RESEARCH REPORT

HEALTH EFFECTS INSTITUTE

Number 181
June 2014

Personal Exposure to Mixtures of Volatile Organic Compounds: Modeling and Further Analysis of the RIOPA Data

Stuart Batterman, Feng-Chiao Su, Shi Li, Bhramar Mukherjee, and Chunrong Jia

Personal Exposure to Mixtures of Volatile Organic Compounds: Modeling and Further Analysis of the RIOPA Data

Stuart Batterman, Feng-Chiao Su, Shi Li,
Bhramar Mukherjee, and Chunrong Jia

with a Critique by the HEI Health Review Committee

Research Report 181
Health Effects Institute
Boston, Massachusetts

Trusted Science • Cleaner Air • Better Health

Publishing history: This document was posted at *www.healtheffects.org* in June 2014.

Citation for document:

Batterman S, Su F-C, 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.

 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 vii

About This Report ix

HEI STATEMENT 1

INVESTIGATORS' REPORT *by Batterman et al.* 3

ABSTRACT 3
- Introduction 3
- Methods 3
- Results 4
- Conclusions 5

INTRODUCTION 5
- Significance of VOC Exposures 5
- VOC Monitoring and Exposure Assessment 6
- High Exposures 6
- Distributions of VOCs in Exposure Mixtures 7
- Assessment of Exposure to VOC Mixtures 7
- Dependencies in VOC Mixtures Analyzed with Copulas 8
- Determinants of VOC Exposures 8

SPECIFIC AIMS 10
- Aim 1 10
- Aim 2 10
- Aim 3 10

DATA SOURCES 11
- Relationships of Indoor, Outdoor, and Personal Air Study 11
- National Health and Nutrition Examination Survey 11

SPECIFIC AIM 1. EXTREME VALUE ANALYSES AND FULL DISTRIBUTION MODELS 13
- Methods: Extreme Value Analyses 13
  - Risk Evaluation for Extreme Value Exposures 13
  - Fitting of Generalized Extreme Value Distributions 13
  - Extreme Value Simulations 14
- Results and Discussion: Extreme Value Analyses 14
  - Outliers 14
  - Predicted Health Risks for Extreme VOC Exposures 14
  - Extreme Value Distributions for the RIOPA Data 16
  - Extreme Value Distributions for the NHANES Data 20
- Methods: Mixture of Normal Distribution Fitting 20
  - Finite Mixture of Normal Distributions 20
  - Dirichlet Process Mixture of Normal Distributions 21
  - Goodness-of-Fit Criteria 22
  - Simulation Study 22

Results and Discussion: Mixture of Normal Distributions for VOCs 22
Single Normal Distributions 22
Finite Mixture of Normal Distributions 23
Nonparametric DPM of Normal Distributions 24
Simulations 26
SPECIFIC AIM 2. DISTRIBUTION MODELS FOR VOC MIXTURES 27
Methods: Identification of Mixtures 27
Positive Matrix Factorization 27
Toxicologic Mode of Action 28
Methods: Dependency Structures of Mixtures 28
Copula Analysis 28
Simulated Joint Distributions 29
Results and Discussion: Selected VOC Mixtures Based on PMF 29
Results and Discussion: Dependency Structures and Joint Distributions of VOC Mixtures 32
Copulas 32
Mixture Fractions 33
Estimated Cancer Risks 35
SPECIFIC AIM 3. IDENTIFICATION OF DETERMINANTS 36
Methods: Time, Compartment, and Source Fractions of VOC Exposure 36
Time Fractions 36
Compartment Fractions 37
Source Fractions 37
Results and Discussion: Time, Compartment, and Source Fractions of VOC Exposure 37
Exposure Fractions by Time 37
Exposure Fractions by Compartment 38
Exposure Fractions by Source 38
Methods: Identification of Determinants of Personal, Home, and Outdoor VOC Exposure 41
Variable Selection 41
Linear Mixed-Effects Models 41
Model Assessment 41
Missing Data 41
Results and Discussion: Identification of Determinants of Personal, Home, and Outdoor VOC Exposure 42
Personal Exposure to Gasoline-Related VOCs 42
Personal Exposure to Odorant and Cleaning-Related VOCs 42
Personal Exposure to Dry-Cleaning and Industry-Related VOCs 45
Summary of Key Personal Exposure Determinants 47
Determinants Identified for Other Compounds Measured in RIOPA 47
Determinants of Indoor VOC Concentrations 47
Determinants of Outdoor VOC Concentrations 50
Common Determinants of Personal, Indoor, and Outdoor Concentrations 50

Assumption of Linearity 54
Model Validation 54
Limitations 54

CONCLUSIONS 54

IMPLICATIONS OF THE FINDINGS 55

ACKNOWLEDGMENTS 56

REFERENCES 56

APPENDIX AVAILABLE ON THE WEB 62

ABOUT THE AUTHORS 62

OTHER PUBLICATIONS RESULTING FROM THIS RESEARCH 63

ABBREVIATIONS AND OTHER TERMS 63

CRITIQUE *by the Health Review Committee* 65

INTRODUCTION 65

AIMS AND APPROACHES 66

BRIEF SUMMARY OF RESULTS 67
Analyses of Determinants 67
Distribution Fitting for Individual VOCs 67
Distribution Fitting for Mixtures of VOCs 67

THE HEALTH REVIEW COMMITTEE'S EVALUATION 67

SUMMARY AND CONCLUSIONS 69

ACKNOWLEDGMENTS 69

REFERENCES 69

Related HEI Publications 71

HEI Board, Committees, and Staff 73

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 half of its core funds from the U.S. Environmental Protection Agency and half from 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 181, *Personal Exposure to Mixtures of Volatile Organic Compounds: Modeling and Further Analysis of the RIOPA Data,* presents a research project funded by the Health Effects Institute and conducted by Dr. Stuart Batterman of the Department of Environmental Health Sciences, School of Public Health, University of Michigan, Ann Arbor, and his colleagues. This 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 Batterman 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 181

Personal Exposure to Mixtures of Volatile Organic Compounds: Modeling and Further Analysis of RIOPA Data

INTRODUCTION

Exposure to various volatile organic compounds (VOCs) has been associated with a wide range of adverse health outcomes. Assessments of exposure and health effects are complicated because many indoor and outdoor sources contribute VOCs, and because certain personal activities and behaviors may influence exposure substantially. Dr. Stuart Batterman of the University of Michigan in Ann Arbor and his colleagues used data from the Relationships of Indoor, Outdoor, and Personal Air (RIOPA) study and, to a lesser extent, the 1999–2000 data from the National Health and Nutrition Examination Survey (NHANES) to identify factors that influence exposure and to characterize exposure distributions for individual VOCs and mixtures, with particular emphasis on extreme values (high exposures).

The original 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 personal, indoor, and outdoor exposure measurements, the investigators collected information on factors that might affect exposures (determinants), such as housing characteristics, personal activities, and geographic and meteorologic information.

The 1999–2000 NHANES obtained personal measurements of VOCs for approximately 650 adult subjects in a U.S. population-based sample.

Aims of the current study were to investigate determinants of exposure to individual VOCs and to characterize the distributions of both individual VOCs and VOC mixtures, with particular emphasis on high exposures.

APPROACH

Fifteen VOCs were analyzed, including benzene, methyl *tert*-butyl ether, 1,4-dichlorobenzene, toluene, tetrachloroethylene, and chloroform.

Exposure determinants for individual VOCs in the RIOPA data set were modeled using linear mixed-effects models adjusted for clustering within cities and among individuals. A different set of determinants was used for each VOC and for each sample type (personal, indoor, outdoor).

Various distribution models were fitted for each VOC using personal exposure data from RIOPA and NHANES. The primary focus in these analyses was characterizing extreme values using generalized extreme value models. Consequently, extreme value analyses were based on small numbers of highly exposed subjects (12 or 24 from RIOPA and 32 or 64 from

What This Study Adds

- Batterman and colleagues used data from RIOPA and NHANES to identify factors that influence exposures and to characterize exposure distributions for VOCs and VOC mixtures, with particular emphasis on high exposures.
- Factors shown to influence personal exposure included city, wind speed, home air exchange rate, number of rooms in the home, attached garage, pumping gas, and other family members showering.
- Distributions other than lognormal models may perform somewhat better in estimating high personal exposures and cancer risks. Interpretation of results was problematic because of the treatment of values below the limit of detection, the deletion of outlier values, and the use of a convenience sample.

This Statement, prepared by the Health Effects Institute, summarizes a research project funded by HEI and conducted by Dr. Stuart Batterman at the University of Michigan, Ann Arbor, and colleagues. Research Report 181 contains both the detailed Investigators' Report and a Critique of the study prepared by the Institute's Health Review Committee.

NHANES), defined as subjects in the top 5% or 10% of the exposure distribution. For three VOCs (chloroform, 1,4-dichlorobenzene, and styrene) Batterman and colleagues fit two types of mixture models: a finite mixture of normal distributions, and a Dirichlet process mixture (DPM) of normal distributions.

Mixtures of VOCs studied in RIOPA were identified using positive matrix factorization and toxicologic modes of action. Then copulas, a class of probability models, were used to characterize the distribution of and dependence among different VOCs in the mixtures.

The performance of the different exposure distribution models was evaluated using goodness-of-fit statistics and simulated data. The investigators also compared cancer risks using standard risk assessment approaches.

MAIN RESULTS AND INTERPRETATION

Significant determinants of mean personal VOC exposure included city, wind speed, home air exchange rate, number of rooms in the home, attached garage, pumping gas, and other family members showering. Similar determinants were identified for indoor exposure, largely because the RIOPA participants spent most of their time at home (average 90%). In contrast, only a few significant determinants were identified for outdoor VOC exposure — city and weather characteristics.

In its independent review of the study, the HEI Health Review Committee noted that the study was well conceived and conducted and that, in particular, the analyses of exposure determinants were a novel and useful contribution to the literature. The Committee thought that their practical applicability was hampered to some extent because the investigators used a different set of possible determinants for each VOC and for each sample type; the exact magnitude of the effect of an exposure determinant could not be readily estimated.

The Committee did not agree with the investigators' treatment of values below the limit of detection (LOD) in the determinant analyses. The method of replacing all values below the LOD for a particular VOC with one single value (½ of the LOD), though commonly used, can cause problems if the number of observations below the LOD is considerable, as in this study. Thus the Committee felt that caution should be exercised in interpreting the determinant analyses for VOCs with high proportions of such values.

The analyses focused on extreme values were considered interesting. The investigators demonstrated that distributions other than the commonly applied lognormal models may perform somewhat better in estimating high personal exposures and cancer risks. In further distribution fitting for individual VOCs, the Committee appreciated that the investigators applied mixture models, which allowed estimation of entire VOC distributions because concentrations below extreme values also affect total population risk. The investigators' interpretation of the analyses for characterizing extreme values was thought to be problematic for several reasons. First, the results were affected by the considerable number of observations below the LOD and the way they were handled. This approach did not affect the extreme value analyses directly, but it did affect the comparison of those distributions with conventional lognormal distributions; the latter were fitted using the full data set, rather than only the top 5% or 10%, as for extreme value analyses. Second, the investigators deleted what they considered to be outliers and other influential values. The Committee thought these deletions were not adequately justified scientifically. Finally, the Committee suggested caution in generalizing the interpretations of the distribution characterizations because the RIOPA data set — which underlies the majority of analyses — was a convenience sample, not a representative population-based sample, as was NHANES.

CONCLUSIONS

Batterman and colleagues used RIOPA and NHANES data to investigate determinants of exposure and to characterize exposure distributions for VOCs and VOC mixtures, with particular emphasis on high exposures. The Committee thought the study was well conceived and conducted and the analyses of determinants were a novel and useful contribution. In the distribution fitting for individual VOCs, the Committee appreciated that the investigators applied mixture models, allowing estimation of entire VOC distributions. The statistical analyses focusing on extreme values were considered interesting, but the interpretation of results was problematic. The applicability in air pollution research of the methods for extreme value analyses developed in this study may be limited and further research is needed.

INVESTIGATORS' REPORT

Personal Exposure to Mixtures of Volatile Organic Compounds: Modeling and Further Analysis of the RIOPA Data

Stuart Batterman, Feng-Chiao Su, Shi Li, Bhramar Mukherjee, and Chunrong Jia

Departments of Environmental Health Sciences (S.B., F-C.S.) and Biostatistics (S.L., B.M.), School of Public Health, University of Michigan, Ann Arbor; School of Public Health, University of Memphis, Tennessee (C.J.)

ABSTRACT

INTRODUCTION

Emission sources of volatile organic compounds (VOCs*) are numerous and widespread in both indoor and outdoor environments. Concentrations of VOCs indoors typically exceed outdoor levels, and most people spend nearly 90% of their time indoors. Thus, indoor sources generally contribute the majority of VOC exposures for most people. VOC exposure has been associated with a wide range of acute and chronic health effects; for example, asthma, respiratory diseases, liver and kidney dysfunction, neurologic impairment, and cancer. Although exposures to most VOCs for most persons fall below health-based guidelines, and long-term trends show decreases in ambient emissions and concentrations, a subset of individuals experience much higher exposures that exceed guidelines. Thus, exposure to VOCs remains an important environmental health concern.

The present understanding of VOC exposures is incomplete. With the exception of a few compounds, concentration and especially exposure data are limited; and like other environmental data, VOC exposure data can show multiple modes, low and high extreme values, and sometimes a large portion of data below method detection limits (MDLs). Field data also show considerable spatial or interpersonal variability, and although evidence is limited, temporal variability seems high. These characteristics can complicate modeling and other analyses aimed at risk assessment, policy actions, and exposure management. In addition to these analytic and statistical issues, exposure typically occurs as a mixture, and mixture components may interact or jointly contribute to adverse effects. However most pollutant regulations, guidelines, and studies remain focused on single compounds, and thus may underestimate cumulative exposures and risks arising from coexposures. In addition, the composition of VOC mixtures has not been thoroughly investigated, and mixture components show varying and complex dependencies. Finally, although many factors are known to affect VOC exposures, many personal, environmental, and socioeconomic determinants remain to be identified, and the significance and applicability of the determinants reported in the literature are uncertain.

To help answer these unresolved questions and overcome limitations of previous analyses, this project used several novel and powerful statistical modeling and analysis techniques and two large data sets. The overall objectives of this project were (1) to identify and characterize exposure distributions (including extreme values), (2) evaluate mixtures (including dependencies), and (3) identify determinants of VOC exposure.

METHODS

VOC data were drawn from two large data sets: the Relationships of Indoor, Outdoor, and Personal Air (RIOPA)

This Investigators' Report is one part of Health Effects Institute Research Report 181, 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. Stuart Batterman, 1420 Washington Heights, Ann Arbor, MI 48109-2029; e-mail: *stuartb@umich.edu*.

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.

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

study (1999–2001) and the National Health and Nutrition Examination Survey (NHANES; 1999–2000). The RIOPA study used a convenience sample to collect outdoor, indoor, and personal exposure measurements in three cities (Elizabeth, NJ; Houston, TX; Los Angeles, CA). In each city, approximately 100 households with adults and children who did not smoke were sampled twice for 18 VOCs. In addition, information about 500 variables associated with exposure was collected. The NHANES used a nationally representative sample and included personal VOC measurements for 851 participants. NHANES sampled 10 VOCs in common with RIOPA. Both studies used similar sampling methods and study periods.

Specific Aim 1. To estimate and model extreme value exposures, extreme value distribution models were fitted to the top 10% and 5% of VOC exposures. Health risks were estimated for individual VOCs and for three VOC mixtures. Simulated extreme value data sets, generated for each VOC and for fitted extreme value and lognormal distributions, were compared with measured concentrations (RIOPA observations) to evaluate each model's goodness of fit.

Mixture distributions were fitted with the conventional finite mixture of normal distributions and the semi-parametric Dirichlet process mixture (DPM) of normal distributions for three individual VOCs (chloroform, 1,4-DCB, and styrene). Goodness of fit for these full distribution models was also evaluated using simulated data.

Specific Aim 2. Mixtures in the RIOPA VOC data set were identified using positive matrix factorization (PMF) and by toxicologic mode of action. Dependency structures of a mixture's components were examined using mixture fractions and were modeled using copulas, which address correlations of multiple components across their entire distributions. Five candidate copulas (Gaussian, t, Gumbel, Clayton, and Frank) were evaluated, and the performance of fitted models was evaluated using simulation and mixture fractions. Cumulative cancer risks were calculated for mixtures, and results from copulas and multivariate lognormal models were compared with risks based on RIOPA observations.

Specific Aim 3. Exposure determinants were identified using stepwise regressions and linear mixed-effects models (LMMs).

RESULTS

Specific Aim 1. Extreme value exposures in RIOPA typically were best fitted by three-parameter generalized extreme value (GEV) distributions, and sometimes by the two-parameter Gumbel distribution. In contrast, lognormal distributions significantly underestimated both the level and likelihood of extreme values. Among the VOCs measured in RIOPA, 1,4-dichlorobenzene (1,4-DCB) was associated with the greatest cancer risks; for example, for the highest 10% of measurements of 1,4-DCB, all individuals had risk levels above 10^{-4}, and 13% of all participants had risk levels above 10^{-2}.

Of the full-distribution models, the finite mixture of normal distributions with two to four clusters and the DPM of normal distributions had superior performance in comparison with the lognormal models. DPM distributions provided slightly better fit than the finite mixture distributions; the advantages of the DPM model were avoiding certain convergence issues associated with the finite mixture distributions, adaptively selecting the number of needed clusters, and providing uncertainty estimates. Although the results apply to the RIOPA data set, GEV distributions and mixture models appear more broadly applicable. These models can be used to simulate VOC distributions, which are neither normally nor lognormally distributed, and they accurately represent the highest exposures, which may have the greatest health significance.

Specific Aim 2. Four VOC mixtures were identified and apportioned by PMF; they represented gasoline vapor, vehicle exhaust, chlorinated solvents and disinfection byproducts, and cleaning products and odorants. The last mixture (cleaning products and odorants) accounted for the largest fraction of an individual's total exposure (average of 42% across RIOPA participants). Often, a single compound dominated a mixture but the mixture fractions were heterogeneous; that is, the fractions of the compounds changed with the concentration of the mixture.

Three VOC mixtures were identified by toxicologic mode of action and represented VOCs associated with hematopoietic, liver, and renal tumors. Estimated lifetime cumulative cancer risks exceeded 10^{-3} for about 10% of RIOPA participants. The dependency structures of the VOC mixtures in the RIOPA data set fitted Gumbel (two mixtures) and t copulas (four mixtures). These copula types emphasize dependencies found in the upper and lower tails of a distribution. The copulas reproduced both risk predictions and exposure fractions with a high degree of accuracy and performed better than multivariate lognormal distributions.

Specific Aim 3. In an analysis focused on the home environment and the outdoor (close to home) environment, home VOC concentrations dominated personal exposures (66% to 78% of the total exposure, depending on VOC); this was largely the result of the amount of time participants spent at home and the fact that indoor concentrations were much higher than outdoor concentrations for most VOCs.

In a different analysis focused on the sources inside the home and outside (but close to the home), it was assumed

that 100% of VOCs from outside sources would penetrate the home. Outdoor VOC sources accounted for 5% (*d*-limonene) to 81% (carbon tetrachloride [CTC]) of the total exposure. Personal exposure and indoor measurements had similar determinants depending on the VOC. Gasoline-related VOCs (e.g., benzene and methyl *tert*-butyl ether [MTBE]) were associated with city, residences with attached garages, pumping gas, wind speed, and home air exchange rate (AER). Odorant and cleaning-related VOCs (e.g., 1,4-DCB and chloroform) also were associated with city, and a residence's AER, size, and family members showering. Dry-cleaning and industry-related VOCs (e.g., tetrachloroethylene [or perchloroethylene, PERC] and trichloroethylene [TCE]) were associated with city, type of water supply to the home, and visits to the dry cleaner. These and other relationships were significant, they explained from 10% to 40% of the variance in the measurements, and are consistent with known emission sources and those reported in the literature. Outdoor concentrations of VOCs had only two determinants in common: city and wind speed. Overall, personal exposure was dominated by the home setting, although a large fraction of indoor VOC concentrations were due to outdoor sources.

City of residence, personal activities, household characteristics, and meteorology were significant determinants.

Concentrations in RIOPA were considerably lower than levels in the nationally representative NHANES for all VOCs except MTBE and 1,4-DCB. Differences between RIOPA and NHANES results can be explained by contrasts between the sampling designs and staging in the two studies, and by differences in the demographics, smoking, employment, occupations, and home locations. A portion of these differences are due to the nature of the convenience (RIOPA) and representative (NHANES) sampling strategies used in the two studies.

CONCLUSIONS

Accurate models for exposure data, which can feature extreme values, multiple modes, data below the MDL, heterogeneous interpollutant dependency structures, and other complex characteristics, are needed to estimate exposures and risks and to develop control and management guidelines and policies. Conventional and novel statistical methods were applied to data drawn from two large studies to understand the nature and significance of VOC exposures. Both extreme value distributions and mixture models were found to provide excellent fit to single VOC compounds (univariate distributions), and copulas may be the method of choice for VOC mixtures (multivariate distributions), especially for the highest exposures, which fit parametric models poorly and which may represent the greatest health risk. The identification of exposure determinants, including the influence of both certain activities (e.g., pumping gas) and environments (e.g., residences), provides information that can be used to manage and reduce exposures. The results obtained using the RIOPA data set add to our understanding of VOC exposures and further investigations using a more representative population and a wider suite of VOCs are suggested to extend and generalize results.

INTRODUCTION

SIGNIFICANCE OF VOC EXPOSURES

Perhaps more so than for other air pollutants, emission sources of VOCs are numerous and widespread in both indoor and outdoor environments (Finlayson-Pitts and Pitts 2000). Important outdoor sources include industrial emissions and other stationary sources, vehicles and other mobile sources, and gasoline service stations and dry cleaners, which are considered area sources (Ling et al. 2011; Maryland Department of the Environment [MDE] 2010). Indoor sources include many building materials, cleaning products, cigarette smoke, adhesives, paint strippers, moth repellents, and water chlorination by-products (Agency for Toxic Substances and Disease Registry [ATSDR] 1997a; Brown 2002; Singer et al. 2006; U.S. Environmental Protection Agency [EPA] 2012c; Wallace et al. 1987; Wallace et al. 1989; Weschler 2011). In the United States and in many other countries, indoor concentrations of VOCs typically exceed outdoor levels (U.S. EPA 2012c). Moreover, most people spend nearly 90% of their time indoors (U.S. EPA 1989). For these two reasons, indoor exposures often constitute a large share, and often the dominant share, of VOC exposures for most individuals who are not exposed in the workplace (occupational exposures were not included in this study). Studies are needed to understand how outdoor and indoor sources contribute to personal exposures of air pollutants, which was a major motivation of funding the RIOPA study (Weisel et al. 2005a; Turpin et al. 2007) and is the focus of this report.

VOC exposure has been associated with a wide range of acute and chronic health effects, including irritation, asthma exacerbation, allergy, respiratory diseases, liver and kidney dysfunction, neurological impairment, and cancer (Kim and Bernstein 2009; Lippy and Turner 1991; Mendell 2007; Rumchev et al. 2007; U.S. EPA 2012a,c). Information regarding toxicity, drawn largely from occupational and animal studies, is available for a number of VOCs. Several elements of this report use the RIOPA VOC measurements

with dose–response information (specifically, the unit risk factor [URF], also called the slope factor) for cancer risk, and the reference concentrations (RfC) for non-cancer endpoints. (In the sections about health risks related to extreme values, for example, lifetime individual excess cancer risks were estimated by multiplying the lifetime [70-year] exposure by the URF specific to the VOC [U.S. EPA 2009].) The estimated risk was compared with acceptable values, which typically range from 10^{-6} to 10^{-4}. Previous work based on the nationally representative 1999–2000 NHANES has shown that exposures to most VOCs for most persons fall below the current guidelines designed to be protective for both acute and chronic (cancer) effects (Jia et al. 2008). However, a subset of individuals are exposed to much higher concentrations that do exceed the guidelines; for example, the estimated lifetime cancer risk from benzene exposure exceeded 10^{-4} for 10% of adults, and exposure to chloroform exceeded the same risk level for 16% of adults. Information on these high exposures is very limited.

VOC MONITORING AND EXPOSURE ASSESSMENT

Personal measurements of pollutant concentrations, obtained using air samplers carried by individuals, are generally believed to provide the most relevant data for analyzing exposure. The RIOPA data set includes such measurements, as well as indoor and outdoor measurements (inside and outside participants' homes) from a relatively large study sample (309 adults and 118 children) (Weisel et al. 2005b). In addition, the VOCs were sampled twice to provide repeated measurements. (For details on the RIOPA data collection, see the section Data Sources.)

VOC monitoring programs in the United States and elsewhere measure only a subset of VOCs. Monitoring often focuses on 1-ring aromatic VOCs (e.g., benzene, toluene, and the xylenes), smaller aliphatic compounds (*n*-hexane and heptane), and a few chlorinated compounds (TCE and CTC). The RIOPA study included several aromatic and chlorinated compounds, as well as *d*-limonene, α-pinene, β-pinene, and MTBE. In general, little information is available regarding levels of and exposures to very volatile VOCs, more polar compounds, and lower volatility VOCs. This report focuses on only the VOCs measured in RIOPA.

HIGH EXPOSURES

As noted, the highest exposures may be most significant in their potential to cause adverse health effects. The assumption of lognormality has been widely applied in analyzing concentration and exposure data. However, lognormal distributions may inadequately characterize the highest values in a data set. For example, VOC distributions can have upper extreme values, which clearly fit neither normal nor lognormal distributions (Jia et al. 2008). In these cases, parametric models will underestimate the highest exposures and risks.

One approach to characterize such extreme values in a data set uses extreme value theory, which describes the probability and magnitude of events with low likelihood (Lenox and Haimes 1996). A variety of extreme value models have been developed, including the Gumbel (Gumbel 1958), the Fréchet (Fisher and Tippett 1928), and the Weibull (Aberg and Guttorp 2008; Ang and Tang 1975) distributions. These three (respectively called type I, II, and III extreme value distributions) belong to the broad class of GEV distributions, which use the three parameters of shape, location, and scale to fit the tails of a distribution (Jenkinson 1955). Extreme value distributions are univariate models (applying to one VOC) and not full-distribution models (applying only to a tail of the distribution).

Despite these limitations, extreme value distributions have many applications. Extreme value theory has been widely applied in engineering (McCormick 1981), finance (Embrechts et al. 1997), hydrology (Engeland et al. 2004; Katz et al. 2002), and other fields. Some, but not many, environmental applications have been published; for example, estimating the likelihood of meteorologic conditions (Hüsler 1983; Sneyer 1983), exceedances of thresholds relevant to dietary intake of pesticides and heavy metals (Paulo et al. 2006; Tressou et al. 2004), concentrations of metals manganese and lead in blood (Batterman et al. 2011), deposition of pollutants in surface soils (Huang and Batterman 2003), and risks of leakage due to pipe corrosion (U.K. Health and Safety Executive 2002). Additional application for air pollutants include the exceedance of air quality standards (Hopke and Paatero 1994; Surman et al. 1987), exposures to ambient air pollutants (Kassomenos et al. 2010), indoor concentrations of radon (Tuia and Kanevski 2008), and VOC exposures in the NHANES data subset mentioned earlier (Jia et al. 2008).

In the Extreme Value Analyses section we apply this theory to the VOC exposure data in the RIOPA data set and provide a critique of the approach. The analysis of extreme values is extended in the Dependency Structure of Mixtures section, which uses copulas to model dependencies among mixture components. This analysis also looks at the upper tail of the distribution, the region that may be critical for health effects assessment and for which simple models and assumptions, such as the lognormal models discussed above, may be ill suited.

DISTRIBUTIONS OF VOCs IN EXPOSURE MIXTURES

Environmental exposures of many VOCs (and other pollutants) at the population level can be viewed as a mixture of distributions. A (typically small) fraction of the population experiences exposure to high concentrations due to specific exposure events, whereas a (typically large) fraction of the population encounters much lower concentrations (Batterman et al. 2011; Jia et al. 2008). For the lower concentrations, measurements often fall below the MDL. These undetectable levels, sometimes referred to as left-censored data, can be treated by substitution, single or multiple imputation, regression on order statistics (modeling using probability plots of known distributions to estimate summary statistics), and laboratory-generated data (using the original data without replacement) (Antweiler and Taylor 2008). The extent of data below MDLs can significantly affect the quality of the results (Antweiler and Taylor 2008; Lubin et al. 2004); and the statistical issues associated with analyzing such data are well known (Krishnamoorthy et al. 2009; Taylor et al. 2001).

Due to the variation in sources of emissions, differences in the settings and environmental factors where exposures occur, and the measurement issues just noted, a mixture distribution of VOC concentrations can have multiple modes, extreme values, and significant portions of data that fall below the MDL and are replaced by a single value. These issues, which can be encountered in exposure and other types of data sets, challenge standard parametric distribution models. Although the GEV distributions discussed above can fit the upper portions of distributions, they do not represent the full distribution of the data. Information on the full distributions of exposure levels is needed to establish exposure–risk guidelines and to estimate risks across a population, to estimate health risks and uncertainty estimates, and to facilitate probabilistic analyses (Hammonds et al. 1994).

Mixtures of distributions, which extend parametric families of distributions to fit data sets that are not adequately fitted by a single common distribution, provide a flexible and powerful approach for representing the distribution of a random variable (McLachlan and Basford 1988; McLachlan and Peel 2000; Titterington et al. 1985). As examples, the finite mixture of normal distributions applies a set of "mixing weights" to a specified and finite number of component distributions; the nonparametric DPM of normal distributions relaxes the need to prespecify the number of component distributions and is potentially advantageous in terms of handling smoothing, modality, and uncertainty (Escobar 1994; Mueller and Quintana 2004). The mixture of normal distributions has been extensively used in a variety of important and practical situations, although environmental applications have been very limited (Burmaster and Wilson 2000; Chu et al. 2005; Razzaghi and Kodell 2000; Taylor et al. 2001).

ASSESSMENT OF EXPOSURE TO VOC MIXTURES

An environmental mixture has been defined as the combination of two or more chemical components, regardless of the sources or the spatial or temporal proximity where exposure occurs (U.S. EPA 1986). Environmental exposures typically involve mixtures of pollutants that occur either simultaneously or sequentially and over both short and long periods. Although interest and concern regarding the cumulative effects of mixtures are growing, most pollutant standards, regulations, and guidelines have historically been, and for the most part remain, focused on single pollutants rather than mixtures of pollutants. Several exceptions are worth noting. Environmental regulations control airborne exposures to particulate matter and diesel exhaust (U.S. EPA 2012a,d); occupational exposure limits exist for gasoline vapor (as well as several of its components, like benzene; American Conference of Governmental Industrial Hygienists [ACGIH] 2012); and drinking water regulations collectively limit the four trihalomethanes (U.S. EPA 2013).

As noted earlier, if components of a mixture can interact or jointly contribute to adverse effects, then estimates of adverse effects and risks based on single compounds rather than the mixture may be underestimated. Effects of exposures to a mixture can be directly assessed using empirical data from the actual mixture of concern, or estimated based on data collected from similar mixtures (ATSDR 2004). However, the most common method is to assume interactions or additive effects among the mixture's components. Following the methods recommended to analyze cumulative risks of mixtures (ATSDR 2004; U.S. EPA 2000, 2003), mixture components can be analyzed as having independent toxicities; this means that each chemical has a different mode of action and that the overall response to the mixture is obtained by adding together the responses to each component (called response addition) (Bliss 1939). For example, the cumulative risks of cancer have been estimated using response addition across 13 VOCs and 6 metals (chromium VI, nickel, arsenic, lead, cadmium, and beryllium) (Sax et al. 2006). If components of a mixture have similar toxic effects or mechanisms, then doses (or concentrations) of each component can be added together (called dose addition). An example of dose addition is the use of a toxic equivalency factor for all the polycyclic aromatic hydrocarbons in a mixture; that is, relate the relative potency of each compound in the mixture to the reference compound (e.g., benzo[*a*]pyrene), and thereby weight the dose or concentration of each compound in order to sum an estimate of the mixture's toxicity (U.S. EPA 1993). U.S. EPA (1986)

suggests that if interaction information is unavailable, then the additive assumption should be adopted.

The understanding and analysis of environmental mixtures can be aided by several additional definitions. Three classes of mixtures have been defined (ATSDR 2004): (1) mixtures of compounds that are generated concurrently from the same process (e.g., by-products of fuel combustion or cigarette smoke); (2) intentional mixtures composed of related compounds typically used to manufacture commercial products (e.g., gasoline); and (3) coincidental mixtures of unrelated compounds that are disposed of or stored and reach the same target population (e.g., metals, solvents, and semivolatile wastes at hazardous waste sites). Generated and intentional mixtures may be common in some settings, like in workplaces and homes. However, exposure to multiple air pollutants emitted from different outdoor sources (e.g., carbon monoxide, particulate matter 2.5 microns or smaller in aerodynamic diameter [$PM_{2.5}$], and benzene from vehicles, and sulfur dioxide from power plants) is very common and can be considered a coincidental mixture. Risk evaluations sometimes define simple and complex mixtures (Feron et al. 1998). Simple mixtures contain a relatively small number of components (fewer than 10); many have been identified and their components well quantified (e.g., medicines and pesticides). In contrast, complex mixtures include many more components and are usually incompletely quantified and highly variable (e.g., gasoline vapor and tobacco smoke).

DEPENDENCIES IN VOC MIXTURES ANALYZED WITH COPULAS

The compositions of mixtures, including the relative concentrations of mixture components, can vary considerably. Dependencies among the components of a mixtures refers to the statistical relationships among the concentrations of the components and possibly to the statistical relationship between each component and the composition of the mixture. In general, the most common indicator of a dependency between two variables is a measure of correlation. The Pearson correlation coefficient (r) assumes that variables are normally distributed (Rodgers and Nicewander 1988); nonparametric correlation measures of dependence — most commonly rank correlation measures using the Spearman ρ and Kendall τ — are robust with respect to outliers and can describe some nonlinear relationships. As noted above, environmental exposures often are not normally distributed, can contain extreme values, and can remain skewed toward the upper concentrations even after log-transformation (Jia et al. 2008). Thus, parametric correlation measures can have significant limitations. Both types of correlation measures (parametric and nonparametric) show only pair-wise dependencies (e.g., not those involving three or more variables), however, and may not be reliable indicators in the presence of nonlinear correlations (Schmidt 2006; Staudt 2010).

Copulas are powerful techniques for modeling dependencies that can overcome the shortcomings of conventional correlation measures. Introduced in 1959 by Sklar, a copula describes the dependency structure of two or more variables across the entire distribution (Frees and Valdez 1998; Sklar 1959). Copulas separate the dependency structures from the marginal distributions of the variables (a major advantage) and thus are unconstrained by marginal distributions. Although unrestricted, the choice of the marginal distribution affects the location and scale structure of copulas (Frees and Valdez 1998).

Although copulas have seldom been used for environmental applications, they have been widely applied in the finance world (especially for derivative pricing and financial risk management) in order to deal with market, credit, and operational risks where classical approaches to describe market and other fluctuations (i.e., using multivariate normal distributions) have been lacking (Cherubini et al. 2004; Jean-Frédéric et al. 2004). As noted earlier, given that environmental exposures often deviate from normal or lognormal distributions and include extreme values (Jia et al. 2008), copulas could be a good tool to explore the dependency structures of mixtures. In earlier work, we showed that several types of copulas — specifically the product, Gumbel, Clayton, Frank, and Gaussian forms — fit bivariate dependency structures of VOC exposures for data taken from the NHANES. The VOCs measured in NHANES showed several types of marginal distributions (e.g., lognormal, Pareto, and Weibull) (Jia et al. 2010). Few other environmental applications have been identified.

