indoor, outdoor, and personal) for each study site (in California, Texas, and New Jersey) by summing the rotated and scaled data for each PC. The mean and 95% confidence intervals for each are plotted in Figure 3. The mean PCS-1 and mean PCS-2 for each location differed by sample type ($P < 0.001$); however, PCS-3 differed by sample type for only California and New Jersey ($P < 0.001$) but not for Texas ($P = 0.563$).

Comparing Elemental Heterogeneity Across Study Sites by Sample Type

For elemental concentrations in personal samples, five PCs with an eigenvalue greater than 1 were extracted. For indoor and outdoor samples, six and five PCs with an eigenvalue greater than 1 were extracted, respectively. The total variance explained by the extracted PCs was 64.3% for personal samples, 65.5% for indoor samples, and 73.0% for outdoor samples.

The loading factors for all elements for PC1–PC3 by sample type are shown in Table B.2 (in Appendix B, available on the Web). Loading factors greater than 0.3 are boldfaced and indicate the elements that explain the PC. For indoor samples, PC1 was not highly loaded, except for Fe; for PC2 the highest loading factors were for Al, Ca, and Si; for PC3 the highest loading factors were for As and Cr. For outdoor samples, no elements were highly loaded on PC1; Al, Br, Ni, S, Se, and V were all highly loaded on PC2; and Ca, Cu, K, and Sn were highly loaded on PC3. For personal samples, the highest loading factors for PC1 were Ca, Fe, Mn, and Ti; for PC2 S, Se, and V; and for PC3 Ag, Br, Cr, and Ni.

Figure 4 shows the point estimate and 95% confidence interval for the PCS for each sample type by study site. Results of ANOVA found that the mean of each PCS-1 were significantly different by study site for indoor ($P < 0.001$), outdoor ($P < 0.001$), and personal ($P = 0.034$) sample types. Mean PCS-2 scores were significantly different across study sites for indoor ($P < 0.001$) and personal samples ($P < 0.001$) but not for outdoor sample ($P = 0.999$). However, the mean PCS-3 scores were significantly different across study sites for all three sample types ($P < 0.001$).

CLUSTERING SUBJECTS

A total of seven clusters of study subjects with similar personal $PM_{2.5}$ elemental composition were identified for the pooled RIOPA dataset. The number of individuals per cluster was small, between 10 (Cluster 7) and 38 (Cluster 1). Overall, the proportion of individuals from each RIOPA study site significantly differed ($P < 0.01$) by cluster (Table 5), suggesting that the pattern of personal exposure to elements varied by study site. In particular, clusters 1

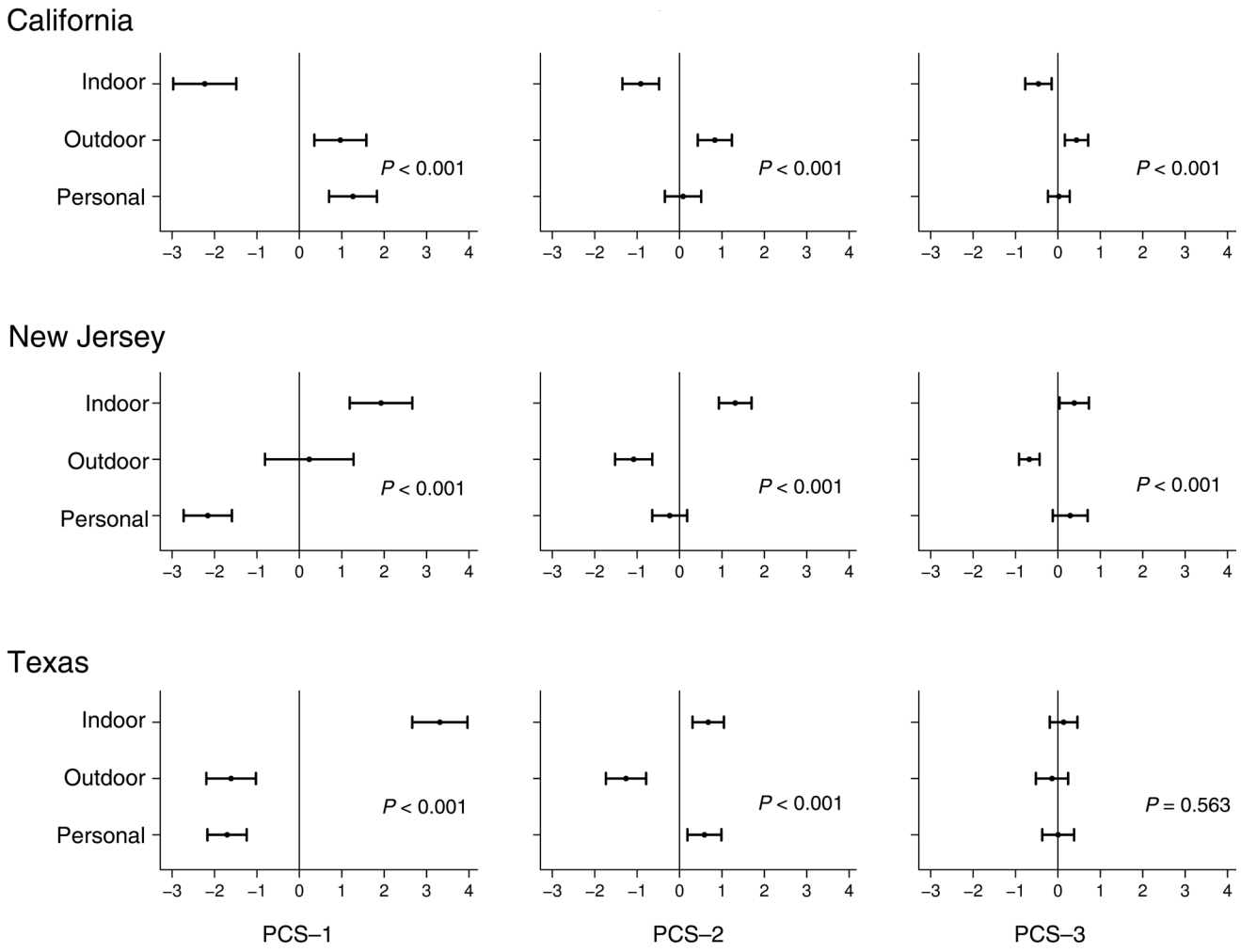

Figure 3. Principal component scores (PCSs) and 95% confidence intervals for the average PC score 1 (PCS-1), average PC score 2 (PCS-2), and average PC score 3 (PCS-3). Comparing sample type by study site.

and 5 are made up of predominantly of participants from California, whereas cluster 2 is chiefly made up of participants from Texas. When the data were stratified by study site, five clusters of study participants were identified in California, three clusters in Texas, and three in New Jersey, suggesting less variability between personal $PM_{2.5}$ elemental composition within individual study sites (results by study site not shown).

In order to visualize the pattern of elements in personal $PM_{2.5}$ for each of the clusters, we calculated *Z*-scores for each element across the seven clusters (Figure 5). In general, cluster 4 had the highest personal concentrations of elements compared with other clusters, and cluster 2 also had higher-than-average personal concentrations for most elements, although not as extreme as those of cluster 4. Cluster 5 had the lowest personal concentrations of elements, and although not as pronounced, cluster 6 also had lower personal concentrations. Clusters 1 and 7 had, overall, average personal elemental concentrations, with a mixture of slightly above or below average elemental concentrations. Cluster 3 was unique, with high concentrations of S, Sb, Se, and V and low concentrations of Al, Ba, and Cl.

In order to describe the concentrations of elements in indoor and outdoor $PM_{2.5}$, we again used heatmaps

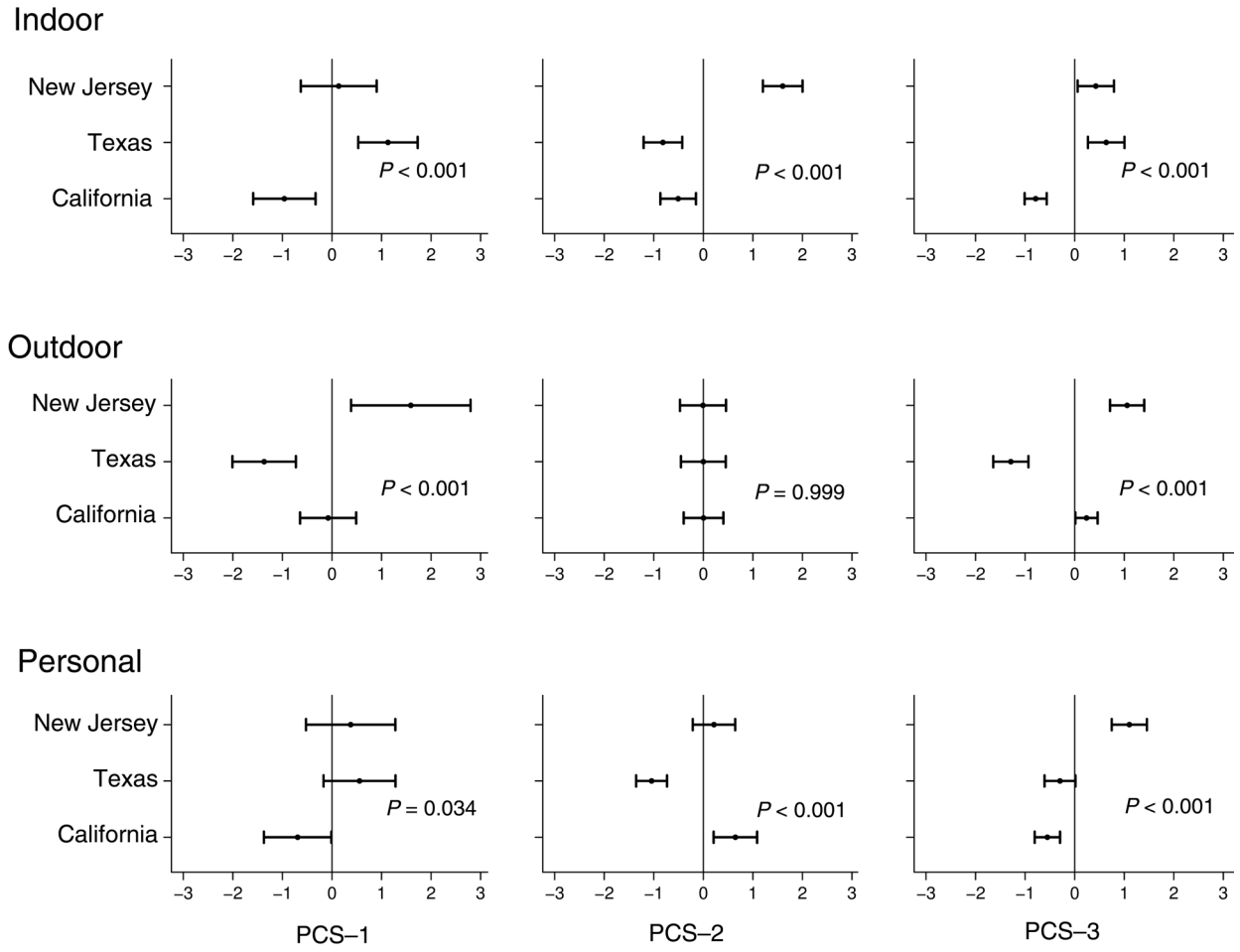

Figure 4. PCSs and 95% confidence intervals for the average PC score 1 (PCS-1), average PC score 2 (PCS-2), and average PC score 3 (PCS-3). Comparing study locations by sample type.

Table 5. Summary of Membership by State for Each Cluster

Cluster	*N*	California *n* (%)	Texas *n* (%)	New Jersey *n* (%)
1	38	24 (63%)	3 (8%)	11 (29%)
2	41	13 (32%)	24 (59%)	4 (10%)
3	12	7 (58%)	1 (8%)	4 (33%)
4	20	4 (20%)	6 (30%)	10 (50%)
5	14	9 (64%)	2 (14%)	3 (21%)
6	33	6 (18%)	14 (42%)	13 (40%)
7	10	5 (50%)	2 (20%)	3 (30%)

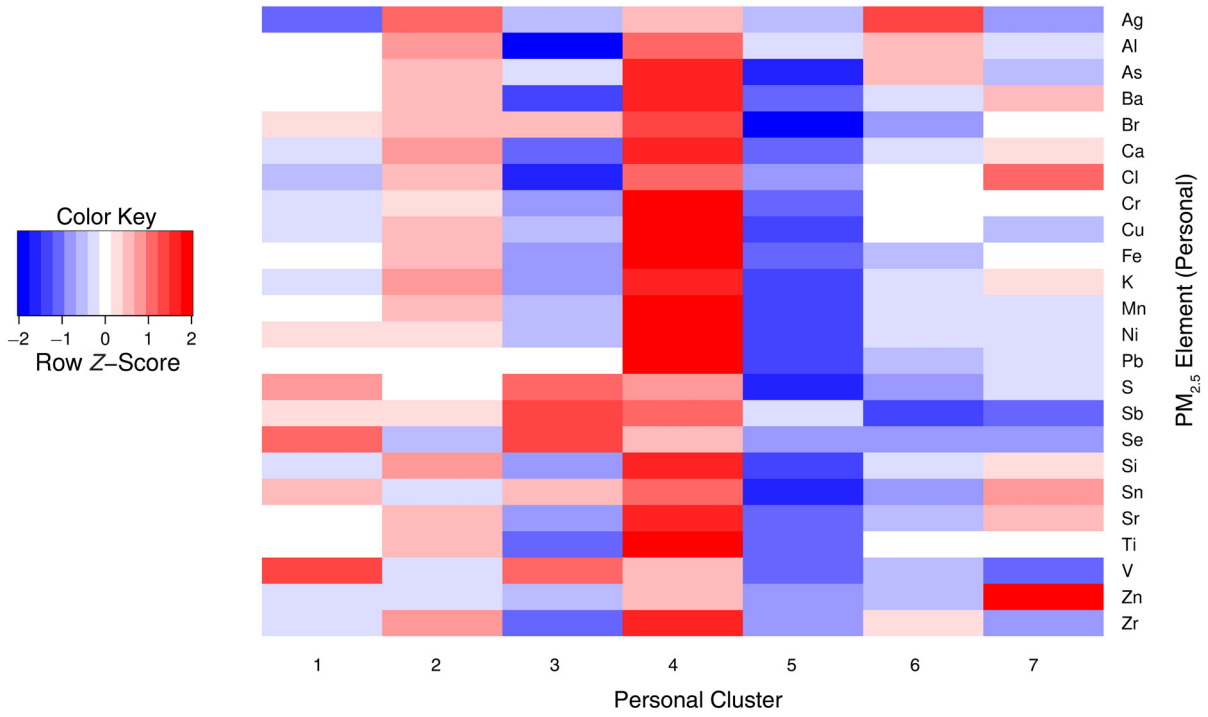

Figure 5. Heatmap depicting the standardized elemental concentrations (*Z*-scores) in personal $PM_{2.5}$ by cluster membership. Darker blue indicates concentrations less than the mean for each element; darker red indicates concentrations greater than the mean.

(Figure 6 and Figure 7). When comparing the personal elemental concentrations of the personal clusters with the indoor and outdoor concentrations found in the personal clusters, the most striking difference was in cluster 4. This cluster showed the highest concentration of personal $PM_{2.5}$ elements and yet contained roughly average concentrations of indoor and outdoor $PM_{2.5}$ elements. As expected, clusters 5 and 6 did contain below-average concentrations of $PM_{2.5}$ elements. Clusters 1 and 3 contained above-average concentrations for both indoor and outdoor $PM_{2.5}$ elements even though these clusters had about average concentrations of personal $PM_{2.5}$ elements. The differences in these sample-type concentrations based on the personal clusters clearly show that using indoor and outdoor samples can be inaccurate when predicting personal elemental $PM_{2.5}$ concentrations. The difference is most pronounced for high and average exposures. Low exposures to elements in $PM_{2.5}$ did not seem to change as much between personal, indoor, and outdoor samples.

In terms of overall $PM_{2.5}$ mass, the same general patterns held true for the clusters as it did for the elemental concentrations (Table 6). Clusters 2 and 4 had the highest personal $PM_{2.5}$ mass, whereas clusters 3, 5, and 6 had the lowest personal $PM_{2.5}$ mass, and clusters 1 and 7 had an about average amount of personal $PM_{2.5}$ mass. Clusters 5 and 6 had the lowest amounts of indoor and outdoor $PM_{2.5}$ mass, and cluster 3 had the highest amounts of indoor and outdoor $PM_{2.5}$ mass.