DETERMINANTS OF VOC EXPOSURES

The phrase "determinants of disease" has been defined as "any factor or variable that can affect the frequency with which a disease occurs in a population" (Putt et al. 1987). Determinants that affect health at individual and community levels can be classified into three groups: the social–economic environment, the physical environment, and a person's individual characteristics and behaviors (World Health Organization [WHO] 2012). In this report, parallels are drawn from these definitions by considering determinants of exposures — factors that affect concentrations of and exposures to pollutants. Like disease determinants, exposure determinants can be grouped into socioeconomic factors (e.g., income level and socioeconomic position), factors related to the physical environment (e.g., meteorology and age of the home), and personal factors (e.g., race or ethnicity and behavior). Although not entirely exclusive, these

groupings provide a structure that may help in understanding and analyzing the factors affecting exposure.

VOC exposures can vary tremendously among individuals. This variation appears to be driven largely by house-to-house variability rather than seasonal, neighborhood, or measurement variability (Jia et al. 2011). In addition, temporal variability may be large at both short and long time scales. Long-term variability includes the actions taken over the past few decades that have reduced emissions of many VOCs (e.g., emission controls and process changes on both stationary and mobile sources) (U.S. EPA 2010b), which partially explain the decline in VOC exposures. Simultaneously, indoor VOC concentrations have fallen in many buildings, a result of reduced or eliminated tobacco smoke, use of low-VOC paints, and other indoor air quality improvements. Short-term variability can include effects of weather, season, and personal activities, and relevant time frames can range from seconds to days. Identification of factors that influence VOC exposures remains incomplete.

A review of 12 studies that examined determinants of VOC exposure is summarized in Appendix Table A.1 (Appendix A is available on the HEI Web site). Below is a summary of those and other studies that emphasizes the results for the general population, although some occupational studies are also included.

Many of the determinants were environmental. Increases in VOC exposures were associated with low AERs and closed windows (D'Souza et al. 2009; Riederer et al. 2009; Sexton et al. 2007; Symanski et al. 2009; Wang et al. 2009); home type (apartment and mobile homes have higher benzene and chloroform levels than single family houses) (Byun et al. 2010; Riederer et al. 2009); fewer years lived in home or newer homes (associated with higher exposure to benzene, toluene, ethylbenzene, and the xylenes [referred to collectively as BTEX]) (D'Souza et al. 2009); and the existence of a fireplace (elevated styrene exposure) (Delgado-Saborit et al. 2009). Also, since chlorine is widely used as a disinfectant to treat public water supplies, households using public water often had higher chloroform (a by-product of chlorine dioxide) exposure than households using well water (D'Souza et al. 2009). In Korea, children had high exposure to traffic-related VOCs (e.g., toluene, ethylbenzene, and *m- & p*-xylenes) in the city with narrower streets and mixed walkways and driveways that increased proximity to traffic (Byun et al. 2010).

VOC exposures clearly are affected by an individual's activities, as shown by many studies (Appendix Table A.1). As examples, smoking and environmental tobacco smoke elevates BTEX and styrene exposures (D'Souza et al. 2009; Delgado-Saborit et al. 2009; Edwards et al. 2001; Kim et al. 2002; Wallace et al. 1989; Wallace 2001), as does being near vehicles (Delgado-Saborit et al. 2009; Hinwood et al. 2007; Kim et al. 2002; Wallace et al. 1989). Pumping gas or being near gasoline increases BTEX and MTBE exposures (D'Souza et al. 2009; Hinwood et al. 2007; Symanski et al. 2009), as does living in a home with an attached garage (D'Souza et al. 2009; Delgado-Saborit et al. 2009; Sexton et al. 2007; Symanski et al. 2009; Wang et al. 2009). The use of paint strippers and thinners also has been associated with higher BTEX exposure (D'Souza et al. 2009; Delgado-Saborit et al. 2009; Symanski et al. 2009). The use of gas heating and gas stoves was associated with increased exposure to aromatic VOCs (e.g., benzene) and a gasoline additive, MTBE (Delgado-Saborit et al. 2009; Kim et al. 2002). (The MTBE associated with an indoor source is unexpected and suggests confounding.) Participation in arts and crafts hobbies increased exposure to toluene, ethylbenzene, and the xylenes (Hinwood et al. 2007), and cooking increased exposure to benzene and toluene in children (Byun et al. 2010). Deodorizer and mothball use increased exposure to 1,4-DCB (D'Souza et al. 2009; Wallace et al. 1989; Wallace 2001) and naphthalene (Batterman et al. 2012). Visiting a dry-cleaner or being near dry-cleaned clothes elevated PERC exposure (D'Souza et al. 2009; Wallace et al. 1989; Wallace 2001). Finally, contact with chlorinated water through drinking tap water, showering and bathing, swimming, and washing dishes and clothes has been shown to increase exposure to chloroform (D'Souza et al. 2009; Sexton et al. 2007; Wallace et al. 1989; Wallace 2001).

A modest number of socioeconomic factors have been identified. VOC exposure has been related to ethnicity; for example, people of Hispanic origin had higher exposure to BTEX, MTBE, and 1,4-DCB; black study participants had higher exposure to 1,4-DCB, PERC, and chloroform (Riederer et al. 2009; Wang et al. 2009); and Mexican participants had higher exposure to benzene and 1,4-DCB when compared with non-Hispanic white participants (Wang et al. 2009). In the NHANES VOC data set, Hispanic and black adults were exposed to higher levels of BTEX, MTBE, and 1,4-DCB than non-Hispanic white participants after controlling for environmental and personal covariates; this suggests possible cultural differences (D'Souza et al. 2009). Occupation clearly affects exposure; for example, higher BTEX exposure has been linked to service station and vehicle repair jobs (Jo and Song 2001), and α-pinene, β-pinene, *d*-limonene, toluene, ethylbenzene, and styrene have been associated with cleaning jobs (Wolkoff et al. 1998). However, effects of occupation on VOC exposures in population studies have rarely been observed. Machine-related jobs have been linked to higher BTEX exposure (D'Souza et al. 2009); and time at work or school has been associated with higher exposure to benzene, ethylbenzene, the xylenes, and PERC (Wang et al. 2009). Increased levels of education and income have been associated with a

decrease in exposures to benzene, 1,4-DCB, PERC, and chloroform (Wang et al. 2009). This might suggest that persons of higher socioeconomic position participate in fewer high-exposure activities (e.g., house cleaning), reside in cleaner homes and neighborhoods (e.g., distant from traffic), and commute to and work in cleaner environments.

In broad terms, many socioeconomic factors are expected to be correlated with environmental factors as yet to be identified; if this is true, those correlated socioeconomic–environmental factors may be more direct determinants of concentrations or exposures.

Although many exposure determinants have been identified, the underlying studies have several limitations and the significance and applicability of the determinants are uncertain. First, many of the studies used small samples; for example, a Birmingham study enrolled only 12 adults (Kim et al. 2002) and a Minneapolis–St. Paul study enrolled 70 adults (Sexton et al. 2007). Observational studies, especially cross-sectional studies, require large sample sizes to disentangle contributions of personal activities and indoor and outdoor environments. Second, the studies had important data gaps; for example, although the NHANES sample was large (personal VOC concentrations measured for 646 individuals) and designed to be nationally representative (National Center for Health Statistics [HCHS] 2012b), outdoor and indoor concentrations, time–activity patterns, and other information were not collected. However, as mentioned, the RIOPA study (Weisel et al. 2005a) collected outdoor, indoor, and personal VOC measurements, along with considerable other information, and it provides a good opportunity to characterize determinants of VOC exposure.

SPECIFIC AIMS

The overall objectives of this project were to (1) identify and characterize exposure distributions (including extreme values), (2) evaluate mixtures (including dependencies among components), and (3) identify determinants of VOC exposure. We used primarily the RIOPA study data set and, when appropriate for comparison, the NHANES data set.

AIM 1

This aim focused on characterizing exposure and risks of exposure to individual VOCs.

A combination of standard and extreme value distributions can best characterize the distribution of pollutant exposures. Work included fitting univariate full distribution models for outdoor, indoor, and personal VOC observations from the RIOPA study, fitting extreme value distributions to the highest 10% and 5% of measurements for each VOC, and estimating risks of extreme value exposures. The results include a comparison of distribution fitting for the RIOPA and NHANES data sets: The main text of this report gives results for personal exposures from RIOPA and some comparisons with NHANES; and Appendix A gives results for indoor and outdoor concentrations from RIOPA, which shared a number of similarities, and further analyses of the NHANES data set.

To Specific Aim 1, we added fitting mixture models in response to comments from the HEI Health Research Committee. For this, we fitted mixture distribution models that could take into account values below detection limits, extreme values, and values in the middle of the distribution.

AIM 2

This aim focused on characterizing exposure and risks of exposure to mixtures of VOCs.

Copulas and other advanced statistical techniques that model multivariate exposure distributions can allow accurate and efficient characterization of VOC mixtures, joint distributions, and dependency structures. This task focused on identifying common and high-priority mixtures of different pollutants and evaluating their effects and risks, especially those for highly exposed individuals. We selected exposure mixtures on the basis of emission sources and toxicity followed by estimating the joint distributions and dependency structures of the mixtures.

AIM 3

This aim focused on characterizing exposure determinants for individual VOCs.

We investigated determinants of VOC exposures to identify the effects of indoor sources (e.g., smoking, attached garages, use of moth repellents), time–activity patterns (e.g., time spent outdoors, in traffic), and socioeconomic, demographic, meteorologic, and other factors. We used LMMs to identify sources and determinants of indoor, outdoor, and personal exposures. We originally proposed to use quantile regression models, but given the repeated measurements available in the RIOPA study, we decided that LMMs would be more effective in identifying exposure determinants.

DATA SOURCES

RELATIONSHIPS OF INDOOR, OUTDOOR, AND PERSONAL AIR STUDY

The RIOPA study included three cities (Elizabeth, NJ; Houston, TX; Los Angeles, CA) that were selected on the basis of differences in the emission sources likely to affect pollutant exposure. Institutional Review Board approval was required and participants' consent forms were obtained. Homes near outdoor emissions were oversampled in order to estimate outdoor contributions to personal exposures (Weisel et al. 2005b). A total of 306 households in which 309 adults and 118 children resided and did not smoke were recruited in the three cities and studied from summer 1999 to spring 2001. Each household and participant was sampled twice about 3 months apart. Outdoor, indoor, and personal air samples were collected using 48-hour sampling periods. VOCs were collected using passive samplers (OVM3500, 3M Company, St. Paul, MN, USA) and analyzed by gas chromatography–mass spectrometry for 18 compounds (benzene [C_6H_6], toluene [C_7H_8], ethylbenzene [C_8H_{10}], *m- & p*-xylenes [C_8H_{10}], *o*-xylene [C_8H_{10}], MTBE [$C_5H_{12}O$], styrene [C_8H_8], 1,4-DCB [$C_6H_4Cl_2$], methylene chloride [MC] [CH_2Cl_2], TCE [C_2HCl_3], PERC [C_2Cl_4], chloroform [$CHCl_3$], CTC [CCl_4], *d*-limonene [$C_{10}H_{16}$], α-pinene [$C_{10}H_{16}$], β-pinene [$C_{10}H_{16}$], 1,3-butadiene [C_4H_6], and chloroprene [C_4H_5Cl]).

Over the course of the original RIOPA study, the following changes were made for data analysis: data for 1,3-butadiene and chloroprene were not reported due to low recovery; and the MC measurements were excluded due to measurement issues (inconsistent blank contributions) (Weisel et al. 2005b). The styrene measurements had higher uncertainty than those for the other VOCs in an inter-laboratory comparison (Weisel et al. 2005c), but were included nevertheless. We defined a new variable — total VOCs — as the sum of the remaining 15 VOCs. MDLs ranged from 0.21 (α-pinene and PERC) to 7.1 (toluene) $\mu g/m^3$. Table 1 shows that detection frequencies for outdoor measurements ranged from 6.3% (β-pinene) to 96.8% (CTC); for indoor measurements they ranged from 25.8% (TCE) to 95.5% (CTC); and for personal measurements they ranged from 31.4% (TCE) to 96.1% (MTBE) (Weisel et al. 2005b). Measurements below the MDLs were replaced with one-half of the MDL. The RIOPA study design is described and other measurements (e.g., carbonyls and $PM_{2.5}$) and results are reported elsewhere (Turpin et al. 2007; Weisel et al. 2005a,b).

The RIOPA investigators administered to their participants three questionnaires that were based on the National Human Exposure Assessment Survey; in total, information was gathered for more than 500 possible exposure determinants. A baseline questionnaire collected demographic and lifestyle factors (e.g., ethnicity, employment, opening windows, and use of deodorizer or fresheners); a technician walk-through questionnaire collected neighborhood and household characteristics (e.g., industrial emissions in neighborhood, type of building, and existence of attached garage); and a third questionnaire collected time–activity information for each participant (e.g., time spent indoors at school or work, frequency of pumping gas, bathing or showering, and gardening; Weisel et al. 2005a). Household AERs and geographic and meteorologic information (e.g., city, outdoor temperature, wind speed, and relative humidity) were also obtained for each household.

NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEY

We compared the RIOPA study cohort with the 1999–2000 cohort of the NHANES, which included personal VOC measurements for 851 participants (NCHS 2012a). (Institutional Review Board and participants' consent forms were obtained for the study.) NHANES used 48- to 72-hour exposure periods and participants were selected using a stratified, multistage cluster design; therefore, analyses used weights to obtain representative national averages. The RIOPA study and NHANES shared 10 VOCs (benzene, toluene, ethylbenzene, *m- & p*-xylenes, *o*-xylene, MTBE, 1,4-DCB, TCE, PERC, and chloroform). Although the study purposes and recruitment strategies differed, NHANES and RIOPA used similar sampling methods and their study periods overlapped. In NHANES, data for four participants were deleted — participant IDs 468 and 578 who had excessively long sampling periods, and participant IDs 3852 and 4076 who had extremely high concentrations of benzene, the xylenes, or toluene; these exclusions are also described by Jia and colleagues (2008). In total, 638–651 observations were available for these analyses. Detection frequencies for the personal exposure measurements ranged from 24.8% (TCE) to 96.4% (*m- & p*-xylenes) (Table 1).

Table 1. Summary of VOC Observations in the RIOPA Study and NHANES[a]

	RIOPA (1999–2001)										NHANES (1999–2000)				
	Outdoor (n = 540)					Indoor (n = 539)					Personal (n = 544)					Personal (n = 638 to 651)				
VOC	DF (%)	Mean	SD	50th	95th	DF (%)	Mean	SD	50th	95th	DF (%)	Mean	SD	50th	95th	DF (%)	Mean	SD	50th	95th
Benzene	75.5	2.15	2.11	1.68	5.16	82.1	3.50	5.15	2.19	10.03	85.5	3.64	5.31	2.39	10.73	65.7	5.29	18.09	2.89	17.82
Toluene	33.9	6.83	6.54	3.56	19.63	70.0	15.26	24.48	10.41	39.79	75.0	19.12	37.31	12.42	50.10	93.3	38.89	123.80	16.16	98.27
Ethylbenzene	75.5	1.28	1.87	0.93	3.04	84.8	2.52	4.74	1.46	7.62	85.8	2.78	5.13	1.68	7.45	93.1	9.23	55.66	2.33	21.01
m- & p-Xylenes	84.7	3.56	4.16	2.49	10.02	89.7	7.32	15.87	4.07	22.18	91.5	8.07	15.49	4.42	22.60	96.4	21.99	122.10	6.03	60.15
o-Xylene	73.3	1.46	3.90	0.96	3.23	81.8	2.47	4.78	1.46	7.24	87.1	2.87	5.59	1.72	7.92	93.2	7.41	31.78	2.16	23.01
MTBE	96.4	8.11	9.99	5.32	22.09	93.9	11.79	27.29	5.98	36.00	96.1	14.77	42.67	7.14	42.05	29.8	5.19	25.94	0.61	23.80
Styrene	16.5	0.58	2.06	0.42	1.29	45.7	1.47	4.24	0.42	5.13	48.5	1.55	4.31	0.42	5.32	—	—	—	—	—
1,4-DCB	28.5	2.15	17.16	0.46	3.66	64.9	68.84	303.76	1.40	343.88	71.5	56.83	229.40	1.88	304.60	68.2	27.18	119.60	2.22	257.30
TCE	20.0	0.34	1.30	0.22	0.80	25.8	0.97	7.19	0.22	1.73	31.4	1.44	10.74	0.22	2.37	24.8	3.43	27.76	0.33	7.21
PERC	68.1	1.02	2.17	0.61	3.17	81.4	1.85	4.53	0.82	6.03	87.5	7.17	112.40	0.89	6.82	69.5	5.20	35.03	0.79	17.17
Chloroform	12.8	0.37	1.43	0.21	0.79	76.7	1.86	2.97	0.94	5.97	82.7	4.25	52.49	1.04	6.36	80.3	2.74	5.36	1.15	10.96
CTC	96.8	0.72	1.31	0.64	1.00	95.5	0.71	0.99	0.62	1.11	95.4	0.80	2.44	0.62	1.06	—	—	—	—	—
d-Limonene	20.4	1.97	6.34	0.64	6.54	86.6	31.59	108.50	9.95	103.10	88.4	41.14	238.90	11.77	112.20	—	—	—	—	—
α-Pinene	24.3	1.31	4.16	1.02	2.23	62.6	7.14	14.76	2.67	26.83	66.3	6.85	16.25	2.88	22.48	—	—	—	—	—
β-Pinene	6.3	0.94	2.15	0.51	1.26	52.2	4.95	11.08	1.27	20.77	56.2	5.53	13.07	1.52	21.46	—	—	—	—	—

[a] Detection frequency (DF) is the estimated proportion of observations above the MDL. Means ± SDs and 50th and 95th percentiles of VOC observations are given in μg/m^3. — indicates not measured.

SPECIFIC AIM 1. EXTREME VALUE ANALYSES AND FULL DISTRIBUTION MODELS

METHODS: EXTREME VALUE ANALYSES

Risk Evaluation for Extreme Value Exposures

Screening-level estimates of cancer risks were estimated using standard approaches. The URFs for the VOCs were taken from the U.S. EPA's Integrated Risk Information System (IRIS; U.S. EPA 2012b), the Office of Environmental Health Hazard Assessment's [OEHHA] Air Toxics Hot Spots Program Risk Assessment Guidelines (OEHHA 2005), or EPA's Cumulative Exposure Project (Caldwell et al. 1998). Each URF and its basis are shown in Appendix Table A.2, along with the RfC and toxic endpoints. URFs were not available for toluene, *m*- & *p*-xylenes, *o*-xylene, *d*-limonene, α-pinene, and β-pinene, and these VOCs were therefore excluded from the extreme value analyses in the current study.

The VOC measurements from the two visits for each adult in the RIOPA study were averaged as an estimate of the long-term exposure concentration. An individual's (*i*) excess lifetime cancer risk for a specific VOC was calculated as:

$$R_i = C_i \, URF_i, \tag{1}$$

where R_i is excess individual lifetime cancer risk (probability), C_i is exposure concentration (μg/m^3), and URF_i is unit risk factor (cancer cases per μg/m^3 change in VOC concentration).

Following guidance for analyzing mixtures (U.S. EPA 2000), risks were calculated by response addition for those VOCs that cause the same toxic effect on the same target organ. In this case, results of equation 1 were summed for each participant for the several chemicals in one of three mixtures: VOCs associated with blood cancers (lymphomas and leukemia), which included benzene, MTBE, 1,4-DCB, TCE, and PERC; VOCs associated with liver and renal tumors, which included ethylbenzene, MTBE, 1,4-DCB, TCE, PERC, chloroform, and CTC; and total VOCs (Borgert et al. 2004; International Agency for Research on Cancer [IARC] 2012). Total VOCs also served as a general indicator of VOC exposure and was used to identify the dominant contributors to risks.

The cumulative risk of exposure to a mixture was computed for each subject by summing the risks of components in the mixture; extreme values of the cumulative risk were taken as the top 10% and top 5% of this sum over all persons.

Fitting of Generalized Extreme Value Distributions

To focus on the health risk of the highest VOC exposures, extreme value distributions were applied to personal VOC exposure observations from the RIOPA study and to the NHANES data set. A broad class of extreme value distributions, the GEV distribution (Jenkinson 1955), was fitted to each extrema data set (the highest 10% and 5% of VOC exposures). The GEV probability density function is expressed as:

$$f_{\xi,\mu,\sigma}(x) = \left[\left(\left\{1+\left[\xi(x-\mu)/\sigma\right]\right\}^{-1-1/\xi}\right)/\sigma\right] \exp\left(-\left\{1+\left[\xi(x-\mu)/\sigma\right]\right\}^{-1/\xi}\right) \text{ if } \xi \neq 0, \tag{2}$$

where ξ is the shape parameter, μ is the location parameter, σ is the scale parameter, and x is observation data. If $\xi > 0$, the GEV distribution belongs to the Fréchet family; if $\xi < 0$, the GEV distribution belongs to the Weibull family (Jenkinson 1955); and if $\xi = 0$, the GEV distribution belongs to the Gumbel family, which permits this simplification of equation 2:

$$f_{0,\mu,\sigma}(x) = \left\{\left[e^{-(x-\mu)/\sigma}\right]/\sigma\right\}\exp\left[-e^{-(x-\mu)/\sigma}\right]. \tag{3}$$

The three parameters of the GEV distribution were determined by maximum likelihood estimate (MLE), and goodness of fit was examined using the Kolmogorov-Smirnov (K-S) test with the null hypothesis that the data subset comes from the GEV distribution. Empirical *P* values were calculated for the repeated (bootstrap) samples in the NHANES weighted data set.

For GEV distribution fitting, only adult personal VOC measurements were estimated because they are likely to be most representative of exposure. We selected adult subjects due to the larger sample size — 544 measurements for 305 participants (299 and 245 measurements at first and second visits, respectively, of which 239 adults had valid samples for both visits). Child exposures were not used due to the smaller sample size and because several households included measurements from several children (only one adult was sampled in a household), which would cause a cluster effect. Since we were most concerned with the risks of long-term exposure (concentrations were too low to study acute effects), we used the average of measurements from the two visits for each participant. Next, we identified outliers (defined as a value that was at least twice that of the next highest observation) and influential observations (identified as observations that clearly altered statistical results). Observations that were both outliers and influential were excluded in subsequent extreme value analyses; these few observations are noted in the Results and Discussion: Extreme Value Analyses section. The sample sizes of the final top 10% and 5% of observed VOC concentrations were 24 and

12, respectively. In addition to the broader GEV distributions, the Gumbel distribution (a two-parameter form of a GEV distribution that includes location and scale parameters) was fitted using probability plots. This approach is described in more detail in Appendix A.

Extreme Value Simulations

For further evaluation, we generated simulated extreme value data sets (n = 10,000) that would match the top 10% and 5% of observations for each personal adult VOC exposure in the RIOPA data set, but that would follow the fitted GEV or Gumbel distribution. Because lognormal distributions are commonly used for exposure data, simulated data (n = 10,000) were also generated for a lognormal distribution to match the full set of RIOPA observations; it was fitted by MLE, and the evaluation was focused on extrema. The simulated data were then compared with the observed data in RIOPA using K-S tests and graphical analyses, and P values were estimated.

Finally, in an analysis oriented toward risk assessment, we compared the fractions of persons with cancer risks that would exceed 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, and 10^{-2} cutoffs based on the three sets of simulated data with the fractions based on the RIOPA observations. These analyses were conducted for both individual VOCs and mixtures.

Distribution fitting, simulations of GEV, Gumbel, and lognormal distributions used the functions gev, rgev, rgumbel, fitdistr, and rlnorm in R, version 2.13.1 (R Development Core Team, Vienna, Austria) and Excel (Microsoft, Redmond, WA).

RESULTS AND DISCUSSION: EXTREME VALUE ANALYSES

Results for the personal exposure analyses are reported here; those for indoor and outdoor exposures are presented in Appendix A. Results for univariate full distribution models for outdoor, indoor, and personal VOCs are shown in Appendix Table A.3 and Figures A.1–A.5.

Outliers

Several observations in the RIOPA data set were identified as both influential (strongly affecting the fitting or not fitting of extreme value distributions) and as outliers (being at least twice as high as the next observation). The averaged personal measurements in RIOPA included six outliers in five different households for five VOCs (ID code NJ090 with MTBE = 440 μg/m^3; NJ063 with TCE = 102 μg/m^3, TX050 with TCE = 97 μg/m^3; NJ063 with PERC = 1,340 μg/m^3; NJ080 with CTC = 20 μg/m^3; TX068 with d-limonene = 1,462 μg/m^3). The two TCE outliers had similar concentrations, but both were four times higher than the next value. For the first-visit measurements in RIOPA (for comparison with NHANES), six influential outliers were identified (CA079 with benzene = 85 μg/m^3, toluene = 641 μg/m^3, chloroform = 1,224 μg/m^3, and d-limonene = 5,114 μg/m^3; NJ063 with PERC = 2,618 μg/m^3; and NJ090 with MTBE = 844 μg/m^3).

Generally, users of secondary data have limited ability to ascertain whether an observation is erroneous or representative, and thus should be cautious in deleting observations. Outliers may be erroneous measurements that were not removed or corrected in earlier quality assurance or data cleaning steps. For example, VOC measurements may be erroneous due to issues in the preparation, transport, collection, storage, or analysis of the sample, as well as to data management. Many of these issues should be detected by quality assurance activities, which would flag or delete such observations in the original data set.

In exposure studies, measurements that are not representative of a subject's exposure also may be considered to be erroneous. For example, at night, a subject might have placed the sampler on their recently polished bed-stand (off-gassing d-limonene and aliphatic VOCs) or near perfume bottles (possibly emitting many aromatic VOCs), or may have worn the sampler on recently dry-cleaned clothes (off-gassing PERC). Such situations could produce high VOC measurements that may not represent the subject's exposure and, for the most part, cannot be identified with the variables in the RIOPA data set.

However, if unusually high observations are found without much replication in the data set (e.g., a very limited number of observations appear highly elevated), then deleting such observations provides relatively little loss in generalization and considerable gain in the likelihood that the analysis is representative. In the present analysis we excluded only a small number of observations (two groups of 6 data each) that met two stringent criteria: both (1) not fitting the extreme value distributions and (2) exceeding by a factor of 2 the next highest observation. We believe we eliminated potentially erroneous measurements (due to either sampling errors or not being representative of exposure), thus improving the modeling of extreme value distributions.

Predicted Health Risks for Extreme VOC Exposures

Estimates of individuals' excess lifetime cancer risks for the median and the 90th and 95th percentile concentrations are shown in Table 2. Using median concentrations, benzene, 1,4-DCB, and chloroform presented the highest (and very similar) risks: 2.0, 2.5, and 2.9 × 10^{-5}, respectively; risks for other VOCs were below 10^{-5}. For the 95th percentile concentrations, the same three VOCs also presented the

Table 2. Predicted Excess Cancer Risks for Adult Participants in the RIOPA Study According to Percentiles of VOC Observations[a]

		Predicted Excess Cancer Cases per Million Population									
					Percentile of VOC Observations						
VOC[b]	Unit Risk per μg/m^3	Mean	SD	Min	25	50	75	90	95	98	Max
Benzene	7.8×10^{-6}	28.4	25.9	4.3[c]	13.5	20.4	32.7	53.0	76.6	134.2	172.6
Ethylbenzene	2.5×10^{-6}	7.1	9.9	0.9[c]	3.0	4.4	7.6	13.0	19.0	43.2	82.9
MTBE	2.6×10^{-7}	3.5	4.6	0.1[c]	1.2	2.1	4.1	6.6	11.6	17.5	37.2
Styrene	2.0×10^{-6}	3.2	6.9	0.3[c]	0.8[c]	1.5	2.6	5.8	12.9	23.9	59.9
1,4-DCB	1.1×10^{-5}	626.5	2223.0	2.4[c]	10.0[c]	24.5	126.0	908.9	3620.7	9518.1	19,167.0
TCE	2.0×10^{-6}	1.4	4.1	0.2[c]	0.2[c]	0.4[c]	0.93	2.2	4.6	16.1	40.9
PERC	5.9×10^{-6}	12.9	25.9	0.7[c]	2.5[c]	5.9	11.8	24.1	47.1	97.5	242.3
Chloroform	2.3×10^{-5}	47.0	62.2	3.2[c]	14.5	28.9	52.6	97.1	147.5	248.8	537.6
CTC	1.5×10^{-5}	9.8	2.9	2.0[c]	8.2	9.3	10.7	12.9	15.0	17.1	27.8
Hematopoietic mixture[d]	NA	680.2	2239.7	12.78	44.89	76.4	180.22	965.4	3651.5	9695.8	19,195.8
Liver and kidney toxicant mixture[e]	NA	714.8	2247.4	20.80	61.25	111.1	265.03	1102.2	3683.6	9723.1	19,222.9
Total VOCs	NA	745.8	2253.9	34.10	83.90	141.1	293.30	1125.0	3710.1	9780.5	19,250.0

[a] n = 239 adults with valid samples at both visits. NA indicates not available.

[b] Of the 15 VOCs studied, 6 did not have unit risks available and were excluded from this set of analyses.

[c] Concentration was based on MDL.

[d] Hematopoietic mixture included benzene, MTBE, 1,4-DCB, TCE, and PERC.

[e] Liver and kidney toxicant mixture included ethylbenzene, MTBE, 1,4-DCB, TCE, PERC, chloroform, and CTC.

highest risks, 1.5×10^{-4}, 3.6×10^{-3}, and 7.7×10^{-5}, respectively; risks above 10^{-5} were also associated with ethylbenzene, MTBE, styrene, PERC, and CTC. Among the VOCs analyzed, 1,4-DCB presented the greatest risks; for example, for the top 10% of observations, all individuals had risks exceeding 10^{-4}, 88% exceeded 10^{-3}, and 13% exceeded 10^{-2} (see Table 4 discussed later). In addition, 1,4-DCB's share of the total carcinogenic risk (the sum of risks across individual VOCs) increased greatly at higher percentiles; for example, 1,4-DCB was associated with 17% of the total VOC risk using median concentrations, 81% using 90th percentile concentrations, and 98% using 95th percentile concentrations (Table 2). As discussed later, the dominance of 1,4-DCB is partly a function of the specific VOCs measured.

Predicted risks for the three VOC mixtures are also shown in Table 2. For the hematopoietic toxicity mixture, the median and 95th percentile risks were 7.6×10^{-5} and 3.7×10^{-3}, respectively, most of which was due to benzene and 1,4-DCB among the five VOCs (benzene, MTBE, 1,4-DCB, TCE, and PERC) in this mixture. For the liver and renal toxicity mixture, the median and 95th percentile risks were 1.1×10^{-4} and 3.7×10^{-3}, respectively, mostly contributed by 1,4-DCB and chloroform among the seven VOCs (ethylbenzene, MTBE, 1,4-DCB, TCE, PERC, chloroform, and CTC) in this mixture.

These risks and hazard quotients represent preliminary screening-level predictions and have several limitations. First, they include only a subset of VOCs among those known or suspected to be toxicants; for example, RIOPA did not include naphthalene, which is associated with anemia (ATSDR 2005a), or reliable measurements of 1,3-butadiene, which is associated with blood and lymphatic system cancers (ATSDR 2009). Second, the two personal exposure measurements averaged together for each RIOPA participant may not be a robust measure of lifetime average exposure. Third, the uncertainty in the RfC and URF is considerable, and the values used are believed to be conservative. Finally, the exposure measurements represent multiday averages; shorter-term exposures (1–24 hours) can be higher and could possibly exceed the RfC or other guidance levels for acute effects.

Extreme Value Distributions for the RIOPA Data

Figure 1 shows distributions of cumulative cancer risks for four VOCs for simulated data matching GEV, Gumbel, and lognormal distributions, as well as the observed RIOPA data. Separate plots are shown for the top 10% and 5% of VOC exposure. (Parameters of GEV distributions fitted to the RIOPA VOC data and goodness-of-fit statistics are provided in Appendix Table A.5, and probability plots depicting the fit to Gumbel distributions are shown in Appendix Figures A.6 through A.9.)

The GEV simulated distributions closely fitted both the top 10% and 5% of RIOPA VOC observations based on K-S tests (Table 3). With the exception of the top 5% of benzene concentrations, the shape parameters of the GEV distributions were close to or larger than 0, indicating Gumbel or Fréchet distributions, and the location and scale parameters reflected the high percentile concentrations (Appendix Table A.5).

Although the GEV distributions closely fitted the extrema for both individual VOCs and the three VOC mixtures, simulations sometimes produced extremely high values that greatly overpredicted maxima (e.g., concentrations > 20,000 μg/m^3). This occurred for the top 10% of ethylbenzene, styrene, 1,4-DCB, TCE, and PERC concentrations, and the top 5% of ethylbenzene, MTBE, styrene, 1,4-DCB, TCE, and chloroform concentrations. These problems were limited to the extreme upper tails of the distributions (e.g., values above the 98th or 99th percentile).

Gumbel distributions fitted extrema for several of the VOCs (e.g., top 10% and 5% of benzene, ethylbenzene, MTBE, styrene, 1,4-DCB, PERC, and chloroform concentrations) based on K-S tests. Sometimes the lowest Gumbel-simulated values (i.e., the lower tail) were lower than the observed data, and some values were even negative (the plots in Figure 1 are truncated and do not make this visible). (For more results related to fitting Gumbel distributions, see Appendix A, Gumbel Distribution Fitting section and Table A.4.)

Lognormal distributions also fitted extrema for several VOCs (e.g., top 10% of benzene and ethylbenzene levels; the top 10% and 5% of MTBE, PERC, and chloroform; and the top 5% of CTC; Table 3). However, these distributions typically diverged from the RIOPA observations, and the upper extreme values were greatly unrepresented (Figure 1). Note that the lognormal distributions were fitted for the full data set, not just the top 10% and 5% used for the GEV and Gumbel distributions.

The predicted fractions of individuals with risks that would exceed 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, and 10^{-2} (risk cutoffs typically used in cancer risk assessment) are listed in Table 4. This analysis was performed for simulated data

Table 3. Adult VOC Distributions for RIOPA Adult VOC Observations Compared with GEV, Gumbel, and Lognormal Distributions of Simulated Data Using K-S Tests[a]

	GEV Simulation				Gumbel Simulation				Lognormal Simulation[b]			
	Top 10%		Top 5%		Top 10%		Top 5%		Top 10%		Top 5%	
VOC	Statistic	*P* Value	Statistic	*P* Value	Statistic	*P* Value	Statistic	*P* Value	Statistic	*P* Value	Statistic	*P* Value
Benzene	0.13	0.823	0.24	0.482	0.17	0.527	0.23	0.549	0.20	0.313	0.40	**0.037**
Ethylbenzene	0.08	0.996	0.14	0.979	0.21	0.228	0.17	0.899	0.22	0.204	0.44	**0.014**
MTBE	0.09	0.987	0.14	0.975	0.27	0.065	0.36	0.083	0.17	0.533	0.26	0.355
Styrene	0.18	0.450	0.15	0.949	0.18	0.423	0.23	0.528	0.41	**0.001**	0.76	**< 0.001**
1,4-DCB	0.10	0.976	0.14	0.970	0.15	0.667	0.15	0.943	0.51	**< 0.001**	0.64	**< 0.001**
TCE	0.10	0.967	0.18	0.822	0.44	**< 0.001**	0.46	**0.014**	0.38	**0.003**	0.65	**< 0.001**
PERC	0.11	0.939	0.11	0.998	0.16	0.603	0.18	0.855	0.18	0.417	0.36	0.067
Chloroform	0.09	0.983	0.17	0.900	0.17	0.467	0.19	0.789	0.13	0.833	0.26	0.357
CTC	0.14	0.747	0.15	0.954	0.47	**< 0.001**	0.52	**0.003**	0.33	**0.011**	0.17	0.816

[a] Goodness of fit was analyzed with the K-S test; K-S results and *P* values are shown. For the top 10% of RIOPA VOC observations, $n = 24$; for the top 5%, $n = 12$. For simulated data sets, $n = 10{,}000$. *P* values < 0.05 are shown in **bold** type; *P* values > 0.05 indicate that no significant difference was found between the two distributions.

[b] Lognormal distributions were fitted for the full data set, not only the top 10% and 5% used for the GEV and Gumbel distributions.

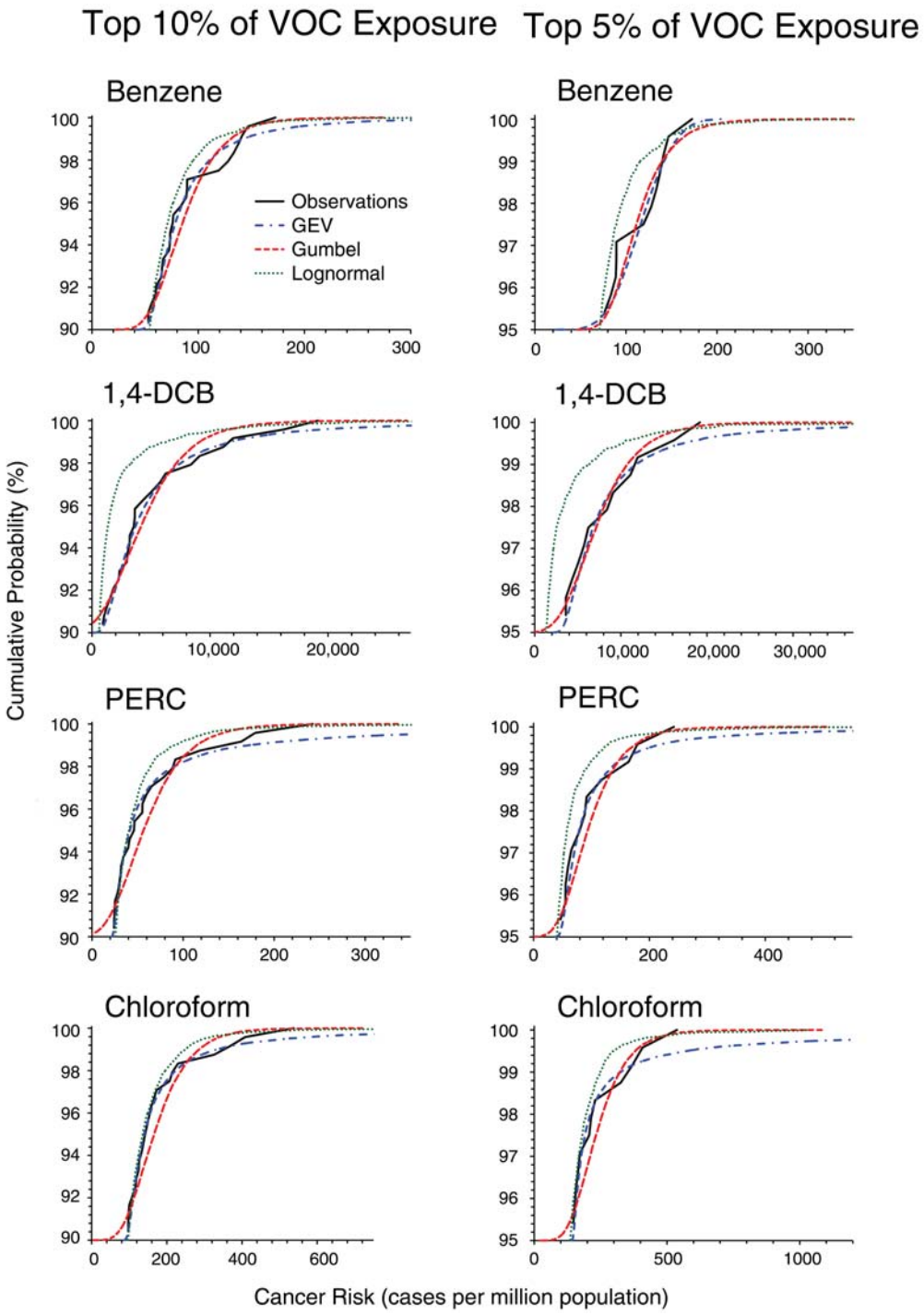

Figure 1. Comparison of cancer risks for top 10% and 5% of VOC exposures using RIOPA observations (n = 24 for top 10% and n = 12 for top 5%), and GEV, Gumbel, and lognormal distributions of simulated data sets (n = 10,000 for simulations).

Table 4. Comparison of Predicted Cancer Risks Based on RIOPA Observations and the GEV, Gumbel, and Lognormal Distributions of Simulated Data[a]

		Percentage of Individuals Estimated to Exceed Indicated Cancer Risk									
		Top 10% of VOC Exposure					Top 5% of VOC Exposure				
VOC	Distribution of RIOPA or Simulated Data	1×10^{-6}	1×10^{-5}	1×10^{-4}	1×10^{-3}	1×10^{-2}	1×10^{-6}	1×10^{-5}	1×10^{-4}	1×10^{-3}	1×10^{-2}
Benzene	Observations	100	100	29	0	0	100	100	58	0	0
	GEV	100	100	26	0	0	100	100	71	0	0
	Gumbel	100	100	31	0	0	100	100	67	0	0
	Lognormal	100	100	18	0	0	100	100	35	0	0
Ethylbenzene	Observations	100	100	0	0	0	100	100	0	0	0
	GEV	100	100	7	1	0	100	100	8	0	0
	Gumbel	100	91	0	0	0	100	98	1	0	0
	Lognormal	100	100	0	0	0	100	100	0	0	0
MTBE	Observations	100	63	0	0	0	100	100	0	0	0
	GEV	100	57	1	0	0	100	100	3	0	0
	Gumbel	98	74	0	0	0	99	87	0	0	0
	Lognormal	100	53	0	0	0	100	100	0	0	0
Styrene	Observations	100	54	0	0	0	100	100	0	0	0
	GEV	100	46	6	1	0	100	100	5	0	0
	Gumbel	96	69	0	0	0	100	93	0	0	0
	Lognormal	100	28	0	0	0	100	55	0	0	0
1,4-DCB	Observations	100	100	100	88	13	100	100	100	100	25
	GEV	100	100	100	96	13	100	100	100	100	27
	Gumbel	96	96	95	89	7	100	100	100	99	24
	Lognormal	100	100	100	65	5	100	100	100	100	10
TCE	Observations	100	21	0	0	0	100	42	0	0	0
	GEV	100	18	2	0	0	100	33	7	2	0
	Gumbel	77	61	1	0	0	83	74	9	0	0
	Lognormal	100	2	0	0	0	100	3	0	0	0
PERC	Observations	100	100	17	0	0	100	100	33	0	0
	GEV	100	100	18	2	0	100	100	32	1	0
	Gumbel	99	96	16	0	0	100	100	44	0	0
	Lognormal	100	100	8	0	0	100	100	16	0	0
Chloroform	Observations	100	100	88	0	0	100	100	100	0	0
	GEV	100	100	93	2	0	100	100	100	6	1
	Gumbel	100	100	86	0	0	100	100	98	0	0
	Lognormal	100	100	93	0	0	100	100	100	0	0
CTC	Observations	100	100	0	0	0	100	100	0	0	0
	GEV	100	100	0	0	0	100	100	1	0	0
	Gumbel	96	81	0	0	0	89	78	4	0	0
	Lognormal	100	100	0	0	0	100	100	0	0	0
Hematopoietic mixture	Observations	100	100	100	96	17	100	100	100	100	33
	GEV	100	100	100	97	14	100	100	100	100	27
	Gumbel	97	97	96	90	10	100	100	100	99	30
	Lognormal	100	100	100	79	2	100	100	100	100	4
Liver and kidney toxicant mixture	Observations	100	100	100	100	17	100	100	100	100	33
	GEV	100	100	100	97	14	100	100	100	100	26
	Gumbel	97	97	97	91	10	100	100	100	99	31
	Lognormal	100	100	100	88	1	100	100	100	100	3
Total VOCs	Observations	100	100	100	100	17	100	100	100	100	33
	GEV	100	100	100	98	13	100	100	100	100	27
	Gumbel	97	97	96	92	11	100	100	100	100	32
	Lognormal	100	100	100	97	1	100	100	100	100	1

[a] For the top 10% of RIOPA values, $n = 24$; for the top 5%, $n = 12$; $n = 10{,}000$ in simulated data sets.

matching GEV, Gumbel, and lognormal exposure distributions, as well as for the top 10% and 5% of the RIOPA exposure observations. The extreme value distributions for simulated data closely matched the distributions for observed data and differences were usually within a few percent. As an example, for the top 10% of the benzene data, the RIOPA observations and the GEV, Gumbel, and lognormal simulations produced individual excess risk levels exceeding 10^{-4} for 29%, 26%, 31%, and 18% of the population, respectively. As a second example, using the top 5% of 1,4-DCB values exceeding 10^{-2}, the corresponding frequencies were 25%, 27%, 24%, and 10%. As noted earlier, GEV simulations sometimes overpredicted at the very highest percentiles (> 99th percentile; seen at the 10^{-4} to 10^{-2} risk level for ethylbenzene, MTBE, styrene, TCE, PERC, chloroform, and CTC) when comparing with the observed data. However, such cases were rare, comprising less than about 1% of the entire data set.

Gumbel distributions also overpredicted higher extrema (data not shown) and underpredicted the lower risks (both top 10% and 5% of data), in part due to its unbounded nature that can generate small and negative values. For example, all (100%) individuals in the RIOPA study had risks exceeding 10^{-6} for MTBE, styrene, 1,4-DCB, TCE, PERC, and CTC, but Gumbel predictions ranged from 77% (TCE) to 99% (PERC), showing the tendency to underpredict.

As noted above, lognormal predictions did not match observations, and the differences could be large; for example, for the top 5% of PERC risks, 33% of the observations exceeded the 10^{-4} risk level, but the lognormal predictions showed percentages less than half of this level. Similar results were seen for benzene, styrene, TCE, and other VOCs.

Overall, these evaluations show that GEV distributions provided a good fit to pollutant and risk extrema for the VOCs and VOC mixtures measured in RIOPA. Occasionally, GEV distributions overpredicted some concentrations and risks, but this was limited to the very highest values. The three-parameter GEV distributions provided better fit than the two-parameter Gumbel distributions. Lognormal distributions provided poor fits to extrema.

Accurate estimates of extreme values of pollutant concentrations and exposures are needed to evaluate risks; the estimates need to include the number of people at risk or highly exposed, the risk to the most-exposed individual, and possibly the burden of disease. Although approaches vary among jurisdictions, toxic air pollutants (like many VOCs) are often regulated according to a risk-based approach; for example, the maximum acceptable excess lifetime cancer risk for any individual or location is limited to a value in the range of 10^{-4} to 10^{-6}. Higher risk levels may be acceptable if exposure is limited to few people, whereas the lower levels may be acceptable if exposure is widespread.

The extreme value distributions investigated here facilitate such analyses. For example, in Figure 1, for the top 5% of 1,4-DCB exposure concentrations, 2% of individuals in the sample (the 98th percentile level on the y axis) would experience a risk (on the x axis) exceeding 7.8×10^{-3} using the GEV model, and the Gumbel model produces a similar estimate (8.1×10^{-3}); but the lognormal model gives only 3.4×10^{-3}, or less than half of the 8.4×10^{-3} risk observed in the RIOPA data set. The GEV (and Gumbel) models allow easy and accurate estimates. As a second example, the number or fraction of individuals who would exceed a specific high-risk level can be better predicted using these models, and again the lognormal model underestimates the observed values; for example, the GEV, Gumbel, and lognormal models predict that 1.33%, 1.22%, and 0.52%, respectively, of the simulated data will exceed a risk of 10^{-2} due to 1,4-DCB exposure compared with 1.67% for the observed data. Overall, lognormal models greatly underestimated the highest risks (generally those exceeding the 98th percentile level), whereas GEV and Gumbel models accurately represented extrema.

Improving the estimates of exposures by fitting appropriate distributions will reduce uncertainties in health risk assessments and in risk-based policies. However, risk estimates will always involve other uncertainties, perhaps most notably those associated with dose–response models (e.g., their functional form and parameters). Although new methods are advancing the ability to provide a rigorous assessment of uncertainties associated with the dose–response models, all sources of uncertainty should be evaluated to understand the role they play in risk calculations.

Our analysis of extreme value distributions has several limitations. First, extreme value distributions describe only the upper tail of a distribution and they cannot be used for the remainder of the distribution. Second, in this study, we defined extrema using a threshold approach (e.g., 90th and 95th percentile concentrations), not the more common block maxima approach (e.g., highest concentration over a given time period or spatial domain). The use of still higher cut-offs (e.g., the 98th percentile) was not feasible due to sample sizes. We recognize that using a threshold to define extrema appears similar to using the peaks-over-threshold approach sometimes used for times series, which are often fitted using a Poisson approximation (Hüsler 2009). In the RIOPA study, because the VOC measurements were neither time series nor block maxima, other selection and fitting approaches could be used. Also, results are based on personal exposure measurements of 15 VOCs in three large cities in the United States using a convenience sample of the population. In particular, the RIOPA participants included only households with residents who did not smoke; locations near outdoor VOC sources were oversampled; and most participants were

unemployed women. Because the sample is not nationally representative, the results may not be broadly applicable to other cities.

Extreme Value Distributions for the NHANES Data

In most cases, the top 10% and top 5% of the NHANES data did not match GEV distributions fitted to either the larger data set (which used sample weights to specify repeat frequencies) or to the smaller (equally sized) data sets that used bootstrap methods and repeated sampling (Appendix Tables A.6 and A.7). Using the latter approach, for example, GEV distributions matched only the top 5% of 1,4-DCB and TCE (marginally significant) based on the A-D tests, but not the K-S tests (Appendix Table A.7). Possibly the two approaches used to incorporate the sampling weights did not decrease the "staircase" nature of the weighted data sets; if so, both approaches may have caused these tests to reject the hypothesis that the original and fitted distributions did not differ. Another possible explanation is that the repeated observations violated the assumption that extreme values are drawn from a set of independent and identically distributed (i.i.d.) samples (Fisher and Tippett 1928).

We tried a third approach — fitting GEV distributions to the unweighted NHANES data — and in this case, goodness-of-fit criteria were met on the basis of both the A-D and K-S tests (Appendix Table A.8). These results suggest that the fitting approaches or possibly the evaluation approaches used for the GEV distributions are not appropriate for weighted data sets.

METHODS: MIXTURE OF NORMAL DISTRIBUTION FITTING

Three VOCs (chloroform, 1,4-DCB, and styrene) were selected to evaluate mixture distributions. These VOCs differ in terms of their distributions, detection frequencies, and other properties. Personal VOC exposure observations for adults in the RIOPA study were selected primarily because the sample size for the adult cohort was largest (n = 544 for each VOC) and because personal samples should best reflect exposure. The original RIOPA investigators used two laboratories to analyze samples. Because the laboratories had different MDLs, all data below the MDL were replaced with a single value using 0.5 of the higher of the laboratories' MDLs. Because the VOC data in RIOPA had many extreme values, the density estimation methods were implemented using logarithms of the concentration value, as described next.

Finite Mixture of Normal Distributions

Finite mixture distributions are commonly used to identify and model subpopulations within an overall population. Rather than identifying the subpopulation that a single observation belongs to, these models assume that the data randomly arise from distributions with certain probabilities. Let $Y = (Y_1, \ldots, Y_n)$ be a random sample of size n from the overall population with the probability density function of Y_i given as $f(y_i)$. Y is assumed to have arisen from a mixture of an initially specified number of distributions. A K-component mixture of distributions supposes that the density of Y_i can be written as

$$f(y_i) = \sum_{k=1}^{K} \lambda_k f_k(y_i), \quad (4)$$

where f_k is the component density of the kth cluster, and λ_k is the corresponding weight with the constraint that $0 \leq \lambda_k \leq 1$ and $\sum_{k=1}^{K} \lambda_k = 1$. In many applications, component densities f_k are assumed to be standard parametric families, such as normal distribution $N(\mu_k, \sigma_k^2)$; then

$$f(y_i) = \sum_{k=1}^{K} \lambda_k N\left(\mu_k, \sigma_k^2\right). \quad (5)$$

The mixture distribution represented by equation 5 is a possible choice for handling concentration and exposure data that can have multiple modes and extreme values. Such mixture distributions are popular choices with attractive properties (Titterington et al. 1985): Since the mixture distributions are constructed as linear combinations of normal distributions, they are computationally and analytically tractable, well behaved in the limiting case, and scalable to higher dimensions.

Mixture distributions can be fitted using many techniques; e.g., graphical methods, the method of moments, MLE, and Bayesian approaches (McLachlan and Peel 2000; Redner and Walker 1984; Titterington et al. 1985). Since closed forms of MLEs, as in equation 4, are not available, mixture distributions are commonly fitted using an expectation maximization (EM) type of algorithm (Dempster et al. 1977; McLachlan and Krishnan 1997; Meng and Pedlow 1992). We used the EM algorithm and a constrained maximum likelihood method to estimate equation 5 with a further constraint that the location of the first cluster (generally the lowest) is below the MDL; i.e., $\mu_1 \leq$ MDL. This constraint ensures that a fitted cluster covers the MDL, which allows it to be interpreted as the subpopulation of the data below the MDL.

An important issue in fitting finite mixture distributions is selecting the number of components, K. Criteria based on penalized likelihood, such as the Akaike information criterion (AIC), have been applied successfully to mixture distributions (McLachlan and Peel 2000). Although this criterion generally favors larger K, considerable practical support for its use is due to its simplicity (Fraley and Raftery 1998). The Bayesian information criterion (BIC) also appears attractive due to its statistical properties and the simplicity of its implementation. Though the BIC always leads to a smaller (or equal) number of components than AIC, the BIC can also lead to an overestimate of the number of clusters regardless of the clusters' separations (Biernacki et al. 2000). In general, with a limited amount of data, a corrected version of AIC (AICc) (Hurvich and Tsai 1989) may be preferable. For these finite mixture distributions, we fitted equation 5 with K = 2 to 5 clusters, and selected the optimal model based on AICc. This analysis was conducted on RIOPA data for each of the three VOCs (chloroform, 1,4-DCB, and styrene).

As a benchmark for comparison, we also fitted the traditional normal distribution, which is a special case of a mixture distribution that uses K = 1. (As noted earlier, log-transformed VOC data were used in all cases.)

The finite mixture distribution model was implemented using the mixtools package (Benaglia et al. 2009) in R (R Foundation for Statistical Computing, Vienna, Austria). This package fits the finite mixture distributions using EM algorithms through the function normalmixEM.

Dirichlet Process Mixture of Normal Distributions

Bayesian density estimation methods using DPM of normal densities have several practical advantages, including optimally trading off local versus global smoothing, assessing modality, and propagating uncertainty on inferences regarding the number of components and thus propagating uncertainty about the density estimate (Escobar 1994; Ferguson 1983; Mueller and Quintana 2004). Instead of prespecifying the number of clusters, these models allow the number of clusters to be chosen in a data-adaptive way. Let $Y_i \sim N(\mu_i, \sigma_i^2)$ and let $(\mu_i, \sigma_i^2) = \theta_i$. The DPM distribution assumes that these normal parameters θ_i follow a random distribution G generated from the Dirichlet process (Ferguson 1973); this can be represented as:

$$\theta_i \mid G \sim G \text{ i.i.d. and } G \mid \alpha, G_0 \sim DP(\alpha\, G_0). \quad (6)$$

$DP(\alpha\, G_0)$ is a Dirichlet process with concentration parameter α and base distribution G_0, which is also known as the prior expectation of G. The precision parameter α determines the concentration of the prior for G around G_0. Blackwell and Macqueen (1973) provided the following representation for the leave-one-out conditional distributions:

$$\theta_i \left| \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n, \sim \frac{\alpha}{n-1+\alpha} G_0 + \frac{1}{n-1+\alpha} \sum_{j \neq i}^{n} I_{\theta_j}(\cdot) \right. . \quad (7)$$

In this approach, $\theta = (\theta_1, \ldots, \theta_n)$ will be reduced to certain K distinct values ($K < n$) with positive probability. From equation 7, two well-known extreme cases of the DPM can be derived. As $\alpha \longrightarrow \infty$, the DPM reduces to a parametric model, namely $\theta_i \sim G_0$ i.i.d. clusters (n), whereas $\alpha \longrightarrow \infty$ implies a common parametric model, namely $\theta_1 = \ldots = \theta_n = \theta^*$ with $\theta^* \sim G_0$ (1 cluster). The baseline distribution G_0 is chosen to be the conjugate normal-inverse-gamma distribution. Hyperpriors could be used on this normal-inverse-gamma distribution to complete the model specification.

The DPM model does not require specification of the number of clusters as is needed for parametric mixture distributions, such as the finite mixture model discussed previously. In practice, suitable values of K will typically be small relative to the sample size n. The implicit prior distribution on K stochastically increases with α and is related to the prior distribution on α (Antoniak 1974). For moderately large n, $E(K \mid \alpha, n) \approx \alpha \log(1 + n/\alpha)$ (Antoniak 1974). A formal assessment of uncertainty regarding the number of components K can be obtained through generated draws from the posterior distribution of K as a part of the Bayesian computation scheme.

For the VOC data, the precision parameter α was chosen to follow a gamma prior distribution, and a sensitivity analysis was conducted with respect to the choice of the gamma parameters. Given the sample size in the test data set (n = 544), for prior information, $\alpha \sim$ Gamma(0.3, 0.4) favors K = 1 to 3 clusters; $\alpha \sim$ Gamma(1.2, 2.5) favors 1 to 5 clusters; $\alpha \sim$ Gamma(2, 1.5) favors 2 to 10 clusters; and $\alpha \sim$ Gamma(5, 2) favors 5 to 20 clusters. A sensitivity analysis was conducted on these prior specifications.

Computational methods were followed that allowed us to evaluate posterior distributions for all model parameters and the number of components, as well as the resulting predictive distributions (Escobar and West 1995). Density estimation using DPM was implemented using the DP package (Jara 2007; Jara et al. 2011) in R (R Foundation for Statistical Computing, Vienna, Austria), which provides posterior draws of all model parameters under a DPM using Markov chain Monte Carlo methods.

Goodness-of-Fit Criteria

Goodness of fit for the density estimation methods was determined by comparing the estimated cumulative distribution function (CDF) $\hat{F}_{\text{est}}$ based on the simulated data with the empirical CDF $\hat{F}_{\text{emp}}$ based on the observed data. Although all of the observed and generated data were used to estimate the CDF by each method, goodness of fit was evaluated using only the data above the MDL in all data sets. Both the mean squared error

$$\text{MSE} = \sum_{i,y_i>MDL}\left[\hat{F}_{\text{emp}}(y_i) - \hat{F}_{\text{est}}(y_i)\right]^2 / \sum_i I(y_i > MDL)$$

and the mean absolute error

$$\text{MAE} = \sum_{i,y_i>MDL}\left|\hat{F}_{\text{emp}}(y_i) - \hat{F}_{\text{est}}(y_i)\right| / \sum_i I(y_i > MDL)$$

were used. The estimated proportion of observations above the MDL (which is often termed the detection frequency) for empirical and estimated distributions was compared.

Simulation Study

For further evaluation of the mixture distributions, several forms of underlying true distributions and varying amounts of left-censored data (values below the MDL) were used as true generation models. Three methods were compared: a single normal distribution; a finite mixture distribution; and a DPM distribution. Two underlying distributions with features similar to the three VOC samples from the RIOPA study were selected: a normal(0.2^2) and a mixture specified as ½ Gamma(3, 1.5) + ½ Uniform(−3, 8). The former is symmetric and the latter is skewed toward upper extreme values, and both had multiple modes when data below the MDL were replaced by 0.5 MDL. The proportion of data below the MDL, P_0, was set to 15%, 30%, and 50% in separate simulations. Goodness-of-fit measures (MSE and MAE as described above) were calculated for each method, target distribution, and choice of P_0. We generated a data set of $n = 1000$ for each simulation under each P_0 setting. The average values of MSE and MAE across 500 simulations are reported.

For the finite mixture distributions, the number of components K was based on the smallest AICc. A convergence problem was encountered when P_0 was high (in the range of 30% to 50%); this may have been because the censored data were set to a single value (0.5 MDL), which resulted in a very small variance of the first (lowest) cluster. In addition, the MLE method for finite mixture models is susceptible to other problems (e.g., nonunique solutions) (McLachlan and Peel 2000; Redner and Walker 1984; Titterington et al. 1985). Thus, data below the MDL was replaced by uniformly generated simulated data from U(0, MDL) if the finite mixture distributions did not converge. In contrast, all of the single normal and DPM method simulations converged.

RESULTS AND DISCUSSION: MIXTURE OF NORMAL DISTRIBUTIONS FOR VOCs

Single Normal Distributions

For chloroform, which was roughly lognormally distributed except that 17% of the observed data were below the MDL, the single normal distribution model fit about as well as the finite mixture (FM) distribution and the DPM distribution (described below) on the basis of MSE and MAE values, and gave a 21% probability of being below the MDL, similar to that for the observed data (Table 5). However, for 1,4-DCB

Table 5. Goodness-of-Fit Statistics for Each Density Estimation Method for Chloroform, 1,4-DCB, and Styrene Data from the RIOPA Study[a]

	Proportion Below MDL				MSE[b]			MAE[b]		
		Predicted								
VOC	Observed	Single Normal	FM	DPM	Single Normal	FM	DPM	Single Normal	FM	DPM
Chloroform	0.17	0.21	0.23	0.23	0.07	0.07	0.08	7.18	6.89	6.95
1,4-DCB	0.34	0.28	0.33	0.33	31.81	0.08	0.04	167.05	7.00	5.30
Styrene	0.66	0.56	0.64	0.64	32.61	0.07	0.04	160.47	6.10	4.27

[a] $n = 544$ for personal adult observations in RIOPA.

[b] MSE and MAE were multiplied by a scalar of 1000 to reflect the significant figure.

and styrene, for which more observed data were below the MDL and the distributions had extreme values, the fit of the single normal distribution model was inferior compared with those of the FM and DPM models. For example, the predicted probability of being below the MDL is 28% and 56% for 1,4-DCB and styrene, respectively, compared with 34% and 66% of observed data, and 33% and 64% estimated by the FM and DPM models. The single normal distribution model overestimated the mean of these VOCs since it underestimated the frequency of data below the MDL.

Finite Mixture of Normal Distributions

Fitted density plots (and component clusters) are shown in the B panels of Figures 2, 3, and 4 for chloroform, 1,4-DCB, and styrene, respectively. The fitted parameters (weight λ_k, location μ_k [mean], and dispersion σ_k^2 [SD]) of each cluster

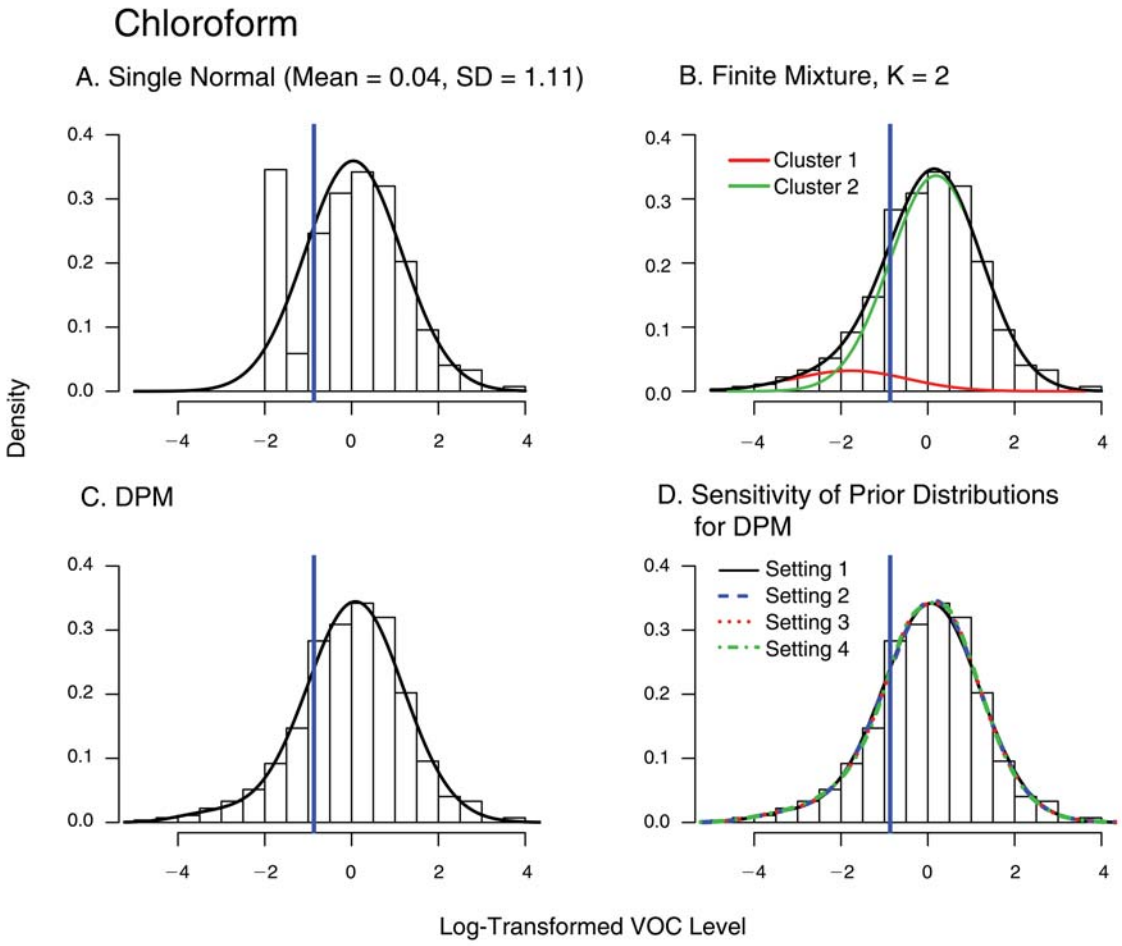

Figure 2. Fitted density plots for chloroform using single normal, FM, and DPM distributions (log scale; *n* = 544 observations for adult participants in RIOPA). The wide bars show RIOPA observations; the blue vertical rule is the MDL; the black curve in the B panel is the fitting distribution (combining all clusters); in the D panel, the four prior distributions of the DPM are shown as settings 1–4 and all of the curves overlap.

K for the FM distribution model are given in Table 6. The optimal Ks (based on the AICc) were 2 for chloroform, 4 for 1,4-DCB, and 3 for styrene. These choices of K clearly reflected the multi-modality and upper-skewness of the VOC data, and the resulting FM distributions closely fitted the observed distributions (Figures 2, 3, and 4). For example, Figure 3B shows the four clusters that fitted the 1,4-DCB data: the first cluster (red) captured the left censoring due to the MDL, the second cluster (green) reflected the majority of the data, the third cluster (blue) reflected the skewness, and the fourth cluster (orange) modeled the extreme values.

Nonparametric DPM of Normal Distributions

Fitted densities using DPM distributions for the three VOCs are shown in the C panels of Figures 2, 3, and 4. This method clearly captured the censoring, upper-skewness, and potential multi-modality of the exposure data. In

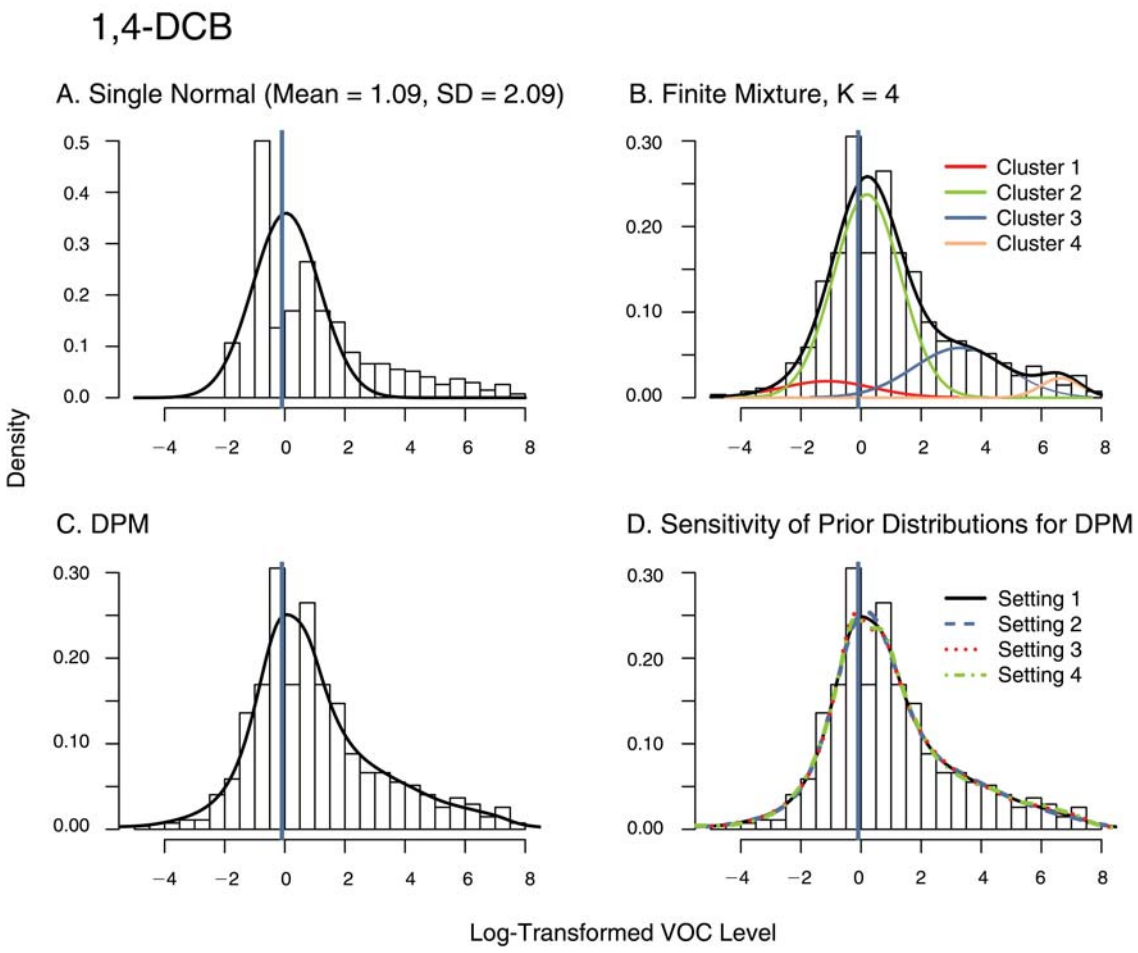

Figure 3. Fitted density plots for 1,4-DCB using single normal, FM, and DPM distributions (log scale; *n* = 544 observations for adult participants in RIOPA). The wide bars show RIOPA observations; the blue vertical rule is the MDL; the black curve in the B panel is the fitting distribution (combining all clusters); in the D panel, the four prior distributions of the DPM are shown as settings 1–4 and all of the curves overlap.

terms of MSE and MAE, the DPM approach attained slightly lower values than the FM distributions except for chloroform (Table 5).