Personal and Home Characteristics Associated with Cluster Membership

We also examined whether personal and home characteristics differed significantly between the seven clusters. Of the personal and home characteristics examined, 13 home characteristics, four geographic characteristics, and two descriptive characteristics were significantly associated with cluster membership (Table 6). Specifically, the air-exchange rate was highest in clusters 1, 3, and 4. The census-tract elevation varied widely across clusters, being highest in cluster 5 (89.9 m) and lowest in cluster 7 (3.5 m). Cluster 7 was made up of people living in largely undeveloped land; it had by far the lowest census-tract population density, census-tract housing density, and percentage of land developed at a high density. Sex also

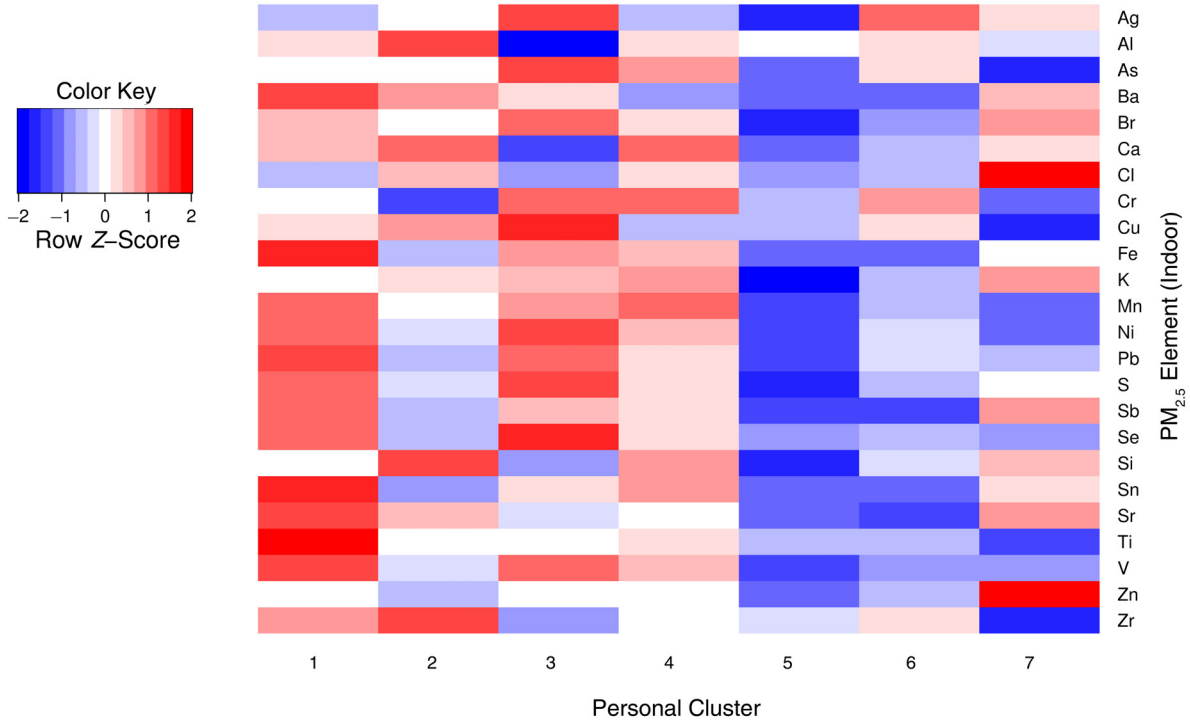

Figure 6. Heatmap depicting the standardized elemental concentrations (Z-scores) in indoor $PM_{2.5}$ by cluster membership. Darker blue indicates concentrations less than the mean for each element; darker red indicates concentrations greater than the mean.

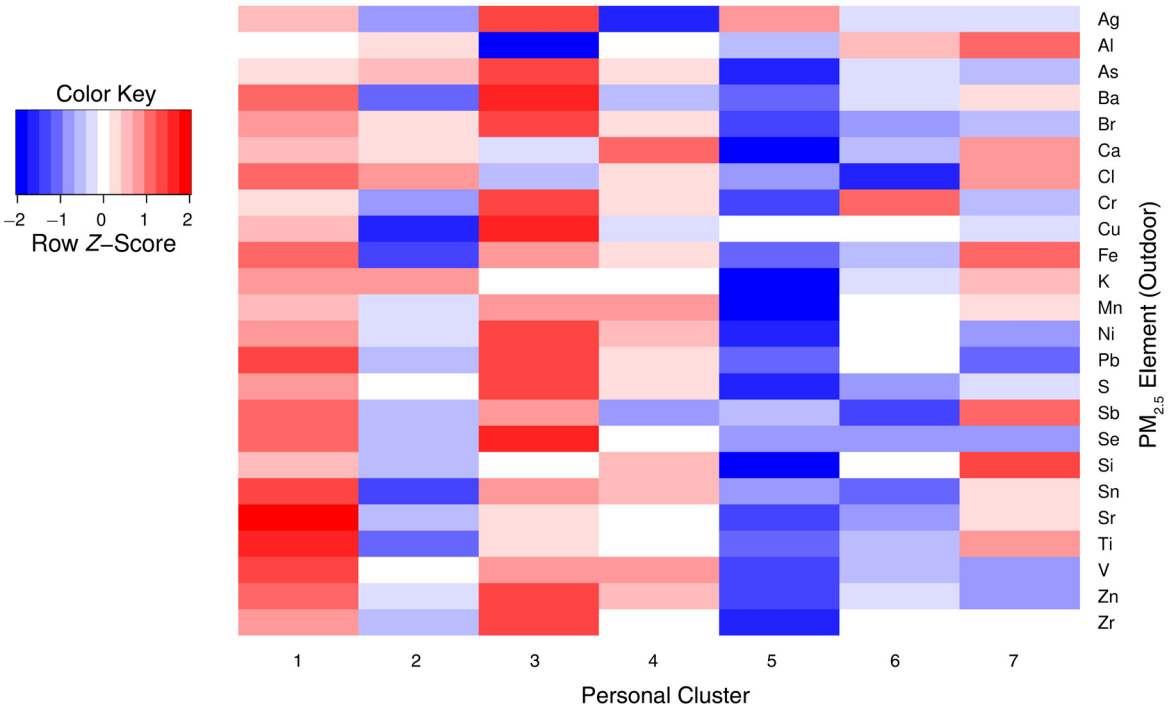

Figure 7. Heatmap depicting the standardized elemental concentrations (Z-scores) in outdoor $PM_{2.5}$ by cluster membership. Darker blue indicates concentrations less than the mean for each element; darker red indicates concentrations greater than the mean.

Table 6. $PM_{2.5}$ Mass and Personal and Home Characteristics by Cluster Membership[a,b]

	Cluster							
	1	2	3	4	5	6	7	*P* value
Indoor Elemental Concentration Characteristics	High	Average	High	Average	Very Low	Low	Average	
Outdoor Elemental Concentration Characteristics	High	Average	High	Average	Very Low	Low	Average	
Personal Elemental Concentration Characteristics	Average	High	High S, Sb, Se, V and low Al, Ba, Cl	Very High	Very Low	Low	Average	
N	38	41	12	20	14	33	10	
Indoor $PM_{2.5}$ ($\mu g/m^3$)	21.9	19.7	23.2	21.9	7.4	14.3	18.8	< 0.001
Outdoor $PM_{2.5}$ ($\mu g/m^3$)	25.7	15.8	26.9	19.1	9.3	12.9	13.3	< 0.001
Personal $PM_{2.5}$ ($\mu g/m^3$)	31.8	42.9	27.7	79.7	17.9	27	37	< 0.001
Air-exchange rate	1.36	0.93	1.14	1.2	0.99	0.7	0.89	0.087
Census-tract elevation (meters)	34.52	17.17	45.53	11.81	89.92	13.73	3.48	< 0.001
Census-tract population density (per sq. km land)	2979.96	1311.91	4063.31	1684.81	2439.6	2415.32	8.07	0.048
Census-tract housing density (per sq. km land)	1052.63	467.72	1479.3	594.86	879.07	924.19	7.02	0.063
Census-tract % high density developed	10.13	9.7	14.18	19.91	4.1	12.2	2.17	0.027
Fraction of study time spent indoors at home	0.62	0.68	0.62	0.66	0.64	0.71	0.7	0.382
Female (%)	19 (54%)	33 (86%)	7 (70%)	15 (83%)	9 (69%)	24 (80%)	4 (50%)	0.03
Year of birth (year)	1956	1956	1953	1949	1951	1951	1946	0.485
Self-identified as Caucasian (%)	8 (22%)	22 (56%)	4 (40%)	5 (27%)	10 (76%)	15 (48%)	4 (44%)	0.01
Education above high school (%)	30 (85%)	28 (75%)	9 (90%)	15 (83%)	13 (100%)	23 (76%)	6 (75%)	0.48
Household yearly income greater than $25K (%)	16 (50%)	19 (52%)	5 (50%)	9 (56%)	9 (75%)	14 (53%)	3 (37%)	0.768
Wood flooring (%)	8 (26%)	4 (11%)	3 (30%)	4 (23%)	3 (33%)	9 (31%)	1 (14%)	0.513
Area rugs (%)	8 (26%)	5 (13%)	1 (10%)	2 (11%)	1 (11%)	7 (2%)	1 (14%)	0.691
Wall-to-wall carpet (%)	18 (60%)	29 (80%)	9 (90%)	11 (64%)	7 (77%)	14 (48%)	6 (85%)	0.053
Presence of basement (%)	9 (30%)	4 (11%)	1 (10%)	8 (47%)	2 (22%)	11 (37%)	2 (28%)	0.071
Nearby emissions (%)	10 (33%)	21 (58%)	1 (10%)	9 (53%)	1 (11%)	12 (41%)	2 (29%)	0.03
Gas station < 200 ft (%)	4 (13%)	3 (8%)	1 (10%)	4 (24%)	1 (11%)	7 (24%)	1 (14%)	0.618
Gas station < 1/2 mile (%)	17 (57%)	23 (64%)	9 (90%)	9 (53%)	8 (89%)	23 (79%)	4 (57%)	0.133
Restaurant < 200 ft (%)	6 (20%)	5 (14%)	3 (30%)	1 (5%)	1 (11%)	5 (17%)	1 (14%)	0.751
Restaurant < 1/2 mile (%)	12 (40%)	20 (56%)	5 (50%)	12 (71%)	8 (89%)	19 (66%)	5 (71%)	0.117
Open windows (%)	32 (91%)	24 (62%)	8 (80%)	11 (61%)	12 (92%)	22 (71%)	7 (78%)	0.041
Air conditioning (%)	22 (63%)	21 (54%)	9 (90%)	11 (61%)	11 (85%)	21 (68%)	4 (44%)	0.183
Fireplace (%)	13 (37%)	11 (28%)	5 (50%)	3 (17%)	6 (46%)	4 (13%)	0 (0%)	0.026
Car for transportation (%)	27 (77%)	33 (85%)	7 (70%)	16 (88%)	11 (85%)	24 (77%)	8 (89%)	0.816
Bus for transportation (%)	3 (9%)	1 (3%)	0 (0%)	1 (6%)	2 (15%)	3 (10%)	0 (0%)	0.566
Pumped gasoline (%)	4 (11%)	9 (22%)	2 (17%)	4 (20%)	2 (14%)	2 (6%)	0 (0%)	0.362

Table continues next page

[a] See Results section for description of the clustering of subjects (Specific Aim 2).

[b] Percentages may not correspond to one another because of differing numbers of missing values for each characteristic.

Table 6 (*continued*). $PM_{2.5}$ Mass and Personal and Home Characteristics by Cluster Membership[a,b]

	Cluster							
	1	2	3	4	5	6	7	*P* value
Use of deodorant, perfume, or hair spray (%)	19 (51%)	35 (85%)	12 (100%)	16 (80%)	11 (79%)	28 (85%)	5 (50%)	0.001
Use of air freshener (%)	12 (32%)	16 (39%)	4 (33%)	5 (25%)	3 (21%)	14 (42%)	4 (40%)	0.765
Traveled on roads (%)	27 (73%)	34 (83%)	7 (58%)	17 (85%)	11 (79%)	27 (82%)	5 (50%)	0.191
Inside with a smoker (%)	3 (8%)	1 (2%)	1 (8%)	3 (15%)	1 (7%)	0 (0%)	1 (10%)	0.356
In a vehicle with a smoker (%)	1 (3%)	1 (2%)	1 (8%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0.598
Swimming (%)	1 (3%)	1 (2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0.888
Used cleaning supplies (%)	10 (27%)	18 (44%)	1 (8%)	10 (50%)	4 (29%)	15 (45%)	0 (0%)	0.017
Open doors or windows at home (%)	30 (81%)	21 (51%)	9 (75%)	16 (80%)	5 (36%)	18 (55%)	7 (70%)	0.011
Took a bath (%)	7 (19%)	5 (12%)	1 (8%)	4 (20%)	3 (21%)	3 (9%)	2 (20%)	0.8
Took a shower (%)	31 (84%)	37 (90%)	12 (100%)	17 (85%)	12 (86%)	30 (91%)	8 (80%)	0.734
Used felt-tip markers (%)	13 (35%)	14 (34%)	4 (33%)	6 (30%)	3 (21%)	4 (12%)	4 (40%)	0.33
Burned candles, oil, or incense in house (%)	5 (14%)	8 (20%)	2 (17%)	3 (15%)	1 (7%)	9 (27%)	5 (50%)	0.142
Use of central air conditioning (%)	4 (11%)	16 (39%)	2 (17%)	2 (10%)	2 (14%)	6 (18%)	2 (20%)	0.048
Use of window-unit air conditioner (%)	7 (19%)	4 (10%)	2 (17%)	7 (35%)	1 (7%)	3 (9%)	0 (0%)	0.079
Use of ceiling or portable fan (%)	17 (46%)	21 (51%)	5 (42%)	6 (30%)	4 (29%)	11 (33%)	2 (20%)	0.36
Use of an exhaust fan (%)	16 (43%)	21 (51%)	5 (42%)	6 (30%)	5 (36%)	19 (58%)	2 (20%)	0.263
Use of central heating (%)	3 (8%)	6 (15%)	1 (8%)	3 (15%)	4 (29%)	7 (21%)	3 (30%)	0.42
Use of dishwasher (%)	5 (14%)	13 (32%)	3 (25%)	1 (5%)	6 (43%)	10 (30%)	4 (40%)	0.065
Use of a clothes washer (%)	19 (51%)	23 (56%)	7 (58%)	10 (50%)	7 (50%)	16 (48%)	5 (50%)	0.994
Use of a clothes dryer (%)	17 (46%)	19 (46%)	5 (42%)	9 (45%)	7 (50%)	15 (45%)	5 (50%)	1
Use of nail polish remover (%)	4 (11%)	5 (12%)	1 (8%)	2 (10%)	0 (0%)	4 (12%)	1 (10%)	0.923
Sprayed pesticides (%)	4 (11%)	4 (9%)	1 (8%)	1 (5%)	1 (7%)	2 (6%)	1 (10%)	0.986
Vacuuming the house (%)	15 (41%)	20 (49%)	5 (42%)	8 (40%)	7 (50%)	13 (39%)	3 (30%)	0.932
Sweeping the house (%)	21 (57%)	23 (56%)	5 (42%)	12 (60%)	7 (50%)	19 (58%)	2 (20%)	0.414
Dusting the house (%)	12 (32%)	15 (37%)	5 (42%)	6 (30%)	3 (21%)	14 (42%)	2 (20%)	0.742
Use of cleaning solutions (%)	15 (41%)	20 (49%)	6 (50%)	10 (50%)	5 (36%)	18 (55%)	4 (40%)	0.874
Gardening (%)	6 (16%)	7 (17%)	0 (0%)	5 (25%)	3 (21%)	6 (18%)	1 (10%)	0.678
Grilling or frying inside the house (%)	15 (41%)	20 (49%)	3 (25%)	12 (60%)	7 (50%)	15 (45%)	3 (30%)	0.5
Motor vehicle parked in attached garage (%)	4 (11%)	7 (17%)	2 (17%)	2 (10%)	5 (36%)	6 (18%)	1 (10%)	0.458
Motor vehicle started in attached garage (%)	4 (11%)	6 (15%)	2 (17%)	2 (10%)	5 (36%)	6 (18%)	2 (20%)	0.473
Diesel vehicle parked around house (%)	13 (37%)	23 (58%)	4 (36%)	10 (53%)	6 (43%)	10 (33%)	3 (30%)	0.355
Nearby or operated a diesel vehicle (%)	19 (51%)	9 (23%)	5 (42%)	5 (25%)	3 (23%)	6 (20%)	2 (20%)	0.056

[a] See Results section for description of the clustering of subjects (Specific Aim 2).