The D panels of Figures 2, 3, and 4 show results of the sensitivity analyses with the four different gamma distributions used as prior distributions for precision parameter α. As noted before, K stochastically increases with α as E(K | α, n) ≈ α log(1 + n/α) for moderately large n (Antoniak 1974). The four prior distributions (shown as settings 1 through 4 in the figures) were informative and formed up to 20 clusters that reflected more specific subject matter information. Estimated densities obtained using the four prior distributions nearly overlapped and showed very similar MSEs and MAEs for each of the VOCs, although the corresponding posterior distribution of the number of clusters varied (Appendix Table A.9). The posterior mean

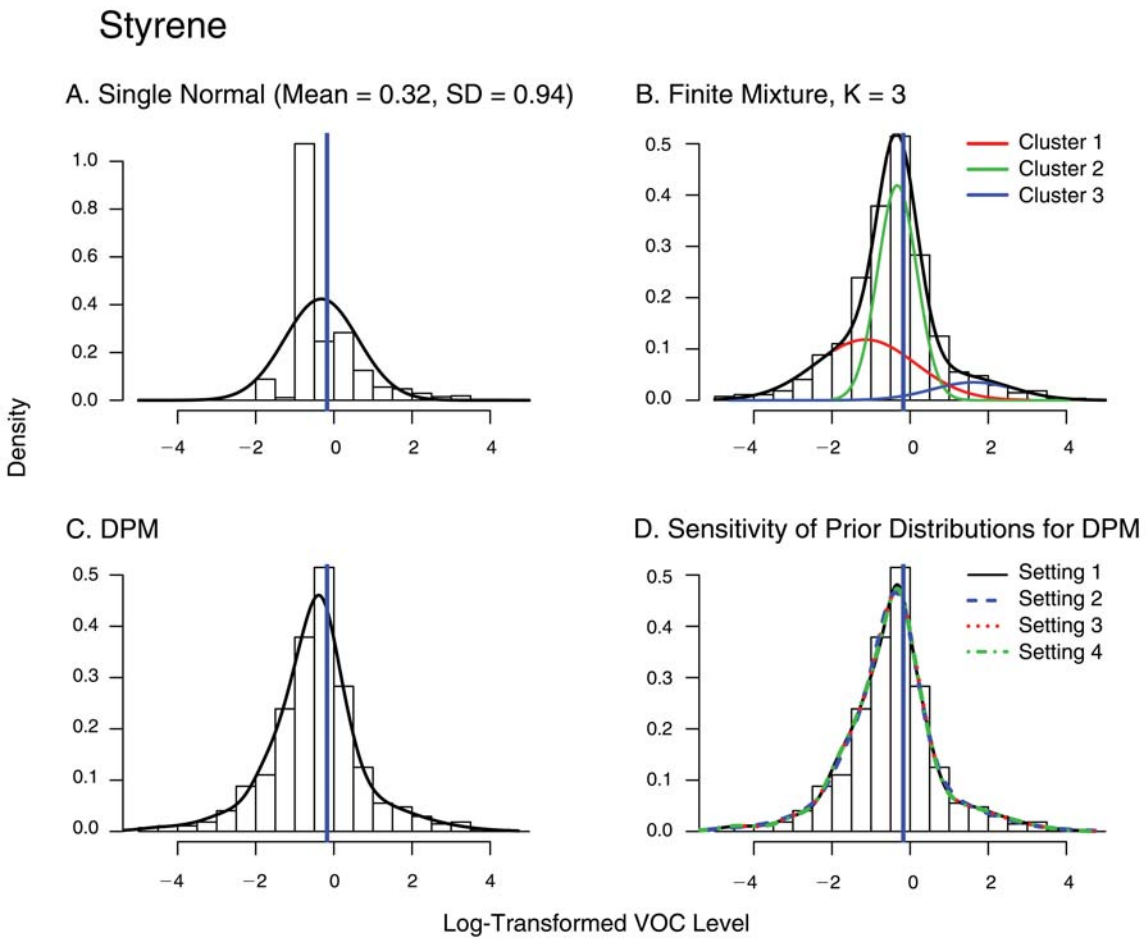

Figure 4. Fitted density plots for styrene using single normal, FM, and DPM distributions (log scale; n = 544 observations for adult participants in RIOPA). The wide bars show RIOPA observations; the blue vertical rule is the MDL; the black curve in the B panel is the fitting distribution (combining all clusters); in the D panel, the four prior distributions of the DPM are shown as settings 1–4 and all of the curves overlap.

Table 6. Fitted Weight, Location, and Dispersion Parameters Under the Finite Mixture Distributions for Chloroform, 1,4-DCB, and Styrene Data from the RIOPA Study[a]

	Chloroform			1,4-DCB			Styrene		
Number of Components	Weight	Location (Mean)	Dispersion (SD)	Weight	Location (Mean)	Dispersion (SD)	Weight	Location (Mean)	Dispersion (SD)
K = 2		**AICc = 1774**			AICc = 2403			AICc = 1735	
Cluster 1	**0.11**	**−1.78**	**1.31**	0.16	−1.05	0.96	0.40	−1.12	1.86
Cluster 2	**0.89**	**0.19**	**1.06**	0.84	1.35	2.23	0.60	−0.40	0.62
K = 3		AICc = 1778			AICc = 2330			**AICc = 1716**	
Cluster 1	0.12	−1.78	1.23	0.12	−1.05	1.58	**0.41**	**−1.12**	**1.31**
Cluster 2	0.60	0.08	0.90	0.63	0.31	1.14	**0.51**	**−0.35**	**0.54**
Cluster 3	0.28	0.55	1.20	0.25	3.84	1.93	**0.08**	**1.82**	**1.01**
K = 4		AICc = 1781			**AICc = 2328**			AICc = 1714	
Cluster 1	0.11	−1.78	1.27	**0.14**	**−1.05**	**1.54**	0.39	−1.12	1.33
Cluster 2	0.07	−0.52	0.25	**0.60**	**0.27**	**1.08**	0.49	−0.37	0.60
Cluster 3	0.05	0.61	0.15	**0.23**	**3.29**	**1.55**	0.04	−0.29	0.08
Cluster 4	0.78	0.24	1.09	**0.04**	**6.64**	**0.67**	0.07	1.90	0.97
K = 5		AICc = 1785			AICc = 2329			AICc = 1722	
Cluster 1	0.11	−1.78	1.26	0.14	−1.05	1.52	0.33	−1.12	1.32
Cluster 2	0.17	−0.39	0.43	0.05	−0.24	0.16	0.05	−1.51	1.28
Cluster 3	0.10	0.60	0.21	0.62	0.48	1.21	0.04	−0.29	0.08
Cluster 4	0.58	0.22	1.21	0.04	6.66	0.66	0.51	−0.37	0.60
Cluster 5	0.04	1.31	0.12	0.16	3.86	1.27	0.08	1.86	0.99

[a] n = 554 observations. The smallest AICc for each VOC is shown in **bold** type; for styrene, K = 3 was chosen over K = 4 due to the small difference in values.

of K under all prior settings of α slightly exceeded the K selected using the AICc (Table 6). The higher K in the DPM model is due to the prior information of α and does not introduce any additional complexity or more model parameters. Given K distinct values among the elements of θ, a larger variance leads to increased dispersion among the K group means, which increases the likelihood of multiple modes and decreases smoothness in the resulting predictive distribution (Escobar and West 1995).

No convergence issues using the DPM method were encountered, and density estimation results were robust given the moderate sample size (n = 544). Another advantage of the DPM method is that a constraint to ensure a cluster below MDL is not required since the sampling scheme given in equation 7 is data driven. As shown in equation 7, the DPM model can handle values below the MDL that are represented as a point mass because a newly sampled value has equal probability $1/(n - 1 + \alpha)$ to be drawn from the observed set of values.

The nonparametric DPM of normal distributions assumes that observed data randomly arise from subdistributions with certain probabilities, as the finite mixture distribution also models. (Again, the subpopulations that individual observations belong to are not identified.) Compared with the finite mixture models, DPM distributions have advantages in providing a formal assessment of uncertainty for all model parameters, including the number of components K, through generated draws from the posterior distribution. With a suitable Dirichlet process prior structure (Escobar and West 1995), these models produce predictive distributions qualitatively similar to kernel techniques, and they allow for differing degrees of smoothing by the choice of priors for precision parameter α. The density estimation results were robust given a moderate sample size (n = 544) without any convergence issues noted.

Simulations

Simulation results, summarized in Table 7, show similar patterns for the MSE and MAE criteria. Both the FM and DPM distributions provided much better fits (reflected as values closer to 0) than a single normal distribution, except that they were only slightly better under distribution 1 with $P_0 = 0.15$. For the single normal distribution, as

Table 7. Goodness-of-Fit Statistics for Each Density Estimation Method for Analyses of Simulated Data Sets[a]

	MSE[c]			MAE[c]		
Proportion of Data Below MDL (P_0)[b]	Single Normal	FM	DPM	Single Normal	FM	DPM
Distribution 1						
0.15	0.09	0.03	0.08	7.65	4.64	7.11
0.30	0.19	0.04	0.08	11.19	4.80	7.29
0.50	0.43	0.05	0.05	16.77	5.26	5.69
Distribution 2						
0.15	1.55	0.10	0.02	32.58	8.19	3.57
0.30	2.53	0.10	0.02	43.69	8.59	3.29
0.50	2.62	0.12	0.02	46.52	8.22	3.28

[a] In the simulated data sets, $n = 1000$ for each P_0.

[b] Distribution 1: Normal(0,2^2); Distribution 2: ½ Gamma(3,1.5) + ½ Uniform(−3,8). Prior distribution on α is Gamma(1.2,2.5).

[c] MSE and MAE are multiplied by a scalar of 1000 to reflect the significant figure.

the fraction P_0 of data below the MDL increased, the lack of fit tended to increase as well (reflected in the values getting higher), whereas the FM and DPM distributions fitted considerably better and without such trend (reflected in the values staying about the same). The DPM distribution showed an advantage of robustness regarding P_0: It fitted equally well, or even better, as P_0 increased. For distribution 1, the FM distribution provided a slightly better fit than the DPM distribution, but this trend can be offset since the prior variance α can be decreased to promote smoothness. In this regard, the DPM model is much more flexible than the FM model. Here, we have used $\alpha \sim$ Gamma(1.2, 2.5), which favors 1 to 5 clusters given our sample size (as the prior information of K). For distribution 2, which is upper-skewed and has extreme values, the DPM model provided a much better fit than the FM model under all settings.

Both types of mixture models are well suited to the RIOPA VOC data containing a large fraction of censored data due to MDLs, extreme values, and multiple modes. They offer clear advantages over parametric full-distribution models and extreme value models, and also appear appropriate for many other types of environmental data, such as concentrations or doses of persistent and emerging compounds and biomarkers. The use of mixture models has the potential to improve the accuracy and realism of models used in a variety of exposure and risk applications, and further environmental applications are warranted.

SPECIFIC AIM 2. DISTRIBUTION MODELS FOR VOC MIXTURES

METHODS: IDENTIFICATION OF MIXTURES

Positive Matrix Factorization

VOC mixtures in the RIOPA data set were selected for our analyses using two approaches: PMF and toxicologic mode of action. The first approach identified common VOC mixtures using PMF, a multivariate analysis that is similar to factor analysis, but with the ability to incorporate uncertainties on each measurement that may potentially reflect sampling errors and MDLs (Anderson et al. 2001; Paatero and Tapper 1994). Based on the uncertainty, variables are modeled as weak or strong; that is, variables with high uncertainties are assigned weak influence and variables with low uncertainties are assigned strong influence. Each VOC was given an uncertainty equal to the measurement precision estimated as the pooled coefficient of variation for duplicate samples (Weisel et al. 2005c). Styrene and total VOCs were designated as weak given their higher uncertainty levels. Measurements below MDLs were retained, but assigned high uncertainties to reduce their influence (U.S. EPA 2008a).

PMF decomposes two matrices from the sample data: (1) a matrix of factor profiles that shows the percentage and mass of each VOC apportioned to the factor; and (2) a matrix of

factor-relative contributions that shows the contributions of each factor to the total concentration of each observation (U.S. EPA 2008a). Because there is no optimal or a priori manner for selecting the number of factors, multiple PMF analyses were conducted using 3, 4, and 5 factors. Each was tested using goodness-of-fit indicators; specifically, scaled residuals were tested using Q values (the sum of squares of the residuals divided by the uncertainties for the concentrations of individual compounds) (Anderson et al. 2001; U.S. EPA 2008a).

To avoid potential biases involved in repeated measurements in further analyses (e.g., cluster effects in copula analyses), PMF analysis was applied to only the personal adult VOC measurements collected at the first visit (n = 299). Note that, unlike the mixtures based on modes of action described next, the PMF-based mixtures were purely data driven due to correlations among components in the dataset.

The PMF analyses used PMF 3.0, a peer-reviewed receptor modeling tool developed by the Environmental Protection Agency's Office of Research and Development (U.S. EPA 2008a).

Toxicologic Mode of Action

The second approach for selecting exposure mixtures evaluated the toxicologic modes of action, which reflect the biochemical pathways and outcomes that may be affected by pollutant exposure (Borgert et al. 2004). Both the acute and the chronic toxicity of mixtures can be affected by factors that alter the absorption, distribution, metabolism, and excretion of a mixture's components and metabolites; for example, chemical-to-chemical reactions; alterations in absorption, protein, DNA binding, and excretion; and inhibition or activation of biotransformation pathways (De Rosa et al. 2001; U.S. EPA 1990). Moreover, exposure to mixtures can result in interactions between mixture components, defined as any circumstance in which two or more components in a mixture result in a different biological response compared with that predicted from the actions of the single chemicals (De Rosa et al. 2001). A biotransformation pathway analysis that evaluates mechanisms and interactions for all of the VOCs measured in RIOPA is not available in the literature and was beyond the scope of this study.

We identified the toxicologic modes of action for target organs and cancer endpoints using a simplified or screening-level assessment; for example, benzene and MTBE are both associated with leukemia and lymphoma, so they were grouped together. Because VOCs can be associated with different cancers, one VOC may be grouped into several mixtures, each representing a different type of cancer. Information regarding cancer endpoints was obtained from IARC monographs (IARC 2013) and the California OEHHA reports (OEHHA 2005), which are summarized in Appendix Table A.2.

The analysis used URFs based on data from human or animal studies for the pertinent cancer endpoints; for example, MTBE's URF was based on animal studies of kidney carcinomas, leukemia, and lymphomas (OEHHA 2005). In the literature, the URFs reflect upper-bound (95th percentile) confidence estimates; thus, estimates of excess lifetime cancer risk for the mixtures (as well as for individual VOCs) may be overestimated. Normally, estimates of URFs seen in the literature were developed for only the most sensitive cancer endpoint; therefore, cumulative toxicity for a single VOC may be overestimated if it is grouped into more than one mixture.

We selected two mixtures that can result in common cancer endpoints: (1) VOCs associated with hematopoietic cancers (lymphomas and leukemia), which include benzene, MTBE, 1,4-DCB, TCE, and PERC (designated mixture B2); and (2) VOCs associated with liver and renal tumors, which include ethylbenzene, MTBE, 1,4-DCB, TCE, PERC, chloroform, and CTC (designated mixtures B1 and B3 described below; Borgert et al. 2004; IARC 2012). The two mode-of-action mixtures contained 5 and 7 components, respectively.

To reduce the number and complexity of analyses of mixtures that contained a large number of components, highly correlated VOCs were grouped together based on their likely emission sources or chemical characteristics. For example, the seven VOCs in the mixture associated with liver and renal tumors were separated into a group of gasoline-related compounds (ethylbenzene and MTBE; designated mixture B1), and a group of chlorinated hydrocarbons (1,4-DCB, TCE, PERC, chloroform, and CTC; designated mixture B3 and includes the same VOCs as in mixture A3; see below).

METHODS: DEPENDENCY STRUCTURES OF MIXTURES

Copula Analysis

Dependency structures of the identified mixtures (using personal adult measurements from the first RIOPA visits [n = 299]) were fitted to five copulas using MLEs (Gaussian, t, Gumbel, Clayton, and Frank). Goodness-of-fit tests were conducted using Akaike and Bayesian information criteria, and the copula with the lowest criterion was chosen as the best-fit dependency structure. Copulas transform the marginal distributions of each variable into a uniform distribution over the interval [0,1]. After this transformation, the

dependency structure is described following reference distributions. Once the dependency structure and marginal distributions are known (or estimated), the joint distribution function is:

$$C(u_1, u_2, \ldots, u_p) = \text{Prob}(U_1 \le u_1, U_2 \le u_2, \ldots, U_p \le u_p), \quad (8)$$

where C is a copula function, U is a uniformly transformed random variable for marginal distribution function $F_i(x_i)$, and p is the number of variables. The joint distribution function can also be expressed as:

$$C\left[F_1(x_1), F_2(x_2), \ldots, F_p(x_p)\right] = F(x_1, x_2, \ldots, x_p). \quad (9)$$

According to the Sklar theorem (1959), if F_i is continuous and x_i is over $[-\infty,\infty]$, then C is unique.

Copulas allow dependency structures to be weighted in different manners and thus can be symmetric or asymmetric (Staudt 2010). The several families and many types of copulas have different origins and properties. The family of elliptical copulas is derived from distributions; for example, the Gaussian copula is from the multivariate normal distribution and the t copula from the multivariate Student t distribution. Given the same correlation coefficient, t copulas provide a better fit to distributions that include extreme values than do the Gaussian copulas; that is, the t copula more accurately models tail dependencies (Schmidt 2006). Among Archimedean copulas, which are stated directly and not derived from distributions, Gumbel copulas emphasize upper tail dependency, Clayton copulas emphasize lower tail dependency, and Frank copulas have no emphasis on tail dependency (i.e., they have symmetrical dependencies on both tails) (Schmidt 2006). The product copula — the simplest copula — shows independence among random variables (Trivedi and Zimmer 2007).

After choosing the best-fit copulas, we generated two sets of objects necessary for simulating joint distributions (discussed in the next section); namely, uniform [0,1] random variables for each component of the mixture that followed the copula-identifying correlations and the copula parameters that were estimated using MLE. The Gaussian copula parameter was the covariance matrix. The t copula used the same matrix plus degrees of freedom. The Gumbel, Clayton, and Frank copulas each used a correlation parameter.

Simulated Joint Distributions

Simulations tested the goodness of fit for the fitted copulas using the uniform random variables and fitted parameters for each copula (described above) as well as marginal distributions fitted using MLE for each VOC (see Appendix A). A large number (n = 1,000) of simulated data were generated for each mixture. Using the simulated data sets, the probabilities that all components in the mixture would exceed 50th, 75th, 90th, and 95th percentile cutoffs were calculated and compared with observations from the RIOPA study. For comparison, we also calculated probability assuming independence among mixture components; for example, the probability of a three-component mixture in which each component exceeded the 90th percentile concentration was 0.001. Because styrene and TCE had low detection frequencies (49% and 31%, respectively), probabilities that all mixture components would exceed the 50th percentile could not be calculated.

To examine the influence of each component in a mixture and any trends that might be associated with an increasing concentration of a component, we calculated mixture fractions — the fraction that each component contributes to a mixture — for both observed and simulated data; we summarized the results using the median fractions in several segments (50th–75th, 75th–90th, 90th–95th, 95th–100th percentiles) for each mixture. Changes in the mixture fractions associated with the total mixture concentration show trends and help reveal the mixture's source (e.g., fractions for generated or intentional mixtures should be constant). Mixtures with consistent mixture fractions across a population or over time are considered to be homogeneous and may represent generated mixtures. In contrast, highly variable or heterogeneous mixture fractions may reflect coincidental mixtures.

For VOC mixtures based on mode of action, cumulative cancer risks were estimated assuming response addition following EPA guidance (U.S. EPA 2000). We also computed the fraction of individuals with cumulative risks that would exceed thresholds of 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, and 10^{-2} and compared them with results obtained from the RIOPA observations and from copula and multivariate lognormal analyses of simulated data. Cumulative probability plots were used to visualize differences between observations and simulations.

Copula fitting and simulations were performed using ModelRisk 5 Industrial edition (Vose Software BVBA, Gent, Belgium). Simulations of multivariate lognormal distributions used RLNORM.RPLUS in R version 2.13.1 (R Development Core Team, Vienna, Austria) and Excel (Microsoft, Redmond, WA).

RESULTS AND DISCUSSION: SELECTED VOC MIXTURES BASED ON PMF

PMF provides a concentration-based approach that can identify generated mixtures and those that arise from a

common source or from correlated sources. However, VOC levels also may reflect common transport and dispersion patterns of contaminants (e.g., due to meteorologic factors) building characteristics (e.g., building AERs), as well as common behavioral patterns (e.g., a tendency to use or tolerate certain types of cleaning products); thus mixtures identified by PMF (or other correlation-based methods) may not be uniquely generated mixtures, but rather a combination of generated and possibly coincidental mixtures. Mixtures identified by PMF should be uncorrelated.

Based on the PMF analysis, four VOC mixtures were identified and designated as mixtures A1–A4 (see Table 8 and Figure 5 for analyses using personal exposure measurements). In the discussion of mixtures below, we refer to a few "marker" VOCs whose concentrations in a mixture were the means to identify a possible source category. (See Appendix Table A.10 and Figures A.10–A.13 for additional PMF results using indoor and outdoor exposure measurements.)

Mixture A1 was identified as gasoline vapor due to its content of benzene (average contribution = 1.4 μg/m^3; Figure 5) and MTBE (11.2 μg/m^3), two highly volatile VOCs. The RIOPA samples, collected from 1999 to 2001, reflect the gasoline composition from a time when benzene levels were higher (benzene content is now, in 2014, limited to 0.62% of a fuel) (U.S. EPA 2007a). At that time MTBE was used in California, New Jersey, and Texas (homes of the three cities studied in RIOPA; U.S. EPA 2008b) and has since been phased out (starting in 2000, fully in 2006 [U.S. EPA 2012b]).

Mixture A2 was designated as vehicle exhaust due to contributions from toluene (4.9 μg/m^3), ethylbenzene (1.9 μg/m^3), *m- & p*-xylenes (5.5 μg/m^3), *o*-xylene (1.7 μg/m^3), and styrene (0.2 μg/m^3). These VOCs are also highly volatile components of gasoline and diesel fuels and exposure occurs directly from the fuels as well as from exhaust emissions (ATSDR 2007, 2010a,b).

Mixture A3 contained several common indoor contaminants, including a moth repellent (1,4-DCB at 0.9 μg/m^3), chlorinated solvents (TCE at 0.2 μg/m^3, PERC at 1.7 μg/m^3, CTC at 0.5 μg/m^3), and a water disinfection by-product (chloroform at 0.8 μg/m^3). These VOCs are fairly specific to these sources; for example, 1,4-DCB is the major ingredient of mothballs (ATSDR 2006), although similar repellents often use naphthalene; PERC is a widely used dry cleaning solvent (ATSDR 1997c); chloroform is a by-product of water disinfection using chlorine dioxide (ATSDR 1997a); and TCE and CTC are industrial degreasers, chemical intermediates, and pesticides (ATSDR 1997b, 2005b).

Mixture A4 contained *d*-limonene (20.5 μg/m^3), α-pinene (1.5 μg/m^3), and β-pinene (2.7 μg/m^3), which are fragrances and solvents indicative of cleaning products and odorants. Both *d*-limonene and the pinenes are widely used fragrance and flavor additives in cleaning products, fresheners, foods and beverages, and other consumer products (IARC 1993; U.S. EPA 2012c).

These four mixtures explained 20.5%, 20.9%, 16.3%, and 42.3%, respectively, of the variation in the total VOC levels in the RIOPA data set (Table 8). PMF is often used for source apportionment, usually for ambient particulate matter, and the factors and apportionments are one of the final results of these approaches. Similar source profiles (gasoline vapor, vehicle exhaust, deodorizer and shower, and dry cleaning) were identified in a study using PMF and the NHANES data set, although NHANES did not measure *d*-limonene, α-pinene, and β-pinene, and in that study the dominant mixtures were gasoline vapor and vehicle exhaust (Jia et al. 2010).

Table 8. Source Categories and Apportionments of VOCs in Mixtures Derived Using PMF and the First-Visit Personal Observations from RIOPA[a]

Mixture	Suggested Source Category	Marker VOCs	Concentration of Marker VOCs (μg/m^3)	Total VOC Concentration (μg/m^3)	Marker VOCs as Fraction of Total VOCs (%)
A1	Gasoline vapor	Benzene and MTBE	12.6	19.9	20.5
A2	Vehicle exhaust	Toluene, ethylbenzene, *m- & p*-xylenes, and styrene	14.2	20.3	20.9
A3	Moth repellents, chlorinated solvents, and disinfection by-products	1,4-DCB, TCE, PERC, chloroform, and CTC	4.1	15.9	16.3
A4	Cleaning products and odorants	*d*-Limonene, α-pinene, and β-pinene	24.7	41.1	42.3

[a] n = 299.

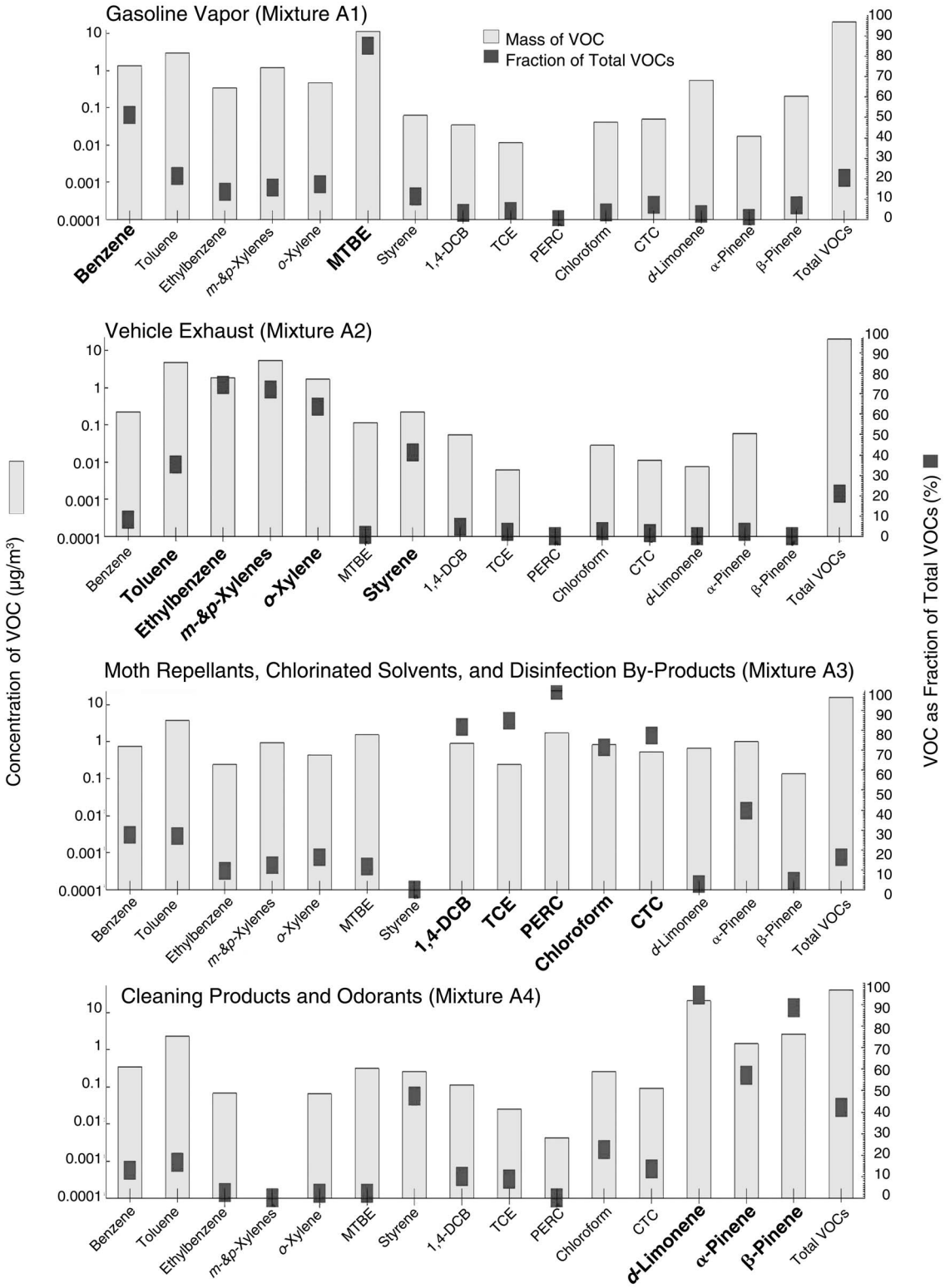

Figure 5. Profiles for mixtures A1–A4 from PMF analyses based on personal exposure measurements of VOCs (n = 299 measurements on the first visit for adult participants in RIOPA). Marker VOCs for each mixture are shown in bold type. Gray bars show the concentration (μg/m^3) of each VOC; black boxes show the percentage of VOC as fraction of total VOCs.

Mixture A4, cleaning products and odorants, explained the largest portion (42.3%) of total VOCs. This large fraction is a result of the relatively small number of VOCs the RIOPA investigators decided to measure, the large fraction of time that most people spent indoors, the common use of these particular VOCs in cleaning products and odorants, and their high concentrations (compared with other VOCs measured in RIOPA). Because many of the RIOPA participants were older (average age 45 years; 24% were 60 years or older) and predominantly female (75%), we suspected that the indoor residential fraction of total exposure would be especially important. Indoor time fractions calculated for the RIOPA participants (which included indoors at home, school, work, and other locations) indicated that they spent an average of 91% of their time indoors. (The indoor time fraction varied by city; e.g., 89%, 92%, and 92% for participants in Los Angeles, Elizabeth and Houston, respectively, $P < 0.0001$.) In summary, the source strength of the A4 mixture and the large amount of time spent indoors explains the dominance of this mixture in terms of its large share of total VOCs.

RESULTS AND DISCUSSION: DEPENDENCY STRUCTURES AND JOINT DISTRIBUTIONS OF VOC MIXTURES

Copulas

Parameters of the marginal distributions, goodness-of-fit statistics, and copula parameters are in Appendix Tables A.12 through A.14. AICs and BICs (Appendix Table A.13) for the different copulas were fairly similar to each other for mixtures A1, A3/B3, A4, and B1. However, for mixtures A2 and B2, the AICs and BICs were much lower for Gaussian and t copulas than for the other three copulas, which suggests that these two performed less well in their ability to describe the dependency structures. Gumbel copulas fitted mixtures A1 and B1 best, both of which had only two marker VOCs, whereas t copulas fitted mixtures A2, A3/B3, A4, and B2 best, each of which had four or more marker VOCs. We previously noted that the VOC exposures in RIOPA tended to have extreme value distributions (Su et al. 2012), and both Gumbel and t copulas better represent extreme values than other copulas (Schmidt 2006).

Fitting results also might have been affected by the detection frequency. For each VOC, data below the MDL were assigned the same value (0.5 MDL). Scatter plots for any two variables that contain many of the same values display in a star shape, which fitted the t copula. In contrast, mixtures A1 and B1 contained at least one marker VOC with very high detection frequencies (e.g., 96% for MTBE), and joint distributions did not show this star shape. Among other mixtures that contained at least two marker VOCs with many undetectable measurements, joint distributions formed star shapes. To explore this explanation, a hypothetical mixture of two VOCs with low detection frequencies (styrene at 49% and α-pinene at 66%) was modeled. In this case, the t copula showed the best fit, suggesting that copula fits are not influenced by the number of mixture components, but that mixtures containing components with low detection frequencies are better fitted by the t copula.

Table 9 shows the observed and predicted probabilities that all components in a mixture would exceed various percentile cut-offs. The differences between the observed and copula-estimated probabilities were generally small. For the binary mixtures A1 and B1 (with only two marker VOCs), differences ranged from 0.001 (A1 at the 90th percentile and B1 at 50th, 75th, and 95th percentiles) to 0.02 (A1 at the 75th percentile). For mixtures with three or more components, differences ranged from 0 (A3/B3 at the 95th percentile, and B2 at the 95th percentile) to 0.12 (A4 at the 50th percentile). These results suggest that copulas have better predictive ability for bivariate distributions than higher-order distributions.

Table 9 also shows the probability that each mixture component exceeds the given percentile cut-offs with the simplistic assumption that the mixture components are uncorrelated (independent). As expected, these probabilities fell far below the observed probabilities, especially at higher percentiles; for example, for mixture A4 at the 90th percentile cut-off, the assumption of independence gives a probability of 0.001 compared with the much larger probability of 0.023 for the observed data. Such differences demonstrate the importance of modeling dependencies in mixtures.

A sensitivity analysis was performed to examine the effect of measurements below MDLs. Four approaches for treating these values were considered using mixture B2 as a test case (benzene, MTBE, 1,4-DCB, TCE, and PERC). The first approach replaced each value by 0.5 MDL (as described earlier). The remaining approaches used three different and plausible distributions to replace values reported as MDLs: (1 and 2) random values based on two distributions censored to give only positive values: N(MDL, 0.5 MDL) and N(0, 0.5 MDL); and (3) random values drawn from a uniform distribution, U(0, MDL). For each of these approaches, five data sets were generated, and the average percentage of individuals that would exceed each specified lifetime cancer risk threshold was calculated for the mixture. The performance of the copula was largely unchanged except in the case of the uniform distribution where the highest risk

Table 9. Probability That the Marker VOCs in a Mixture Would All Exceed Various Percentile Cut-Offs for Observed and Simulated Data

				Probability		
Mixture[a]	Marker VOCs	Copula[b]	Percentile	Observed[c] (n = 299)	Copula (n = 1000)	Uncorrelated[d] (n = 1000)
A1	Benzene and MTBE	Gumbel	50	0.3545	0.3470	0.2500
			75	0.1371	0.1550	0.0625
			90	0.0502	0.0510	0.0100
			95	0.0201	0.0250	0.0025
A2	Toluene, ethylbenzene, *m- & p*-xylenes, and styrene	t	50	NC	0.1950	0.0625
			75	0.0635	0.0500	0.0039
			90	0.0134	0.0110	0.0001
			95	0.0033	0.0040	0
A3, B3	1,4-DCB, TCE, PERC, chloroform, and CTC	t	50	NC	0.0820	0.0313
			75	0.0067	0.0040	0.0010
			90	0.0033	0	0
			95	0	0	0
A4	*d*-Limonene, α-pinene, and β-pinene	t	50	0.3244	0.2070	0.1250
			75	0.1171	0.0480	0.0156
			90	0.0234	0.0060	0.0010
			95	0.0100	0.0030	0.0001
B1	Ethylbenzene and MTBE	Gumbel	50	0.3478	0.3490	0.0625
			75	0.1438	0.1430	0.0039
			90	0.0435	0.0510	0.0001
			95	0.0234	0.0240	0.0000
B2	Benzene, MTBE, 1,4-DCB, TCE, and PERC	t	50	NC	0.0630	0.0313
			75	0.0067	0.0060	0.0010
			90	0.0033	0	0
			95	0	0	0

[a] An A indicates the mixture was identified by PMF; a B indicates the mixture was identified by mode of action.

[b] Copulas were determined as best-fit; see Appendix Table A.13.

[c] NC indicates not calculated because styrene and TCE had detection frequencies < 50%.

[d] Analysis assumed that all components in a mixture were uncorrelated. Results were calculated by the equation of independent probability.

levels were overpredicted. As before, the multivariate log-normal simulations slightly overestimated the likelihood of risks at the 10^{-4} level and underpredicted higher risk levels. These results suggest that copula predictions were not strongly affected by MDLs and that our copula results are robust.

Gumbel and Gaussian copulas have previously been shown to best fit VOCs in the NHANES data that were highly correlated (Jia et al. 2010). However, the earlier study examined only bivariate mixtures and did not consider t copulas, which we found in this study to best fit much of the RIOPA data. However, the present study did find the same dependency structure as found in earlier analyses of NHANES for the benzene and MTBE mixture (Gumbel copulas).

Mixture Fractions

Median mixture fractions for RIOPA observed data and simulated data analyzed with copulas are shown in Table 10 for various percentile cut-offs. Often, a single compound dominated the mixture (shown in bold in the table); for example, MTBE accounted for 78% to 94% of

Table 10. Median Mixture Fractions Based on RIOPA Observations and Copula Analyses with Simulated Data[a]

		Median Mixture Fraction by Percentile Range[c]							
		Observed Data				Predicted by Best-Fit Copula[d]			
Mixture[b]	VOCs	50–75	75–90	90–95	95–100	50–75	75–90	90–95	95–100
A1	Benzene	0.222	0.150	0.169	0.099	0.179	0.177	0.137	0.173
	MTBE	**0.778**	**0.850**	**0.831**	**0.901**	**0.821**	**0.823**	**0.863**	**0.827**
A2	**Toluene**	**0.578**	**0.555**	**0.571**	**0.484**	**0.557**	**0.572**	**0.533**	**0.547**
	Ethylbenzene	0.072	0.071	0.085	0.083	0.073	0.072	0.080	0.074
	m- & p-Xylenes	0.300	0.316	0.328	0.368	0.303	0.280	0.298	0.291
	Styrene	0.024	0.020	0.019	0.012	0.038	0.040	0.037	0.038
A3, B3	**1,4-DCB**	**0.333**	**0.842**	**0.972**	**0.993**	**0.447**	**0.786**	**0.968**	**0.994**
	TCE	0.026	0.009	0.001	0.000	0.031	0.010	0.002	0.000
	PERC	0.165	0.032	0.005	0.001	0.128	0.031	0.009	0.001
	Chloroform	0.180	0.053	0.015	0.003	0.134	0.052	0.013	0.001
	CTC	0.065	0.023	0.005	0.001	0.069	0.024	0.006	0.001
A4	***d*-Limonene**	**0.667**	**0.661**	**0.754**	**0.765**	**0.720**	**0.751**	**0.825**	**0.850**
	α-Pinene	0.204	0.149	0.100	0.080	0.176	0.127	0.102	0.041
	β-Pinene	0.078	0.099	0.143	0.120	0.061	0.055	0.026	0.029
B1	Ethylbenzene	0.156	0.125	0.106	0.062	0.154	0.117	0.106	0.083
	MTBE	**0.844**	**0.875**	**0.894**	**0.938**	**0.846**	**0.883**	**0.894**	**0.917**
B2	Benzene	0.118	0.062	0.019	0.004	0.093	0.068	0.022	0.004
	MTBE	**0.606**	0.347	0.054	0.009	**0.552**	**0.515**	0.159	0.023
	1,4-DCB	0.134	**0.411**	**0.857**	**0.982**	0.127	0.170	**0.484**	**0.943**
	TCE	0.010	0.005	0.001	0.000	0.009	0.005	0.003	0.001
	PERC	0.054	0.019	0.004	0.001	0.031	0.016	0.012	0.001

[a] n = 299 for RIOPA and n = 1000 for simulated data.

[b] An A indicates the mixture was identified by PMF; a B indicates the mixture was identified by mode of action.

[c] Mixture fractions may not sum to 1. Dominant mixture faction is shown in **bold**.

[d] Copula simulations used the fitted marginal distributions shown in Appendix Table A.12 and the best-fit copula type shown in Table 9 and Appendix Table A.13.

the exposure in mixtures A1 and B1 for both observations and copula simulations. For each of the dominant components, the mixture fractions for the copula simulations matched the mixture fractions for the observed data in all mixtures at all levels, with one exception: mixture B2 at the 75th–90th percentile level showed 1,4-DCB as the dominant component for the observed data and MTBE for the copula simulations.

VOCs with strong indoor sources (e.g., 1,4-DCB and *d*-limonene) dominated mixtures A3 and A4, respectively, and their fractions increased with increasing percentiles. For example, the median fractions of 1,4-DCB in mixture A3 for 50th–75th percentile observations and simulations were 0.33 and 0.45, respectively; these both increased to 0.99 at the 95th–100th percentile. These results reflect the extreme values we found for 1,4-DCB. In contrast, mixture fractions varied little for mixtures A1, A2, and B1; for example, toluene was the dominant component in mixture A2 with mixture fractions of 0.58 and 0.56 for observations and simulations, respectively, at the 50th–75th percentile level, and 0.57 and 0.53, respectively, at the 90th–95th percentile. Consistent mixture fractions may suggest generated mixtures as compared with other types where compositions are more variable.

Mixture B2 shifted composition at upper percentiles; for example, the MTBE mixture fractions were 0.61 and 0.55 at the 50th–75th percentile levels for observations and simulations, respectively, but 1,4-DCB was dominant at the 95th–100th percentiles with mixture fractions of 0.98 and 0.94, respectively. These results show that mixtures such as B2 may be very heterogeneous with compositions that differ by exposure level. Note that this mixture was selected based on the mode-of-action for the component VOCs in increasing liver and renal cancer and not on the basis of common sources or high correlations. Mixture B2 may be considered an "incidental" mixture as it likely combines VOCs from different sources.

Mixtures A3/B3 and B2 were selected to investigate whether the mixture fractions estimated by the copulas were driven by copula type or by the marginal distribution of the components in the mixture. Both mixtures were simulated for five types of copulas, all using the same set of marginal distributions. (For these simulations, marginal distributions are shown in Appendix Table A.12 and mixture fractions in Appendix Table A.15.) For mixture A3/B3, the analysis revealed only small changes in median fractions; for example, 1,4-DCB remained the dominant component at high exposure levels, and its mixture fraction increased with percentile regardless of the copula applied. Mixture B2 showed larger differences between median fractions for the (best-fit) t and other copulas, and the dominant VOC at the 90th–95th percentile level differed among copulas; for example, the dominant VOCs were 1,4-DCB for the t and Clayton copulas, but MTBE for the Gaussian, Gumbel, and Frank copulas. Even though the t and Clayton copulas identified 1,4-DCB, its mixture fraction varied from 0.47 to 0.70 in the two copulas. This highlights the importance of the type of copula, not just the marginal distributions of the VOC components.

Estimated Cancer Risks

Estimated cancer risks for the mode-of-action mixtures B1, B2, and B3 are shown in Table 11 for observed data and simulations using copulas and multivariate lognormal distributions. Based on the observed data, VOC mixtures can present rather high cancer risks; for example, about 10% of RIOPA participants had exposures to mixtures B2 and B3 associated with a 10^{-3} or higher lifetime cancer risk. Mixture B1 posed lower risks; for example, a 25% chance of exceeding a risk of 10^{-5} and a 1% chance of exceeding 10^{-4}. For mixture B2, 3% of participants exceeded a very high risk level, 10^{-2}. Similar results were seen for mixture B3.

For each mixture, the copula simulations gave risk predictions that were generally similar to that for observations, although there is notable divergence at the highest risk levels, particularly for mixture B3 (Table 11). The highest risks ($> 10^{-3}$) were underestimated by both the copulas and the lognormal simulations, although copulas had smaller errors. For mixture B1, the lognormal simulation slightly overestimated the chance of exceeding a risk of 10^{-5}, but underpredicted higher risks. For mixture B2, the lognormal simulation again overestimated low to moderate risks

Table 11. Comparison of Predicted Cancer Risks for Three VOC Mixtures (B1–B3) Based on RIOPA Observations, Copulas, and Lognormal Distributions of Simulated Data[a]

Mixture	VOC	Type of Data and Model	Percentage of Individuals Estimated to Exceed Indicated Cancer Risk				
			1×10^{-6}	1×10^{-5}	1×10^{-4}	1×10^{-3}	1×10^{-2}
B1	Ethylbenzene and MTBE	Observations	100.0	25.4	1.0	0.0	0.0
		Copula simulations	97.5	27.1	0.6	0.0	0.0
		Lognormal simulations	96.9	32.0	0.0	0.0	0.0
B2	Benzene, MTBE, 1,4-DCB, TCE, and PERC	Observations	100.0	100.0	34.8	9.7	3.0
		Copula simulations	100.0	99.5	35.9	6.6	1.6
		Lognormal simulations	100.0	99.2	40.1	5.6	0.7
B3	1,4-DCB, TCE, PERC, chloroform, and CTC	Observations	100.0	100.0	44.5	11.0	3.3
		Copula simulations	100.0	99.8	44.8	9.5	1.9
		Lognormal simulations	100.0	99.7	53.6	6.7	0.2

[a] $n = 299$ for observations; $n = 1{,}000$ for simulations.

(10^{-4}); both copula and lognormal simulations underestimated the highest risks (10^{-3} to 10^{-2}). For mixture B3, the lognormal simulations significantly underestimated the highest cancer risks (10^{-2}). The cumulative probability plot (Figure 6) shows that the copulas sometimes overpredicted the highest values (information not seen in Table 11); for example, the highest observed risk for mixture B3 was 3.0×10^{-2} whereas the highest copula simulation was 8.1×10^{-2}. However, such cases were rare (< 1% of cases).

This analysis suggests that lognormal distributions are a poor choice to represent extreme values. It also highlights several important differences between predictions using lognormal distributions and copulas. Copulas can use any marginal distribution for each mixture component and the simulations used the best-fit marginal distribution (both type and parameters) for each VOC. This increases the flexibility and can improve fit for marginal distributions. However, the copula simulations propagate any mismatches in the marginal distributions, which may explain the underprediction of the higher risk levels. In addition, copulas permit asymmetric dependency structures that can emphasize extreme values or other portions of the distribution that display "local" dependencies; for example, mixture B1 fit the Gumbel copula, which emphasizes upper tail dependencies. Furthermore, copulas performed better than multivariate lognormal models in all cases, although the copula predictions also diverged from observations above the 95th percentile.

SPECIFIC AIM 3. IDENTIFICATION OF DETERMINANTS

METHODS: TIME, COMPARTMENT, AND SOURCE FRACTIONS OF VOC EXPOSURE

The data used for these analyses included 544 adult personal observations (n = 299 and n = 245 for first and second visits, respectively); 554 indoor observations (inside the home; n = 303 and n = 251); 555 outdoor observations (directly outside the home; n = 302 and n = 253); and participants' time–activity questionnaires (n = 532).

Time Fractions

The sampling time and time a participant spent in different locations (outdoors in neighborhood, outdoors out of neighborhood, indoors at home, indoors at school or work, other indoors, transportation, and unknown) were calculated for each participant from their time–activity

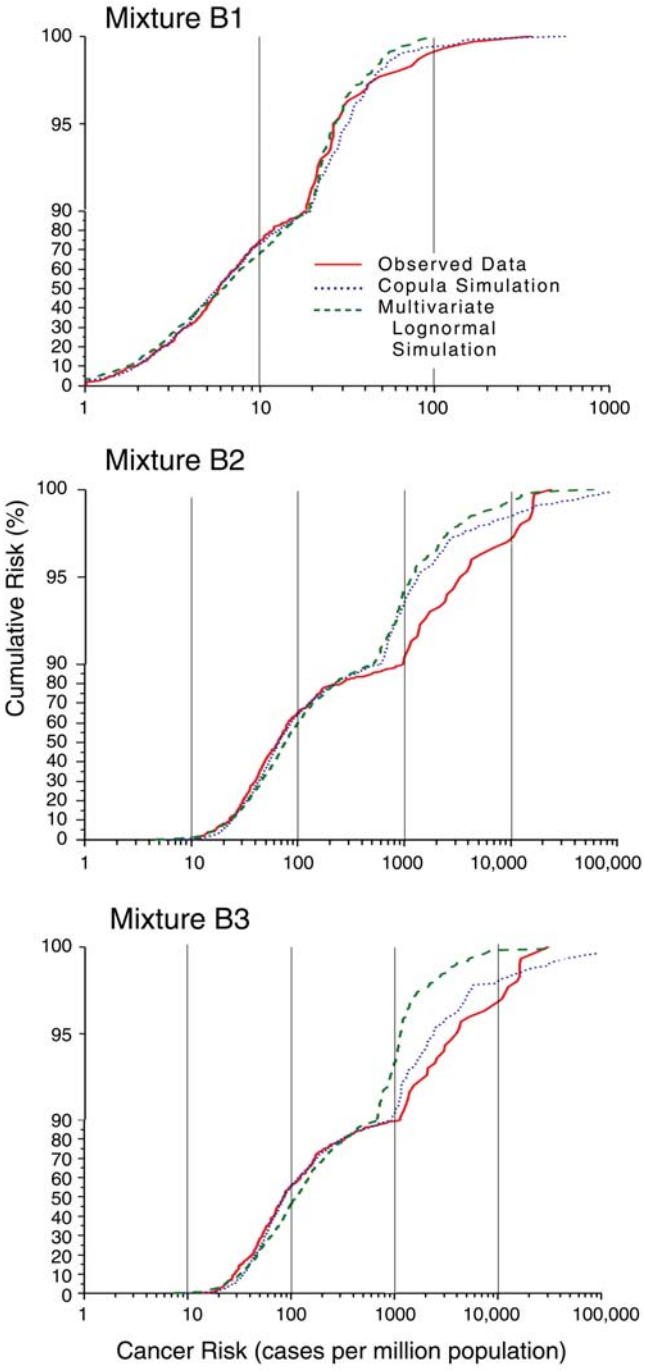

Figure 6. Probability plots of cumulative cancer risks for mixtures of VOC exposure based on observations from the RIOPA study (n = 299 for measurements on the first visit for adult participants), and copula and multivariate lognormal simulations (n = 1,000). The y axis scales emphasize differences at upper percentiles. Note that the scales on the x axes differ among panels

questionnaires. The fraction of unidentified or missing time was calculated as:

$$F_{t,miss} = (T_{total} - T_{outdoor} - T_{indoor} - T_{transit}) \ / \ T_{total}, \quad (10)$$

where T_{total} is total time, $T_{outdoor}$ is time spent outdoors in any environment, T_{indoor} is time spent indoors in any environment, and $T_{transit}$ is time spent in transit (all time measured in minutes). Some observations (n = 60) for participants who had missing time fractions that exceeded 0.25 were excluded from fraction analyses.

Compartment Fractions

An individual's total, cumulative, or potential exposure is often represented as the sum of the concentration–time products across all compartments or microenvironments in a given time period. We defined two compartments of exposure as outside the home ($F_{outdoor\_C}$) and inside the home (F_{home_C}). The outdoor compartment, $F_{outdoor\_C}$, was calculated for each participant as:

$$F_{outdoor\_C} = (C_{outdoor}\ T_{neighborhood}) / (C_{personal}\ T_{total}), \quad (11)$$

where $F_{outdoor\_C}$ is the fraction of a person's exposure due to being outdoors in their neighborhood, $C_{outdoor}$ is the concentration (μg/m^3) of a residential outdoor VOC, $T_{neighborhood}$ is time (minutes) spent outdoors in neighborhood, and $C_{personal}$ is personal VOC exposure (μg/m^3).

Similarly, the home compartment, F_{home_C}, was calculated as:

$$F_{home\_C} = (C_{home}\ T_{home}) / (C_{personal}\ T_{total}), \quad (12)$$

where F_{home_C} is the fraction of a person's exposure due to being in their home, C_{home} is the concentration (μg/m^3) of a VOC in the home, T_{home} is time (minutes) spent indoors at home, and $C_{personal}$ is personal VOC exposure (μg/m^3).

Source Fractions

In a separate analysis, we defined two broad sources of exposure as outside the home ($F_{outdoor\_S}$) and inside the home (F_{home_S}). These analyses assumed 100% penetration efficiency for outdoor VOCs entering a residence and 0% loss rate or VOC decay. These exposure sources were calculated as:

$$F_{outdoor\_S} = \left[C_{outdoor}(T_{home} + T_{neighborhood})\right] / (C_{personal}\ T_{total}), \quad (13)$$

and

$$F_{home\_S} = \left[(C_{home} - C_{outdoor})\ (T_{home})\right] / (C_{personal}\ T_{total}), \quad (14)$$

where $F_{outdoor\_S}$ is the fraction of a person's exposure attributable to VOCs from outdoor sources in the neighborhood, and F_{home_S} is the fraction of a person's exposure attributable to VOC sources in the home.

Exposure fractions were stratified by city and by warm (May to October) and cool (November to April) seasons. Differences among cities and between seasons were evaluated using Kruskal-Wallis (K-W) tests.

RESULTS AND DISCUSSION: TIME, COMPARTMENT, AND SOURCE FRACTIONS OF VOC EXPOSURE

Exposure Fractions by Time

Figure 7 displays the average time fractions participants spent outdoors, indoors, and in transit. Indoor time fractions (including time spent at home, time spent at school

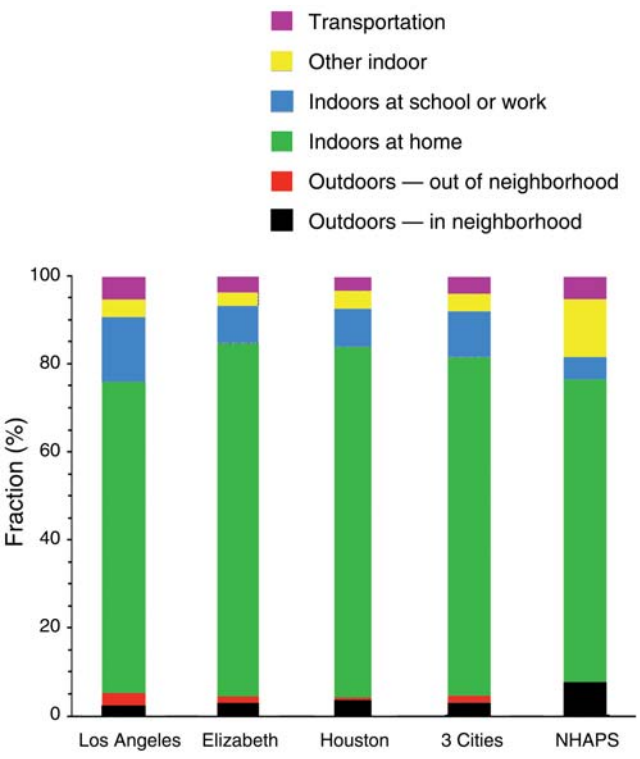

Figure 7. Mean time fractions spent in six locations for RIOPA participants (n = 544) based on their time–activity questionnaires. Results from RIOPA are compared with participants interviewed for the National Human Activity Pattern Survey (n = 9,196; see Klepeis et al. 2001).

or work, and time spent indoors elsewhere) averaged 89% in Los Angeles, 92% in Elizabeth, and 92% in Houston ($P < 0.001$); participants in Los Angeles spent the least time at home (71%, 80%, and 80% for the three cities, $P < 0.001$). This is explained in part by the lower unemployment rate in Los Angeles. Little time was spent outdoors, including time within and outside of their neighborhoods (fractions averaging 5.1% in Los Angeles, 4.5% in Elizabeth, and 4.3% in Houston; $P = 0.650$). Similarly, time spent in transit was small (5.5%, 3.6%, and 3.6% in the three cities, respectively, $P < 0.001$).

Figure 7 shows the RIOPA time fractions for each city and the three cities together compared with a nationally representative sample from the National Human Activity Pattern Survey (NHAPS), a probability-based telephone interview survey conducted from 1992 to 1994 that collected 24-hour time–activity, demographic, and exposure-related information from 9,196 respondents (Klepeis et al. 2001). NHAPS respondents spent more time outdoors (7.6%) than the RIOPA participants (4.6%), and less time indoors (87% NHAPS vs. 91% RIOPA) and at home (69% NHAPS vs. 77% RIOPA). RIOPA's eligibility criterion that participants be home at least 10 hours per day (Weisel et al. 2005b) may have increased the numbers of women (75% RIOPA vs. 54% NHAPS), participants over 64 years of age (18% RIOPA vs. 14% NHAPS), and unemployed (53% RIOPA vs. 35% NHAPS). In support of this result, NHAPS shows somewhat more time in transit and less time at school or work. Both RIOPA and NHAPS reflect the well-known pattern that most individuals spend an overwhelming fraction of time at home.

Exposure Fractions by Compartment

F_{home_C} exceeded 1 for 11% to 20% of the indoor VOC observations ($n = 52$ to 98, depending on the VOC), exceeded 1.25 for 5% to 11% ($n = 25$ to 53), and exceeded 1.5 for 2% to 8% ($n = 11$ to 39). Sampling error might explain a large part of the divergence from the assumptions in our model. It could be argued that cases in which $F_{home\_C} > 1$ should be excluded from analyses, but given the importance of indoor exposures, it was reasonable to assume that $F_{home\_C} \approx 1$ and $F_{outdoor\_C} \approx 0$ in such cases. Therefore, in our analyses, we excluded values of $F_{home\_C} > 1.25$.

On the basis of compartment fraction analyses, F_{home_C} VOC concentrations dominated personal exposures; for example, median F_{home_C} fractions ranged from 0.66 for MTBE to 0.78 for α-pinene, and the 95th percentile values approached or were higher than 1 for all VOCs (Table 12 and Appendix Figure A.14). The importance of the F_{home_C} is unsurprising since RIOPA participants spent most (median of 77%) of their time at home and since indoor concentrations of most VOCs were much higher than outdoor concentrations.

F_{home_C} differed among the three cities and by city for most VOCs (except toluene, *o*-xylene, 1,4-DCB, PERC, *d*-limonene, and β-pinene) and by season for two VOCs (benzene and MTBE; Table 13). The median F_{home_C} among the three cities was highest in Houston (0.68 to 0.81) for most VOCs (except benzene, styrene, PERC, and *d*-limonene; see Table 12). A city effect on indoor concentrations of VOCs is likely a result of differences in emission sources, meteorology, and household characteristics (e.g., presence of attached garage) among the three cities; seasonal effects may also be affected by lifestyle factors (e.g., opening windows and using air conditioners).

On the basis of compartment fraction analyses, $F_{outdoor\_C}$ VOC concentrations contributed very little to personal exposures; median fractions of $F_{outdoor\_C}$ were less than 0.01 for all VOCs except CTC (which was at 0.01; Table 12). (Even 95th percentile values of $F_{outdoor\_C}$ fell below 0.15 [Appendix Figure A.14].) A low $F_{outdoor\_C}$ is likely a result of both the little time spent outdoors (as discussed above) and the low outdoor VOC concentrations. Because many VOCs (8 of 15; toluene, styrene, 1,4-DCB, TCE, chloroform, *d*-limonene, α-pinene, β-pinene) had low detection frequencies (< 60%), the outdoor fractions are approximate. $F_{outdoor\_C}$ differed ($P < 0.05$) by season for all VOCs and by city for over half of the VOCs (benzene, toluene, *m- & p*-xylenes, *o*-xylene, MTBE, TCE, PERC, and CTC) (Table 13).

Exposure Fractions by Source

We assumed outdoor VOCs entered a residence with 100% penetration efficiency; therefore, $F_{outdoor\_S}$ fractions demonstrate the importance of outdoor sources of VOCs. With the exception of VOCs with strong indoor sources (1,4-DCB, chloroform, *d*-limonene, and α-pinene), $F_{outdoor\_S}$ exceeded F_{home_S}; for example, $F_{outdoor\_S} > 0.60$ for benzene, MTBE, TCE and CTC (Table 12). Thus, on the basis of source fraction analyses, and with the exception of VOCs with strong indoor sources, outdoor sources were the major contributor to personal exposures.

This conclusion may seem surprising given that many VOC studies have shown elevated indoor/outdoor (I/O) concentration ratios, thus implicating indoor VOC sources. However, unless I/O ratios exceed 2, outdoor sources will provide over half of the exposure (the lowest possible estimate based on an individual spending 100% time indoors). For the RIOPA participant spending an average of 91% of their time indoors, I/O ratios must exceed 2.1 for indoor sources to dominate exposure. Of the VOCs studied in RIOPA, median I/O ratios were 2.6, 4.4, 12.9, and 3.2 only for 1,4-DCB, chloroform, *d*-limonene,

Table 12. Median Compartment and Source Fractions of Personal VOC Exposure in RIOPA[a]

	Compartments								Sources	
	Median F_{home_C}[b]				Median $F_{outdoor\_C}$[c]				Median F_{home_S}[d]	Median $F_{outdoor\_S}$[e]
VOC	All	CA	NJ	TX	All	CA	NJ	TX	All	All
Benzene	0.72	0.64	0.76	0.73	< 0.01	< 0.01	< 0.01	< 0.01	0.05	0.61
Toluene	0.66	0.63	0.67	0.68	< 0.01	< 0.01	< 0.01	< 0.01	0.18	0.38
Ethylbenzene	0.69	0.64	0.68	0.73	< 0.01	< 0.01	< 0.01	< 0.01	0.15	0.48
m- & *p*-Xylenes	0.68	0.64	0.67	0.75	< 0.01	< 0.01	< 0.01	< 0.01	0.13	0.48
o-Xylene	0.69	0.65	0.67	0.71	< 0.01	< 0.01	< 0.01	< 0.01	0.10	0.50
MTBE	0.66	0.63	0.58	0.72	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	0.67
Styrene	0.74	0.72	0.79	0.72	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	0.58
1,4-DCB	0.72	0.67	0.73	0.76	< 0.01	< 0.01	< 0.01	< 0.01	0.29	0.23
TCE	0.74	0.66	0.74	0.80	< 0.01	< 0.01	< 0.01	0.01	< 0.01	0.76
PERC	0.71	0.69	0.75	0.71	< 0.01	< 0.01	< 0.01	< 0.01	0.03	0.56
Chloroform	0.74	0.74	0.70	0.81	< 0.01	< 0.01	< 0.01	< 0.01	0.57	0.16
CTC	0.75	0.72	0.74	0.79	0.01	< 0.01	0.01	0.02	< 0.01	0.81
d-Limonene	0.71	0.72	0.67	0.71	< 0.01	< 0.01	< 0.01	< 0.01	0.60	0.05
α-Pinene	0.78	0.79	0.74	0.81	< 0.01	< 0.01	< 0.01	< 0.01	0.45	0.22
β-Pinene	0.76	0.76	0.73	0.78	< 0.01	< 0.01	< 0.01	< 0.01	0.19	0.34

[a] n = 484 personal observations; 60 observations were excluded for participants who had missing time fractions > 0.25.

[b] $F_{home\_C} = (C_{home} \times T_{home}) / (C_{personal} \times T_{total})$.

[c] $F_{outdoor\_C} = (C_{outdoor} \times T_{neighborhood}) / (C_{personal} \times T_{total})$.

[d] $F_{home\_S} = ([C_{home} - C_{outdoor}] \times [T_{home}]) / (C_{personal} \times T_{total})$.

[e] $F_{outdoor\_S} = (C_{outdoor} \times [T_{home} + T_{neighborhood}]) / (C_{personal} \times T_{total})$.

Table 13. Differences in VOC Compartment Fractions Between Cities and Seasons in RIOPA[a]

	F_{home_C}[b]							$F_{outdoor\_C}$[c]						
	Among Cities			Between Seasons				Among Cities			Between Seasons			
VOC	City	Hot Season	Cool Season	Season	CA	NJ	TX	City	Hot Season	Cool Season	Season	CA	NJ	TX
Benzene	**0.001**	0.053	**0.007**	**0.014**	0.337	0.118	**0.039**	**0.021**	0.406	**0.005**	**0.011**	0.053	**0.016**	0.701
Toluene	0.138	0.415	0.258	0.950	0.908	0.603	0.695	**0.017**	0.462	**0.004**	**0.008**	**0.038**	**0.032**	0.572
Ethylbenzene	**0.028**	**0.017**	0.487	0.110	**0.021**	0.614	0.772	0.125	0.676	**0.030**	**0.007**	0.062	**0.026**	0.476
m- & p- Xylenes	**0.013**	**0.041**	0.234	0.412	0.552	0.371	0.948	**0.049**	0.492	**0.022**	**0.009**	0.060	**0.047**	0.377
o-Xylene	0.072	**0.041**	0.808	0.597	0.371	0.427	0.463	**0.029**	0.433	**0.014**	**0.010**	0.056	**0.042**	0.395
MTBE	**0.004**	**0.006**	0.555	**0.001**	**0.009**	0.098	0.157	**0.006**	0.316	**0.002**	**0.013**	0.064	**0.013**	0.661
Styrene	**0.039**	**0.008**	0.780	0.068	0.377	**0.032**	0.847	0.196	0.313	0.056	**0.002**	**0.050**	**0.005**	0.490
1,4-DCB	0.255	0.151	0.810	0.467	0.075	0.772	0.940	0.409	0.816	0.072	**0.004**	0.051	**0.010**	0.509
TCE	**0.000**	0.052	**0.003**	0.798	0.276	0.554	0.468	**0.000**	**0.039**	**0.001**	**0.004**	0.061	**0.013**	0.240
PERC	0.329	0.534	0.642	0.358	0.456	0.604	0.563	**0.016**	0.274	**0.010**	**0.006**	**0.046**	**0.020**	0.527
Chloroform	**0.001**	**0.006**	0.138	0.280	0.439	0.404	0.921	0.108	0.376	**0.023**	**0.021**	0.084	**0.019**	0.970
CTC	**0.000**	**0.003**	**0.001**	0.526	0.980	**0.024**	0.454	**0.001**	0.063	**0.001**	**0.007**	0.059	**0.004**	0.618
d-Limonene	0.053	0.259	0.096	0.827	0.272	0.454	0.767	0.721	0.145	0.292	**0.002**	**0.012**	**0.015**	0.476
α-Pinene	**0.017**	0.063	0.220	0.629	0.840	0.920	0.423	0.076	**0.006**	0.532	**0.004**	**0.040**	**0.003**	0.910
β-Pinene	0.175	0.663	0.183	0.302	0.504	0.844	0.232	0.108	0.547	**0.005**	**0.018**	0.063	**0.006**	0.721

[a] *P* values for K-W tests are shown. *P* values ≤ 0.05 (**in bold**) indicate that at least one of the medians in the group is different from the others. n = 427–455 for F_{home_C}, and 480–481 for $F_{outdoor\_C}$, depending on the VOC.

[b] $F_{home\_C} = (C_{home} \times T_{home}) / (C_{personal} \times T_{total})$.

[c] $F_{outdoor\_C} = (C_{outdoor} \times T_{neighborhood}) / (C_{personal} \times T_{total})$.

and α-pinene, respectively. For all other VOCs, outdoor sources contributed most of the personal exposure. This conclusion parallels prior RIOPA analyses in showing that indoor and personal $PM_{2.5}$ exposures were mostly due to outdoor sources (Meng et al. 2007; Meng et al. 2009; Polidori et al. 2006).

METHODS: IDENTIFICATION OF DETERMINANTS OF PERSONAL, HOME, AND OUTDOOR VOC EXPOSURE

Variable Selection

As an initial step to identify possible exposure determinants, each of the 527 RIOPA variables was used in univariate regression models with outdoor, indoor, and personal VOC observations as dependent variables. These models used six VOCs (benzene, toluene, MTBE, 1,4-DCB, PERC, and chloroform) that were selected to represent a range of VOCs and potential emission sources. Next, variables that attained statistical significance ($P < 0.05$) in the univariate models were used in forward stepwise multivariate regression models with selection based on the Schwarz BIC. Although this reduced the number of variables, the resulting parameter estimates were approximate since these models did not account for possible correlations due to clustering and nesting (e.g., two seasonal samples for most participants).

Linear Mixed-Effects Models

LMMs that incorporated fixed and random effects and repeated measures (Krueger and Tian 2004) were estimated for outdoor, indoor, and personal observations using the variables selected by the stepwise models. These models also incorporated several variables with strong theoretical support or of special interest (e.g., city, ethnicity, and presence of an attached garage). Two-way interactions among variables with city, use of air-conditioner, opened windows, and time spent indoors at home were also evaluated. Using log-transformed VOC concentrations, random intercepts, nested effects for city, and interactions, the LMMs were expressed as:

$$\log(C_{ti}) = (\beta_0 + b_{0i}) + \beta_1\,\mathrm{Visit}_t + \beta_2\,\mathrm{City} + \ldots + \beta_n X_n + \varepsilon_{ti}, \quad (15)$$

where C_{ti} is VOC concentration (μg/m^3) at time t for individual i, β is model coefficient for fixed effects, b is random deviation from the overall fixed effects, Visit_t is sample collected at time t, X is other covariates, and ϵ_{ti} is random error of the VOC concentrations at time t for individual i. Since the LMMs used log-transformed VOCs, the effect size for each explanatory variable was calculated as follows:

$$\text{Effect size} = e^{(\beta U)}, \quad (16)$$

where e is exponential and U = 1 for categorical variables and the interquartile range (IQR) for continuous variables.

To maintain a sufficient sample size, variables with 400 observations or more were included in the final LMMs. Separate LMMs were developed for the 15 VOCs, and grouped into three categories based on common determinants: gasoline-related VOCs (BTEX, MTBE, and styrene); odorant and cleaning-related VOCs (1,4-DCB, chloroform, *d*-limonene, α-pinene, and β-pinene); and dry-cleaning and industry-related VOCs (TCE, PERC, and CTC). Note that some determinant analyses were performed using VOC mixtures instead of individual VOCs (Appendix Table A.11.)

Model Assessment

The following steps were taken to help verify model results. Partial residual plots were examined to assess the linearity and fit of continuous variables (e.g., wind speed and household AERs). Transformations (e.g., log or reciprocal transformation) were tested for variables showing nonlinear relationships. Because the reduction in residual variance (R^2) attributable to fixed-effects variables cannot be directly obtained from the SAS procedure, R^2 was estimated as:

$$R^2 = \left(\sigma^2_{\mathrm{int}} - \sigma^2_{\mathrm{full}}\right) / \sigma^2_{\mathrm{int}}, \quad (17)$$

where σ^2_{int} is the variance of the intercept-only model, and σ^2_{full} is the variance of the full model. Here, R^2 indicates the difference of variance between reduced (i.e., intercept-only) and full (i.e., with predictor variables) models.

Missing Data

Candidate variables in the LMMs typically had 50 to 100 missing observations. The effect of missing data was evaluated using multiple imputation (MI), and results from the datasets with imputed values (no missing data) were compared with results from the original data set. Three VOCs for each sample type were selected for this comparison with different numbers of missing observations; for example, for personal observations, styrene, benzene, and *d*-limonene were selected with 3%, 20%, and 28% of observations missing, respectively. Differences between the original and MI data sets were computed as the relative change in model estimates of β.

The results of this comparison (Appendix Tables A.16 through A.18) demonstrated that although models using imputed data tended to have smaller (more statistically significant) *P* values, the differences between the data set with imputed values and the original data set were not large. Also, the model parameters themselves did not show obvious biases. Differences tended to increase with the fraction of missing data, although changes were generally small and, among the nine models tested, only outdoor and personal benzene and indoor PERC had three parameters change by more than 30%. Because missing data did not greatly affect the LMM results, subsequent analyses did not use MI.

Most analyses used SAS 9.2 (SAS Institute, Cary, North Carolina, USA). Variable selection used proc glmselect, LMMs used proc mixed, and MI analyses used proc mi and proc mianalyze. Partial residual plots were drawn in R version 2.13.1 (R Development Core Team, Vienna, Austria). Relative changes were calculated using Excel (Microsoft, Redmond, WA).

RESULTS AND DISCUSSION: IDENTIFICATION OF DETERMINANTS OF PERSONAL, HOME, AND OUTDOOR VOC EXPOSURE

Personal Exposure to Gasoline-Related VOCs

BTEX, MTBE, and styrene, all components of gasoline and vehicle exhaust, shared several exposure determinants (Table 14; see also Appendix Table A.19). Increased exposures were associated with living in Houston, homes with attached garages, and pumping gas; decreased exposures were associated with higher wind speeds and higher home AERs. Interestingly, lower exposures of toluene, ethylbenzene, and *o*-xylene were found for participants who reported cooking activities during the sampling period, possibly because these individuals drove less for food-related activities (such as dining out). Indeed, participants who reported cooking activities spent less time in cars with closed windows (mean time spent = 71 min) than those who did not report cooking activities (mean time spent = 88 min, *P* value of *t* test = 0.038). (No differences were seen for time in cars with open windows or for total travel time.)

The literature supports these findings for BTEX, MTBE, and styrene (Appendix Table A.1). In Houston, important outdoor VOC sources included petrochemical facilities and vehicles (Weisel et al. 2005b); and pumping gasoline has been shown to elevate personal exposure to BTEX in cold weather in Alaska (Backer et al. 1997); and attached garages are known sources of gasoline-related aromatics in homes (Batterman et al. 2007; D'Souza et al. 2009; Delgado-Saborit et al. 2009; Sexton et al. 2007; Symanski et al. 2009; Wang et al. 2009). The effects of both pumping gas and attached garages on gasoline-related VOCs were also seen in NHANES (Symanski et al. 2009). Concentrations from outdoor sources (e.g., vehicle exhaust) are diluted by wind (U.S. EPA 2010a), so higher wind speeds may lower exposures. The home AER, which accounts for infiltration and ventilation and which depends on wind speed (U.S. EPA 2011), influences indoor concentrations and thus personal exposures for those pollutants from indoor sources.

Cooking-related activities have been shown to increase indoor and personal concentrations of several VOCs (e.g., benzene and toluene; Byun et al. 2010; Clobes et al. 1992). However in the current RIOPA analyses, negative associations were seen between cooking and personal exposures to toluene, ethylbenzene, and *o*-xylene. This inconsistency could be explained by statistical chance, although the explanation offered above — that participants without cooking activity traveled more to dine out during which time they were exposed to gasoline-related VOCs — appears reasonable. The RIOPA data do not allow further analysis, but we speculate that visits to drive-through fast food facilities where vehicles are queued up and idling may be a particularly important source of VOC exposure.

Personal Exposure to Odorant and Cleaning-Related VOCs

Four determinants were found for the group of odorant and cleaning-related VOCs (1,4-DCB, chloroform, *d*-limonene, α-pinene, and β-pinene) (Table 15 and Appendix Table A.20). Like the gasoline-related VOCs, Houston participants had higher exposure to these VOCs. AERs were negatively associated with VOC exposures, reflecting dilution of indoor sources. Participants in larger homes (more rooms) tended to have lower exposure to 1,4-DCB, chloroform, *d*-limonene and α-pinene. Interestingly, the behavior of other household members was associated with personal exposure; for example, non-participants showering during the sampling period was associated with higher exposures of chloroform, *d*-limonene, α-pinene, and β-pinene.