[b] Percentages may not correspond to one another because of differing numbers of missing values for each characteristic.

differed across clusters; clusters 1 and 7, for example, had more males than the other clusters. Cluster 1 had the lowest percentage of white participants (23%); cluster 5 had the highest percentage of white participants (76%).

The presence of wall-to-wall carpeting in the home differed according to cluster, with cluster 6 having a lower amount of homes in that category compared with the other clusters. Clusters 4 and 6 also had more homes with basements (47% and 38%, respectively).

The presence of open windows during a home visit was low for both clusters 2 and 4 (62% and 61%, respectively). However, the same measure, when self-reported by the study participant, did not show the same patterns. Also differing by cluster was the percentage of homes with a fireplace. Clusters 4, 6, and 7 had fewer homes with fireplaces than did the homes in the other clusters. Air-conditioning use (both central and window unit) differed by cluster too. Homes in cluster 2 were more likely to have used central air conditioning during the study, and homes in cluster 4 were much more likely to have window-unit air conditioning during the study. Finally, at least a fifth of the subjects in each cluster had been near or operated a diesel vehicle during the study period, but clusters 1 and 3 had by far the largest percentages (51% and 42%, respectively).

LINEAR REGRESSION

Our initial approach included the development of 24 linear regression models corresponding to each element in the personal $PM_{2.5}$ samples. Indoor and outdoor $PM_{2.5}$ elemental concentrations of each element were used as covariates to model the personal $PM_{2.5}$ concentration of that same element (Equation 2). Results of these linear models for the pooled RIOPA dataset are shown in Table 7. The variability explained by indoor and outdoor elemental concentrations ranged from 0.03 (Sb) to 0.80 (S), although in general, concentrations of elements in the indoor $PM_{2.5}$ were significant predictors of personal concentrations. Only eight elements in outdoor $PM_{2.5}$ (Ag, As, Cl, S, Se, Sn, Ti, and Zn) were significantly associated ($P < 0.05$) with personal $PM_{2.5}$ in models that also included indoor concentration as a predictor.

In order to examine whether these results were modified by time spent indoors, we repeated the 24 regression models and included both the reported fraction of time spent indoors during the sampling period and the interaction between time spent indoors and the indoor concentration of each element (Equation 3). As seen in Table 7, time spent indoors significantly modified the associations between indoor concentrations of Ca, Fe, Ti, and Zn and personal exposure to these same elements.

Does the Inclusion of Indoor Elemental Concentrations Improve the Prediction of Personal Exposure Compared with Outdoor Elemental Concentrations Only?

Our preliminary modeling approach described above was conducted using both indoor and outdoor elemental concentrations to predict personal exposure for each element in separate models. However, these data are not typically available in epidemiologic studies. Thus, based on models containing information about all $PM_{2.5}$ elements, the results of our Outdoor Only regression model (Equation 4) were compared with those from a model containing both outdoor and indoor elemental concentrations (the Outdoor + Indoor model, described in Equation 5). As shown in Figure 8, the Outdoor + Indoor model's CV R^2 was less than 0.3 for most elements. The inclusion of indoor elemental concentrations significantly increased the CV R^2 for nine elements: Ba, Ca, Cl, Cu, K, Sn, Sr, V, and Zn. However, the CV R^2 remained less than 0.5 for all of these except S and V. For the remaining 14 elements, the inclusion of indoor $PM_{2.5}$ elemental composition did not significantly improve the CV R^2.

Does the Inclusion of Personal and Home Characteristics Further Improve the Prediction of Personal Exposure Compared with Outdoor and Indoor Elemental Concentrations?

Of all the personal and home characteristics considered for model inclusion, 23 characteristics (e.g., time spent indoors, the presence of a basement, nearby industrial emissions, existence of a kerosene heater, existence of a gas heater, having a wood stove, having a fireplace, the air-exchange rate, the elevation of the centroid of the census tract, census-tract population, census-tract housing density, and the percentage of highly developed land in census tract) met our screening criteria (i.e., being univariately associated with $\geq$ six elements in personal $PM_{2.5}$). These personal and home characteristics were therefore included in a model (i.e., the Outdoor + Indoor + Home and Personal Characteristics model) that also contained indoor and outdoor elemental concentrations. The results of the model were compared with those of the Outdoor + Indoor model (Figure 8). The CV R^2 significantly increased only for Pb and caused a decrease (but not significantly different) in CV R^2 in 16 of the elements (Ag, Al, As, Ba, Br, Ca, Cr, K, S, Sb, Se, Si, Sn, Sr, V, and Zn).

RANDOM FOREST ANALYSIS

In addition to the linear regression approach for addressing the questions of whether indoor $PM_{2.5}$ data and other personal and home characteristics helped to predict elemental concentrations in personal exposure better than

outdoor $PM_{2.5}$ data alone, we also pursued the machine-learning technique known as random forest analysis. We used this technique to predict personal $PM_{2.5}$ exposure and quantitatively compared all three models using CV R^2. In our analyses, we used random forests to predict elemental concentrations in personal $PM_{2.5}$ based on all outdoor and indoor elemental concentrations as well as the personal and home characteristics listed in Table 4.

Does the Inclusion of Indoor Elemental Concentrations Improve the Prediction of Personal Exposure Compared with Outdoor Elemental Concentrations Only?

Figure 9 shows the CV R^2 for the Outdoor Only and Outdoor + Indoor random forest models as well as their 95% confidence intervals. The addition of indoor $PM_{2.5}$ concentrations did not significantly improve the random forest model's predictions of personal $PM_{2.5}$ concentrations for any of the elements.

Does the Inclusion of Personal and Home Characteristics Further Improve the Prediction of Personal Exposure Compared with Outdoor and Indoor Elemental Concentrations?

Figure 9 also shows the CV R^2 for the Outdoor + Indoor and the Outdoor + Indoor + Personal and Home Characteristics random forest models as well as their confidence intervals. The addition of personal and home characteristics did not significantly improve the random forest model's predictions of concentrations in personal $PM_{2.5}$.

Table 7. Association Between Elemental Concentrations in Indoor and Outdoor $PM_{2.5}$ and in Personal $PM_{2.5}$ (from Initial Models)[a]

	Model with Indoor and Outdoor Concentrations Only (Equation 2)			Model with Indoor, Outdoor, and Time Spent at Home and Interaction Between Indoor Concentrations and Time Spent at Home (Equation 3)				
Element	R^2	$\beta_{(indoor)}$	$\beta_{(outdoor)}$	R^2	$\beta_{(indoor)}$	$\beta_{(outdoor)}$	$\beta_{(time\ spent\ at\ home)}$	$\beta_{(indoor\ time \times spent\ at\ home)}$
Ag	0.09	**0.24**	**0.30**	0.10	0.16	**0.26**	0.32	0.14
Al	0.33	**0.56**	−0.03	0.33	0.26	−0.04	−0.87	0.43
As	0.21	**0.40**	**0.22**	0.21	0.39	**0.20**	0.21	−0.01
Ba	0.06	**0.29**	−0.06	0.07	0.65	−0.08	1.35	−0.57
Br	0.42	**0.65**	0.02	0.42	**0.76**	0.01	0.10	−0.15
Ca	0.31	**0.63**	0.07	0.35	−0.46	0.04	**−6.82**	**1.74**
Cl	0.29	**0.44**	**−0.08**	0.32	0.39	−0.08	0.79	0.05
Cr	0.36	**0.55**	0.14	0.38	0.27	0.13	**−0.89**	0.44
Cu	0.28	**0.63**	−0.09	0.28	0.30	−0.11	−1.10	0.47
Fe	0.15	**0.42**	−0.03	0.18	−0.31	−0.05	**−4.85**	**1.06**
K	0.36	**0.54**	0.01	0.37	0.14	0.00	−2.30	0.62
Mn	0.20	**0.54**	0.02	0.20	0.45	0.02	−0.04	0.13
Ni	0.15	**0.29**	0.15	0.16	0.45	0.16	−0.16	−0.27
Pb	0.21	**0.53**	0.09	0.22	−0.02	0.06	−1.32	0.83
S	0.80	**0.72**	**0.10**	0.81	**0.58**	**0.11**	−1.35	0.23
Sb	0.03	**0.20**	0.01	0.03	0.30	0.00	−0.13	−0.18
Se	0.62	**0.53**	**0.22**	0.62	**0.64**	**0.22**	−0.08	−0.19
Si	0.14	**0.49**	−0.04	0.14	−0.08	−0.05	−3.08	0.77
Sn	0.44	**0.53**	**0.14**	0.44	**0.77**	**0.15**	0.75	−0.33
Sr	0.29	**0.84**	−0.11	0.29	0.61	−0.09	−0.19	0.35
Ti	0.10	**0.33**	**−0.14**	0.14	−0.67	**−0.15**	**−3.02**	**1.54**
V	0.75	**0.77**	0.11	0.75	**0.56**	0.13	−0.54	0.28
Zn	0.46	**0.64**	**−0.24**	0.49	−0.07	**−0.18**	**−2.63**	**0.89**
Zr	0.15	**1.09**	0.09	0.16	**1.95**	0.09	0.94	−1.29

[a] **β** values in **bold** indicate $P < 0.05$

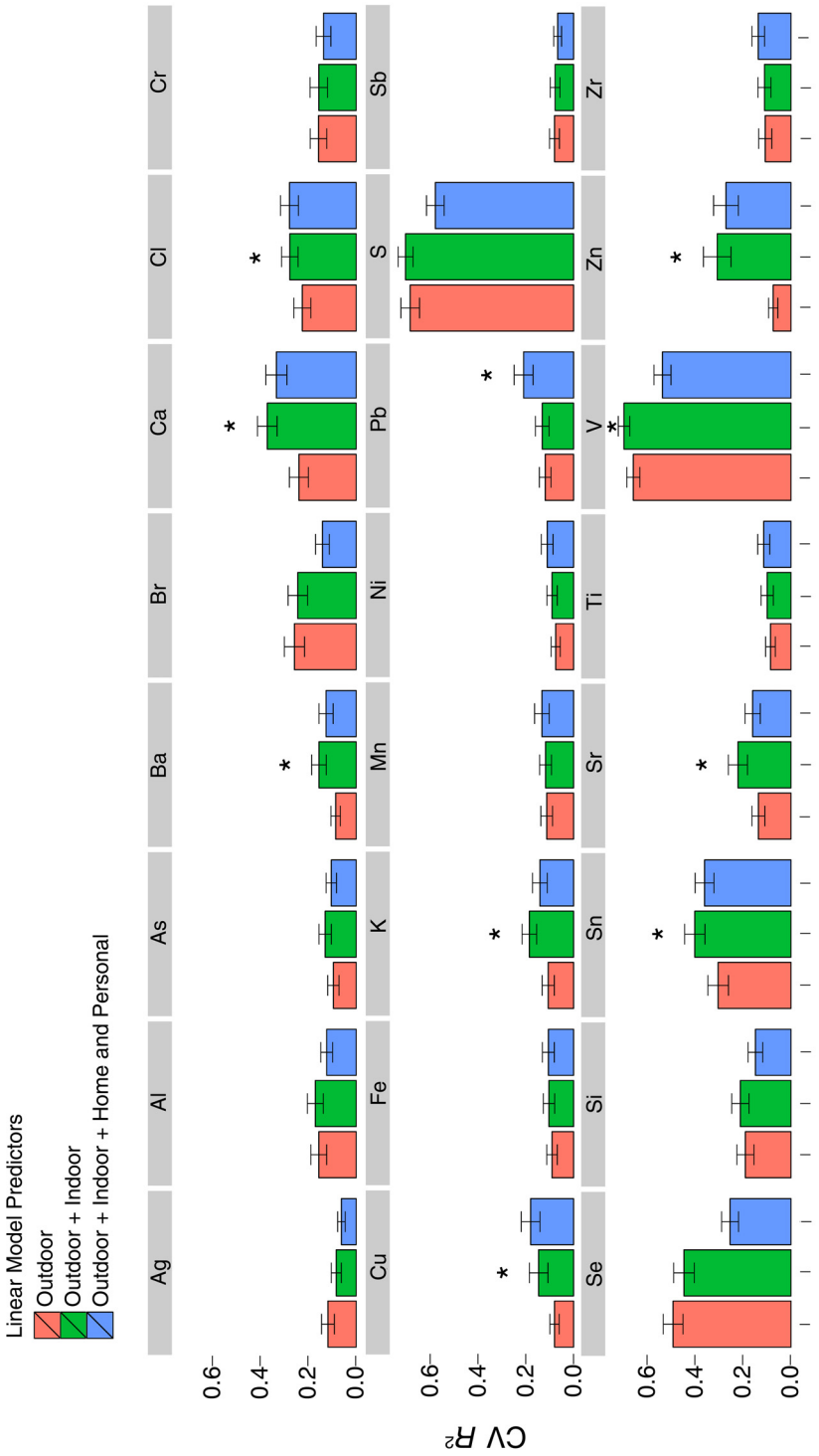

Figure 8. Comparison of the three personal $PM_{2.5}$ predictions for all three linear regression models for each element. Stars above the green bars represent those elements for which the CV R^2 increased when indoor concentrations were added to the outdoor concentrations. Stars above the blue bars represent those elements for which the CV R^2 increased when home and personal characteristics were added to the indoor and outdoor concentrations.

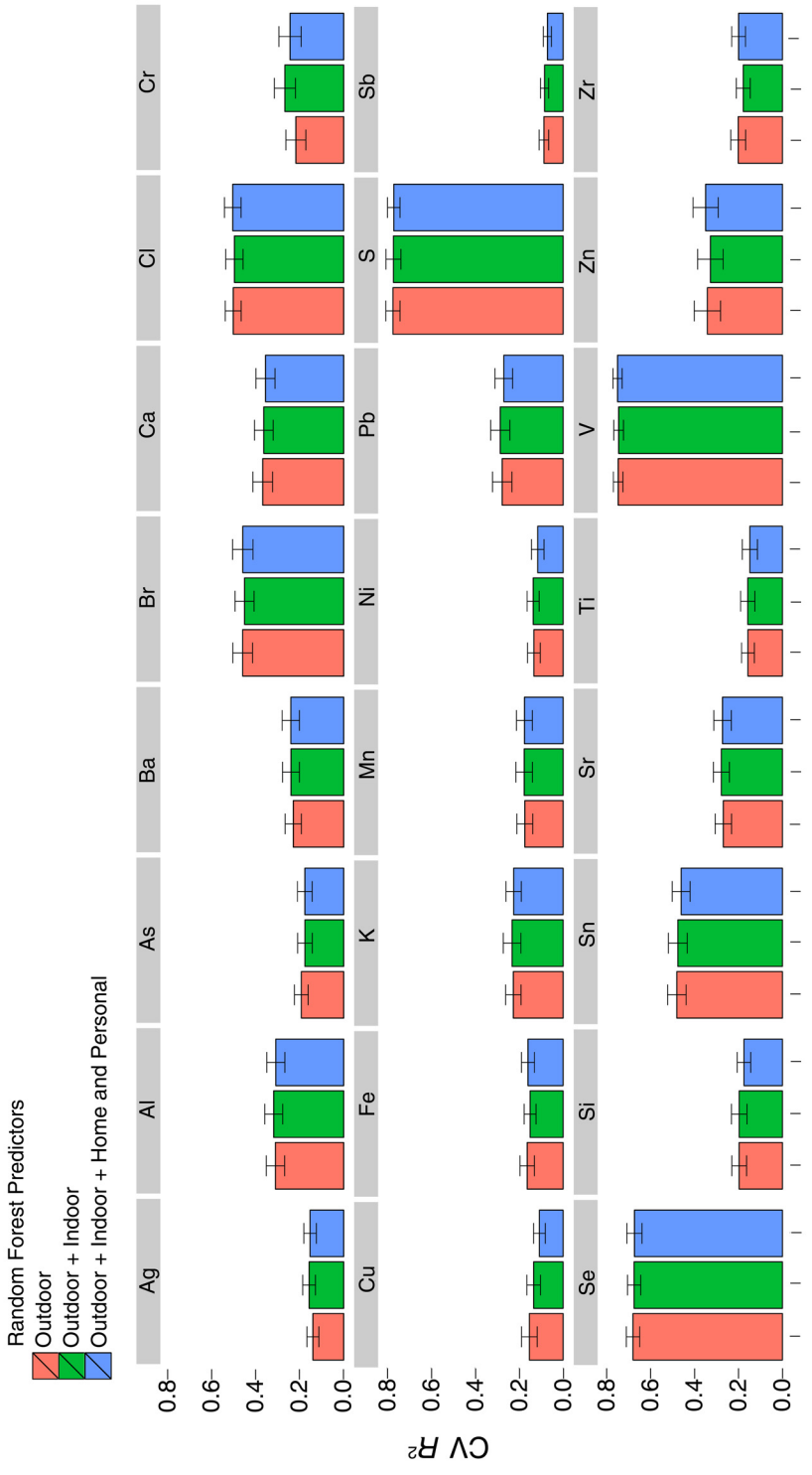

Figure 9. Comparison of the three personal $PM_{2.5}$ predictions for all three random forest models for each element.

DISCUSSION AND CONCLUSIONS

The goals of this study were (1) to explore the relationships between elemental concentrations in concurrently sampled indoor, outdoor, and personal $PM_{2.5}$, (2) to determine if groups of individuals with similar exposure to elements in personal $PM_{2.5}$ samples could be identified and, if so, elucidate personal and home factors associated with these groups, and (3) to examine the appropriateness of outdoor measurements of elemental components of $PM_{2.5}$ as surrogates for personal exposure. In addition, we sought to determine if concentrations of elements in indoor $PM_{2.5}$ samples markedly increased the prediction of personal exposure and, if so, for which elements. In order to accomplish these goals, we combined traditional statistical methods with novel approaches, including model-based clustering and machine-learning techniques, to address our primary research questions.

The results of our analyses showed that personal concentrations of $PM_{2.5}$ and most elements were significantly correlated with, though frequently exceeded, both indoor and outdoor measurements. As anticipated, indoor concentrations were in general more highly correlated with personal exposure than were outdoor concentrations, consistent with participants' report of spending 66% of their time indoors at home. Overall, for the combined RIOPA participants, the median ratio of I/P and O/P concentrations for most elements was consistently less than or approximately equal to one. These results suggest that there were other sources of personal exposure besides those in the home and in the neighborhood air. Exposures occurring in these other microenvironments, which include transit locations, are likely to have resulted in higher concentrations of elemental components of $PM_{2.5}$ than in the participants' homes, but these other microenvironments were not measured.

Few studies have examined the heterogeneous mixture of elements in $PM_{2.5}$ and the differences in these elemental mixtures across geographic locations. To our knowledge, this is the first study to determine whether the mixture of elements in indoor, outdoor, and personal $PM_{2.5}$ varied within a single study site. By using PCA, we found that the mixture of elements in indoor, outdoor, and personal $PM_{2.5}$ samples, as a whole, varied significantly by study site, reflecting different emission sources and personal activities in the three geographic locations. In addition, within an urban area, the elemental composition of $PM_{2.5}$ across indoor, outdoor, and personal samples for the same individual was significantly different, suggesting different composition of $PM_{2.5}$ and sources of elements in each microenvironment.

Using model-based clustering techniques, we were able to identify groups of individuals whose personal $PM_{2.5}$ composition were similar to one another. Not surprisingly, we identified more clusters (n = 7) of similar personal $PM_{2.5}$ composition when analyzing the pooled RIOPA dataset than when clustering individuals by study site (n = 3–5), because it is known that there are different air pollution sources and personal activities in the three geographic locations. We were able to identify home and personal characteristics associated with the clusters of RIOPA participants, such as the presence of a basement or fireplace in the home, presence of open windows, presence of air conditioning, and census-tract-level variables such as the elevation and proportion of developed land. Our analysis also revealed subsets of RIOPA participants whose personal exposure to $PM_{2.5}$ and elemental concentrations were significantly higher than both their indoor and outdoor concentrations or vice versa.

Using our initial linear regression models, we found that indoor measurements were more significantly associated with personal concentrations for most elements than were outdoor measurements, consistent with the results from our O/P and I/P correlation analyses. Indeed, when including indoor concentrations in the linear regression models, outdoor concentrations were significantly associated with personal exposure for only Ag, As, Cl, S, Se, Sn, Ti, and Zn. Participant report of time spent at home was a significant modifier of the association between indoor and personal concentrations for four of the elements. Results showed that inclusion of indoor concentrations improved the models for prediction of personal concentrations of nine elements, but home and personal characteristics did not improve the models for personal concentrations. A more detailed discussion of each of our analyses and the implications of this research for epidemiologic studies are provided below.

CORRELATIONS AND RATIOS OF ELEMENTAL CONCENTRATIONS IN INDOOR–PERSONAL AND OUTDOOR–PERSONAL SAMPLES

There have been few studies of concurrent exposures to the elemental components of $PM_{2.5}$ in indoor and outdoor samples with respect to the elemental composition of $PM_{2.5}$ in personal samples. A similar approach by Janssen and colleagues (2005) in a study of elderly subjects with cardiovascular disease in two European cites found higher correlations ($r > 0.7$) between $PM_{2.5}$ in both outdoor and indoor air to personal samples than those found in the RIOPA dataset. In the three RIOPA study sites, the highest I–P correlation for $PM_{2.5}$ ($r = 0.65$) was found in Texas. All other $PM_{2.5}$ correlations in our study were between 0.09

and 0.47. These differences may be a result of varying time–activity patterns of the study participants, housing factors, nearby sources, or other unknown factors. For example, subjects enrolled in the Janssen and colleagues (2005) study were seniors who may have been less mobile than RIOPA participants and spent most of their time indoors and in their neighborhood.

With respect to elemental components of $PM_{2.5}$, we observed, similarly to Janssen and colleagues (2005), that concentrations of S, Se, and V were highly correlated ($r >$ 0.65). This was likely because there are few indoor sources of these elements and limited outdoor spatial variation (Janssen et al. 2005; Martuzevicius et al. 2004). We observed much lower O–P correlations for Zn (r = 0.14–0.29), Fe (r = 0.01–0.30), K (r = −0.04–0.37), and Ca (r = 0.05–0.29) (Figures A.1–A.3, Appendix A, available on the Web) than did Janssen and colleagues (2005). I–P correlations of these elements, although lower than those reported by Janssen and colleagues (2005), were higher than the O–P correlations in all three RIOPA study sites. Indoor activities (including cleaning, humidifier use, or cooking) may have been in part responsible for the differences. Another potential explanation for the observed differences may be the sampled populations and, specifically, the oversampling of women and homes with nearby outdoor traffic, industrial activities, or other sources of exposure in the RIOPA study. Overall, the correlations between elemental concentrations in I–P and O–P samples in the RIOPA study, though significant, were lower than those observed in similar studies with different populations. Elements considered to be primarily of ambient origin with few indoor sources (S, Se, and V) had the highest observed correlations in both I–P and O–P samples. Of note, O–P correlations of the concentration of many elements, including As, Mn, Ni, and Zn, previously identified in epidemiologic studies as being associated with increased morbidity and mortality, were lower than the I–P correlations, and these correlations varied by study site.

The results of our correlation analyses were in general consistent with those of previous studies and demonstrated that, across the three study sites, personal exposures to $PM_{2.5}$ and some elemental components correlated with both indoor and outdoor stationary measurements. Also, higher I–P correlation varied by site, with higher correlations found in Texas. These results were consistent with the time–location pattern — that is, the Texas participants spent more time indoors than those in California and New Jersey did. Additionally, average air-exchange rates were lower in Texas than those in California. Across all sampling sites, correlations between indoor concentrations of all elements, with the exception of Ag, were better correlated with personal exposure than were outdoor concentrations (Figure 1). Some interesting trends emerged, however, when examining the I–P and O–P correlations by study site. In California, the I–P and O–P concentration correlations were more similar to each other than were the correlations of concentrations in either New Jersey or Texas, where the I–P concentration correlations were greater than O–P correlations for many elements. One potential reason for this finding is that air-exchange rates were higher and the percentage of time spent indoors was lower for participants in California compared with those in Texas and New Jersey. The consistent exceptions to this trend were for S and V, which both had similar I–P and O–P correlations, likely reflecting the lack of indoor sources of these elements.

COMPARING ELEMENTAL HETEROGENEITY USING PRINCIPAL COMPONENT ANALYSIS

The results of our PCAs supported two conclusions: (1) that the composition of elements in $PM_{2.5}$ found in indoor, outdoor, and personal samples varied significantly by study site (Figure 4) and (2) that the overall composition of elements in $PM_{2.5}$ varied significantly by sample type (indoor, outdoor, and personal) at each of the study sites (Figure 3). It is not surprising that the mixture of elements in indoor, outdoor, and personal $PM_{2.5}$ differed by study site, given that the major contributing sources of outdoor (and likely, indoor) $PM_{2.5}$ varied across the three study sites (Weisel 2005a). We thus expected the composition of $PM_{2.5}$ to vary by study site, regardless of sample type. It is somewhat surprising, however, that the overall composition of elements in $PM_{2.5}$ was significantly different within a given study site for indoor, outdoor, and personal samples (Figure 3). Although we have been unable to the identify individual elements that may be responsible for the observed differences in composition, by examining the loading factors for the PCs by study site we could identify the elements whose loading factors were the primary contributors to the PCS, suggesting that the mixtures of these elements (and their sources) were responsible for the observed differences in PCS. Differences in the concentrations of the mixtures of highly loaded elements between indoor, outdoor, and personal samples were likely responsible for the significant differences in $PM_{2.5}$ composition we observed in PCS-1 (Figure 3). Many of these elements were also identified as being important predictors of elemental heterogeneity in outdoor $PM_{2.5}$ by Han and colleagues (2012). Potential outdoor sources of these elements include crustal minerals and the resuspended road dust, which may explain differences in these elements between indoor, outdoor, and personal samples in California. Many

of these elements, including Al, Ca, Cu, Si, and Zn, also have indoor sources, including personal care and cleaning products, which may have led to differences between personal exposure and outdoor and indoor $PM_{2.5}$ samples (Levy et al. 2010). In particular, cooking may be an important factor related to exposure to elements, including Al, Fe, Ca, and Si. These elements have been shown to be related to indoor combustion and especially to cooking (Levy et al. 2010). Although these elements were not highly loaded on PCs compared across sample sites, they were the most highly loaded on the PC1 and PC2 of indoor samples (Table B.1, Appendix B, available on the Web), suggesting that differences in these partly explain the variability in indoor measurements.

When comparing the PC results across sample types within study sites (Table B.1, Appendix B, available on the Web), the first PC for all three locations was not highly loaded by any elements. This is usually the case, because the first PC typically represents a weighted average, and the remaining PCs represent contrasts. PCS-1 significantly differed across sample types for all three locations. In New Jersey, all three sample types appeared to be different, suggesting that the overall average $PM_{2.5}$ elemental composition was different for each sample type, whereas in California and Texas, the outdoor and personal sample types were more closely related compared with the indoor sample type.

PC2 was primarily explained by S, Se, and V for all three locations and also by Br for California and New Jersey but not Texas. S is thought to come from coal combustion (Hammond, et al. 2008; Lall and Thurston 2006) but could also from diesel combustion (Spencer et al. 2006) related to traffic, given that these measurements were taken before the use of low-sulfur diesel fuel became prevalent. Se is also attributed to coal combustion (Hammond et al. 2008; Lee et al. 2008; Ogulei et al. 2006). V is mainly hypothesized to come from local or regional oil combustion and refinery activities (Hammond et al. 2008; Lall and Thurston 2006; Li et al. 2004; Qin et al. 2006). S, Se, and V are also the three elements pointed out earlier as having the highest I–P and O–P correlations and the I/P and O/P ratios closest to unity. Thus, PC2 can be thought of as representing mainly ambient $PM_{2.5}$ components of oil and coal combustion from largely outdoor sources such as refinery activities and industrial processes. PCS-2 was found to differ significantly among the indoor, outdoor, and personal sample types for all three locations (Figure 3). In California and New Jersey, the PCS for all three sample types were different, suggesting that oil and coal combustion contributed differently for all three sample types. This could be because exposures to these sources are common outside the home and because these particles are thought to be robust and associated with long-range transport, such that they could then spread throughout the home into personal airspace. However, in Texas, indoor and personal PCS-2 were very similar, whereas outdoor PCS-2 was different, suggesting that the contribution of oil and coal combustion was similar for indoor and personal exposures. This similarity could be because the hotter weather in Texas is associated with more time spent indoors and there is less exchange of air from outdoors to indoors.

PC3 in California was largely explained by Cl, Cr, and Ni. Cl and Cr are usually attributed to brake and tire dust as well as soil and road dust (Schauer et al. 2006). Thus PC3 in California can be thought of as largely driven by non-tailpipe traffic emissions. In California, PCS-3 differed for each of the indoor, outdoor, and personal sample types, again suggesting that road dust contributed differently to the elemental composition of each sample type. PC3 in New Jersey was most heavily loaded by As, Cl, K, and Zn. The domination by As and Cl again suggests roadway dust sources, but the more heavily loaded K and Zn suggest that PC3 in New Jersey was also associated with biomass and garbage burning (Lee et al. 2008; Qin et al. 2006; Rizzo and Scheff 2007) as well as with steel production (Hammond et al. 2008; Pekney et al. 2006; Rizzo and Scheff 2007). In our analyses, PCS-3 differed by sample type, but indoor and personal sample types were more similar compared with the outdoor sample type. This likely means that the burning of biomass and garbage as well as steel production contributed the most to outdoor $PM_{2.5}$ concentrations and not as much to indoor and personal concentrations. This is to be expected because most of the PM from biomass sources and steel production is usually from a few dominant sources and might not penetrate the indoor airspace of a home. Finally, PC3 in Texas was highly loaded by As, Br, Cl, and Sn. Although the contribution of road dust is indicated by As and Cl, the sources of PC3 that produce Br and Sn are not completely clear. The absence of a known dominant source for PC3 may explain why PCS-3 was not found to differ between indoor, outdoor, and personal sample types.

Our application of PCA differs from previous studies that used this approach to address the issues of multiple correlated exposures and to identify potential sources of chemical constituents in PM. These studies used PCA primarily as a method to identify sources of environmental pollutants based on the examination of constituents whose PC loading factors were elevated. For example, this method was used in Arizona to determine that soil and combustion sources were the primary contributors to $PM_{2.5}$ (Upadhyay et al. 2011) and in the Mexican state of

Guadalajara to examine differences in sources of $PM_{2.5}$ between locations with industrial and vehicular sources (Saldarriaga-Norena et al. 2009). PCA has also been used in Istanbul, Turkey, to identify components of PM (Onat et al. 2012). In contrast to these studies, our objective was to use PCs and the derived PCS to examine the overall heterogeneity of the mixture of elements in $PM_{2.5}$. This application was previously proposed by Han and colleagues (2012) to identify significant differences in the composition of metals in $PM_{2.5}$ between eight U.S. counties, but to our knowledge, this is the first application of PCS to examine the heterogeneity in elemental composition between indoor, outdoor, and personal $PM_{2.5}$ samples within geographic regions.

There are some limitations, however, to our PCA approach. Although PCA allows us to compare and test the compositional differences in $PM_{2.5}$ by sample type within a study site and also across study sites by sample type, we were unable to identify the specific elements or sources that primarily contributed to these differences. In contrast to previous studies, examining the loading factors for the PCs, both by study site and sample type, did not reveal noticeably elevated loading factors for most individual elements. Rather, multiple elements in most PCs had slightly higher than average weights, making it difficult to readily identify potential sources. This is likely a result of the RIOPA study design, which was intentionally conducted in cities with distinct sources of $PM_{2.5}$, resulting in multiple sources of elements whose contribution diluted the overall weights of the PCs. The strength of this approach, however, is its ability to identify significant differences in the $PM_{2.5}$ elemental composition, as a whole, between sample types within study locations. The implication of these findings is that, even within a given urban area, the mixture of elements to which individuals are exposed varies between indoor, outdoor, and personal samples. Future research investigating the joint effects of multiple simultaneous exposures to constituents in $PM_{2.5}$ should be aware that the mixture of elements in $PM_{2.5}$ of outdoor, and even indoor, samples does not necessarily reflect the mixture of elements to which an individual is actually exposed. As epidemiologic and statistical methods are developed to examine multiple exposures, researchers should consider not only the impact of exposure misclassification with respect to both estimated exposure levels, but also the impact of errors in estimating the mixtures of exposures themselves; given the significant differences in PCS between sample types within study sites, our PCA results suggest that the elemental composition of outdoor $PM_{2.5}$ does not accurately represent the overall mixture of elements to which individuals are exposed.

CLUSTERING SUBJECTS BASED ON THE ELEMENTAL COMPOSITION OF PERSONAL $PM_{2.5}$

To our knowledge, the current study is the first application of model-based clustering to identify groups of individuals based on their personal exposure to the mixture of elements in $PM_{2.5}$. Clusters were generated based on personal $PM_{2.5}$ exposures, and as such, each cluster may be considered a signature or pattern of personal exposure to the mixture of elements in personal $PM_{2.5}$. Because the RIOPA study sites and participants were chosen to maximize the differences in sources and types of exposure across the study sites, it was not surprising that a total of seven clusters from the combined RIOPA dataset were identified, with fewer clusters (three to five) found for individual sites.