The odorant and cleaning-related VOCs are primarily released by indoor sources, such as mothballs, air fresheners, cleansers, and chlorinated water (ATSDR 1997a, 2006; Chin et al. 2013; U.S. EPA 2012a). Thus, the use and storage of these products can affect exposure. Also, since these VOCs arise mainly from indoor sources, AER is expected to be a determinant (Mudarri 2010). The identification of the number of rooms, a suggestion of home size, as a determinant may reflect additional mixing in large homes that lowers concentrations compared with approximately the same product use in smaller homes. We have

Table 14. Results of Linear Mixed-Effects Models for Personal Exposure to Gasoline-Related VOCs[a]

Variable	Group or Unit of Change[b]	Benzene β	Benzene 95% CI	Toluene β	Toluene 95% CI	Ethylbenzene β	Ethylbenzene 95% CI	*m*- & *p*-Xylenes β	*m*- & *p*-Xylenes 95% CI	*o*-Xylene β	*o*-Xylene 95% CI	MTBE β	MTBE 95% CI	Styrene β	Styrene 95% CI
Intercept		2.21	1.41, 3.01	3.74	3.01, 4.47	1.41	0.59, 2.23	2.23	1.50, 2.96	0.78	0.21, 1.35	1.82	1.19, 2.45	1.09	0.44, 1.74
Visit	1	−0.03	−0.17, 0.11	0.12	−0.06, 0.30	−0.14	−0.30, 0.02	−0.08	−0.24, 0.08	−0.07	−0.21, 0.07	0.06	−0.14, 0.26	0.07	−0.09, 0.23
	2	Reference		Reference		Reference		Reference		Reference		Reference		Reference	
City	Los Angeles	−0.83	−1.07, −0.59	0.08	−0.14, 0.30	−0.37	−0.64, −0.10	−0.29	−0.56, −0.02	−0.06	−0.31, 0.19	−0.35	−0.66, −0.04	−0.23	−0.45, −0.01
	Elizabeth	−0.37	−0.64, −0.10	0.06	−0.19, 0.31	−0.16	−0.51, 0.19	−0.25	−0.62, 0.12	−0.17	−0.50, 0.16	0.07	−0.32, 0.46	−0.11	−0.31, 0.09
	Houston	Reference		Reference		Reference		Reference		Reference		Reference		Reference	
Attached garage	No	−0.19	−0.37, −0.01	−0.72	−1.21, −0.23	−0.36	−0.60, −0.12	−0.36	−0.60, −0.12	−0.35	−0.57, −0.13	−0.36	−0.60, −0.12	−0.42	−0.91, 0.07
Cooking	No			0.22	0.04, 0.40	0.17	0.01, 0.33	0.15	−0.03, 0.33	0.20	0.04, 0.36				
Education	< High school	0.15	−0.09, 0.39												
	High school	−0.08	−0.28, 0.12												
	> College	Reference													
Ethnicity	White					−0.13	−0.42, 0.16	−0.23	−0.54, 0.08	−0.21	−0.48, 0.06				
	Mexican					0.19	−0.18, 0.56	0.07	−0.30, 0.44	0.12	−0.21, 0.45				
	Hispanic					0.30	−0.07, 0.67	0.27	−0.12, 0.66	0.35	0.00, 0.70				
	Other					Reference		Reference		Reference					
Heating fuel	Electricity	0.20	−0.15, 0.55												
	Gas	0.42	0.11, 0.73												
	Oil and wood	Reference													
Indoor temperature	°C	−0.04	−0.06, −0.02												
Inverse wind speed	1/knot	4.20	3.16, 5.24			3.16	1.81, 4.51	2.84	1.45, 4.23	2.54	1.32, 3.76	5.86	4.21, 7.51		
Log-transformed AER	1/hour			−0.30	−0.40, −0.20	−0.17	−0.29, −0.05	−0.21	−0.33, −0.09	−0.14	−0.24, −0.04	−0.09	−0.23, 0.05		
Number of floors		−0.15	−0.23, −0.07									−0.20	−0.32, −0.08		
Number of rooms		−0.10	−0.16, −0.04											−0.09	−0.13, −0.05
Open doors or windows	No									0.22	0.02, 0.42			0.20	0.02, 0.38
Pumping gas	No	−0.16	−0.32, 0.00			−0.24	−0.46, −0.02	−0.22	−0.44, 0.00	−0.28	−0.48, −0.08	−0.34	−0.59, −0.09		
Renovation in the past year	No			−0.30	−0.50, −0.10										
Time spent in home	minutes			0.00	−0.0004, 0	0.00	−0.0004, 0							0.00	−0.0005, 0
Unemployed	No											0.23	−0.01, 0.47		
Using air cleaning devices	No					−0.27	−0.62, 0.08	−0.42	−0.77, −0.07	−0.38	−0.69, −0.07	−0.35	−0.74, 0.04		
Using nail polish remover	No			−0.29	−0.62, 0.04	−0.39	−0.70, −0.08	−0.33	−0.66, 0.00						
Wore powder, spray, or perfume	No											0.41	0.17, 0.65		

[a] n = 400 to 530 observations depending on the model.

[b] For dichotomous variables, the reference group is "Yes".

Table 15. Results of Linear Mixed-Effects Models for Personal Exposure to Odorant and Cleaning-Related VOCs[a]

Variable	Group or Unit of Change[b]	1,4-DCB		Chloroform		*d*-Limonene		α-Pinene		β-Pinene	
		β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI
Intercept		3.50	1.97, 5.03	1.34	0.42, 2.26	3.62	2.86, 4.38	2.42	1.93, 2.91	1.57	0.71, 2.43
Visit	1	0.33	0.06, 0.60	0.15	−0.03, 0.33	0.10	−0.19, 0.39	0.18	0.04, 0.32	0.08	−0.12, 0.28
	2		Reference		Reference		Reference		Reference		Reference
City	Los Angeles	−1.10	−1.69, −0.51	−0.45	−0.76, −0.14	−0.82	−1.19, −0.45	−0.71	−0.96, −0.46	−1.16	−1.45, −0.87
	Elizabeth	−0.81	−1.42, −0.20	−0.06	−0.39, 0.27	−1.12	−1.55, −0.69	−0.59	−0.86, −0.32	−1.06	−1.39, −0.73
	Houston		Reference		Reference		Reference		Reference		Reference
Air conditioning	No	0.54	0.09, 0.99					−0.51	−0.71, −0.31	−0.20	−0.45, 0.05
Ambient relative humidity	%			−0.01	−0.02, 0.00					−0.01	−0.02, 0.00
Furniture refinisher in neighborhood	No	−1.30	−2.28, −0.32								
Waxing or polishing furniture	No	−0.81	−1.46, −0.16								
Keeping dogs or cats	No							0.15	−0.05, 0.35	0.29	0.07, 0.51
Log-transformed AER	1/hour			−0.41	−0.53, −0.29	−0.33	−0.49, −0.17	−0.40	−0.50, −0.30	−0.31	−0.45, −0.17
Not using fresheners or candles	No									0.32	−0.03, 0.67
Number of rooms		−0.14	−0.28 0.00	−0.12	−0.20, −0.04	−0.13	−0.21, −0.05	−0.10	−0.16, −0.04		
Open doors or windows	No	0.42	0.03, 0.81							0.22	−0.02, 0.46
Other family members took showers	No			−0.39	−0.68, −0.10	−0.80	−1.15, −0.45	−0.41	−0.65, −0.17	−0.35	−0.62, −0.08
Outdoor swimming pool or hot tub	No							−0.31	−0.56, −0.06		
Using heating at	< 64°F	0.76	0.25, 1.27								
	64–70°F	−0.03	−0.50, 0.44								
	> 70°F		Reference								
Ownership of the home	No			0.30	0.03, 0.57						
Pets indoors	No			0.32	0.08, 0.56						
Renovation in the past year	No					−0.45	−0.74, −0.16				
Restaurants or bakery in neighborhood	No	−0.63	−1.16, −0.10								
Unemployed	No					−0.35	−0.66, −0.04				
Using a clothes washer	No	0.53	0.16, 0.90								
Using dishwashers	No			−0.25	−0.50, 0.00						
Using other heaters (not central heating)	No					0.55	0.02, 1.08				

[a] n = 393 to 433 observations depending on the model.

[b] For dichotomous variables, the reference group is "Yes".

previously noted that in low-income households, which are usually smaller and sometimes crowded, there may be a tendency to try to mask odors using heavier applications of cleaners and fragrances that would increase concentrations (Chin et al. 2013). In our current analyses, the number of rooms in a home was positively associated with household income ($\beta = 0.79$, P value < 0.001), and thus socioeconomic factors may be an indirect or interacting factor associated with high exposures of odorant and cleaning-related VOCs. However, in this study no association between household income and VOC exposures was found. The effect of employment on *d*-limonene exposure might result from unemployed participants spending more time at home (2,278 and 2,000 minutes for unemployed and employed participants, respectively; P value < 0.001) and possibly engaged in chores that increased their contact with cleaners and odorants.

Chloroform is a byproduct produced when chlorine is used as a water disinfectant, thus drinking water, contacting water (e.g., bathing), and inhaling water vapor can increase exposure (ATSDR 1997a). Elevated chloroform concentrations in a room adjoining a bathroom during showering has been noted and called "secondary shower exposure" (Gordon et al. 2006). Such secondary exposure is consistent with findings that chloroform exposure increased when another family member showered. However, bathing or showering by the RIOPA participants themselves did not affect their exposure. Similar (negative) results with showering were found for the 1999–2000 NHANES data set, possibly due to a lack of variance in showering-related variables since most (85%) participants showered during the sampling period (Riederer et al. 2009). The same explanation may apply to the present study in which 87% of participants showered during the sampling period. In addition, participants were instructed not to get the samplers wet, and they may have removed them outside the bathroom (Weisel et al. 2005b).

The effect of city can be attributable to several factors, including differences in outdoor emission sources (e.g., industry and traffic), meteorologic factors that affect both dispersion and emissions of outdoor pollutants, systematic differences in building AERs, and demographic and cultural factors (Weisel et al. 2005b). For example, outdoor temperatures were considerably warmer in Houston during the sampling period, compared with Los Angeles and Elizabeth (averaging 22.3 $\pm$ 7.5 in Houston, 18.6 $\pm$ 4.7 in Los Angeles, and 14.6 $\pm$ 8.6°C in Elizabeth; P value < 0.001). Higher temperatures increase vapor pressures, permeation rates, and evaporation rates, possibly producing higher concentrations. Since a fraction of odorant and cleaning-related VOCs arise from volatilization and sublimation from indoor sources, indoor temperatures were also important. Indoor temperatures showed less variation and fewer differences were not significant (averaging 23.3 $\pm$ 2.6 in Los Angeles, 23.9 $\pm$ 2.6 in Elizabeth, and 24.0 $\pm$ 3.4°C in Houston; P value = 0.052).

Personal Exposure to Dry-Cleaning and Industry-Related VOCs

The dry-cleaning and industrial emissions group had three VOCs (TCE, PERC, and CTC), which were affected by city and household water source (Table 16; see also Appendix Table A.21). Elizabeth and Los Angeles participants had the highest TCE and PERC exposures, but Houston participants had the highest CTC exposure. Public water supplies were associated with lower TCE exposure, but higher CTC exposure.

As expected, PERC exposures increased by visiting a dry cleaner. This solvent has been widely used for dry cleaning clothes; exposure occurs when visiting dry cleaning establishments and storing dry cleaned clothes at home, whether or not the clothes are wrapped in plastic (Sherlach et al. 2011). PERC exposures were higher among employed participants. Since PERC has been widely used in industry as a degreaser and also has been added to products such as adhesives and paint removers (ATSDR 1997c), employed participants may have more chances to be exposed. The city effect may be related to population density: Los Angeles and Elizabeth have densities higher than Houston (Weisel et al. 2005b), which may lead to more dry cleaners and elevated ambient concentrations. The outdoor PERC levels were higher in Los Angeles and Elizabeth than in Houston (median was 1.29 in Los Angeles, 0.74 in Elizabeth, and 0.11 μg/m^3 in Houston; P value < 0.001).

TCE has also been used extensively as a degreaser, paint remover, adhesive, and chemical intermediate (ATSDR 1997b). Exposure may increase if TCE-containing consumer or home products are present (e.g., vinyl siding, glue, and car stain removers; U.S. EPA 2007b). In addition, TCE is sometimes found in contaminated soils and groundwater; thus participants in households near to subsurface or surface contaminated soils may be exposed indoors through soil vapor intrusion and through water consumption if a local well (especially a private well without water monitoring or treatment) is the water source. In the RIOPA data set, the TCE detection frequency was only 31%; thus only the higher levels were quantified. In consequence, TCE results may not be robust.

Most commercial uses of CTC were phased out by 1986 due to this chemical's toxicity and persistence, and industrial emissions also have been limited under the Clean Air Act Amendments of 1990 (ATSDR 2005b). Previously, CTC

Table 16. Results of Linear Mixed-Effects Models for Personal Exposure to Dry-Cleaning and Industrial-Related VOCs[a]

Variable	Group or Unit of Change[b]	TCE		PERC		CTC	
		β	95% CI	β	95% CI	β	95% CI
Intercept		−0.79	−1.61, 0.03	−0.48	−1.44, 0.48	−0.64	−1.09, −0.19
Visit	1	0.18	0.04, 0.32	0.19	−0.01, 0.39	−0.01	−0.07, 0.05
	2		Reference		Reference		Reference
City	Los Angeles	0.66	0.39, 0.93	0.58	0.23, 0.93	−0.17	−0.31, −0.03
	Elizabeth	1.23	0.96, 1.50	0.54	0.07, 1.01	−0.11	−0.25, 0.03
	Houston		Reference		Reference		Reference
Ambient relative humidity	%			−0.01	−0.03, 0.01		
Ethnicity	White			−0.12	−0.49, 0.25		
	Mexican			−0.48	−0.93, −0.03		
	Hispanic			0.06	−0.41, 0.53		
	Other				Reference		
Having a fireplace	No					−0.13	−0.27, 0.01
Indoor temperature	°C	−0.03	−0.05, −0.01			0.01	−0.01, 0.03
Inverse wind speed	1/knot			4.87	3.24, 6.50		
Log-transformed AER	1/hour			−0.20	−0.34, −0.06		
Not using fresheners or candles	No					−0.20	−0.36, −0.04
Restaurants or bakery in neighborhood	No	0.26	0.01, 0.51				
Source of household water	Public	−0.58	−1.11, −0.05			0.50	0.23, 0.77
Sweeping indoors	No			0.19	−0.05, 0.43		
Time spent in closed cars	minutes	0.00	0.0008, 0.0028				
Unemployed	No			0.42	0.17, 0.67		
Using air cleaning devices	No					−0.19	−0.35, −0.03
Vinyl, asbestos or other siding	No	−0.25	−0.50, 0.00				
Visited dry cleaners during past week	No			−0.63	−0.92, −0.34		

[a] n = 400 to 446 observations depending on the model.

[b] For dichotomous variables, the reference group is "Yes".

had been used in medical treatment and as a component in fire extinguishers, fumigants, and pesticides. Currently, CTC use is permitted only in a few industrial processes for which there is no effective substitute. CTC is globally distributed at generally low levels with little spatial variation, except near contaminated areas where levels increase. The variation in CTC exposures among the RIOPA participants was limited as was the prediction ability of the model.

Summary of Key Personal Exposure Determinants

The most common and significant determinants of personal VOC exposure were city, inverse wind speed, log-transformed AER, number of rooms, presence of an attached garage, and pumping gas. Inverse wind speed was positively associated with log-transformed benzene, ethylbenzene, *m-& p*-xylenes, *o*-xylene, MTBE, and PERC. Log-transformed AER was negatively associated with log-transformed toluene, ethylbenzene, *m- & p*-xylenes, *o*-xylene, PERC, chloroform, *d*-limonene, α-pinene, and β-pinene. Participants living in larger homes (more rooms) had lower exposures to benzene, styrene, 1,4-DCB, chloroform, *d*-limonene, and α-pinene; those in homes with attached garages were exposed to higher levels of benzene, toluene, ethylbenzene, *m- & p*-xylenes, *o*-xylene, and MTBE. Participants who pumped gas had higher exposures to benzene, ethylbenzene, *m- & p*-xylenes, *o*-xylene, and MTBE. Although the effects varied, participants in Houston usually had higher exposures than participants in Los Angeles and Elizabeth. The effect of employment lowered *d*-limonene exposure but increased PERC exposure (Tables 15 and 16). These effects were significant and based on LMMs, which controlled for clustering and repeated measures. As discussed later, the LMMs explained 0.003 (CTC) to 0.4 (β-pinene) of the variance in personal exposure.

Few significant interactions among VOC determinants were found. For example, two-way interactions (including city and air-conditioner use, city and opened doors, city and opened windows, and city and time spent indoors at home) were generally not significant in the LMMs. Thus, interaction terms were not retained in the final models.

Determinants Identified for Other Compounds Measured in RIOPA

Several reports have identified determinants of personal exposure for chemicals other than VOCs measured in RIOPA, including carbonyls (Liu et al. 2007) and $PM_{2.5}$ (Meng et al. 2009). For personal exposures to carbonyls, yard and gardening activities were associated with increasing acetaldehyde, acetone, crotonaldehyde, and hexaldehyde; being near automobile emissions and gasoline was related to higher exposure to acetaldehyde, benzaldehyde, and methylglyoxal. Formaldehyde was associated with perfume and nail polish and remover use. For personal exposures to $PM_{2.5}$, sweeping and woodworking resulted in higher exposure. In addition, several elements in $PM_{2.5}$ were associated with the use of fireplaces (sulfur and vanadium), incense burning (sulfur and potassium), and outdoor cooking (calcium, potassium, and vanadium).

Determinants of Indoor VOC Concentrations

An analysis using LMMs and parallel to that performed for personal samples was conducted for the indoor VOC observations. Given the correlation between indoor and personal exposure observations, it is not surprising that many of the same factors were identified as determinants (Tables 17, 18, and 19). Most of the VOCs were affected by city and several household characteristics. Among household characteristics, AER was negatively associated with indoor levels of toluene, *m- & p*-xylenes, *o*-xylene, styrene, chloroform, *d*-limonene, α-pinene, β-pinene, TCE, and PERC. Larger homes (more rooms) were associated with decreased concentrations of benzene, toluene, *m- & p*-xylenes, *o*-xylene, styrene, 1,4-DCB, *d*-limonene, and α-pinene. BTEX (except for toluene) and MTBE increased with the presence of attached garages. Again, city effect varied by VOC, although Houston had the highest levels of VOCs except for MTBE, TCE, and PERC, which were highest in Elizabeth.

Two meteorologic factors were negatively associated with indoor VOC levels: ambient relative humidity with toluene, ethylbenzene, *m- & p*-xylenes, *o*-xylene, styrene, β-pinene, and chloroform; and wind speed with ethylbenzene, *m- & p*-xylenes, *o*-xylene, MTBE, styrene, and PERC. Higher wind speed was expected to dilute outdoor concentrations from local sources and to affect AERs as noted earlier. Outdoor relative humidity may be a surrogate for seasonal effects and weather (e.g., precipitation), possibly representing the effects of fronts or low pressure systems with good dispersion or effective cleansing. In contrast, indoor temperature showed the opposite effects on two indoor VOCs, benzene and chloroform: higher indoor temperatures were associated with lower levels of benzene and higher levels of chloroform. This may reflect higher volatilization rates for chloroform, which is due largely to indoor sources.

Table 17. Results of Linear Mixed-Effects Models for Indoor Levels of Gasoline-Related VOCs[a]

Variable	Group or Unit of Change[b]	Benzene β	Benzene 95% CI	Toluene β	Toluene 95% CI	Ethylbenzene β	Ethylbenzene 95% CI	*m*- & *p*-Xylenes β	*m*- & *p*-Xylenes 95% CI	*o*-Xylene β	*o*-Xylene 95% CI	MTBE β	MTBE 95% CI	Styrene β	Styrene 95% CI
Intercept		2.57	1.79, 3.35	3.88	2.86, 4.90	0.58	−0.36, 1.52	2.65	1.69, 3.61	1.47	0.59, 2.35	1.46	0.73, 2.19	1.10	0.22, 1.98
Visit	1	−0.22	−0.38, −0.06	0.26	0.10, 0.42	−0.09	−0.27, 0.09	0.07	−0.11, 0.25	0.11	−0.05, 0.27	−0.10	−0.30, 0.10	0.12	−0.06, 0.30
	2	Reference		Reference		Reference		Reference		Reference		Reference		Reference	
City	CA	−0.52	−0.76, −0.28	−0.18	−0.43, 0.07	−0.49	−0.78, −0.20	−0.22	−0.53, 0.09	−0.03	−0.30, 0.24	−0.22	−0.53, 0.09	−0.28	−0.53, −0.03
	NJ	−0.81	−1.06, −0.56	−0.09	−0.34, 0.16	−0.30	−0.67, 0.07	−0.22	−0.63, 0.19	−0.21	−0.58, 0.16	0.18	−0.29, 0.65	−0.01	−0.28, 0.26
	TX	Reference		Reference		Reference		Reference		Reference		Reference		Reference	
Ambient relative humidity	%			−0.01	−0.02, 0.00	−0.01	−0.02, 0.00	−0.01	−0.02, 0.00	−0.01	−0.02, 0.00			−0.02	−0.03, −0.01
Attached garage	No	−0.23	−0.41, −0.05			−0.38	−0.62, −0.14	−0.37	−0.59, −0.15	−0.39	−0.59, −0.19	−0.62	−0.87, −0.37	0.28	0.04, 0.52
Cement and other flooring	No			0.22	0.00, 0.44										
Central heat	No			−0.12	−0.30, 0.06										
Education	< High school	0.34	0.09, 0.59												
	High school	0.03	−0.17, 0.23												
	> College	Reference													
Ethnicity	White					−0.23	−0.52, 0.06	−0.19	−0.52, 0.14	−0.14	−0.43, 0.15	−0.04	−0.39, 0.31		
	Mexican					0.19	−0.16, 0.54	0.19	−0.20, 0.58	0.24	−0.11, 0.59	0.53	0.12, 0.94		
	Hispanic					0.06	−0.31, 0.43	0.16	−0.25, 0.57	0.27	−0.10, 0.64	0.21	−0.22, 0.64		
	Other					Reference		Reference		Reference		Reference			
Heating fuel	Electricity					0.47	0.04, 0.90								
	Gas					0.48	0.09, 0.87								
	Oil and wood					Reference									
Indoor temperature	°C	−0.04	−0.06, −0.02	−0.02	−0.04, 0.00										
Inverse wind speed	1/knot					3.21	1.84. 4.58	3.06	1.59, 4.53	2.92	1.57, 4.27	6.13	4.52, 7.74	2.74	1.35, 4.13
Log-transformed AER	1/hour			−0.34	−0.44, −0.24			−0.19	−0.31, −0.07	−0.18	−0.30, −0.06			−0.20	−0.32, −0.08
Number of floors												−0.13	−0.23, −0.03		
Number of rooms		−0.10	−0.16, −0.04	−0.06	−0.12, 0.00			−0.08	−0.14, −0.02	−0.09	−0.15, −0.03			−0.09	−0.15, −0.03
Open doors or windows	No							0.20	−0.02, 0.42	0.23	0.03, 0.43			0.18	−0.04, 0.40
Professional cleaning	No	0.19	−0.03, 0.41			0.20	−0.04, 0.44								
Time spent indoors at home	min			0.00	−0.0003, 0									0.00	−0.0004, 0
Type of building	Single-family home											0.15	−0.14, 0.44		
	Mobile home											−0.31	−0.80, 0.18		
	Apartment or townhouse											Reference			
Unemployed	No	0.17	−0.03, 0.37			0.27	0.07, 0.47								
Use of candles or incense	No							−0.25	−0.45, −0.05	−0.22	−0.40, −0.04				
Using air cleaning devices	No					−0.41	−0.74, −0.08	−0.61	−0.98, −0.24	−0.50	−0.83, −0.17	−0.40	−0.81, 0.01		

[a] n = 387 to 455 observations depending on the model.

[b] For dichotomous variables, the reference group is “Yes”.

.

Table 18. Results of Linear Mixed-Effects Models for Indoor Levels of Odorant and Cleaning-Related VOCs[a]

Variable	Group or Unit of Change[b]	1,4-DCB		Chloroform		*d*-Limonene		α-Pinene		β-Pinene	
		β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI
Intercept		4.22	2.45, 5.98	−0.41	−1.47, 0.65	3.57	2.81, 4.33	2.48	2.07, 2.89	1.59	0.75, 2.43
Visit	1	0.32	0.02, 0.62	−0.04	−0.24, 0.16	−0.11	−0.40, 0.18	0.12	−0.02, 0.26	0.12	−0.06, 0.30
	2	Reference		Reference		Reference		Reference		Reference	
City	CA	−0.97	−1.58, −0.37	−0.26	−0.55, 0.03	−0.46	−0.83, −0.09	−0.47	−0.71, −0.23	−1.01	−1.30, −0.72
	NJ	−0.61	−1.42, 0.20	−0.22	−0.53, 0.09	−0.53	−1.10, 0.04	−0.62	−0.87, −0.37	−1.20	−1.51, −0.89
	TX	Reference		Reference		Reference		Reference		Reference	
Ambient relative humidity	%			−0.01	−0.02, 0.00					−0.01	−0.02, 0.00
Ethnicity	White	−0.42	−1.19, 0.36			0.19	−0.26, 0.64				
	Mexican	0.44	−0.46, 1.34			0.78	0.25, 1.31				
	Hispanic	−0.29	−1.17, 0.60			0.06	−0.53, 0.65				
	Other	Reference				Reference					
Furniture or floor was waxed or polished	No	−0.91	−1.58, −0.24								
Furniture refinisher in neighborhood	No	−1.39	−2.43, −0.35								
Indoor temperature	°C			0.05	0.01, 0.09						
Keeping dogs or cats	No									0.35	0.13, 0.57
Log-transformed AER	1/hour			−0.54	−0.66, −0.42	−0.43	−0.61, −0.25	−0.46	−0.56, −0.36	−0.34	−0.46, −0.22
Not using fresheners	No									0.37	0.02, 0.72
Number of rooms		−0.14	−0.28, 0.00			−0.13	−0.21, −0.05	−0.07	−0.13, −0.01		
Open doors or windows	No			0.16	−0.08, 0.40					0.18	−0.06, 0.42
Other family members took showers	No			−0.40	−0.65, −0.15	−0.76	−1.13, −0.39	−0.55	−0.77, −0.33	−0.34	−0.61, −0.07
Outdoor temperature when heating starts	< 64°F	0.53	−0.04, 1.09								
	64 to 70°F	−0.04	−0.54, 0.46								
	> 70°F	Reference									
Ownership of the home	No			0.59	0.35, 0.83						
Pets indoors	No			0.30	0.08, 0.52						
Renovation in the past year	No					−0.34	−0.63, −0.05				
Spending awake time at 1st floor	Yes							−0.39	−0.63, −0.15		
Using a clothes washer	No	0.68	0.28, 1.09								
Using central air conditioning	No					−0.44	−0.77, −0.11	−0.62	−0.84, −0.40	−0.27	−0.51, −0.03
Using cleaning solutions	No			−0.20	−0.40, 0.00						
Using dishwashers	No			−0.34	−0.58, −0.10						
Using mothballs	No	−0.40	−1.02, 0.21								

[a] n = 409 to 494 observations depending on the model.

[b] For dichotomous variables, the reference group is "Yes".

Table 19. Results of Linear Mixed-Effects Models for Indoor Levels of Dry-Cleaning and Industrial-Related VOCs[a]

Variable	Group or Unit of Change[b]	TCE		PERC		CTC	
		β	95% CI	β	95% CI	β	95% CI
Intercept		−0.88	−1.45, −0.30	−1.99	−2.48, −1.50	−0.70	−0.90, −0.50
Visit	1	0.19	0.06, 0.32	0.05	−0.13, 0.23	0.05	−0.05, 0.15
	2	Reference		Reference		Reference	
City	CA	0.71	0.46, 0.96	0.98	0.65, 1.31	−0.06	−0.16, 0.04
	NJ	1.10	0.85, 1.34	1.20	0.89, 1.51	−0.11	−0.23, 0.01
	TX	Reference		Reference		Reference	
Cooking	No			0.20	0.02, 0.38		
Having a fireplace	No					0.11	0.01, 0.21
Inverse wind speed	1/knot			4.00	2.47, 5.53		
Log-transformed AER	1/hour	−0.17	−0.27, −0.07	−0.30	−0.42, −0.18		
Professional cleaning	No			−0.28	−0.53, −0.03		
Source of household water	Public	−0.49	−0.94, −0.04				
Sweeping indoors	No			0.16	−0.04, 0.36		
Unemployed	No			0.24	0.00, 0.48		
Using central air conditioning	No					−0.11	−0.21, −0.01
Using other heaters	No	−0.34	−0.61, −0.07			0.15	−0.01, 0.31
Using nail polish remover	No	−0.31	−0.60, −0.02				
Vacuuming	No			0.26	0.06, 0.46	0.12	0.04, 0.20
Vinyl, asbestos or other siding	No	−0.22	−0.44, 0.00	0.38	0.13, 0.63		
Visited dry cleaners	No			−0.34	−0.61, −0.07		

[a] n = 400 to 472 observations depending on the model.

[b] For dichotomous variables, the reference group is "Yes".

Determinants of Outdoor VOC Concentrations

Outdoor concentrations were affected by city and three meteorologic variables (Tables 20, 21, and 22). Ambient relative humidity was negatively associated with concentrations of benzene, ethylbenzene, *m- & p*-xylenes, *o*-xylene, MTBE, styrene, and β-pinene levels. Wind speed was negatively associated with concentrations of benzene, toluene, ethylbenzene, *m- & p*-xylenes, *o*-xylene, MTBE, styrene, α-pinene, TCE, and PERC. Effects of city and outdoor temperature depended on the VOC. For example, Houston had the highest concentrations for benzene, *m- & p*-xylenes and β-pinene, which may be due to the crowded petrochemical industry (Weisel et al. 2005b).

Additional neighborhood-scale determinants seem both possible and likely (e.g., proximity to emission sources). However, these could not be distinguished, possibly due to the low detection frequencies of most of the outdoor VOCs.

Common Determinants of Personal, Indoor, and Outdoor Concentrations

Two factors affected personal, indoor, and outdoor levels: city and wind speed. Three factors affected both personal and indoor levels: AER, number of rooms, and attached garage. The fact that five common determinants affected concentrations of most personal and indoor VOC observations indicates the importance of indoor (at home) exposures and reflects the large amount of time RIOPA participants spent indoors at home. It is important to note that exposures in the home environment arise from both indoor and outdoor VOC sources, and the source-oriented exposure fractions indicated that outdoor sources were responsible for most exposure to all but four VOCs when 100% penetration efficiency for outdoor VOCs entering a residence was assumed. Thus, although exposure occurs mostly at home, both indoor and outdoor sources are important contributors (Sexton et al. 2007).

Table 20. Results of Linear Mixed-Effects Models for Outdoor Levels of Gasoline-Related VOCs[a]

Variable	Group or Unit of Change[b]	Benzene		Toluene		Ethylbenzene		*m,p*-Xylene		*o*-Xylene		MTBE		Styrene	
		β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI
Intercept		0.39	−0.37, 1.15	1.17	0.86, 1.48	−0.13	−0.93, 0.67	−0.04	−0.90, 0.82	−0.05	−0.66, 0.56	2.13	1.27, 2.99	−1.24	−1.68, −0.80
Visit	1	−0.03	−0.15, 0.09	0.26	0.14, 0.38	−0.03	−0.17, 0.11	0.06	−0.08, 0.20	0.09	−0.05, 0.23	−0.06	−0.24, 0.12	−0.03	−0.13, 0.07
	2	Reference		Reference		Reference		Reference		Reference		Reference		Reference	
City	CA	−0.56	−0.76, −0.36	0.06	−0.12 0.24	0.01	−0.19, 0.21	−0.25	−0.50, 0.00	−0.03	−0.25, 0.19	0.02	−0.25, 0.29	0.56	0.40, 0.71
	NJ	−0.76	−1.01, −0.51	−0.16	−0.36, 0.04	0.05	−0.22, 0.32	−0.19	−0.48, 0.10	−0.05	−0.32, 0.22	−0.09	−0.44, 0.26	0.74	0.52, 0.95
	TX	Reference		Reference		Reference		Reference		Reference		Reference		Reference	
Ambient relative humidity	%	−0.01	−0.02, −0.01			−0.02	−0.02, −0.01	−0.01	−0.02, −0.01	−0.02	−0.02, −0.01	−0.02	−0.03, −0.01	−0.01	−0.01, 0.00
Attached garage	No			−0.15	−0.29, −0.01										
Cooking	No													−0.09	−0.19, 0.01
Crawl space	No	−0.17	−0.35, 0.01												
Ethnicity	White	−0.20	−0.42, 0.02			−0.21	−0.43, 0.01	−0.16	−0.41, 0.09	−0.14	−0.36, 0.08	−0.14	−0.43, 0.15	−0.03	−0.19, 0.13
	Mexican	0.10	−0.15, 0.35			0.24	−0.01, 0.49	0.09	−0.20, 0.38	0.15	−0.10, 0.40	0.43	0.10, 0.76	0.27	0.07, 0.47
	Hispanic	−0.02	−0.27, 0.23			0.08	−0.19, 0.35	0.03	−0.26, 0.32	0.08	−0.19, 0.35	−0.02	−0.37, 0.33	−0.07	−0.27, 0.13
	Other	Reference				Reference		Reference		Reference		Reference		Reference	
Foundation of slab	No			0.15	0.03, 0.27										
Gardening	No	0.17	0.01, 0.33												
Home volume	m^3							0.00	−0.0012 −0.0002	0.00	−0.001, 0.0				
Inverse wind speed	1/knot	4.18	3.20, 5.16	1.91	0.95, 2.87	4.69	3.63, 5.75	5.65	4.51, 6.79	5.50	4.44, 6.56	5.63	4.18, 7.08	3.10	2.32, 3.88
Near diesel vehicles	No	−0.20	−0.32, −0.08												
Pets	No			−0.27	−0.41, −0.13										
Number of floors										−0.08	−0.14, −0.02				
Number of rooms						−0.10	−0.14, −0.06								
Open doors or windows	No													−0.14	−0.26, −0.02
Other family members took showers	No			−0.38	−0.54, −0.22										
Outdoor temperature	Q1	0.39	0.21, 0.57	0.26	0.08, 0.44	0.22	0.02, 0.42	0.29	0.09, 0.49	0.23	0.03, 0.43	−0.03	−0.28, 0.22		
	Q2	0.33	0.15, 0.51	0.27	0.11, 0.43	0.26	0.08, 0.44	0.22	0.02, 0.42	0.22	0.04, 0.40	0.24	−0.01, 0.49		
	Q3	−0.01	−0.19, 0.17	−0.10	−0.26, 0.06	0.08	−0.10, 0.26	0.02	−0.18, 0.22	−0.02	−0.20, 0.16	0.01	−0.24, 0.26		
	Q4	Reference		Reference		Reference		Reference		Reference		Reference			
Ownership of the home	No							0.21	0.03, 0.39	0.23	0.07, 0.39				
Pets indoors	No							0.14	−0.02, 0.30	0.15	0.01, 0.29				
Professional cleaning	No			0.14	−0.02, 0.30										
Tobacco products smoked in home	No	0.66	0.15, 1.17			0.64	0.11, 1.17	0.79	0.22, 1.36						
Type of building	Single-family home													−0.08	−0.20, 0.04
	Mobile home													0.19	−0.01, 0.39
	Apartment or townhouse													Reference	
Unvented appliances in basement	No											−0.46	−0.85, −0.07		

[a] n = 439 to 457 observations depending on the model.

[b] For dichotomous variables, the reference group is "Yes".

Table 21. Results of Linear Mixed-Effects Models for Outdoor Levels of Odorant and Cleaning-Related VOCs[a]

Variable	Group or Unit of Change[b]	1,4-DCB		Chloroform		*d*-Limonene		α-Pinene		β-Pinene	
		β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI
Intercept		−0.50	−0.96, −0.05	−1.63	−1.93, −1.32	−0.05	−0.49, 0.39	−1.27	−1.62, −0.92	0.26	−0.02, 0.55
Visit	1	0.03	−0.15, 0.20	0.05	−0.05, 0.16	0.11	−0.06, 0.27	0.01	−0.13, 0.15	0.00	−0.08, 0.07
	2	Reference		Reference		Reference		Reference		Reference	
City	CA	0.80	0.57, 1.03	0.44	0.28, 0.60	0.98	0.74, 1.21	1.43	1.23, 1.63	−0.60	−0.69, −0.50
	NJ	0.63	0.34, 0.93	0.37	0.16, 0.59	0.55	0.27, 0.83	1.38	1.15, 1.60	−0.77	−0.89, −0.64
	TX	Reference		Reference		Reference		Reference		Reference	
Air conditioning	No	0.35	0.15, 0.55								
Ambient relative humidity	%									0.00	−0.01, 0.00
Attached garage	No			−0.20	−0.34, −0.06						
Cement and other flooring	No			0.16	0.02, 0.30						
Detached garage or carport	No									0.08	0.00, 0.16
Ethnicity	White			0.12	−0.06, 0.30						
	Mexican			0.27	0.07, 0.47						
	Hispanic			0.08	−0.14, 0.30						
	Other			Reference							
Furniture refinisher in neighborhood	No	−0.77	−1.18, −0.36								
Inverse wind speed	1/knot							1.56	0.50, 2.62		
Pets	No			−0.17	−0.29, −0.05	−0.20	−0.40, 0.00			−0.09	−0.17, −0.01
Not using fresheners	No			−0.17	−0.35, 0.01	−0.28	−0.55, −0.01				
Number of floors				0.08	0.04, 0.12	0.09	0.01, 0.17				
Outdoor temperature	Q1	−0.36	−0.61, −0.11	−0.21	−0.35, −0.07	−0.47	−0.71, −0.23	−0.23	−0.43, −0.03		
	Q2	−0.32	−0.56, −0.08	−0.17	−0.31, −0.03	−0.52	−0.74, −0.30	−0.29	−0.47, −0.11		
	Q3	0.03	−0.21, 0.27	−0.10	−0.24, 0.04	−0.21	−0.43, 0.01	−0.13	−0.31, 0.05		
	Q4	Reference		Reference		Reference		Reference			
Outdoor temperature	°C									0.01	0.00, 0.01
Professional cleaning	No					−0.31	−0.53, −0.09	−0.20	−0.38, −0.02		
Type of building	Single-family home							−0.12	−0.28, 0.04		
	Mobile home							0.11	−0.13, 0.35		
	Apartment or townhouse							Reference			
Using mothballs	No	0.15	−0.03, 0.33								
Using cloth dryers	No							0.13	−0.01, 0.27		
Wore any powder, hair spray, or perfume	No									0.07	−0.01, 0.15

[a] n = 437 to 470 observations depending on the model.

[b] For dichotomous variables, the reference group is "Yes".

.

Table 22. Results of Linear Mixed-Effects Models for Outdoor Levels of Dry-Cleaning and Industrial-Related VOCs[a]

Variable	Group or Unit of Change[b]	TCE		PERC		CTC	
		β	95% CI	β	95% CI	β	95% CI
Intercept		−1.86	−2.13, −1.59	−2.26	−2.65, −1.87	−0.29	−0.45, −0.13
Visit	1	0.14	0.06, 0.22	−0.03	−0.19, 0.13	0.06	0.00, 0.12
	2		Reference		Reference		Reference
City	CA	0.46	0.34, 0.58	1.40	1.18, 1.62	−0.01	−0.11, 0.09
	NJ	0.80	0.66, 0.94	1.20	0.96, 1.44	−0.08	−0.20, 0.04
	TX		Reference		Reference		Reference
Dry cleaners in neighborhood	No			−0.16	−0.32, 0.00		
Inverse wind speed	1/knot	0.74	0.09, 1.39	4.63	3.47, 5.79		
Pets	No	−0.12	−0.22, −0.02	−0.22	−0.40, −0.04		
Not using fresheners	No					−0.15	−0.27, −0.03
Number of carpeted rooms				−0.05	−0.09, −0.01		
Number of floors		−0.03	−0.07, 0.01				
Open doors or windows	No	−0.08	−0.18, 0.02			0.06	−0.02, 0.14
Outdoor temperature	Q1	0.13	0.01, 0.25	0.15	−0.07, 0.37		
	Q2	0.17	0.05, 0.29	0.33	0.13, 0.53		
	Q3	0.03	−0.09, 0.15	0.00	−0.19, 0.20		
	Q4		Reference		Reference		
Type of building	Single-family home					−0.12	−0.20, −0.04
	Mobile home					−0.12	−0.24, 0.00
	Apartment or townhouse						Reference
Unvented appliances in basement	No	−0.23	−0.41, −0.05				
Vacuuming	No			0.16	0.02, 0.30		

[a] n = 402 to 461 observations depending on the model.

[b] For dichotomous variables, the reference group is "Yes".

Other determinants (not measured in this study) might affect exposure to several VOCs, especially those in common VOC mixtures (e.g., BTEX from gasoline-related sources). In addition, many VOCs arise from different sources that have different determinants other than the five common ones identified above (city, wind speed, AER, number of rooms, and attached garage). These five determinants either act on indoor concentrations from indoor sources (e.g., AERs and home size) or affect outdoor concentrations (and subsequently indoor levels) from neighborhood or urban VOC sources (e.g., city and wind speed).

The magnitude of coefficient estimates for determinants across sample types (personal, indoor, outdoor) was generally consistent with the exposure source. For example, personal and indoor concentrations of benzene were negatively associated with the presence of an attached garage (β = −0.19 for personal and −0.23 for indoor; Tables 14 and 17); the difference between these values could reflect, for example, that participants were not home all day. The same pattern was observed for personal and indoor 1,4-DCB concentrations for waxing/polishing furniture (β = −0.81 and −0.91; Tables 15 and 18), and the same explanation

may apply. In contrast, exposures related to personal activities (e.g., exposure to PERC and visiting dry-cleaners) often showed larger coefficients (absolute value) for personal compared with indoor exposures ($\beta = -0.63$ and -0.34; Tables 16 and 19).

Assumption of Linearity

The assumption of linearity for the continuous covariates in the LMMs (wind speed, ambient relative humidity, indoor temperature, AER, and time spent indoors at home) was evaluated using partial residual plots, which accounted for the effects of all other covariates. Plots for wind speed and AER suggested some nonlinearities with log-transformed VOC concentrations (panels A, C, and E in Appendix Figure A.15). Several transformations of these variables were attempted, and near-linear relationships were achieved using the reciprocal of wind speed and the logarithm of AER (panels B, D, and F in Appendix Figure A.15). Inverse wind speed can be supported based on dilution or mass balance principles, which apply to sources with emission rates that are independent of the wind speed. For buildings with internal emission sources, the AER is proportional to the air flow through the building, so again the reciprocal of the AER was expected to be linearly related to indoor concentrations. However, indoor concentrations are affected by many factors and AERs are measured with error. The log AER, rather than 1/AER, would tend to diminish the effect of both very large and very small AERs, and the fit with this transformation suggests that the measured AERs included some outliers. Still, the expected relationship was seen; that is, indoor concentrations of VOCs with strong indoor sources (e.g., chloroform and *d*-limonene) decreased as AERs increased (Table 18).

Model Validation

The estimated fraction of variance (R^2) attributable to fixed-effect variables in the LMMs for each VOC and each sample type (personal, indoor, outdoor) is shown in Appendix Table A.22. For personal exposures, R^2 ranged from 0.003 (CTC) to 0.40 (β-pinene); for indoor observations, R^2 ranged from 0.09 (toluene) to 0.42 (PERC); and for outdoor concentrations, R^2 values ranged from 0.17 (1,4-DCB) to 0.65 (PERC). Generally, more of the variance for the outdoor observations was explained. In all three sample types, VOCs with specific emission sources (e.g., PERC from dry cleaners and α-pinene from cleaning products and fresheners) had the largest R^2 among the 15 VOCs. In contrast, VOCs used in many commercial products that were also components of exhaust and from other sources (e.g., toluene) had small R^2 across the three sample types. The LMMs explained only a portion of the variance in the data set. Some of the variance is random and some is due to errors in measurement and model specifications; therefore, it is likely that the LMMs are incomplete models in the sense that other (unknown) variables and other (also unknown) interactions among the variables affect exposure. Further, effects of short-term activities (e.g., cooking) may not be observable with 48-hour integrated observations. However, low R^2 values do not invalidate the identification or significance of the determinants identified in this study.

Limitations

This analysis of exposure determinants has several limitations. First, missing data in the RIOPA data set decreased the sample size and reduced statistical power. To address this issue, we excluded variables with sample sizes below 400 (i.e., with over 150 missing observations) from the LMMs, and evaluated the use of MI. For the nine models tested, MI results showed that the impact of missing data would probably not be substantial.

Second, measurements below MDLs might somehow distort results. For the indoor and personal observations, styrene and TCE had detection frequencies below 50%; for the outdoor observations, toluene, styrene, TCE, 1,4-DCB, chloroform, *d*-limonene, and the both pinenes had low detection frequencies (see Table 1). As a result, determinants of the outdoor VOCs should be interpreted cautiously.

A third issue is the nature of the RIOPA sample. As mentioned earlier, RIOPA observations were taken in three U.S. cities that have specific and somewhat different emission sources; homes near outdoor emission sources were oversampled (Weisel et al. 2005b); and a convenience sample of participants was selected. Because a convenience sample is not (nationally) representative, the results may not be broadly applicable. However, many of our findings are paralleled in studies that used regional or national data sets, and thus most results appear relevant.

CONCLUSIONS

This project drew on the outdoor, indoor, and personal VOC observations from two large data sets, RIOPA and NHANES, and used several novel and powerful statistical modeling and analysis techniques to identify and characterize exposure distributions, risks, mixtures, dependencies of the components in mixtures, and exposure determinants.

The results of the extreme value analyses showed that the highest exposures in RIOPA, which can be the most significant in terms of health risks, closely fitted GEV distributions and, in many cases, Gumbel distributions (a

reduced form of the GEV distribution). In contrast, lognormal distributions, the usual distributional assumption, underestimated concentrations and risks from extreme values. Despite the importance of extreme value exposures, few studies have fitted distributions or otherwise characterized such extrema. Better ways to accurately characterize pollutant distributions and to predict the number of individuals that would be expected to exceed risk-based exposure guidelines or other criteria are needed. Extreme value distributions will be useful in impact and policy analyses to describe concentrations, exposures, and risks.

Although extreme value distributions can represent the upper tail of exposure and risk distributions, they do not fit the full distribution of most environmental data, which can have multiple modes, extreme values, left-censoring, and other features. Compared with parametric distributions, the finite mixture and DPM distributions were shown to have superior performance in fitting VOC exposure data with extreme values or with a large fraction of data below the MDLs. The optimal number of distributions (K) needed for the finite mixture models ranged from 2 to 4, depending on VOCs. Distributions from the DPM model provided slightly better fits than those from the finite mixture model. The DPM model has advantages by characterizing uncertainty around the number of components and by providing a formal assessment of uncertainty for all model parameters through the posterior distribution. The method adapts to a spectrum of departures from standard model assumptions and provides robust estimates of the exposure density, even with left censoring (due to the MDL).

Many VOCs have similar emission sources and toxicologic effects, which highlights the need to understand and evaluate exposure to mixtures. VOC mixtures in the RIOPA data set were identified using PMF analyses and toxicologic modes of action. The VOC emission sources identified using PMF included gasoline vapor (mixture A1), vehicle exhaust (mixture A2), moth repellents, chlorinated solvents, and water disinfection by-products (mixture A3), and cleaning products and odorants (mixture A4). These four mixtures were affected by city, ethnicity, and AERs. The influence of environmental factors and personal activities was also shown for certain mixtures; for example, mixture A1 was associated with attached garages and self-service pumping gas. Three additional mixtures based on cancer endpoints were identified for liver and renal tumors (mixtures B1 and A3/B3) and hematopoietic cancers (mixture B2).

Dependencies among mixture components were described using copulas, which showed a high degree of accuracy and flexibility, including the ability to represent asymmetrical dependency structures. The dependency structures of four mixtures in RIOPA were best described by the t copula; two other mixtures best fitted Gumbel copulas, which capture dependency structures of distributions containing extreme values. In all cases, the copulas clearly provided better fits than multivariate lognormal distributions. Copulas can provide accurate estimates and simulations for the joint distribution of pollutants across the full range of concentrations, and they faithfully represent the correlation in the tails of the distributions. Thus, copulas may be the method of choice for estimating cumulative risks of exposure to mixtures, particularly for the highest exposures or extreme events, which fit lognormal distributions poorly and may represent the greatest risk.

LMMs were used to identify determinants of VOC exposures in the RIOPA data set. Determinants of personal exposures of VOCs included city, personal activities (e.g., pumping gas and visiting dry cleaners), household characteristics (e.g., AERs, number of rooms, attached garages), and meteorology (e.g., wind speed). Similar determinants were found for indoor concentrations. Most of these determinants were consistent with previous studies (e.g., BTEX and attached garages, and PERC and visiting dry cleaners). Several new determinants were identified, including city, other family member showering, and residence size. With the exception of four VOCs with strong indoor sources, most exposure resulted from outdoor sources. Further investigation using a more representative population and a wider suite of VOCs would extend and generalize these results.

IMPLICATIONS OF THE FINDINGS

This report highlights several critical issues in exposure science relevant to public health that have received relatively little attention. These issues were addressed using several advanced statistical approaches and the RIOPA VOC data set. These methods performed well, and they deserve more widespread consideration and application.

First, the highest exposure events do not fit default distributional assumptions, such as lognormal distributions, but they can be described using extreme value analyses. Since the highest exposures may be the ones most relevant to health risks, they frequently become the determinants or drivers of environmental decisions and policies. We suggest that these high concentrations and exposures need to be more accurately characterized and modeled, possibly using the extreme value theory to establish exposure guidelines and to estimate risks across a population.

Second, single (parametric) distributions may not accurately fit exposure data, which contain features such as multiple modes, extreme values, and left censoring. The finite mixture and DPM models, provided much better distribution fits to the RIOPA VOC data set than did lognormal distributions. These full-distribution models offer several

advantages over parametric distribution models, and they appear appropriate for other types of environmental data (e.g., persistent or emerging compounds). The use of mixture models can improve the accuracy and realism of models used in a variety of exposure and risk applications and further environmental applications are warranted.

Third, copulas were used to estimate dependency structures in mixtures of VOCs. The RIOPA data set showed complex dependencies; for example, the dominant VOC in a mixture often changed as the mixture concentration increased. Copula methods have many strengths: they overcome shortcomings of traditional methods that address only pair-wise correlations (e.g., correlation coefficients); allow the use of any marginal distribution; permit asymmetrical dependency structures; and decouple the dependency structure from the marginal distribution. These are essential considerations for cumulative exposure and risk assessment and copulas are a powerful tool in this application, especially for high-concentration mixtures that may pose the greatest risks.

Fourth, the analysis of exposure determinants in this report suggests several interventions that can help prevent or reduce VOC exposures. Since participants in the RIOPA study spent over 90% of their time at home, and since the home compartment contributes an average of 60% of an individual's total VOC exposure, minimizing indoor VOC levels will decrease exposure. In addition, VOC exposures can be reduced by modifying activities that contribute significantly to VOC exposure (e.g., pumping gasoline and visiting dry cleaners) and by addressing environmental factors that influence VOC exposures (e.g., attached garages, outdoor VOC sources).

ACKNOWLEDGMENTS

The authors would like to thank Geoffrey Sunshine, Hanna Boogaard, Virgi Hepner, and the HEI Review Committee for their assistance in producing this report.

REFERENCES

Aberg S, Guttorp P. 2008. Distribution of the maximum in air pollution fields. Environmetrics 19:183–208.

Agency for Toxic Substances and Disease Registry. 1997a. Toxicological profile for chloroform. Atlanta, GA:ATSDR. Available: *www.atsdr.cdc.gov/ToxProfiles/tp.asp?id=53&tid=16.*

Agency for Toxic Substances and Disease Registry. 1997b. Toxicological profile for tetrachloroethylene. Atlanta, GA: ATSDR. Available: *www.atsdr.cdc.gov/ToxProfiles/tp.asp?id=265&tid=48.*

Agency for Toxic Substances and Disease Registry. 1997c. Toxicological profile for trichloroethylene. Atlanta, GA:ATSDR. Available: *www.atsdr.cdc.gov/toxprofiles/tp.asp?id=173&tid=30.*

Agency for Toxic Substances and Disease Registry. 2004. Guidance manual for the assessment of joint toxic action of chemical mixtures. Atlanta, GA:ATSDR. Available: *www.atsdr.cdc.gov/interactionprofiles/ipga.html*

Agency for Toxic Substances and Disease Registry. 2005a. Toxicological profile for carbon tetrachloride. Atlanta, GA:ATSDR. Available: *www.atsdr.cdc.gov/toxprofiles/tp30.pdf.*

Agency for Toxic Substances and Disease Registry. 2005b. Toxicological profile for naphthalene, 1-methylnaphthalene, and 2-methylnaphthalene. Atlanta, GA:ATSDR. Available: *www.atsdr.cdc.gov/toxprofiles/ tp67.pdf.*

Agency for Toxic Substances and Disease Registry. 2006. Toxicological profile for dichlorobenzenes. Atlanta, GA:ATSDR. Available: *www.atsdr.cdc.gov/toxprofiles/tp.asp?id=704&tid=126.*

Agency for Toxic Substances and Disease Registry. 2007. Toxicological profile for xylene. Atlanta, GA:ATSDR. Available: *www.atsdr.cdc.gov/toxprofiles/tp71.pdf.*

Agency for Toxic Substances and Disease Registry. 2009. Toxicological profile for 1,3-butadiene. Atlanta, GA:ATSDR. Available: *www.atsdr.cdc.gov/ToxProfiles/tp28.pdf.*

Agency for Toxic Substances and Disease Registry. 2010a. Toxicological profile for styrene. Atlanta, GA:ATSDR. Available: *www.atsdr.cdc.gov/PHS/PHS.asp?id=419&tid=74.*

Agency for Toxic Substances and Disease Registry. 2010b. Toxicological profile for ethylbenzene. Atlanta, GA:ATSDR. Available: *www.atsdr.cdc.gov/toxprofiles/tp110.pdf.*

American Conference of Governmental Industrial Hygienists. 2012. TLVs and BEIs. Cincinnati:ACGIH.

Anderson MJ, Miller SL, Milford JB. 2001. Source apportionment of exposure to toxic volatile organic compounds using positive matrix factorization. J Expo Anal Environ Epidemiol 11:295–307.

Ang A, Tang WH. 1975. Probability Concepts in Engineering Planning and Design: Decision, Risk and Reliability. New York:John Wiley & Sons.

Antoniak CE. 1974. Mixtures of Dirichlet processes with applications to nonparametric problems. Ann Stat 2:1152–1174.

Antweiler R, Taylor H. 2008. Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets: I. Summary statistics. Environ Sci Technol 42:3732–3738.

Backer LC, Egeland GM, Ashley DL, Lawryk NJ, Weisel CP, White MC, et al. 1997. Exposure to regular gasoline and ethanol oxyfuel during refueling in Alaska. Environ Health Perspect 105:850–855.

Batterman S, Chin JY, Jia C, Godwin C, Parker E, Robins T, et al. 2012. Sources, concentrations, and risks of naphthalene in indoor and outdoor air. Indoor Air 22:266–278.

Batterman S, Jia C, Hatzivasilis G. 2007. Migration of volatile organic compounds from attached garages to residences: a major exposure source. Environ Res 104:224–240.

Batterman S, Su FC, Jia C, Naidoo RN, Robins T, Naik I. 2011. Manganese and lead in children's blood and airborne particulate matter in Durban, South Africa. Sci Total Environ 409:1058–1068.

Benaglia T, Chauveau D, Hunter DR, Young DS. 2009. Mixtools: An R package for analyzing mixture models. J Stat Softw 32.

Biernacki C, Celeux C, Govaert G. 2000. Assessing a mixture model for clustering with the integrated completed likelihood. IEEE Trans Pattern Anal Mach Intell 22:719–725.

Blackwell D, MacQueen JB. 1973. Ferguson distributions via Pólya urn schemes. Ann Stat 1:353–355.

Bliss C. 1939. The toxicity of poisons applied jointly. Ann Appl Biol 26:585–615.

Borgert CJ, Quill TF, McCarty LS, Mason AM. 2004. Can mode of action predict mixture toxicity for risk assessment? Toxicol Appl Pharmacol 201:85–96.

Brown SK. 2002. Volatile organic pollutants in new and established buildings in Melbourne, Australia. Indoor Air 12:55–63.

Burmaster DE, Wilson AM. 2000. Fitting second-order finite mixture models to data with many censored values using maximum likelihood estimation. Risk Anal 20:261–272.

Byun H, Ryu K, Jang K, Bae H, Kim D, Shin H, et al. 2010. Socioeconomic and personal behavioral factors affecting children's exposure to VOCs in urban areas in Korea. J Environ Monit 12:524–535.

Caldwell JC, Woodruff TJ, Morello-Frosch R, Axelrad DA. 1998. Application of health information to hazardous air pollutants modeled in EPA's cumulative exposure project. Toxicol Ind Health 14:429–454.

Cherubini U, Luciano E, Vecchiato W. 2004. Copula Methods in Finance. New York:John Wiley & Sons.

Chin JY, Godwin C, Jia C, Robins T, Lewis T, Parker E, et al. 2013. Concentrations and risks of *p*-dichlorobenzene in indoor and outdoor air. Indoor Air 23:40–49.

Chu H, Kensler TW, Muñoz A. 2005. Assessing the effect of interventions in the context of mixture distributions with detection limits. Stat Med 24:2053–2067.

Clobes AL, Ananth GP, Hood AL, Schroeder JA, Lee KA. 1992. Human activities as sources of volatile organic compounds in residential environments. Ann NY Acad Sci 641:79–86.

D'Souza JC, Jia C, Mukherjee B, Batterman S. 2009. Ethnicity, housing and personal factors as determinants of VOC exposures. Atmos Environ 43:2884–2892.

De Rosa CT, Hansen H, Wilbur S, Pohl HR, El-Masri HA, Mumtaz MM. 2001. Interactions. In: Patty's Toxicology. John Wiley & Sons.

Delgado-Saborit JM, Aquilina NJ, Meddings C, Baker S, Harrison RM. 2009. Model development and validation of personal exposure to volatile organic compound concentrations. Environ Health Perspect 117.

Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser B Stat Methodol 39:1–38.

Edwards RD, Jurvelin J, Saarela K, Jantunen M. 2001. VOC concentrations measured in personal samples and residential indoor, outdoor and workplace microenvironments in EXPOLIS–Helsinki, Finland. Atmos Environ 35:4531–4543.

Embrechts P, Klüppelberg C, Mikosch T. 1997. Modelling extremal events for insurance and finance. New York: Springer.

Engeland K, Hisdal H, Frigessi A. 2004. Practical extreme value modelling of hydrological floods and droughts: a case study. Extremes 7:5–30.

Escobar MD. 1994. Estimating normal means with a Dirichlet process prior. J Am Stat Assoc 89:268–277.

Escobar MD, West M. 1995. Bayesian density estimation and inference using mixtures. J Am Stat Assoc 90:577–588.

Ferguson TS. 1973. A Bayesian analysis of some nonparametric problems. Ann Stat 1:209–230.

Ferguson TS. 1983. Bayesian density estimation by mixtures of normal distributions. In: Recent Advances in Statistics (M. Rizvi, J. Rustagi, and D. Siegmund, eds). New York:Academic Press, 287–302.

Feron V, Groten J, van Bladeren P. 1998. Exposure of humans to complex chemical mixtures: hazard identification and risk assessment. Arch Toxicol Suppl 20:363–373.

Finlayson-Pitts BJ, Pitts Jr JN. 2000. Indoor air pollution: sources, levels, chemistry, and fates. In: Chemistry of the Upper and Lower Atmosphere. San Diego:Academic Press, 844–870.

Fisher R, Tippett L. 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. Math Proc Camb Philos Soc 24:180–190.

Fraley C, Raftery AE. 1998. How many clusters? Which clustering method? Answers via model-based cluster analysis. Comp J 41:578–588.

Frees E, Valdez E. 1998. Understanding relationships using copulas. NAAJ 2:1–25.

Gordon S, Brinkman M, Ashley D, Blount B, Lyu C, Masters J, et al. 2006. Changes in breath trihalomethane levels resulting from household water-use activities. Environ Health Perspect 114:514–521.

Gumbel EJ. 1958. Statistics of Extremes. New York: Columbia University Press.

Hammonds JS, Hoffman FO, Bartell SM. 1994. An introductory guide to uncertainty analysis in environmental and health risk assessment. ES/ER/TM-35. Washington, DC:U.S. Environmental Protection Agency.

Hinwood AL, Rodriguez C, Runnion T, Farrar D, Murray F, Horton A, et al. 2007. Risk factors for increased BTEX exposure in four Australian cities. Chemosphere 66:533–541.

Hopke PK, Paatero P. 1994. Extreme-value estimation applied to aerosol size distributions and related environmental problems. J Res Natl Inst Stand Technol 99:361–367.

Huang Y-L, Batterman S. 2003. An extreme value analysis of pollutant concentrations in surface soils due to atmospheric deposition. Hum Ecol Risk Assess 9:1729–1746.

Hüsler J. 1983. Frost data: a case study on extreme values of non-stationary sequences. In: Statistical Extremes and Applications (De Oliveira JT, ed). Boston, MA:D. Reidel Publishing Co.

Hüsler J. 2009. Extreme value analysis in biometrics. Biometrical J 51:252–272.

Hurvich CM, Tsai CL. 1989. Regression and time series model selection in small samples. Biometrika 76:297–307.

International Agency for Research on Cancer. 1993. Some Naturally Occurring Substances: Food Items and Constituents, Heterocyclic Aromatic Amines and Mycotoxins. Lyon, France:World Health Organization, International Agency for Research on Cancer.

International Agency for Research on Cancer. 2012. Agents Classified by the IARC Monographs. Lyon, France:World Health Organization, International Agency for Research on Cancer. Available: *http://monographs.iarc.fr/ENG/Classification/index.php.*

International Agency for Research on Cancer. 2013. Agents Classified by the IARC Monographs. Lyon, France:World Health Organization, International Agency for Research on Cancer. Available: *http://monographs.iarc.fr/ENG/Classification/index.php.*

Jara A. 2007. Applied Bayesian non- and semi-parametric inference using DPpackage. R News 7:17–26.

Jara A, Hanson T, Quintana FA, Müller P, Rosner GL. 2011. DPpackage: Bayesian semi- and nonparametric modeling in R. J Stat Softw 40:1–30.

Jean-Frédéric J, Gaël R, Thierry R. 2004. Financial applications of copula functions. In: Risk Measures for the 21st Century (Szegö G, ed). John Wiley & Sons.

Jenkinson AF. 1955. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Q J Roy Astron Soc 81:158–171.

Jia C, Batterman S, D'Souza J. 2010. Copulas and other multivariate models of personal exposures to VOC mixtures. Hum Ecol Risk Assess 16: 873-900.

Jia C, Batterman SA, Relyea GE. 2012. Variability of indoor and outdoor VOC measurements: an analysis using variance components. Environ Pollut 169:152-159.

Jia C, D'Souza J, Batterman S. 2008. Distributions of personal VOC exposures: a population-based analysis. Environ Int 34:922–931.

Jo W-K, Song K-B. 2001. Exposure to volatile organic compounds for individuals with occupations associated with potential exposure to motor vehicle exhaust and/or gasoline vapor emissions. Sci Total Environ 269:25–37.

Kassomenos P, Lykoudis S, Chaloulakou A. 2010. A tool for determining urban emission characteristics to be used in exposure assessment. Environ Int 36:281–289.

Katz RW, Parlange MB, Naveau P. 2002. Statistics of extremes in hydrology. Adv Water Resour 25:1287–1304.

Kim H, Bernstein J. 2009. Air pollution and allergic disease. Curr Allergy Asthma Rep 9:128–133.

Kim YM, Harrad S, Harrison RM. 2002. Levels and sources of personal inhalation exposure to volatile organic compounds. Environ Sci Technol 36:5405–5410.

Kinney P, Chillrud S, Ramstrom S, Ross J, Spengler J. 2002. Exposures to multiple air toxics in New York City. Environ Health Perspect 110:539–546.

Klepeis NE, Nelson WC, Ott WR, Robinson JP, Tsang AM, Switzer P, et al. 2001. The National Human Activity Pattern Survey (NHAPS): A resource for assessing exposure to environmental pollutants. J Expo Anal Env Epi 11:231–252.

Krishnamoorthy K, Mallick A, Mathew T. 2009. Model-based imputation approach for data analysis in the presence of non-detects. Ann Occup Hyg 53:249–263.

Krueger C, Tian L. 2004. A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. Biol Res Nurs 6:151–157.

Lenox MJ, Haimes YY. 1996. The constrained extremal distribution selection method. Risk Anal 16:161–176.

Ling Z, Guo H, Cheng H, Yu Y. 2011. Sources of ambient volatile organic compounds and their contributions to photochemical ozone formation at a site in the Pearl River Delta, Southern China. Environ Pollut 159:2310–2319.

Lippy BE, Turner RW. 1991. Complex mixtures in industrial workspaces: lessons for indoor air quality evaluations. Environ Health Perspect 95:81–83.

Liu W, Zhang J, Korn LR, Zhang L, Weisel CP, Turpin B, et al. 2007. Predicting personal exposure to airborne carbonyls using residential measurements and time/activity data. Atmos Environ 41:5280–5288.

Lubin J, Colt J, Camann D, Davis S, Cerhan J, Severson R, et al. 2004. Epidemiologic evaluation of measurement data in the presence of detection limits. Environ Health Perspect 112:1691–1696.

Maryland Department of the Environment. 2010. Top Ten Sources of Volatile Organic Compound in the Baltimore Area 2002. Maryland Department of the Environment. Available: *www.mde.state.md.us/Air/air_information/TopTenVOC.asp.*

McCormick NJ. 1981. Reliability and Risk Analysis: Methods and Nuclear Power Applications. San Diego, CA:Academic Press.

McLachlan GJ, Basford KE. 1988. Mixture Models: Inference and Applications to Clustering. New York:M. Dekker.

McLachlan GJ, Krishnan T. 1997. The EM Algorithm and Extensions. New York:John Wiley & Sons.

McLachlan GJ, Peel D. 2000. Finite Mixture Models. New York:John Wiley & Sons.

Mendell MJ. 2007. Indoor residential chemical emissions as risk factors for respiratory and allergic effects in children: a review. Indoor Air 17:259–277.

Meng QY, Spector D, Colome S, Turpin B. 2009. Determinants of indoor and personal exposure to $PM_{2.5}$ of indoor and outdoor origin during the RIOPA study. Atmos Environ 43:5750–5758.

Meng QY, Turpin BJ, Lee JH, Polidori A, Weisel CP, Morandi M, et al. 2007. How does infiltration behavior modify the composition of ambient $PM_{2.5}$ in indoor spaces? An analysis of RIOPA data. Environ Sci Technol 41:7315–7321.

Meng XL, Pedlow S. 1992. EM: a bibliographic review with missing articles. In: Proceedings of the Statistical Computing Section of the American Statistical Association:24–27.

Mudarri DH. 2010. Building Codes and Indoor Air Quality. Available: *www.epa.gov/iaq/pdfs/building_codes_and_iaqpdf.*

Mueller P, Quintana F. 2004. Nonparametric Bayesian data analysis. Stat Sci 19:95–110.

National Center for Health Statistics. 2012a. National Health and Nutrition Examination Survey. 1999–2000 data documentation, codebook, and frequencies: volatile organic compounds (VOC) (Lab21). Hyattsville, MD:U.S. Department of Health and Human Services, Centers for

Disease Control and Prevention. Available: *www.cdc.gov/nchs/nhanes/nhanes1999–2000/LAB21.htm.*

National Center for Health Statistics. 2012b. National Health and Nutrition Examination Survey. 1999–2000 laboratory files. Hyattsville, MD:U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. Available: *www.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Laboratory&CycleBeginYear=1999.*

Office of Environmental Health Hazard Assessment. 2005. Air toxics hot spots program risk assessment guidelines, part II: technical support document for describing available cancer potency factors. Sacramento, CA:California Environmental Protection Agency, OEHHA, Air Toxicology and Epidemiology Section.

Paatero P, Tapper U. 1994. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. Environmetrics 5:111–126.

Paulo MJ, van der Voet H, Wood JC, Marion GR, van Klaveren JD. 2006. Analysis of multivariate extreme intakes of food chemicals. Food Chem Toxicol 44:994–1005.

Polidori A, Turpin B, Meng QY, Lee JH, Weisel C, Morandi M, et al. 2006. Fine organic particulate matter dominates indoor-generated $PM_{2.5}$ in RIOPA homes. J Expo Sci Environ Epidemiol 16:321–331.

Putt S, Shaw A, Woods A, Tyler L, AD J. 1987. Veterinary Epidemiology and Economics in Africa: A Manual for Use in the Design and Appraisal of Livestock Health Policy. Addis Ababa, Ethiopia:International Livestock Centre for Africa.

Razzaghi M, Kodell RL. 2000. Risk assessment for quantitative responses using a mixture model. Biometrics 56: 519–527.

Redner R, Walker H. 1984. Mixture densities, maximum likelihood and the EM algorithm. SIAM Rev 26:195–239.

Riederer A, Bartell S, Ryan P. 2009. Predictors of personal air concentrations of chloroform among U.S. adults in NHANES 1999–2000. J Expo Sci Environ Epidemiol 19:248–259.

Rodgers JL, Nicewander WA. 1988. Thirteen ways to look at the correlation coefficient. Am Stat 42:59–66.

Rumchev K, Brown H, Spickett J. 2007. Volatile organic compounds: do they present a risk to our health? Rev Environ Health 22:39–55.

Sax S, Bennett D, Chillrud S, Ross J, Kinney P, Spengler J. 2006. A cancer risk assessment of inner-city teenagers living in New York City and Los Angeles. Environ Health Perspect 114:1558–1566.

Schmidt T. 2006. Coping with Copulas. In: Copulas from Theory to Applications in Finance (Rank J, ed). London, UK:Risk Books.

Sexton K, Mongin SJ, Adgate JL, Pratt GC, Ramachandran G, Stock TH, et al. 2007. Estimating volatile organic compound concentrations in selected microenvironments using time-activity and personal exposure data. J Toxicol Environ Health, Part A 70:465–476.

Sherlach KS, Gorka AP, Dantzler A, Roepe PD. 2011. Quantification of perchloroethylene residues in dry-cleaned fabrics. Environ Toxicol Chem 30:2481–2487.

Singer BC, Destaillats H, Hodgson AT, Nazaroff WW. 2006. Cleaning products and air fresheners: emissions and resulting concentrations of glycol ethers and terpenoids. Indoor Air 16:179–191.

Sklar A. 1959. Fonctions de répartition à *n* dimensions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris 8:229–231.

Sneyer R. 1983. Extremes in meteorology. In: Statistical Extremes and Applications (De Oliveira JT, ed). Boston, MA:D. Reidel Publishing Co.

Staudt A. 2010. Tail Risk, Systemic Risk and Copulas. Casualty Actuarial Society E-Forum 2. Available: *www.casact.org/pubs/forum/10fforumpt2/Staudt.pdf.*

Su F-C, Jia C, Batterman S. 2012. Extreme value analyses of VOC exposures and risks: a comparison of RIOPA and NHANES datasets. Atmos Environ 62:97–106.

Su F-C, Mukherjee B, Batterman S. 2011. Trends of VOC exposures among a nationally representative sample: analysis of the NHANES 1988 through 2004 data sets. Atmos Environ 45:4858–4867.