By examining heat maps of the *Z* scores (Figure 5), elements that differentiate cluster membership may be identified. These may also suggest significant sources (or lack thereof) of personal exposure for individuals in each cluster. Using this approach, we identified seven subgroups of RIOPA participants distinguished by the concentration of elements in personal $PM_{2.5}$. Clusters 1 and 7 (Figure 5) may be considered the groups of participants whose exposure were, in general, the average for the RIOPA study. Participants in clusters 5 and 6 had lower than average exposure, particularly those in cluster 5. These results are also consistent with the concentrations of these elements in indoor and outdoor $PM_{2.5}$ for this cluster (Figures 6 and 7).

Cluster 3 was unique, with high concentrations of S, Sb, Se, and V and low concentrations of Al, Ba, and Cl. Because Al and Ba are associated with soil and road dust (Lee et al. 2008; Li et al. 2004; Qin et al. 2006; Rizzo and Scheff 2007; Zhao et al. 2006) and S, Se, and V are associated with oil and fuel combustion (Hammond et al. 2008; Lall and Thurston 2006), this pattern of concentrations suggested that subjects in this cluster were exposed to fuel or oil combustion without the usually associated exposure to traffic and road dust. These subjects, who otherwise experienced low concentrations of all elements in $PM_{2.5}$, may have been exposed to high levels of S, Se, and V from non-traffic sources of fuel combustion.

Interestingly, the clusters with high personal exposures (clusters 2 and 4) had approximately average exposures for all elements for both indoor and outdoor sample types. Cluster 1, which had about average personal exposures for all elements, had high exposures across most elements for both indoor and outdoor sample types. In contrast, cluster 7, which had average personal exposure for all elements, also showed average outdoor concentrations but showed a mixture of high and low indoor elemental concentrations.

High Cl and Zn concentrations and low As, Cu, and Zr concentrations defined the indoor pattern of cluster 7. Clusters 1 and 7 were similar with respect to personal $PM_{2.5}$ concentrations but differed with respect to indoor and outdoor $PM_{2.5}$ concentrations, suggesting other factors. Finally, the indoor and outdoor concentrations of cluster 3 were similar to the personal patterns in cluster 3, showing high concentrations of S, Se, and V along with low concentrations of Al. However, rather than the other elements being slightly below average (as they were for the personal exposure), they were all above average for indoor and outdoor exposures. This difference suggested that subjects in cluster 3 could be classified as being highly exposed for almost every element in $PM_{2.5}$ when measured using either indoor or outdoor sampling but actually had high personal exposures to only S, Se, and V. This coincided with the conclusion drawn from the personal exposures that subjects in cluster 3 have high exposure to these elements from sources other than traffic. It is possible that subjects in this cluster lived in areas with high $PM_{2.5}$ concentrations because of the presence of both combustion and non-combustion traffic activities but, either through a higher amount of time spent indoors or a lower exchange of air from outside to inside, managed to maintain a lower personal exposure to combustion-related $PM_{2.5}$ than their indoor and outdoor sample concentrations would indicate. Of the 24 elements analyzed, 20 of the indoor elemental concentrations and 15 of the outdoor elemental concentrations differed significantly by cluster.

Overall, the patterns of indoor and outdoor elemental composition of $PM_{2.5}$ based on the clusters formed on personal $PM_{2.5}$ elemental composition exposure suggested that there were nuanced patterns of $PM_{2.5}$ exposure and that using only outdoor (or indoor) measurements may not accurately capture these specific differences in elemental composition. Subjects with lower concentrations of $PM_{2.5}$ elements overall were likely to have very similar indoor, outdoor, and personal concentrations, whereas subjects with high indoor and outdoor $PM_{2.5}$ concentrations did not necessarily have high personal $PM_{2.5}$ concentrations and vice versa.

Although a limitation of any cluster analysis is the subjectivity of its interpretation, the comparison of these clusters with respect to indoor, outdoor, and personal elemental $PM_{2.5}$ concentrations has proved valuable even if the clusters were possibly mischaracterized.

PERSONAL AND HOME CHARACTERISTICS ASSOCIATED WITH CLUSTER MEMBERSHIP

The air-exchange rate and other census-level variables (e.g., elevation and population and housing densities) varied significantly across cluster. Most outstanding was cluster 7, which had a very low census-tract elevation, low population and housing density, and mostly non-developed land. Although clusters 1 and 7 had similar about-average personal $PM_{2.5}$ elemental concentrations, cluster 1 had the highest indoor and outdoor $PM_{2.5}$ elemental concentration means of any cluster, whereas cluster 7 had average indoor and outdoor personal $PM_{2.5}$ elemental concentrations. Cluster 1 had the highest census-tract housing density, and a high percentage of land that was highly developed yet it also had the highest air-exchange rate of any cluster. These characteristics could explain why the subjects in cluster 1 experienced high indoor and outdoor $PM_{2.5}$ concentrations but had about average personal $PM_{2.5}$ concentrations. Cluster 3, which was the unique cluster with high indoor and outdoor exposures but only high personal concentrations of S, Se, and V, had the highest census-tract housing density and population density but an approximately average percentage of land classified as developed at a high density. Individuals in this cluster may have had industrial sources of $PM_{2.5}$ nearby that were not related to traffic and vehicle-fuel combustion. Interestingly, the fraction of time spent at home did not vary by cluster.

Of the personal characteristics (age, sex, race, education, and income), only sex and being Caucasian differed significantly by cluster (Table 6). The cohort was overall mostly female (73%), but clusters 1 and 7 were approximately an even mix between male and female. The percentage of subjects who self-identified as Caucasian overall was 44%, but this also differed significantly by cluster. Clusters 1 and 4 were largely less Caucasian than the overall average (22% and 27%, respectively), whereas cluster 5 was highly Caucasian (76%). These trends agree with previous reports that PM exposure is more common among non-whites and people with lower socioeconomic status (Bell and Ebisu 2012), although it is interesting to note that neither education nor family income differed with respect to cluster membership.

Twelve home and personal characteristics differed significantly by cluster (Table 6). The use of central air conditioning was elevated in cluster 2, whereas the use of window-unit air conditioning was elevated in cluster 4. These two clusters had the highest personal $PM_{2.5}$ concentrations but had only average indoor and outdoor $PM_{2.5}$ concentrations. Additionally, the presence of open windows during a home visit was low for both clusters 2 and 4 (62% and 61%, respectively). However, the same measure, when self-reported by the study participant, did not show the same patterns. Clusters 2 and 4 also had a relatively high number of nearby emission sources and a high use of

cleaning supplies. The increased exposure to possible $PM_{2.5}$ sources coupled with a high use of air conditioning and a low use of open windows could explain why these clusters had high or very high personal elemental $PM_{2.5}$ concentrations but only average indoor and outdoor elemental $PM_{2.5}$ concentrations. Clusters 1 and 3 both had a high number of study participants that had been nearby or operated a diesel vehicle. This could explain the disagreement of personal with indoor and outdoor $PM_{2.5}$ concentrations for these clusters.

Additional research should be conducted to further elucidate the activities and exposures related to increased personal exposure and changes in the composition of elements between indoor, outdoor and personal $PM_{2.5}$. In addition, the clustering approach we used may be applied to other studies of multiple exposures and health effects to identify and describe groups of individuals with common patterns of exposure and determine if these groups are associated with specific health outcomes.

IMPROVING THE PREDICTION OF PERSONAL EXPOSURE USING INDOOR $PM_{2.5}$ AND PERSONAL AND HOME CHARACTERISTICS

The results of our analyses have demonstrated that personal exposure to $PM_{2.5}$ exceeds levels measured at outdoor monitoring sites. In addition, earlier studies found that the composition of $PM_{2.5}$ varied across geographic regions because the sources of exposure, weather, and other factors were different (Bell et al. 2007, 2011). Few studies, however, have examined whether the composition of outdoor $PM_{2.5}$ accurately reflects the composition of personal $PM_{2.5}$. To inform future exposure assessment for epidemiologic studies, we sought to determine whether the addition of indoor $PM_{2.5}$ elemental concentrations or personal and home characteristics would improve the prediction of personal exposure compared with the use of outdoor concentrations alone . Our hypothesis that time spent indoors would significantly modify the association between indoor and personal concentrations proved to be correct for Ca, Fe, Ti, and Zn. Given the contribution of indoor concentrations to personal exposure and the modification of personal exposure by time spent at home, investigators should query study participants about the amount of time they spend in their home when using outdoor measurements as a proxy for personal exposure.

We found nine elements (Ba, Ca, Cl, Cu, K, Sn, Sr, V, and Zn) for which the inclusion of indoor $PM_{2.5}$ information significantly increased the prediction accuracy of personal $PM_{2.5}$ elemental concentrations. However, the increase in CV R^2 was marginal (less than 0.3) for all of these, and overall the model R^2 was less than 0.5 for all elements except V and S. No clear pattern emerged with respect to specific sources or types of elements that might benefit the most from the inclusion of indoor $PM_{2.5}$ data. Of these elements, only Ca, Cu, K, V, and Zn are known to have significant indoor sources, primarily cooking (Ca) and personal-care products (Ca, Cu, K, V, and Zn). These results are consistent with our correlation analyses (Figure 1) in that these elements also had significantly higher observed I–P concentration correlations compared with O–P concentration correlations. Similarly, the I/O and O/P ratios of these elements were mostly less than 1, suggesting increased personal concentrations compared with both indoor and outdoor concentrations.

The further inclusion of personal and home characteristics in the model using indoor and outdoor elemental concentrations improved the model fit only for Pb. In fact, 16 of the elements (Ag, Al, As, Ba, Br, Ca, Cr, K, S, Sb, Se, Si, Sn, Sr, V, and Zn) actually had less predictive accuracy when personal and home characteristics were added to the linear model. The improvement in model fit when including personal and home characteristics only for Pb may be indicative of Pb in household dust from the earlier use of lead-based paint. The fact that model fit did not improve for any other elements further supports the possibility that personal activities, rather than housing characteristics, are the primary driver of personal exposure levels.

Although the linear models did show improvement in prediction of personal exposure after addition of indoor and personal and home characteristics for some elements in $PM_{2.5}$, the random forest analyses did not show improvement for any elements when any of those variables were added. Random forest models can deal better with a higher number of predictors that are possibly correlated as well take into account complex interactions and nonlinear behaviors of elemental concentrations when used as predictors. Because of these advantages, it is likely that random forest analysis can more accurately model personal elemental concentrations than linear models, further corroborating our findings that the addition of indoor concentrations or home and personal characteristics did not improve predictive accuracy. The increase of predictive accuracy in the linear regression models was likely related to model instability from collinearity — indeed, some elements showed significantly worse predictive accuracy in the linear model, whereas in the random forest analysis, the predictive accuracy was not significantly different for any of the elements.

Additional evidence that our random forest analysis was able to fit the exposure relationships more accurately could be seen in that the random forest analysis better predicted $PM_{2.5}$ concentrations for almost all of the elements

considered and with every combination of predictors we tried than did the linear model. When using the Outdoor Only model, random forest analysis yielded better predictions than did the linear model for all elements except Ag and Si. Likewise, when using the Outdoor + Indoor model, random forest analysis performed better than the linear model for all elements except Ca, Cu, K, Sb, Si, and Zn. Finally, when using the Outdoor + Indoor + Home and Personal Characteristics model, random forest analysis performed better than the linear model for all elements except Ca, Ni, Sb, Si, and Ti, although the linear model did perform better for Cu.

A side implication of these results is that epidemiologic researchers should seek to improve their predictive personal exposure accuracy by using nonlinear modeling methods to capture complex and nonlinear interactions as well as to deal with the common scenario where the number of possible predictors outnumbers the sample size. These methods appear to be less sensitive to the selection of underlying predictors and will therefore be more robust when predicting unknown exposures.

IMPLICATIONS OF FINDINGS

The majority of studies linking PM to adverse health outcomes have used the mass concentration of $PM_{2.5}$ as the primary exposure metric. This measure is most frequently obtained from ambient central monitoring stations maintained for regulatory purposes and is particularly used for studies across geographic regions. However, observed differences in risk estimates across regions suggest that the total mass concentration of PM may not be the most relevant measure of PM toxicity (Pun et al. 2014). Rather, the toxicity of PM is likely a function of the various sizes of PM found in $PM_{2.5}$ and of the complex mixtures of PM components, including sulfate, nitrate, mineral dust, elemental and organic carbon, polycyclic aromatic hydrocarbons, and other chemical constituents.

In recognition of the fact that the composition of $PM_{2.5}$ plays an important role in its toxicity and health effects, more recent epidemiologic studies have examined the association between specific components of $PM_{2.5}$ and various health outcomes. Results of these studies suggest that combustion-related elements, including elemental carbon, Ni, V, and Zn, play an important role in associated cardiovascular and respiratory morbidity and mortality (Bell et al. 2009; Franklin et al. 2008; Ostro et al. 2007, 2008, 2009; Patel et al. 2009; Pun et al. 2014). Al, a crustal material associated with resuspended soil and dust, unpaved roads, and construction activities, is also associated with nonaccidental deaths, respiratory hospitalizations, and low birth weight (Bell et al. 2010; Franklin et al. 2008; Pun et al. 2014).

An underlying assumption of epidemiologic studies using central-site ambient air monitoring is that the concentration data obtained at outdoor monitoring stations, frequently limited in number and geography, are representative of personal exposure. Although $PM_{2.5}$ mass concentrations have previously been shown to have little spatial variability across countywide regions (Dominici et al. 2006), elemental components, particularly those related to traffic sources, have considerably higher spatial variability (Martuzevicius et al. 2004). We thus hypothesized that concentrations of elements obtained at outdoor monitoring stations may not accurately reflect personal exposure. Our results do not support our hypothesis in the case of some individual elements. However, as future epidemiologic studies consider multiple pollutants simultaneously, our results suggest that the overall mixture of elements in outdoor $PM_{2.5}$ is not representative of the mixture of elements in personal air samples.

In order to examine our results in the context of studies of health effects associated with individual elements in $PM_{2.5}$, we selected five elements (Al, Ni, Si, S, and V) previously identified as components of $PM_{2.5}$ that are associated with adverse cardiovascular, respiratory, and birth outcomes (Table 8). Of these, S and V are primarily the result of long-range transport of PM from coal combustion, specifically from secondary formation (S) or from regional transport of PM emitted from oil combustion (V). Our analyses showed that for these elements, both I–P and O–P concentration correlations were high ($r > 0.8$). I/P ratios of each were close to 1, whereas O/P ratios were greater than 1, which was consistent with previous studies (Hsu et al. 2012). The inclusion of indoor $PM_{2.5}$ data did improve the prediction of personal exposure to V but not to S in linear regression models. However, this improvement was marginal (CV R^2 of 0.65 to 0.77). Collectively, our results are consistent with previous studies (Hsu et al. 2012; Janssen et al. 2005), and we concluded that outdoor monitoring is likely to be an appropriate surrogate of personal exposure for these elements and others where regional and long-range transport are the primary sources of exposure. However, it should be noted that concentrations of S and V outdoors are likely to overestimate personal exposure levels, and when possible, time spent indoors should be considered in order to reduce the overestimation of personal exposure.

Al and Si are crustal elements whose outdoor sources are primarily soil and road dust and represent the coarser fraction of $PM_{2.5}$. For both of these elements, we found the I–P concentration correlation to be approximately twice that of

Table 8. Summary of Health Effects Associated with Selected Elements in PM

Element	Potential Sources	Health Outcome	Location	Reference
Al	Crustal material, resuspended dust soil, unpaved roads, and construction	Low birth weight	New England	Bell et al. 2010
		Nonaccidental death	25 U.S. communities	Franklin et al. 2008
		Respiratory hospitalizations	Hong Kong, China	Pun et al. 2014
S	Coal emissions and secondary pollutant associated with regional sources	Low birth weight	New England	Bell et al. 2010
		Nonaccidental death	25 U.S. communities	Franklin et al. 2008
Si	Soil and road dust, cooking, and resuspension of indoor dust	Low birth weight	New England	Bell et al. 2010
		Nonaccidental death	25 U.S. communities	Franklin et al. 2008
V	Oil combustion	Low birth weight	New England	Bell et al. 2010
		Childhood wheezing	New York City	Patel et al. 2009
		Cardiovascular and respiratory hospitalizations	106 U.S. counties	Bell et al. 2009
Ni	Oil combustion and industrial emissions	Low birth weight	New England	Bell et al. 2010
		Childhood wheezing	New York City	Patel et al. 2009
		Nonaccidental death	25 U.S. communities	Franklin et al. 2008
		Cardiovascular and respiratory hospitalizations	106 U.S. counties	Bell et al. 2009
		Cardiovascular hospitalizations	Hong Kong, China	Pun et al. 2014

the O–P correlation, and the I/P and O/P ratios indicated that personal concentrations significantly exceeded both indoor and outdoor concentrations. The inclusion of indoor elemental concentrations did not significantly improve the prediction of personal concentrations of Al and Si. Similarly, the addition of personal and home characteristics did not markedly improve the association between indoor and outdoor $PM_{2.5}$ on personal concentrations.

Ni is a transition metal whose primary source is the combustion of oil. Previous studies have found weak correlations between concentrations of Ni at monitoring locations and outdoor home sites but strong correlations between Ni I–P and O–P samples (Hsu et al. 2012). Similarly, we found I–P and O–P correlations to be similar to but less than those observed for S, a marker of long-range transport. Personal concentrations of Ni were greater than both indoor and outdoor concentrations. Neither the linear model nor random forest approaches suggested that the inclusion of indoor concentrations or home and personal characteristics would improve the prediction of personal exposures.

Our clustering analyses yielded several significant findings. First, there were individuals whose personal activities resulted in significantly increased personal exposure to $PM_{2.5}$; neither the total mass nor composition of $PM_{2.5}$ was reflected in indoor or personal $PM_{2.5}$ samples. In our analyses, cluster 4 had significantly increased concentrations of $PM_{2.5}$ and many elements in personal samples, despite the outdoor concentrations of $PM_{2.5}$ being less than or similar to that of other clusters. This highlights the importance of personal sampling when possible, because we were unable to clearly identify other personal and home characteristics that predicted membership in this cluster. Secondly, the fraction of time spent indoors was similar

across all clusters, suggesting that indoor sources and exposure occurring outside of the home were significant contributors to the overall composition of $PM_{2.5}$. Finally, examining those elements frequently associated with health effects (Al, Ni, Si, S, and V) across the seven clusters, we observed that all five were differentiated across the clusters. Thus, personal exposure to these varied and frequently occurred simultaneously with other elements.

There are some limitations to our analyses that should be considered. First, elemental carbon, a specific marker of traffic exposure previously linked to adverse health outcomes (Bell et al. 2009; Epstein et al. 2012; Ostro et al. 2007), was not included in the RIOPA database in personal samples. Therefore, we were unable to evaluate how well outdoor concentrations of elemental carbon reflected personal exposure. Secondly, our analyses used data from the RIOPA study, which recruited individuals to capture different air pollution sources and weather conditions in the three cities. In addition, participants in the RIOPA study were nonsmoking, and residences were clustered within geographic areas in each study site. Nevertheless, our analyses were conducted on concurrent indoor, outdoor, and personal samples in order to compare the differences among them.

Another limitation to our study is that the indoor, outdoor, and personal sampling was conducted concurrently over the same, relatively short sampling period. Therefore, the results of this study address spatial rather than temporal variability in $PM_{2.5}$ elemental composition. The composition of $PM_{2.5}$ has been shown to vary temporally by season (Bell et al. 2008) and may also vary over short time periods. However, we were unable to specifically examine short-term temporal variability in the RIOPA dataset, and our results should be considered only in the context of spatial variability in $PM_{2.5}$ composition.

With respect to the spatial contrasts observed in the RIOPA dataset, it should also be noted that the RIOPA study collected outdoor $PM_{2.5}$ at the participants' homes rather than at central monitoring sites. We expected the outdoor $PM_{2.5}$ composition at the RIOPA participants' homes to be better correlated with the composition observed in the sampled indoor and personal $PM_{2.5}$ compared with $PM_{2.5}$ composition at the central monitoring sites. However, future research is required to confirm our findings using elemental composition data from central monitoring site in the three RIOPA cities.

SUMMARY OF KEY FINDINGS

In summary, our results support the following three conclusions:

1. The inclusion of indoor monitoring offered minimal improvement in the modeling of personal exposure for individual elements compared with outdoor monitoring at the home alone. This conclusion is based on the results of our linear and random forest analyses, which showed marginal increases in model R^2 when including information on indoor concentrations of elements.
2. Personal and home characteristics did not confer additional benefit in the prediction of personal exposure to individual elements. This finding is likely a result of personal activity patterns and behaviors that influenced personal exposure to a greater extent than demographic and housing factors. Future research incorporating personal exposure measurements and detailed data on personal activities are required to elucidate specific activities that result in increased personal exposure to individual elements.
3. The overall composition of $PM_{2.5}$ varied significantly between indoor, outdoor, and personal $PM_{2.5}$ samples.

Our PCA and cluster results support the conclusion that, in contrast to our findings with individual elements, the overall composition of elements in $PM_{2.5}$ does vary significantly by sample type. This finding has implications for future time-series studies using central monitoring sites if multiple pollutants or elements are considered simultaneously. Although the data used in these analyses compared outdoor $PM_{2.5}$ collected at the home to indoor and personal $PM_{2.5}$, it is likely that the composition of $PM_{2.5}$ at central monitoring sites also does not accurately represent the overall elemental composition of personal $PM_{2.5}$ exposure.

Future studies examining co-exposure to multiple elements in $PM_{2.5}$ simultaneously should be cautious in the use of central monitoring sites because of the likelihood of exposure misclassification with respect to the composition of elements in personal $PM_{2.5}$ for individual study subjects.

REFERENCES

Bell ML, Belanger K, Ebisu K, Gent JF, Lee HJ, Koutrakis P, et al. 2010. Prenatal exposure to fine particulate matter and birth weight: variations by particulate constituents and sources. Epidemiology 21:884–891.

Bell ML, Dominici F, Ebisu K, Zeger SL, Samet JM. 2007. Spatial and temporal variation in PM(2.5) chemical composition in the United States for health effects studies. Environ Health Perspect 115:989–995.

Bell ML, Ebisu K. 2012. Environmental inequality in exposures to airborne particulate matter components in the United States. Environ Health Perspect 120: 699–1704.

Bell ML, Ebisu K, Peng RD. 2011. Community-level spatial heterogeneity of chemical constituent levels of fine particulates and implications for epidemiological research. J Expo Sci Environ Epidemiol 21:372–384.

Bell ML, Ebisu K, Peng RD, Samet JM, Dominici F. 2009. Hospital admissions and chemical composition of fine particle air pollution. Am J Respir Crit Care Med 179: 1115–1120.

Bell ML, Ebisu K, Peng RD, Walker J, Samet JM, Zeger SL, et al. 2008. Seasonal and regional short-term effects of fine particles on hospital admissions in 202 US counties, 1999–2005. Am J Epidemiol 168:1301–1310.

Brunekreef B, Janssen N, de Hartog JJ, Oldenwening M, Meliefste K, Hoek G, et al. 2005. Personal, Indoor, and Outdoor Exposures to $PM_{2.5}$ and Its Components for Groups of Cardiovascular Patients in Amsterdam and Helsinki. Research Report 127. Boston, MA:Health Effects Institute.

Dockery DW. 2009. Health effects of particulate air pollution. Ann Epidemiol 19:257–263.

Dominici F, Peng RD, Bell ML, Pham L, McDermott A, Zeger SL, et al. 2006. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. JAMA 295:1127–1134.

Epstein TG, Ryan PH, LeMasters GK, Bernstein CK, Levin LS, Bernstein JA, et al. 2012. Poor asthma control and exposure to traffic pollutants and obesity in older adults. Ann Allergy Asthma Immunol 108:423–428.

Fraley C, Raftery AE. 2002. Model-based clustering, discriminant analysis, and density estimation. J Am Stat Assoc 97:611–631.

Fraley C, Raftery AE, Murphy BT, Scrucca L. 2012. MCLUST Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. Technical Report No. 597. Seattle, WA:University of Washington.

Franklin M, Koutrakis P, Schwartz P. 2008. The role of particle composition on the association between $PM_{2.5}$ and mortality. Epidemiology 19:680–689.

Hammond DM, Dvonch JT, Keeler GJ, Parker EA, Kamal AS, Barres JA, et al. 2008. Sources of ambient fine particulate matter at two community sites in Detroit, Michigan. Atmos Environ 42:720–732.

Han I, Mihalic JN, Ramos-Bonilla JP, Rule AM, Polyak LM, Peng RD, et al. 2012. Assessment of heterogeneity of metal composition of fine particulate matter collected from eight U.S. counties using principal component analysis. J Air Waste Manage Assoc 62:773–782.

Hsu SI, Ito K, Kendall M, Lippmann M. 2012. Factors affecting personal exposure to thoracic and fine particles and their components. J Expo Sci Environ Epidemiol 22:439–447.

Janssen NA, Lanki T, Hoek G, Vallius M, de Hartog JJ, Van Grieken R, et al. 2005. Associations between ambient, personal, and indoor exposure to fine particulate matter constituents in Dutch and Finnish panels of cardiovascular patients. Occup Environ Med 62:868–877.

Lall R, Thurston GD. 2006. Identifying and quantifying transported vs. local sources of New York City $PM_{2.5}$ fine particulate matter air pollution. Atmos Environ 40: S333–S346.

Lee S, Liu W, Wang Y, Russell AG, Edgerton ES. 2008. Source apportionment of $PM_{2.5}$: Comparing PMF and CMB results for four ambient monitoring sites in the southeastern United States. Atmos Environ 42:4126–4137.

Levy JI, Clougherty JE, Baxter LK, Houseman EA, Paciorek CJ. 2010. Evaluating Heterogeneity in Indoor and Outdoor Air Pollution Using Land-Use Regression and Constrained Factor Analysis. Research Report 152. Boston, MA:Health Effects Institute.

Li Z, Hopke PK, Husain L, Qureshi S, Dutkiewicz VA, Schwab JJ, et al. 2004. Sources of fine particle composition in New York City. Atmos Environ 38:6521–6529.

Liaw A, Wiener M. 2002. Classification and Regression by randomForest. R News 2:18–22.

Martuzevicius D, Grinshpun SA, Reponen T, Górny RL, Shukla R, Lockey J, et al. 2004. Spatial and temporal variations of $PM_{2.5}$ concentration and composition throughout an urban area with high freeway density—the Greater Cincinnati study. Atmos Environ 38:1091–1105.

National Research Council of the National Academies. 2004. IV. Continuing research progress. In: Research Priorities for Airborne Particulate Matter. Washington, DC:National Academies Press.

Nerriere É, Zmirou-Navier D, Blanchard O, Momas I, Ladner J, Le Moullec Y, et al. 2005. Can we use fixed ambient air monitors to estimate population long-term exposure to air pollutants? The case of spatial variability in the Genotox ER study. Environ Res 97:32–42.

Ogulei D, Hopke PK, Zhou L, Pancras JP, Nair N, Ondov JM. 2006. Source apportionment of Baltimore aerosol from combined size distribution and chemical composition data. Atmos Environ 40:396–410.

Onat B, Alver Sahin U, Bayat C. 2012. Assessment of particulate matter in the urban atmosphere: Size distribution, metal composition and source characterization using principal component analysis. J Environ Monit 14:1400–1409.

Ostro B, Feng WY, Broadwin R, Green S, Lipsett M. 2007. The effects of components of fine particulate air pollution on mortality in California: Results from CALFINE. Environ Health Perspect 115:13–19.

Ostro BD, Feng WY, Broadwin R, Malig BJ, Green RS, Lipsett MJ. 2008. The impact of components of fine particulate matter on cardiovascular mortality in susceptible subpopulations. Occup Environ Med 65:750–756.

Ostro B, Roth L, Malig B, Marty M. 2009. The effects of fine particle components on respiratory hospital admissions in children. Environ Health Perspect 117:475–480.

Ozkaynak H, Xue J, Spengler J, Wallace L, Pellizzari E, Jenkins P. 1995. Personal exposure to airborne particles and metals: Results from the Particle TEAM study in Riverside, California. J Expo Anal Environ Epidemiol 6:57–78.

Patel MM, Hoepner L, Garfinkel R, Chillrud S, Reyes A, Quinn JW, et al. 2009. Ambient metals, elemental carbon, and wheeze and cough in New York City children through 24 months of age. Am J Respir Crit Care Med 180:1107–1113.

Pekney NJ, Davidson CI, Robinson A, Zhou LM, Hopke P, Eatough D, et al. 2006. Major source categories for $PM_{2.5}$ in Pittsburgh using PMF and UNMIX. Aerosol Sci. Technol. 40:910–924.

Peng RD, Bell ML, Geyh AS, McDermott A, Zeger SL, Samet JM, et al. 2009. Emergency admissions for cardiovascular and respiratory diseases and the chemical composition of fine particle air pollution. Environ Health Perspect 117:957–963.

Pun VC, Yu IT-S, Qiu H, Ho K-F, Sun Z, Louie PK, et al. 2014. Short-term associations of cause-specific emergency hospitalizations and particulate matter chemical components in Hong Kong. Am J Epidemiol 179:1086–1095.

Qin Y, Kim E, Hopke PK. 2006. The concentrations and sources of $PM_{2.5}$ in metropolitan New York City. Atmos Environ 40:312–332.

Rizzo MJ, Scheff PA. 2007. Fine particulate source apportionment using data from the U.S. EPA speciation trends network in Chicago, Illinois: Comparison of two source apportionment models. Atmos Environ 41: 6276–6288.

Saldarriaga-Norena H, Hernandez-Mena L, Ramirez-Muniz M, Carbajal-Romero P, Cosio-Ramirez R, Esquivel-Hernandez B. 2009. Characterization of trace metals of risk to human health in airborne particulate matter ($PM_{2.5}$) at two sites in Guadalajara, Mexico. J Environ Monit 11:887–894.

Schauer JJ, Lough GC, Shafer MM, Christensen WF, Arndt MF, DeMinter JT, et al. 2006. Characterization of Metals Emitted from Motor Vehicles. Research Report 133. Boston, MA:Health Effects Institute.

Schlesinger RB, Kunzli N, Hidy GM, Gotschi T, Jerrett M. 2006. The health relevance of ambient particulate matter characteristics: coherence of toxicological and epidemiological inferences. Inhal Toxicol 18:95–125.

Simkhovich BZ, Kleinman MT, Kloner RA. 2008. Air pollution and cardiovascular injury epidemiology, toxicology, and mechanisms. J Am Coll Cardiol 52:719–726.

Spencer MT, Shields LG, Sodeman DA, Toner SM, Prather KA. 2006. Comparison of oil and fuel particle chemical signatures with particle emissions from heavy and light duty vehicles. Atmos Environ 40:5224–5235.

Turpin BJ, Weisel CP, Morandi M, Colome S, Stock T, Eisenreich S, et al. 2007. Part II. Analyses of concentrations of particulate matter species. In: Relationships of Indoor, Outdoor, and Personal Air (RIOPA): Research Report 130. Boston, MA:Health Effects Institute.

Upadhyay N, Clements A, Fraser M, Herckes P. 2011. Chemical speciation of $PM_{2.5}$ and PM_{10} in south Phoenix, AZ, USA. J Air Waste Manage Assoc 61:302–310.

Weisel CP, Zhang J, Turpin BJ, Morandi MT, Colome S, Stock TH, et al. 2005a. Part I: Collection methods and descriptive analyses. In: Relationships of Indoor, Outdoor, and Personal Air (RIOPA). Research Report 130. Boston, MA:Health Effects Institute.

Weisel CP, Zhang J, Turpin BJ, Morandi MT, Colome S, Stock TH, et al. 2005b. Relationship of Indoor, Outdoor and Personal Air (RIOPA) study: Study design, methods and quality assurance/control results. J Expo Anal Environ Epidemiol 15:123–137.

Yu CH, Fan Z-H, Meng Q, Zhu X, Korn L, Bonanno LJ. 2011. Spatial/temporal variations of elemental carbon, organic carbon, and trace elements in PM_{10} and the impact of land-use patterns on community air pollution in Paterson, NJ. J Air Waste Manage Assoc 61:673–688.

Zanobetti A, Franklin M, Koutrakis P, Schwartz J. 2009. Fine particulate air pollution and its components in association with cause-specific emergency admissions. Environ Health 8:58.

Zhao W, Hopke PK, Norris G, Williams R, Paatero P. 2006. Source apportionment and analysis on ambient and personal exposure samples with a combined receptor model and an adaptive blank estimation strategy. Atmos Environ 40:3788–3801.

APPENDICES AVAILABLE ON THE WEB

Appendices A and B contain supplemental material not included in the printed report. They are available on the HEI Web site *http://pubs.healtheffects.org*.