Surman PG, Bodero J, Simpson RW. 1987. The prediction of the numbers of violations of standards and the frequency of air pollution episodes using extreme value theory. Atmos Environ 21:1843–1848.

Symanski E, Stock TH, Tee PG, Chan W. 2009. Demographic, residential, and behavioral determinants of elevated exposures to benzene, toluene, ethylbenzene, and xylenes among the U.S. Population: Results from 1999–2000 NHANES. J Toxicol Environ Health, Part A 72:903–912.

Taylor DJ, Kupper LL, Rappaport SM, Lyles RH. 2001. A mixture model for occupational exposure mean testing with a limit of detection. Biometrics 57:681–688.

Titterington DM, Smith AFM, Makov UE. 1985. Statistical analysis of finite mixture distributions. Chichester, U.K:John Wiley & Sons.

Tressou J, Crépet A, Bertail P, Feinberg MH, Leblanc JC. 2004. Probabilistic exposure assessment to food chemicals based on extreme value theory. Application to heavy metals from fish and sea products. Food Chem Toxicol 42:1349–1358.

Trivedi P, Zimmer D. 2007. Copula Modeling: An Introduction for Practitioners. Hackensack, NJ:World Scientific Publishing.

Tuia D, Kanevski M. 2008. Indoor radon distribution in Switzerland: lognormality and extreme value theory. J Environ Radioact 99:649–657.

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.

U.K. Health and Safety Executive. 2002. Guidelines for Use of Statistics for Analysis of Sample Inspection of Corrosion. United Kingdom:The Health and Safety Executive.

U.S. Environmental Protection Agency. 1986. Guidelines for the health risk assessment of chemical mixtures. Washington, DC:U.S. EPA. Available: *http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=22567#Download.*

U.S. Environmental Protection Agency. 1989. Report to Congress on Indoor Air Quality: Volume 2. EPA/400/1-89/001C. Washington, DC:U.S. EPA.

U.S. Environmental Protection Agency. 1990. Technical support document on risk assessment of chemical mixtures. Washington, DC:U.S. EPA. Available: *http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=35770.*

U.S. Environmental Protection Agency. 1993. Provisional guidance for quantitative risk assessment of polycyclic aromatic hydrocarbons. Washington, DC:U.S. EPA. Available: *http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=49732.*

U.S. Environmental Protection Agency. 2000. Supplementary guidance for conducting health risk assessment of chemical mixtures. Washington, DC: U.S EPA. Available: *http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=20533.*

U.S. Environmental Protection Agency. 2003. Framework for cumulative risk assessment. Washington, DC:U.S. EPA. Available: *www.epa.gov/raf/publications/pdfs/frmwrk_cum_risk_assmnt.pdf.*

U.S. Environmental Protection Agency. 2007a. Mobile source air toxics — regulations. Washington, DC:U.S. EPA. Available: *www.epa.gov/oms/toxics-regs.htm.*

U.S. Environmental Protection Agency. 2007b. Trichloroethylene (TCE) TEACH chemical summary. Available: *www.epa.gov/teach/chem_summ/TCE_summary.pdf.*

U.S. Environmental Protection Agency. 2008a. EPA positive matrix factorization (PMF) 3.0 fundamentals & user guide. Available: *www.epa.gov/heasd/products/pmf/EPA%20PMF%2030%20User%20Guide%20v16_092208_finalpdf.*

U.S. Environmental Protection Agency. 2008b. Regulatory determinations support document for selected contaminants from the second drinking water contaminant candidate list (ccl 2). Drinking Water Contaminant Candidate List (CCL 2). Washington, DC:U.S. EPA. Available: *www.epa.gov/ogwdw/ccl/pdfs/reg_determine2/report_ccl2-reg2_supportdocument_full.pdf.*

U.S. Environmental Protection Agency. 2009. Integrated risk information system (IRIS) glossary/acronyms and abbreviations. Washington, DC:U.S. EPA. Available: *www.epa.gov/iris/help_gloss.htm.*

U.S. Environmental Protection Agency. 2010a. Control emissions technologies — transport & dispersion of air pollutants. Available: *www.epa.gov/apti/course422/ce1.html.*

U.S. Environmental Protection Agency. 2010b. Volatile organic compounds emissions. Available: *http://cfpub.epa.gov/eroe/indexcfm?fuseaction=detailviewInd&lv=listlistByAlpha&r=209842&subtop=341.*

U.S. Environmental Protection Agency. 2011. Indoor air. Available: *http://cfpub.epa.gov/eroe/indexcfm?fuseaction=listlistBySubTopic&ch=46&s=343.*

U.S. Environmental Protection Agency. 2012a. An introduction to indoor air quality: volatile organic compounds (VOCS). Washington, DC:U.S. EPA. Available: *www.epa.gov/iaq/voc.html.*

U.S. Environmental Protection Agency. 2012b. Integrated risk information system (IRIS). Washington, DC:U.S. EPA. Available: *www.epa.gov/IRIS/index.html.*

U.S. Environmental Protection Agency. 2012c. MTBE — recommendations and actions. Washington, DC:U.S. EPA. Available: *www.epa.gov/mtbe/action.htm*.

U.S. Environmental Protection Agency. 2012d. National Ambient Air Quality Standards (NAAQS). Available: *www.epa.gov/air/criteria.html.*

U.S. Environmental Protection Agency. 2013. Basic information about disinfection byproducts in drinking water: Total trihalomethanes, haloacetic acids, bromate, and chlorite. Available: *http://water.epa.gov/drink/contaminants/basicinformation/disinfectionbyproducts.cfm.*

Wallace LA. 2001. Human exposure to volatile organic pollutants: implications for indoor air studies. Annu Rev Energy Environ 26:269–301.

Wallace LA, Pellizzari E, Leaderer B, Zelon H, Sheldon L. 1987. Emissions of volatile organic compounds from building materials and consumer products. Atmos Environ 21:385–393.

Wallace LA, Pellizzari ED, Hartwell TD, Davis V, Michael LC, Whitmore RW. 1989. The influence of personal activities on exposure to volatile organic compounds. Environ Res 50:37–55.

Wang S-W, Majeed MA, Chu P-L, Lin H-C. 2009. Characterizing relationships between personal exposures to VOCs and socioeconomic, demographic, behavioral variables. Atmos Environ 43:2296–2302.

Weisel CP, Zhang J, Turpin B, 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 Env Epid 15:123–137.

Weisel CP, Zhang J, Turpin B, Morandi MT, Colome S, Stock TH, et al. 2005c. Commentary by the Special Review Panel. In: Relationships of Indoor, Outdoor, and Personal Air (RIOPA); Part I. Collection methods and descriptive analyses. Research Report 130. Boston, MA:Health Effects Institute.

Weschler CJ. 2011. Chemistry in indoor environments: 20 years of research. Indoor Air 21:205–218.

Wolkoff P, Schneider T, Kildesø J, Degerth R, Jaroszewski M, Schunk H. 1998. Risk in cleaning: chemical and physical exposure. Sci Total Environ 215:135–156.

World Health Organization. 2012. The Determinants of Health. Geneva, Switzerland:WHO. Available: *www.who.int/hia/evidence/doh/en/.*

APPENDIX AVAILABLE ON THE WEB

Appendix A. Supplemental Information contains additional material not included in the printed report and is available on the HEI Web site *http://pubs.healtheffects.org*.

ABOUT THE AUTHORS

Stuart Batterman is a professor of Environmental Health Sciences and professor of Civil and Environmental Engineering at the University of Michigan. He completed his Ph.D. (1986) from the Massachusetts Institute of Technology. His research and teaching interests address environmental impact assessment, human exposure and health risk assessment, and environmental management. He has over three decades of experience in air quality, exposure assessment, modeling, measurements, data interpretation, laboratory and field measurements, and related analyses. As the principal investigator on this project, he provided overall supervision and direction, assisted in data interpretation and manuscript preparation, and was responsible for quality control.

Feng-Chiao Su is a post-doctoral scholar in the Department of Environmental Health Sciences at the University of Michigan. She obtained her M.S. (2006) from National Taiwan University, Taipei, Taiwan, and her Ph.D. from the University of Michigan in 2013. Her research interests focus on the nature and significance of exposures to air pollutants that occur in environmental settings. She worked as a research assistant on this project, conducted most of the analyses, and prepared reports and manuscripts. This work constitutes her Ph.D. research topic.

Shi Li earned both his M.S. (2010) and Ph.D. (2013) in biostatistics from the University of Michigan. His research interests focus on Bayesian analysis of hierarchical modeling of epidemiologic and environmental data and health-related problems. In this project, he helped with statistical analyses.

Bhramar Mukherjee is a professor in the Department of Biostatistics at the University of Michigan and was previously in the Department of Statistics at the University of

Florida. She completed her Ph.D. (2001) from Purdue University. Her principal research interests lie in Bayesian methods in environmental epidemiology and studies of gene–environment interaction. She is also interested in modeling missingness in exposure, categorical data models, Bayesian nonparametrics, and the general area of statistical inference under outcome/exposure–dependent sampling schemes. She provided guidance regarding application of statistical methods and other statistical and software support.

Chunrong Jia is an assistant professor of Environmental Health at the School of Public Health, University of Memphis, Tennessee. He completed his Ph.D. (2007) from the University of Michigan under Professor Batterman's tutelage. His research interests include theoretical work and applied laboratory and field studies in Environmental Health Sciences, including the application of gas chromatography–mass spectrometry, measurement of VOCs in indoor and ambient air, indoor air quality assessment and management, and statistical and modeling methods. Dr. Jia helped in data analyses.

OTHER PUBLICATIONS RESULTING FROM THIS RESEARCH

Su FC, Batterman S. 2014. Modeling and analysis of personal exposures to VOC mixtures using copulas. Environ Int 63:236-245.

Li S, Batterman S, Su FC, Mukherjee B. 2013. Addressing extrema and censoring in pollutant and exposure data using mixture of normal distributions. Atmos Environ 77:464–473.

Su FC, Mukherjee B, Batterman S. 2013. Determinants of personal, indoor, and outdoor VOC concentrations: an analysis of the RIOPA data. Environ Res 126:192-203.

Su FC, Jia C, Batterman S. 2012. Extreme value analyses of VOC exposures and risks: A comparison of RIOPA and NHANES datasets. Atmos Environ 62:97–106.

ABBREVIATIONS AND OTHER TERMS

1,4-DCB	1,4-dichlorobenzene
A-D	Anderson-Darling
AER	air exchange rate
AIC	Akaike information criterion
BIC	Bayesian information criterion
BTEX	benzene, toluene, ethylbenzene, xylenes
CDF	cumulative distribution function
CTC	carbon tetrachloride
DPM	Dirichlet process mixture
EM	expectation maximization
EPA	Environmental Protection Agency
ETS	environmental tobacco smoke
GEV	generalized extreme value
IQR	inter-quartile range
K-S	Kolmogorov-Smirnov
K-W	Kruskal-Wallis
LMM	linear mixed-effects model
MAE	mean absolute error
MC	methylene chloride
MDL	method detection limit
MEC	mobile examination center
MI	multiple imputation
MLE	maximum likelihood estimate
MSE	mean squared error
MTBE	methyl *tert*-butyl ether
NHANES	National Health and Nutrition Examination Survey
NHAPS	National Human Activity Pattern Survey
OEHHA	Office of Environmental Health Hazard Assessment
PAH	polycyclic aromatic hydrocarbon
PBDE	polybrominated diphenyl ether
PERC	tetrachloroethylene (or perchloroethylene)
$PM_{2.5}$	particulate matter $\leq$ 2.5 μm in aerodynamic diameter
PMF	positive matrix factorization
R^2	reduction in residual variance
RfC	reference concentration
RIOPA	Relationships of Indoor, Outdoor, and Personal Air
SBC	Schwarz Bayesian Information Criterion
TCE	trichloroethylene
TVOC	total volatile organic compound
URF	unit risk factor
VOC	volatile organic compound

CRITIQUE
Health Review Committee

HEI

Research Report 181, *Personal Exposure to Mixtures of Volatile Organic Compounds: Modeling and Further Analysis of the RIOPA Data,* S. Batterman et al.

INTRODUCTION

Exposure to various volatile organic compounds (VOCs*) has been associated with a wide range of acute and chronic health effects including asthma, respiratory diseases, liver and kidney dysfunction, neurologic impairment, and cancer (U.S. Environmental Protection Agency [EPA] 2012b). VOCs include a very large number of chemicals, some of which are on the priority lists of hazardous mobile-source and urban air toxics developed by the EPA (HEI Air Toxics Review Panel 2007.).

Assessments of exposure and adverse health outcomes are complicated by the fact that there are many indoor and outdoor sources of VOCs. In addition, certain personal activities and behaviors (e.g., smoking, driving, cooking, and use of cleaning products) contribute substantially to VOC exposure. As a result, exposure varies widely across the population and among locations, even though outdoor VOC levels are generally low. Some of the earliest studies that provided detailed information about personal, indoor, and outdoor concentrations of VOCs in the United States were the Total Exposure Assessment Methodology studies in the 1980s (e.g., Clayton et al. 1993; Wallace 1987; Wallace et al. 1991). In the mid-1990s, the European Exposure Assessment Project conducted measurements of personal exposure to VOCs, which included concurrent measurements inside homes, outside homes, and at work in four European cities (Edwards et al. 2005).

The Relationships of Indoor, Outdoor, and Personal Air (RIOPA) study was co-funded by HEI and the National Urban Air Toxic Research Center in 1999 to better define the relationships among indoor, outdoor, and personal exposure concentrations of air pollutants including particulate matter and VOCs (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 subjects who did not smoke and who lived in homes located at various distances from air pollution sources were selected. Homes close to sources were preferentially sampled. Personal, indoor, and outdoor 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, were collected using three detailed questionnaires. Household air exchange rates and geographic and meteorologic information were obtained as well.

In 2008, HEI published Request for Applications 08-1, "Relationships of Indoor, Outdoor, and Personal Air (RIOPA): Further Analyses of the RIOPA Study Data." This RFA sought proposals to perform further 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. Stuart Batterman of the University of Michigan submitted an application to HEI, in which he proposed a 2-year study to identify factors (determinants) that may affect exposure and to characterize various exposure distributions for both individual VOCs and VOC mixtures, with particular emphasis on extreme values (high exposures). The proposed study would extend previous work by Batterman and colleagues (Jia et al. 2008) in which they analyzed different exposure distributions of VOCs using the 1999–2000 data set from the National Health and Nutrition Examination Survey (NHANES). The NHANES 1999–2000 project obtained personal measurements of VOCs for approximately 650 adult subjects from a population-based sample in the United States. The analyses by Jia and associates (2008) used a stratified multistage cluster design and applied weights in order to obtain representative population-based measurements. The proposed study would use RIOPA data and, to a lesser extent, NHANES 1999–2000 data.

Dr. Batterman's 2-year study, "Modeling and Analysis of Personal Exposures to Pollutant Mixtures: Further Analysis of the RIOPA Data," began in September 2010. Total expenditures were $160,890. The draft Investigators' Report from Batterman and colleagues was received for review in February 2013. A revised report, received in August 2013, was accepted for publication in October 2013. 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.

The HEI Health Research Committee was interested in Batterman's proposal because of the potential to advance the exposure assessment and modeling field by exploring new modeling approaches. The focus on extreme values was also of interest because they represent the highest exposure scenarios and thereby indicate a possible need for specific interventions. The use of data from a population-based sample (NHANES) in addition to the RIOPA data set was also considered an advantage. The Committee recommended the study for funding and the project started in September 2010.

HEI also funded a second study under RFA 08-1, in which Dr. Patrick Ryan of the Cincinnati Children's Hospital Medical Center in Cincinnati, Ohio, would analyze the elemental composition of indoor, outdoor, and personal $PM_{2.5}$ samples (this study will be published late in 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 perspective.

AIMS AND APPROACHES

The aims of the current study were to:

- Investigate exposure determinants for individual VOCs;
- Characterize exposure to individual VOCs using full and extreme value distributions in particular; and
- Identify and characterize exposure to mixtures of VOCs using advanced statistical techniques.

Of the 16 VOCs analyzed in the original RIOPA data set, 15 were included in the current analyses (benzene, toluene, ethylbenzene, *m- & p*-xylene, *o*-xylene, methyl *tert*-butyl ether [MTBE], styrene, 1,4-dichlorobenzene [1,4-DCB], trichloroethylene [TCE], tetrachloroethylene, chloroform, carbon tetrachloride, *d*-limonene, α-pinene, and β-pinene). Methylene chloride was omitted because of inconsistent blank values.

Exposure determinants for individual VOCs in the RIOPA data set were modeled using linear mixed-effects models adjusted for clustering within cities and among individuals. Determinant models were run separately for personal, indoor, and outdoor exposure samples. The number of observations in a model ranged from 400 to 530, depending on the exact model. Because the RIOPA data set contained more than 500 possible exposure determinants, the number of determinants was reduced by considering only those that attained statistical significance ($P < 0.05$) in univariate models and when entered in forward stepwise multivariate regression models. As a result, a different set of determinants was used for each VOC and for each sample type (personal, indoor, outdoor).

To characterize the distributions of individual VOCs, Batterman and his team fitted various distribution models for each VOC using both RIOPA and NHANES data. Distribution models have different properties related to how they treat low and high (extreme) values, right-skewing of data, and the occurrence of multiple modes in the data — all features that are often present in VOC exposure data. The primary focus in these analyses was on characterizing extreme values using generalized extreme value (GEV) models, which use shape, location, and scale parameters to fit the tails of a distribution. First, Batterman and his colleagues calculated average values of the two personal samples for each of the 239 adult RIOPA subjects with complete data (only one sample per person was collected in the NHANES data set). Then they selected for further analysis only those subjects with high average exposures, defined as the subjects in the top 5% and top 10% of the exposure distribution. As a consequence, extreme value analyses were based on relatively small numbers of subjects: 12 or 24 using RIOPA data and 32 or 64 using NHANES data. For three individual VOCs (chloroform, 1,4-DCB, and styrene) Batterman and colleagues fit two types of mixture models: a finite mixture of normal distributions, and a Dirichlet process mixture (DPM) of normal distributions. The finite mixture approach required that the number of clusters be preselected from a limited set of options (two clusters for chloroform, four for 1,4-DCB, and three for styrene were selected), whereas the DPM adaptively determined the number of clusters based on the data.

The two steps taken by the investigators to address the distribution of mixtures were to identify mixtures of VOCs in the RIOPA data set and then estimate their joint distributions and dependency structures. They used two approaches to identify mixtures: positive matrix factorization and toxicologic modes of action. The first is a statistical approach similar to factor analysis and is based on actual correlations among VOCs in the data set. The latter approach reflects common biochemical pathways and health outcomes associated with certain VOCs based on prior knowledge. In the second step, copulas (a class of probability models; e.g., t, Gumbel) were used to characterize the distribution of and dependencies among different VOC components of the mixtures previously identified. Unlike standard multivariate models (such as multivariate normal or lognormal), copulas are able to distinguish between the dependencies (or correlations among components in the mixture) and the marginal distribution of components in the mixture. This is an advantage because standard multivariate models may not fit data well, especially for extreme values.

In addition, copulas are flexible in many statistical properties; for example, they allow the dependency structure to be weighted in different manners. For the copula analyses, the RIOPA data were restricted to personal measurements from adults during the first visit in order to avoid effects of clustering within individuals, which would complicate the analyses.

The performance of the different exposure distribution models was evaluated using goodness-of-fit statistics and simulations. The investigators also calculated lifetime cancer risks using standard risk assessment approaches with unit risk factors for cancer (increased risk of getting cancer per unit increase of exposure; obtained primarily from the EPA) and exposure data that were based on either measured or modeled values. The investigators also fitted lognormal distributions as a benchmark for comparison purposes.

BRIEF SUMMARY OF RESULTS

ANALYSES OF DETERMINANTS

- Personal exposure. Several significant determinants of mean personal VOC exposure were, among others, an increase in wind speed, an increase in home air exchange rate, and having more rooms in the home were each associated with a decrease in VOC exposure. The presence of an attached garage, the subject pumping gas, and other family members showering were each associated with an increase in VOC exposure. Living in Houston was often associated with a higher level of VOC exposure than was living in the other two cities (Los Angeles and Elizabeth).
- Indoor exposure. Determinants similar to those for personal exposure were identified for indoor exposure to VOCs; the similarity was largely due to the fact that RIOPA participants spent the vast majority of their time at home (on average about 90%).
- Outdoor exposure. In contrast, only a few significant determinants were identified for outdoor VOC exposure, namely city and weather characteristics (wind speed, outdoor temperature, and ambient relative humidity).

DISTRIBUTION FITTING FOR INDIVIDUAL VOCS

- The commonly applied lognormal distributions typically underestimated high personal exposure to VOCs and associated cancer risks.
- Other distributions, such as the GEV, that are tailored to fit extreme values performed somewhat better than lognormal distributions in estimating high personal exposures and associated cancer risks for individual VOCs. The results of extreme value modeling were inconsistent across analytic approaches in the NHANES data set. This was probably because of the complicated sample-weighting scheme needed for NHANES analyses in order to be a representative sample.
- The mixture models, both the finite mixture of normal distributions and the DPM, captured the observed concentration data well and generally performed better than the single normal distributions.

DISTRIBUTION FITTING FOR MIXTURES OF VOCs

- Four mixtures were identified by positive matrix factorization using personal measurements: two mixtures (chlorinated solvents, and cleaning products and odorants) represented indoor sources, and two mixtures (gasoline, and vehicle exhaust) represented outdoor sources. The indoor sources (in which 1,4-DCB and *d*-limonene dominated the two mixtures) explained most of the variation in total VOC exposure. Three mixtures were identified by toxicologic mode of action to represent VOCs associated with hematopoietic cancers and liver and renal tumors.
- As in the results for individual VOCs, distributions other than the multivariate lognormal distribution, such as the t and Gumbel copulas, performed somewhat better in estimating exposures and associated cancer risks of VOC mixtures.

THE HEALTH REVIEW COMMITTEE'S EVALUATION

In its independent review of the study by Batterman and colleagues, the HEI Health Review Committee concluded that the study adds to a better understanding and characterization of VOC exposures. The members noted that the study was well conceived and conducted, and the investigators were thorough in the analyses they performed. In particular, the analyses of exposure determinants represent a novel and useful contribution to the literature. Exposure determinants that were identified — including city, wind speed, home air exchange rate, number of rooms in the home, whether the home had an attached garage, whether the subject pumped gas, and whether other family members were showering — largely agreed with reports from previous studies (D'Souza et al. 2009; Sexton et al. 2007; Symanski et al. 2009; Wang et al. 2009).

However, the Committee thought that the practical applicability of the determinant analyses was hampered to some

extent because the investigators used a different set of possible determinants for each VOC — even when they came from the same source, such as the gasoline-related VOCS — and for each sample type (personal, indoor, outdoor); this meant that the exact magnitude of the effect of an exposure determinant could not be readily estimated. For example, the effect of a certain determinant could not be readily compared across various gasoline-related VOCs because it might have been included for one VOC but excluded for another (e.g., the variable "time spent at home" was included for personal benzene exposure but not for personal toluene exposure). As a result, the sets of determinants used to adjust the coefficients of a particular determinant differed among pollutants. For the same reasons, one cannot readily compare the effects of a certain determinant for a given VOC across sample types. The Committee thought that the modeling decisions were primarily driven by prediction criteria (i.e., how well the final determinant model predicted the exposure concentrations), rather than by estimating and understanding common or major exposure determinants.

In addition, the Committee did not agree with the investigators' treatment of values below the limit of detection (LOD) in the determinant analyses. The method adopted by the investigators of replacing all values below the LOD for a particular VOC with one single value (½ of the LOD), though commonly used, can cause problems if the number of observations below the LOD is considerable. For example, for six outdoor VOCs (styrene, chloroform, *d*-limonene, TCE, α-pinene, and β-pinene), 75% or more of the samples were below the LOD in the RIOPA data set (see Table 1 in the Investigators' Report). Some studies have shown bias in estimates when 10% or more of the measurements were below the LOD and were replaced with a single value (see, for example, Lubin et al. 2004). The Committee noted that alternative methods could have been applied to treat values below the LOD, such as regression on order statistics and nonparametric techniques, both of which have been shown to perform better (Antweiler and Taylor 2008). In view of the Committee's concern about the way values below LOD were treated in the current study, caution should be exercised in interpreting the results of the determinant analyses for VOCs with high proportions of such values; this was especially true for outdoor samples, for which many values were below the LOD.

The Committee appreciated that two approaches were applied to identify and distinguish different VOC mixtures, namely an exposure-based approach (positive matrix factorization) and a risk-based approach (toxicologic mode of action). Considering the interactions and correlations among components in complex atmospheres, both the research and policy communities are increasingly interested in evaluating exposure mixtures and their determinants in addition to evaluating single compounds (Johns et al. 2012; U.S. EPA 2012a). Although the investigators conducted some preliminary analyses to evaluate exposure determinants of VOC mixtures (see Table A.15 in Appendix A of the Investigators' Report, which is available on the HEI Web site), this was not as extensive as the determinant analyses for individual VOCs. For example, the analyses did not take into account dependencies among components in the mixture. Thus, the Committee thought that future analyses should pursue exposure determinants of VOC mixtures more fully.

For characterizing individual VOCs, the investigators demonstrated that distributions other than the commonly applied lognormal models, such as the GEV model, performed somewhat better in estimating high personal exposures and associated cancer risks. Similar conclusions can be drawn regarding the characterization of VOC mixtures, in which copulas performed somewhat better than multivariate lognormal models. In further distribution fitting for individual VOCs, the Committee appreciated that the investigators applied mixture models, which allowed entire VOC distributions to be estimated because concentrations below extreme values also affect total population risk. The mixture models performed generally better than a single lognormal distribution, particularly for characterizing right-skewed multi-modal distributions with a large fraction of values below the LOD. However, the Committee felt that more research is needed to determine the broad applicability of mixture models for exposure and risk modeling.

Although the Committee considered the statistical analyses focused on characterizing extreme values to be interesting, the investigators' interpretation of the results was thought to be problematic for several reasons. First, because the number of observations below the LOD was sometimes considerable (see above), the commonly used but simplistic treatment of those values could have affected the results. The Committee agreed with the investigators that this approach did not affect the extreme value analyses directly. However, they thought that it did affect the comparison of those distributions with conventional lognormal distributions, because the lognormal distributions were fitted using the full data set, as opposed to the top 10% or top 5% of the exposure distribution that was used for the extreme value analyses. In contrast, their treatment of values below the LOD had less impact on the mixture distribution fit. Second, in the analyses of extreme values, the investigators deleted six observations that they considered to be outliers and other influential values. Although the investigators gave an explanation as to why they thought those

observations might not reflect a subject's exposure (e.g., placing the personal sampler near perfume bottles at night), the Committee thought these deletions were not adequately justified in the absence of specific information about such irregularities in the subjects' questionnaires. They concluded that extreme value analyses are by definition based on small numbers of extreme observations, and thus one should be very cautious in deleting outliers.

Finally, the Committee suggested caution in generalizing the authors' interpretations of the distribution characterizations because the RIOPA data set — which underlies the majority of analyses — was a convenience sample, rather than a representative population-based sample, as was NHANES. The RIOPA data set includes only three U.S. cities, and homes close to sources were preferentially sampled. In addition, only households in which no one smoked were selected, and participants were older (average age 45 years; 24% 60 years or older) and were predominantly female (75%) and unemployed (53%) (Turpin et al. 2007; Weisel et al. 2005). RIOPA differed from NHANES in exposure-sampling protocols, target populations, and locations studied in the United States. Some of these differences are evident in the marginal distributions in the two studies — when comparing personal VOC measurements in RIOPA with those from NHANES, values from RIOPA were considerably lower for all measured VOCs except MTBE and 1,4-DCB, which were higher (see Table 1 in the Investigators' Report).

The Committee also believed that the applicability in air pollution research of the methods for extreme value analyses developed in this study may be limited and further research is needed. In large data sets, such as NHANES, one can read the values of interest (e.g., percentage of data above some high value) more or less directly from the data, so not much is to be gained by using an estimate based on parametric distribution. In small data sets, fitting an extreme value distribution is likely to lead to imprecise estimates because of the small sample size. In addition, new and improved methods are needed before extreme value analyses can be further applied in data sets that require reweighting for proper inference, as was the case for the NHANES data. Furthermore, for risk assessment purposes, the Committee thought that other sources of uncertainty (such as extrapolation across species and dose levels) associated with calculating lifetime cancer risks related to VOC exposure would easily overwhelm the uncertainty associated with using one particular exposure distribution rather than another. However, the latter was clearly well beyond the scope of this research.

SUMMARY AND CONCLUSIONS

In summary, Dr. Batterman and colleagues used data from the RIOPA study to investigate determinants of exposure to VOCs, such as housing characteristics and personal activities. In addition, using data from RIOPA and, to a lesser extent, from NHANES, they characterized various distribution models for individual VOCs and VOC mixtures. In its independent review of the study, the HEI Review Committee noted that the study was well conceived and conducted, and that the analyses of exposure determinants in particular represent a novel and useful contribution to the literature. In the distribution fitting for individual VOCs, the Committee appreciated the application of mixture models allowing estimation of entire VOC distributions. The statistical analyses focused on characterizing extreme values were considered interesting; however, the investigators' interpretation of the results was thought to be problematic because of the treatment of values below the LOD, the deletion of outlier values, and the use of a convenience sample. The applicability in air pollution research of the methods for extreme value analyses that were developed in this study may be limited and further research is needed.

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 Hanna Boogaard for her assistance in preparing its Critique, to Virgi Hepner for science editing of this Report and its Critique, and to Hope Green, Fred Howe, Carol Moyer, and Ruth Shaw for their roles in preparing this Research Report for publication.

REFERENCES

Antweiler RC, Taylor HE. 2008. Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets: I. Summary statistics. Environ Sci Technol 42:3732–3738.

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 Env Epid 3:227–250.

D'Souza JC, Jia C, Mukherjee B, Batterman S. 2009. Ethnicity, housing and personal factors as determinants of VOC exposures. Atmos Environ 43:2884–2892.

Edwards RD, Schweizer C, Jantunen M, Lai HK, Bayer-Oglesby L, Katsouyanni K, et al. 2005. Personal exposures to VOC in the upper end of the distribution—relationships to indoor, outdoor and workplace concentrations. Atmos Environ 39:2299–2307.

HEI Air Toxics Review Panel. 2007. Mobile-Source Air Toxics: A Critical Review of the Literature on Exposure and Health Effects. Special Report 16. Boston, MA:Health Effects Institute.

Jia C, D'Souza J, Batterman S. 2008. Distributions of personal VOC exposures: a population-based analysis. Environ Int 34:922–931.

Johns DO, Stanek LW, Walker K, Benromdhane S, Hubbell B, Ross M, et al. 2012. Practical advancement of multipollutant scientific and risk assessment approaches for ambient air pollution. Environ Health Perspect 120:1238–1242.

Lubin JH, Colt JS, Camann D, Davis S, Cerhan JR, Severson RK, et al. 2004. Epidemiologic evaluation of measurement data in the presence of detection limits. Environ Health Perspect 112:1691–1696.

Sexton K, Mongin SJ, Adgate JL, Pratt GC, Ramachandran G, Stock TH, et al. 2007. Estimating volatile organic compound concentrations in selected microenvironments using time–activity and personal exposure data. J Toxicol Environ Health A 70:465–476.

Symanski E, Stock TH, Tee PG, Chan W. 2009. Demographic, residential, and behavioral determinants of elevated exposures to benzene, toluene, ethylbenzene, and xylenes among the U.S. population: Results from 1999–2000 NHANES. J Toxicol Environ Health A 72:915–924.

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.

U.S. Environmental Protection Agency. 2012a. Air, Climate, and Energy Strategic Research Action Plan 2012–2016. Available: *www.epa.gov/research/docs/strapace2012.pdf.*

U.S. Environmental Protection Agency. 2012b. Integrated Risk Information System (IRIS). Available: *www.epa.gov/iris/.*

Wallace LA. 1987. The Total Exposure Assessment Methodology (TEAM) Study: Summary and Analysis: Volume 1. EPA/600/6-87/002a. Washington, DC:U.S. Environmental Protection Agency. Available: *http://exposurescience.org/pub/reports/TEAM_Study_book_1987.pdf.*

Wallace L, Nelson W, Ziegenfus R, Pellizzari E, Michael L, Whitmore R, et al. 1991. The Los Angeles TEAM study: personal exposures, indoor–outdoor air concentrations, and breath concentrations of 25 volatile organic compounds. J Expo Anal Env Epid 1:157–192.

Wang S-W, Majeed MA, Chu P-L, Lin H-C. 2009. Characterizing relationships between personal exposures to VOCs and socioeconomic, demographic, behavioral variables. Atmos Environ 43:2296–2302.

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

RELATED HEI PUBLICATIONS: AIR TOXICS, MIXTURES, AND METHODS

Number	Title	Principal Investigator	Date*
Research Reports			
172	Potential Air Toxics Hot Spots in Truck Terminals and Cabs	T.J. Smith	2012
160	Personal and Ambient Exposures to Air Toxics in Camden, New Jersey	P.J. Lioy	2011
158	Air Toxics Exposure from Vehicle Emissions at a U.S. Border Crossing: Buffalo Peace Bridge Study	J.D. Spengler	2011
156	Concentrations of Air Toxics in Motor Vehicle–Dominated Environments	E.M. Fujita	2011
150	Mutagenicity of Stereochemical Configurations of 1,3-Butadiene Epoxy Metabolites in Human Cells	R.Q. Meng	2010
143	Measurement and Modeling of Exposure to Selected Air Toxics for Health Effects Studies and Verification by Biomarkers	R.M. Harrison	2009
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
116	Biomarkers in Czech Workers Exposed to 1,3-Butadiene: A Transitional Epidemiologic Study	R.J. Albertini	2003
115	Validation and Evaluation of Biomarkers in Workers Exposed to Benzene in China	Q. Qu	2003
HEI Special Reports			
17	Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects		2010
16	Mobile-Source Air Toxics: A Critical Review of the Literature on Exposure and Health Effects		2007
HEI Communications			
4	Theoretical Approaches to Analyzing Complex Mixtures		1996

* *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*

Archibald Cox, Founding Chair *1980–2001*

Donald Kennedy, Vice Chair Emeritus *Editor-in-Chief Emeritus,* Science*; President Emeritus and Bing Professor of Biological Sciences, Stanford University*

Health Research Committee

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

David Christiani *Elkan Blout Professor of Environmental Genetics, Harvard School of Public Health*

Francesca Dominici *Professor of Biostatistics and Senior Associate Dean for Research, Harvard 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*

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*

Richard L. Smith *Director, Statistical and Applied Mathematical Sciences Institute, University of North Carolina–Chapel Hill*

James A. Swenberg *Kenan Distinguished Professor of Environmental Sciences, Department of Environmental Sciences and Engineering, University of North Carolina–Chapel Hill*

HEI BOARD, COMMITTEES, and STAFF

Health Review Committee

Homer A. Boushey, Chair *Professor of Medicine, Department of Medicine, University of California–San Francisco*

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*

Stephanie London *Senior Investigator, Epidemiology Branch, National Institute of Environmental Health Sciences*

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

Armistead Russell *Howard T. Tellepsen Chair of Civil and Environmental Engineering, School of Civil and Environmental Engineering, Georgia Institute of Technology*

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*

Barbara Gale *Director of Publications*

Jacqueline C. Rutledge *Director of Finance and Administration*

April Rieger *Corporate Secretary*

Kate Adams *Senior Scientist*

Hanna Boogaard *Staff Scientist*

Aaron J. Cohen *Principal Scientist*

Maria G. Costantini *Principal Scientist*

Philip J. DeMarco *Compliance Manager*

Hope Green *Editorial Assistant*

L. Virgi Hepner *Senior Science Editor*

Anny Luu *Administrative Assistant*

Francine Marmenout *Senior Executive Assistant*

Nicholas Moustakas *Policy Associate*

Hilary Selby Polk *Senior Science Editor*

Jacqueline Presedo *Research Assistant*

Evan Rosenberg *Staff Accountant*

Margarita Shablya *Science Administrative Assistant*

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 181
June 2014