Appendix A. Descriptive Statistics for Elemental PM Concentrations by Study Site

Appendix B. Principal Component Loading Factors

ABOUT THE AUTHORS

Patrick H. Ryan, Ph.D., is an associate professor of pediatrics and environmental health at Cincinnati Children's Hospital Medical Center and the University of Cincinnati College of Medicine in Cincinnati, Ohio. He received his M.S. and Ph.D. in environmental health and epidemiology from the University of Cincinnati and a B.S. in mathematics from Xavier University in Cincinnati. His research interests include the assessment and impact of traffic-related air pollution on respiratory and neurobehavioral development in childhood and the use of personal monitoring for ultrafine particle exposure.

Cole Brokamp, B.S., is a graduate student in biostatistics at the University of Cincinnati Department of Environmental Health in Cincinnati, Ohio. He received his B.S. from the College of Engineering at the University of Cincinnati. His research interests include developing and applying machine-learning and data-mining approaches for biomedical data.

Zhi-Hua (Tina) Fan, Ph.D., is an associate professor in the Exposure Science Division at the Department of Environmental and Occupational Medicine at Robert Wood Johnson Medical School, Rutgers University in Piscataway, New Jersey. She received her B.S. and M.S. in environmental chemistry from Peking University, in Beijing, China, in 1986 and 1989, respectively, and received her Ph.D. in atmospheric chemistry from the Department of Environmental Sciences and Engineering in the School of Public Health at the University of North Carolina–Chapel Hill in 1995. Her research interests include characterization of personal and community exposure to environmental contaminants, particularly air pollutants in socioeconomically disadvantaged communities, and the identification of the sources of exposures and the factors that may affect those exposures.

M. B. Rao, Ph.D., is a professor of biostatistics in the Department of Environmental Health at the University of Cincinnati in Cincinnati, Ohio. He received a B.S. and M.S. in statistics from the University of Madras in Chennai, India, and a Ph.D. in statistics from the Indian Statistical Institute in Chennai, India. He is a fellow of the American Association for the Advancement of Science and the American Statistical Association. His research interests include applications of statistical methodologies in bioinformatics, classification and clustering techniques in medical diagnostics, longitudinal data analysis, and false discovery rates.

OTHER PUBLICATION RESULTING FROM THIS RESEARCH

Brokamp C, Rao MB, Fan Z-H, Ryan PH. 2015. Does the elemental composition of indoor and outdoor $PM_{2.5}$ accurately represent the elemental composition of personal $PM_{2.5}$? Atmos Environ 101:226–234.

ABBREVIATIONS

ANOVA	analysis of variance
CV	cross-validated
ICP–MS	inductively coupled plasma mass spectrometry
I/P	indoor/personal
O/P	outdoor/personal
PC	principal component
PCA	principal component analysis
PCS	principal component score
PM	particulate matter
$PM_{2.5}$	PM having an aerodynamic diameter ≤ 2.5 μm
RIOPA	Relationships of Indoor, Outdoor, and Personal Air
XRF	X-ray fluorescence

ABBREVIATIONS OF CHEMICAL ELEMENTS

Ag	silver
Al	aluminum
As	arsenic
Ba	barium
Br	bromine
Ca	calcium
Cl	chlorine
Cr	chromium
Cu	copper
Fe	iron
K	potassium
Mn	manganese
Ni	nickel
Pb	lead
S	sulfur
Sb	antimony
Se	selenium
Si	silicon
Sn	tin
Sr	strontium
Ti	titanium
V	vanadium
Zn	zinc
Zr	zirconium

CRITIQUE

Health Review Committee

Research Report 185, *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*, P.H. Ryan et al.

INTRODUCTION

Particulate matter with a diameter of less than 2.5 µm ($PM_{2.5}$*) has been associated with adverse health effects (Hoek et al. 2013; Rückerl et al. 2011). However, $PM_{2.5}$ is a complex mixture of many components that vary in composition and size and originate from a variety of outdoor sources. Ambient concentrations of these components vary greatly across time and space because of differences in and proximity to sources and differences in weather and topography. Because the composition of PM is complex, there has long been a question as to whether some of its components are of greater public health concern than others. HEI's National Particle Component Toxicity initiative (NPACT), however, found that no component or source of $PM_{2.5}$ can so far be eliminated (Lippmann et al. 2013; Vedal et al. 2013); a similar conclusion was drawn by Stanek and colleagues (2011) in a review of earlier published studies on the same topic.

Assessments of exposure to PM and its components and of their associated adverse health outcomes are further complicated by the fact that there are also indoor sources of PM. In addition, certain personal activities and behaviors (e.g., smoking, driving, and cooking) contribute substantially to exposure.

The Relationships of Indoor, Outdoor, and Personal Air (RIOPA) study was co-funded by HEI and the National Urban Air Toxics Research Center in 1999 to better define the relationships among indoor, outdoor, and personal exposure concentrations of air pollutants, including $PM_{2.5}$ and its components (Turpin et al. 2007; Weisel et al. 2005). The study was conducted in three cities with different air pollution sources and weather conditions: Los Angeles, California (dominated by mobile sources); Houston, Texas (dominated by large industrial sources); and Elizabeth, New Jersey (a mixture of mobile and industrial sources). In each city, convenience samples of approximately 100 participants who did not smoke and who lived in homes located at various distances from air pollution sources were selected. Homes close to air pollution sources were preferentially sampled. Indoor, outdoor (directly outside the home), and personal air pollution samples were collected during two 48-hour sampling periods in various seasons (approximately three months apart) between summer 1999 and spring 2001. Information on personal activities and factors that might affect exposures, such as housing characteristics, was collected using three detailed questionnaires. Household air-exchange rates and geographic and meteorologic information were obtained as well. Initial methods, features of the population, a data quality description, and descriptive analyses of the volatile organic compounds (VOCs), carbonyls, and $PM_{2.5}$ measurements and speciation for each city and the three cities combined have been published (Turpin et al. 2007; Weisel et al. 2005). Only a few earlier studies have been published that provided detailed information about the elemental composition of simultaneously obtained indoor, outdoor, and personal concentrations of $PM_{2.5}$ (e.g., Clayton et al. 1993; Janssen et al. 2005; Long and Sarnat 2004; Sarnat et al. 2006).

In 2008, HEI issued Request for Applications (RFA) 08-1, "Relationships of Indoor, Outdoor, and Personal Air (RIOPA): Further Analyses of the RIOPA Study Data." This RFA sought proposals to conduct more detailed analyses of the RIOPA study data in order to address additional questions about exposure to air pollution as a function of weather, housing characteristics, and distance from sources. Exploration of methodologic issues using the RIOPA data set was encouraged as well. In response, Dr. Patrick H. Ryan, then at the University of Cincinnati, in Cincinnati, Ohio, submitted an application to HEI in which he proposed a two-year study to analyze the elemental composition of

Dr. Patrick H. Ryan's 2-year study, "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," began in August 2010. Total expenditures were $110,600. The draft Investigators' Report from Ryan and colleagues was received for review in April 2014. A revised report, received in December 2014 and further revised in May 2015, was accepted for publication in June 2015. During the review process, the HEI Health Review Committee and the investigators had the opportunity to exchange comments and to clarify issues in both the Investigators' Report and the Review Committee's Critique.

* A list of abbreviations and other terms appears at the end of the Investigators' Report.

indoor, outdoor, and personal $PM_{2.5}$ samples in greater detail than was done in the original RIOPA analyses. The HEI Research Committee was interested in Ryan's proposal because of the potential to address important questions relevant to epidemiologic studies, but felt that some of the proposed statistical analyses were not appropriate. The Committee recommended that mixed-modeling approaches (in order to allow adjusting for clustering in the data) and principal component analysis (PCA) be considered. After revisions to the proposal, the Committee recommended Dr. Ryan's study for funding, and the project started in August 2010. Completion was delayed because of his move to Cincinnati Children's Hospital and other reasons.

HEI also funded a second study under RFA 08-1, in which Dr. Stuart Batterman of the University of Michigan identified factors that influence exposure to individual VOCs and VOC mixtures and characterized various exposure distributions for them, with particular emphasis on high exposures. The report of this study has been published (Batterman et al. 2014).

This critique is intended to aid the sponsors of HEI and the public by highlighting both the strengths and limitations of the study and by placing the Investigators' Report into scientific and regulatory context.

APPROACH

The aims of the study were to

1. Explore the relationships among the elemental compositions of indoor, outdoor, and personal $PM_{2.5}$ samples;
2. Identify clusters of individuals with similar exposure to various mixtures of elements in personal $PM_{2.5}$ samples and examine personal and home characteristics associated with these clusters; and
3. Investigate whether indoor exposure, outdoor exposure, or personal and home characteristics predict personal exposure to elements in $PM_{2.5}$.

The RIOPA data were analyzed as a cross-sectional study with one simultaneously obtained set of sample types (indoor, outdoor, and personal) per person available for analysis. Analyses were limited to 168 adults — one per household — with complete data. Data on 24 different elements that had detectable values in at least 70% of the personal samples were used for the analyses. Detectable values below the analytical limit of detection were retained. Nondetectable values were replaced by zero. In all but the descriptive analyses, values were transformed as log(1 + value).

To address aim 1, descriptive analyses were conducted, including calculation of outdoor/personal ratios, indoor/personal ratios, and Spearman correlation coefficients. In addition, PCA (a data-reduction procedure in which the original correlated data are transformed into uncorrelated components) was conducted to assess the heterogeneity of the elemental composition of $PM_{2.5}$. Data for the combined sample types (indoor, outdoor, and personal) by city and for the combined cities (Los Angeles, Houston, and Elizabeth) by sample type were used. Principal components with eigenvalues of greater than 1 were extracted. Analysis of variance was used to test whether the principal components differed by sample types or cities.

To address aim 2, model-based cluster analyses (using the R package mclust) were conducted to group RIOPA participants with similar exposures to mixtures of elements. Four cluster analyses were conducted, with personal samples from the three cities combined and from each city separately. Using the Bayesian information criterion, the optimal number of clusters as well as the most likely model form were identified. Differences in personal and home characteristics across the clusters were examined.

To address aim 3, both linear regression and random forest analyses were conducted to predict personal exposure to PM components. In initial linear regression models, personal exposure was predicted for each element separately with outdoor and indoor measurements as the predictor variables in one model and, additionally, by including time spent indoors and an interaction term with the indoor measurements in the model. Additional regression models were run to predict total personal exposure, with the following predictor variables: (1) outdoor measurements (24 variables), (2) outdoor and indoor measurements (48 variables), and (3) outdoor and indoor measurements and personal and home characteristics (71 variables) selected using an initial screening procedure. Screening criteria for inclusion in the final prediction model were developed; the criteria were (1) that a characteristic should be associated with the personal sample in a univariate model ($P < 0.10$) for six or more elements and (2) that information about the characteristic should be available for at least 90% of the participants. Similar analyses were performed using random forest analyses.

A random forest analyses is a machine-learning technique that aggregates many single regression trees using a different bootstrapped sample of the data for each tree. The performance of the regression models was tested using an out-of-sample squared correlation coefficient between the observed and predicted values estimated from 10-fold cross-validation (i.e., the data were split into a training set and a test set; the model was built on the training set,

evaluated on the test set, and repeated 10 times; and then the estimated errors were averaged). Finally a *t*-test was used to assess the differences of the squared correlation coefficients between the models. *P* values less than 0.05 were considered significant.

SUMMARY OF RESULTS

- On average, personal exposures were higher than outdoor and indoor concentrations for $PM_{2.5}$ and most of its elements. The exceptions were S and V, with outdoor and indoor concentrations consistently higher than personal exposures in pooled and city-specific analyses. For these elements, the highest indoor–personal and outdoor–personal correlations were also reported (> 0.7).
- PCA showed that the elemental composition of the $PM_{2.5}$ was heterogeneous across cities and sample types (indoor, outdoor, and personal).
- Model-based cluster analyses using personal samples from the three cities combined yielded limited meaningful insights. No details were given about those analyses by city in the report.
- Outdoor concentrations adequately predicted personal exposures only for a limited number of elements (e.g., S and V), mainly those associated with long-range PM transport and with few known indoor sources. For the other elements, outdoor concentrations did not predict personal exposure well.
- Compared with the use of outdoor elemental concentrations alone in the model, the addition of indoor concentration data did not improve the prediction of personal exposure for most of the elements. Results differed substantially between the linear regression analyses and the random forest analyses. In the linear regression analyses, inclusion of indoor concentrations significantly improved the prediction of personal exposure for nine of the 24 elements (Ba, Ca, Cl, Cu, K, Sn, Sr, V, and Zn). Note that no analyses were run to assess how well indoor concentrations alone predicted personal exposures.
- The addition of personal and home characteristics to models with both outdoor and indoor concentrations improved the prediction of personal exposure for only one element (Pb) and only in linear regression analyses.

HEALTH REVIEW COMMITTEE'S EVALUATION

In its independent review of the study, the HEI Health Review Committee noted that the authors had conducted extensive analyses on data from 168 adult participants in the original RIOPA study 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; applications of traditional approaches in a unique way (i.e., PCA to examine the heterogeneity of elemental composition across cities and sample types); and novel applications (i.e., random forest analyses). The authors presented the basic concepts behind some complex statistical methods quite clearly, and the Committee appreciated the use of a variety of approaches to obtain better insight into this unique data set. However, the Committee also identified several important issues with the analytical 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. PCA results in the study showed, as expected, notable differences in air pollution across the cities. Therefore, the current study's PCA, which combined data from the three cities, as well as the model-based cluster analyses, which combined personal samples from the three cities, provided limited meaningful insights. In addition, participant indoor, outdoor, and personal samples are typically not independent; not correcting for these correlations affects the ability to compare across sample types. Lastly, the analyses did not account for differences in elemental concentrations caused by season. Many studies, including RIOPA, have reported substantial differences in air pollution levels by season as well as in the relationships among indoor, outdoor, and personal exposures (Kioumourtzoglou et al. 2014; Weisel et al. 2005).

The Committee questioned in particular the linear regression analyses, because the number of predictor variables in the regression models was large (71 in the final model) compared with the number of observations (N = 168), which can be problematic in a model intended for prediction and inference testing at the same time. Each added variable, when not needed to improve model fit, can also reduce the ability of a model to predict the outcome variable properly. In addition, although potentially useful for improving prediction, each added variable can introduce multi-collinearity problems if it is correlated with other variables; this can destabilize the model and make drawing conclusions about regression coefficients challenging (Kutner et al. 2004).

The results from the random forest analyses — which were notably different from those of the linear regression analyses — were not affected by the issues described above. Unlike traditional regression models, random forest models have several attractive properties, including the ability to deal with a large number of correlated predictor variables and to adequately capture complex interactions and nonlinear relationships among predictors (Hastie et al. 2009). However, although the Committee would have expected the random forest analyses to uncover more nuanced or additional relationships among sample types and personal and home characteristics in the prediction of personal exposures (compared with the use of outdoor elemental concentrations alone), none were reported. A possible explanation of the differences between the results of the linear regression and random forest analyses, apart from the issues described above, could be the influence of outlier values, which were abundantly present in the RIOPA data (see Investigators' Report Tables 1–3). Random forest analyses are less sensitive to outlier values than are linear regression analyses. However, the investigators did not evaluate the influence of outliers on their results; this would have been a useful addition to the study.

HEI's goal in funding this study was to explore how well outdoor concentrations could serve as a surrogate for personal exposures, because personal sampling is difficult and expensive and, therefore, not practical on a large scale. Given that there are many indoor sources of air pollution, one of the goals of this study was to parse the contribution of outdoor concentrations to indoor concentrations and personal exposures. Unfortunately, the Ryan study provides limited insight on these questions. In part, the design of the RIOPA study (in which Dr. Ryan was not involved) limited some of the analyses he could perform. Dr. Ryan analyzed RIOPA data as a cross-sectional study, although two samples per person were available. Time-series studies, which are common in air pollution epidemiology, use temporal variation in outdoor concentrations as a surrogate for temporal variation in personal exposures. Therefore, the results from the current study seem most appropriately applied to cross-sectional and cohort studies, which use spatial variation in outdoor concentrations. Additionally, RIOPA collected data from outdoor measurements from monitors placed directly outside the home; outdoor measurements from central site monitors — typically used in epidemiologic studies — were not obtained either by the original investigators or Dr. Ryan. These measurements would have been a useful addition to the data set and analyses. At the very least, it would have been useful in the current study to more directly explore the relationship between outdoor and personal concentrations by, for instance, removing participants where indoor sources were particularly influential. Another issue that limits application of these results more broadly for epidemiology is that the RIOPA study was performed in a convenience sample rather than a representative, population-based sample. This was, of course, known before the Ryan study began and was discussed in the commentaries accompanying the original HEI RIOPA studies (Turpin et al. 2007; Weisel et al. 2005).

The current study's analyses show that outdoor concentrations did not represent personal exposures well for elements other than those associated with long-range transport, such as S and V, and that the collection of personal exposure measurements therefore may still be important for understanding exposure to specific particle species. It has been shown previously that S (or sulfate) is a good metric for estimating personal exposure to $PM_{2.5}$ of ambient origin because it has few known indoor sources and a high infiltration rate because it is found predominantly in the submicrometer fraction of $PM_{2.5}$ (e.g., Jansen et al. 2005 and Sarnat et al. 2002). Correlations between indoor–outdoor and personal–outdoor samples are generally higher for repeated measurements compared with a cross-sectional sample, given the larger differences between participants than within participants for factors affecting exposure. This could be one reason why the correlations reported by Janssen and colleagues (2005) and Montagne and colleagues (2014) were higher than those found by Ryan and colleagues. However, typically lower correlations can be expected when comparing central site concentrations with personal and indoor concentrations rather than with concentrations just outside the home (e.g., Montagne et al. 2014). Comparing the current results with those of other studies is difficult because of the issues described above. It should also be noted that personal exposure measurements are unfeasible in large epidemiologic studies. However, the association between ambient concentrations and personal exposure can be investigated in smaller representative subgroups of participants, and therefore conducting validation studies like RIOPA remains useful.

An important aim of doing such validation studies would be to quantify exposure measurement error and, ultimately, allow researchers to take this into account in health analyses. This remains an important area of research (Sheppard et al. 2012). A number of methods to assess and quantify exposure measurement error have recently been proposed, such as risk-set regression calibration for time-varying exposures (Hart et al. 2015) based on a pooled analysis of several personal exposure studies (Kioumourtzoglou et al. 2014), imputation techniques based on meta-analysis of studies reporting correlations between personal and ambient exposures (Holliday et al.

2014), and methods to correct measurement error for "spatially incompatible" data (i.e., different distributions of locations of monitors and participants) (Szpiro and Paciorek 2013). Typically, the adjustment of health effect estimates for exposure measurement error leads to alterations in the effect estimates and the widening of their 95% confidence intervals. More research in this area is clearly needed. Characterization of exposure measurement error will be an important part of the studies HEI plans to fund under the recently issued RFA 14-3 "Assessing adverse health effects of long-term exposure to low levels of ambient air pollution."

SUMMARY AND 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., PCA); and a novel approach (i.e., random forest analyses). 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 the analyses were not adjusted for clustering and correlation within cities, among individuals, and between seasons; the number of predictor variables was large compared with the size of the data set; and the results could have been influenced by outlier values. Performing 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.

ACKNOWLEDGMENTS

The Health Review Committee thanks the ad hoc reviewers for their help in evaluating the scientific merit of the Investigators' Report. The Committee is also grateful to Geoffrey Sunshine for his oversight of the study, to Kate Adams for the initial review of the study, to Hanna Boogaard for her assistance in preparing its Critique, to Mary Brennan for science editing of this Report and its Critique, and to Hope Green, Fred Howe, Hilary Selby Polk, and Ruth Shaw for their roles in preparing this Research Report for publication.

REFERENCES

Batterman S, Su FC, Li S, Mukherjee B, Jia C. 2014. Personal Exposure to Mixtures of Volatile Organic Compounds: Modeling and further analysis of the RIOPA data. Research Report 181. Boston, MA:Health Effects Institute.

Clayton CA, Perritt RL, Pellizzari ED, Thomas KW, Whitmore RW, Wallace LA, et al. 1993. Particle total exposure assessment methodology (PTEAM) study: Distributions of aerosol and elemental concentrations in personal, indoor, and outdoor air samples in a southern California community. J Expo Anal Environ Epidemiol 3:227–250.

Hart JE, Liao X, Hong B, Puett RC, Yanosky JD, Suh H, et al. 2015. The association of long-term exposure to $PM_{2.5}$ on all-cause mortality in the Nurses' Health Study and the impact of measurement-error correction. Environ Health 14:38.

Hastie T, Tibshirani R, Friedman J. 2009. The Elements of Statistical Learning: Data mining, inference, and prediction. 2nd edit. Berlin:Springer.

Hoek G, Krishnan RM, Beelen R, Peters A, Ostro B, Brunekreef B, et al. 2013. Long-term air pollution exposure and cardio-respiratory mortality: A review. Environ Health 12:43.

Holliday KM, Avery CL, Poole C, McGraw K, Williams R, Liao D, et al. 2014. Estimating personal exposures from ambient air pollution measures: Using meta-analysis to assess measurement error. Epidemiology 25:35–43.

Janssen NA, Lanki T, Hoek G, Vallius M, de Hartog JJ, Van Grieken R, et al. 2005. Associations between ambient, personal, and indoor exposure to fine particulate matter constituents in Dutch and Finnish panels of cardiovascular patients. Occup Environ Med 62:868–877.

Kioumourtzoglou MA, Spiegelman D, Szpiro AA, Sheppard L, Kaufman JD, Yanosky JD, et al. 2014. Exposure measurement error in $PM_{2.5}$ health effects studies: A pooled analysis of eight personal exposure validation studies. Environ Health 13:2.

Kutner MH, Nachtsheim CJ, Neter J, Li W. 2004. Applied Linear Statistical Models. 5th edit. New York:McGraw-Hill/Irwin Series Operations and Decision Sciences.

Lippmann M, Chen LC, Gordon T, Ito K, Thurston GD. 2013. National Particle Component Toxicity (NPACT) Initiative: Integrated epidemiologic and toxicologic studies of the health effects of particulate matter components. Research Report 177. Boston, MA:Health Effects Institute.

Long CM, Sarnat JA. 2004. Indoor-outdoor relationships and infiltration behavior of elemental components of outdoor $PM_{2.5}$ for Boston-area homes. Aerosol Sci Technol 38:91–104.

Montagne D, Hoek G, Nieuwenhuijsen M, Lanki T, Siponen T, Portella M, et al. 2014. Temporal associations of ambient $PM_{2.5}$ elemental concentrations with indoor and personal concentrations. Atmos Environ 86:203–211.

Rückerl R, Schneider A, Breitner S, Cyrys J, Peters A. 2011. Health effects of particulate air pollution: A review of epidemiological evidence. Inhal Toxicol 23:555–592.

Sarnat JA, Long CM, Koutrakis P, Coull BA, Schwartz J, Suh HH. 2002. Using sulfur as a tracer of outdoor fine particulate matter. Environ Sci Technol 36:5305–5314.

Sarnat SE, Coull BA, Schwartz J, Gold DR, Suh HH. 2006. Factors affecting the association between ambient concentrations and personal exposures to particles and gases. Environ Health Perspect 114:649–654.

Sheppard L, Burnett RT, Szpiro AA, Kim SY, Jerrett M, Pope CA 3rd, et al. 2012. Confounding and exposure measurement error in air pollution epidemiology. Air Qual Atmos Health 5:203–216.

Stanek LW, Sacks JD, Dutton SJ, Dubois J-JB. 2011. Attributing health effects to apportioned components and sources of particulate matter: An evaluation of collective results. Atmos Environ 45:5655–5663.

Szpiro AA, Paciorek CJ. 2013. Measurement error in two-stage analyses, with application to air pollution epidemiology. Environmetrics 24:501–517.

Turpin BJ, Weisel CP, Morandi M, Colome S, Stock T, Eisenreich S, et al. 2007. Relationships of Indoor, Outdoor, and Personal air (RIOPA): Part II. Analyses of concentrations of particulate matter species. Research Report 130. Boston, MA:Health Effects Institute.

Vedal S, Campen MJ, McDonald JD, Kaufman JD, Larson TV, Sampson PD, et al. 2013. National Particle Component Toxicity (NPACT) Initiative Report on Cardiovascular Effects. Research Report 178. Boston, MA:Health Effects Institute.

Weisel CP, Zhang J, Turpin BJ, Morandi MT, Colome S, Stock TH, et al. 2005. Relationships of Indoor, Outdoor, and Personal Air (RIOPA). Part I. Collection methods and descriptive analyses. Research Report 130. Boston, MA:Health Effects Institute.

RELATED HEI PUBLICATIONS:

PARTICULATE MATTER, MIXTURES, AND PERSONAL EXPOSURE

Number	Title	Principal Investigator	Date
Research Reports			
183	New Statistical Methods for Analyzing Multiple Pollutants, Sources, and Health Outcomes	B.A. Coull, E.S. Park	2015
181	Personal Exposure to Mixtures of Volatile Organic Compounds: Modeling and Further Analysis of the RIOPA Data	S. Batterman	2014
179	Development and Application of an Aerosol Screening Model for Size-Resolved Urban Aerosols	C.O. Stanier	2014
178	National Particle Component Toxicity (NPACT) Initiative Report on Cardiovascular Effects	S. Vedal	2013
153	Improved Source Apportionment and Speciation of Low-Volume Particulate Matter Samples	J.J. Schauer	2010
152	Evaluating Heterogeneity in Indoor and Outdoor Air Pollution Using Land-Use Regression and Constrained Factor Analysis	J.I. Levy	2010
142	Air Pollution and Health: A European and North American Approach	K. Katsouyanni	2009
140	Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality	D. Krewski	2009
139	Effects of Long-Term Exposure to Traffic-Related Air Pollution on Respiratory and Cardiovascular Mortality in the Netherlands: The NLCS-AIR Study	B. Brunekreef	2009
138	Health Effects of Real-World Exposure to Diesel Exhaust in Persons with Asthma	J. Zhang	2009
137	The Influence of Improved Air Quality on Mortality Risks in Erfurt, Germany	A. Peters	2009
131	Characterization of Particulate and Gas Exposures of Sensitive Subpopulations Living in Baltimore and Boston	P. Koutrakis	2005
130	Relationships of Indoor, Outdoor, and Personal Air (RIOPA)		
	Part I. Collection Methods and Descriptive Analyses	C.P. Weisel	2005
	Part II. Analyses of Concentrations of Particulate Matter Species	B.J. Turpin	2007
127	Personal, Indoor, and Outdoor Exposures to $PM_{2.5}$ and Its Components for Groups of Cardiovascular Patients in Amsterdam and Helsinki	B. Brunekreef	2005
124	Particulate Air Pollution and Nonfatal Cardiac Events		2005
	Part I. Air Pollution, Personal Activities, and Onset of Myocardial Infarction in a Case–Crossover Study	A. Peters	
	Part II. Association of Air Pollution with Confirmed Arrhythmias Recorded by Implanted Defibrillators	D. Dockery	
122	Evaluation of a Personal and Microenvironmental Aerosol Speciation Sampler (PMASS)	A. Geyh	2004
114	A Personal Particle Speciation Sampler	S. Hering	2003

Continues next page

Copies of these reports can be obtained from HEI; pdf's are available for free downloading at http://pubs.healtheffects.org.

RELATED HEI PUBLICATIONS:

PARTICULATE MATTER, MIXTURES, AND PERSONAL EXPOSURE

Number	Title	Principal Investigator	Date
99	A Case–Crossover Analysis of Fine Particulate Matter Air Pollution and Out-of-Hospital Sudden Cardiac Arrest	H. Checkoway	2000
98	Daily Mortality and Fine and Ultrafine Particles in Erfurt, Germany *Part I.* Role of Particle Number and Particle Mass	H-E. Wichmann	2000
95	Association of Particulate Matter Components with Daily Mortality and Morbidity in Urban Populations	M. Lippmann	2000
94	National Morbidity, Mortality, and Air Pollution Study		
	Part I. Methods and Methodologic Issues	J.M. Samet	2000
	Part II: Morbidity and Mortality from Air Pollution in the United States	J.M. Samet	2000
	Part III: Concentration–Response Curves and Thresholds for the 20 Largest US Cities	M.J. Daniels	2004
	Part IV: Hierarchical Bivariate Time-Series Models—A Combined Analysis of PM_{10} Effects on Hospitalization and Mortality	F. Dominici	2005
91	Mechanisms of Morbidity and Mortality from Exposure to Ambient Air Particles	J.J. Godleski	2000
Special Reports			
17	Traffic-Related Air Pollution: A Critical Review of the Literature		2010
	Revised Analyses of Time-Series Studies of Air Pollution and Health		2003

Copies of these reports can be obtained from HEI; pdf's are available for free downloading at http://pubs.healtheffects.org.

HEI BOARD, COMMITTEES, and STAFF

Board of Directors

Richard F. Celeste, Chair *President Emeritus, Colorado College*

Sherwood Boehlert *Of Counsel, Accord Group; Former Chair, U.S. House of Representatives Science Committee*

Enriqueta Bond *President Emerita, Burroughs Wellcome Fund*

Purnell W. Choppin *President Emeritus, Howard Hughes Medical Institute*

Michael T. Clegg *Professor of Biological Sciences, University of California–Irvine*

Jared L. Cohon *President Emeritus and Professor, Civil and Environmental Engineering and Engineering and Public Policy, Carnegie Mellon University*

Stephen Corman *President, Corman Enterprises*

Linda Rosenstock *Dean Emerita and Professor of Health Policy and Management, Environmental Health Sciences and Medicine, University of California–Los Angeles*

Henry Schacht *Managing Director, Warburg Pincus; Former Chairman and Chief Executive Officer, Lucent Technologies*

Warren M. Washington *Senior Scientist, National Center for Atmospheric Research; Former Chair, National Science Board*

Health Research Committee

David L. Eaton, Chair *Dean and Vice Provost of the Graduate School, University of Washington–Seattle*

Jeffrey R. Brook *Senior Research Scientist, Air Quality Research Division, Environment Canada, and Assistant Professor, University of Toronto, Canada*

David Christiani *Elkan Blout Professor of Environmental Genetics, Harvard T.H. Chan School of Public Health*

Francesca Dominici *Professor of Biostatistics and Senior Associate Dean for Research, Harvard T.H. Chan School of Public Health*

David E. Foster *Phil and Jean Myers Professor Emeritus, Department of Mechanical Engineering, Engine Research Center, University of Wisconsin–Madison*

Uwe Heinrich *Professor, Hannover Medical School; Executive Director, Fraunhofer Institute for Toxicology and Experimental Medicine, Hanover, Germany*

Amy H. Herring *Carol Remmer Angle Distinguished Professor of Children's Environmental Health, and Associate Chair, Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina–Chapel Hill*

Barbara Hoffmann *Professor of Environmental Epidemiology and Head of Environmental Epidemiology of Aging, IUF-Leibniz Research Institute for Environmental Medicine, and Professor, Medical Faculty, Heinrich Heine University of Düsseldorf, Germany*

Allen L. Robinson *Raymond J. Lane Distinguished Professor and Head, Department of Mechanical Engineering, and Professor, Department of Engineering and Public Policy, Carnegie Mellon University*

Ivan Rusyn *Professor, Department of Veterinary Integrative Biosciences, Texas A&M University*

HEI BOARD, COMMITTEES, and STAFF

Health Review Committee

James A. Merchant, Chair *Professor and Founding Dean, College of Public Health, University of Iowa*

Michael Brauer *Professor, School of Environmental Health, University of British Columbia, Canada*

Bert Brunekreef *Professor of Environmental Epidemiology, Institute of Risk Assessment Sciences, University of Utrecht, the Netherlands*

Mark W. Frampton *Professor of Medicine and Environmental Medicine, University of Rochester Medical Center*

Jana B. Milford *Professor, Department of Mechanical Engineering and Environmental Engineering Program, University of Colorado–Boulder*

Roger D. Peng *Associate Professor, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health*

Lianne Sheppard *Professor of Biostatistics, School of Public Health, University of Washington–Seattle*

Officers and Staff

Daniel S. Greenbaum *President*

Robert M. O'Keefe *Vice President*

Rashid Shaikh *Director of Science*

Jacqueline C. Rutledge *Director of Finance and Administration*

April Rieger *Corporate Secretary*

Zachary Abbott *Research Assistant*

Sarah Benckart *Science Administration Assistant*

Hanna Boogaard *Staff Scientist*

Aaron J. Cohen *Principal Scientist*

Maria G. Costantini *Principal Scientist*

Philip J. DeMarco *Compliance Manager*

Hope Green *Publications Associate*

Anny Luu *Executive Assistant*

Hilary Selby Polk *Managing Editor*

Evan Rosenberg *Staff Accountant*

Robert A. Shavers *Operations Manager*

Geoffrey H. Sunshine *Senior Scientist*

Annemoon M.M. van Erp *Managing Scientist*

Donna J. Vorhees *Senior Scientist*

Katherine Walker *Senior Scientist*

HEALTH
EFFECTS
INSTITUTE

101 Federal Street, Suite 500
Boston, MA 02110, USA
+1-617-488-2300
www.healtheffects.org

RESEARCH
REPORT

Number 185
December 2015